

11-1-2004

Multivariate And Multistrata Nonparametric Tests: The NonParametric Combination Method

Livio Corain

University of Padova, livio.corain@unipd.it

Luigi Salmaso

University of Padova

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Corain, Livio and Salmaso, Luigi (2004) "Multivariate And Multistrata Nonparametric Tests: The NonParametric Combination Method," *Journal of Modern Applied Statistical Methods*: Vol. 3 : Iss. 2 , Article 16.

DOI: [10.22237/jmasm/1099268160](https://doi.org/10.22237/jmasm/1099268160)

Multivariate And Multistrata Nonparametric Tests: The NonParametric Combination Method

Livio Corain Luigi Salmaso
Department of Management and Engineering
University of Padova

Researchers and practitioners in many scientific disciplines and industrial fields are often faced with complex problems when dealing with comparisons between two or more groups using classical parametric methods. The data arising from real problems rarely are in agreement with stringent parametric assumptions. The NonParametric Combination (NPC) methodology frees the researcher from stringent assumptions of parametric methods and allows a more flexible analysis, both in terms of specification of multivariate hypotheses and in terms of the nature of the variables involved in the analysis. An outline of NPC methodology is given, along with case studies.

Key words: Multivariate Inference, nonparametric combination, permutation test

Introduction

From a methodological point of view, when comparing NonParametric Combination (NPC) Test methodology to unconditional parametric testing it should be remembered that the latter suffers from the constraint that it is appropriate and applicable only when a set of conditions concerned with the likelihood model are all satisfied (Pesarin, 2002). Only if all conditions are jointly satisfied is the extension of inferential results to the population possible and appropriate. Otherwise when these conditions fail, especially if selection-bias procedures are used for data collection processes as in most real applications, most parametric inferential extensions are generally wrong or misleading.

Moreover, when all the above conditions are satisfied, in practice other assumptions regarding the validity of the parametric method, such as normality, are rarely

satisfied. Consequent inferences, when not improper, are necessarily approximated and their approximations are often difficult to assess. However, there are circumstances in which conditional testing procedures may be unavoidable as in the case of multivariate problems, when some variables are categorical and others are quantitative or when multivariate alternatives are subjected to order restrictions (for a detailed list of these circumstances see Pesarin, 2002). A short outline of the implementation of NPC methodology follows.

Brief overview of the NPC methodology

Without loss of generality, let us refer to a one-way MANOVA layout. The data structure is defined as follows. Denote by \mathbf{X} an $(n \times k)$ data set:

$$\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_j, \dots, \mathbf{X}_c] = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_k],$$

where $\mathbf{X}_j, j=1, \dots, C, (C \geq 2)$ represents the j -th $n_j \times k$ group, $n_j \geq 2$ and $\sum_j n_j = n$, and \mathbf{X}_i is the i -th univariate aspect of $\mathbf{X}, i=1, \dots, k (k \geq 1)$; moreover let \mathbf{X}_{ji} represent the i -th univariate aspect of \mathbf{X}_j .

In the context of NonParametric Combination (NPC) of Dependent Permutation Tests a set of conditions should be jointly satisfied:

Livio Corain is a doctoral candidate at the University of Padova. The Email address is: livio.corain@unipd.it. Luigi Salmaso is an Associate Professor of statistics at the University of Padova. His research interests include multivariate analysis, design of experiments, nonlinear regression, and nonparametric statistics.

1) suppose that for $\mathbf{X}=[\mathbf{X}_1, \dots, \mathbf{X}_C]'$ an appropriate probabilistic k -dimensional distribution structure P exists, $P_j \in F, j=1, \dots, C$, belonging to a (possibly non-specified) family F of non-degenerate probability distributions.

2) the null hypothesis H_0 states the equality in distribution of the multivariate distribution of the k variables in all C groups:

$$H_0 : [P_1 = \dots = P_C] = \left[\mathbf{X}_1^d = \dots = \mathbf{X}_C^d \right].$$

Null hypothesis H_0 implies the exchangeability of the individual data vector with respect to the groups. Moreover H_0 is supposed to be properly decomposed into k sub-hypotheses $H_{0i}, i=1, \dots, k$, each appropriate for partial (univariate) aspects, thus H_0 (multivariate) is true if all the H_{0i} (univariate) are jointly true:

$$H_0 : \left[\bigcap_{i=1}^k X_{1i}^d = \dots = X_{Ci}^d \right] = \left[\bigcap_{i=1}^k H_{0i} \right].$$

H_0 is called the *global* or *overall null hypothesis*, and $H_{0i}, i=1, \dots, k$, are called the *partial null hypotheses*.

3) The alternative hypothesis H_1 is represented by the union of partial H_{1i} sub-alternatives:

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

so H_1 is true if at least one of sub-alternatives is true. In this context, H_1 is called the *global* or *overall alternative*, and $H_{1i}, i=1, \dots, k$, are called the *partial alternatives*.

4) let $\mathbf{T}=\mathbf{T}(\mathbf{X})$ represent a k -dimensional vector of test statistics, $k \geq 1$, whose components $T_i=T_i(X_i), i=1, \dots, k$, represent the partial univariate and non-degenerate *partial test* appropriate for testing the sub-hypothesis H_{0i} against H_{1i} . Without loss of generality, all partial tests are assumed to be marginally unbiased, consistent and significant for large values (for more details, see Pesarin, 2001).

At this point, in order to test the global null hypothesis H_0 , the key idea comes from the partial (univariate) tests which are focused on k partial aspects, and then, combining them with an appropriate combining function, from a global (multivariate) test which is referred to as the global null hypothesis.

However, before introducing the combination methodology, we should observe that in most real problems, when the sample size is great enough, there is a clash over the problem of computational difficulties in calculating the conditional permutation space. This means it is not possible to calculate the exact p -value of observed statistic T_{i0} . This is overcome by using the CMCP (Conditional Monte Carlo Procedure).

The CMCP on the pooled data set X is a random simulation of all possible permutations of the same data under H_0 (for more details refer to Pesarin, 2001). Hence, in order to obtain an estimate of the permutation distribution under H_0 of all test statistics, a CMCP can be used. Every resampling without replacement \mathbf{X}^* from the pooled data set \mathbf{X} actually consists of a random attribution of individual data vectors to the C samples. In every \mathbf{X}_r^* resampling, $r=1, \dots, B$, the k partial tests are calculated to obtain the set of values $[T_{ir}^*=T(\mathbf{X}_{ir}^*), i=1, \dots, k; r=1, \dots, B]$, the B independent random resamplings.

It should be emphasized that CMCP only considers permutations of individual data vectors, so that all underlying dependence relations that are present in the component variables are preserved. From this point of view, the CMCP is essentially a multivariate procedure.

The two-phases algorithm

Once the hypothesis system is defined and an appropriate set of k statistics $T_i=T_i(X_i), i=1, \dots, k$, the natural way to test the global null hypothesis consists of two sequential phases:

- 1) performing k partial tests;
- 2) combining them in a second-order global test.

It should be pointed out that this two-step procedure can be characterized by several

intermediate combinations if there is a more complex data configuration where the most interesting cases are given by testing in presence of stratification, closed-testing, multi aspect testing and repeated measures.

Assuming that the partial tests have real values and are marginally unbiased, consistent and significant for large values, then the first phase consists in:

1a. calculating the k -vector of observed values of test statistics \mathbf{T}_0 :

$$\mathbf{T}_0 = \mathbf{T}(\mathbf{X}) = [T_{i0}(X_i), i=1, \dots, k];$$

1b. considering a data permutation of \mathbf{X} by a random resampling \mathbf{X}_r^* , in order to randomly assign every individual data vector to a proper group and then calculate the vector statistics \mathbf{T}_r^* :

$$\mathbf{T}_r^* = \mathbf{T}_r^*(\mathbf{X}_r^*) = [T_{ir}^*(X_{ir}^*), i=1, \dots, k];$$

1c. carrying out B independent repetitions of step 1.b; the result is a set \mathbf{T}^* of $B \times k$ CMC

$$\mathbf{T}^* = [\mathbf{T}_r^*, r=1, \dots, B] = [\mathbf{T}_1^*, \dots, \mathbf{T}_r^*, \dots, \mathbf{T}_B^*]'$$

is thus a random sampling from the permutation k -variate distribution of vector test statistics \mathbf{T} ;

1d. the k -variate EDF (Empirical Distribution Function) $\hat{F}_B(\mathbf{z} | \mathbf{X})$

$$\hat{F}_B(\mathbf{z} | \mathbf{X}) = [1/2 + \sum_r \mathbf{I}(\mathbf{T}_r^* \leq \mathbf{z})] / (B+1), \forall \mathbf{z} \in \mathbb{R}^k,$$

where $\mathbf{I}(\cdot)$ is the indicator function, and gives an estimate of the corresponding k -dimensional permutation distribution $F_B(\mathbf{z} | \mathbf{X})$ of \mathbf{T} . Moreover

$$\hat{L}_i(z | \mathbf{X}) = [1/2 + \sum_r \mathbf{I}(T_{ir}^* \geq z)] / (B+1), i=1, \dots, k,$$

gives an estimate $\forall z \in \mathbb{R}^1$ of the marginal permutation *significance level function*

$$L_i(z | \mathbf{X}) = \Pr \{T_i^* \geq z | \mathbf{X}\};$$

thus

$$\hat{L}_i(T_{i0} | \mathbf{X}) = \hat{\lambda}_i$$

gives an estimate of the marginal p -value $\lambda_i = \Pr \{T_i^* \geq T_{i0} | \mathbf{X}\}$ relative to test $T_i, i=1, \dots, k$. All these are unbiased and consistent estimates of corresponding true values;

1e. if $\hat{\lambda}_i < \alpha$, the null hypothesis H_{0i} relating to the i -th variable is rejected at the significance level α .

The second phase, based on a nonparametric combination of the dependent tests previously obtained, consists in the following steps:

2a. the combined observed value of the second-order test is evaluated through the same CMC results as the first phase, and is given by:

$$\mathbf{T}_0'' = \psi(\hat{\lambda}_1, \dots, \hat{\lambda}_k);$$

2b. the r -th combined value of vector statistics (step 1.d) is then calculated by:

$$\mathbf{T}_r'' = \psi(\hat{\lambda}_{1r}^*, \dots, \hat{\lambda}_{kr}^*),$$

where $\hat{\lambda}_{ir}^* = \hat{L}_i(T_{ir}^* | \mathbf{X}), i=1, \dots, k, r=1, \dots, B$;

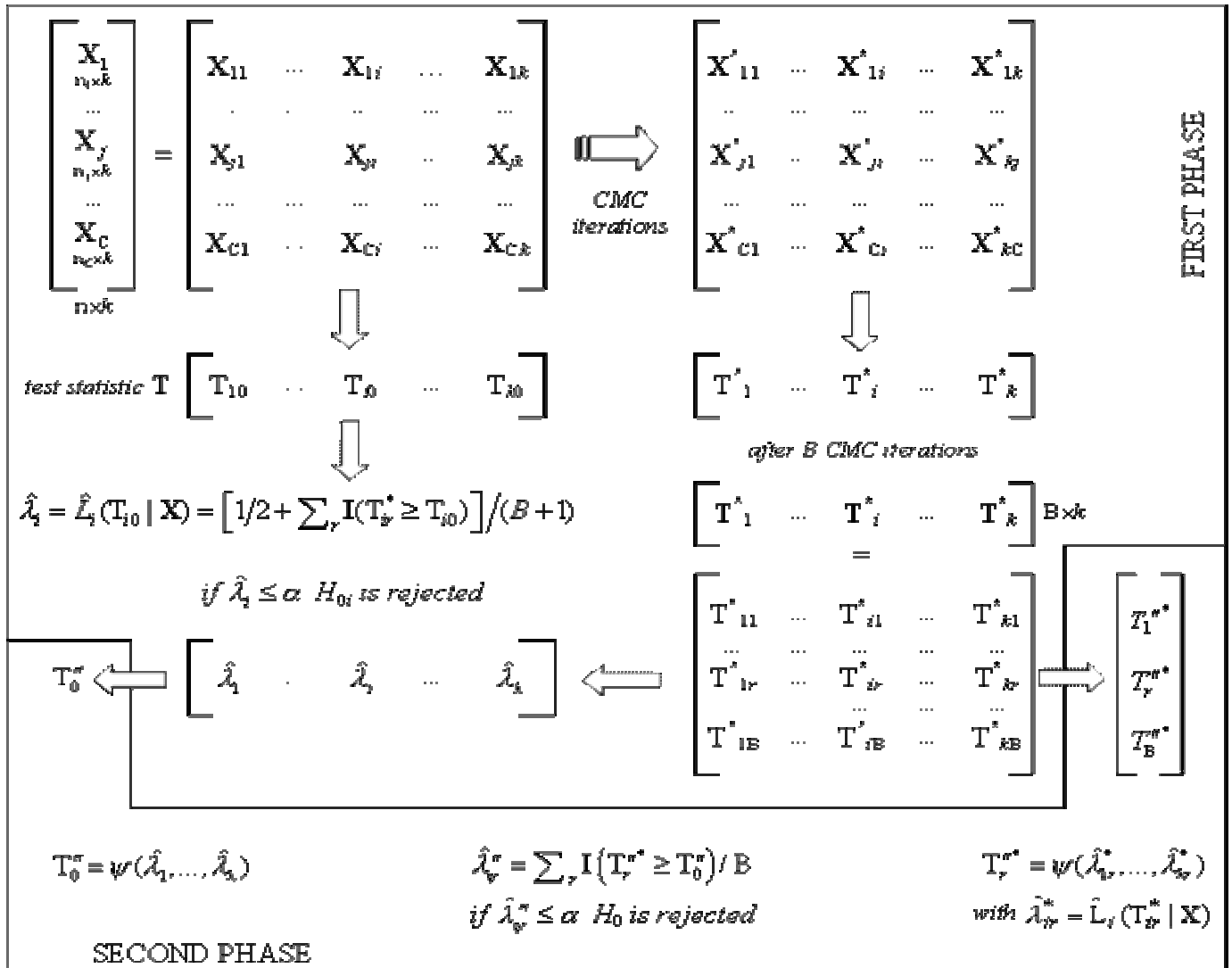
2c. hence, the p -value of combined test \mathbf{T}'' is estimated as:

$$\lambda_{\psi}'' = \sum_r \mathbf{I}(\mathbf{T}_r'' \geq \mathbf{T}_0'') / B;$$

2d. if $\lambda_{\psi}'' \leq \alpha$, the global null hypothesis H_0 is rejected at significant level α ; where ψ is an appropriate combining function.

Figure 1 summarizes graphically the complete framework of NPC solution. Remember that, in order to preserve the underlying dependence relations among variables, permutations must always be carried out on individual data vectors, so that all component variables and partial tests must be jointly analyzed.

Figure 1. Graphical description of two-phase NPC solution.



It can be seen that under the general null hypothesis the CMC procedure allows a consistent estimation of the permutation distributions, both marginal and combined, of the k partial tests. In the nonparametric combination procedure, Fisher's combination function is usually considered, principally for its good properties which are both finite and asymptotic (Pesarin, 2001). Of course, if it were considered appropriate, it would be possible to take into consideration any other combining function (Folks, 1984; Pesarin, 2001). The com-

bined test is unbiased and consistent; it also has interesting asymptotic properties.

A general characterization of the class of combining functions is given by the following three main features for the combining function ψ :

- a) it must be non-increasing in each argument:

$$\psi(\dots, \lambda_i, \dots) \geq \psi(\dots, \lambda'_i, \dots) \text{ if } \lambda_i < \lambda'_i, \\ i \in \{1, \dots, k\};$$

- b) it must attain its supreme value $\bar{\psi}$, possibly non finite, even when only one argument reaches zero:

$$\psi(\dots, \lambda_i, \dots) \rightarrow \bar{\psi} \text{ if } \lambda_i \rightarrow 0, \\ i \in \{1, \dots, k\};$$

- c) $\forall \alpha > 0$, the critical value of every ψ is assumed to be finite and strictly smaller than the supreme value:

$$T_\alpha'' < \bar{\psi}.$$

The above properties define the class C of combining functions. Some of the functions most often used to combine independent tests (e.g., Fisher, Lancaster, Liptak, Tippett, Mahalanobis) are included in this class. If in the overall analysis distinguishing the importance of partial tests by using appropriate weights opportunely fixed: $w_i \geq 0$, $i = 1, \dots, k$, with at least one strong inequality is considered more suitable, then the combined test using the Fisher combination is:

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

Nested combinations

Suppose that the k variables describing the testing problem can be classified into $m_1 < k$ classes according to some meaningful criteria. Moreover, the m_1 classes could themselves be put together in a further grouping, obtaining $m_2 < m_1$ classes and so on. After $T < k$ steps, this nested classification rule leads to only one final class which includes all variables. It is clear that in such a situation, before carrying out the global test by nonparametric combination of k partial tests, it is more appropriate to introduce T intermediate combination phases that reflect the meaningful classification rules.

This nested procedure can be represented by a graph (Figure 2) in which, from top to bottom, each node indicates a partial test (the corresponding p -value is displayed), and each arch indicates a nonparametric combination into a higher order test. Note that it is not necessary for all partial tests to be involved in

every phase. Some could be included after a given phase.

Features of Software *NPC Test 2.0*

NPC Test 2.0 (see details online at www.methodologica.it) implements completely NPC methodology offering both flexibility and a user-friendly interface. The available multivariate analyses are *Two or C Samples with Dependent Variables* (highlighting the dependence among responses) and *Two or C Samples with Repeated Measures*.

Data sets may be either created and manipulated inside the program on a normal spreadsheet or may be pasted or directly imported from the most utilized formats (see Figure 3).

The reader should be reminded that in NPC Test there are no limitations in the number of observations with respect to the number of variables, i.e., there are no problems regarding a possible lack of degrees of freedom. It is possible to consider one or more stratification factors in order to solve problems with extremely complex experimental designs.

All kinds of variables are dealt with (numeric or continuous, nominal, ordered categorical, or binary; see Figure 4) each provided with an appropriate set of test statistics should they also be suitable for an effective managing of missing values.

The testing procedure is easily performed by following a three step wizard where at first the user is requested to define the sample and the strata, then he has to specify the variables under testing and the test statistic (Figure 5) and finally he has to select a suitable Nonparametric Combination in to perform the global test. Four different functions for combining nonparametrically the partial tests are available: Fisher, Liptak, Tippet and Direct (Figure 6).

We highlight that every partial alternative hypothesis may be specified as being either one or two tailed. Moreover there is the possibility of testing both aspect X and X^2 of the same variable so multi-aspect testing (Pesarin, 2001) is also obtainable. Finally all performed tests are kept in an effective report that can easily be integrated and customised by means of an efficient text editor (Figure 7, 8).

Figure 2. Graphical representation of nested combinations.

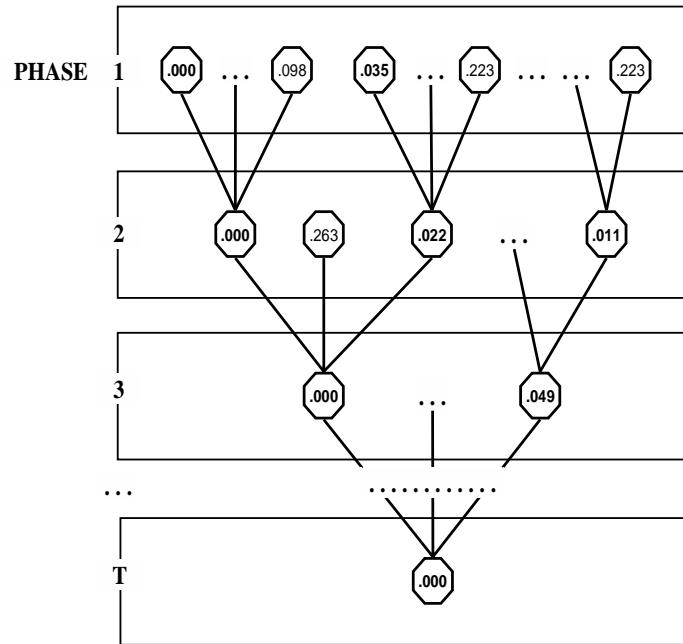


Figure 3. NPC Test's interface for data management.

NPC Test - [C:\Programmi\NPCTest\Examples\Setig.npc]

File Edit Test Help About

	group	strata	D	U	L	R
1	SPECIFIC	1	1	0	11	0
2	SPECIFIC	1	0	0	?	1
3	SPECIFIC	1	0	1	?	0
4	SPECIFIC	1	0	1	?	0
5	SPECIFIC	1	1	1	19	0
6	SPECIFIC	1	0	1	?	0
7	SPECIFIC	1	0	0	?	1
8	SPECIFIC	1	0	0	?	1
9	SPECIFIC	1	0	0	39	1
10	SPECIFIC	1	0	0	?	1
11	SPECIFIC	1	0	1	25	1
12	SPECIFIC	1	0	0	?	1
13	SPECIFIC	1	1	0	10	0
14	SPECIFIC	1	1	1	19	0
15	SPECIFIC	1	1	1	23	0
16	SPECIFIC	2	0	0	10	1
17	SPECIFIC	2	0	1	?	0
18	SPECIFIC	2	0	1	45	1
19	SPECIFIC	2	1	0	12	0

Table Definition:

Rows:

Columns:

OK Cancel

Figure 4. Type of variable definition.

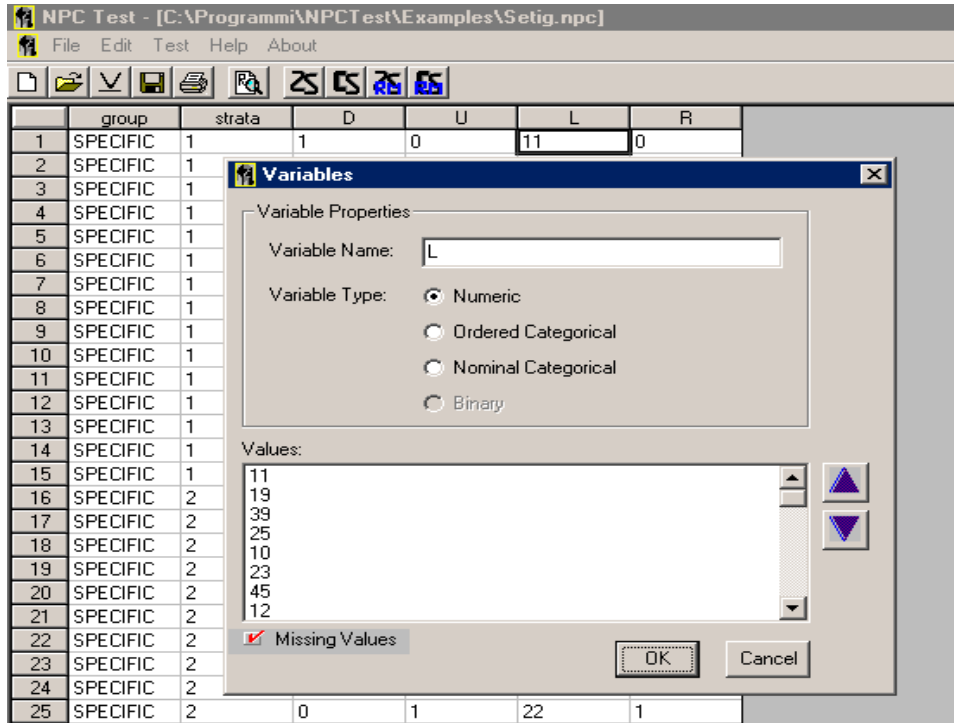


Figure 5. Partial tests definition.

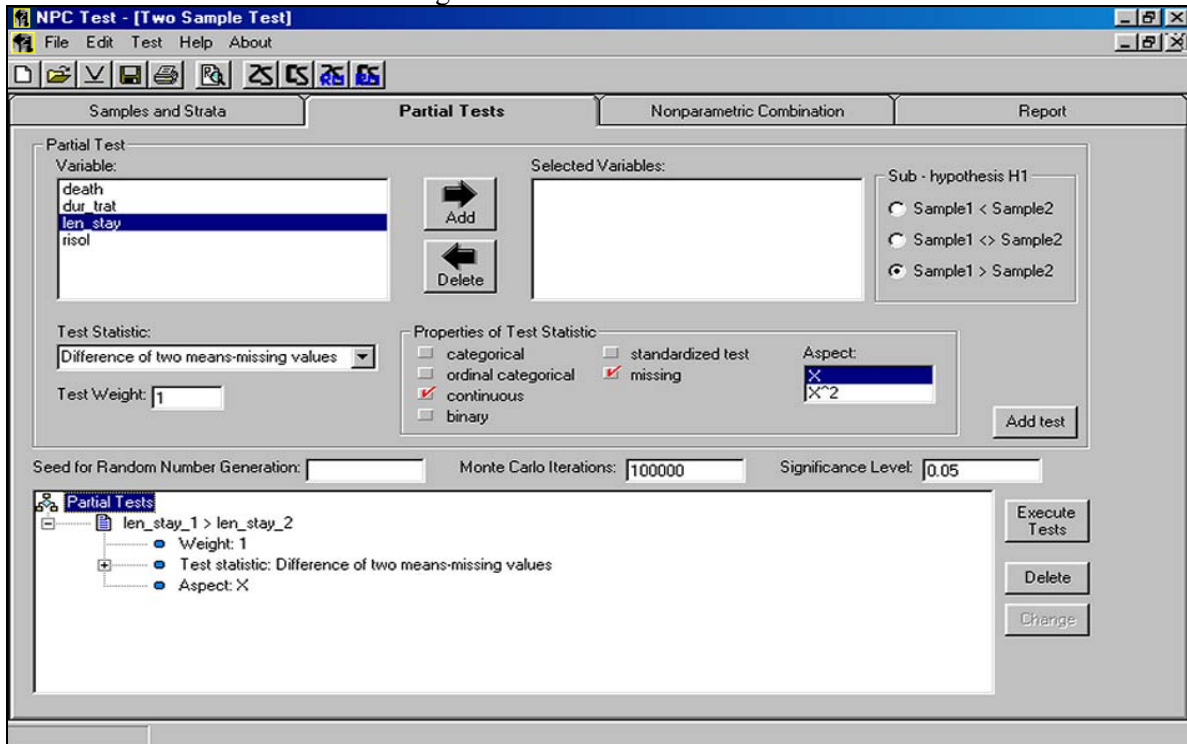


Figure 6. Nonparametric combination.

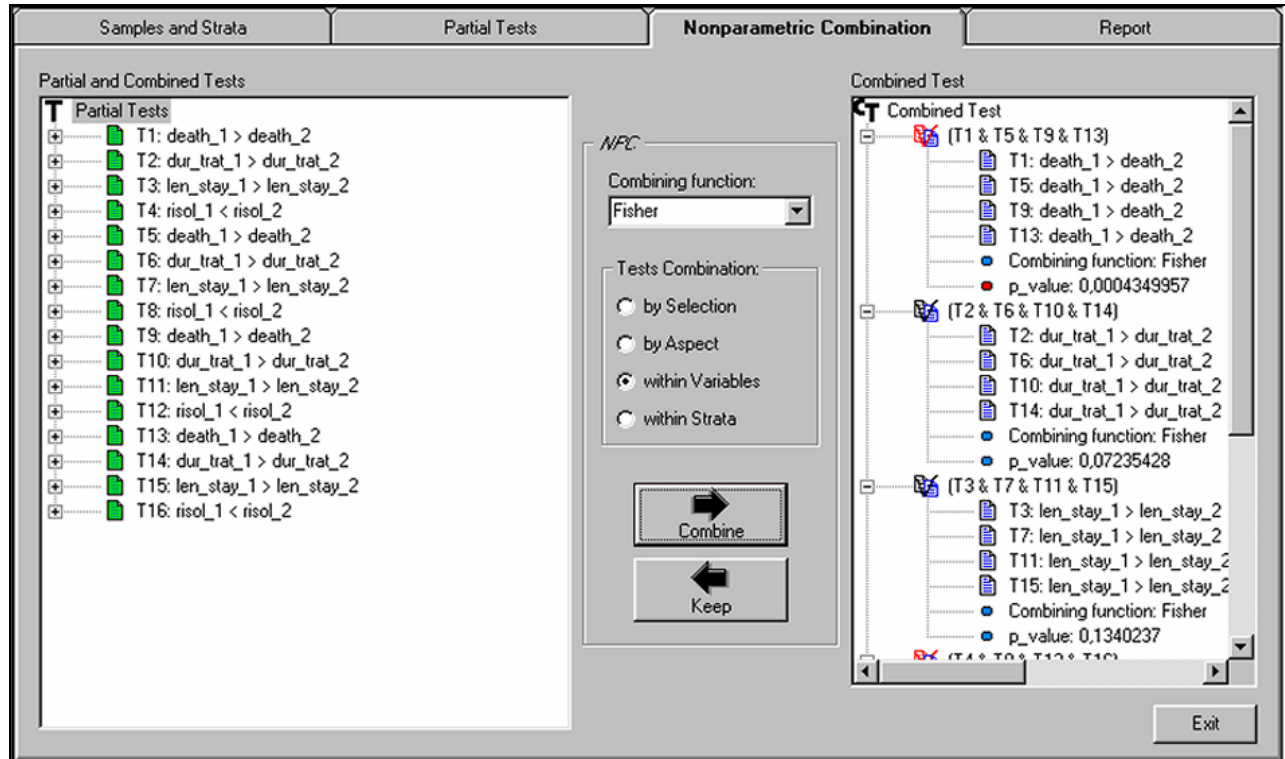


Figure 7. Performed tests in the report.

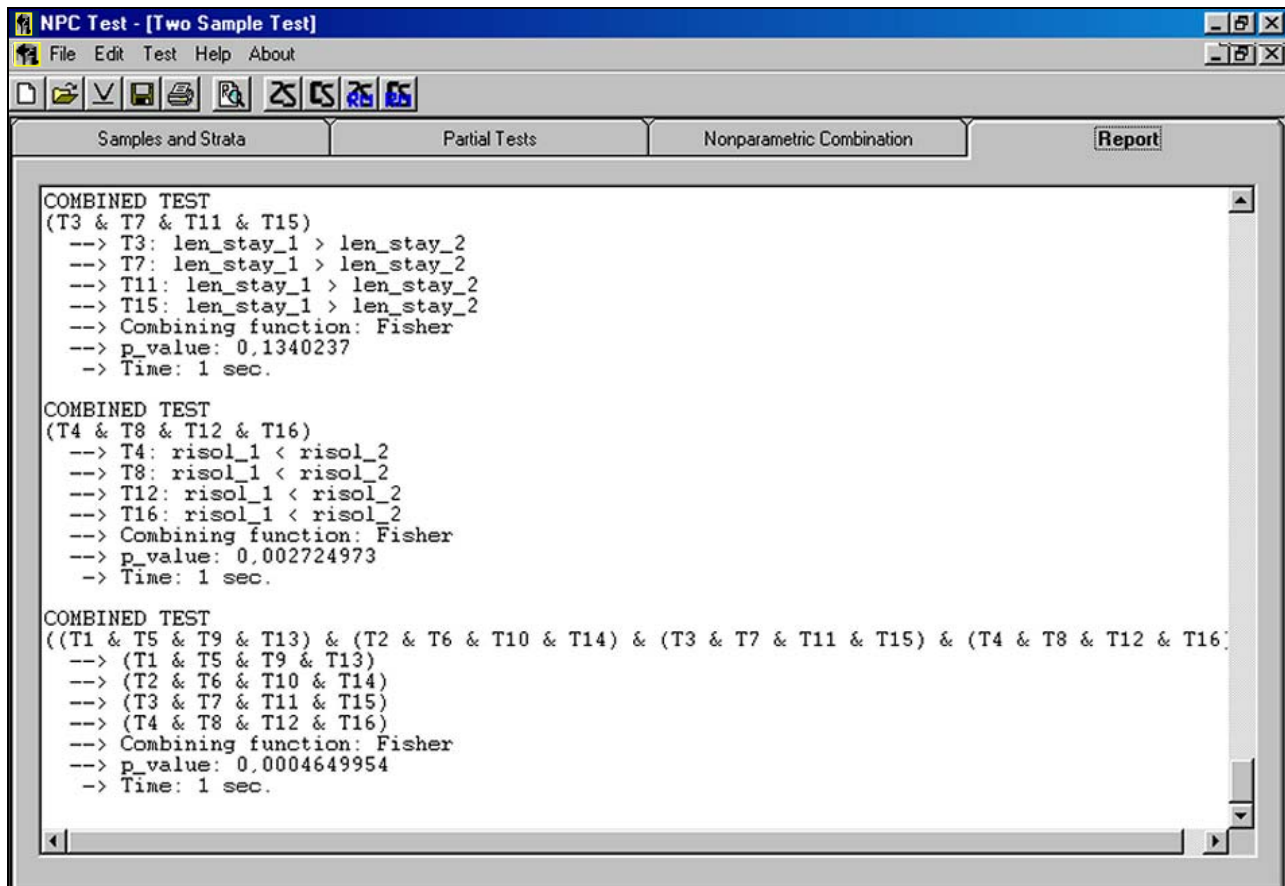
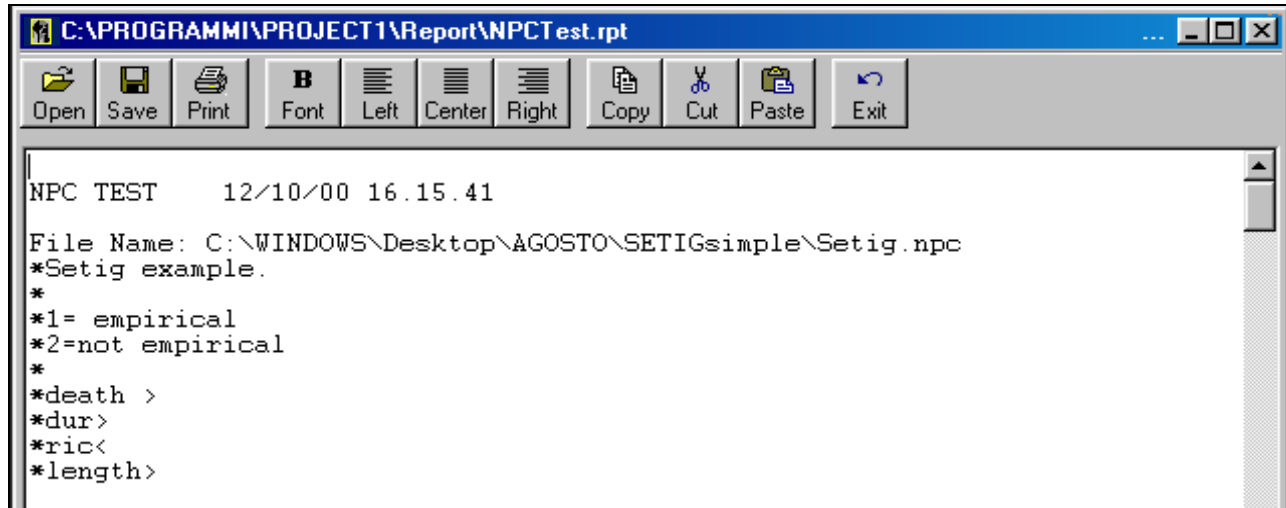


Figure 8. The report file editor.



Case studies: developing successful products and comparing two respiratory drugs

In order to better illustrate the NPC Test methodology let us develop two real case studies in the field of Management and Biostatistics. What does distinguish the best firms in the new product development (NPD) process? Over the past decade the New Product Development (NPD) process has been analysed in a number of works, both from an academic and a practitioner's point of view (Booz, Allen & Hamilton 1982; Madique & Zirger, 1984; Link, 1987; Cooper, 1990, 1993; Cooper & Kleinschidt, 1993; 1995; Pittiglio, Rabin, Todd & McGrath, 1995; Griffin, 1997, 1998). These works aimed at identifying NPD performance drivers, that is to say, all those practices, specific process configurations and internal business contexts which underlie the achievement of superior performances and company objectives.

However, these studies were carried out in different contexts and used both different measures of success and different methods of analysis. Griffin and Page (1993), in their literature review, identified 75 different measures previously used in papers on this topic, and classified them in the following groups: customer acceptance, financial performance, product level measures, firm based measures and program measures.

In general terms, in different industries and market types (i.e. B2C versus B2B) the relationship between drivers and performances and the appropriate set of measures of success to be considered may be different. For example, in a B2B marketplace a supplier involved in NP design, can be successful if the supplier is able to meet the specific needs of the client at a low cost and to carry out the task within an established time (Ragatz, Handfield & Scannell, 1997; Droge, Jayaram & Vickery, 2000). A company which produces industrial goods must consider the specific requirements of the customers and offer customized or semi-customized products. This can be done by using approaches and practices in NP development; making an effort to develop a partnership with customers (Hartley, Zirger & Kamath, 1997; Swink & Mabert, 2000; Tuten & Urban, 2001).

Recent studies have laid emphasis on the configuration of different drivers distinguishing between Best and Rest at a company level, considering the whole of the product the company developed in the last three or five years, i.e. the development program. Griffin (1997, 1998), for example, considered the NP program over a five year period and to do so, divided the sample on the basis of three sets of measures: market and financial success, relative success of the program in terms of meeting its objectives and, overall industry

success. Companies were classified as best when they were in the top third of their industry for NPD success and, also, were above the mean of the entire sample regarding the relative success of the program and market – financial success.

Context of the study, framework and key variables

This study aims at identifying the differences in driver configurations between successful and unsuccessful companies working in a B2B marketplace in two specific industries (Machinery Manufacturing, SIC35, and Electrical, Electronic Machinery, Equipment and Supplies, SIC36). We have considered all products developed and launched onto the market by each company in the last three years. Successful companies were defined as those above the median position for the global ranking of both the performances of the new product on the market and of the performances of the NPD process. This study has considered many different drivers: practices and processes, strategic guide and internal environment which support NP development.

The research considers companies which develop and produce industrial goods such as machinery, equipment and appliances to sell to other companies which use them in their production processes, or products, modules and components which will be incorporated into the client company's final products (in other words, these companies have other companies as clients, so their operations and businesses are conditioned by, for example: 1) the importance of the interaction between customer and supplier, so the NP department plays an important role in designing products based on the specific needs of the customer; 2) a limited number of customers with different requirements, 3) a short distribution channel and often direct sales; 4) a different and sometimes more critical role of marketing and promotion compared to a B2C environment; 5) customization or semi-customization of products; 6) a limited number of competitors (often companies that work in a niche or specialized market).

In this study, six categories of variables are considered, including performance and

driver use measures, referring to a three year NPD program:

PERFORMANCE

- NPD Operational Performances (IP, Internal Performances);
- Market, Products and Financial success (EP, External Performances).

DRIVER

- Product Architecture Approach;
- Organizational Mechanisms of NPD;
- Development Process of NPD;
- Strategic Capabilities.

Operational Performances (IP)

Operational Performances are those that depend on the NPD process, practices and environment support. Three types of performances are considered and are related to the time and quality dimensions of the development.

- Launch on Time;
- Time to Market Reduction;
- Product quality capability.

Market, Products and Financial success (EP)

The variables belonging to this category and considered in the present study are:

- Meet Profit Goals;
- Overall Product Success;
- Meet Revenue Goals.

Product Architecture Approach

The technical approach on product architecture.

- Standardization;
- Modularization;
- Platform.

Organizational Mechanisms

Organizational mechanisms refer to a set of techniques used during the various phases of the development process. Some of them concern technological aspects, others are concerned with organizational practices (PM, team, integration etc.).

- Project Manager Use;
- Customer Involvement (multi-item scale);
- Integration Design – Marketing;

- Integration Design – Manufacturing;
- Supplier Involvement (multi-item scale);
- Team Use.

Development Process

An NPD Process concerns the phases of the development itself and the overlapping level between these phases. The variables measure in how many cases during the development program each phase or approach has been used.

- Product Concept Development;
- Product Concept Test;
- Preliminary Design (multi-item scale);
- Late Engineering Changes (i.e. Early modifications);
- Overlapping Approach.

Strategic Capabilities

NP performances and success do not only depend on best practices and well defined process but also on the internal environment which supports NP development. This support can come from the management of the company (top management support, strategic guide) and from the capabilities of the employees.

- Up – Front Capabilities (VOC) (multi-item scale);
- Top Management Support;
- NP Strategic Guide (multi-item scale);
- Company Innovation Culture;
- Technological Capabilities (multi-item scale).

Distinguishing Best and Rest companies on the basis of high or low values in PI and PE, we obtain eight different groups (Figure 9):

- Best companies in PI (labeled BX);
- Rest companies in PI (labeled RX);
- Best companies in PE (labeled XB);
- Rest companies in PE (labeled XR);
- Best companies in both PI and PE (labeled BB);
- Rest companies in both PI and PE (labeled RR);
- Best companies in PI and Rest in PE (labeled BR);
- Rest companies in PI and Best in PE (labeled RB).

Among the set of all possible comparisons, after selecting only those more interesting from a research point of view (Figure 10), it is hypothesized that:

- H1: BX companies have higher level of drivers than RX companies;
- H2: XB companies have higher level of drivers than XR companies;
- H3: RB companies have higher level of drivers than RR companies;
- H3: BB companies have higher level of drivers than RR companies;
- H5: BB companies have higher level of drivers than BR companies.

In the empirical analysis conducted during the year 2000, we considered all NPs marketed from 1997 to 1999 by each company: this was defined as the NPD program. Market, product and financial measures of success refer to the results obtained as a result of the NPD program. For operational performances we considered the percentage of new products that have obtained high operational performances. As regards the drivers, in almost all cases we asked the company the percentage of projects which had adopted a certain driver. In other cases (i.e., capabilities and internal culture) we obtained the level of presence in the company as a whole, because it is practically impossible to discern the adoption percentage among projects for this type of variable.

Data and information were gathered through a questionnaire mailed to Italian manufacturing companies working in the B2B market in the mechanical and electronic sectors (SIC codes 35 & 36), with more than 100 and less than 1000 employees and a revenue of more than 20 billion Lire per year (approximately 10 million Euro). The addresses of the companies we mailed the questionnaire to were taken from Dun & Bradstreet's Business to Business database. The questionnaire was addressed to the new product development department manager. Phone assistance was provided to ensure that the information gathered was both complete and correct and some managers were interviewed. The sample was made up of 85 companies. Table 1 shows the composition of the sample used for the data analysis.

Figure 9. Graphical representation of Best and Rest definition.

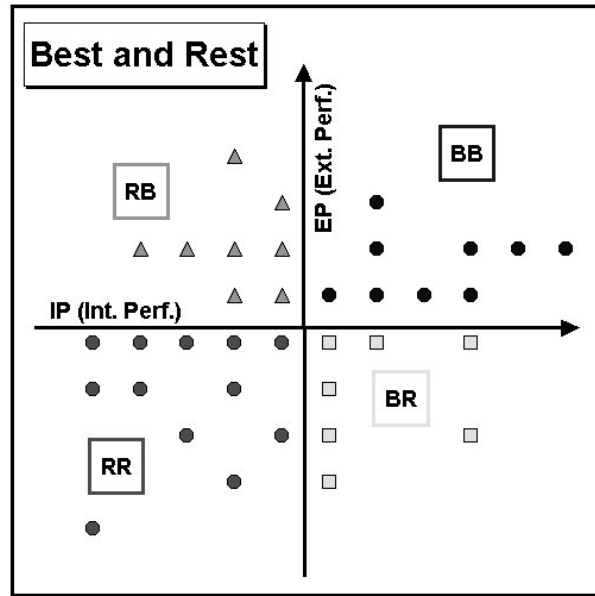


Figure 10. Graphical representation of research hypotheses.

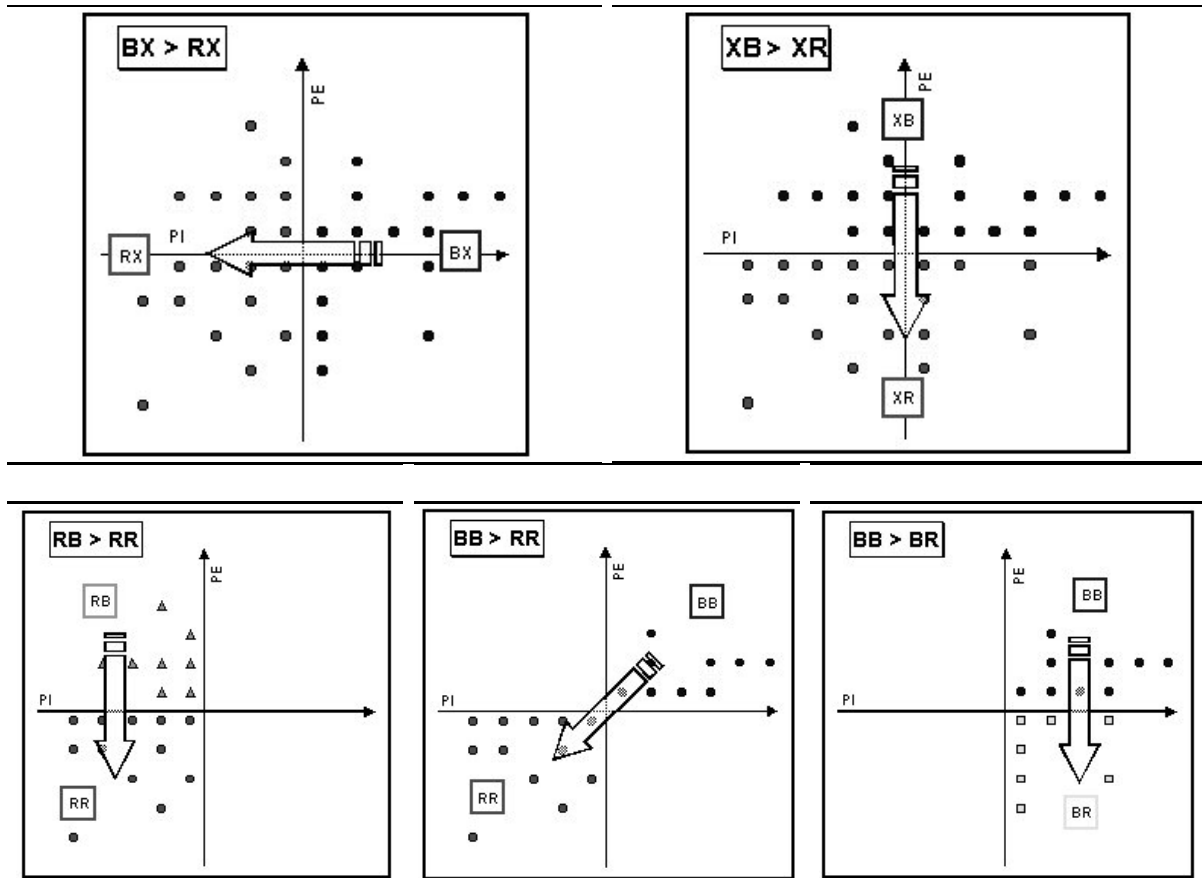


Table 1. Sample used for the data analysis.

Code	Description	N
SIC35	Machinery Manufacturing	60
SIC36	Electrical, Electronic Machinery, Equipment & Supplies	25
Total Sample Size		85

During the three year period considered (1997-1999), the firms launched a total of about 900 new products classified by the companies themselves as follows:

- 41% new products for new markets;
- 33% partially or totally substitute products;
- 26% products with significant improvements with respect to existing ones.

Best and Rest definition: the NonParametric Combination (NPC) of dependent rankings method

In many real situations we encounter the need to compare entities of a different nature (products, services, companies, behavior and so on) in order to obtain a ranking among the considered statistical units. If the comparison is based on only one feature the result is obtained in a trivial way but difficulties may arise when we are dealing with two or more informative variables jointly. We can build up as many rankings as the number of features we are dealing with. Apart from the case where units occupy the same position in every ranking, the need to summarize a set of rankings into one single global ranking arises.

The main purpose of the method (Pesarin, 2000) is to obtain a single ranking criterion for the statistical units under study, which summarizes many starting partial (univariate) criteria. This method is defined as nonparametric since it needs neither the knowledge of the underlying statistical distribution for the variables being studied, nor the dependence structure among variables, apart from the assumption that all dependences are monotonic regressions.

Methodology

Given a multivariate phenomenon $\mathbf{X}=[X_1, X_2, \dots, X_k]$, observed on N statistical units, and once we have calculated the k partial rankings R_1, R_2, \dots, R_k , starting from the variables $X_i, i=1, \dots, k$, each one being informative about a partial aspect of phenomenon \mathbf{X} , we want to build up a *global combined ranking* Y :

$$Y = \psi(X_1, X_2, \dots, X_k; w_1, w_2, \dots, w_k), \psi: \mathbb{R}^{2k} \rightarrow \mathbb{R}^1$$

where ψ is a real function allowing us to combine the partial dependent rankings and where w_1, w_2, \dots, w_k is a set of weights, defined on the basis of technological, functional or economic considerations, which measure the relative degree of importance among the k aspects of \mathbf{X} .

In order to build up Y , a set of minimal reasonable conditions related to the variables $X_i, i=1, \dots, k$ are introduced:

1) For each of the k informative variables a partial ordering criterion is well established, in the sense that large is better; if it is not so, it is possible to recode the variables by means of any appropriate transformation φ :

a) if large is worse $\Rightarrow \varphi(X)=1/X$ or $\varphi(X)=-X$;

b) if δ is better (central target value) $\Rightarrow \varphi(X)=|X-\delta|$;

2) Regression relationships within the k informative variables are monotonic (increasing or decreasing)

3) The marginal distribution of each informative variable is non-degenerate.

Moreover, further assumptions need not be made, either on the statistical distribution of the informative variables, or on their dependence structure. Finally, notice that there is no need to assume the continuity of $X_i, i=1, \dots, k$, so that the probability of ex-equo can be different from zero.

Define the set of variables X_i as $\{Z_{ji}, i=1, \dots, k, j=1, \dots, N\}$, possibly after proper transformations. Without loss of generality, they are assumed to behave in accordance with the rule “large is better”. In this setting, we consider the rank transformations R_{ji} (*partial rankings*):

$$\{R_{ji} = R(Z_{ji}) = \#(Z_{ji} \geq Z_{hi}), i=1, \dots, k, j, h=1, \dots, N\}.$$

Associated with these ranks are the scores:

$$\left\{ \lambda_{ji} = \frac{R_{ji} + 0.5}{N + 1}, i = 1, \dots, k, j = 1, \dots, N \right\}.$$

Once a combining function ψ (for details of combining functions see paragraph 2.1 above) has been chosen, we compute the transformation

$$\psi: \{Y_j = \psi(\lambda_{j1}, \dots, \lambda_{jk}, w_1, \dots, w_k), j=1, \dots, N\},$$

and finally, applying the rank transformation, we obtain the global combined ranking Y :

$$\{Y_j = R(Y_j) = \#(Y_j \geq Y_h), j, h=1, \dots, N\}.$$

In the global ranking Y , each statistical units is ranked in a unique way, by taking into consideration the whole set of the k informative variables.

The method of nonparametric combination of dependent rankings has proved to be particularly useful for the problem of finding a meaningful classification criterion for the sample in groups, distinguishing companies which develop successful products from those which develop less successful products from the point of view of both Internal and External performances.

In fact, once the method is applied to the two sets of variables, the first measuring the *Market, Products and Financial* success (EP, External Performances) and the second measuring the *NPD Operational Performances* (IP, Internal Performances), obtain two global rankings of the companies, taking into account all success criteria.

Therefore, in these two global combined rankings the successful companies in External and Internal Performances were those in the upper positions while the worst companies were

those in the lower positions. As a discrimination rule, adopt the median positions. Those companies with a position above the median position in the global ranking were chosen as Best companies in EP and IP, and the remaining companies were labelled as Rest companies.

As a sensitivity analysis we performed an NPC testing procedure to verify whether the division was significant or not, that is to say whether Best companies in IP revealed a significantly higher level of operational variables and Best companies in EP revealed a significantly higher level of success variables.

As the associated p -values in Table 2 show, we can verify that at a 5% significance α -level the Best companies in IP are characterized by higher levels in all three operational variables and in the global test, taking into account the multivariate distribution of all three variables. In the same way the Best companies in EP are characterized by higher levels in all three success criteria and in the global test, taking into account the multivariate distribution of all three variables.

<i>Internal Performances (IP)</i>			Global
Launch on Time	Time to Market Red.	Quality Capability	
.000	.000	.015	.000

<i>External Performances (EP)</i>			Global
Meet Profit Goals	Overall Prod. Succ.	Meet Rev. Goals	
.000	.000	.000	.000

Table 2. Sensitivity analysis for testing the division in Best and Rest for both internal and external performances.

By simultaneously crossing the two rankings, the sample was divided into four classes, i.e. BB, BR, RB and RR (the first letter represents the internal performances), as shown in the Table 3. This final classification into four groups has been used to test the research questions.

Results

The NPC Test aims to identify the significant differences of the considered variables which characterize two specific groups. A p -value table is presented below for each of the five tested hypotheses (we use the graph representation only for the first hypothesis (Figure 11), where for sake of clarity a gray node means a significant p -value at a 5% α -level), reflecting the nested data set configuration in correspondence to the three variables classification: (1) multi-item scale variables, (2) variables belonging to the same driver group and (3) a final grouping which considers all driver groups together. As a result, the testing procedure is split up into the following phases:

- 1.1) is only for multi-item variables (if they are included in the driver group), performing the partial tests and
- 1.2) combines them into a single second order combined test;
- 2.) performs the other partial tests, in each group, for the remaining variables and
- 3.) combines them within the driver group, along with the combinations of step 1.2, obtaining a third order combined test of all variables within a driver group;
- 4.) finally, combines the four combined tests from step 2.2 (one for each driver group) in a global final test which is informative on the global null hypothesis.

In order to make the detecting of relevant differences easier, only significant p -values at 5% α -level have been printed in Table 4. The results suggest the Best companies on Market/Financial performances use the Architecture Approach more than the others. In particular, it is interesting to note that this group of variables does not discriminate between Best and Rest regarding the Operational Performances.

In other words it seems that an extensive use of product architecture related practices, such as the development of a product platform upon which to develop an entire new product line, the standardization of components to reduce production costs, modularity to offer a greater variety of products to the customer while

at the same time containing the internal variety the company has to deal with, allows the company to overcome any deficiencies in Operational Performances. This result is easier to understand if you consider the fact that some variables, which may influence external performances, have not been considered in the present study. These variables, such as for example the cost of the product on the market, are in turn influenced by company choices about the product architecture.

However, the main result is the great difference between the various groups in the Strategic Capability variables, and in particular the existence of a shared development strategy, well-defined development objectives and high technological capabilities. These variables represent the most noticeable difference between the various Best-Rest comparisons previously performed. In other words, strategic capabilities can help to achieve superior performance both on the operational and market/financial side. Perhaps these are the variables the companies have to act on in order to reach superior NPD performances, according to previous literature on this topic (see, for example, Griffin, 1998; and Cooper & Kleinschmidt, 1993).

A comparison between two different respiratory drugs

With the aim of comparing the features of two different respiratory drugs, a sample of 226 patients was recruited and then randomized into two distinct groups: group A, treated with a new drug labelled with A, and group B, treated with an old usual drug B. The purpose of the study is to establish, whether the new drug A is better than B, stressing the multivariate nature of the clinical comparison. In fact, we wish to make a decision on the basis of the three measured clinical end-points: D_MAT, D_SER and COMPL.

The first two clinical parameters are numeric variables which quantify the patient's health by means of a measure of respiratory airways expanding: D_MAT, is the difference, measured at noon, between the average of respiratory airways expanding evaluated two weeks before treatment (wash-out phase) and six weeks after treatment, and D_SER, the same difference, measured in the afternoon. The last

clinical parameter (COMPL) is a binary measure of therapy finishing: with this variable we can evaluate whether treatment A has a better degree of tolerance than B.

The null hypothesis states that A and B present no differences in their benefits, that is the equality in distribution of the multivariate distribution of the 3 responses in both groups:

$$H_0 : [P_A = P_B] = \left[\mathbf{X}_A \stackrel{d}{=} \mathbf{X}_B \right]$$

where \mathbf{X}_A and \mathbf{X}_B represent the multivariate random variables underlying group A and B, and P_A and P_B are the corresponding probability functions.

In the context of nonparametric combination, H_0 is supposed to be properly decomposed into

$$H_0 : \left[\begin{array}{l} D_MAT_A \stackrel{d}{=} D_MAT_B \\ \cap \left[D_SER_A \stackrel{d}{=} D_SER_B \right] \cap \left[COMPL_A \stackrel{d}{=} COMPL_B \right] \end{array} \right],$$

thus H_0 (multivariate) is true if all the H_{0i} (univariate) are jointly true.

The alternative hypothesis H_1 is represented by:

$$H_1 : \left[\begin{array}{l} D_MAT_A \stackrel{d}{>} D_MAT_B \\ \cup \left[D_SER_A \stackrel{d}{>} D_SER_B \right] \cup \left[COMPL_A \stackrel{d}{>} COMPL_B \right] \end{array} \right],$$

where $\stackrel{d}{>}$ means stochastic dominance. With 10000 CMC iterations results are shown in table 5.

It is concluded that from a multivariate point of view treatment A is better than B at 1% α -level. In order to take multiplicity into account, the FWE (Family Wise Error rate) must be considered to draw inferential conclusions, not only for the global test, but also for partial tests. At present, one of the best procedures is the *Closed Testing* (see e.g. Westfall et al., 1999).

For details on closed testing procedures with NPC, the reader should consider Finos et al. (2001). In this case-study, closed testing through NPC Test provided the result shown in Figure 12. Hence, after considering closed testing p-value corrections, the D_MAT and D_SER are both found to be significant, the first at 1% α -level and the second at 5% α -level.

The analysis can be extended by considering the same problem of the comparison between A and B treatments with the inclusion of a possible confounding factor, i. e., the patient's age. In order to do so we stratify the sample into Y, 4-8 year-old patients, and by O, 9-13 year-old patients.

In this way the hypothesis system is rewritten as:

$$H_0 : \bigcap_{j=Y,O} \left\{ \bigcap_{i=1}^k \left[X_{Ai} \stackrel{d}{=} X_{Bi} \right] \right\},$$

against the alternative

$$H_1 : \bigcup_{j=Y,O} \left\{ \bigcup_{i=1}^k \left[X_{Ai} \stackrel{d}{>} X_{Bi} \right] \right\}.$$

Notice that when we decide to aim our analysis at strata, we add a second step into our two-phase algorithm with the within-strata combination. Results are provided in Table 6.

The closed testing correction was also performed. Apart from a clinical interpretation of results which we do not consider here, it is worth noting that very complete information is provided by NPC Test analysis. Since the global test is significant at 1% α -level we are also able to identify:

- that only stratum O contributed to the overall significance;
- variables D_MAT and D_SER within stratum O contribute to the stratum significance.

Table 3. The four group definition.

EP External Perf.	IP (Internal Performances)		Tot.	Group	N.
	B	R			
B	25	13	38	BB	25
R	24	23	47	BR	24
Tot.	49	36	85	RB	13
				RR	23
				Tot.	85

Table 4. P-value table for each of the five tested hypotheses.

Driver / DRIVER GROUP	H1 BX>RX	H2 XB>XR	H3 RB > RR	H4 BB > RR	H5 BB > BR
Standardisation		.010			
Modularisation		.001	.001	.032	.019
Platform		.027	.039	.030	
PROD. ARCHITECTURE APPR.		.002	.006	.044	
Project Manager Use					
Customer Involvement 1					
Customer Involvement 2					
Customer Involvement					
Integration Design - Marketing					
Integration Design - Manufacturing					
Supplier Involvement 1					
Supplier Involvement 2					.022
Supplier Involvement					
Teame Use					
ORGANISATIONAL MECHANISMS					
Product Concept Development					
Product Concept Test					.024
Product Concetpt					.041
Pre-Design 1					
Pre-Design 2					
Pre-Design					
Late Engineering Changes					
Overlapping Approach					
DEVELOPMENT PROCESS					
Up – Front Capabilities 1	.003			.000	.019
Up – Front Capabilities 2	.004			.004	
Up – Front Capabilities	.002			.000	.024
Top Management Support					.008
NP Strategic Guide 1	.010	.036		.009	
NP Strategic Guide 2	.010			.010	
NP Strategic Guide 3	.017				
NP Strategic Guide	.003			.002	
Company Innovation Culture		.004		.002	.003
Technological Capabilities 1	.023	.000		.000	.001
Technological Capabilities 2	.008	.034		.000	
Technological Capabilities	.008	.001		.000	.006
STRATEGICAL CAPABILITIES	.010	.003		.000	.012
GLOBAL	.037	.010	.022	.000	.044

Table 5.

A > B		D_MAT	D_SER	COMPL	GLOBAL
	<i>p-value</i>	0.0009	0.0014	0.0660	0.0003

Figure 11. Graphical representation of testing hypothesis BB>RR.

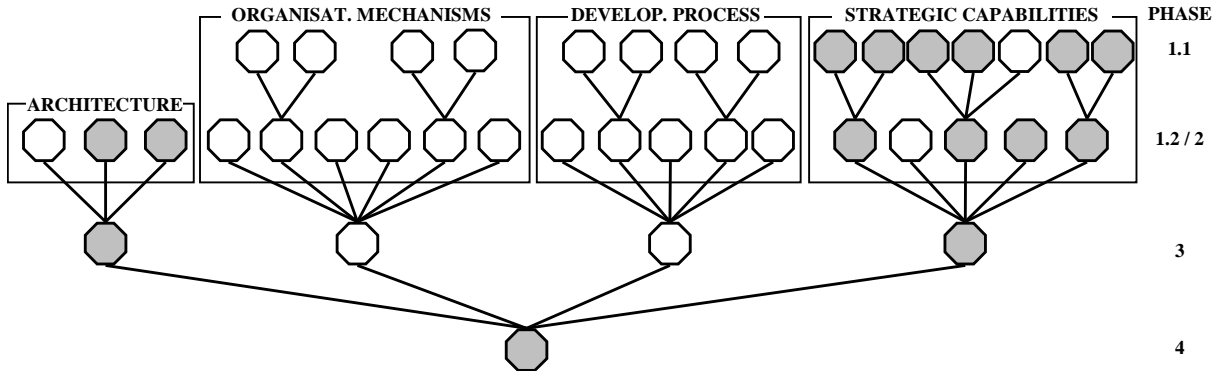


Figure 12: Closed testing procedure performed by NPC Test.

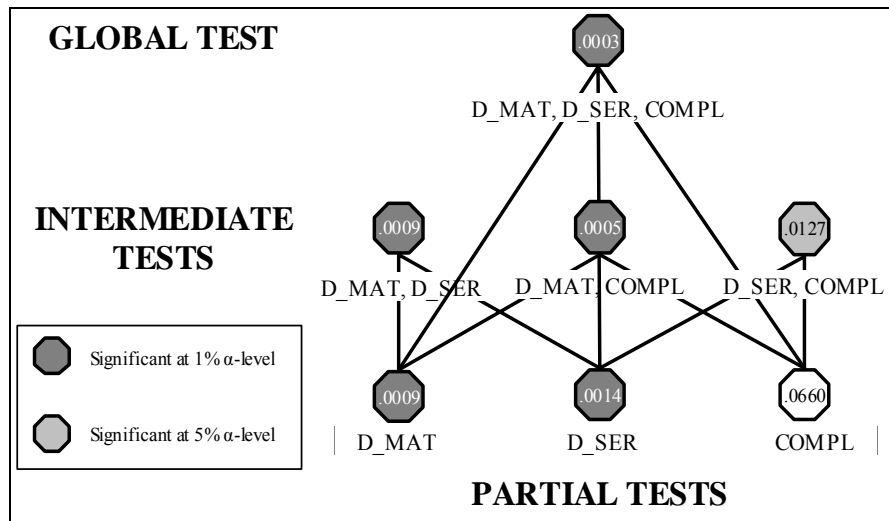


Table 6: P-values of partial, within-strata and global test considering stratification by patient’s age.

A > B	AGE	D_MAT	D_SER	COMPL	COMBINED
	Y: 4-8	0.0572	0.0520	0.1768	0.0520
	O: 9-13	0.0046	0.0353	0.2303	0.0068
	GLOBAL				0.0095

References

- Booz, Allen, & Hamilton. (1982). *New products management for the 1980s*. NY: Booz, Allen & Hamilton.
- Cooper, R. G. (1990, November - December). New products: what distinguishes the winners. *Research and Technology Management*, p. 27-31.
- Cooper, R. G. (1993). *Winning at new products: accelerating the process from idea to launch*. Reading, MA: Addison-Wesley.
- Cooper, R. G., & Kleinschmidt, E. J. (1993). New product success in the chemical industry. *Industrial Marketing Management*, 22 (2), 85-99.
- Droge, C., Jayaram, J., & Vickery, S. K. (2000). The ability to minimize the timing of new product development and introduction: an examination of antecedent factors in the north american automobile supplier industry. *Journal of Product Innovation Management*, 17 (1), 24-40.
- Finos, L., Pesarin, F., & Salmaso, L. (2001). *Test Combinati per il controllo della molteplicità mediante procedure di Closed Testing*. Dipartimento di Statistica, Università di Padova, Italy; Working Paper n.8/2001.
- Folks, J. L. (1984). *Combinations of independent tests*. In P.R. Krishnaiah & P. K. Sen (Eds.), *Handbook of Statistics*, 4, 113-121, Amsterdam: North-Holland.
- Griffin, A. (1997). PDMA research on new product development practices: Updating trends and benchmarking best practices. *Journal of Product Innovation Management*, 14 (6), 429-458.
- Griffin, A. (1998). *Drivers of new product development success: multivariate results from PDMA's best practices research*. In: EIASM - 5th International Product Development Management Conference, Como, I, p. 491-502.
- Griffin, A., & Page, A. L. (1993). An interim report on measuring product development success and factors. *Journal of Product Innovation Management*, 10 (3), 291-308.
- Hartley, J. L., Zirger, B. J., & Kamath, R. R. (1997). Managing the buyer - supplier interface for on-time performance in product development. *Journal of Operations Management*, 15 (1), 57-70.
- Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods*. (2nd Ed.). New York: Wiley.
- Joe, H. (1997) *Multivariate models and dependence concepts*. London: Chapman & Hall.
- Link, P. L. (1987). Keys to new product success and failure. *Industrial Marketing Management*, 16, 109-118.
- Madique, M. A., & Zirger, B. J. (1985). A study of success and failure in product innovation: The case of the u.s. electronics industry. *IEEE Transactions on Engineering Management*, 31 (4), 192-203.
- Methodologica Srl. (2001). *NPC Test© 2.0 - Statistical software for multivariate permutation tests*. Methodologica srl (<http://www.methodologica.it>), Treviso, Italy.
- Nelson, L. S. (1992). A randomization test for ordered alternatives. *Journal of Quality Technology*, 24, 51-53.
- Pesarin, F., & Lago, A. (2000). Nonparametric combination of dependent rankings with application to the quality assessment of industrial products. *Metron*, LVIII, 1-2.
- Pesarin, F. (2001). *Multivariate permutation tests with applications in biostatistics*. Chichester: Wiley.
- Pesarin, F. (2002, in press). Extending conditional inferences to unconditional ones. *Journal of the Italian Statistical Society*. 11, 161-173.
- Pittiglio, Rabin, Todd, & McGrath, (1995). *Product development leadership for technology - based companies: Measurement and management - a prelude to action.*, Weston, MA: Pittiglio Rabin Todd & McGrath.
- Ragatz, G. L., Handfield, R. B., & Scannell, T. V. (1997). Success factors for integration suppliers into new product development. *Journal of Product Innovation Management*, 14, 190-202.
- Swink, M. L., & Mabert, V. A. (2000). product development partnership: balancing the needs of OEMs and suppliers. *Business Horizons*, 43 (3), 59-68.
- Tuten, T. L., & Urban, D. J. (2001). An Expanded Model of Business-to-Business Partnership Formation and Success. *Industrial Marketing Management*, 30 (2), 149-164.
- Westfall, P., Tobias, R., Rom, D., Wolfinger, R., & Hochberg, Y. (1999). *Multiple Comparisons and Multiple Tests using the SAS®*, SAS Books by Users.