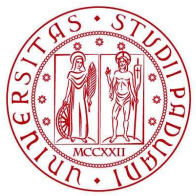

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXII

INNOVATIVE APPROACHES FOR THE CONSTRUCTION OF ORTHOGONAL ARRAYS AND PERMUTATION TESTS IN COMPLEX EXPERIMENTAL DESIGNS

Direttore della Scuola: Ch.ma Prof.ssa Alessandra Salvan

Supervisore: Ch.mo Prof. Luigi Salmaso

Co-supervisori: Ch.mo Prof. Fortunato Pesarin, Ch.mo Dott. Roberto Fontana

Dottoranda: Susanna Ragazzi

Data 30/06/2011

To my family

La capacità di ben giudicare e di distinguere il vero dal falso, che è ciò che propriamente si chiama buon senso o ragione, è naturalmente uguale in tutti gli uomini; così che la diversità delle nostre opinioni non deriva dal fatto che gli uni sono più ragionevoli degli altri, ma solo dal fatto che percorriamo diverse vie di pensiero, e non prendiamo in considerazione le stesse cose.

Cartesio *Discours de la methode* A.T.II, p. 88

Sommario

Il presente lavoro di tesi è stato finanziato da un borsa erogata dalla Fondazione Cariparo sul progetto di ricerca intitolato "*The evaluation of customer satisfaction to improve product and services by new statistical methods of Conjoint Analysis and Design of Experiments*". Il lavoro si è concentrato principalmente su due aspetti pienamente pertinenti al progetto sopracitato ovvero la costruzione di Orthogonal Arrays (OAs), funzionale alla Conjoint Analysis, e lo sviluppo di nuove tecniche inferenziali nell'ambito della pianificazione degli esperimenti. Si è ritenuto pertanto di inserire come titolo della presente tesi "*Innovative approaches for the construction orthogonal arrays and permutation tests in complex experimental designs*" in modo sia da rispecchiare in modo puntuale i contenuti dell'elaborato di tesi sia per evidenziare quali parti del progetto iniziale sono state maggiormente approfondite. Molti sono i contributi sulle problematiche di esistenza e costruzione degli OAs, qui si propone un approccio costruttivo basato sui concetti di *counting polynomial functions*; dopo un lavoro di formalizzazione dei principali concetti e nozioni, si procede alla generazione di numerosi OAs considerando diverse numerosità di livelli e di forza. Particolare attenzione è posta nella costruzione di classi nonisomorfe di OAs attraverso le quali si propone un metodo per l'analisi statistica di piani multifattoriali a due ed a tre livelli non replicati. Le costruzioni dei disegni nonisomorfi sono basate sui concetti di *algebraic strata*, evitando così la computazione in campo complesso. Per quanto concerne lo sviluppo di metodologie inferenziali non parametriche nei disegni sperimentali, si sono presi in considerazione i disegni Randomized Complete Block (RCB) ed i disegni Split-plot (SP). Per entrambi i disegni dopo una revisione critica e sistematica dei metodi inferenziali proposti in letteratura si propongono applicazioni originali dei test di permutazione. Viene considerato, per i disegni RCB , sia il caso univariato che multivariato, proponendo soluzioni basate su più fasi di combinazione dei test allo scopo di ottenere una soluzioe globale ed anche di provveddere ad una metodologia per la correzzione della molteplicità dei p-value. Sono stati inoltre effettuati studi di validazione dei test al fine di confrontare il comportamento, in termini di potenza, dei test parametrici e non parametrici da letteratura con quelli proposti. Viene inoltre effettuata un'applicazione reale del metodo nell'ambito della sperimentazione di nuove possibili fragranze di un determinato detergente con quella presente attualmente sul mercato. Per i disegni SP si valutano procedure di permutazione non parametriche basate sulle simmetrie dei blocchi (SYP) e combinazioni dirette dei test parziali al fine di testare i possibili effetti dei fattori principali e delle loro interazioni. Tramite uno studio di simulazione si valutano in potenza e sotto H_0 tali test, comparandoli con test noti da letteratura. Sulla base dei risultati ottenuti, possiamo affermare che i metodi proposti forniscono una valida alternativa e soluzioni efficienti per problemi di analisi nella pianificazione degli esperimenti in presenza di non normalità delle componenti d'errore, cosa del resto frequente in studi di customer satisfaction, ed in presenza di fattori bloccati.

Abstract

This thesis work has been funded by a Fondazione Cariparo's scolarship for the research project entitled on "*The evaluation of customer satisfaction to improve product and services by new statistical methods of Conjoint Analysis and Design of Experiments*". The work is principally focused on two aspects of the cited project which are the construction of Orthogonal Arrays (OAs), useful for Conjoint Analysis, and the development of new inferential techniques in planning of experiments. We consequently considered to insert the following title "*Innovative approaches for the construction of orthogonal arrays and permutation tests in complex experimental designs* "in order to give a clear representation of the contents of the thesis work and to highlight the parts, of the initial project, which have been deepened. There has been wide contribution on OAs' existence and construction issues; in this work we propose a constructive approach based on the concept of *counting polynomial functions*; after a formalization of the main concepts and notions, we generate several OAs with different number of levels and strength. Particular attention is given to the construction of nonisomorphic classes of OAs, using these designs we propose a method for the statistical analysis in two-level and three-level unreplicated multifactorial designs. As for the development of inferential nonparametric methods for the analysis of data in experimental designs, we take into consideration the Randomized Complete Block (RCB) and Split-plot (SP) designs. After a critic and systematic revision of the inferential methods proposed in literature, we propose original applications of the permutation tests. As for the RCB designs, we took into consideration both the univaried and the multivaried case, proposing solutions based on more phases of the combination of the tests in order to obtain a global solution and to provide a method to adjust for the multiplicity of the p-value. Furthermore, we made studies for the validation of the adequacy of the proposed methods and we compared in terms of power the behavior of parametric and nonparametric tests from the literature with the proposed procedures. We also made an actual application of the method in the experimentation of new possible fragrances of a determined detergent with the one currently present on the market. For the SP designs we evaluated nonparametric permutation procedures based on the symmetries of the blocks (SYP) and direct combination of the partial tests in order to test the possible effects of the whole plot factors and their interactions. Through a simulation study we evaluated such tests in power and under H_0, comparing them with known tests from the literature. Based on the results, we can affirm that the proposed methods provide a valid alternative and efficient solutions to analysis problems in experiment planning in presence of non-normality of the error components, which is frequent in customer satisfaction studies and in presence of blocked factors.

Contents

Chapter 1

Introduction

1.1 Overview

The initial objective of this thesis work was to give methodological contributes in permutation inferential techniques for testing active effects in experimental designs. The designs took into account were the unreplicated (fractional) multifactorial designs, Randomized Complete Block designs, with particular focus in the case of ordered categorical response variables, and Split-plot designs. The methodologies developed substantially differ from the usual permutation techniques, because, in the major part of cases, the underlying assumptions of the permutation methods are not justified.

When considering unreplicated designs it is not possible to obtain an estimate of the variance of the errors, therefore the usual inferential techniques, aimed at identifying the significantly active factors, are unsuitable. In this work we provided a permutation test for active effects in unreplicated multifactorial designs considering classes of nonisomorphic orthogonal arrays (Arboretti, Fontana and Ragazzi (2011)). We proposed a solution based on extensions of the *inequivalent matrices permutation* testing procedures (Basso et al. (2004)) in order to obtain separate permutation tests for the effects in unreplicated fractional factorial designs. Rather than permuting responses of a single observed design matrix (or, equivalently, permuting rows of the design matrix keeping the responses fixed), matrices are exchanged in order to obtain the permutation distribution. In order to validate the proposed method we performed a Monte Carlo simulation study and we found out that the permutation tests appear to be a valid solution for testing effects, in particular when the usual normality assumptions cannot be justified. We showed our methodologies on three-level unreplicated fractional factorial designs but the approach suggested is general, it is valid for mixed fractional factorial designs (designs in which factors can assume different levels' number) and for testing the interactions between factors (if mixed nonisomorphic orthogonal arrays of strength greater than three are available!). The methodologies suggested strictly depend on the number of nonisomorphic orthogonal arrays available for the experimentation. In order to extend the proposed method when

there are only a few nonisomorphic orthogonal designs we experimented a modified procedure. Given a list of k_1 nonisomorphic arrays, for each of its members we randomly generate, by row permutations, a certain number r of isomorphic arrays. Then the hybrid class made by $k = k_1 \cdot r$ arrays replaces the class of nonisomorphic orthogonal arrays and the steps of the procedure remain the same. Also this procedure has been validate by Monte Carlo simulation study.

Orthogonal arrays can be considered essentials in all statistic techniques, as for example Conjoint Analysis, in which it is needed to use fractional factorial designs assuming the orthogonality between factors, i.e. all combination of factor levels appear equally often. The major advantage to use orthogonal arrays is to give the possibility to estimate low-order interactions between factors, in particular each orthogonal array of strength t specifies the only fraction of resolution $t+1$. In this work we generated several orthogonal arrays using recent algebraic method based on the *counting polynomial functions* (Fontana et al. (2000), Pistone and Rogantin (2008)(a)), by these function it is possible to construct fractions without restrictions on the cardinality of the levels (such as prime or prime powers) but implies computation in the complex field. In order to avoid this computational problem we constructed orthogonal arrays using the concepts of *algebraic strata*. By this method we provided a full list of nonisomorphic orthogonal arrays of strength two and three with a given size (Arboretti, Fontana and Ragazzi (2011)).

The assumption of exchangeability of data with respect to *groups*, or *blocks*, according the so-called treatment levels of the experiment, under the null hypothesis is required in the usual permutation testing procedures (Pesarin (2001)), but this cannot be assumed in Split-plot designs. Split-plot design may be refer to a common experimental setting where a particular type of restricted randomization has occurred during a planned experiment. In this work we presented a new method to perform inference on Split-plot experiments via combination-based permutation tests (Corain, Ragazzi and Salmaso (2010)). This novel nonparametric approach has been studied and validated using a Monte Carlo simulation study where we compared it with the parametric and nonparametric procedures proposed by literature. Results suggest that in each experimental situation where normality is hard to justify and especially when errors have heavy-tailed distribution, the proposed nonparametric procedure can be considered as a valid solution.

It is well known that a best permutation test for all population distributions P does not generally exist, because the most powerful unbiased permutation test is a function of the population distribution P which is assumed to be unknown (Pesarin (2001)). In this work we considered the Randomized Complete Block design in case of an ordered categorical response variable, which is the typical reference setting in many psychometric studies and we compared via a Monte Carlo simulation study several combination-based permutation test statistics (Arboretti, Corain and Ragazzi (2010)). We found out that the Multi-focus statistic (Pesarin and Salmaso (2010)) using the Fisher's combining function appears to be the more powerful solution which we proved also to be better under non normal errors than traditional parametric and rank-based nonparametric counterparts. Multivariate extensions of the combination-based test statistics for the comparison of

several treatments in the multivariate Randomized Complete Block designs are introduced (Arboretti, Corain and Ragazzi (2011)). Several tests for the multivariate Randomized Complete Block designs, including MANOVA procedure, are compared with the method proposed via a Monte Carlo simulation study. The method has also been applied to a real case study in the field of sensorial testing studies, the results suggest that in each experimental situation where normality is hard to justify and especially when errors have heavy-tailed distributions, the proposed nonparametric procedure can be considered as a valid solution.

1.2 Original contributions in the thesis

The following list concerns the original results developed during the Ph.D. thesis:

- After a work of formalization of the main concepts related to the orthogonal arrays, we discuss a method of construction based on the *counting polynomial functions* (Fontana et al. (2000); Pistone and Rogantin (2008)(a)). Through a number of recent papers, the original ideas have been developed into a technology based on symbolic computation software, which we wished to illustrate with reference to original works and typical examples. We propose the generation of several orthogonal arrays using these polynomial functions in order to evaluate the real applicability of the method purposed. Fractional factorial designs that satisfy a set of conditions in terms of orthogonality between factors have been described as the zero-set of a system of polynomial equations whose indeterminate are the complex coefficients of the counting polynomial functions. We present a list of orthogonal arrays generated using this method considering multifactorial designs with different number of factors' levels and strength with the aim to examine the adequacy and generality of the method proposed.

- The classes of nonisomorphic orthogonal arrays with a given size and strength have been investigated. In particular we propose the generation of classes of nonisomorphic orthogonal arrays with two and three-levels (Arboretti, Fontana and Ragazzi (2011)). The generation can be seen as an extension of a recent algebraic approach based on the concepts of *algebraic strata* (Fontana and Pistone (2011)). The generation problem is reduced into finding non-negative integer solutions to a system of linear equations while avoiding the computation with complex numbers. These nonisomorphic orthogonal arrays are used for the statistical analysis of unreplicated experiments.

- We propose a novel permutation solution for testing effects in two-levels and three-levels multifactorial unreplicated designs (Arboretti, Fontana and Ragazzi (2011)). This procedure is based on the Inequivalent Matrices Permuation Testing (IMPt) approach (Basso et al. (2004)). IMPt approach allows us to separately test all effects by exchanging matrices instead of permuting the vector of responses. The matrices

implemented are the novel classes of nonisomorphic orthogonal arrays constructed. In order to validate the IMPt procedures we perform a Monte Carlo simulation study for both two-level and three-level designs where the innovative procedures have been compared with the traditional counterpart. A modified $IMPt$ procedure has been introduced in case of a reduced number of nonisomorphic designs. This hybrid approach has been validate by a Monte Carlo simulation study.

- We have compared several combination-based permutation tests for testing hypotheses within the framework of Randomized Complete Block design (Arboretti, Corain and Ragazzi (2010)). The method is based on extensions of the Nonparametric Combination of Dependent Permutation tests (Pesarin (2001)). After a critical revision of the methods proposed in the literature, we compare such methods and the one proposed through an exhaustive simulation study in terms of power and under H_0. We refer to a typical psychometric study when using ordered categorical response variable within the framework of Randomized Complete Block design such as in the case of sensorial evaluations. We propose the iterated combination procedure: based on the results of the simulation study we find out that the proposed iterated combination-based permutation procedure is a very effective way to improve the power of permutation tests.

- In order to supply a complete study of this method, we also suggest a multivariate extension of the same (Arboretti, Corain and Ragazzi (2011)). In this case we focus on the categorical case and, in order to validate the purposed procedures, we perform a comparative simulation study with parametric and nonparametric competitors. Results suggest that in each experimental situation where normality of the supposed underlying continuous model is hard to justify and especially when errors have heavy-tailed distributions, the proposed nonparametric procedure can be considered as a valid solution.

- We propose an application of the proposed methods to a real case study: the research and development division of a famous detergent company wants to compare 5 potential fragrances of a certain detergent with the one that is currently being sold on the market. For this experiment we included 7 panelists that gave a score through a Likert scale on couple comparisons of the fragrances.

- For Split-plot designs, we proceed by giving a new method to test the possible effects of the main-plot factors (whole-plot) and their interactions in the permutation framework (Corain, Ragazzi and Salmaso (2010)). Given the structure of the design, permutation approach seems not be possible: in this experimental setting it is not possible to assume the exchangeability of the errors due to one exceeding error component inside the model. We then thought to arrange partial tests based on the symmetries of the blocks to then combine them through direct function combination. Based on the simulation studies these methods seem to be efficient in terms of power

and competitive when the errors seem to have symmetrical distributions.

- In order to test the effects of the sub-plot factors (split-plot) we proceeded by extending the method based on the synchronized permutations: according to the simulation study such methods appear to be a valid method in all those cases where it is not possible to assume normality of the components of errors and homogeneity of the variances.

Chapter 2

Orthogonal arrays

2.1 Theoretical background

A problem which often arises in the design of an experiment is that of ascertaining the effect of quantitative or qualitative alterations in the various components upon some measurable characteristic of the complete assembly. Consider for example the case of m fertilizers each of which can be applied at s different levels. To carry out a *full* factorial experiment would require s^m plot or *assemblies* to use a more general term applicable to any type of experiment, agronomic, industrial or otherwise.

Typically in design of experiments, the *levels* (or *symbols*) indicate the settings of the *factors* or variables whose effects on a response of interest are to be studied. The number of assemblies, or the number of level or treatment combinations are usually called *runs* (or observations) of the experiment. In what follows we refer to the terms *factors*, *levels* and *runs* of the design.

Full factorial designs involve all possible combinations of factors' levels so, if there are a large number of factors, there isn't even the possibility to obtain a response to each of the possible level combinations. Any collection of the level combinations does not use all possible combination is called a *Fractional Factorial* or simply a *Fraction*. Among fractional factorial designs *orthogonal arrays* (OAs) are of particular interest for their appreciable statistical properties. Rao (1947) has shown that if the design is an OA of strength t, then the estimates of main effects and interactions are unaffected by interactions of order greater than one and less than $t-1$, but the estimate of error is enhanced by their presence. If t is even, we can measure interactions up to order $t/2$ supposing the higher order interactions absent. When t is odd we can measure interactions up to order $(t-1)/2$.

Thus an array of appropriate strength must be chosen to handle those interactions which are deemed important.

Let us consider an experiment which includes m factors such that p_1 factors have s_1 levels, p_2 factors have s_2 levels ..... and p_m factors have s_m levels. Following the

definition of Hedayat et al. (1999) we can define an OA as follows:

Definition 1. *An* $OA(N, s_1^{p_1} \cdot \ldots \cdot s_m^{p_m}, t)$*, is an array of size* $N \times m$*, where* $m = p_1 + \ldots + p_m$ *is the total number of factors, in which the first* p_1 *columns have* s_1 *symbols, the next* p_2 *columns have* s_2 *symbols and so on, with the property that in any* $N \times t$ *subarray every possible* t*-tuple occurs* λ *times as a row.*

In general we assume that if the index (λ) of the OA is not specified then it is equal to one. The parameter t represents the strength of the OA. So, for example, the Table 2.1 (Hedayat et al. (1999)) represents an OA of size 12×11 with entries from $S = \{0, 1\}$ where $N = 12$ is the number of runs, $m = 11$ is the number of variables (factors), $s = 2$ is the number of levels (symbols) of each factor and $t = 2$ is the strength.

Strength equal two means that every 12×2 subarray of A contains each couple based on $S = \{0, 1\}$ exactly the same number of times (in this case three times).

In general the rows of the matrix (N) represent the tests that have to be performed and the columns (m) represent the different variables while the entries of the array specify the levels that the variables must assume.

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	1	1	0	1	0	0	0
0	1	1	1	0	1	1	0	1	0	0
0	0	1	1	1	0	1	1	0	1	0
0	0	0	1	1	1	0	1	1	0	1
1	0	0	0	1	1	1	0	1	1	0
0	1	0	0	0	1	1	1	0	1	1
1	0	1	0	0	0	1	1	1	0	1
1	1	0	1	0	0	0	1	1	1	0
0	1	1	0	1	0	0	0	1	1	1
1	0	1	1	0	1	0	0	0	1	1
1	1	0	1	1	0	1	0	0	0	1

Table 2.1: OA with 11 factors, two levels and strength two.

In Definition 1 we note that the factors can assume different number of levels, usually, these types of OAs, are known as *mixed* (or *asymmetrical*) OAs.

In this work we focus on the construction of *fixed* (or *symmetrical*) OAs where all the factors assume the same number of levels, but in this Section we present the general theory related to the bound parameters for their existence.

In what follows we refer to the terms *fixed* or *mixed*. We can define a *fixed* OA as in Definition 1 where $s_1 = s_2 = \ldots = s_m = s$. Let S be a set of s levels (or symbols), in general it has structure of finite field or a Galois field but it can have also different

structures (for example the complex field $\mathbb{C}$ - see Chapter 3). We refer to the values N, m, s and t as the *parameters* of a given OA.

Hedayat et al. (1999) showed that the following properties of OAs follow immediately from Definition 1:

1. $\lambda = \frac{N}{s^t}$;
2. if a matrix A is an OA of strength t then permuting the runs of A results an OA with the same parameters;
3. if a matrix A is an OA of strength t then permuting the levels of any factor results an OA with the same parameters;
4. if a matrix A is an orthogonal array of strength t and index λ then it is also an OAs of strength $t' < t$ with index $\lambda' = \lambda s^{t-t'}$;
5. let A be an $OA(N, m, s, t)$ then any $N \times m'$ subarray of A is an $OA(N, m', s, t')$ where $t' = \min\{m', t\}$;
6. if $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ is an $OA(N, m, s, t)$, and A_1 is an $OA(N_1, m, s, t_1)$ then A_2 is an $OA(N_2, m, s, t_2)$ where $N_2 = N - N_1$ and $t_2 \geq \min\{t, t_1\}$.

In real-life applications, it's important to use the OA which contains the minimum number of rows, because there are practical limitations on the number of the runs of the experiment, and on the other hand for a given number of runs we want the strength to be as large as possible; in many applications this is set equal to 2, 3 of 4.

The Rao's inequalities (Rao (1947) and Rao (1973)) explain the upper bounds for the parameters of an OA, in particular they show what is the minimum number of runs N necessary for its construction, given the number of factors, levels and strength. The following Propositions explain the Rao's inequality for fixed and mixed OAs.

Proposition 2. *The parameter of an $OA(N, m, s, t)$ satisfy the following inequalities:*

- $N \geq \sum_{k=0}^{i} \binom{n}{k} (s-1)^k$ *if* $t = 2i$;
- $N \geq \sum_{k=0}^{i} \binom{n}{k} (s-1)^k + \binom{n-1}{i} (s-1)^{i+1}$ *if* $t = 2i+1$.

Proposition 3. *The parameter of an $OA(N, s_1^{p_1 \cdot \ldots \cdot s_m^{p_m}}, t$ where $s_1 \leq \ldots \leq s_m$ satisfy the following inequalities:*

- $N \geq \sum_{u=0}^{i} \sum_{I_u(m)} \binom{n_1}{i_1} \cdot \ldots \cdot \binom{n_m}{i_m} (s_1-1)^{i_1} \cdot \ldots \cdot (s_m-1)^{i_m}$ *if* $t = 2i$;

- $N \geq \sum_{u=0}^{i} \sum_{I_u(m)} \binom{n_1}{i_1} \cdot \ldots \cdot \binom{n_m}{i_m} (s_1-1)^{i_1} \cdot \ldots \cdot (s_m-1)^{i_m} + \sum_{I_i(m)} \binom{n_1}{i_1} \cdot \ldots \cdot \binom{n_{m-1}}{i_{m-1}} \binom{n_m-1}{i_m} (s_1-1)^{i_1} \cdot \ldots \cdot (s_{m-1}-1)^{i_{m-1}} (s_m-1)^{i_{m+1}}$ *if* $t = 2i+1$*;*

where $I_u(m) = \{(i_1, \ldots, i_m) : i_1 \geq 0, \ldots, i_m \geq 0, \sum_{l=1}^{m} i_l = m\}$ *with* $u \geq 0$ *and* $m \geq 1$ *integers.*

We observe that the number N of runs required for the OA is often prohibitive. Suppose for example that it is required to find an OA with strength two ($t = 2$) with five factors ($m = 5$) in which two factors have four levels and three factors have three levels. By the Rao's inequalities we know that the dimension of the OA will be or $N = \text{lcm}(3 \times 3, 3 \times 4, 4 \times 4) = 144$ (where lcm notes the lowest common multiple) or a multiple of 144. This effort only allows to estimate 13 ($= 1 + 3 \times 2 + 2 \times 3$) parameters in the analysis model (excepted the variance).

So, an important question in the construction of OAs, is what the minimal size of an array is when all other parameters are fixed.

Among the first lower bounds on the size of an OA for fixed and mixed levels were those developed by Rao (Rao (1947) and Rao (1973)). Various other bounds have been developed since then; see Hedayat et al. (1999) and Collombier (1996) for a survey of different inequalities that are known thus far.

Bierbrauer (1995) found a lower bound on the runs of a fixed OA which he found to be better than Rao's for some values of the strength, recently, Diestelkamp (2004) provided an analogues inequality for OAs having mixed levels.

Here we recall the well known results related to the lower bound of fixed and mixed OAs which will be used throughout the work.

Bierbrauer (1995) provided the following bound on the size N of a fixed OA.

Theorem 4. *Assume that an* $OA(N, m, s, t)$ *exists, then:*

$$N \geq s^m \left(1 - \frac{(s-1)m}{s(t+1)}\right). \tag{2.1}$$

We observe that the inequality (2.1) is much easier to compute than Rao's bound and Diestalkamp (2004) provided a list of parameter combinations for $OA(N, m, s, t)$ for which the bound in (2.1) is sharper than Rao's; we also observe that the inequality (2.1) leads to a larger minimum value for N than Rao's. In particular Diestalkamp (2004) suggested that (2.1) yields a minimum value for N that is at least as good as Rao's whenever $s > 2$ and $t < m < \frac{t(s+1)}{s-1}$.

If we consider a mixed $OA(N, s_1 \cdot \ldots \cdot s_m, t)$ we observe that the size N of the array must be a multiple of $s_{i_1} \cdots s_{i_t}$ for any set $I = \{i_1, \ldots, i_t\} \subset \{1, \ldots, m\}$. N is a multiple of lcm, follows that:

$$N \geq \text{lcm}\left\{s_{i_1} \cdots s_{i_t} : 0 < i_1 < \ldots < i_t < m\right\}. \tag{2.2}$$

Proposition 5. *The dimension N of an $OA(N, s_1^{p_1} \cdot \ldots \cdot s_m^{p_m}, 2)$ is a multiple of the all products*

$$p_i \times p_j \quad \forall (i,j) \in \{1, \ldots, m\}^2 : i \neq j. \tag{2.3}$$

Proof. By definition of OA we know that in the matrix $N \times t$, of order $p_i \times p_j$, its elements are all equals and of sum N. This implies that N is necessary a multiple of $p_i \times p_j \quad \forall i \neq j$. □

The number of runs of an OA of strength $t = 3$ is given in the following proposition (due to Collombier (1996)).

Proposition 6. *Let us consider m factors with $s_1, \ldots, s_m$ levels, where $s_m = \max(\{s_i\})$. The dimension N of an OA of strength $t = 3$ verifies the following inequality:*

$$N \geq s_m(\sum_{i=1}^{m}(s_i - 1) - (s_m - 2)). \tag{2.4}$$

Proof. See Collombier (1996) pp 72 – 74. □

Considering fixed OAs we can rewrite the Proposition 6 in the following way:

Corollary 7. *The number of runs of an (fixed) OA of strength $t = 3$ verifies:*

$$N \geq s((m-1)(s-1)+1). \tag{2.5}$$

Proof. Consider $s_1 = \cdots = s_m = s$, using Proposition 6 we have that:

$$\begin{aligned} N \geq \quad & s(\textstyle\sum_{i=1}^{m}(s_i - 1) - (s_m - 2)) = \\ & = s(s_1 - 1 + \ldots + s_m - 1) - (s_m - 2)) = \\ & = s(m(s-1) - (s-2)) = s((m-1)(s-1)+1). \end{aligned} \tag{2.6}$$

□

From Corollary 7 follows that the number of factors m of an OA of strength $t = 3$ satisfies:

$$m \leq \frac{s^2 + N - 2}{s^2 - s}. \tag{2.7}$$

The index λ of a fixed orthogonal array is equal to:

$$\lambda = \frac{N}{s^t}, \tag{2.8}$$

so rewriting (2.7) in terms of λ, s and m we have that:

$$m \leq [\frac{\lambda s^2 - 1}{s - 1}] + 1, \tag{2.9}$$

where, with the notation $[x]$, we intend the integer part of a real x.

For the OAs of strength $t = 4$ the number of the runs is $N = \lambda s^4$ with λ integer, in particular:

$$N \geq 1 + m(s-1) + \frac{m(m-1)}{2}(s-1)^2, \tag{2.10}$$

this implies that:

$$m \leq [\frac{1}{2(s-1)}((s-3) + ((s-3)^2 + 8(N-1))^{1/2})]. \tag{2.11}$$

We observe that the lower bounds for the number of runs can be considered as a function in terms of m, s and t denoted by $f(m, s, t)$ and it is not completely determine of the maximal number of factors m in any OA, noted by $g(N, s, t)$.

The relation between these bounds can be explained as follows:

$$f(m, s, t) = \min \{N : g(N, s, t) \geq m\}, \tag{2.12}$$

$$g(N, s, t) \leq \max \{m : f(m, s, t) \leq N\}. \tag{2.13}$$

The generalization of the Rao's bound for mixed OAs can be found in Diestelkamp (2004) and in Brouwer et al. (2006).

Theorem 8. *Brouwer et al. (2006). Let $m \geq t \geq 1$ and assume an $OA(N, s_1 \cdot \ldots \cdot s_m, t)$ exists.*

- *if t is even, then:*

$$N \geq \sum_{j=0}^{t/2} \sum_{|I|=j} \prod_{i \in I} (s_i - 1), \tag{2.14}$$

- *if t is odd, then:*

$$N \geq 1 + \sum_{j=1}^{(t-1)/2} \sum_{|I|=j} \prod_{(} s_i - 1) + \max_j \left[(s_j - 1) \sum_{|I|=\frac{t-1}{2}, j \notin I} \prod_{i \in I} (s_i - 1) \right], \tag{2.15}$$

where $I = \{i_1, \ldots, i_t\} \subset \{1, \ldots, m\}$.

For a complete overview on the methods available for the construction of OAs we remind to Hedayat et al. (1999) and Collombier (1996); in this work we propose the construction of OAs (Chapter 3) based on a recent algebraic approach based on the polynomial counting functions (Fontana et al. (2000), Pistone and Rogantin (2008)(a)) in accordance with the previous Propositions and Theorems in order to quantify the parameters of the OA. In particular, in our constructions, we fixe the number of factors, levels and strength and we find the minimum number of runs according to (2.1) and Theorem 8, supposed that the researched OA exists.

2.2 Nonisomorphic orthogonal arrays

Two arrays are said to be combinatorially isomorphic if one array can be obtained by permuting rows, columns, or factor levels of the other array (see e.g. Hedayat et al. (1999)).

It would be highly desirable to have an enumeration method obtaining, for given parameters, a minimum complete set of nonisomorphic *OAs*. Such a set has a single representative for each isomorphism class.

If the factors are all qualitative, one would then have to choose the best array according to some optimality criterion, assign the factors to the columns such that prior knowledge on the factor's activities is incorporated as good as is feasible, and to assign the factor levels to the symbols in a column at random. For qualitative factors at more than two levels, one would have to consider additionally the non-equivalent ways to assign factor levels to the symbols, because the levels now are ordered.

When more than one OA are available for the factorial experiment, the experiment can use a suitable criterion to choose among the candidate arrays. The most commonly used criteria are those of resolution and aberration, along with their generalization (Hedayat et al. (1999)).

An essential element in establishing a minimum complete set of *OAs* is in proving that all of the arrays are indeed *nonisomorphic* to each other.

Definition 9. *Two orthogonal arrays are said isomorphic if one can be obtained from the other by a sequence of permutations of the columns, the rows and the levels of each factor.*

In Arboretti, Fontana and Ragazzi (2011) has been explored the use of nonisomorphic *OAs* for nonparametric testing of factorial effects. In Chapter 4 of this thesis we report a part of it.

A special class of *OAs* are those with parameters $OA(N, m, 2, t)$ which define orthogonal factorial designs with N runs, m two-level columns and strength $t \geq 2$. These *OAs* have received great interest in the last decade (see Cheng (1995)) and many authors have contributed to the investigation and identification of the full list of nonisomorphic *OAs* with the same parameters.

Evangelaras et al. (2007)(a) contributed to this problem by providing the full list of nonisomorphic orthogonal arrays when $12 \leq N \leq 24$ and $3 \leq m \leq 6$ when $28 \leq N \leq 40$ and $3 \leq m \leq 5$ as well as those with $44 \leq N \leq 64$ and $3 \leq m \leq 4$. However, orthogonal arrays with three levels have not been sufficiently explored in this manner. Evangelaras et al. (2007)(b) define a class of nonisomorphic three level orthogonal arrays with 18 runs and $3 \leq m \leq 7$. Recently Schoen et al. (2010) provided an algorithm to enumerate a minimum complete set of combinatorially nonisomorphic orthogonal arrays of given strength t, run-size N and level-numbers of the factors. This allowed the generation of an extremely rich library of orthogonal arrays.

The main use of nonisomorphic *OAs* consists into identify a simple method for the enumeration of *OAs* of a given size, as all arrays can easily be generated once all non-

isomorphic arrays are available. Despite the importance of the (very difficult) problem of the *OAs*' enumeration (see for example Stufken and Tang (2007)), in this work we suggest to use the nonisomorphic *OAs* for the statistical analysis of *unreplicated* experiments, i.e. there is a single observation for level combination. The statistical analysis of unreplicated designs presents a challenge. This is because we can estimate all the effects but there are no degree of freedom left to estimate the error variance.

In this context many solutions are proposed in the literature, but they are only for two-level fractional factorial design; in Chapter 4 we present a novel solution for the analysis of unreplicated factorial designs within the framework of the permutation tests using the nonisomorphic *OAs*.

The key idea is to construct nonisomorphic orthogonal arrays (Chapter 3) in case of two-level and three-level multifactorial experiments using algebraic *strata*, then find the approximate permutation distribution of the test statistic (adopted for testing active effects) exchanging nonisomorphic arrays instead of permuting the vector of responses (Basso et al. (2004)). This approach usually extend the usual concept of *permutation*.

The concept of *strata* for the generation of fractional factorial design was introduced in Fontana and Pistone (2010); in this way the generation problem is reduced into finding non-negative integer solutions to a system of linear equations. The method seems to be a valid alternative because it's possible to take in consideration any number of factors' levels. In the cited paper a list of nonisomorphic *OAs* is presented.

2.3 Orthogonal arrays as fractional factorial designs

The importance of the *OAs* in statistical analysis regards the possibility to estimate low-order interactions between factors. Suppose for example an experiment which includes m factors, noted by $A_1, \ldots, A_m$ with s levels, the level combinations can be represented by the m-tuples $(\alpha_1, \ldots, \alpha_m)$ where, using the integer coding of factor levels, $0 \leq \alpha_j \leq s-1$, with $j = 1, \ldots, m$.

The set L of the all level combination is:

$$L = \{\alpha = (\alpha_1, \ldots, \alpha_m) : \alpha_j = 0, \ldots, n_j - 1,\ j = 1, \ldots, m\}; \tag{2.16}$$

and $M = \#L = s^m$ is the number of elements of L or *cardinality* of L.

The related statistical model includes the main and interaction effects of order l such that $l \leq s^m$. Considering N experimental units which are made available for the experiment, and each unit will be assigned to a particular level of combination α. By r_α we denote the number of experimental units which are assigned to level combination α.

The statistical model can be represented in the following way:

$$Y_{\alpha k} = \mu_\alpha + \epsilon_{\alpha k} \text{ with } k = 1, \ldots, r_\alpha,\ \alpha \in L, \tag{2.17}$$

where $Y_{\alpha k}$ denotes the random variable corresponding to the response of the k-th unit that is assigned to level combination α, μ_α denotes a nonrandom population mean for possible

observations with level combination α, and $\epsilon_{\alpha k}$ denotes a nonobservable random deviation from the population mean μ_α for the k-th unit that receives α.

We assume that the $\epsilon_{\alpha k}$'s (which are also referred to as random errors), have mean 0 and constant variance σ^2 and we will henceforth also assume that the $Y_{\alpha k}$'s are uncorrelated.

We observe that if systematic differences among the experimental units are known or suspected, it might be useful to group the units into *blocks* of similar units. This implies, in general, a modification of the commonly used method of randomization for assigning units to level combinations and this would also result in an additional term in model 2.17 to account for possible differences between blocks. We discuss the blocking in Chapters 5 and 6 referring to *Randomized Complete Block* (RCB) and *Split-Plot* (SP) designs, we focus on the inference procedures for testing effects in these designs and we propose new nonparametric statistical methods for testing (active) effects (Arboretti, Corain, Ragazzi (2010) and Corain, Ragazzi, Salmaso (2010)) within the framework of the nonparametric combination of permutation tests (Pesarin (2001)).

The model (2.17) can be represented in matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\epsilon}, \tag{2.18}$$

where X is an $N \times M$ matrix of zeros and ones whose entries are obtained as follows: we label the columns by the level combinations, in the *standard order* (precedence of one level combination over another is decided on a lexicographic basis with the primary criterion being the level of the last factor, with ties in the level of the last factor being decided by the next-to-last factor and so on) and the rows by the subscripts of the corresponding entries in Y; the entry in position $(\alpha k, \alpha')$ is equal to one if $\alpha = \alpha'$.

If there are a large number of factors, the value of M can be enormous, so the use of fractional factorial designs becomes essential.

Definition 10. *(Pistone et al. (1996)) Let us consider a design that involves m factors, noted by $A_1, \ldots, A_m$, so that any factor A_j has s_j levels. The full factorial design can be defined as the cartesian product of factors:*

$$\mathcal{D} = A_1 \times \ldots \times A_m. \tag{2.19}$$

Definition 11. *A fraction $\mathcal{F}$ is a multiset of $\mathcal{D}$ ($\mathcal{F}$ could be any subset of $\mathcal{D}$ without any restriction on its run size).*

In Chapter 3 we discuss the algebraic description of the full factorial and fractional factorial designs and we propose a method based on particular *polynomial functions* for the construction of the orthogonal arrays.

In general we can define a fraction of resolution R in the following way:

Definition 12. *A fraction factorial design is of resolution R if all factorial effects up to order $\lfloor (R-1)/2 \rfloor$ are estimable, under the assumption that all factorial effects of order $\lfloor (R+1)/2 \rfloor$ and higher are zero, with the additional convention that if R is odd then the general mean is also estimable and if R is even then the general mean is not of interest for estimation.*

The notion of resolution and strength are related. The relation between orthogonal arrays and fractional factorial designs can be resumed in the following Proposition:

Proposition 13. *Any Orthogonal Array specifies a (fractional) factorial designs and the orthogonal arrays of strength t are the only fractions of resolution $t + 1$.*

The fractions give the possibility of confounding the estimation of effects of single terms with the effects of interactions terms, the following list concern the properties of the fractions of resolution $R = t + 1$:

- Fractions of resolution III ($t = 2$): no main effect is aliased with the other main effects. Main effects are aliased with interaction effects of order two and interaction effects are aliased between each others;
- Fractions of resolution IV ($t = 3$): no main effect is aliased with the other main effects. The interaction effects of order two are aliased between each others;
- Fractions of resolution V ($t = 4$): no main effect is aliased with the other main effects or with interaction effects of order two. The interaction effects of order two are aliased between interaction effects of order three.

We can affirm that the importance of the OAs cannot be overstated. They are directly useful as fractional factorial in factorial experiments arising from scientific and technological investigations. They are intimately related to many other combinatorial objects such as Hadamard matrices, orthogonal Latin squares and error-correcting codes.

The most fundamental questions concerning OAs are when they exist and how to construct them if they do exist. Indeed, these are the two main themes of Hedayat et al. (1999). A comprehensive discussion on the use of orthogonal arrays as fractional factorial designs is given by Wu and Hamada (2000).

In this work we contribute to the construction of OAs using the *counting polynomial functions* (Fontana et al. (2000) and Pistone and Rogantin (2008)(a)) and we use recent algebraic approach (based on the concept of *algebraic strata* (Fontana and Pistone (2010)) in order to construct nonisomorphic OAs for a given size and strength.

We introduce also a 'novel' application of the nonisomorphic OAs: we use these arrays in order to give a novel permutation approach for statistical inference in unreplicated fractional factorial designs (Arboretti, Fontana and Ragazzi (2011)).

Chapter 3

Generation of orthogonal arrays

3.1 Introduction

In this Chapter we review the basic facts about the algebraic study of full factorial and fractional factorial designs and level coding. A general reference to the main concepts of algebra and methods of algebraic statistics for designing experiments we use is Kreuzer and Robbiano (2000). Though a number of recent papers, the original ideas have developed into a technology based on symbolic computation software, which we wish to illustrate with reference to original works and typical examples.

We suggest two algebraic methods for the construction of OAs: the first is based on the *counting polynomial functions* (we mainly refer to Fontana et al. (2000) and Pistone and Rogantin (2008)(a)) and the second is based on the concepts of *algebraic strata* (Fontana and Pistone (2010)).

Using counting polynomial functions we present the generation of several fixed orthogonal arrays (Section 3.3) for a given size N, considering different values of strength t and of levels' factors. This approach holds for a wide class of designs: regular fractional factorial, mixed-level fractional factorials, Sudoku designs, fixed and mixed orthogonal arrays. The main advantage of the proposed method is the possibility to consider different numbers of factors without restrictions (such as prime or prime power) on the number of the levels. Using this approach the fractional factorial design consists in the set of solutions of counting polynomial functions that satisfy a set of orthogonality conditions, implying computation with complex numbers. In order to avoid this computational problem we use a second method based on the *strata*: in this way the problem of finding fractional factorial designs that satisfy a set of orthogonality conditions becomes the problem of finding non-negative integer solutions to a system of linear equations. Using *algebraic strata* we generate the class of all the OAs of a given size and strength, then, from this class we extract all the nonisomorphic OAs.

The present Chapter is organized as follows. Section 3.2 is devoted to the formalization of the algebraic descriptions of full factorial and fractional factorial designs with focus on

the level coding. In Section 3.3 we present the generation of several fixed OAs using counting polynomial functions while in Section 3.4 we present the formalization of the algebraic method based on the strata utilized for the generation of the nonisomorphic OAs (Arboretti, Fontana and Ragazzi (2011)). Finally, Section is dedicated to conclusions and final remarks.

3.2 Algebraic description of full factorial and fractional factorial designs

Let us consider an experiment which includes m factors, $A_1, \ldots, A_m$, where each factor A_j has s_j levels for $j = 1, \ldots, m$. The full factorial design $\mathcal{D}$ can be defined as the algebraic variety:

$$\mathcal{D} = \{x_j \mid d_j(x_j) = 0 \ \forall j = 1, \ldots, m\}, \tag{3.1}$$

where $d_1(x_1), \ldots, d_m(x_m)$ are the univariate polynomials in the polynomial ring on the field k in m variables containing $\mathbb{Q}$ noted by $k[x_1, \ldots, x_m]$ and x_j, $i = 1, \ldots, s_j$; $j = 1, \ldots, m$ are the solutions of each equation $d_j(x_j)$ which are all distinct and they belong to an extension K of the field k.

As each equation contains one single indeterminate, the solution set of the system is the Cartesian product of each set of solutions (Definition 10, Chapter 2). The *ideal* of $\mathcal{D}$ is the ideal generated by the polynomials d_j given by the set of all the polynomials which vanishes on $\mathcal{D}$. We refer to this ideal as the ideal of the full factorial design, noted by $I(\mathcal{D})$.

Each polynomial $f \in I(\mathcal{D})$ is of the form $f = \sum_{j=1}^{m} h_j d_j$ where $h_j \in k[x_1, \ldots, x_m]$, $j = 1, \ldots, m$.

The code of factor levels, in general, has made with numbers (especially when the levels are ordered), classical examples of level coding are:

1. *Integer coding (a)*
   The levels takes values in the additive group $\mathbb{Z}_{s_j}$: $0, 1, \ldots, s_j - 1$ then

   $$d_j(x) = x(x-1)(x-2) \cdot \ldots \cdot (x - (s_j - 1)), \tag{3.2}$$

   in this case $k = K = \mathbb{Q}$.

2. *Integer coding (b)*
   The levels l_{ij} with $i = 1, \ldots, s_j$ and $j = 1, \ldots, m$ take values in the following way:

   $$l_{ij} = (2i - s_j - 1)/2 \ \text{ if } \ s_j \ \text{ odd},$$
   $$l_{ij} = 2i - s_j - 1 \ \text{ if } \ s_j \ \text{ even}.$$

   So for example in a 3^m factorial design we have that:

$$d_j(x) = x(x^2 - 1), \tag{3.3}$$

and $k = K = \mathbb{Q}$.

3. *Complex coding*
   If the levels are the s_jth roots of the unity then:

$$d_j(x) = x^{s_j} - 1, \tag{3.4}$$

and $k = \mathbb{Q}$. K is an extension of $\mathbb{Q}$ containing the field $\mathbb{Q}[x]/I(\mathcal{D})$ (as for example $\mathbb{C}$).

Using the first coding we define the class of *Regular Fractions* (Fontana et al. (2000)) when $s_j = s$ and s is a prime. Regular fractions are clearly defined and have been widely studied. Such regular fractions are characterized by their simple confounding structure and defining relations. Wu and Hamada (2000) call a design regular if any two factorial effects either estimated independently or are completely confounded.

The general setting where factors assume more than two levels it requires precise specification of the factorial effects under consideration. However the definitions for regular fractions for the case of mixed-level factorial designs have only recently appeared in the literature (Pistone and Rogantin (2008) (a)).

We observe that the study of regular fractions is connected with the study of OAs, in fact the Theorem 8.2.2 of Dey and Mukerjee (1999) states that each regular fraction is an OA of a strength that is related to its solution. We observe that Dey and Mukerjee (1999) restricted the attention only for fixed fraction with a number of levels equal to prime power. Van De Ven and Bucchianico (2009) extended the definition of regular fraction for mixed fractions and for number of level not equal to a prime power.

The second coding is the result of the orthogonalization of the linear term with respect to the constant term. Fontana et al. (2000), Ye (2003) and Tang and Deng (1999) use the coding -1 and $+1$ for binary factors in the multiplicative group isomorphic to $\mathbb{Z}_2$.

The complex code of factors' levels was introduced for the first time in Bayley (1983). Pistone and Rogantin (2008)(a) discussed the complex code of factors' levels. Here, we provide the methodology adopted only when $s_j = s$ and refer to the cited paper for the general theory.

The s levels of a factor are the complex solutions of the equation $\zeta^s = 1$:

$$\omega_h^{(s)} = \exp\left(i\frac{2\pi}{s}k\right), \quad k = 0, \ldots, s-1. \tag{3.5}$$

We observe that the complex roots of units are all distinct (this is because if we consider the complex polynomial $f(\zeta) = \zeta^s - 1$ for $s \geq 2$ we have that the derivative polynomial $f'(\zeta) = s\zeta^{s-1}$ has only one root that is zero, this implies that $f(\zeta)$ and $f'(\zeta)$ don't have common roots hence all the roots are distinct) and $\left(\omega_1^{(s)}\right)^k = \omega_k^{(s)}$.

We denote such a factor with s levels by $\Omega_s = \left\{\omega_0^{(s)}, \ldots, \omega_{s-1}^{(s)}\right\}$ where Ω_s denotes the multiplicative group consisting of the sth complex roots of units; the full factorial design in complex code will be:

$$\mathcal{D} = \Omega_{s_1} \times \ldots \times \Omega_{s_m}. \tag{3.6}$$

The full factorial design L in integer coding (Definition 10, Chapter 2) and the full factorial $\mathcal{D}$ are in turn finite Abelian groups under the operations of elementwise addition and multiplication respectively. The following Proposition asserts the isomorphic relations between the groups $\mathbb{Z}_s$ and Ω_s.

Proposition 14. *A function ϕ such that:*

$$\begin{array}{rccc} \phi: & \mathbb{Z}_s & \leftrightarrow & \Omega_s \\ & k & \leftrightarrow & \omega_k, \end{array} \tag{3.7}$$

is a group isomorphism of the additive group of $\mathbb{Z}_s$ on the multiplicative group Ω_s.

We observe that ϕ is an isomorphism between cyclic groups of the same order s. Using the isomorphism (3.7) the set of the all level combination L (2.16) is both the full factorial design with integer coding and the exponent set of the complex coded design $\mathcal{D}$ and also any element of L is both a treatment combination in the integer coding and a multi-exponent of an interaction term.

Example 15. *Let us consider an experiment which includes three factors with two levels. The full factorial design using the integer coding is:*

$$L = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\},$$

using the complex coding the full factorial design is the group:
$\mathcal{D} = \{(1,1,1),(1,1,-1),(1,-1,1),(1,-1,-1),(-1,1,1),(-1,1,-1),(-1,-1,1),(-1,-1,-1)\}$.

Example 16. *Consider an $2^3 4^1$ factorial design, the full factorial in integer coding (Table 3.1) has 32 runs coded by the element in the set $L = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4$:*
Using the complex code of the factors' levels (Table 3.2), the runs are coded by the element in the set $\mathcal{D} = \Omega_2 \times \Omega_2 \times \Omega_2 \times \Omega_4$ where $\Omega_2 = \left\{\omega_0^{(2)}, \omega_1^{(2)}\right\} = \{1, -1\}$ and
$\Omega_4 = \left\{\omega_0^{(4)}, \omega_1^{(4)}, \omega_2^{(4)}, \omega_3^{(4)}\right\} = \{1, i(= i \operatorname{sen} \pi/2), -1(= \cos \pi), -i(= i \operatorname{sen} 3/2\pi)\}$. *So the elements of the group $\mathcal{D}$ (Table 3.2) represent the runs of the experiment.*

Using the complex code of factors' levels the full factorial design $\mathcal{D}$ coincides with the set of solutions of the polynomial equations:

$$f_j(\zeta_j) = \zeta_j^s - 1 = 0, \ j = 1, \ldots, m; \tag{3.8}$$

A_1	A_2	A_3	A_4
0	0	0	0
0	0	0	1
0	0	0	2
0	0	0	3
0	0	1	0
0	0	1	1
0	0	1	2
0	0	1	3
⋮	⋮	⋮	⋮
1	1	1	3

Table 3.1: A $2^3 4^1$ full factorial design in integer coding.

A_1	A_2	A_3	A_4
1	1	1	1
1	1	1	i
1	1	1	-1
1	1	1	$-i$
1	1	-1	1
1	1	-1	i
1	1	-1	-1
1	1	-1	$-i$
⋮	⋮	⋮	⋮
-1	-1	-1	$-i$

Table 3.2: A $2^3 4^1$ full factorial design in complex coding.

where ζ_j is a point of the full factorial design in complex code. Such a design allows the identification of all polynomial responses of the form:

$$\sum_{\alpha \in L} \theta_\alpha X^\alpha, \tag{3.9}$$

with $X^\alpha = X_1^{\alpha_1} \dots X_m^{\alpha_m}$ $\theta_\alpha \in \mathbb{C}$.

In this way we can consider a complex response f as a function defined on $\mathcal{D}$ with values in $\mathbb{C}$ and considered as the restriction to $\mathcal{D}$ of a complex polynomial.

A factor X_j (or simple term) is defined as:

$$X_j : \mathcal{D} \ni (\zeta_1, \dots, \zeta_m) \longmapsto \zeta_i, \tag{3.10}$$

the interaction term $X_1^{\alpha_1} \times \ldots \times X_m^{\alpha_m}$ is the function:

$$X^\alpha : \mathcal{D} \ni (\zeta_1, \ldots, \zeta_m) \longmapsto \zeta_1^{\alpha_1} \ldots \zeta_m^{\alpha_m}. \tag{3.11}$$

The function X^α is a response called monomial or interaction term, and we say that X^α has order of interaction equal to k if k factors are involved (if the m-tuple α has k non-null entries).

The set of all monomial responses in the model (3.9) is noted by:

$$Est(\mathcal{D}) = \{X^\alpha : \alpha \in L\}, \tag{3.12}$$

and it is an orthonormal basis of the complex responses on the design $\mathcal{D}$.

If f is a response defined on $\mathcal{D}$ then its mean value is:

$$E_{\mathcal{D}}(f) = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{D}} f(\zeta). \tag{3.13}$$

Definition 17. *The set $\mathcal{C}(\mathcal{D})$ of the all complex responses is a complex Hilbert space with the Hermitian product $f \cdot g = E_{\mathcal{D}}(f\overline{g})$.*

$$\mathcal{C}(\mathcal{D}) = \left\{ \sum_{\alpha \in L} \theta_\alpha X^\alpha : \theta_\alpha \in \mathbb{C} \right\}, \tag{3.14}$$

where $\overline{g}$ represents the conjugate of the response g:

$$\overline{g(\zeta)} = \sum_{\alpha \in L} \overline{\theta_\alpha X^\alpha}(\zeta) = \sum_{\alpha \in L} \overline{\theta_{[-\alpha]}} X^\alpha(\zeta). \tag{3.15}$$

Pistone and Rogantin (2008)(a) showed that the coefficients of each response f are uniquely defined by:

$$\theta_\alpha = E_{\mathcal{D}}(f\overline{X^\alpha}). \tag{3.16}$$

We can describe a fractional factorial design by its Counting Function, or if it is a single replicate fraction, by its Indicator Function (introduced by Fontana et al. (1997)) (we refer to them as counting polynomial functions).

Definition 18. *The Indicator Function F of a fraction $\mathcal{F}$ is a response defined on $\mathcal{D}$ such that:*

$$F(\zeta) = 1 \text{ if } \zeta \in \mathcal{F} \text{ and } F(\zeta) = 0 \text{ if } \zeta \in (\mathcal{D} \setminus \mathcal{F}). \tag{3.17}$$

Using the theory of Gröbner basis and algebraic geometry (Kreuzer and Robbiano (2000)), Fontana el al. (2000) show that each indicator function has a unique polynomial representation:

$$F(\zeta) = \sum_{\alpha \in L} b_\alpha X^\alpha(\zeta) \quad \zeta \in \mathcal{D} \quad b_\alpha \in \mathbb{C}, \tag{3.18}$$

where b_α are the coefficients of the representation of F on $\mathcal{D}$.

Definition 19. *The Counting Function R of a fraction $\mathcal{F}$ with replicates is a response defined on $\mathcal{D}$ such that for each $\zeta \in \mathcal{D}$, $R(\zeta)$ equals the number of appearances of ζ in the fraction. The coefficients of the representation of R on $\mathcal{D}$ are denoted by c_α.*

$$R(\zeta) = \sum_{\alpha \in L} c_\alpha X^\alpha(\zeta) \quad \zeta \in \mathcal{D}. \tag{3.19}$$

A fraction is fully described by its indicator or counting function. The design with replicates associated to a counting function can be considered a multi-subset $\mathcal{F}$ of the design $\mathcal{D}$, or an array with repeated rows. If R is a polynomial function then R is a counting function of a fraction with replicates up to r if and only if $R(R-1)\dots(R-r) = 0$ on $\mathcal{D}$.

A function F is an indicator function if and only if $F^2 - F = 0$, and $F - 1 = 0$ is a generating equation of the fraction $\mathcal{F}$. The indicator function and the counting function are real valued so $\overline{b_\alpha} = b_{[-\alpha]}$ and $\overline{c_\alpha} = c_{[-\alpha]}$. The coefficients b_α and c_α are:

$$\frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{F}} \overline{X^\alpha(\zeta)}, \tag{3.20}$$

in particular b_0 and c_0 are the ratio between the number of points of the fraction and those of the full design.

A response f is centered on a fraction if the mean value of f on the fraction $\mathcal{F}$ is zero, where the mean value of a fraction f on the fraction $\mathcal{F}$ is defined as:

$$E_{\mathcal{F}}(f) = \frac{1}{\#\mathcal{F}} \sum_{\zeta \in \mathcal{F}} f(\zeta). \tag{3.21}$$

We say that two responses f and g are orthogonal if the mean value of $f\overline{g}$ on the fraction is zero. For *orthogonal* here we refer to vector orthogonality and not to the factor orthogonality, but for the choice of the complex coding the two orthogonalities are essentially equivalent (for more detail we refer to Pistone and Rogantin (2008)(a)).

3.3 Construction of orthogonal arrays using counting polynomial functions

In this Section we present the *OAs* in relation to the counting polynomial functions. The aim of our present discussion is to give the theoretical supports for the generation of *OAs* based on the counting polynomial models. The basic idea is that a fraction can algebraically described using its counting polynomial function. This device allows a full generalization of the mixed and fixed orthogonal arrays. The relation between *OAs* and indicator polynomial functions has already been introduced in Fontana et al. (2000), Pistone and Rogantin (2008)(a), in this Section we propose the generation of orthogonal

arrays using extensively symbolic algebra software, especially Mathematica. This one and other similar software systems can perform exact computations on computable number field.

Let us consider now I a non-empty set of $\{1,\ldots,m\}$ and J its complement set then $\mathcal{D} = \mathcal{D}_I \times \mathcal{D}_J$ where $\mathcal{D}_I$ is the full factorial design over I and $\mathcal{D}_J$ is the full factorial design over J. If $\mathcal{F}$ is a fraction of $\mathcal{D}$ then $\mathcal{F}_\mathcal{I} \times \mathcal{F}_\mathcal{J}$ will be its projections.

1. A fraction $\mathcal{F}$ fully projects on the I-factors, if the projection $\mathcal{F}_I$ is a full factorial design where each point is replicate the same number s of times.

2. A complex code fraction $\mathcal{F}$ is a mixed orthogonal array of strength t if it fully projects on any I-factors with $\#I = t$.

Consider now the counting function R of the fraction $\mathcal{F}$. The follow proposition characterizes the counting function of an OA of strength t.

Proposition 20. *A fraction $\mathcal{F}$ is an orthogonal array of strength t if and only if all the coefficients of its counting function up to order t are zero.*

Here we focus on the orthogonal arrays without replicates, i.e there is only one observation for any rows of the design matrix. We can describe an orthogonal array without replicates with its indicator polynomial function. We remember that if F is the indicator function of a fraction $\mathcal{F}$, b_α are the coefficients of the representation of F on $\mathcal{D}$ and F can be represented as:

$$F(\zeta) = \sum_{\alpha \in L} b_\alpha X^\alpha(\zeta) \ \zeta \in \mathcal{D}, \ b_\alpha \in \mathbb{C}. \tag{3.22}$$

The following Proposition summarize theoretical results inherent to the coefficients of the counting polynomial functions formally presented in Pistone and Rogantin (2008)(a).

Proposition 21. *The coefficients b_α and c_α of the indicator function F and the counting function R are:*

$$b_\alpha = c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{F}} \overline{X^\alpha(\zeta)}, \tag{3.23}$$

so b_0 and c_0 are the ratio between the number of points of the fraction and those of the design. In a fraction without replicates the coefficients of F are:

$$b_\alpha = \sum_{\beta \in L} b_\beta b_{[\alpha-\beta]}. \tag{3.24}$$

Proof. We proof (3.23).

$$\sum_{\zeta \in \mathcal{F}} \overline{X^\alpha(\zeta)} \sum_{\zeta \in \mathcal{D}} R\overline{X^\alpha(\zeta)} = \sum_{\zeta \in \mathcal{D}} \sum_{\beta \in L} c_\beta X^\beta(\zeta) \overline{X^\alpha(\zeta)} \sum_{\zeta \in \mathcal{D}} c_\alpha = \#\mathcal{D}. \tag{3.25}$$

The proof of (3.24) follows from $F = F^2$:

$$\sum_{\alpha} b_\alpha X^\alpha = \sum_{\beta} b_\beta X^\beta \sum_{\gamma} b_\gamma X^\gamma = \sum_{\alpha} \sum_{[\beta+\gamma]=\alpha} b_\beta b_\gamma X^\alpha = \sum_{\alpha} \sum_{\beta} b_\beta b_{[\alpha-\beta]} X^\alpha. \qquad (3.26)$$

□

Considering for example a full factorial design including three factors with two level, an orthogonal array of strength $t = 2$ and $N = 4$ runs is the matrix: where the s levels

A_1	A_2	A_3
1	1	1
1	-1	-1
-1	1	-1
-1	-1	1

Table 3.3: An orthogonal array of strength two

are coded with the two roots of the unit. In this simple case the monomial responses are 1, X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3 $X_1X_2X_3$ and L (the exponent set) is:
$\{(0,0,0),(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}$.

If we consider a fraction $\mathcal{F}$ given by the points of the orthogonal array described in Table 3.3, all monomial responses on $\mathcal{F}$ and their values on the points are in Table 3.4:

ζ	1	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$
$(1,1,1)$	1	1	1	1	1	1	1	1
$(1,-1,-1)$	1	1	-1	-1	-1	-1	1	1
$(-1,1,-1)$	1	-1	1	-1	-1	1	-1	1
$(-1,-1,1)$	1	-1	-1	1	1	-1	-1	1

Table 3.4: Example of an orthogonal array

The counting function associated to the orthogonal array in Table 3.3 has the coefficients all equal zero up to order two and it is equal to:

$$R = \frac{1}{4}(1 + X_1X_2X_3).$$

In this work we used counting polynomial functions in order to construct fixed orthogonal arrays, according the lower bound for the OA's size explained in the previous Chapter.

A simple algorithm to generate orthogonal arrays can be summarize in the following steps:

1. Compute the lower bound N according to (8).
2. Given the levels $\{0, 1, \ldots, s-1\}$ and the number of factors m, compute the exponent set L. This exponent set has size (n, m).
3. For $i = 1$ to N, we map b to any row of L.
4. For $i = 2$ to n, we write the system of equations on b_α in accordance with (3.24): $b_i = \sum_{j \neq i \in L} b_i b_{[j-i]}$. We label any equation by $g[i]$ with $i = 0, \ldots, n-1$.
5. Put $b_0 = \frac{N}{n}$;
6. Put the orthogonality conditions: $b_i = 0$ for any i of order up to t, $i \neq (0, \ldots, 0)$;
7. Put the polynomial equation equal to zero.
8. Sample a solution among the set of solutions.

Step 2. concerns the computation of the exponent set that is the full factorial design in integer coding. We observe that not all symbolic softwares provide a specific instruction to generate all possible combination of elements from a list. In Mathematica step 1. and 2. can be easily done using the following function:

```
L = Tuples[b[0,1,2,3,...,s-1],m];
```

The elements of `b[0,1,2,3,...,s-1]` are treated as distinct, so that the function `Tuples` for a list of length s gives output of length s^m. As first remark we highlight the problem of the information overload, in fact the dimension of L is often prohibitive in case of a large number of factors and levels. The other steps of the algorithm can easily implement in Mathematica (and in all symbolic software) with particular attention to Steps 8., it attempts to solve a set of polynomial equations with the orthogonality conditions as expressed in 6. In this case we suggest to use the following Mathematica instruction:

```
Solve[{g[0]==0,g[1]==0,...,g[n-1]==0,b[0,...,0]==N/n},
      {b[0,...,0],...,b[s-1,...,s-1]}];
```

The function `Solve[eqns,vars]` attempts to solve the set of equations `eqns` (in our case `g[0]==0,...,g[n-1]==0,b[0,...,0]==0`) for the variables `vars` (in our case `b[0,...,0],...,b[s-1,...,s-1]`); in this way we find the set of all solutions of the polynomial equations implying computational time overload. So we suggest to use the following function:

```
FindInstance[g[0]==0 && g[1]==0 && ...&& g[n-1]==0 && b[0,...,0]==N/n,
     {b[0,...,0],...,b[s-1,...,s-1]}];
```

Using `FindInstance[eqns,vars]` we find a solution instance of a system of equations. The following list concerns the orthogonal arrays constructed using the counting polynomial functions.

- Orthogonal Arrays of the form $OA(N, s^m, 2)$ with $s = 2, 3$ and 4
  - $OA(4, 3, 2, 2)$
  - $OA(8, 7, 2, 2)$
  - $OA(9, 4, 3, 2)$
  - $OA(16, 5, 4, 2)$
- Orthogonal Arrays of the form $OA(N, s^m, 3)$ with $s = 2$ and 3
  - $OA(8, 4, 2, 3)$
  - $OA(16, 8, 2, 3)$
  - $OA(54, 5, 3, 3)$

3.4 Two-level and three-level nonisomorphic orthogonal arrays using strata

In this Section we propose the generation of nonisomorphic two-level and three-level orthogonal arrays using a recent algebraic method based on *strata*.

We mainly refer to Fontana and Pistone (2010). In that paper the general case of mixed level fractional factorial designs with no restriction on the number of levels of each factor is analysed. In the next Sections we will work with orthogonal arrays with $s = 2$ or $s = 3$ levels, so, for this reason, we only describe the methodology for fractional factorial designs where all the factors have the same number s of levels and s is a prime number.

Let us consider an experiment which includes m factors with s level, where any factor A_j, $j = 1, \ldots, m$ is coded with the s-th roots of the unity, so that $A_j \equiv \Omega_s = \{\omega_0, \ldots, \omega_{s-1}\}$, $\omega_k = exp(\sqrt{-1}\frac{2\pi}{s}k)$, $k = 0, \ldots, s-1$. The full factorial design $\mathcal{D}$ is given by the cartesian product $\Omega_s \times \ldots \times \Omega_s$.

A fraction $\mathcal{F}$ can be defined as a multiset of $\mathcal{D}$. In Section 3.3 polynomial counting functions are used to represent multiplicity functions of orthogonal fractional factorial designs. Given a fraction $\mathcal{F}$ the counting function R of $\mathcal{F}$ can be represented as a complex polynomial defined over $\mathcal{D}$; for each $\zeta \equiv (\zeta_1, \ldots, \zeta_m) \in \mathcal{D}$, $R(\zeta)$ coincides with the number of appearances of a point ζ in the fraction (Definition 19).

We now express the coefficients c_α (Equation 3.30) using *strata*. The indicator function of all the points ζ of $\mathcal{D}$ is defined as:

$$1_\zeta : \mathcal{D} \ni \tau \longmapsto \begin{cases} 1 & \zeta = \tau \\ 0 & \zeta \neq \tau \end{cases} \tag{3.27}$$

In this way the counting function can be written as:

$$R = \sum_{\zeta \in \mathcal{D}} \gamma_\zeta 1_\zeta, \tag{3.28}$$

where

$$\gamma_\zeta \equiv R(\zeta) \in \{0, 1, \ldots, n, \ldots\}. \tag{3.29}$$

The coefficients of the counting function $c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{F}} \overline{X^\alpha(\zeta)}$ are related to the γ_ζ by:

$$c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{D}} \gamma_\zeta \overline{X^\alpha(\zeta)}. \tag{3.30}$$

The following propositions (Pistone and Rogantin (2008)(a)) regard two properties which hold for the full factorial design. We present them for the particular case in which all the factors have the same number s of levels and s is prime.

Proposition 22. *Let X_j^r be a simple term. Over $\mathcal{D}$, the term X_j^r*

- *will be constantly equal to 1, if $r = 0$;*
- *will take all the values of $\Omega_s = \{\omega_0, \ldots, \omega_{s-1}\}$ equally often, if $r \geq 1$.*

Proposition 23. *Let X^α an interaction term, $X^\alpha = X_1^{\alpha_1} \ldots X_m^{\alpha_m}$. Over $\mathcal{D}$, the term X^α*

- *will be constantly equal to 1, if $\alpha = \underline{0} \equiv (0, \ldots, 0)$;*
- *will take all the values of $\Omega_s = \{\omega_0, \ldots, \omega_{s-1}\}$ equally often if $\alpha \neq \underline{0}$.*

Now we can define the *strata* which are associated to simple or interaction terms.

Definition 24. *Let us consider a simple or interaction term $X^\alpha, \alpha \in L = \mathbb{Z}_s \times \ldots \times \mathbb{Z}_s$, $\alpha \neq \underline{0}$. The full factorial design $\mathcal{D}$ is partitioned into the following strata:*

$$D_h^\alpha = \left\{ \zeta \in \mathcal{D} : \overline{X^\alpha(\zeta)} = \omega_h \right\}, \tag{3.31}$$

where $\omega_h \in \Omega_s$ according to the Propositions 22 and 23.

The number of points $n_{\alpha,h}$ of the fraction $\mathcal{F}$ that are in the stratum D_h^α is:

$$n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} \gamma_\zeta, \quad h = 0, \ldots, s-1. \tag{3.32}$$

Let us consider a fraction $\mathcal{F}$ of the full factorial design $\mathcal{D}$ with counting function $R = \sum_{\alpha \in L} c_\alpha X^\alpha$. From Equation (3.30), we can write the coefficients $c_\alpha,\ \alpha \neq \underline{0}$, in terms of $n_{\alpha,h}, h = 0, \ldots, s-1$:

$$c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{D}} \gamma_\zeta \overline{X^\alpha(\zeta)} = \frac{1}{\#\mathcal{D}} \sum_{h=0}^{s-1} \omega_h \sum_{\zeta \in D_h^\alpha} \gamma_\zeta = \frac{1}{\#\mathcal{D}} \sum_{h=0}^{s-1} n_{\alpha,h} \omega_h. \tag{3.33}$$

We also observe that $c_{0,\ldots,0}$ is equal to the ratio between the number of points of $\mathcal{F}$ and the number of points of $\mathcal{D}$, $\frac{\#\mathcal{F}}{s}$. Pistone and Rogantin (2008)(a) showed that a fraction is an orthogonal array of strength t if and only if all the coefficients c_α of the polynomial counting function R up to order t are zero:

$$c_\alpha = 0 \quad \forall \alpha \in \mathcal{C}, \tag{3.34}$$

where $\mathcal{C} = \{\alpha \in L : 0 < \|\alpha\| \leq t\}$ and $\|\alpha\|$ is the number of non null elements of α. Proposition 6 in Fontana et al. (2010) obtained conditions on $n_{\alpha,h}$ that make $c_\alpha = 0$ for the counting function of the fraction $\mathcal{F}$.

For the case that we are considering, i.e. all the factors have the same number s of levels and s is prime, this Proposition asserts that $c_\alpha = 0$ if, and only if, $n_{\alpha,0} = n_{\alpha,1} = \ldots = n_{\alpha,s-1}$. It follows that, using *strata*, it is possible to express the condition $c_\alpha = 0$ as:

$$\sum_{\zeta \in \mathcal{D}_h^\alpha} \gamma_\zeta = \lambda_\alpha, \quad h = 0, \ldots, s-1, \tag{3.35}$$

where λ_α is a constant that, in general, does not depend on h.

It is easy to verify that, in this case, λ_α, being equal to $\frac{\#\mathcal{F}}{s}$, does not depend on α either. Then we denote λ_α simply by λ. If we vary $\alpha \in \mathcal{C}$ we can obtain a homogeneous system of linear equations:

$$\tilde{A}\tilde{Y} = 0, \tag{3.36}$$

where $\tilde{A} = [A, -\underline{1}]$, $\tilde{Y} = (Y, \lambda)$, A is the $(\#\mathcal{C} \times s^m)$ matrix whose rows contain the values of the indicator function of the strata, Y is the s^m column vector whose entries are the values of the counting function over $\mathcal{D}$, λ will be equal to $\frac{\#\mathcal{F}}{s}$ and $\underline{1}$ is the s^m column vector of 1.

It is now straightforward to verify that the union of two orthogonal arrays, $\mathcal{F}_1 \in OA(n_1, m, s, t)$ with counting function Y_1 and $\mathcal{F}_2 \in OA(n_2, m, s, t)$ with counting function Y_2, is an orthogonal array $\mathcal{F}_1 \cup \mathcal{F}_2 \in OA(n_1 + n_2, m, s, t)$ with counting function $Y_1 + Y_2$. Indeed $\tilde{A}(\widetilde{Y_1 + Y_2}) = \tilde{A}(\tilde{Y_1}) + \tilde{A}(\tilde{Y_2}) = 0$.

Given a homogenous linear system $\tilde{A}\tilde{Y} = 0$ where the unknowns $\tilde{Y}$ are positive integers, the Hilbert Basis (Schrijver (1986)) is a minimal set of generators of all the positive integer vectors $\tilde{Y}$ that are solutions of $\tilde{A}\tilde{Y} = 0$.

In this way any counting function Y of an orthogonal array $\mathcal{F} \in OA(n, m, s, t)$ is a linear combination of the generators with positive or null integer coefficients. Using $4ti2$ to compute the Hilbert Basis, Fontana and Pistone (2010), considered the following cases:

- $OA(n, 5, 2, 2)$; the Hilbert Basis contains the counting functions of $26,142$ orthogonal arrays. They can be classified by the number of points:

number of points	8	12	16	20	24	28	32	36	Total
number of OAs	60	224	162	960	7,680	8,384	5,760	2,912	26,142

- $OA(n, 3, 3, 2)$; the Hilbert Basis contains the counting functions of 66 orthogonal arrays, 12 have 9 points, all different, and 54 have 18 points, one replicated, i.e. the support size is equal to 17.

Let us consider the first case, $OA(n, 5, 2, 2)$. Let us suppose that we want to generate all the orthogonal arrays with 24 runs. We have to take:

- the $7,680$ elements of the Hilbert Basis that correspond to orthogonal arrays of size 24;
- all the counting functions that are the sum of the elements of the Hilbert Basis and have a total number of points equal to 24. This can be done:
  - adding 3 elements, each one with 8 points; we obtain $32,620$ different orthogonal arrays;
  - adding 2 elements, each one with 12 points; we obtain $22,044$ different orthogonal arrays;
  - adding 2 elements, one with 8 points and the other with 16 points; we obtain $7,800$ different orthogonal arrays;

In total we get $7,680 + 32,620 + 22,044 + 7,800 = 70,144$ OAs. If we only consider the different ones, we obtain $55,284$ orthogonal arrays. From this set we extract a list of nonisomorphic OAs. We implement an algorithm based on the one presented in Evangelaras et al. (2006). It should be noted that representing an orthogonal array using its counting function simplifies the algorithm because counting functions are invariant for row permutations. We obtain 63 nonisomorphic OAs (Arboretti, Fontana and Ragazzi (2011)).

We adopt the same approach for $OA(n, 4, 3, 3)$. The Hilbert Basis contains $131,892$ different OAs. If we consider all the OAs that contain 54 points and that are nonisomorphic this number is reduced to 7 (Arboretti, Fontana and Ragazzi (2011)).

3.5 Final remarks

In this Chapter we explored the algebraic methods for describe full factorial and fractional factorial designs.

The method proposed (based on the counting polynomial functions) appears to be a valid tool for the construction of OAs, the major advantage is that there are no restriction to the number of levels of the factors and so both fixed and mixed OAs can be considered under the same framework.

Using indicator functions we can also provide the generation of OAs when there is only one observation for any row of the matrix, in this work we focus on the generation of unreplicated OAs because in many experimental situations, in particular in a nonparametric framework, these designs appear very often.

We observe also that the problem which occurs in the generation of OAs is the information overload, so fact when there is a great number of factor and levels the computation of the exponent set is often prohibitive.

In any case the method, using the complex code of factor levels, gives the possibility to find OAs with strengths greater then two, with different number of factors and levels. The problem to construct OAs with strength greater than two and with a great number of factors and levels is still an open problem.

We also use a new methodology, based on algebraic strata, to generate the class of all the orthogonal arrays of given size and strength. From this class we extract all the nonisomorphic orthogonal arrays. The method is based on finding positive integer solutions of linear system of equations avoiding computations with complex numbers. We focus on the generation of nonisomorphic orthogonal arrays but the proposed approach can be applied to all fractional factorial designs, including mixed level, and without restriction on the number of levels of each factor.

Chapter 4

Unreplicated multifactorial designs

4.1 Introduction

The aim of this Chapter is to present a novel nonparametric approach for testing active effects in unreplicated fractional factorial designs through nonisomorphic OAs (Arboretti, Fontana and Ragazzi (2011)). The suggested method is based on extensions of the *inequivalent matrices permutation* (IMP) testing procedures (Basso et al. (2004)) and it provides separate permutation tests for effects in two-level and three-level unreplicated multifactorial designs. The key idea is to obtain the permutation distributions exchanging matrices, which are elements of the class composed of nonisomorphic OAs of a given size and strength available for the experiment, instead of permuting responses of a single observed design matrix. This extend the usual approach to permutation techniques. For a detailed discussion of the general theory of the nonparametric permutation testing procedures we refer to Pesarin (2001), for the aim of our present discussion we only review the basic concepts and notations of permutation tests within the framework of experimental designs (Section 4.2).

Analysis of unreplicated factorial and fractional factorial designs has been an extensively researched subject. To avoid misunderstanding it should be pointed out that our approach can be adopted not only for unreplicated designs but also when some replications are available.

We show our methodology on OAs with 2 or 3 levels but it can be used on the wider class of orthogonal fractional factorial designs, including mixed level designs. Using the classes of nonisomorphic two-level and three-level OAs generated using *strata* (Chapter 3, Section 3.4) we provide permutation tests for active effects in both two-level and three-level unreplicated multifactorial designs.

There are many numerical methods for analyzing unreplicated factorial experiments. Hamada and Balakrishnan (1994) compared a large number of methods for orthogonal two-level designs using an extensive simulation study. Of the direct methods they compared, most performed similarly with respect to power. They recommended two methods based

on ease of calculation. One of these two methods was developed by Lenth (1989).

Another two-stage method developed by Loh (1992) had a familiar graphical representation based on the normal probability plot. Although this method was not included in the extensive simulation study conducted by Balakrishnan and Hamada (1994), Loh (1992) indicated that it was more powerful than Lenth's method.

Within a permutation framework several contributions are developed for the analysis of unreplicated two-level factorial and fractional factorial designs. Pesarin (2001) provides permutation tests for two-level fractional factorial designs using the concepts of *realigning observations* and *paired permutations*. The method provides exact separate tests but it is only for the two-level designs and for the m main effects.

Basso et al. (2004) suggest using the IMP testing procedures for the analysis of unreplicated two-level designs. The method obtains the permutation distributions by exchanging nonisomorphic OAs instead of permuting responses of a single observed array. In that paper a simulation study is performed in order to compare the IMP test with Lenth's and the parametric F test, showing a better performance of the IMP test with respect to both other tests in the case of non normal error distributions.

We observe that all these methods are for the analysis of two-level multifactorial unreplicated experiments.

In Section 4.3 we present a generalization of the IMP testing procedures for testing effects in three-level unreplicated multifactorial experiments . In the Simulation Study Section (Section 4.4) we show the IMP test for both two-level and three-level designs using classes composed by:

- 63 nonisomorphic OAs with parameters $N = 24$, $m = 5$, $s = t = 2$ generated using *strata* (Section 3.4, Chapter 3);
- 7 nonisomorphic OAs with parameters $N = 54$, $m = 4$, $s = t = 3$ generated using *strata* (Section 3.4, Chapter 3);
- 711 nonisomorphic OAs with parameters $N = 27$, $m = 4$, $s = 3$ and $t = 2$ (Schoen et al (2010)).

4.2 Background and rationale for permutation tests

The first descriptions of permutation tests for linear statistical models (including analysis of variance and regression) can be traced back to the early half of this century in the work of Fisher (1935). Such tests are computationally intensive, however, and the use of these tests as opposed to the traditional normal-theory tests did not receive much attention in the natural and behavioral sciences until much later, with the emergence of widely accessible computer power (Edgington (1995)).

There is general agreement concerning an appropriate method of permutation for exact tests of hypotheses in one-way analysis of variance (ANOVA) or simple linear regression

(or, more simply, tests for the relationship between two variables, e.g., Pesarin (2001)). This is not the case, however, for tests of individual factors in the context of multiple linear regression or multifactorial designs. A clear description of the permutation methods for testing effects within the framework of experimental designs is given in Pesarin (2001). It is possible to assume, in general, that the observed data set $\mathbf{Y}$ is usually obtained by an experiment performed n times on a population variable and it is generally partitioned into groups (or blocks) according to the treatment levels of the experiment.

Pesarin (2001) describes and shows the statistical properties of the permutation tests (such as: exactness, unbiasedness, similarity, consistency and invariance) when the conceptual notion of the *exchangeability* among the units under H_0 is assumed to be valid. We observe that the assumption of exchangeability under H_0 is equivalent to the assumption of independent and identically distributed (IID) random errors. We observe that the assumption of the exchangeability among units is typical referred to the concept of *exchangeability of data with respect to groups or blocks* (*restricted* permutation approach - Pesarin (2001)). The following Proposition formalizes the concept of exchangeability.

Proposition 25. *A $n \times 1$ random vector $\mathbf{Y}$ has an exchangeable distribution if and only if any permutation of $\mathbf{Y}$ has the same distribution.*

In this way the variance matrix Σ of $\mathbf{Y}^* = P_\pi \mathbf{Y}$, where P_π is a permutation matrix, must be the same as that of $\mathbf{Y}$. So we can say that:

$$P_\pi \Sigma P_\pi^T = \Sigma, \tag{4.1}$$

for any matrix $P_\pi \in \mathcal{P}_e$, the set of $n \times n$ permutation matrices.

We also observe that if we intend to make inferences from our sample to a wider population, then an added assumption is necessary, namely, any effect of treatments on the observed units is the same as the effect of treatments on the set of units in a wider specified population. Such an assumption is weaker than the assumption that we actually have a random sample from the population.

In this way we do not need to assume what kind of population distribution our particular errors are obtained from (normal or otherwise), only that whatever that population is, the errors associated with the units we have are IID. Note that this means that a test by permutation does not avoid the assumption of homogeneity of error variances (the observation units must be exchangeable) under the null hypothesis.

Let us consider two points $\mathbf{Y}^*$ and $\mathbf{Y}'$ of the *permutation sample space* Ξ (which is the set of all the point $\mathbf{Y}^*$ such that the likelihood ratio $f_P^n(\mathbf{Y})/f_P^n(\mathbf{Y}^*) = \rho(\mathbf{Y}, \mathbf{Y}^*)$ is not dependent on the density f_P^n for whatever continous distribution P which belongs to a nonparametric family of distribution $\mathcal{P}$) then, in the null hypothesis H_0, the conditional probability of a generic point $\mathbf{Y}^*$ is given by:

$$Pr(\mathbf{Y}^* = \mathbf{Y}' \mid \Xi) = \frac{\#[\mathbf{Y}^* = \mathbf{Y}',\ \mathbf{Y}^* \in \Xi]}{\#[\mathbf{Y}^* \in \Xi]}, \tag{4.2}$$

which is P-independent (invariance properties). Pesarin (2001) shows that this probability, if there are no ties in the data set and if permutations correspond to permutation of the arguments, is equal to $\frac{1}{n!}$.

It's well known that the data set $\mathbf{Y}$ is always a set of sufficient statistics in H_0 for whatever distribution. In a permutation framework we assume that the data set *partitioned* into *groups* (or *blocks*) is now the set of sufficient statistics. If we consider $\mathbf{Y}_1$ and $\mathbf{Y}_2$ two separate and independent data set with sample size n_1 and n_2 respectively then as the likelihood associated with the pooled data set is $f_P^n(\mathbf{Y}) = f_{P_1}^{n_1}(\mathbf{Y_1}) \cdot f_{P_2}^{n_2}(\mathbf{Y}_2)$ from the sufficiency principle it follows that the data set partitioned into groups $(\mathbf{Y}_1, \mathbf{Y}_2)$ is the set of sufficient statistic. This implies that no datum from $\mathbf{Y}_1$ can be exchanged with any other from $\mathbf{Y}_2$ because under the alternative the permutations are permitted only within groups.

It is a common misconception that a permutation test has *no assumptions*. However, exchangeability must either be assumed or it must be ensured by virtue of an a priori random allocation of units in an experiment.

Considering these conceptual notions it is possible to consider permutational approaches in experimental planning: Pesarin (2001) provides separate exact permutation tests for replicated factorial and fractional factorial designs, based on the concepts of *synchronized permutations* and for unreplicated two-level factorial designs (based on the concepts of *realigning observations* and *paired permutations*).

In unreplicated factorial designs a restricted permutation approach (in which the exchangeability of the units respect to the treatment levels holds in H_0) seems not to be possible, this is because the experiment is performed only one time and this implies the impossibility to find the permutation distributions obtained by permuting the units among the treatment levels of the response variable.

So in this context the key condition of exchangeability of observations doesn't hold. In order to give a permutation solution for testing effects in unreplicated multifactorial designs we need to look for approximate solutions based on the residuals, in particular in Section 4.3 we provide an extension of the IMP procedures for testing active effects in three-level multifactorial designs, which is essentially based on permutation of residuals.

Commenges (2003) shows that when the exchangeability doesn't hold one may try to find a transformation which achieves approximate exchangeability then an appropriate permutation test can be done. In the cited paper it is shown that if we consider a statistic $T = \phi(\mathbf{Y})$ it may be possible to find V such that $\widetilde{\mathbf{Y}} = V(\mathbf{Y})$ is exchangeable and to write $T = \widetilde{\phi}(\widetilde{\mathbf{Y}})$. In other cases Commenges (2003) restricts that $\widetilde{\mathbf{Y}}$ has an exchangeable variance matrix (*second moment exchangeability*) and shows the privileged role of residuals.

Let us consider for example the usual linear model in a 2^m unreplicated factorial designs:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4.3}$$

where $\mathbf{Y}$ is a vector of responses, $\mathbf{X}$ is the design matrix, $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_m]^T$ is a vector of unknown parameters and $\boldsymbol{\varepsilon}$ is a vector of errors which are assumed to be uncorrelated with

the others and with same common variance σ^2. We observe that the Y_i, for $i = 1, \ldots, n$, do not have the same prediction. Assume (for the moment) that the parameters of the model are known and let us consider affine transformations $g_i = a_i + Y_i + b_i$. So we have that:

$$E(a_i Y_i + b_i) = \lambda \text{ for some } \lambda, \tag{4.4}$$

hence

$$b_i = \lambda - a_i E(Y_i) \text{ so that } g_i(Y_i) = a_i[Y_i - E(Y_i)] + \lambda. \tag{4.5}$$

This is a linear form of residuals plus a constant; for linear or quadratic forms, the tests do not depend on λ so that if we consider affine transformations we are led to ordinary residuals. In this way we can consider solutions based on residuals for testing active effects in unreplicated multifactorial designs which, being affine transformations (Commenges (2003)), preserve the first and second-moment exchangeability.

4.3 Permutation solution for testing active effects in unreplicated multifactorial designs

We will now discuss an extension of the IMP solution (Basso et al. (2004)) for testing the effects in unreplicated three-level factorial designs (Arboretti, Corain and Ragazzi (2011)). The IMP testing procedures allow us to separately test all the effects by exchanging matrices instead of permuting the vector of responses. Without loss of generality, we mainly refer to experimental designs when there are m factors, each at three levels and when only the significance of the main effects of the factors is being tested. The generalisation to mixed level fractional factorial designs and to interactions testing is straightforward.

The main effects of factors $A_1, \ldots, A_m$ can be expressed through the parameters $\beta_1, \ldots, \beta_{2m}$; indeed for each factor we have two parameters that, in the case of quantitative factors, can be related to the linear and to the quadratic component. We denote by β_i and β_{i+m} the parameters that represent the main effect of factor A_i, $i = 1, \ldots, m$.

When only the main effects are under analysis, the design matrix $\mathbf{X}$ corresponding to a fractional factorial design $\mathcal{F}$ with m factors, each with 3 levels, and N runs can be defined as:

$$\mathbf{X} = [\mathbf{I}\ \mathbf{M}_{\mathcal{D}}], \tag{4.6}$$

where $\mathbf{I}$ is the $N \times 1$ column vector whose entries are all equal to 1 and that represent the general mean contribution, $\mathbf{M}_{\mathcal{D}}$ is a $N \times 2 \cdot m$ matrix whose columns, when the factors are quantitative, represent the linear and quadratic effects of the m factors.

The usual linear model (fixed effects) is:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4.7}$$

where $\mathbf{Y}$ is a vector of responses, $\mathbf{X}$ is the design matrix defined in (4.6), $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{2m}]^T$ is a vector of unknown parameters and $\boldsymbol{\varepsilon}$ is a vector of errors. As in Basso et al. (2004)

the errors are assumed to be uncorrelated with other errors, with zero means and equal variances.

When the information matrix $\mathbf{X}^T\mathbf{X}$ is of full rank, the only OLS (ordinary least square) estimate of parameter $\boldsymbol{\beta}$ is given by:

$$\widehat{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = [\widehat{b}_0, \ldots, \widehat{b}_{2m}] \tag{4.8}$$

where $\widehat{b}_0$ is the mean of observed responses and $\widehat{b}_i$, $i = 1, \ldots, 2m$ are the contrasts estimating the effects. In order to obtain uncorrelated estimators of β we represent linear and quadratic effects using orthogonal polynomials as reported in Table 4.1. In the case of qualitative factors they will simply represent orthogonal contrasts.

For two level designs we simply take $\mathbf{M}_{\mathcal{D}} \equiv \mathcal{D}$, being $\omega_0 = 1$ and $\omega_1 = -1$. We refer

Level	Linear	Quadratic
ω_0	-1	-1
ω_1	0	2
ω_2	1	-1

Table 4.1: Values of the orthogonal polynomials for three levels.

to a design matrix built using the values given in Table 4.1 as an *unnormalized design matrix*.

Usually the experimenter is interested in separately testing all the parameters. There are $2m$ parameters and consequently $2m$ null separate sub-hypotheses of interest:

$$H_{0\beta_i} : \{\beta_i = 0\}\,, \; H_{1\beta_i} : \{\beta_i \neq 0\}\,, \tag{4.9}$$

irrespective of whether the remaining $H_{0\beta_j}$ are true or not, $i, j = 1, \ldots, 2m$; $i \neq j$.

Once the experiment has been performed according to a fraction $\mathcal{F}$, and assuming that k nonisomorphic OAs are available, the IMP testing procedure can be summarized in the following steps:

1. obtain the estimates in decreasing order $\left|\widehat{b}^{obs}_{(1)}\right| \geq \left|\widehat{b}^{obs}_{(2)}\right| \geq \ldots \geq \left|\widehat{b}^{obs}_{(2m)}\right|$, where *obs* stands for *observed*;
2. put $\widetilde{\mathbf{Y}} = \mathbf{Y} - \sum_{j \neq s} \widehat{b}_{(j)}\mathbf{X}_{(j)}$, where $\widehat{b}_{(j)}$ represents the estimate of the jth ordered observed effect different from the one considered and $\mathbf{X}_{(j)}$ are the ordered columns of the design matrix $\mathbf{X}$ from which the effect of the corresponding factor was estimated (for further details refer to Basso et al., 2004);
3. for $q = 2, \ldots, k$ obtain the estimates $\widehat{b}^q_1, \widehat{b}^q_2, \ldots, \widehat{b}^q_{2m}$ from $\widetilde{\mathbf{Y}}$, using $\mathbf{X}^{(q)}$, the design matrix corresponding to the q-th nonisomorphic OA;
4. put $\widehat{b}^q_1 = \max_j \left\{\left|\widehat{b}^q_{(j)}\right|\right\}$;

5. calculate the p-value for $\widehat{b}^{obs}_{(s)} : p_{(s)} = \frac{\#[\widehat{b}^{q}_{(1)} \geq \widehat{b}^{obs}_{(s)}\ ]}{k}$;

6. repeat steps 2., 3., 4. and 5. for $s = 1, \ldots, 2m$.

In the present work we apply the IMP testing procedures for testing the main effects using the class composed of 63 nonisomorphic $OA(24, 5, 2, 2)$ (generated as described in Section 3.4, Chapter 3) and the class of 711 nonisomorphic $OA(27, 4, 3, 2)$ (Schoen et al. (2010)).

In order to extend the proposed method to when there are only a few nonisomorphic designs, like the 7 nonisomorphic $OA(54, 4, 3, 3)$, we experiment a modified IMP procedure. Given a list of k_1 nonisomorphic arrays, for each of its members we randomly generate, by row permutations, a certain number r of isomorphic arrays. Then the hybrid class made by $k = k_1 \cdot r$ arrays replaces the class of nonisomorphic OAs and the steps of the algorithm remain the same.

4.4 Simulation study

In this Section we report a simulation study performed in order:

- to validate the IMP testing procedures on two-level and three-level factorial designs; we consider
  - $OA(24, 5, 2, 2)$, linear effects;
  - $OA(27, 4, 3, 2)$, linear and quadratic effects.
- to experiment the modified IMP testing; we consider
  - $OA(54, 4, 3, 3)$, linear and quadratic effects.

The focus of the analysis is on the power of IMP testing procedures and on the behaviour of the tests under the null hypotheses. We compare the IMP testing with the standard parametric F test.

The simulation program generates $1,000$ independent experiments considering two different distributions for the error terms $\boldsymbol{\varepsilon}$: the standardized normal distribution $N(0, 1)$, as often occurs in the literature, and the Cauchy distribution $Ca(0, 1)$, to get an idea of the robustness of the test when errors have heavy tails.

To assess the power of the test, the simulation setting proposed in (Basso et al. (2004)) is as reported in the following Table 4.2:

In the simulation study, we use *normalized* design matrices, i.e. design matrices where each column is obtained dividing the corresponding column of the unnormalized matrix by the square root of its norm. For the sake of comparison with (Basso et al. (2004)) we multiply the suggested values of Table 4.2 by the square root of the norms of the columns of the unnormalized design matrix.

$OA(24,5,2,2)$					
Distribution	β_1	β_2	β_3	β_4	β_5
Normal	1	$\frac{3}{4}$	$\frac{1}{2}$	0	0
Cauchy	2	$\frac{3}{2}$	1	0	0

Table 4.2: Values of active effects as in (Basso et al. (2004))

We observe that, using this approach for OAs with three levels, the parameters related to the quadratic components are higher than those referring to the linear part. This allows us to explore the behaviour of the IMP testing procedure for many different values of the alternative.

The following Table 4.3 reports the absolute values of the effects that we used in our study.

Distribution	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Normal	$\sqrt{24}$	$\frac{3}{4}\sqrt{24}$	$\frac{1}{2}\sqrt{24}$	0	0			
Cauchy	$2\sqrt{24}$	$\frac{3}{2}\sqrt{24}$	$\sqrt{24}$	0	0			
	Values of active effects in the $OA(24,5,2,2)$ study							
Normal	$\sqrt{18}$	$\frac{3}{4}\sqrt{18}$	$\frac{1}{2}\sqrt{18}$	0	$\sqrt{54}$	$\frac{3}{4}\sqrt{54}$	$\frac{1}{2}\sqrt{54}$	0
Cauchy	$2\sqrt{18}$	$\frac{3}{2}\sqrt{18}$	$\sqrt{18}$	0	$2\sqrt{54}$	$\frac{3}{2}\sqrt{54}$	$\sqrt{54}$	0
	Values of active effects in the $OA(27,4,3,2)$ study							
Normal	$\sqrt{36}$	$\frac{3}{4}\sqrt{36}$	$\frac{1}{2}\sqrt{36}$	0	$\sqrt{108}$	$\frac{3}{4}\sqrt{108}$	$\frac{1}{2}\sqrt{108}$	0
Cauchy	$2\sqrt{36}$	$\frac{3}{2}\sqrt{36}$	$\sqrt{36}$	0	$2\sqrt{108}$	$\frac{3}{2}\sqrt{108}$	$\sqrt{108}$	0
	Values of active effects in the $OA(54,4,3,3)$ study							

Table 4.3: Values of active effects in the simulation study

The significance levels which can actually be achieved have steps equal to $1/k$ where k is the number of nonisomorphic OAs.

Tables 4.4, 4.5 and 4.6 report the results of the simulation study performed on the $OA(24,5,2,2)$ arrays. Table 4.4 reports the behaviour of the tests under H_0 while Table 4.5 and 4.6 report the rejection rates of the IMP test under the alternative as defined in Table 4.3.

We observe that our results are very similar to those presented in (Basso et al., 2004); it is worth noting that:

1. as the significance levels increase, the absolute value of the gap between the observed probability of type I error and the desired significance level decreases and for values of α larger than around 20% the test looks conservative;

2. the power of the IMP test is close to that of the F test when errors are normally

distributed and is higher than that of the F test when the errors follow a Cauchy distribution.

	Normal errors					Cauchy errors				
α	β_1	β_2	β_3	β_4	β_5	β_1	β_2	β_3	β_4	β_5
0.0159	0.07	0.069	0.031	0.028	0.042	0.044	0.051	0.029	0.017	0.019
0.0317	0.07	0.069	0.031	0.028	0.042	0.044	0.051	0.029	0.017	0.019
0.0476	0.088	0.085	0.039	0.034	0.057	0.061	0.062	0.039	0.026	0.036
0.0635	0.112	0.113	0.058	0.055	0.084	0.096	0.09	0.054	0.043	0.061
0.0794	0.123	0.125	0.063	0.061	0.103	0.106	0.098	0.062	0.049	0.071
0.0952	0.123	0.125	0.063	0.061	0.103	0.106	0.098	0.062	0.049	0.071
0.1111	0.138	0.13	0.072	0.065	0.108	0.115	0.105	0.073	0.057	0.084
0.127	0.156	0.146	0.086	0.086	0.139	0.139	0.127	0.09	0.071	0.095
0.1429	0.167	0.154	0.092	0.089	0.149	0.146	0.141	0.098	0.078	0.102
0.1587	0.167	0.154	0.092	0.089	0.149	0.146	0.141	0.098	0.078	0.102
0.1746	0.179	0.164	0.1	0.096	0.163	0.155	0.156	0.105	0.083	0.117
0.1905	0.195	0.18	0.118	0.104	0.189	0.181	0.179	0.116	0.09	0.135
0.2063	0.195	0.18	0.118	0.104	0.189	0.181	0.179	0.116	0.09	0.135
0.2222	0.202	0.193	0.125	0.109	0.201	0.191	0.189	0.12	0.093	0.146
0.2381	0.218	0.208	0.141	0.126	0.234	0.201	0.208	0.148	0.109	0.173
0.254	0.23	0.215	0.148	0.134	0.244	0.212	0.22	0.153	0.116	0.182
0.2698	0.23	0.215	0.148	0.134	0.244	0.212	0.22	0.153	0.116	0.182
0.2857	0.239	0.224	0.159	0.139	0.255	0.219	0.232	0.163	0.125	0.19
0.3016	0.259	0.238	0.169	0.157	0.268	0.237	0.254	0.182	0.139	0.214
0.3175	0.266	0.244	0.18	0.161	0.282	0.245	0.264	0.193	0.147	0.222

Table 4.4: $OA(24, 5, 2, 2)$: Behaviour of IMP tests under H_0

	Normal errors					Cauchy errors				
α	β_1	β_2	β_3	β_4	β_5	β_1	β_2	β_3	β_4	β_5
0.0159	0.966	0.859	0.469	0.025	0.043	0.480	0.370	0.162	0.014	0.018
0.0317	0.966	0.859	0.469	0.025	0.043	0.480	0.370	0.162	0.014	0.018
0.0476	0.969	0.875	0.518	0.033	0.054	0.503	0.391	0.182	0.025	0.038
0.0635	0.969	0.883	0.616	0.054	0.081	0.532	0.416	0.219	0.042	0.062
0.0794	0.969	0.888	0.639	0.059	0.101	0.549	0.426	0.231	0.047	0.071
0.0952	0.969	0.888	0.639	0.059	0.101	0.549	0.426	0.231	0.047	0.071
0.1111	0.969	0.889	0.656	0.063	0.111	0.556	0.436	0.241	0.057	0.082

Table 4.5: $OA(24, 5, 2, 2)$: Powers of IMP tests

	Normal errors					Cauchy errors				
α	β_1	β_2	β_3	β_4	β_5	β_1	β_2	β_3	β_4	β_5
0.0159	0.984	0.826	0.44	0.012	0.018	0.306	0.212	0.092	0.002	0.005
0.0317	0.994	0.897	0.554	0.032	0.041	0.368	0.28	0.135	0.011	0.012
0.0476	0.996	0.931	0.619	0.047	0.063	0.411	0.331	0.174	0.02	0.021
0.0635	0.997	0.947	0.672	0.061	0.079	0.447	0.364	0.211	0.029	0.029
0.0794	0.998	0.958	0.717	0.076	0.094	0.479	0.388	0.236	0.041	0.039
0.0952	0.998	0.966	0.748	0.093	0.115	0.496	0.409	0.263	0.056	0.049
0.1111	0.999	0.971	0.776	0.107	0.127	0.517	0.428	0.283	0.071	0.061

Table 4.6: $OA(24, 5, 2, 2)$: Powers of F tests

Rejection rates under null hypotheses								
α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.0014	0.0070	0.0110	0.0010	0.0030	0.0160	0.0140	0.0040	0.0040
0.0028	0.0170	0.0160	0.0020	0.0030	0.0220	0.0150	0.0040	0.0070
0.0042	0.0220	0.0190	0.0030	0.0030	0.0220	0.0180	0.0050	0.0090
0.0113	0.0360	0.0290	0.0090	0.0080	0.0310	0.0230	0.0110	0.0180
0.0211	0.0480	0.0380	0.0140	0.0180	0.0360	0.0340	0.0170	0.0270
0.0309	0.0500	0.0410	0.0150	0.0220	0.0400	0.0390	0.0280	0.0320
0.0408	0.0550	0.0500	0.0240	0.0300	0.0490	0.0470	0.0350	0.0400
0.0506	0.0640	0.0550	0.0310	0.0390	0.0540	0.0530	0.0390	0.0440
0.0605	0.0680	0.0610	0.0380	0.0510	0.0640	0.0630	0.0430	0.0530
0.0703	0.0730	0.0700	0.0420	0.0520	0.0710	0.0680	0.0450	0.0550
0.0802	0.0750	0.0730	0.0470	0.0550	0.0760	0.0730	0.0500	0.0580
0.0900	0.0790	0.0750	0.0510	0.0580	0.0810	0.0760	0.0550	0.0620
0.1013	0.0870	0.0820	0.0570	0.0610	0.0850	0.0820	0.0590	0.0640

Table 4.7: $OA(27, 4, 3, 2)$: Powers of IMP tests under H_0 - Normal errors

Rejection rates under null hypotheses								
α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.0014	0.0210	0.0140	0.0000	0.0020	0.0130	0.0100	0.0010	0.001
0.0028	0.0240	0.0170	0.0020	0.0040	0.0180	0.0150	0.0060	0.001
0.0042	0.0290	0.0200	0.0020	0.0040	0.0200	0.0180	0.0070	0.001
0.0113	0.0360	0.0290	0.0030	0.0080	0.0290	0.0250	0.0120	0.007
0.0211	0.0400	0.0370	0.0070	0.0110	0.0370	0.0340	0.0130	0.015
0.0309	0.0510	0.0420	0.0160	0.0170	0.0430	0.0420	0.0190	0.021
0.0408	0.0610	0.0500	0.0190	0.0210	0.0490	0.0490	0.0240	0.027
0.0506	0.0680	0.0520	0.0210	0.0240	0.0580	0.0560	0.0270	0.031
0.0605	0.0750	0.0610	0.0230	0.0290	0.0620	0.0630	0.0340	0.037
0.0703	0.0790	0.0660	0.0270	0.0320	0.0640	0.0650	0.0360	0.040
0.0802	0.0860	0.0730	0.0340	0.0360	0.0660	0.0680	0.0390	0.046
0.0900	0.0910	0.0750	0.0390	0.0410	0.0690	0.0720	0.0490	0.048
0.1013	0.0970	0.0790	0.0410	0.0450	0.0780	0.0740	0.0520	0.058

Table 4.8: $OA(27, 4, 3, 2)$: Powers of IMP tests under H_0 - Cauchy errors

Rejection rates under alternatives

α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.0014	0.8310	0.5710	0.0910	0.0030	0.9460	0.9090	0.5270	0.0040
0.0028	0.8440	0.6070	0.1250	0.0030	0.9460	0.9120	0.5970	0.0070
0.0042	0.8490	0.6320	0.1470	0.0030	0.9460	0.9130	0.6280	0.0090
0.0113	0.8700	0.6940	0.2380	0.0080	0.9460	0.9180	0.8010	0.0180
0.0211	0.8760	0.7260	0.2920	0.0180	0.9460	0.9180	0.8430	0.0270
0.0309	0.8800	0.7410	0.3250	0.0220	0.9460	0.9180	0.8610	0.0320
0.0408	0.8830	0.7560	0.3610	0.0300	0.9460	0.9180	0.8920	0.0400
0.0506	0.8840	0.7680	0.3810	0.0390	0.9460	0.9180	0.8980	0.0440
0.0605	0.8840	0.7760	0.4120	0.0510	0.9460	0.9180	0.9110	0.0530
0.0703	0.8840	0.7820	0.4320	0.0520	0.9460	0.9180	0.9220	0.0550
0.0802	0.8850	0.7870	0.4510	0.0550	0.9460	0.9180	0.9260	0.0580
0.0900	0.8850	0.7880	0.4620	0.0580	0.9460	0.9180	0.9330	0.0620
0.1013	0.8850	0.7910	0.4850	0.0610	0.9460	0.9180	0.9380	0.0640

Table 4.9: $OA(27, 4, 3, 2)$: Powers of IMP tests under Normal errors

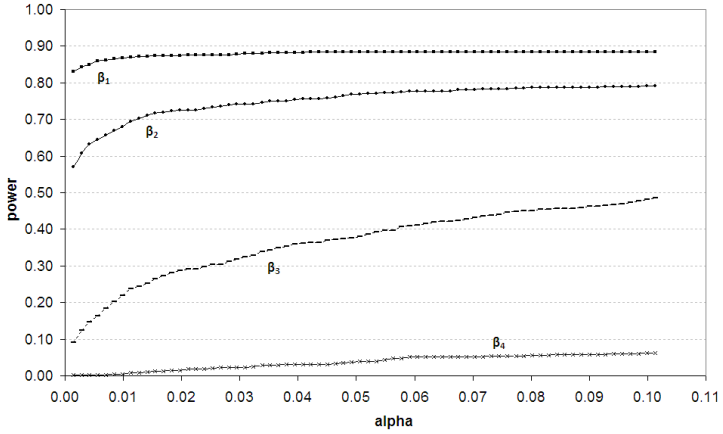

Figure 4.1: $OA(27, 4, 3, 2)$ Powers of IMP tests - Normal error terms

Rejection rates under alternatives

α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.0014	0.683	0.292	0.084	0.001	1	0.935	0.517	0.002
0.0028	0.78	0.401	0.134	0.002	1	0.962	0.613	0.002
0.0042	0.823	0.455	0.174	0.003	1	0.974	0.678	0.006
0.0113	0.915	0.622	0.294	0.015	1	0.987	0.809	0.015
0.0211	0.954	0.731	0.384	0.021	1	0.996	0.875	0.021
0.0309	0.971	0.799	0.448	0.028	1	0.998	0.895	0.032
0.0408	0.98	0.839	0.481	0.041	1	0.999	0.921	0.043
0.0506	0.986	0.871	0.523	0.05	1	0.999	0.937	0.054
0.0605	0.989	0.883	0.546	0.06	1	0.999	0.948	0.059
0.0703	0.989	0.896	0.575	0.068	1	0.999	0.956	0.064
0.0802	0.991	0.91	0.603	0.081	1	0.999	0.963	0.072
0.0900	0.993	0.925	0.626	0.088	1	0.999	0.965	0.081
0.1013	0.993	0.93	0.646	0.107	1	1	0.969	0.093

Table 4.10: $OA(27, 4, 3, 2)$: Powers of F tests under Normal errors

Rejection rates under alternatives

α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.0014	0.2700	0.1750	0.0230	0.0020	0.5030	0.3750	0.1020	0.0010
0.0028	0.2880	0.1860	0.0280	0.0040	0.5180	0.3890	0.1240	0.0010
0.0042	0.3040	0.2020	0.0360	0.0040	0.5250	0.4000	0.1330	0.0010
0.0113	0.3240	0.2390	0.0640	0.0080	0.5500	0.4260	0.2000	0.0070
0.0211	0.3400	0.2520	0.0800	0.0110	0.5630	0.4470	0.2310	0.0150
0.0309	0.3530	0.2620	0.0920	0.0170	0.5690	0.4570	0.2530	0.0210
0.0408	0.3690	0.2730	0.1070	0.0210	0.5800	0.4650	0.2740	0.0270
0.0506	0.3770	0.2760	0.1210	0.0240	0.5880	0.4710	0.2880	0.0310
0.0605	0.3820	0.2860	0.1370	0.0290	0.5900	0.4760	0.2980	0.0370
0.0703	0.3900	0.2980	0.1440	0.0320	0.5930	0.4810	0.3100	0.0400
0.0802	0.3940	0.3090	0.1580	0.0360	0.5980	0.4860	0.3180	0.0460
0.0900	0.3970	0.3140	0.1640	0.0410	0.6000	0.4930	0.3220	0.0480
0.1013	0.4100	0.3210	0.1720	0.0450	0.6060	0.5020	0.3320	0.0580

Table 4.11: $OA(27, 4, 3, 2)$: Powers of IMP tests under Cauchy errors

Rejection rates under alternatives

α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.0014	0.104	0.047	0.017	0	0.311	0.187	0.075	0
0.0028	0.133	0.066	0.027	0.001	0.343	0.215	0.099	0
0.0042	0.155	0.077	0.035	0.002	0.366	0.237	0.117	0.002
0.0113	0.223	0.124	0.057	0.005	0.421	0.301	0.176	0.005
0.0211	0.263	0.165	0.091	0.01	0.475	0.347	0.217	0.011
0.0309	0.3	0.21	0.115	0.015	0.507	0.377	0.241	0.013
0.0408	0.326	0.233	0.128	0.022	0.53	0.41	0.264	0.018
0.0506	0.343	0.249	0.148	0.026	0.554	0.432	0.284	0.026
0.0605	0.36	0.274	0.167	0.033	0.568	0.451	0.303	0.032
0.0703	0.381	0.295	0.181	0.041	0.582	0.472	0.318	0.039
0.0802	0.394	0.313	0.197	0.048	0.595	0.49	0.336	0.05
0.0900	0.406	0.324	0.213	0.059	0.61	0.511	0.348	0.058
0.1013	0.419	0.345	0.228	0.076	0.618	0.525	0.37	0.066

Table 4.12: $OA(27, 4, 3, 2)$: Powers of F tests under Cauchy errors

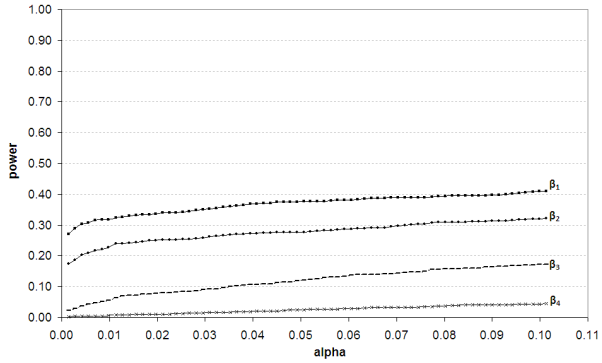

Figure 4.2: $OA(27, 4, 3, 2)$ Powers of IMP tests - Cauchy error terms

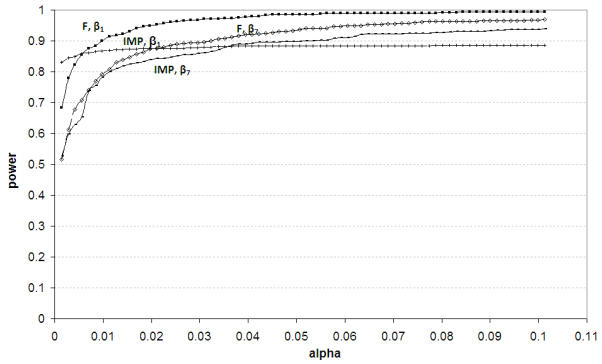

Figure 4.3: $OA(27, 4, 3, 2)$ IMP vs F test- Normal error terms

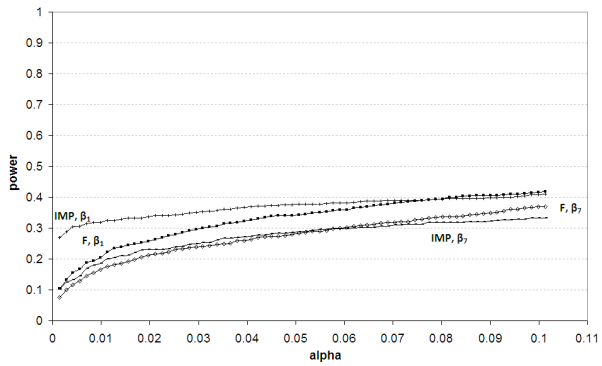

Figure 4.4: $OA(27, 4, 3, 2)$ IMP vs F test - Cauchy error terms

Rejection rates under null hypotheses

α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.010	0.003	0.002	0.001	0.000	0.003	0.002	0.004	0.003
0.020	0.006	0.004	0.002	0.005	0.008	0.003	0.006	0.003
0.030	0.006	0.006	0.004	0.008	0.009	0.003	0.008	0.004
0.040	0.008	0.008	0.005	0.013	0.011	0.004	0.010	0.006
0.050	0.009	0.009	0.005	0.015	0.012	0.005	0.011	0.007
0.060	0.015	0.010	0.006	0.015	0.013	0.007	0.014	0.010
0.070	0.017	0.013	0.007	0.016	0.014	0.007	0.014	0.012
0.080	0.018	0.015	0.007	0.016	0.016	0.009	0.017	0.013
0.090	0.024	0.016	0.010	0.018	0.017	0.012	0.019	0.014
0.100	0.027	0.020	0.013	0.020	0.018	0.013	0.022	0.014
Normal Errors								
0.010	0.000	0.003	0.001	0.001	0.001	0.000	0.004	0.000
0.020	0.004	0.004	0.004	0.002	0.002	0.002	0.006	0.001
0.030	0.008	0.004	0.004	0.002	0.006	0.003	0.010	0.001
0.040	0.009	0.004	0.005	0.002	0.007	0.007	0.012	0.001
0.050	0.011	0.007	0.008	0.004	0.008	0.012	0.014	0.003
0.060	0.011	0.009	0.010	0.004	0.010	0.014	0.015	0.007
0.070	0.015	0.010	0.010	0.006	0.011	0.015	0.016	0.008
0.080	0.016	0.011	0.012	0.010	0.011	0.017	0.019	0.008
0.090	0.022	0.011	0.014	0.011	0.011	0.019	0.020	0.009
0.100	0.024	0.012	0.016	0.011	0.013	0.019	0.022	0.010
Cauchy Errors								

Table 4.13: $OA(54, 4, 3, 3)$: Powers of IMP tests under H_0

Rejection rates under alternatives

α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.010	0.995	0.885	0.392	0.000	1.000	1.000	0.973	0.003
0.020	0.999	0.921	0.487	0.005	1.000	1.000	0.983	0.003
0.030	0.999	0.935	0.538	0.008	1.000	1.000	0.987	0.004
0.040	1.000	0.954	0.575	0.013	1.000	1.000	0.990	0.006
0.050	1.000	0.964	0.610	0.015	1.000	1.000	0.991	0.007
0.060	1.000	0.966	0.633	0.015	1.000	1.000	0.993	0.010
0.070	1.000	0.969	0.646	0.016	1.000	1.000	0.994	0.012
0.080	1.000	0.977	0.664	0.016	1.000	1.000	0.994	0.013
0.090	1.000	0.980	0.681	0.018	1.000	1.000	0.995	0.014
0.100	1.000	0.982	0.695	0.020	1.000	1.000	0.995	0.014
Normal Errors								
0.010	0.210	0.122	0.031	0.001	0.455	0.312	0.167	0.000
0.020	0.245	0.157	0.046	0.002	0.477	0.356	0.195	0.001
0.030	0.264	0.169	0.058	0.002	0.495	0.371	0.211	0.001
0.040	0.279	0.186	0.069	0.002	0.505	0.384	0.227	0.001
0.050	0.289	0.197	0.074	0.004	0.512	0.397	0.239	0.003
0.060	0.301	0.203	0.075	0.004	0.529	0.409	0.248	0.007
0.070	0.306	0.212	0.081	0.006	0.537	0.422	0.258	0.008
0.080	0.317	0.219	0.091	0.010	0.543	0.436	0.263	0.008
0.090	0.320	0.225	0.099	0.011	0.547	0.440	0.266	0.009
0.100	0.326	0.232	0.101	0.011	0.552	0.445	0.269	0.010
Cauchy Errors								

Table 4.14: $OA(54, 4, 3, 3)$: Powers of IMP tests

Tables 4.7, 4.8, 4.9, 4.10, 4.11 and 4.12 report the simulation study results considering the 711 nonisomorphic OAs $OA(27, 4, 3, 2)$.

In Figure 4.1 and 4.2 we report the power of the tests on the linear parameters as a function of the α levels, $\alpha \leq 10\%$. From Tables 4.7 and 4.8, we observe that, for both the error distributions, the test is conservative for α greater than around 7%.

The IMP testing procedure is based on the permutations of inequivalent matrices and so, it does not only permute the observations within a certain configuration of the factors. In this sense IMP provides an approximate test and this justifies the gap between the desired and the observed significance levels. The power simulations show the effectiveness of the IMP test. Indeed, from the first nominal level, the active effects are detected. We note that with Cauchy errors the IMP test works well, both under H_0 (Table 4.8) and H_1 (Table 4.11), in particular the inactive effects under H_1 maintain the same significance levels observed under H_0. Analogously to the $OA(24, 5, 2, 2)$ case, the IMP test has a performance close to that of the F test when errors are normally distributed (Figure 4.3) and performs better than the F test when errors follow a Cauchy distribution (Figure 4.4). We can also observe that the methodology suggested provides separate tests, one for each of the relating sub-hypotheses of null effect (this implies that for each factor we find two separate permutation tests related to the linear and quadratic component respectively), and within the framework of the permutation tests it is possible to take into consideration a phase of nonparametric combination methodology of the IMP tests. Finally in Tables 4.13 and 4.14 we report the simulation study results performed in order to experiment the proposed method when the number of nonisomorphic arrays is quite small; indeed there are only 7 nonisomorphic $OA(54, 4, 3, 3)$. We use a modified IMP test procedure. For each of the 7 nonisomorphic arrays, we randomly generate, by row permutations, 100 arrays. We get 700 arrays that we use as the input of the IMP algorithm. We can see that results look promising both in terms of accordance with the significance level and in terms of power.

4.5 Final remarks

Using nonisomorphic OAs generated using strata (Chapter 3) and from the literature (Schoen et al. (2010)) we provided separate tests for testing the main effects in unreplicated experiments in the framework of permutation tests. The basic idea was to exchange matrices instead of permuting the vector of responses (Basso et al. (2004)). The approach remains unchanged for testing interactions. As suggested by the simulation study, we can confirm that the proposed inferential solution is particulary useful for testing the effects when errors are not normally distributed. The applicability of the method in practical situations can be very high. Indeed, as we said, the method proposed by (Fontana and Pistone (2010)) covers the generation of all fractional factorial designs and for OAs in particular the recent catalogue of nonisomorphic OAs created by (Schoen et al. (2010)) offers a really wide range of designs. When the number of nonisomorphic OAs is not

sufficient (let us say $k \leq 50$) the modified IMP test seems to be a valid tool. This will be part of further research development.

Chapter 5

Randomized Complete Block designs

5.1 Introduction

In many scientific disciplines and industrial fields researchers and practitioners are often faced with theoretical and practical problems multivariate in nature within the framework of Randomized Complete Block (RCB) design with ordered categorical response variables when dealing with comparisons between two or more treatments.

For example, this situations can arise very often in the field of sensorial testing study, where especially in the food and body care industry, several useful experimental performance indicators are individual sensorial evaluations provided by trained people (panelists) during a so-called sensory test (Meilgaard et al. (2006)). Within this framework the experimental design typically handles panelists as blocks.

In general, the requirement to take into consideration a RCB design occurs when the experimental units are heterogeneous, hence the notion of blocking is used to control the extraneous sources of variability. The major criteria of blocking are characteristics associated with the experimental material and the experimental setting. The purpose of blocking is to sort experimental units into blocks, so that the variation within a block is "minimized"while that among blocks is "maximized". An effective blocking not only yields more precise results than an experimental design of comparable size without blocking, but also increases the range of validity of the experimental results.

It is well known that a best permutation test for all population distributions P does not generally exist, because the most powerful unbiased permutation test is a function of the population distribution P which is assumed, in the framework of the nonparametric inference, to be unknown (Pesarin (2001)). In this work we focus our attention on the RCB designs in case of ordered categorical response variables, which are typical in sensorial studies.

We propose, within the framework of the Nonparametric Combination (NPC) of De-

pendent Permutation Tests several combination-based permutation test statistics (Section 5.4) for the RCB design, especially in case of ordered categorical variables used for sensorial studies (Arboretti, Corain and Ragazzi (2010)). In order to validate the proposed method we present a Monte Carlo simulation study (Section 5.6), the focus is on the powers of the permutation testing procedures which have been compared with the traditional parametric and nonparametric competitors (Section 5.3). We find out that the Multifocus statistic using Fisher's combining function appears to be the more powerful solution which we proved also to be better under non normal errors than traditional parametric and rank-based nonparametric counterparts. We also propose a generalization of the combination-based permutation testing in the multivariate case (Section 5.7) (Arboretti, Corain and Ragazzi (2011)) and via an other comparative simulation study (Section 5.9) we find out that the Anderson-Darling statistic using the Fisher's combining function (for all phases of combination) represents a valid solution. Some examples in the field of the sensorial studies of the combination-based permutation approach are presented in Section 5.10. We conclude, Section 5.11, with a discussion and some directions of current and future research.

5.2 The set of hypotheses

Let us consider $\mathbf{X}$ a response categorical variable which the support is partitioned into $m \geq 2$ ordered classes $\{A_\gamma; \gamma = 1, \ldots, m\}$, in the sense that relationships such as $A_\gamma < A_\lambda$ have a clear meaning for every pair of subscripts γ, λ such that $1 \leq \gamma < \lambda \leq m$. The classes A_γ represent either quantitative or qualitative categories in according to the nature of $\mathbf{X}$.

In this context let us assume that a statistical model for $\mathbf{X}$ exists given by the tern $(\boldsymbol{\chi}, \mathcal{B}, P \in \mathcal{P})$, where $\boldsymbol{\chi}$ is the sample space of $\mathbf{X}$, $\mathcal{B}$ is an algebra of events and $\mathcal{P}$ is a nonparametric family of non-degenerate probability distributions on $(\boldsymbol{\chi}, \mathcal{B})$. Let us assume that data are classified according to the C treatments and to the n blocks according to the experimental design adopted for the experimentation. In this context we suppose that experimental units are randomly assigned to the C treatments ($C > 2$) and exactly one unit is assigned to each of the C treatments, the experimental design is developed with the aim at comparing the C treatments.

In what follows we refer to X_{ij} as the univariate categorical response variables where the subscripts $i = 1, \ldots, n$ and $j = 1, \ldots, C$ are related to the n blocks and C treatments respectively. If we consider C independent random samples $\mathbf{X}_j = \{X_{ji}; i = 1, \ldots, n; j = 1, \ldots, C\}$ the testing hypotheses can be formally described as follows:

$$H_0 : \left\{ X_{i1} \stackrel{\mathrm{d}}{=} \ldots \stackrel{\mathrm{d}}{=} X_{iC} \right\} = \{F_{i1}(A_\gamma) = \ldots = F_{iC}(A_\gamma), \gamma = 1, \ldots, m\} \ \forall i, \tag{5.1}$$

$$H_1 : \left\{ (X_{i1k} \overset{\mathrm{d}}{\neq} \ldots \overset{\mathrm{d}}{\neq} X_{iC}) \right\} = \left\{ \bigcup_{\gamma=1}^{m} (F_{i1}(A_\gamma) \neq \ldots \neq F_{iC}(A_\gamma)) \right\} \text{ for at least one } i, \tag{5.2}$$

for $i = 1, \ldots, n$, $\gamma = 1, \ldots, m$ and where $F_{ij} = Pr\{X_{ij} \leq A_\gamma\}$ with $j = 1, \ldots, C$ play the role of cumulative distribution functions (CDFs) for ordered categorical variables X_{ij}. We observe that (5.2) defines the stochastic non-dominance of X_{ij} with respect to X_{ih} with $j < h; j, h = 1, \ldots, C$.

Formally this problem is related to the problem of goodness-of-fit testing for ordered categorical variables and it's well known that when the sample sizes are finite and some nuisance parameters of the underlying multinomial distribution are unknown and are estimated from data, best solution very rarely exist for general hypotheses. In this context we present some solution from the viewpoint of permutation testing, in particular considering alternatives of the so-called stochastic non-dominance type for ordered categorical variables. We observe that the underlying response model is similar to the the model for generalized stochastic effects (Pesarin (2001)) and as an example of a possible solution within the permutation framework for stochastic ordering problems we remind the the works of Finos et al. (2007) and Finos et al. (2008).

Let us assume that there exists a nonparametric function ϕ such that:

$$X_{ij} = \phi(Y_{ij}) \overset{\mathrm{d}}{=} \phi(Y_{ih} + \Delta_j) \; j, h = 1, \ldots, C \; j \neq h; \; i = 1, \ldots, n; \tag{5.3}$$

where Y_{ij} represents underlying real-valued response, ϕ is a function which transforms Y_{ij} into ordered categorical data X_{ij}, and Δ_j represents non-negative stochastic effects.

The function ϕ can be defined as a nonparametric function defined on the set Γ (i.e. the set of the real-valued quantitative responses) in the set $\mathcal{O}$ (i.e. the set of the categorical responses) which maps any point of Γ in $\mathcal{O}$ such that $\phi(Y_{ij} + \Delta_j) \geq \phi(Y_{ij})$, for any $j = 1, \ldots, C$ and $i = 1, \ldots, n$.

This analogy allow us to extend the use of terminology adopted for quantitative variables to the case of ordered categorical variables, we also observe that this notation is suitable for simulation algorithms, when underlying continuous models are supposed to be generated before transformation into ordered classes. Conceptually continuous variables are sometimes treated as continuous and other times, especially in biomedical and social science research, as categorical. In this work we *categorize* (or *discretize*) continuous variables, as expressed in (5.3).

Following this rationale, let us focus our attention on the (continuous) variable $\mathbf{Y} = \{Y_{11}, \ldots, Y_{nC}\}$ related to a $n \times C$ matrix of responses. Let us consider an experimental design where there are n blocks and, within each block, experimental units are randomly assigned to the C ($C > 2$) treatments and exactly one experimental unit is assigned to each of the C treatments.

The statistical model (with fixed effects) for the randomized complete block (RCB)

design can be represented as follows:

$$Y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim IID(0, \sigma^2)\ i = 1, \cdots, n, j = 1, \cdots, C, \tag{5.4}$$

where β_i, τ_j and Y_{ij}, are respectively the effect of the i-th block, the effect of the j-th treatment and the response variable for the i-th block and the j-th treatment.

The random term ε_{ij} represents the experimental error with zero mean, variance σ^2 and unknown continuous distribution P. The usual side-conditions for effects are given by the constrains $\sum_i \beta_i = \sum_j \tau_j = 0$.

Model (5.4) is also called the *effects model* (Montgomery, (2001)). If we define $\mu_j = \mu + \tau_j$, $j = 1, \ldots, C$, an alternative representation of model (5.4) is the *means model*, i.e.

$$Y_{ij} = \mu_j + \beta_i + \varepsilon_{ij}. \tag{5.5}$$

The resulting inferential problem of interest is concerned with the following hypotheses: $H_0 : \{\tau_j = 0, \forall j\}$, against $H_1 : \{\exists j : \tau_j \neq 0\}$.

Note that this hypothesis is referred to a global test; if H_0 is rejected, it is of interest to perform inference on each pairwise comparison between couples of treatments, i.e. $H_{0(jh)} : \tau_j = \tau_h,\ \ j, h = 1, \ldots, C$, $j \neq h$, against $H_{1(jh)} : \tau_j \neq \tau_h$; with reference to model (5.5), an equivalent representation of $H_{0(jh)}$ is the following: $H_{0(jh)} : \mu_j - \mu_h = 0,\ \ j, h = 1, \ldots, C$, $j \neq h$, against $H_{1(jh)} : \mu_j - \mu_h \neq 0$.

We recall that in the framework of RCB designs there is usually no interest in testing the block effect which is handled as a nuisance factor. Note that, since no interaction effect between treatments and blocks is here supposed to exist, expressions (5.4) and (5.5) do not consider any interaction effect.

5.3 Traditional parametric and nonparametric testing procedures for the RCB design

In the framework of traditional parametric methods, when assuming random normal components, it is appropriate to test the equality of all treatment means by using the traditional F statistic:

$$F = \frac{SS_{Treatments}/(C-1)}{SS_E/(n-1)(C-1)}, \tag{5.6}$$

where:

$$SS_{Treatments} = n\sum_{j=1}^{C}(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})^2,\ SS_E = \sum_{i=1}^{n}\sum_{j=1}^{C}(Y_{ij} - \overline{Y}_{\cdot j} - \overline{Y}_{i\cdot} + \overline{Y}_{\cdot\cdot})^2,$$

and $\overline{Y}_{\cdot j}$ is the mean of the n experimental units in the j-th treatment, $\overline{Y}_{i\cdot}$ is the block mean for the i-th block, and $\overline{Y}_{\cdot\cdot}$ is the overall mean. The F statistic is distributed as

$F_{C-1,(C-1)(n-1)}$ if the null hypothesis H_0 is true, hence we would reject H_0, at the significance level α, if $F_0 > F_{\alpha;(C-1),(C-1)(n-1)}$.

If the analysis indicates a significant difference in treatment means, we are usually interested in multiple comparisons to find out which treatment means differ. That is, if the global null hypothesis H_0 is rejected we would consider the post-hoc set of $C(C-1)/2$ individual $H_{0(jh)}$ null hypotheses. Under normality, Bonferroni adjusted t-tests or Tukey's tests are the most recommended procedures. We recall that when carrying out multiple testing, there should be a formal guarantee against incorrect decisions.

The so called *multiplicity* problem is particularly relevant in multiple comparison problems, since omitting to consider the multiplicity issue can often cause biased statistical analysis (Westfall et al. (2000)).

Since the normality assumption is often questionable, if we do not assume the normality of random errors, we can take into consideration a nonparametric approach. In the framework of nonparametric rank-based testing procedures, one of the earlier tests has been proposed by Friedman (1937).

A general form of the Friedman's statistic T, which incorporates a correction for ties (Hollander and Wolfe, (1999)), is given by:

$$T = \frac{(C-1)\sum_{j=1}^{C}[R_{+j} - n(C+1)/2]^2}{\sum_{i=1}^{n}\sum_{j=1}^{C}(R_{ij})^2 - nC(C+1)^2/4}, \tag{5.7}$$

where R_{ij} is the rank of Y_{ij} among the experimental units in block i and $R_{+j} = \sum_j R_{ij}$ is the sum of the ranks for the j-th treatment over the n blocks. Under the null hypothesis, the R_{+j}'s should be close to $n(C+1)/2$ which is the average of the R_{+j}. Since T has an asymptotic Chi-square distribution with $C-1$ degree of freedom, we would reject the null hypothesis H_0 if $T_0 > \chi_{\alpha,(C-1)^2}$.

After rejection of H_0, the comparisons between pairs of treatments can be performed via absolute differences of the sums of within-blocks ranks. This set of values have to be compared with an appropriate value r_α which is function of C and n. For small values of C and n, r_α has been tabulated whereas, as n tends to infinity, it can be approximated by the distribution of the range of independent standard normal variables. This procedure, called *Wilcoxon-Nemenyi-McDonald-Thompson procedure* (Hollander and Wolfe (1999)), has been designed in order to maintain an appropriate Maximum Experimentwise Error Rate (MEER) α, where MEER is defined as the probability to reject at least one true hypotheses in the set of $C(C-1)/2$ individual $H_{0(jh)}$ null hypotheses.

Following Lehmann and D'abrera (2006), formula (5.7) can be replaced by:

$$T = nd'\Sigma_0^{-1}d, \tag{5.8}$$

where $\Sigma_0 = (\sigma_{jj'})$ is the covariance matrix under the null hypothesis of $R_i = (R_{i1}, \ldots, R_{i,C-1})$,

that is the rank order of the first $C-1$ treatments, and

$$d' = \left[R_{+1} - (C+1)/2, R_{+2} - (C+1)/2, \ldots, R_{+(C-1)} - (C+1)/2\right],$$

where $R_{+j} = \sum_j R_{ij}$.

Sepansky (2007) suggests a modification of (5.8), by the following test statistic:

$$T_P = nd'\widehat{\Sigma}^{-1}d, \tag{5.9}$$

where $\widehat{\Sigma}^{-1} = (s_{jj'})$ is the sample covariance matrix of the R_i.

Note that T_P is an Hotelling-type T^2 statistic and its limiting null distribution is the χ^2 distribution with $C-1$ degrees of freedom (see Anderson (2003) and Sepansky (2007)) examines also the covariance matrix in the test statistic (5.9) when the number of blocks or sample size is small and he claims that the null hypothesis of no treatment difference should be rejected when the sample covariance matrix is singular.

It is worth noting that while the Friedman test statistic is well defined when n is less than C, T_P is not since the sample covariance matrix is singular for all possible data matrices in this case. The idea of Sepansky of rejecting the null hypothesis when the sample covariance matrix is questionable and he does not support this statement with any kind of formal proof and the motivation he provided is quite debatable. Moreover, the simulation results presented by author clearly show that, especially for small values of n, his test statistic does not maintain the nominal levels under the null hypothesis. Hence, this proposal might be unreliable to properly perform inference for RCB designs.

Another approach, refereed as *Aligned Rank Test* (Lehmann and D'Abrera (2006)), is to make all blocks comparable so that comparisons between treatments in different blocks are meaningful. This can be done by subtracting the median or mean value of the experimental units in the block from all experimental units in that block. After this alignment is completed, the aligned experimental units are ranked over all blocks and treatments. It can be shown that under the null hypothesis the following statistic is a χ^2_{C-1} for large samples:

$$S = \frac{(C-1)n^2 \sum_{j=1}^{C} (\overline{R}_{\cdot j} - \overline{R}^2_{\cdot\cdot})^2}{\sum_{i=1}^{n}\sum_{j=1}^{C} (R_{ij} - \overline{R}_{i\cdot})^2}, \tag{5.10}$$

where now R_{ij} denotes the aligned rank for Y_{ij}, $\overline{R}_{i\cdot}$ is the average rank for the i-th block, $\overline{R}_{\cdot j}$ is the average rank for the j-th treatment and $\overline{R}_{\cdot\cdot}$ is the overall average rank.

In the literature there are a few other test statistics proposed for the RCB design. Among others, Quade (1979) proposed a test based on within-block rankings that gives greater weights to blocks that have greater variability. However, since several simulations studies (Fawcett and Salter (1984), Groggel (1987)) have shown that the Quade's procedure is not well performing in some situations, hence as suggested by O'Gorman (2001), it will be not included in the simulations we will present afterwards in this work. O'Gorman

(2001) reviews and evaluates several tests for RCB design, including the F-test, Friedman's test, and a few Aligned Rank Tests. His simulations show that Friedman's test has low power compared with the Aligned Rank Tests if the number of treatments does not exceed six and a novel aligned rank-based F-test proposed by the author shows relatively high power for several skewed distributions if there is a large number of experimental units.

5.4 Combination-based permutation solution for the RCB design

When dealing with complex designs conditional nonparametric methods can represent a reasonable approach. We recall that traditional unconditional parametric testing methods may be available, appropriate and effective only when a set of restrictive conditions are satisfied. Accordingly, just as there are circumstances in which unconditional parametric testing procedures may be appropriate, there are others where they may be unsuitable or even impossible to be properly applied. In conditional testing procedures, provided that exchangeability of data with respect to groups is satisfied in the null hypothesis, permutation methods play a central role. This is because they allow for quite efficient solutions, are useful when dealing with many difficult problems, provide clear interpretations of inferential results, and allow for weak extensions of conditional to unconditional inferences. For a detailed discussion on the topic of the comparison between permutation conditional inferences with traditional unconditional inferences we refer to Pesarin (2001).

In this Section 5.4 we propose a novel solution for the whole set of hypotheses of interest within the nonparametric framework of NonParametric Combination (NPC) of dependent permutation tests.

In this context we refer to a typical sensorial testing study so the observed data $\mathbf{X}$ is partitioned into m ordered classes, so $\mathbf{X}$ is usually organized in $C \times m$ contingency table. We note that H_0 (5.1) implies that the data of the C treatments are exchangeable, so that the *permutation testing principle* (Pesarin (2001)) may be properly applied. This implies taking into consideration the permutation sample space $\Xi_{/\mathbf{X}}$ generated by all permutations of pooled data set $\mathbf{X}$, that is, the set of all possible tables in which marginal frequencies are fixed.

In order to avoid computational problems, using the nonparametric function ϕ (5.3) which formally *categorizes* the continuous variable $\mathbf{Y}$ into the variable $\mathbf{X}$, we now consider the set of hypotheses statistics and the related combination-based permutation solutions in reference to the *categorized* variable $\mathbf{Y}$. In order to better explain the proposed approach let us denote an $(n \times C)$ data set $\mathbf{Y}$ as:

$$\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_j, \ldots, \mathbf{Y}_C] = \begin{bmatrix} Y_{11} & \ldots & Y_{1j} & \ldots & Y_{1C} \\ \vdots & \ddots & \vdots & & \vdots \\ Y_{i1} & \ldots & Y_{ij} & \ldots & Y_{iC} \\ \vdots & & \vdots & \ddots & \vdots \\ Y_{n1} & \ldots & Y_{nj} & \ldots & Y_{nC} \end{bmatrix}$$

where Y_{ij} represents the ijth observed response for ith block and jth treatment, with $i = 1, \ldots, n$ and $j = 1, \ldots, C$ ($C \geq 2$).

In the framework of NonParametric Combination (NPC) of dependent permutation tests we suppose that, if the global null hypothesis H_0 is true, the assumption of exchangeability of random errors within the same block holds. Hence, the following set of mild conditions should be jointly satisfied:

i) Suppose that for $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_C]$ an appropriate distribution P_j exists, $P_j \in \mathcal{P}$, $j = 1, \ldots, C$, belonging to a (possibly non-specified) family $\mathcal{P}$ (a family of non-degenerate probability distributions).

ii) The null hypothesis H_0 states the equality in distribution of the response variable in all C groups:

$$H_0 : [P_1, \ldots, P_C] = \left[\mathbf{Y}_1 \stackrel{d}{=} \ldots \stackrel{d}{=} \mathbf{Y}_C\right], \tag{5.11}$$

null hypothesis H_0 implies the exchangeability, within each block, of the individual data with respect to the C groups. Moreover H_0 is supposed to be properly decomposed into $C \times (C-1)/2$ sub-hypotheses $H_{0(jh)}$, $j, h = 1, \ldots, C$, $j \neq h$, each one related to the jhth pairwise comparison between couples of treatments:

$$H_0 : \left[\bigcap_{\substack{j,h=1 \\ j \neq h}}^{C} \mathbf{Y}_j \stackrel{d}{=} \mathbf{Y}_h\right] = \left[\bigcap_{\substack{j,h=1 \\ j \neq h}}^{C} H_{0(jh)}\right], \tag{5.12}$$

H_0 is called the *global* or *overall null hypothesis*, and $H_{0(jh)}$, $j, h = 1, \ldots, C$, $j \neq h$, are the *partial null hypotheses*.

iii) The alternative hypothesis H_1 is represented by the union of partial $H_{1(jh)}$ sub-alternatives:

$$H_1 : \left[\bigcup_{\substack{j,h=1 \\ j \neq h}}^{C} H_{1(jh)}\right] = \left[\bigcup_{\substack{j,h=1 \\ j \neq h}}^{C} H_{1(jh)}\right], \tag{5.13}$$

so that 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 $H_{1(jh)}$, $j, h = 1, \ldots, C$, $j \neq h$, are called the partial alternatives.

iv) Let $\mathbf{T} = \mathbf{T}(\mathbf{Y})$ represent a vector of test statistics, whose components $T_{(jh)}$, $j, h = 1, \ldots, C$, $j \neq h$, represent the partial univariate and non-degenerate *partial test* appropriate for testing the sub-hypothesis $H_{0(jh)}$ against $H_{1(jh)}$.

The assumptions regarding the set of all partial tests $\mathbf{T} = \left\{T_{(jh)}, j, h = 1, \ldots, C, j \neq h\right\}$ necessary for nonparametric combination are:

- all permutation partial test are marginally unbiased and significant for large values, so they are stochastically larger in H_1 than in H_0;
- all permutation partial tests $T_{(jh)}$ are consistent, that is,

$$Pr\left\{T_{(jh)} \geq T_{(jh)\alpha} \mid H_{1(jh)}\right\} \rightarrow 1, \forall\ \alpha > 0, j, h = 1, \ldots, C,\ j \neq h,$$

as n tends to infinity, where $T_{(jh)\alpha}$, which is assumed to be finite, is the critical value of $T_{(jh)}$ at level α.

The first assumption formally implies that,

$$Pr\left\{T_{(jh)} \geq T_{(jh)\alpha} \mid \mathbf{Y}, H_{1(jh)}\right\} \geq \alpha, \forall\ \alpha > 0, j, h = 1, \ldots, C, j \neq h;$$

$$Pr\left\{T_{(jh)} \leq z \mid H_{0(jh)}\right\} = Pr\left\{T_{(jh)} \leq z \mid \mathbf{Y}, H_{0(jh)} \cap H_{jh}^{\dagger}\right\} \geq$$

$$Pr\left\{T_{(jh)} \leq z \mid \mathbf{Y}, H_{1(jh)}\right\} = Pr\left\{T_{(jh)} \leq z \mid H_{1(jh)} \cap H_{(jh)}^{\dagger}\right\};$$

with $j, h = 1, \ldots, C\ \forall z \in \mathbb{R}^1$ where irrelevance with respect to the complementary set of hypotheses $H_{(jh)}^{\dagger} : \left\{H_{0(jh)} \cup H_{1(jh)}\right\}$ means that it does not matter which among $H_{0(jh)}$ and $H_{1(jh)}$, $j, h \neq i$, is true when testing for the ith sub-hypothesis.

The first assumption implies that the set of p - values $\left\{\lambda_{(jh)}, j, h = 1, \ldots, C, j \neq h\right\}$ associated with the partial tests in $\mathbf{T}$, are positively dependent in the alternative. In order to obtain global traditional consistency it suffices that at least one partial test is consistent (Pesarin (2001)).

As far as the partial test statistic $T_{(jh)}$ is concerned, since we are referring to a m-categories ordered response variable, it seems appropriate to take into consideration the following proposal:

- Multi-focus statistic: this approach suggests to decompose the categorical response variable of interest into m binary variables each one related to one category of the response; in this way it is possible to refer to a further decomposition of the sub-hypothesis $H_{0(jh)}$ into m additional sub-hypothesis each one suitable for testing the

equality in distribution of each one of the k category of the ordered categorical response variable; this is done via a set of m Chi-squared based test statistics calculated from $2m \times 2$ contingency sub-tables to be then combined into a final test statistic; note that a choice on the way to combine the k additional statistics have to be done;

- Anderson-Darling statistic:

$$T^*_{AD} = \sum_{r=1}^{m-1} (N^*_{hr} - N^*_{jr})[2\frac{N_{\cdot r}}{n}(\frac{2n - N_{\cdot r}}{2n})\frac{n^2}{2n-1}]^{-\frac{1}{2}}, \tag{5.14}$$

where $N_{\cdot r} = N_{jr} + N_{hr} = N^*_{jr} + N^*_{hr}$ are the observed and the permutation cumulative frequencies in which $N^*_{sr} = \sum_{q \leq r} f^*_{sq}, r = 1, \ldots, C-1, s = j, h.$ $f^*_{sr} = \#\left(X^*_{jr} \in A_r\right)$, $j = 1, \ldots, C$ and $r = 1, \ldots, m$ are permutation frequencies related to the classes A_r.

- Anderson-Darling statistic based on squared-values:

$$T^*_{AD2} = \sum_{r=1}^{m-1} (N^*_{hr} - N^*_{jr})^2 [N_{\cdot r} \times (2n - N_{\cdot r})]^{-1}. \tag{5.15}$$

- Kolmogorov-Smirnov statistic:

$$T^*_{KS} = sup(F^*_{hr} - F^*_{jr}), \tag{5.16}$$

where $F^*_{sr} = N^*_{sr}/n,\ r = 1, \ldots, m; s = j, h.$

At this point, once we have selected one of the above proposed test statistic $T_{(jh)}$ (each one related to the hypotheses $H_{0(jh)}$ with the goal of comparing the j-th and h-th treatment, the idea is to combine all them with an appropriate combining function, in order to test the global null hypothesis H_0.

However, we should observe that in most real problems when the number of blocks is large enough, there might be computational difficulties in calculating the conditional permutation distribution. This means that it is not possible to calculate the exact p-value of observed statistic $T_{(jh)0}$. This drawback is overcome by using the Conditional Monte Carlo (CMC) Procedure. The CMC on the pooled data set $\mathbf{Y}$ is a random simulation of all possible permutations of the same data under H_0. Hence, in order to obtain an estimate of the permutation distribution under H_0 of all test statistics, a CMC can be used. It should be emphasized that CMC only considers permutations of individual data vectors within each individual block, so that all underlying dependence relations which are present in the component variables are preserved.

5.5 A suitable algorithm

A suitable algorithm for calculating the proposed permutation test is composed of the following steps:

a) For each pairwise comparison between couples of treatments calculate the vector of the observed values of test statistics ${}^{o}\mathbf{T}(\mathbf{Y})$, whose components ${}^{o}T_{jh} = T(\mathbf{Y}_j, \mathbf{Y}_h)$, $j,\, h = 1, \ldots, C,\, j \neq h$, are appropriate for testing the sub-hypothesis $H_{0(jh)}$ against $H_{1(jh)}$.

b) Consider $\mathbf{Y}^*$ as a permutation of the data set $\mathbf{Y}$, carried out within each ith block in order to preserve the dependence structure of data, then calculate the permutation value of the test statistics:

$$T^*_{jh} = T(\mathbf{Y}^*_j, \mathbf{Y}^*_h),\ j,\, h = 1, \ldots, C,\ j \neq h.$$

c) Carry out B independent repetitions (i.e. CMC iterations) of step (b). The set of CMC results $\{{}_{b}T^*_{jh},\ b = 1, \ldots, B\}$ is thus a random sampling from the null permutation distribution of the test statistics.

d) Obtain the p-value from each partial sub-hypothesis $H_{0(jh)}$:

$$\lambda_{jh} = \#(T^*_{jh} \geq^{o} T_{jh})/B, \quad b = 1, \ldots, B,\ j,\, h = 1, \ldots, C,\ j \neq h.$$

e) The combined observed value of the global or overall null hypothesis H_0 is:

$${}^{o}T'' = \psi(\lambda_{11}, \ldots, \lambda_{(C-1)C}).$$

f) The combined value is then computed by:

$$T''^{*} = \psi(\lambda^*_{11}, \ldots, \lambda^*_{(C-1)C})$$

where $\lambda^*_{jh} = \#(T''^{*}_{jh} \geq {}_{b}T''^{*}_{jh})/B, \quad b = 1, \ldots, B.$

g) The global p-value is computed as:

$$\lambda'' = \#(T''^{*} \geq {}^{o}T'')/B, \quad b = 1, \ldots, B.$$

It can be seen that under the general null hypothesis the CMC procedure provides a consistent estimation of the permutation distributions, both marginal and combined, of the $C(C-1)$ partial tests. A general characterization of the class of combining functions is given by the following three main features for the combining function $\psi : [0,1]^{C(C-1)} \to \mathbb{R}$:

1. It must be non-increasing in each argument:

$$\psi(\ldots,\lambda_s,\ldots) \geq \psi(\ldots,\lambda'_s,\ldots) \text{ if } \lambda_s < \lambda'_s,\ s \in \{1,\ldots,S\}.$$

2. It must attain its supreme value, possibly not finite, even when only one argument reaches zero:

$$\psi(\ldots,\lambda_s,\ldots) \to \overline{\psi} \text{ if } \lambda_s \to 0,\ s \in \{1,\ldots,S\}.$$

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

$$T''_\alpha < \overline{\psi},$$

where λ_i is the p-value of the i-th partial test.

The above properties define the class C of combining functions. In Pesarin (2001) it is provided that: (i) if the partial permutation tests are exacts, then the combined test T_ψ is exact; (ii) if all partial permutation tests are marginally (separately) unbiased then T_ψ is unbiased; (iii) if both partial tests are marginally unbiased and at least one is consistent, then T_ψ is consistent.

Some of the functions most often used to combine partial dependent tests are:

- Fisher combining function:

$$T_F = -2\sum_i log(\lambda_i). \tag{5.17}$$

- Liptak combining function:

$$T_L = \sum_i \Psi^{-1}(1-\lambda_i), \tag{5.18}$$

where Ψ is the standard normal cumulative distribution function.

- Tippett combining function:

$$T_T = \max_i(1-\lambda_i). \tag{5.19}$$

- Liptak combining function with logistic transformations of the p-values:

$$T_P = \sum_i log[(1-\lambda_i)/\lambda_i]. \tag{5.20}$$

We observe that an iterated combination approach could overcome the problem of possible instability of any specific combination function. This solution could be particularly useful even in case of ordered categorical response variables; in fact, due to the finite support induced by the test statistics, the achievable p-value levels take 'jumping'values and this behaviour could imply possible p-value estimates which are different when applying different combining functions.

A general algorithm able to implement the iterated combination approach can be described as follows:

- choose a set of combining functions, for example Fisher, Liptak and Tippett, and separately apply all of them to obtain a global p-value, following the proposed algorithm for calculating the permutation test (see steps f and g); the set of calculated global p-values will slightly differ one another;

- separately apply again the set of chosen combining functions to the global p-values obtained in the previous step; the set of the second-iteration global p-values will differ one another but less than in the first iteration;

- iteratively repeat the previous step until the set of the iterated global p-values will take approximately the same value. The convergence is guaranteed thanks to the property of asymptotic equivalence of the combining functions (see Pesarin and Salmaso (2010)).

5.6 Comparative simulation study

In order to validate the proposed method and to evaluate its performance in comparison with either the traditional parametric (F test) and the nonparametric approach (Friedman and Aligned Rank Tests (MAR)), in this Section we perform a comparative simulation study. The goal is focused either on the global test H_0 and on the related pairwise comparisons (hypotheses $H_{0(jh)}$).

The real context we are referring to is a typical sensorial study where the number of blocks (panel lists) usually ranges around $10 - 15$ people and the sensorial evaluation is provided with a Likert $1 - 5$ rating ordinal scale, where we suppose that the 0.5 scores are admitted as well. Note that we are actually considering a 9 point ordered categorical response variable. The nonparametric function (5.3) and the relation between the non-dominance stochastic model and (5.2) allow us to consider the following setting:

- 1,000 independent simulations;

- number of blocks: n=6,10,20; number of treatment: C=3,5,7;

- block effect β_i, $i = 1, \ldots, n$, generated from a discrete uniform distribution with values (-1,- 0.5,0,0.5,1);

- with reference to model (2), the treatment effects μ_j, $j = 1, \cdots, C$, are set in the following way:

  - three treatments ($C = 3$): $\mu_1 = 1$, $\mu_2 = 2$ and $\mu_3 = 4$;
  - five treatments ($C = 5$): $\mu_1 = 1$, $\mu_2 = 1.5$, $\mu_3 = 2$, $\mu_4 = 3$ and $\mu_5 = 4$;
  - seven treatments ($C = 7$): $\mu_1 = 1$, $\mu_2 = 1.5$, $\mu_3 = 2$, $\mu_4 = 2.5$, $\mu_5 = 3$, $\mu_6 = 3.5$ and $\mu_7 = 4$;

- we generate the underlying (univariate) continuous variables Y_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, C$ with three types of random errors: normal, exponential (as an example of an asymmetric distribution) and Student's t with 2 degree of freedom (as an example of an heavy tailed distribution). The variability of random errors has been calibrated to the value of $\sigma = 2$, with the aim of properly reveal and compare the power among the considered procedures. Using the nonparametric function ϕ expressed in 5.3 we obtain the categorical variables $X_{jir} =$, $i = 1, \ldots, n$, $j = 1, \ldots, C$ and $r = 1, \ldots, m$ where the transformation consists to round the random errors to the nearest integer. In this way we obtain the categorical variables X_{ijr} with respective number of ordered classes $m = C$.

For each simulation we performed the permutation tests (with 1,000 CMC) using all the considered statistics: Multi-focus (with Fisher combining function: MF-F; with Tippett combining function: MF-T; with Liptak combining function: MF-F), Anderson-Darling statistic (AD), Anderson-Darling statistic based on squared-values (AD2) and Kolmogorov-Smirnov statistic (KS). As far as the combining function is concerned, we considered either the Fisher and the Tippett combining function. The considered significance level was $\alpha = 0.05$.

All results of the comparisons among combination-based permutation statistics for the RCB design are reported in Table 5.1, Table 5.2 and Table 5.3, where we added to the statistic's label (MF-F, MF-T, MF-L, AD, AD2, KS) the term 'w F'or 'w T'to indicate that we used the Fisher or the Tippett combining function respectively. It can be proved that the combined permutation test obtained using Fisher, Liptak or Tippett combining functions are *admissible* combinations, i.e. it does not exist any other type of combination which is uniformly more powerful (Pesarin and Salmaso (2010)).

As parametric and nonparametric counterparts, we included in the simulation study the traditional F-test, the Friedman test (5.7) and finally the Mean Aligned Rank (Mean AR) test proposed by O'Gorman (2001). Results for the RCB design parametric and nonparametric counterparts, along with for the iterated combination-based permutation statistics (applied to MF-F, KS and AD2), are reported in Table 5.4.

Table 5.5 reports the rejection rates of the pairwise comparisons of the tests. Note that, in order to be able to properly compare the performances of the compared procedures with different values of C (i.e. no. of treatments), rejection rates of pairwise comparisons are presented in terms of *delta* (δ), that is of the true differences (in term of σ) between treatment effects, where delta is defined as:

$$\delta_{jh} = \tau_j - \tau_h, j, h = 1, \ldots, C;\ j \neq h. \tag{5.21}$$

For example we get $\delta = 1\sigma$ for $C = 3$ from the difference between μ_2 and μ_1, whereas we get $\delta = 1\sigma$ for $C = 5$ from the differences $\mu_3 - \mu_1$, $\mu_4 - \mu_3$, $\mu_5 - \mu_4$.

In case of rejection of the global null hypothesis H_{0k}, in order to perform the treatment pairwise comparisons, we considered permutation tests for two paired samples. Least

Significant Difference (LSD) for the difference of mean ranks and t-tests as post-hoc procedures respectively for Friedman test and F-test and MAR have been considered as well.

We recall that all post-hoc pairwise procedures should take into account for the problem of multiplicity (Westfall et al. (2000)) hence they have to be well defined in order to maintain at the desired α-level the type I error probability of the main global hypothesis H_0. For this goal, for permutation tests we adopted a multiplicity correction strategy by using the closed testing approach (Westfall et al. (2000)) via Tippett combining function (i.e. the so called $minP$ procedure, Westfall et al. (2000)) which is particularly suitable to be implemented within the framework of permutation tests, while for all other pairwise procedures we adopted the Bonferroni correction.

As first remark for the simulation study, we observe that under the null hypothesis all procedures appear to behave properly according to the nominal level (Tables 5.1, 5.2, 5.3 and 5.4). From a general point of view, as expected, the power for the global hypothesis increases as the number of blocks increases. With reference on the comparison among combination-based statistics, first of all it should be noted that the Tippett combining function does not perform satisfactorily because it produces systematically less powerful results than statistics using the Fisher combining function. This is not surprising and agrees with results of Finos and Salmaso (2004), in fact the Fisher combination is able to 'synergistically'exploit the behaviour of many hypotheses even when the effects are not very prominent. It is worth noting that Multi-focus type statistics generally perform very good and more or less similarly one each other, with a slight preference on the Fisher-based Multi-focus statistic. Both Anderson-Darling statistics are not satisfactory; actually they appear to be the worst statistics, while the Kolmogorov-Smirnov statistic is less powerful than Multi-focus statistics but it performs similarly in case of Student's t random errors.

Finally, a general weakness of all combination-based statistics is that they are somewhat conservative when the number of blocks is less than 10, as we can realize from the comparison with the traditional RCB statistics. This is surely due to the relative low number of possible permutations and consequently of different values that the test statistic can assume.

As far as the comparison with traditional RCB statistics, obviously the F-test shows a better behaviour under normality, but in case of exponential errors and particularly of Student's t errors, all nonparametric procedures show a greater power. Among rank-based nonparametric tests, the worst one is the Friedman test whereas a good behaviour is provided by the Mean Aligned Rank test. It should be noted that Friedman test is actually not satisfactory when data have ties as in case of ordered categorical variables we considered in this Chapter. In fact, the continuity correction proposed by several authors is valid only asymptotically and for finite samples it does not provides a conservative test (Lehmann and D'Abrera (2006)). Note that the approach of iterated combination turns out to be very effective. In fact, Table 5.4 shows that rejection rates are always increasing when compared with results from Tables 5.1, 5.2 and 5.3.

Finally, as highlighted by Figures 5.1, 5.2 and 5.3, when comparing the best combination-based permutation statistic, i.e. the iterated Multi-focus with Fisher combining function,

it appears that it is actually the more powerful solution under non normal errors when the number of blocks is equal or greater than ten.

		H_1 Rejection Rates			H_0 Nominal Levels		
Statistic	n	C			C		
		3	5	7	3	5	7
MF-F w T	6	.252	.202	.130	.022	.022	.024
	10	.692	.594	.580	.048	.046	.033
	20	.974	.955	.965	.046	.042	.039
MF-T w T	6	.097	.103	.117	.003	.005	.032
	10	.548	.484	.513	.040	.037	.045
	20	.923	.894	.903	.042	.043	.038
MF-L w T	6	.239	.210	.122	.023	.027	.024
	10	.688	.611	.563	.055	.045	.043
	20	.968	.959	.964	.048	.047	.046
KS w T	6	.097	.101	.123	.003	.004	.010
	10	.506	.447	.495	.040	.028	.027
	20	.906	.888	.910	.038	.049	.039
AD2 w T	6	.013	.008	.005	.002	.002	.001
	10	.612	.506	.479	.040	.032	.034
	20	.951	.926	.945	.042	.039	.031
AD w T	6	.018	.006	.003	.001	.000	.001
	10	.598	.508	.453	.042	.039	.036
	20	.948	.927	.941	.045	.038	.031
MF-F w F	6	.293	.282	.290	.031	.027	.030
	10	.733	.690	.723	.059	.051	.045
	20	.971	.977	.982	.048	.041	.045
MF-T w F	6	.202	.228	.245	.022	.016	.032
	10	.628	.600	.639	.064	.058	.045
	20	.951	.945	.961	.049	.046	.038
MF-L w F	6	.308	.279	.294	.024	.027	.028
	10	.749	.691	.734	.059	.051	.044
	20	.968	.982	.975	.048	.035	.051
KS w F	6	.193	.225	.229	.014	.016	.032
	10	.608	.589	.636	.060	.054	.043
	20	.938	.951	.962	.047	.046	.042
AD2 w F	6	.143	.150	.142	.008	.011	.009
	10	.655	.601	.643	.056	.048	.026
	20	.965	.962	.969	.037	.042	.042
AD w F	6	.148	.150	.149	.009	.009	.010
	10	.660	.598	.648	.059	.047	.030
	20	.963	.956	.968	.041	.040	.037

Table 5.1: Global Test with Fisher(F) and Tippett(T) combining function - Normal Errors

		H_1 Rejection Rates			H_0 Nominal Levels		
Statistic	n	C			C		
		3	5	7	3	5	7
MF-F w T	6	.401	.434	.422	.024	.021	.029
	10	.856	.865	.891	.040	.055	.060
	20	.998	1.000	1.000	.053	.050	.042
MF-T w T	6	.304	.376	.388	.016	.017	.025
	10	.846	.857	.871	.031	.049	.062
	20	.998	1.000	1.000	.046	.034	.044
MF-L w T	6	.407	.436	.425	.023	.020	.025
	10	.798	.814	.831	.044	.051	.061
	20	.981	.969	.972	.044	.050	.041
KS w T	6	.267	.348	.401	.009	.016	.032
	10	.823	.840	.857	.028	.052	.055
	20	.998	.998	.998	.049	.041	.043
AD2 w T	6	.177	.207	.203	.005	.005	.007
	10	.637	.645	.689	.036	.037	.037
	20	.953	.952	.959	.035	.039	.039
AD w T	6	.179	.208	.198	.005	.007	.007
	10	.591	.586	.635	.036	.035	.035
	20	.882	.883	.879	.038	.038	.034
MF-F w F	6	.337	.298	.173	.019	.015	.030
	10	.798	.796	.749	.042	.049	.056
	20	.997	.996	.991	.044	.050	.051
MF-T w F	6	.161	.189	.191	.005	.002	.007
	10	.767	.740	.721	.021	.033	.045
	20	.994	1	.991	.041	.039	.054
MF-L w F	6	.329	.303	.167	.021	.014	.027
	10	.727	.742	.711	.048	.052	.051
	20	.963	.932	.936	.049	.057	.046
KS w F	6	.158	.187	.184	.002	.004	.008
	10	.733	.713	.710	.019	.029	.042
	20	.991	.993	.987	.033	.041	.051
AD2 w F	6	.013	.018	.007	.000	.000	.000
	10	.559	.543	.504	.031	.030	.041
	20	.930	.902	.888	.039	.039	.043
AD w F	6	.015	.018	.009	.002	.001	.003
	10	.591	.586	.635	.034	.032	.034
	20	.882	.883	.879	.038	.046	.036

Table 5.2: Global Test with Fisher(F) and Tippett(T) combining function - Exponential Errors

		H_1 Rejection Rates			H_0 Nominal Levels		
Statistic	n	C			C		
		3	5	7	3	5	7
MF-F w T	6	0.145	0.106	0.066	0.024	0.028	0.017
	10	0.382	0.372	0.317	0.053	0.042	0.051
	20	0.737	0.695	0.701	0.044	0.042	0.05
MF-T w T	6	0.057	0.06	0.085	0.007	0.009	0.011
	10	0.319	0.351	0.334	0.033	0.042	0.043
	20	0.714	0.682	0.7	0.049	0.04	0.050
MF-L w T	6	0.145	0.094	0.074	0.025	0.023	0.016
	10	0.335	0.313	0.308	0.053	0.055	0.052
	20	0.602	0.513	0.521	0.048	0.052	0.05
KS w T	6	0.048	0.056	0.069	0.006	0.01	0.008
	10	0.297	0.328	0.309	0.025	0.046	0.043
	20	0.730	0.699	0.699	0.034	0.038	0.039
AD2 w T	6	0.010	0.002	0.000	0.005	0.001	0.001
	10	0.313	0.285	0.246	0.039	0.038	0.032
	20	0.691	0.66	0.655	0.046	0.037	0.037
AD w T	6	0.006	0.004	0.000	0.003	0.001	0.000
	10	0.306	0.279	0.240	0.041	0.037	0.035
	20	0.618	0.582	0.541	0.039	0.045	0.033
MF-F w F	6	0.169	0.163	0.168	0.025	0.03	0.025
	10	0.399	0.429	0.417	0.061	0.051	0.048
	20	0.76	0.734	0.744	0.046	0.053	0.050
MF-T w F	6	0.116	0.136	0.161	0.033	0.024	0.024
	10	0.379	0.427	0.39	0.045	0.049	0.049
	20	0.742	0.726	0.766	0.049	0.041	0.052
MF-L w F	6	0.169	0.165	0.16	0.03	0.036	0.026
	10	0.369	0.385	0.384	0.056	0.056	0.044
	20	0.641	0.604	0.593	0.042	0.055	0.050
KS w F	6	0.105	0.142	0.158	0.02	0.014	0.026
	10	0.363	0.423	0.4	0.043	0.053	0.044
	20	0.784	0.767	0.796	0.053	0.041	0.049
AD2 w F	6	0.078	0.078	0.09	0.008	0.007	0.004
	10	0.347	0.359	0.357	0.043	0.041	0.036
	20	0.733	0.718	0.745	0.043	0.032	0.038
AD w F	6	0.079	0.085	0.086	0.009	0.007	0.006
	10	0.326	0.346	0.341	0.046	0.043	0.038
	20	0.670	0.647	0.649	0.042	0.04	0.030

Table 5.3: Global Test with Fisher(F) and Tippett(T) combining function - Student's t Errors

		H_1 Rejection Rates			H_0 Nominal Levels		
Statistic	n	C			C		
		3	5	7	3	5	7
F	6	0.535	0.506	0.508	0.050	0.043	0.038
	10	0.811	0.774	0.811	0.061	0.060	0.038
	20	0.986	0.993	0.987	0.05	0.042	0.049
Friedman	6	0.424	0.400	0.402	0.044	0.041	0.032
	10	0.702	0.667	0.711	0.047	0.054	0.041
	20	0.957	0.962	0.976	0.047	0.04	0.041
MAR	6	0.459	0.438	0.456	0.051	0.051	0.044
	10	0.719	0.696	0.731	0.054	0.06	0.045
	20	0.957	0.965	0.978	0.047	0.043	0.042
Normal							
F	6	0.559	0.565	0.547	0.054	0.034	0.041
	10	0.813	0.788	0.839	0.046	0.05	0.052
	20	0.984	0.975	0.987	0.038	0.04	0.044
Friedman	6	0.547	0.583	0.553	0.058	0.031	0.043
	10	0.803	0.852	0.890	0.042	0.039	0.044
	20	0.993	0.996	0.997	0.045	0.049	0.047
MAR	6	0.582	0.615	0.590	0.070	0.044	0.056
	10	0.818	0.873	0.900	0.047	0.042	0.049
	20	0.993	0.997	0.998	0.047	0.058	0.047
Exponential							
F	6	0.209	0.178	0.158	0.033	0.035	0.031
	10	0.256	0.257	0.242	0.046	0.037	0.037
	20	0.491	0.422	0.415	0.033	0.04	0.031
Friedman	6	0.223	0.209	0.201	0.045	0.035	0.029
	10	0.344	0.378	0.398	0.047	0.063	0.041
	20	0.706	0.732	0.753	0.042	0.051	0.05
MAR	6	0.245	0.238	0.237	0.053	0.045	0.037
	10	0.367	0.400	0.429	0.049	0.07	0.046
	20	0.72	0.745	0.764	0.046	0.056	0.051
Student's t							

Table 5.4: Global Test - Competitors

H_1 Rejection Rates

Stat.	n	C=3			C=5			C=7		
		delta								
		1	2	3	1	2	3	1	2	3
	6	0.034	0.175	0.393	0.02	0.107	0.309	0.024	0.057	0.245
F	10	0.092	0.377	0.761	0.036	0.236	0.632	0.048	0.127	0.533
	20	0.191	0.73	0.977	0.1	0.614	0.968	0.14	0.357	0.941
	6	0.017	0.094	0.292	0.005	0.043	0.164	0.005	0.02	0.127
Fr.	10	0.032	0.22	0.609	0.009	0.116	0.454	0.018	0.058	0.363
	20	0.066	0.53	0.935	0.043	0.417	0.878	0.078	0.223	0.878
	6	0.058	0.206	0.365	0.025	0.107	0.277	0.022	0.055	0.206
MAR	10	0.104	0.37	0.657	0.033	0.202	0.536	0.04	0.105	0.455
	20	0.174	0.651	0.943	0.086	0.532	0.912	0.127	0.315	0.899
	6	0.018	0.083	0.185	0.347	0.007	0.031	0.06	0.363	0.008
AD	10	0.091	0.35	0.643	0.721	0.027	0.135	0.357	0.729	0.036
	20	0.259	0.761	0.973	0.988	0.113	0.566	0.935	0.991	0.158
		Normal								
	6	0.057	0.252	0.489	0.024	0.123	0.375	0.03	0.069	0.31
F	10	0.104	0.39	0.756	0.049	0.253	0.62	0.058	0.141	0.563
	20	0.211	0.728	0.972	0.103	0.6	0.955	0.144	0.369	0.931
	6	0.018	0.092	0.46	0.005	0.054	0.326	0.009	0.027	0.261
Fr.	10	0.041	0.241	0.785	0.014	0.188	0.68	0.036	0.111	0.618
	20	0.139	0.657	0.993	0.05	0.603	0.987	0.134	0.4	0.979
	6	0.137	0.296	0.545	0.035	0.177	0.462	0.046	0.095	0.404
MAR	10	0.207	0.47	0.82	0.059	0.36	0.769	0.087	0.221	0.734
	20	0.325	0.84	0.995	0.14	0.767	0.993	0.222	0.538	0.99
	6	0.028	0.155	0.26	0.449	0.012	0.033	0.085	0.478	0.016
AD	10	0.156	0.486	0.792	0.914	0.058	0.253	0.545	0.94	0.086
	20	0.314	0.768	0.966	0.994	0.15	0.602	0.917	0.995	0.238
		Exponential								
	6	0.019	0.05	0.134	0.006	0.026	0.067	0.006	0.008	0.029
F	10	0.037	0.079	0.203	0.011	0.042	0.133	0.006	0.015	0.072
	20	0.052	0.195	0.395	0.017	0.08	0.237	0.012	0.033	0.144
	6	0.013	0.044	0.154	0.004	0.017	0.071	0.003	0.01	0.052
Fr.	10	0.031	0.101	0.254	0.008	0.049	0.183	0.007	0.022	0.115
	20	0.058	0.265	0.624	0.024	0.155	0.501	0.03	0.086	0.443
	6	0.049	0.091	0.192	0.011	0.045	0.116	0.013	0.022	0.086
MAR	10	0.068	0.155	0.294	0.018	0.082	0.231	0.014	0.034	0.162
	20	0.095	0.316	0.641	0.04	0.209	0.543	0.047	0.111	0.495
	6	0.018	0.029	0.089	0.146	0.006	0.014	0.029	0.172	0.007
AD	10	0.066	0.155	0.312	0.428	0.014	0.057	0.171	0.412	0.011
	20	0.096	0.266	0.458	0.559	0.032	0.134	0.329	0.593	0.037
		Student's t								

Table 5.5: Pairwise Comparisons

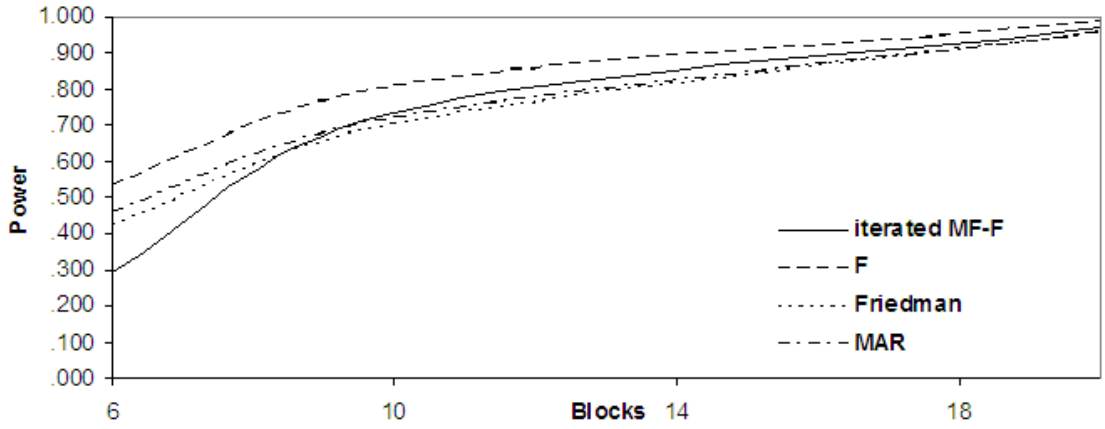

Figure 5.1: $C = 3$ normal errors. Power comparison among iterated Fisher MultiFocus permutation test and traditional counterparts.

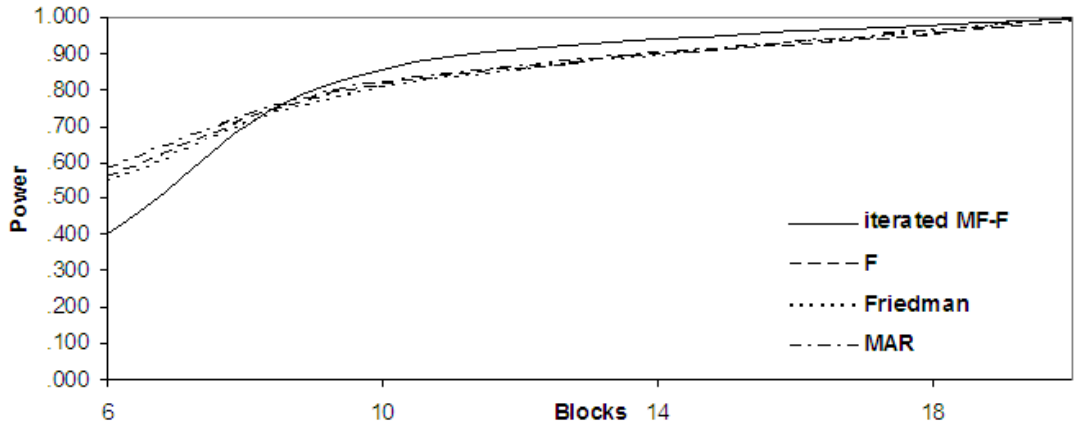

Figure 5.2: $C = 3$ exponential errors. Power comparison among iterated Fisher MultiFocus permutation test and traditional counterparts.

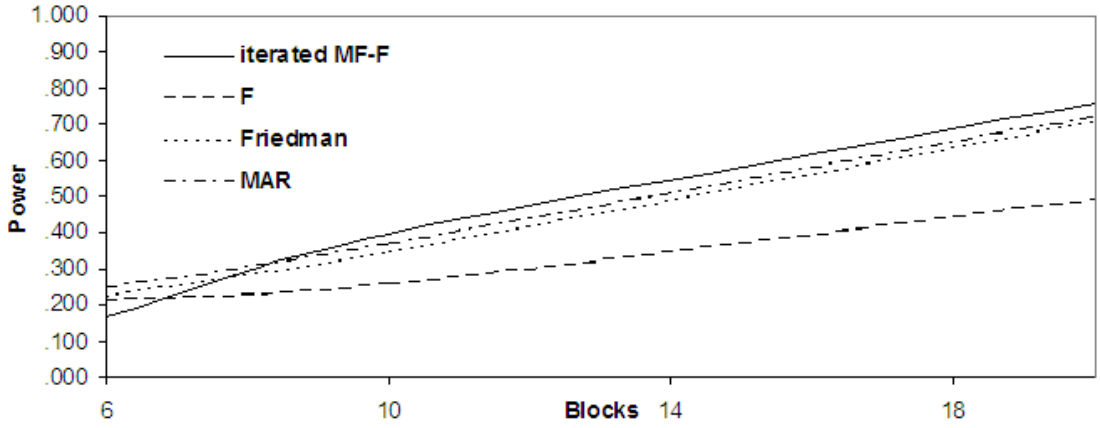

Figure 5.3: $C = 3$ Student's t errors. Power comparison among iterated Fisher MultiFocus permutation test and traditional counterparts.

5.7 Multivariate extensions

The proposed combination-based permutation approach (described in Section 5.4), can be easily extended via nonparametric combination methodology also to the multivariate case, i.e. when the response variable is multidimensional. In this case the univariate components could be mixed, it means that some could be continuous and/or ordered categorical. Also in this case we focus on ordered categorical response variables proposing a generalization of the combination-based permutation approach for testing effects within the framework of the multivariate RCB designs, considering an additive phase of combination of dependent univariate permutation tests. Following the rationale explained in Section 5.2, let us consider $\mathbf{X}$ as a response categorical variable whose support is partitioned into $m \geq 2$ ordered classes $\{A_\gamma; \gamma = 1, \ldots, m\}$, in the sense that relationships such as $A_\gamma < A_\lambda$ have a clear meaning for every pair of subscripts γ, λ such that $1 \leq \gamma < \lambda \leq m$. Let us also assume that a statistical model for $\mathbf{X}$ exists given by the tern $(\boldsymbol{\chi}, \mathcal{B}, P \in \mathcal{P})$, where $\boldsymbol{\chi}$ is the sample space of X, $\mathcal{B}$ is an algebra of events and $\mathcal{P}$ is a nonparametric family of non-degenerate probability distributions on $(\boldsymbol{\chi}, \mathcal{B})$. Let us assume that data are classified according to C treatments and n blocks with reference to the adopted RCB design and also the result of the experiment is characterized by the simultaneous observation of p response variables. Hence, the multivariate categorical response variable $\mathbf{X}$ is related to a p-dimensional vector of responses. The experimental design is developed with the aim at comparing the C treatments with respect to p-different response variables.

In what follows we refer to X_{ijk} as the univariate categorical response variable where the subscripts $i = 1, \ldots, n$, $j = 1, \ldots, C$ and $k = 1, \ldots, p$ are related to the n blocks, C

treatments and p variables respectively. In this case, considering C independent random samples $\mathbf{X}_j = \{\mathbf{X}_{ji}; i = 1, \ldots, n; j = 1, \ldots, C\}$, the testing hypotheses expressed in (5.1) and (5.2) can be rewritten as follows:

$$H_0 : \left\{ \bigcap_{k=1}^{p} X_{i1k} \overset{\mathrm{d}}{=} \ldots \overset{\mathrm{d}}{=} X_{iCk} \right\} = \left\{ \bigcap_{k=1}^{p} (F_{i1k}(A_\gamma) = \ldots = F_{iCk}(A_\gamma)), \gamma = 1, \ldots, m \right\} \ \forall i, \tag{5.22}$$

$$H_1 : \left\{ \bigcup_{k=1}^{p} (X_{i1k} \overset{\mathrm{d}}{\neq} \ldots \overset{\mathrm{d}}{\neq} X_{iCk}) \right\} = \left\{ \bigcup_{k=1}^{p} \bigcup_{\gamma=1}^{m} (F_{i1k}(A_\gamma) \neq \ldots \neq F_{iCk}(A_\gamma)) \right\} \text{ for at least one } i, \tag{5.23}$$

for $i = 1, \ldots, n$, $\gamma = 1, \ldots, m$, where $F_{ijk} = Pr\{X_{ijk} \leq A_\gamma\}$ with $j = 1, \ldots, C$ are the cumulative distribution functions (CDFs) for ordered categorical variables X_{ijk}. Also in this case (5.23) defines the non-dominance stochastic of X_{ijk} with respect to X_{ihk} with $j < h; j, h = 1, \ldots, C$ and defines the inequality distribution of X_{ijk} with respect to X_{ihk} with $j \neq h; j, h = 1, \ldots, C$.

Also in this case we conceptually *categorize* a multidimensional continuous variable $\mathbf{Y}$ by a (nonparametric) function ϕ defined as follows:

$$\mathbf{X}_{ij} = \phi(\mathbf{Y}_{ij}) \overset{\mathrm{d}}{=} \phi(\mathbf{Y}_{ih} + \boldsymbol{\Delta}_j) \ j, h = 1, \ldots, C \ j \neq h \ i = 1, \ldots, n; \tag{5.24}$$

where $\mathbf{Y}_{ij}$ represents underlying real-valued p-vector of responses, ϕ is a function which transforms Y_{ijk} into ordered categorical data X_{ijk}, and $\boldsymbol{\Delta}_j$ represents non-negative stochastic effects. So the function ϕ is a nonparametric function defined on the set Γ (i.e. the set of the p-vector of numerical responses) in the set $\mathcal{O}$ (i.e. the set of the of the p-vector of categorical responses) which maps any point of Γ in $\mathcal{O}$ such that $\phi(Y_{ijk} + \boldsymbol{\Delta}_j) \geq \phi(Y_{ijk})$, for any $j = 1, \ldots, C$, $k = 1, \ldots, p$ and $i = 1, \ldots, n$. As in the univariate case, we use this analogy in order to use the terminology adopted for quantitative variables to the case of ordered categorical variables. Now we focus our attention on the multivariate (continuous) variable $\mathbf{Y}$ related to a p-dimensional vector of responses and we consider an experimental design with n blocks and, within each block, experimental units are randomly assigned to the C ($C > 2$) treatments and exactly one experimental unit is assigned to each of the C treatments. We can represent the $\mathbf{Y}$ data as a matrix of size $n \times C \times p$ where n represents the number of blocks, C represents the number of treatments and p represents the number of the univariate response variables. The p-variate observations can be presented in an $n \times C$ table:

$$\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_j, \ldots, \mathbf{Y}_C] = \begin{bmatrix} \mathbf{Y}_{11} & \ldots & \mathbf{Y}_{1j} & \ldots & \mathbf{Y}_{1C} \\ & \ddots & & & \\ \mathbf{Y}_{i1} & \ldots & \mathbf{Y}_{ij} & \ldots & \mathbf{Y}_{iC} \\ & & & \ddots & \\ \mathbf{Y}_{n1} & \ldots & \mathbf{Y}_{nj} & \ldots & \mathbf{Y}_{nC} \end{bmatrix}$$

where $\mathbf{Y}_{ij} = [Y_{ij1}, \ldots, Y_{ijk}, \ldots, Y_{ijp}]^T$ represents the ij-th observed $p \times 1$ response for i-th block and j-th variable, $i = 1, \ldots, n$, $j = 1, \ldots, C$ ($C > 2$). Similarly to the model (5.4) the multivariate data can be described as follows:

$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\beta}_i + \boldsymbol{\tau}_j + \boldsymbol{\varepsilon}_{ij},\ \boldsymbol{\varepsilon}_{ij} \sim IID(\mathbf{0}, \boldsymbol{\Sigma})\ i = 1, \ldots, n,\ j = 1, \ldots, C \tag{5.25}$$

where: $\boldsymbol{\beta}_i$, $\boldsymbol{\tau}_j$, and $\mathbf{Y}_{ij}$, are respectively the effect of the i-th block, the effect of the j-th treatment and $\mathbf{Y}_{ij}$ is the p-dimensional multivariate response variable for the i-th block and the j-th treatment. The random term $\boldsymbol{\varepsilon}_{ij}$ represents a p-vector of experimental errors with zero mean, variance/covariance matrix $\boldsymbol{\Sigma}$ and unknown continuous distribution P. The usual side-conditions for effects are given by the constrains $\sum_i \boldsymbol{\beta}_i = \sum_j \boldsymbol{\tau}_j = \mathbf{0}$.

The effect model for the k-th univariate component of $\mathbf{Y}_{ij}$ can be written as:

$$Y_{ijk} = \mu_k + \beta_{ik} + \tau_{jk} + \varepsilon_{ijk}\ i = 1, \ldots, n;\ j = 1, \ldots, C;\ k = 1, \ldots, p. \tag{5.26}$$

Within the framework of NPC Test methodology, we remark that model (5.26) could be rewritten in a more general form considering the interaction between effects and blocks. This can be done because blocks play the role of strata, therefore it is as if each block would provide a separated test so that it is not required that $\boldsymbol{\tau_j}$'s have to be equal among blocks (under the alternative nor under the null hypothesis). In addition it is not required that errors have to be independent and identically distributed among block/subjects, but just among trials within the same block. In this setting model (5.25) fall to be as a special case of it. Note that the parametric counterparts does not extend to interaction so easily. In the framework of NPC Test we suppose that, if the global null hypothesis H_0 is true, the hypothesis of exchangeability of random errors within the same block holds.

Also in this case the *global null hypothesis* H_0 states the equality in distribution of the the multivariate distribution of the p variables in all C groups, and it is supposed to be properly decomposed into $C \times (C-1)/2$ sub-hypotheses $H_{0(jk)}$ (*partial null hypotheses*), $j, h = 1, \ldots, C$, $j \neq h$; each one related to the jh-th pair-wise comparison between couples of treatments:

$$H_0 : \left[\mathbf{Y}_1 \overset{\mathrm{d}}{=} \ldots \overset{\mathrm{d}}{=} \mathbf{Y}_C\right] = \left[\bigcap_{\substack{j,h=1 \\ j \neq h}}^{C} (\mathbf{Y}_j \overset{\mathrm{d}}{=} \mathbf{Y}_h)\right] = \left[\bigcap_{\substack{j,h=1 \\ j \neq h}}^{C} H_{0(jh)}\right]. \tag{5.27}$$

Finally, each $H_{0(jh)}$ is supposed to be properly decomposed into p sub-hypotheses $H_{0k(jh)}$, $k = 1, \ldots, p$, each appropriate for partial (univariate) aspects, thus H_0 (multivariate) is true if all the $H_{0k(jh)}$ (univariate) are jointly true:

$$H_0 : \left[\bigcap_{\substack{j,h=1 \\ j \neq h}}^{C} \bigcap_{k=1}^{p} (Y_{ijk} \overset{\mathrm{d}}{=} Y_{ihk})\right] = \left[\bigcap_{\substack{j,h=1 \\ j \neq h}}^{C} \bigcap_{k=1}^{p} H_{0k(jh)}\right],\ \forall i = 1, \ldots, n. \tag{5.28}$$

The *global alternative* hypothesis H_1 is represented by the union of *partial* $H_{1k(jh)}$ sub-alternatives:

$$H_1 : \left[\bigcup_{\substack{j,h=1 \\ j \neq h}}^{C} H_{1(jh)} \right] = \left[\bigcup_{\substack{j,h=1 \\ j \neq h}}^{C} \bigcup_{k=1}^{p} H_{1k(jh)} \right]; \tag{5.29}$$

so hat H_1 is true if at least one of sub-alternatives is true.

Following the idea described in the univariate case, we want to test the global null hypothesis H_0 and the $C \times (C-1)/2$ hypotheses $H_{0(jh)}$. The key idea comes from the partial (univariate) tests which are focused on p-th partial aspects $H_{0k(jh)}$, and then, combining them with an appropriate combining function, firstly to test $H_{0(jh)}$, $j, h = 1, \ldots, C$, $j \neq h$, and finally to test the global (multivariate) test which is referred to as the global null hypothesis H_0.

Any permutation testing procedure suggested for the univariate case (Section 5.4) can be used, in this work we use the Anderson-Darling statistic (5.16) and the well-known Pearson statistic:

$$P^{*2} = \sum_{jh} \frac{n_j \cdot [\frac{N^*_{jh}}{n_j} - \widehat{p}_h]^2}{\widehat{p}_h}, \tag{5.30}$$

where N^*_{jh} is described in (5.14) and $\widehat{p}_h = \frac{f_{.h}}{n}$ where by $f_{.h}$ we intend the marginal frequencies of the observation in the hth treatment.

We observe that in the null hypothesis H_0 the permutation distribution of (5.30), in case of large samples, is well approximated by that of a central chi-square with $C-1$ degree of freedom. By the fact that the null distribution of (5.30) depends only on exchangeable errors, it can be evaluated by a CMC method in order to obtain an exact permutation test. In general the (5.30) is the most popular test statistic for non-dominance alternative and non-ordered categorical variables while for ordered categorical variables is inadequate.

So we suggest an alternative, within the framework of permutation tests, of the Pearson statistic for ordered categorical variables given by:

$$P^{*2}_f = \sum_{j=1}^{k} (f^*_{hj} - \frac{n_h . f_{.j}}{n})^2 [f_{.j}(n - f_{.j})]^{-1}. \tag{5.31}$$

This test statistic, again Anderson-Darling approach, in H_0 is equivalent to the sum of standardized squared summands, except for a permutationally invariant coefficient and its behaviour is very close to that of the traditional chi-square. In order to obtain an estimate of the permutation distribution under H_0 of all test a CMC procedure can be used. Every resampling without replacement $\mathbf{Y}^*$ from the data set $\mathbf{Y}$ actually consists of a random attribution of the individual block data vectors to the C treatments. In every $\mathbf{Y}^*_r$ resampling, $r = 1, \ldots, B$, the $S = p \times C \times (C-1)/2$ partial tests $T_{k(jh)}$ are calculated to obtain the set of values $[T^*_{sr} = T(Y^*_{sr}), s = 1, \ldots, S; r = 1, \ldots, B]$, from the B independent

random resamplings. It should be emphasized that CMCP only considers permutations of individual data vectors within each individual block, so that all underlying dependence relations which are present in the component variables are preserved.

5.8 Some testing procedures from the literature for the multivariate RCB design

Of sure, when assuming the hypothesis of multivariate normality for random error components, the inferential problem can be solved by means of MANOVA procedure (Montgomery (2004)). Möttönen et al. (2003) purpose an extension of the Friedman and Page tests for the multivariate case, such procedures usually imply univariate rank, in the multivariate case they are replaced by spatial rank tests. The spatial centered rank of the observation, $\mathbf{r}_i$ among a $n \times C$ p-variate data-set $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ is defined as:

$$\mathbf{r}_i = 1/n \sum_{q=1}^{n} sgn(\mathbf{Y}_i - \mathbf{Y}_q), \tag{5.32}$$

where

$$sgn\,(\mathbf{Y}) = \left\{ \begin{array}{cc} \|\mathbf{Y}\|^{-1}\,\mathbf{Y} & \text{if } \mathbf{Y} \neq \mathbf{0} \\ \mathbf{0} & \text{if } \mathbf{Y} = \mathbf{0} \end{array} \right\}. \tag{5.33}$$

The $\mathbf{r}_{ij}$ defined as the centered rank of the observation with the j-th treatment in the i-th block, is possible to write $\mathbf{r} = (\mathbf{r}_1^T, \ldots, \mathbf{r}_{.C-1}^T)^T = \sum_i \mathbf{r}_i$ where $\mathbf{r}_i = (\mathbf{r}_{i1}^T, \ldots, \mathbf{r}_{i,C-1}^T)^T$ is the concatenated vector of ranks vector. The p-variate observations can be presented in an $n \times C$ table, if we consider the concatenated random vectors of the observations in each blocks $\mathbf{Y}_i = (\mathbf{Y}_{i1} \ldots \mathbf{Y}_{iC})^T$, $i = 1, \ldots, n$, we can assume the independence of $\mathbf{Y}_i$ with CDF's $F_i(\mathbf{Y}_{i1} - \tau_1, \ldots, \mathbf{Y}_{iC} - \tau_C)$, where F_i are permutation invariant and the τ_j are the treatment effects, so, within the blocks, the variables $\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{iC}$ are exchangeable under the null model.

In this framework the permutation MANOVA test (Möttönen et al. (2003)) is given by:

$$Q = \frac{C-1}{nC} \mathbf{Z}_{.j}^T D^{-1} \mathbf{Z}_{.j}, \tag{5.34}$$

where

$$\mathbf{Z}_i = (\mathbf{Z}_{i1} \ldots \mathbf{Z}_{i,C-1})^T = \left(\mathbf{Z}_{i1} - \overline{\mathbf{Z}}_{i.} \ldots \mathbf{Z}_{i,C-1} - \overline{\mathbf{Z}}_{i.}\right)^T, \tag{5.35}$$

denotes the $(C-1)p$-vector of the centered response values in the i-th block, then the sum of the centered response values over the blocks is given by:

$$\mathbf{Z} = (\mathbf{Z}_{.1}, \ldots, \mathbf{Z}_{.,C-1}) = \sum_{i=1}^{n} \mathbf{Z}_i. \tag{5.36}$$

The matrix D is given by:

$$D = \frac{1}{nC}\sum_{i=1}^{n}\sum_{j=1}^{C}\mathbf{Z}_{ij}\mathbf{Z}_{ij}^{T} \tag{5.37}$$

Möttönen et al. (2003) show that if the sequence $(\mathbf{Y}_i)$ is such that the second and third moments of the permutation random variables $\mathbf{Z}_i^*$ are uniformly bounded with $D \to \Sigma$ where Σ is of full rank then, under H_0 the limiting permutation distribution of Q^* is $\chi^2_{p(C-1)}$. In order to obtain the spatial rank tests Möttönen et al. (2003) suggest to replace the multivariate blockwise centered response vectors $\mathbf{Z}_{ij}$ by multivariate blockwise centered rank vectors. The multivariate centered ranks $\mathbf{r}_{ij}$ is the centered rank of the observation with the jth treatment among the observations in the ith block. The rank can be displayed in a table as follows:

	Treatments			
Blocks	1	2	...	C
1	$\mathbf{r}_{11}$	$\mathbf{r}_{12}$	...	$\mathbf{r}_{1C}$
2	$\mathbf{r}_{21}$	$\mathbf{r}_{22}$	...	$\mathbf{r}_{2C}$
⋮	⋮	⋮	⋱	⋮
n	$\mathbf{r}_{n1}$	$\mathbf{r}_{n2}$	...	$\mathbf{r}_{nC}$

(5.38)

If we consider the concatenated vector of rank vectors in the ith block $(\mathbf{r}_i = (\mathbf{r}_{i1}^T, \ldots, \mathbf{r}_{i,C-1}^T))$ the vector of the rank sums over the blocks will be $\mathbf{r} = (\mathbf{r}_{\cdot 1}^T, \ldots, \mathbf{r}_{\cdot,C-1}^T) = \sum_i \mathbf{r}_i$. The permutation random variable of $\mathbf{r}$ is $\mathbf{r}^*$ and of $\mathbf{r}_{ij}$ is $\mathbf{r}_{ij}^*$ ($\mathbf{r}_{i1}, \ldots, \mathbf{r}_{iC}$ are exchangeable within the blocks). $E^*(\mathbf{r}_{ij}^*) = \mathbf{0}$, $Var^*(\mathbf{r}_{ij}^*) = \frac{1}{C}\sum_{j=1}^{C}\mathbf{r}_{ij}\mathbf{r}_{\cdot ij}^T$, $Cov^*(\mathbf{r}_{ij}^*, \mathbf{r}_{ij''}^*) = \frac{1}{C(C-1)}\sum_{j=1}^{C}\mathbf{r}_{\cdot ij}\mathbf{r}_{\cdot ij}^T$. We obtain that $E^*(\mathbf{r}^*) = \mathbf{0}$, $S_r^{-1} = Var^*(\mathbf{r}^*) = \frac{n}{C-1}(C(\mathbf{I}_{C-1}) - (\mathbf{J}_{C-1}) \otimes (\mathbf{G}_r)$ where $(\mathbf{G}_r) = \frac{1}{nC}\sum_{i=1}^{n}\sum_{j=1}^{C}\mathbf{r}_{\cdot ij}\mathbf{r}_{\cdot ij}^T$ and $\otimes$ is the symbol for the Kronecker product. Multivariate extensions of the well known Friendman's and Page's tests are defined as the Multivariate Friedman Statistic given by $Q_r = \mathbf{r}^T\mathbf{S}_r^{-1}\mathbf{r}$ and the Multivariate Page test statistic defined as $L = \mathbf{r}_{\cdot 1} + 2\mathbf{r}_{\cdot e} + \ldots + C\mathbf{r}_{\cdot C}$.

Möttönen et al. (2003) show that under H_0 the limiting permutation distribution of Q_r^* is a χ^2-distribution with $p(C-1)$ degrees of freedom and the limiting permutation distribution of $P^* = \frac{12}{nC^2(C+1)}\mathbf{L}^{*T}G_r^{-1}\mathbf{L}^*$ **has a central** χ^2-distribution with p degrees of freedom.

5.9 Comparative simulation study

The permutation solution for the multivariate RCB have been evaluated with the parametric (MANOVA) and nonparametric testing procedures (Multivariate Friedman tests

and Multivariate Page test) by a comparative simulation study. The simulation study is designed in the following way:

- 1,000 independent simulations;
- number of blocks: $n = 5, 10, 20$; number of treatments: $C = 2, 3, 4, 5$;
- block effect β_i, $i = 1, \ldots, n$, is generated from discrete uniform distribution (independently for all univariate components);
- number of response variables is $p = 3$;
- number of response variables where treatment effect is active: $1, 2, 3$; by 'active variable' we mean that for the kth variable, $k = 1, 2, 3$, true means differ from each other (following the pattern detailed below) so that $\exists k : \tau_{jk} \neq 0, j = 1, \ldots, C, k = 1, 2, 3$; instead, we say that the treatment is not active for the kth variable when all τ_{jk} are set equal to zero;
- at univariate level the treatment effects are set as follows:
  - two treatments ($C = 2$): $\mu_1 = 0$ and $\mu_2 = 3$;
  - three treatments ($C = 3$): $\mu_1 = 0$, $\mu_2 = 1$ and $\mu_3 = 3$;
  - four treatments ($C = 4$): $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$ and $\mu_4 = 3$;
  - five treatments ($C = 5$): $\mu_1 = 0$, $\mu_2 = 0.5$, $\mu_3 = 1$,$\mu_4 = 2$ and $\mu_4 = 3$;
- three types of random errors: normal, skew normal (as an example of an asymmetric distribution) and Student's t with 3 degrees of freedom (as an example of a heavy-tailed distribution); each one of the $p = 3$ random component ε_{ijk}, $i = 1, \ldots, n$, $j = 1, \ldots, C$, $k = 1, 2, 3$, are generated with a variance/covariance structure as follows:

$$\Sigma = \begin{bmatrix} 1.25 & -0.5 & 0.5 \\ -0.50 & 1.25 & -0.25 \\ 0.5 & -.25 & 1.25 \end{bmatrix} \tag{5.39}$$

  finally, in order to better represent a genuine ordinal scale, before being added to the true effects the random errors were rounded to the nearest integer;
- the considered significance α-level is equal to 0.05.

We perform a multivariate permutation test ($1,000$ conditional Monte Carlo iterations, CMC) for each simulation using three types of combining function: Fisher, Tippett and the second version of Liptak combing function; we also perform the parametric MANOVA two way layout (for the RCB) and the nonparametric Multivariate Friedman Test and Multivariate Page test as counterparts (Section 5.11).

Tables 5.6, 5.7 and 5.8 report the rejection rates and nominal levels of the permutation tests under the three different distributions of the error components (multivariate normal errors in Table 5.6, multivariate Student's t errors in Table 5.7 and multivariate Skew-normal errors in Table 5.8). In the first column we report the type of test statistic, Anderson-Darling statistic(AD) (5.14) and Pearson statistic for ordered categorical variables (Pears) (5.31). We added to the statistic's label the terms 'FwF', 'LwL'and 'TwT' to indicate that we used the Fisher, Tippett and the logistic version of the Liptak combining function for the first and second phase of combination (so for example 'FwF' means that we used the Fisher combination function in all the phases of combination). We underline that in the Tables 5.6, 5.7 and 5.8 the term *Var.* specifies the number of active variables which occurs in the experiment.

As first remark of the simulation study we observe that, from a general point of view, all the procedures under the null hypothesis appear to behave properly according to the nominal level. We observe also that, the power of the global test increases when the number of blocks and treatments increases.

From Tables 5.6, 5.7 and 5.8 we note that the Anderson-Darling statistic with the Fisher and the Liptak (second version) combining functions appears to be more powerful than the others permutation test statistics implemented, we observe also that among the permutation tests the use of the Tippett combining function seems to produce a loss of power.

In Figures 5.4, 5.5 and 5.6 we report the power behaviour tests under multivariate normal errors (Figure 5.4), multivariate Student's t errors (Figure 5.5) and multivariate Skew-normal errors (Figure 5.6) of the combination-based permutations testing procedure (Permutation 1 and Permutation2), of the traditional MANOVA procedure (MANOVA) and of the nonparametric competitors multivariate Friedman and multivariate Page tests. Note that with the statistic's label 'Permutation1' and 'Permutation2' we refer to the Pearson test statistic for ordered categorical variables and to the Anderson-Darling test statistic respectively, using the Fisher's combination function for all the phases of combination.

We observe that all the combination-based statistics provide greater power than the traditional MANOVA under non normal errors. As expected,MANOVA procedure has a good behaviour under normality of the error components. As shown in Figure 5.4, the combination-based permutation provide greater power than all the other procedure under multivariate Student's t errors. The results suggest that all the combination-based permutation testings with the Fisher and the Liptak (second version) combining functions seems to be satisfactory, in particular the Anderson-Darling statistic using the Fisher's combination function. Among the nonparametric competitors the worst one is the multivariate Friedman test statistic, but we observe that it seems to be a valid solution under asymmetrical distributions of the error components. In general we can say that permutation tests provide greater powers under multivariate Student's t errors, while under multivariate

Skew-normal errors they seem loss power respect to the nonparametric competitors.

		Rejection Rates												Nominal Levels			
		$C=2$			$C=3$			$C=4$			$C=5$						
		Var.			Var.			Var.			Var.			C			
Test	n	1	2	3	1	2	3	1	2	3	1	2	3	2	3	4	5
	5	.012	.288	.303	.012	.292	.307	.012	.296	.312	.013	.301	.317	.013	.013	.013	.013
Pears.	10	.570	.662	.666	.579	.672	.676	.588	.683	.687	.596	.693	.697	.022	.022	.023	.023
(LwL)	20	.667	.691	.793	.678	.701	.805	.688	.711	.817	.698	.722	.829	.043	.044	.044	.045
	5	.012	.290	.305	.012	.295	.310	.012	.299	.315	.013	.303	.319	.013	.013	.013	.013
AD	10	.575	.668	.672	.584	.678	.682	.593	.689	.693	.602	.699	.703	.022	.023	.023	.023
(LwL)	20	.673	.697	.800	.683	.928	.972	.852	.972	.972	.704	.728	.836	.043	.044	.045	.045
	5	.011	.262	.276	.011	.266	.280	.011	.270	.284	.011	.274	.289	.012	.012	.012	.012
Pears.	10	.520	.604	.608	.528	.613	.617	.536	.622	.626	.544	.632	.636	.020	.020	.021	.021
(TwT)	20	.609	.630	.723	.618	.639	.734	.627	.649	.745	.637	.658	.756	.039	.040	.040	.041
	5	.011	.265	.278	.011	.269	.283	.011	.273	.287	.012	.277	.291	.012	.012	.012	.012
AD	10	.525	.609	.613	.532	.619	.622	.540	.628	.632	.549	.637	.641	.020	.021	.021	.021
(TwT)	20	.614	.635	.729	.623	.847	.887	.777	.887	.887	.642	.664	.763	.040	.040	.041	.041
	5	.012	.287	.302	.012	.291	.307	.012	.296	.311	.013	.300	.316	.013	.013	.013	.013
Pear.	10	.569	.661	.665	.578	.671	.675	.586	.681	.685	.595	.691	.695	.022	.022	.023	.023
(FwF)	20	.666	.689	.791	.676	.788	.803	.686	.801	.815	.696	.889	.890	.043	.044	.044	.045
	5	.012	.290	.305	.012	.294	.309	.012	.298	.314	.013	.303	.319	.013	.013	.013	.013
AD	10	.574	.667	.671	.583	.677	.681	.591	.687	.691	.600	.697	.701	.022	.023	.023	.023
(FwF)	20	.672	.695	.798	.682	.801	.970	.850	.879	.970	.703	.901	.834	.043	.044	.045	.045

Table 5.6: Permutation Tests: Rejection rates and nominal levels under multivariate normal errors

		Rejection Rates												Nominal Levels			
		$C=2$			$C=3$			$C=4$			$C=5$						
		Var.			Var.			Var.			Var.			C			
Test	n	1	2	3	1	2	3	1	2	3	1	2	3	2	3	4	5
	5	.013	.301	.316	.013	.305	.321	.013	.310	.326	.013	.314	.331	.013	.014	.014	.014
Pear.	10	.596	.692	.697	.605	.703	.707	.614	.713	.718	.623	.724	.728	.023	.023	.024	.024
(LwL)	20	.698	.722	.829	.708	.962	1.000	.882	1.000	1.000	.729	.755	.866	.045	.046	.046	.046
	5	.013	.301	.317	.013	.306	.322	.013	.310	.327	.013	.315	.332	.013	.014	.014	.014
AD	10	.597	.694	.698	.606	.704	.709	.615	.715	.719	.625	.726	.730	.023	.023	.024	.024
(LwL)	20	.699	.723	.830	.710	.964	1.000	.884	1.000	1.000	.731	.756	.868	.045	.046	.047	.047
	5	.011	.274	.288	.012	.278	.293	.012	.282	.297	.012	.287	.302	.012	.012	.013	.013
Pear.	10	.543	.631	.635	.552	.641	.645	.560	.650	.654	.568	.660	.664	.021	.021	.022	.022
(TwT)	20	.636	.658	.756	.646	.877	.918	.805	.918	.918	.665	.688	.790	.041	.042	.042	.042
	5	.011	.275	.289	.012	.279	.293	.012	.283	.298	.012	.287	.302	.012	.012	.013	.013
AD	10	.545	.633	.637	.553	.642	.646	.561	.652	.656	.570	.662	.666	.021	.021	.022	.022
(TwT)	20	.638	.660	.757	.647	.879	.921	.806	.921	.921	.667	.690	.792	.041	.042	.042	.042
	5	.012	.297	.313	.013	.301	.317	.013	.306	.322	.013	.311	.327	.013	.013	.014	.014
Pear.	10	.589	.684	.688	.598	.694	.699	.607	.705	.709	.616	.715	.720	.023	.023	.023	.023
(FwF)	20	.689	.713	.819	.700	.917	.995	.872	.956	.995	.721	.966	.996	.045	.045	.046	.046
	5	.012	.298	.313	.013	.302	.318	.013	.307	.323	.013	.311	.328	.013	.013	.014	.014
AD	10	.590	.686	.690	.599	.696	.700	.608	.706	.711	.617	.717	.721	.023	.023	.024	.024
(FwF)	20	.691	.765	.821	.701	.953	.997	.874	.997	.997	.722	.991	1.000	.045	.045	.046	.046

Table 5.7: Permutation Tests: Rejection rates and nominal levels under multivariate Student's t errors

		Rejection Rates												Nominal Levels			
		$C=2$			$C=3$			$C=4$			$C=5$						
		Var.			Var.			Var.			Var.			C			
Test	n	1	2	3	1	2	3	1	2	3	1	2	3	2	3	4	5
	5	.012	.286	.301	.012	.290	.305	.012	.294	.310	.012	.299	.314	.013	.013	.013	.013
Pear.	10	.566	.658	.662	.575	.668	.672	.584	.678	.680	.592	.688	.692	.022	.022	.023	.023
(LwL)	20	.663	.686	.787	.673	.914	.957	.839	.957	.957	.693	.717	.823	.043	.043	.044	.044
	5	.012	.286	.301	.012	.291	.306	.012	.295	.310	.013	.300	.315	.013	.013	.013	.013
AD	10	.568	.660	.664	.576	.670	.674	.585	.680	.684	.594	.690	.694	.022	.022	.023	.023
(LwL)	20	.665	.688	.789	.675	.916	.960	.841	.960	.960	.695	.719	.825	.043	.044	.044	.044
	5	.011	.261	.274	.011	.264	.278	.011	.268	.282	.011	.272	.287	.012	.012	.012	.012
Pear.	10	.517	.600	.604	.524	.609	.613	.532	.618	.622	.540	.628	.631	.020	.020	.021	.021
(TwT)	20	.605	.626	.718	.614	.834	.873	.765	.873	.873	.632	.654	.751	.039	.040	.040	.040
	5	.011	.261	.275	.011	.265	.279	.011	.269	.283	.011	.273	.287	.012	.012	.012	.012
AD	10	.518	.602	.605	.526	.611	.614	.533	.620	.623	.541	.629	.633	.020	.020	.021	.021
(TwT)	20	.606	.627	.720	.615	.836	.875	.767	.875	.875	.634	.656	.753	.039	.040	.040	.040
	5	.012	.261	.296	.012	.285	.300	.012	.289	.304	.012	.294	.309	.012	.013	.013	.013
Pear.	10	.557	.647	.651	.565	.657	.661	.574	.666	.670	.582	.676	.681	.022	.022	.022	.022
(FwF)	20	.652	.699	.774	.662	.918	.941	.824	.970	.975	.682	.958	1.000	.042	.043	.043	.043
	5	.012	.269	.296	.012	.286	.301	.012	.290	.305	.012	.294	.310	.012	.013	.013	.013
AD	10	.558	.652	.662	.567	.658	.662	.575	.668	.672	.584	.678	.682	.022	.022	.022	.022
(FwF)	20	.653	.703	.776	.663	.934	.943	.826	.978	.980	.683	.960	1.000	.042	.043	.043	.043

Table 5.8: Permutation Tests: Rejection rates and nominal levels under multivariate Skew-normal errors

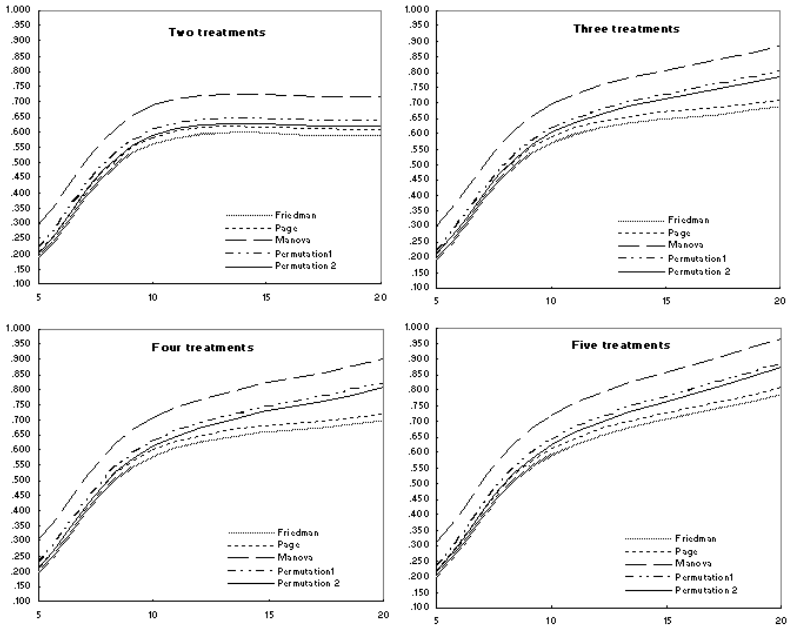

Figure 5.4: Power Comparison - multivariate normal errors

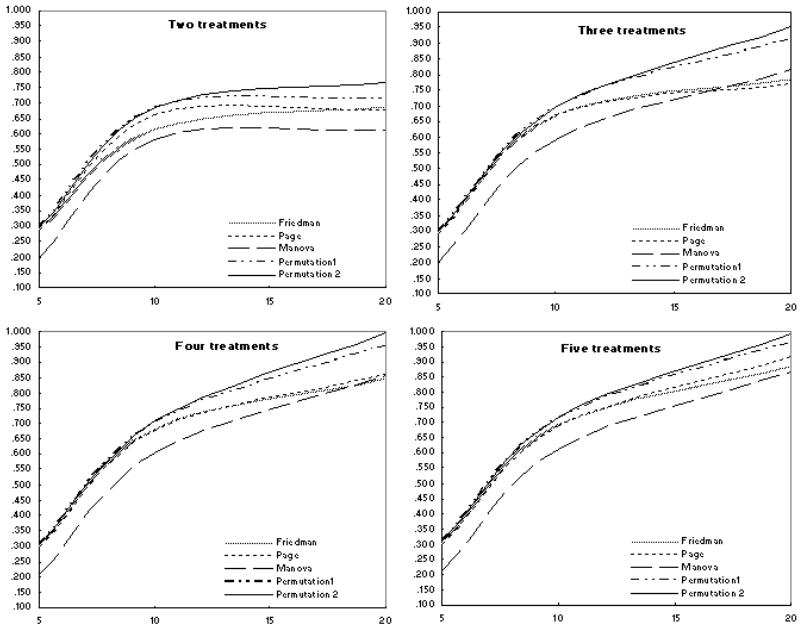

Figure 5.5: Power Comparison - multivariate Student's t errors

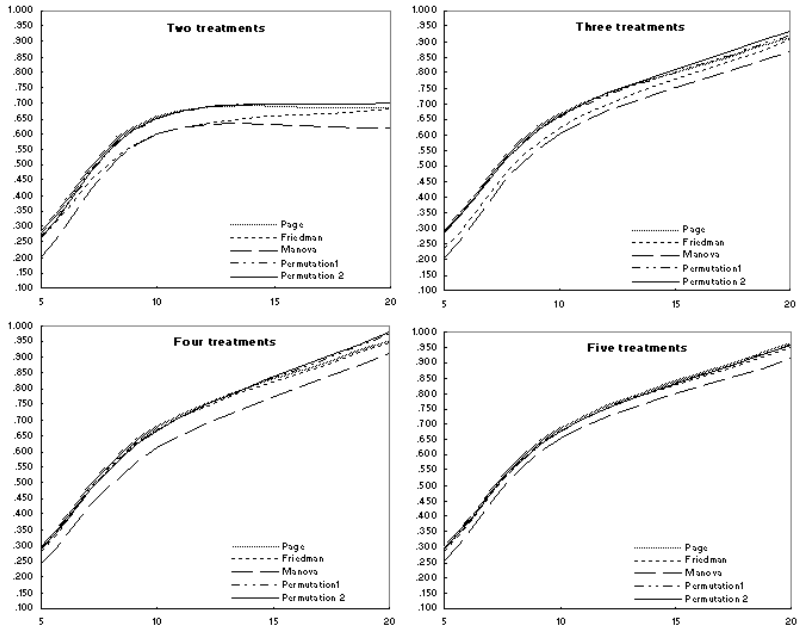

Figure 5.6: Power Comparison - multivariate Skew-normal errors

5.10 Examples of the permutation methodologies for RCB designs in sensorial evaluation studies

The comparison of the flavour of meat from three breeds of geese

As an example of the combination-based permutation approach for testing effects in the (univariate) RCB design we now deal with a real case study proposed in the literature.

Suppose, as in Lamond (1970), that we wish to compare the flavour of meat from three breeds of geese X, Y, and Z on a five point scale with categories ranging from 'excellent' to 'very poor' and that the data from eight consumers shown in Table 5.9 are obtained, where we have labeled the ordered categories as 1-5 scores. When applying the considered RCB procedures to meat flavour data we can obtain results reported in Table 5.10, where we performed pairwise comparisons only if the global test had been rejected ($\alpha = 0.05$).It is interesting to observe that not all procedures agree to reject the global null hypothesis ($\alpha = 0.05$). Moreover, the application of the iterated combination approach implies a strengthening of evidence against the null hypothesis of equal flavour of meat from three breeds of geese.

Consumer	X	Y	Z	Consumer	X	Y	Z
1	3	2	3	5	2	4	2
2	4	5	4	6	1	3	3
3	3	2	3	7	2	5	4
4	1	4	2	8	2	5	2

Table 5.9: Category ratings for meat flavour for three breeds of geese

Test	p-value	
F	.028	
Friedman	.152	
MAR	.158	
		iterated comb.
MF-F	.037	.025
MF-L	.028	.022
MF-T	.204	.122
AD	.201	.164
AD2	.153	.119

Table 5.10: p-values by testing procedure from flavour meat data.

An application to a real case study

The R&D division of a home-care company is studying 5 possible new fragrances (labeled r, s, t, v, w) of a given detergent to compare with their own presently marketed product (labeled x). The experiment is designed as follows: after testing one given product (using sense of smell), the panelist assigned three different scores to it, describing the three most important aspects of the product: Strength – S $(1-5$ points), Pleasantness – $P(1-5)$, Appropriateness – A(Yes, No).

The same experiment is replicated under different assessment conditions (Bloom - B, Dry - D, Long - L, Neat - N and Wet - W), which should represent the situations in which the final customers will make use of the product.

The graph in Figure 4.7 shows the described complex RCB design with its comparison reference between fragrance r and x.
Figure 4.7 also displays the idea behind the solution to this problem that we propose using NPC Test methodology: for comparison between fragrance r and x, first of all a set of 15 univariate permutation tests is computed, where each test ${}_iT_j$, $i = B, D, L, N, W$, $j = S, P, A$, takes into account for comparison of one given aspect (Strength – S, Pleasantness – P, Appropriateness – A) within a given condition (Bloom - B, Dry - D, Long - L, Neat - N and Wet - W).

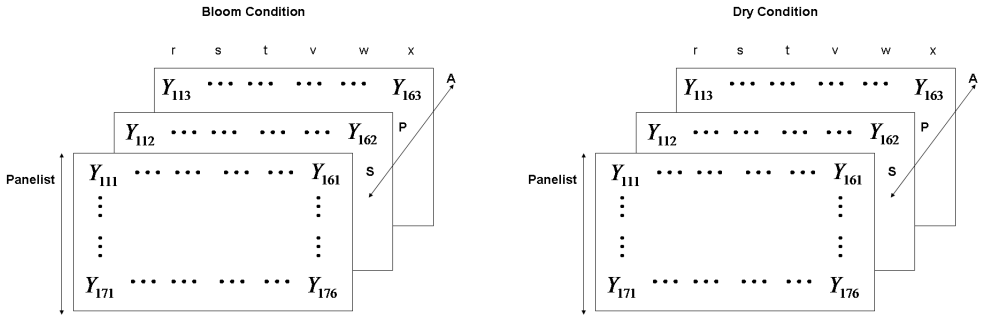

Figure 4.7: Experimental design and scheme of the permutation solution.

The next step relates to a multivariate comparison of fragrance r and x for each condition, which can be resolved by test ${}_iT$, $i = B, D, L, N, W$, obtained via nonparametric combination of the univariate test for the three aspects. The final step is a global multivariate test T, obtained by a final further nonparametric combination of the five tests ${}_iT$, $i = B, D, L, N, W$.

As an application of the proposed solution to real data, Table 5.11 displays the whole set of results from a real experiment with 7 panelists where we compare five fragrances with fragrance x and univariate permutation p-values have been corrected by multiplicity.

Table 5.11: Results for comparison between five new fragrances (r,s,t,v,w) with the presently marketed product (x)

Procedure	Glob.	r vs x	s vs x	t vs x	v vs x	w vs x
MANOVA	.039	.032	.032	.032	.032	.032
Friedman	.039					
Page	.039					
Permutation	.011	.027	.042	.059	.015	.000

5.11 Final remarks

We presented a combination-based permutation solution for hypothesis testing within the framework of the randomized complete block design. The proposed solution may suggest to practitioners in the field of evaluation for educational services and quality of products an effective approach, especially when using ordered categorical variables, such as in the case of sensorial evaluations.

As confirmed by the presented simulation studies, the nonparametric tests are certainly good alternatives, in particular respect to the traditional parametric procedures.

In fact, even in case of normality, the power of permutation tests is nearly the same as that of the parametric tests, while in case of asymmetric or heavy tailed error distributions permutation tests can provide higher power.

Hence, in each practical situation where the normality assumption is hard to justify, the proposed nonparametric procedure can be considered a valid solution.

Finally, as suggested by the real case study, a possible way to improve power of permutation tests is to better investigate the role of the combining functions. Note that our proposed permutation test applies a combining function two times: at first in order to combine the partial pairwise permutation tests to obtain a global test, then we apply a combining function in order to perform a suitable multiplicity correction strategy for pairwise permutation p-values.

The generalization of the combination-based permutation approaches in the case of multidimensional response variables is presented.

As suggested by the developed case study, we can confirm that the proposed solution is a good alternative in particular, as confirmed by the comparative simulation study performed, the proposed approach is quite flexible and may provide both partial and global indicators for investigating the performance of new products to be developed.

Chapter 6

Split-plot designs

6.1 Introduction

Split-plot (SP) designs result when a particular type of restricted randomization has occurred during a planned experiment. A simple factorial experiment can result in a SP-type design because of the way the experiment was in fact carried out. For example, in many industrial experiments some of the factors of interest may be time-cost expensive while the remaining factors are easy to set up. As a result, the order in which the treatment combinations for the experiment are run is determined by the ordering of these time-cost expensive factors. Likewise, an experiment might involve experimental units which need to be processed as a batch for one or more of the factors in a particular treatment combination. Finally, in some situations experimental units might be processed individually, one right after the other, for the same treatment combination without resetting the factor settings for that treatment combination. Classic SP experiments are usually run in replicates in a number of blocks, so that the general structure of these experiments involves several treatment effects (that are considered fixed) and block effects (that are considered random); given that the SP structure is of a mixed type, there is an additional variance term component to be considered when testing for effects. In general, there are three principal structures in a SP design which are to be taken into account:

1. Treatment Structure: as with other factorial designs, there is usually a factorial structure where all treatment combinations occur in the design. The analysis model will have terms corresponding to the main effects and interactions of the factors.

2. Experimental Units Structure: the key feature of the split-plot design is that there are two sizes of experimental units, i.e. main-plots and sub-plots. The main-plots can be arranged in either a completely randomized design (CRD) or in blocks as part of a randomized complete block (RCB) design. The sub-plots are always smaller portions of the main-plots.

3. Randomization Structure: two separate randomizations are carried out in a split-plot design. At the upper level, the main-plot treatments are randomly assigned to main-plot units. This may be done as a CRD or as an RCB. Usually, an equal number of replicates of each main-plot treatment is carried out. At the lower level, the sub-plot treatments are randomly assigned within each main-plot and the randomization is done independently for each main-plot.

The aim of this work is to present a novel permutation approach for testing effects in the framework of SP designs (Corain, Ragazzi and Salmaso (2010)), the innovative approach is formally presented in Section 6.4. Section 6.2 is devoted to the formalization of SP designs. In Section 6.3 an overview on testing procedures for split plot experiments is presented. Section 6.5 illustrates a comparative simulation study where the proposed permutation approach is compared with the parametric and nonparametric counterparts proposed in the literature. Finally, Section 6.7 is dedicated to conclusions and final remarks.

6.2 Split-plot designs

Consider the case of a SP experiment which includes a factor A with I levels, and a factor B with J levels. Consider A as a whole-plot factor and B as a sub-plot factor (Naes et al., (2006)). As usual in SP designs there are r blocks (or replications). The design can be represented by the following model:

$$Y_{ijl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_l + E_{il} + \varepsilon_{ijl}, \tag{6.1}$$

with $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $l = 1, \ldots, r$; μ is the general mean; α_i are the effects of the whole-plot factor (fixed), β_j are the effects of the sub-plot factor (fixed), $(\alpha\beta)_{ij}$ are the interaction effects between A and B, δ_l are the replicate/block effects (random effect) and E_{il} is the whole-plot random error term which accounts for interaction between α_i and block l (with $Var(E_{il}) = \sigma^2_{whole}$). The term ε_{ijl} is the regular (or sub-plot) error with variance $Var(\varepsilon_{ijl}) = \sigma^2_{sub}$. Given that some of the effects are random, the testing procedure for the different effects are based on different error terms. For this model, the whole-plot main effects are tested against the whole-plot mean square (MS) error, while the sub-plot effects and interactions between whole-plots and sub-plots are tested against the regular residual or sub-plot MS. In general it is possible to assume that all split-plot models can be cast into the following general model (Naes et al. (2006)):

$$\mathbf{Y} = \mathbf{Db} + \boldsymbol{\delta} + \mathbf{e}, \tag{6.2}$$

where $\mathbf{Y}$ is the vector of response measurements, $\mathbf{D}$ is the design matrix for the experimental factors, $\mathbf{b}$ is the vector of corresponding regression coefficients, $\boldsymbol{\delta}$ is the vector of whole-plot error terms and $\mathbf{e}$ is the vector of sub-plot errors. The terms $\boldsymbol{\delta}$ and $\mathbf{e}$ can be combined into one error term: $\mathbf{e}^* = \boldsymbol{\delta} + \mathbf{e}$. Furthermore, we assume that all elements in $\mathbf{e}$ are uncorrelated and uncorrelated between elements of $\boldsymbol{\delta}$; while the elements in $\boldsymbol{\delta}$ are

identical within the sample whole-plot and uncorrelated between whole-plots. Therefore, $Var(\boldsymbol{\delta}) = \sigma^2_{whole}\mathbf{J} + \sigma^2_{sub}\mathbf{I}$ where $\mathbf{J}$ is a $(p \times p)$ diagonal matrix with ones in the blocks and zero elsewhere and $\mathbf{I}$ is a $(p \times p)$ identity matrix ($p = I \cdot J \cdot r$ is the dimension of Y). Each block in the block diagonal matrix corresponds to the elements within the same whole-plot. We observe that, in the model (6.1), a random block effect is also present in the model and this factor will then be a part of the design matrix $\mathbf{D}$. The general mixed linear model (6.3) avoids this ambiguity:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu}+\mathbf{e}, \tag{6.3}$$

where $\mathbf{Xb}$ represents only the fixed effects of the design, $\mathbf{Zu}$ represents the random effects and $\mathbf{e}$ is the vector of uncorrelated residual errors. The covariance matrix of $\mathbf{Zu} + \mathbf{e}$ is identical to the one in equation (6.2) because there is no random replicate effect in the model.

6.3 An overview of testing procedures for split-plot experiments

The simplest way to analyze SP designs is to ignore the split-plot structure in the design, an approach which has been investigated by Letsinger et al. (1996) and by Kowalsky et al. (2002). Essentially the regular ordinary least squares (OLS) estimate is used for $\mathbf{b}$ (equation (6.2)) and the covariance matrix is calculated as $(\mathbf{X}^T\mathbf{X})^{-1}\sigma^2_{sub}$. The results indicate that when the whole-plot error is smaller than the residual error, this approach is suggested, but when the ratio between σ^2_{whole} and σ^2_{sub} is larger than unity, the OLS should not be used.

The classic testing approach for SP is based on suitable ANOVA procedures and F tests (Naes et al. (2006); Montgomery (2001)). The sub-plot error variance is obtained by MS_{sub} (residual mean square) and the whole-plot error variance is obtained by using the difference between the sum of squares for the saturated and reduced models obtained only from the fixed effects believed to be important (Letsinger et al. (1996)). The expected corresponding $E(MS_{dif})$ is equal to:

$$E(MS_{dif}) = \sigma^2_{sub} + m\sigma^2_{whole}, \tag{6.4}$$

where m is the number of sub-plot experiments per whole-plot combination. After calculations we find that

$$\widehat{\sigma}^2_{sub} = (MS_{dif} - MS_{sub})/m, \tag{6.5}$$

is an unbiased estimator for the whole-plot error variance (for more details see Naes et al. (2006) and Montgomery (2001)). When it is not possible to assume that there are not interactions between whole-plot and split-plot factors, and when it is not possible to assume compound symmetry (the cumulative distribution function of Y_{ijl} is invariant under permutations of the error components among themselves), Koch (1969) gives a

summary of the parametric and nonparametric procedures in such a case. In that paper Koch considers only the situation in which the standard assumptions regarding normality and variance homogeneity may not hold. With reference to the procedures suggested by Koch (1969), the ν different treatments correspond in model (6.1) to the I whole-plot treatments, the p different conditions correspond to the J split-plot treatments, and finally the N subjects correspond to the all combination factor levels $(I \times J)$. Now consider an experiment that involves N subjects, ν treatment groups and p conditions. Each of N subjects has been randomly assigned to one of ν treatment groups and an observation has been made on his or her response to each of p conditions. Let $Y_{ij}^{(k)}$ be the response of the j-th individual in the i-th treatment group to the k-th condition where $i = 1, \ldots, \nu$; $j = 1, \ldots, n_i$; $k = 1, \ldots, p$; so $N = \sum_{i=1}^{\nu} n_i$.

The data matrix will be:

$$\mathbf{Y}_{ij}^T = (Y_{ij}^{(1)}, \ldots, Y_{ij}^{(p)}), \tag{6.6}$$

for $i = 1, \ldots, \nu$ and $j = 1, \ldots, n_i$. The cumulative distribution function (CDF) of $\mathbf{Y}_{ij}$ is continuous and p-variate, denoted by $F_i(\mathbf{y}) = G(\mathbf{y} - \mathbf{m}_i)$ with $\mathbf{m}_i' = (m_i^{(1)}, \ldots, m_i^{(p)})$ representing the respective median vectors and indicative of the locations of the distributions while G is a function that characterizes their shape. The different $\mathbf{Y}_{ij}$ are statistically independent. The parametric analysis requires the basic assumption that G is a p-variate normal distribution with a null mean and unknown positive definite covariance matrix $\mathbf{\Sigma}$. This implies that $\mathbf{Y}_{ij}$ are independently distributed according to the multivariate normal distribution, i.e. $\mathbf{Y}_{ij} \sim N(\mathbf{m}_i, \mathbf{\Sigma})$, we note the homoscedasticity assumption.

The nonparametric analysis does not require the above normality assumption. Within this framework we want to test:

$$H_{ot} : \mathbf{m}_1 = \ldots = \mathbf{m}_\nu. \tag{6.7}$$

To this purpose, if continuity is assumed, we may use the following statistic:

$$L = (\frac{N-1}{N}) \sum_i n_i (\overline{R}_{i.} - \frac{N+1}{2} j)' V_N^{-1} (\overline{R}_{i.} - \frac{N+1}{2} j), \tag{6.8}$$

which is a generalization of the well-known Kruskal-Wallis test. By $\overline{R}_{i.}^{(k)}$ we mean the average ranks, so $\overline{R}_{i.}^{(k)} = 1/n_i \sum_j R_{ij}^{(k)}$ and

$$R_{ij}^{(k)} = \left[\text{Rank of } Y_{ij}^{(k)} \text{in the set } \left\{Y_{11}^{(k)}, \ldots, Y_{\nu n_\nu}^{(k)}\right\}\right], \tag{6.9}$$

where ties are handled by the mid-rank test. When H_{0t} is true, all vectors $\mathbf{Y}_{ij}$ have the same distribution. Hence, the joint distribution of all vectors is invariant under the different possible assignments of the vectors to treatment groups. This invariance generates a set of $N!$ equally likely realizations for the rank vectors $\mathbf{R}_{ij}$. Under this conditional distribution (P_{Nt}) we have:

$$E(\mathbf{R}_{ij} \mid P_{Nt}) = \frac{N+1}{2} j, \tag{6.10}$$

$$Var(\mathbf{R}_{ij} \mid P_{Nt}) = \frac{1}{N}\sum_i \sum_j (R_{ij}\frac{N+1}{2}j)(R_{ij}\frac{N+1}{2}j) \equiv \mathbf{V}_N, \tag{6.11}$$

where $(i^{'}, j^{'}) \neq (i, j)$ and we assume that $\mathbf{V}_N$ is non-singular.

Under the permutation approach, model (6.8) has equally likely (conditional) realizations and this may be used as a basis for a permutationally distribution free test for H_{0t}. The permutation distribution of L is approximately a chi-square with $p(\nu - 1)$ degrees of freedom. The data matrix can be viewed as a set of mixed models, each of which corresponds to a particular treatment (whole-plot). We now wish to test that in each of these there are no condition effects, so the null hypothesis is:

$$H_{oc} : m_1^{(1)} = \ldots = m_\nu^{(p)}. \tag{6.12}$$

This states that within each of the treatments there are no condition effects. First of all we require the assumption of diagonal symmetry. We first make the transformation of variables $\mathbf{U}_{ij} = \mathbf{C}_1\mathbf{Y}_{ij}$ where $\mathbf{C}_1$ is a $(p-1)\times p$ matrix whose rows are linearly independent contrasts (so $\mathbf{C}_1\mathbf{j} = 0$ if $\mathbf{j}$ is a vector - dimension p - of ones); we observe that if H_{0c} is true then $\mathbf{U}_{ij} \sim N(\mathbf{0}_{p-1}, \mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_1')$.

If the sample size (ni) is not too small, Koch (1969) suggests using tests that use the information associated with all signed rank tests (Wilcoxon (1949)) corresponding to all possible pairs of conditions. They have the form:

$$W_{n_i}^* = T_{n_i}^{*\prime} C_1' \left[C_1 V_{n_i}^* C_1'\right]^{-1} C_1 T_{n_i}^*, \tag{6.13}$$

where:

$$T_{n_i}^{(k)*} = 1/n_i \sum_j S_{ij}^{(k)},$$

$$S_{ij}^{(k)} = \sum_{k'=1}^{p} \left\{\text{sign}(Y_{ij}^{(k)} - Y_{ij}^{(k')})\right\} \left[\text{Rank of} \left|Y_{ij}^{(k)} - Y_{ij}^{(k')}\right|\right]$$

$$xmin \left\{\left|Y_{ij}^{(k)} - Y_{ij}^{(k')}\right|, \ldots, \left|Y_{in_i}^{(k)} - Y_{in_i}^{(k')}\right|\right\},$$

$$v_{n_i;kk'}^* = 1/n^2 \sum_j S_{ij}^{(k)} S_{ij}^{(k')}, \qquad V_{n_i}^* = (v_{n_i;kk'}^*)$$

For large values of n_i the statistic W_{ni}^* is a chi-square with $(p-1)$ degrees of freedom. The underlying permutation model P_{N_c} is due to the 2^{n_i} equally likely (conditional) realizations associated with sign invariance imposed by the condition of diagonal symmetry.

Hence, we can consider the statistic:

$$W_{N,\nu}^* = \sum_i W_{n_i}^*, \tag{6.14}$$

so that when H_{oc} is true, for large values of n_i, $W^*_{N,\nu}$ is a chi-square with $\nu(p-1)$ degrees of freedom, but if n_i is small, then $W^*_{N,\nu}$ is only approximately a chi-square with $\nu(p-1)$ degrees of freedom. In the case of n_i being too small, which implies singularities in $V^*_{n_i}$ and difficulties in the chi-square approximation, $W^*_{n_i}$ can be modified applying it to the sample obtained from treatment groups, thus obtaining a nonparametric analogue of the parametric statistic F in the multivariate case. This statistic (W^*_N) is computed in a manner similar to that used to obtain (6.13) but applied to the sample of subjects obtained by the pooling treatment groups. The behaviour of this statistic is a chi-square with $(p-1)$ degrees of freedom when H_{oc} is true. When the distributions within the respective treatment groups may differ in shape as well as location, the statistic W^*_N is not a valid test for H_{oc}. In this situation Koch (1969) proposes the following statistic:

$$W = \widetilde{T}'_N C'_1 \left[C_1 \widetilde{V}_N C'_1 \right]^{-1} C_1 \widetilde{T}_N, \tag{6.15}$$

where:

$$\widetilde{T}^{(k)}_N = \sum_i \sum_j \widetilde{R}^{(k)}_{ij},$$

$$\widetilde{V}_{N;kk'} = \frac{1}{N^2} \sum_i \sum_j (\widetilde{R}^{(k)}_{ij} - \frac{p+1}{2})(\widetilde{R}^{(k')}_{ij} - \frac{p+1}{2}) \text{ and } V^*_N = (V^*_{N;kk'}),$$

$$\widetilde{R}^{(k)}_{ij} = \left[\text{Rank of } Y^{(k)}_{ij} \text{ in the set } \left\{ Y^{(p)}_{ij}, \ldots, Y^{(p)}_{ij} \right\} \right].$$

When H_{oc} is true, we can do an exact test based on (6.15) in the permutation model P_{Nc}, but if N is large the behaviour of (6.15) is approximately a chi-square distribution with $(p-1)$ degrees of freedom.

6.4 A new approach to analyze split-plot designs

A novel permutation solution for testing whole-plot and split-plot main effects and their interactions is here presented. We start saying that, in general, an exact permutation solution for testing whole-plot main effects and their interactions cannot be done, due to the fact that the exchangeability, under null hypothesis, of the error components doesn't hold. In this Section we suggest a new method to perform inference on split-plot experiments via combination-based permutation tests (Pesarin and Salmaso (2010)).

The method is composed essentially of two parts: the first is based on a new symmetry-based permutation approach (SYP) for testing the effects of whole-plot factors and the interactions between whole-plot and split-plot factors; the second is based on the synchronized permutation approach (SYN)- for more detail see Basso et al. (2007) and Salmaso (2003)- to test the effects of split-plot factors and their interactions.

The example that generated our present discussion of SP experiments occurred in an agricultural setting (see Naes et al. 2006). In this instance there are two different fertilizers

(factor B), and two systems of irrigation (factor A) are to be used on two different varieties of potato (factor C). The yield per square hectare is the response variable. There are two blocks large enough to test each variety of potato with all fertilizers and both systems of irrigation. It is possible to split each block into smaller sections but a full randomization of the combinations is not possible - the same irrigation can only be used for strip sections of the whole experimental area. The experimental design is a typical SP design where the positions of the two systems of irrigation are randomized within the blocks and the varieties of potato and the fertilizers are also randomized within the blocks. In this design the systems of irrigation are considered to be a whole-plot and the others a split-plot. The main-plot treatments follow a CRD or RCB design. Each of these plots is then split into sub-plots, one for each of the varieties. Each variety is then randomly assigned to a sub-plot within each main plot. The assignment is performed randomly within each main plot, and independently of other main plots, i.e. the main plots are treated as blocks as far as the varieties are concerned. Looking at it from two perspectives, the main-plot treatments by themselves follow a simple CRD design, while the sub-plot treatments mimic an RCB design. Combining the two together produces the split-plot design. Recognizing the two types of designs that are combined is the key to analyzing these designs. There are two different sizes of experimental units. The main-plot factor (A) is randomly assigned to large (main) plots. Consequently, variation at the main-plot level is what limits detection of effects. The sub-plot factor (B) is randomly assigned to smaller (split) plots within each main plot. Consequently, both the blocking by main-plot and the smaller split-plot experimental unit must be considered.

The following model is assumed to describe the data adequately:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \delta_l + E_{il} + \varepsilon_{ijkl}, \quad (6.16)$$

with i, j, k, l are set equal to 1 or 2. The fixed effects are: μ (overall mean), α_i (main effect to the i-th level of A - whole-plot effect), β_j (main effect to the j-th level of B - sub-plot effect), γ_k (main effect to the k-th level of C - sub-plot effect), $(\alpha\beta)_{ij}$ (interaction effect corresponding to the i-th level of A and j-th level of B), $(\alpha\gamma)_{ik}$ (interaction effect corresponding to the i-th level of A and k-th level of C), $(\beta\gamma)_{jk}$ (interaction effect corresponding to the j-th level of B and k-th level of C) and $(\alpha\beta\gamma)_{ijk}$ (interaction effect corresponding to the i-th level of A, j-th level of B and k-th level of C). The random effects are: δ_l (block effect), E_{il} (random effect corresponding to the i-th system of irrigation and l-th block - whole-plot error term) and ε_{ijkl}, (random effect corresponding to the i-th level of A, j-th level of B, k-th level of C and l-th block - residual error term). In this context we assume that E_{il} are IID with unknown continuous distribution function P - symmetrically distributed around 0 - with mean 0 and $Var(E_{il}) = \sigma^2_{whole}$ and also ε_{ijkl} are i.i.d with unknown continuous distribution function P - symmetrically distributed around 0 - with mean 0 and $Var(\varepsilon_{ijkl}) = \sigma^2_{sub}$. We observe that it is also possible to incorporate random interactions between sub-plots and replicates in the model, but these are often considered to be negligible and are therefore omitted here. There are also the usual side

conditions:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0, \tag{6.17}$$

$$\sum_i (\alpha\beta)_{ij} = 0 \text{ for any j } \sum_j (\alpha\beta)_{ij} = 0 \text{ for any i}, \tag{6.18}$$

$$\sum_i (\alpha\gamma)_{ik} = 0 \text{ for any k } \sum_k (\alpha\gamma)_{ik} = 0 \text{ for any i}, \tag{6.19}$$

$$\sum_j (\beta\gamma)_{jk} = 0 \text{ for any k } \sum_k (\beta\gamma)_{jk} = 0 \text{ for any j}, \tag{6.20}$$

$$\sum_i (\alpha\beta\gamma)_{ijk} = 0. \tag{6.21}$$

The following Table 6.1 is an illustration of the agricultural example with two blocks. The symbol '+' represents level 1 of the factors (A, B and C) while the symbol '-' represents level 2 of the factors involved in the experiment.

Table 6.1: Experimental design. The first variable is the whole-plot variable and the last two are the sub-plot variables.

Block	A	B	C	Block	A	B	C
I	+	+	+	II	−	−	−
I	+	+	−	II	−	−	+
I	+	−	+	II	−	+	−
I	+	−	−	II	−	+	+

The over-null hypothesis state the null-effects of all factors and it c may be written as:

$$\begin{aligned} H_0 : &\{(\alpha = 0) \cap (\beta = 0) \cap (\gamma = 0)\} \cap \\ &\cap \{(\alpha\beta = 0) \cap (\alpha\gamma = 0) \cap (\beta\gamma = 0) \cap (\alpha\beta\gamma) = 0\}. \end{aligned}$$

We want to separately test the sub-null hypotheses $\{H_{0A} : \alpha = 0\}$;
$\{H_{0B} : \beta = 0\}$; $\{H_{0AB} : (\alpha\beta) = 0\}$; $\{H_{0AC} : (\alpha\gamma)_{ik} = 0\}$ and
$\{H_{0ABC} : (\alpha\beta\gamma)_{ijk} = 0\}$. In order to simplify the notations we can rewrite the model (6.1) as follows:

$$Y_{ijkl} = \mu_{ijk} + \delta_l + \sigma_i \widetilde{\varepsilon}_{ijkl}, \tag{6.22}$$

where $\widetilde{\varepsilon}_{ijkl} = E_{il} + \varepsilon_{ijkl}$ are the random errors which are assumed to be *symmetrically* distributed around the origin with unknown continuous distribution P_i $(i = 1, 2)$, $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$ are unknown location parameters and $\sigma_i = \sigma_{wp} + \sigma_{sp}; i = 1, 2$ are unknown scale coefficients which are not assumed to be equal. Also we note by $\mathbf{Y}_1$ and $\mathbf{Y}_2$ the data vectors in the first and second block respectively

(for example $\mathbf{Y}_1 = (Y_{1111}, Y_{1211}, Y_{1121}, Y_{1221})$). We note that, in our present discussion, the levels of the whole-plot factor A coincide with the blocks.

In this context, it must be observed that in H_0 data are not exchangeable, so that the permutation testing principle (Pesarin (2001)) is to some extent not exactly applicable. Thus, in order to give a permutation solution, we have to look for approximate solutions.

Over-null hypothesis (referring to the model (6.22)) can be rewritten as:

$$H_0 = \left\{ \frac{\mathbf{Y}_1 - \mu}{\sigma_1} \stackrel{d}{=} \frac{\mathbf{Y}_2 - \mu}{\sigma_2};\ \sigma_1 \neq \sigma_2 \right\}, \tag{6.23}$$

where in addition we assume that the univariate responses Y_{ijkl} $(i, l = 1, 2)$ take values on the sample space χ, and are symmetrically distributed around the common location parameter μ. The unconditional distributions of $X_{ijkl} = (Y_{ijkl} - \widetilde{Y})$, where $\widetilde{Y}$ indicates the pooled sampling median as $\widetilde{Y} = (Y_{(n/2)} + Y_{(1+n/2)})/2$ if n (where $n = n_1 + n_2$ is the total number of the observations) is even and as $\widetilde{Y} = (Y_{(n+1)/2}$ if n is odd, where $Y_{(i)}$ indicates the ith order statistic; are symmetrically distributed around the origin if and only if H_0 is true. On the other hand, the conditional distributions of X_{ijkl} conditional on the pooled median $\widetilde{Y}$ are such that:

$$Pr\left\{X_{ijkl} < z \mid \widetilde{Y} = -t\right\} = Pr\left\{X_{ijkl} > z \mid \widetilde{Y} = t\right\},\ \forall z, t \in \mathbb{R}, \tag{6.24}$$

hence they are mutually symmetric, so they are only slightly asymmetric and their asymmetry vanishes as sample size increases.

We can observe that the distributions of X_{ijkl} in H_0 are invariant with respect to the common location parameter μ, but they depend on scale coefficients σ_i and on underlying distributions P_i.

The key idea is to work within a permutation framework by conditioning with respect to a set of jointly sufficient statistics in H_0.

Suppose now that f_{P_1} and f_{P_2} are the densities of two sampling distributions related to the two populations P_1 and P_2, relative to the same dominating measure Ψ and $\mathbf{Y}_1$, $\mathbf{Y}_2$ are the two separate data sets with sample size n_1 and n_2 respectively (in our discussion $n_1 = n_2$). So as the likelihood associated with the pooled data set is:

$$f_P^n(\mathbf{Y}) = f_{P_1}^{n1}(\mathbf{Y_1}) \cdot f_{P_2}^{n2}(\mathbf{Y_2}), \tag{6.25}$$

from the sufficiency principle it follows that the data set partitioned into two blocks, $(\mathbf{Y}_1, \mathbf{Y}_2)$ is now the set of sufficient statistic. In fact, by joint invariance of the likelihood ratio with respect to f_{P_1} and f_{P_2}, the permutation sample space of $\mathbf{Y}$ is $\Xi_{/\mathbf{Y}_1} \times \Xi_{/\mathbf{Y}_2}$ where $\Xi_{/\mathbf{Y}_1}$ and $\Xi_{/\mathbf{Y}_2}$ are the partial permutation sample space related to $\mathbf{Y}_1$ and $\mathbf{Y}_2$ respectively. Consequently, conditionally no datum from $\mathbf{Y}_1$ can be exchanged with any other from $\mathbf{Y}_2$, because permutations are permitted only within blocks, separately.

The idea is to find a pair of conditional permutation tests for separately testing symmetry conditionally on $(\mathbf{Y}_1, \mathbf{Y}_2)$ in accordance with the theory of permutation testing for symmetry (Pesarin (2001)) followed by a suitable combination.

6.4.1 The permutation approach in testing symmetry

In reference to the motivating example (model (6.1)) we can write the sub-hypothesis of no whole-plot main effects as follows:

$$H_{0A} : \{\alpha = 0\} \ \ vs \ H_{1A} : \{\alpha \neq 0\}\,, \tag{6.26}$$

or equivalently (model (6.22))

$$H_{0A} : \{\mu_{ijkl} = \mu + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}\}\,,$$

$$vs$$

$$H_{1A} : \{\mu_{ijkl} \neq \mu + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}\}\,.$$

Let us consider the data on the first block $\mathbf{Y}_1$, that is a sample of n_1 IID observations from a continuous variable Y_1 with unknown distribution P_1 on the real line.

The responses Y_{ijkl} $(l = 1, 2)$ take values on the sample space χ and are symmetrically distributed around the common location parameter μ, so the unconditional distributions of $X_{ijkl} = (Y_{ijkl} - \widetilde{Y})$ (where $\widetilde{Y}$ is the pooled sampling median) are symmetrically distributed around the origin if and only if H_{0A} is true.

We can consider the variables X_{ijkl} as the differences $X_{Aijkl} - X_{Bijkl}$ where A and B are two occasion of measurements with $j, k = 1, 2$ and $i = l = 1$. Pesarin (2001) shows that the distribution of differences X_{ijkl} is symmetric around the origin if $X_{Aijkl} \overset{d}{=} X_{Bijkl}$.

Let us consider $\mathbf{t} = (t_1, \ldots, t_{n1})$ and $\mathbf{t}' = (t'_1, \ldots, t'_{n1})$ two points of the sample space, they lie in the same orbit of a minimal sufficient statistic for any density f_{P1} symmetric around zero if and only if the likelihood ratio

$$\frac{f_{P1}(t_1) \cdot \ldots \cdot f_{P_1}(t'_{n_1})}{f_{P_1}(t'_1) \cdot \ldots \cdot f_{P_1}(t'_{n_1})} = \rho_{f_{P_1}}(\mathbf{t}, \mathbf{t}'), \tag{6.27}$$

does not depend on f_{P1}. Supposing the symmetry of f_{P1}, the ratio $\rho_{f_{P1}}(\mathbf{t}, \mathbf{t}')$ is f_{P_1} independent if $f_{P_1}(t_i) = f_{P_1}(t'_i)$ $(i = 1, \ldots, n_1)$ except for an irrelevant permutation of units, i.e. $t_i = \pm t'_i$. In this way the set of points $\chi_{/\mathbf{Y}_1} \in \mathbb{R}^{n_1}$ which contains the same amount of information with respect to P_1 as that contained in $\mathbf{Y}_1$ is made by all points obtained by giving *signs* in all possible way to the elements of $\mathbf{Y}_1$.

Considering a finite group $\mathbf{G}$ of transformations of signs we have that:

$$\mathbf{X}_1^\star = g^\star(\mathbf{X}_1 = \left\{X_{1111} \cdot S_1^\star, \ldots, X_{1n_1n_11} \cdot S_{n_1}^\star\right\};\ n_1 = 2, \tag{6.28}$$

where

$$\left\{S_1^\star, \ldots, S_{n_1}^\star\right\} \in [-1, 1]^{n_1}, \tag{6.29}$$

is a vector of signs. So a suitable test for H_{0A} (6.26) is given by $T = \sum_{jk} X_{1jk1}$, the permutation distribution of which is obtained by considering the permutation support

$\mathcal{T}(\mathbf{X}_1) = \left\{T(\mathbf{X}_1^*) = \sum_{jk} X_{1jk1} \cdot S_1^* : g \in \mathbf{G}\right\}$ where signs are assigned in all possible ways with equal probability. So within the permutation approach in order to test H_{0A} we are looking for two separate partial tests of symmetry, one from each block, followed by a suitable combination. These partial tests can be defined as:

$$^{I}T_A^* = \sum_{ijkl} X_{ijkl} \cdot S_{ijkl}^*, \tag{6.30}$$

where the notation ^{I} means "block 1" and the notation ^{II} means "block 2"and where $\mathbf{S}^* = \left\{S_{ijkl}^*; i = j = k = l = 1, 2\right\}$ is a random sample of i.i.d observations from variable S such that:

$$P(S = \pm 1) = 1/2. \tag{6.31}$$

So we have that:

$$^{I}T_A^* = \sum_{jk} X_{1jk1} \cdot S_{1jk1}^*, \tag{6.32}$$

$$^{II}T_A^* = \sum_{jk} X_{2jk2} \cdot S_{2jk2}^*. \tag{6.33}$$

After calculation and for conditions (6.17),(6.18),(6.19),(6.20),(6.21) we find:

$$^{I}T_A^* = 4 \cdot (\mu + \alpha + \sum_{jk} \widetilde{\varepsilon}_{1jk1}) \cdot (+1), \tag{6.34}$$

$$^{II}T_A^* = 4 \cdot (\mu - \alpha + \sum_{jk} \widetilde{\varepsilon}_{2jk2}) \cdot (-1). \tag{6.35}$$

In these formulations, $\widetilde{\varepsilon}_{ijkl}$ indicates the sum of all error components ($\widetilde{\varepsilon}_{ijkl} = E_{il} + \varepsilon_{ijkl}$). The two tests are independent and we can combine two tests into one (for more details see Pesarin (2001)).

A possible combining function is a linear form (because the tests are independent) given by:

$$\psi = \left|^{II}T_A^* +^{I} T_A^*\right|. \tag{6.36}$$

After calculations we find that:

$$T_A^* = 4(2\alpha + \sum_{jk} (\widetilde{\varepsilon}_{1jk1} - \widetilde{\varepsilon}_{2jk2})). \tag{6.37}$$

So we have found a statistic for testing the effects of whole-plot factor A that depends only on effect α and on linear combinations of errors. Due to conditioning on a sample quantity, the test statistic T_A^* is only approximately exact for testing symmetry. In a similar way, suitable test statistics for testing on H_{0AB}, H_{0AC}, H_{0ABC} can be obtained.

6.4.2 A synchronized-based permutation solution for testing the sub-plot main effects and interactions

As far as the split-plot structure is concerned, there is a complete randomization within the blocks, so the sub-plot treatment mimics an RCB. For testing H_{0B}, H_{0C} and H_{0BC} we suggest using a synchronized permutation approach, which leads to three exact solutions (Basso et al. (2007)). In particular, for each $H_{0.}$ we can find four intermediate statistics. For example, considering the tests for H_{0B}, we have two partial tests in the first block:

$$^{I}T_{B|1} = (Y_{1111} - Y_{1211}), \tag{6.38}$$

$$^{I}T_{B|2} = (Y_{1121} - Y_{1221}), \tag{6.39}$$

and another two partial tests in the second block:

$$^{II}T_{B|1} = (Y_{2112} - Y_{2212}), \tag{6.40}$$

$$^{II}T_{B|2} = (Y_{2122} - Y_{2222}). \tag{6.41}$$

In the first block v_1^* data are randomly selected from the strip $A_1B_1C_1$ and exchanged with v_1^* data randomly selected from the strip $A_1B_2C_1$ and also v_2^* data are randomly selected from the strip $A_1B_1C_2$ and exchanged with v_2^* data randomly selected from the strip $A_1B_2C_2$.

In the second block v_3^* data are randomly selected from the strip $A_2B_1C_1$ and exchanged with v_3^* data randomly selected from the strip $A_2B_2C_1$ and also v_4^* data are randomly selected from the strip $A_2B_1C_2$ and exchanged with v_4^* data randomly selected from the strip $A_2B_2C_2$.

The permutation structures of the four intermediate statistics are:

$$^{I}T^*_{B|1} = 2(n - 2v_1^*)(\beta + \beta\gamma) + n(\overline{\tilde{\varepsilon}}^*_{1111} - \overline{\tilde{\varepsilon}}^*_{1211}), \tag{6.42}$$

$$^{I}T^*_{B|2} = 2(n - 2v_2^*)(\beta - \beta\gamma) + n(\overline{\tilde{\varepsilon}}^*_{1121} - \overline{\tilde{\varepsilon}}^*_{1221}), \tag{6.43}$$

$$^{II}T^*_{B|1} = 2(n - 2v_3^*)(\beta + \beta\gamma) + n(\overline{\tilde{\varepsilon}}^*_{2112} - \overline{\tilde{\varepsilon}}^*_{2212}), \tag{6.44}$$

$$^{II}T^*_{B|2} = 2(n - 2v_4^*)(\beta + \beta\gamma) + n(\overline{\tilde{\varepsilon}}^*_{2122} - \overline{\tilde{\varepsilon}}^*_{2222}), \tag{6.45}$$

where $\overline{\tilde{\varepsilon}}^*_{ijkl}$ are the sampling means of permutation errors.

Synchronizing the permutations of the two statistics in the first block (that is $v_1^* = v_2^* = v^*$) and the two statistics in the second block ($v_3^* = v_4^* = v^*$) then:

$$^{I}T^*_{B} = 4(n - 2v^*)(\beta) + n(\sum_k \overline{\tilde{\varepsilon}}^*_{11k1} - \sum_k \overline{\tilde{\varepsilon}}^*_{12k1}, \tag{6.46}$$

$$^{II}T^*_{B} = 4(n - 2v^*)(\beta) + n(\sum_k \overline{\tilde{\varepsilon}}^*_{21k2} - \sum_k \overline{\tilde{\varepsilon}}^*_{22k2}). \tag{6.47}$$

These two tests depend only on the effect of B and on a linear combination of exchangeable (within blocks) errors, so that they give two exact tests for H_{0B} independent of the truth of the other sub-null hypotheses.

The next step is how to combine these two tests, i.e. find out a suitable combining function ψ:

$$\psi({}^{I}T_B^*, {}^{II}T_B^*). \tag{6.48}$$

For the independence of these tests, we suggest using Fisher's combining function (Pesarin (2001)).

The same strategy we used to obtain the solution for B can be applied to derive permutation test statistics for C and BC.

6.5 Simulation study

In this Section, in order to validate the suggested method, we report on a simulation study performed to validate symmetry permutation (SYP) testing on split-plot designs. The reference setting for the present simulation study concerns a split-plot design with three factors (model (6.16)), where both whole-plot factor A and split-plot factors B and C have two levels each. The emphasis is on the power of SYP testing procedures: a comparison with the parametric (F) test and the nonparametric counterparts is also presented.

The simulation program generates 1000 independent experiments for each of the following combination of cases:

- Two, four and six replications.
- The whole-plot error term (E_{il}) and the regular errors (ε_{ijkl}) are independently generated from a normal, exponential and 2 d.f. Student's t distributions.

The standard deviation for random split-plot error (σ^2_{sub}) was held at 0.5 and the additional whole-plot component (σ^2_{whole}) was 1. Note that the simulation mimics the situation described in the previous Section. The main effects and interaction effects are set as follows: $\alpha = 2$, $\beta = 1$, $\gamma = 0.5$, $\alpha\beta = 1.5$, $\alpha\gamma = 1.25$, $\beta\gamma = 0.75$ and $\alpha\beta\gamma = 0.25$.

Tables 6.2, 6.3, 6.4 and 6.5 report a summary of the rejection rates ($\alpha = 0.05$) and nominal levels for the hypotheses H_{0A}, H_{0AB}, H_{0AC}, H_{0ABC} obtained using the symmetry-based permutation testing procedure (SYP), ANOVA testing procedure for SP designs (F) and the nonparametric counterparts L (6.8) and W (6.15); while tables 6.6, 6.7 and 6.8 report the rejection rates and nominal levels for the hypotheses H_{0B}, H_{0C}, H_{0BC} obtained using the synchronized-based permutation testing procedure (SYN) and the parametric and nonparametric counterparts (F, L and W).

The first column reports the type of statistic, the second the number of replications and then the estimated power under the three types of errors. Nominal levels are also reported, in case we set all effects equal to zero.

From Figure 6.1 we note that for all whole-plot hypotheses the permutation SYP statistic appears to be more powerful than the others under heavy-tailed errors (Student's t), while in the normal case F is obviously better, but we observe that when the number of replicates is relatively small, the SYP statistic is a good alternative.

The rejection rates of the second order interaction tests between whole-plot factor A and split-plot factors B, C are represented in Figures 6.2 and 6.3; while the rejection rates of the third order interaction tests are reported in Figure 6.4.

Note that although Koch (1969) suggests applying the L statistic for whole-plot testing and W statistic for split-plot testing, we decided to consider both statistics when testing for all effects. The aim was to quantify the loss of power of each statistic for the inappropriate cases.

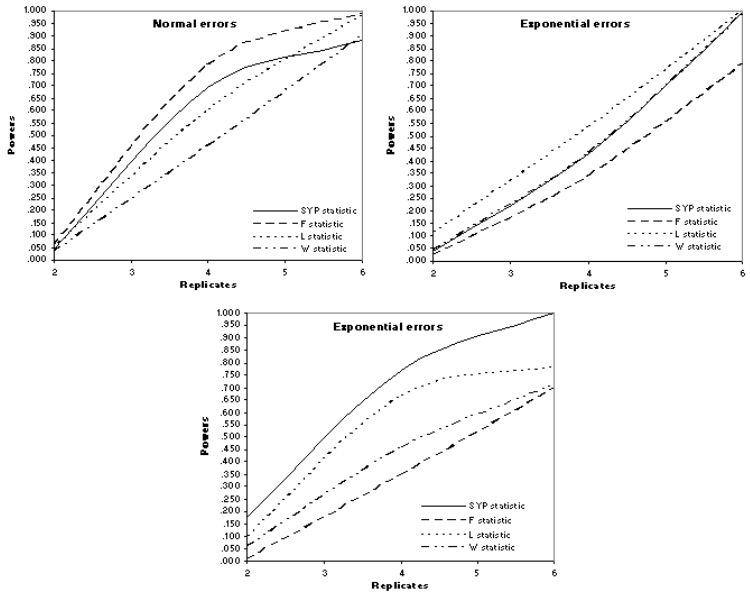

Figure 6.1: Test on Whole-Plot factor (A).

		H_1 **(Rejection Rate)**			H_0 **(Nominal Level)**		
		Error			**Error**		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.044	.176	.036	.016	.031	.036
SYP	4	.689	.768	.432	.033	.040	.056
	6	.879	1.000	.996	.045	.048	.044
	2	.056	.011	.023	.017	.021	.017
F	4	.780	.350	.340	.023	.035	.048
	6	.989	.700	.789	.031	.056	.040
	2	.050	.099	.112	.009	.011	.005
L	4	.600	.667	.536	.014	.012	.015
	6	.980	.785	.999	.028	.035	.047
	2	.032	.056	.045	.013	.018	.021
W	4	.458	.457	.430	.022	.025	.039
	6	.900	.711	.988	.041	.033	.050

Table 6.2: Power simulation study in a SP design: test on whole-plot factor (A).

		H_1 **(Rejection Rate)**			H_0 **(Nominal Level)**		
		Error			**Error**		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.067	.107	.065	.016	.030	.023
SYP	4	.546	.678	.439	.029	.045	.059
	6	.992	.993	.821	.032	.049	.068
	2	.034	.014	.030	.019	.003	.023
F	4	.590	.189	.214	.021	.014	.038
	6	.998	.563	.667	.024	.057	.044
	2	.007	.067	.056	.002	.010	.013
L	4	.430	.389	.578	.023	.027	.031
	6	.789	.891	.996	.040	.060	.046
	2	.056	.073	052	.001	.004	.003
W	4	.354	.334	.512	.014	.011	.016
	6	.698	.765	.931	.035	.049	.043

Table 6.3: Power simulation study in a SP Design: test on interaction between whole-plot and split-plot factor (AB).

		H_1 **(Rejection Rate)**			H_0 **(Nominal Level)**		
		Error			**Error**		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.040	.105	.087	.020	.007	.028
SYP	4	.356	.543	.349	.034	.012	.056
	6	.964	.999	.775	.047	.046	.057
	2	.045	.011	.024	.000	.002	.036
F	4	.445	.198	.289	.015	.019	.044
	6	.999	.670	.558	.045	.037	.048
	2	.003	.090	.045	.006	.013	.012
L	4	.400	.443	.568	.010	.025	.034
	6	.947	.887	1.000	.034	.041	.047
	2	.045	.067	.069	.015	.009	.016
W	4	.325	.324	.467	.034	.035	.038
	6	.900	.879	.976	.048	.046	.050

Table 6.4: Power simulation study in a SP Design: test on interaction between whole-plot and split-plot factor(AC).

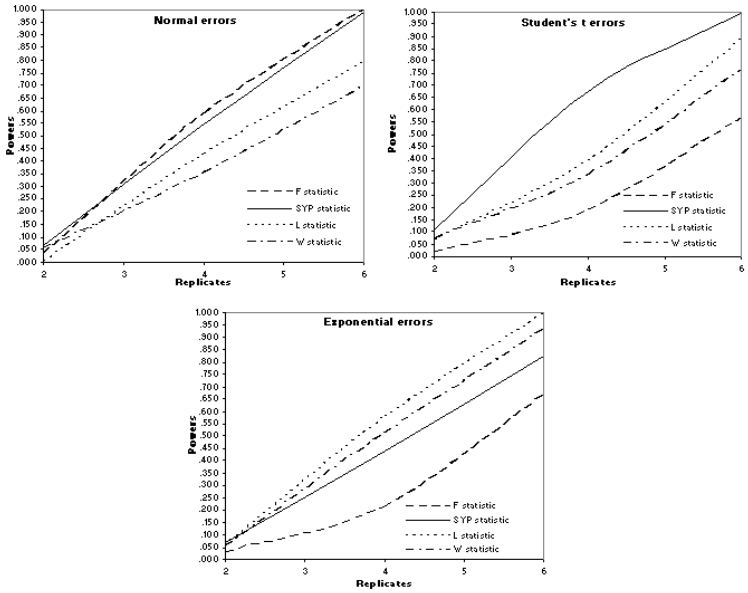

Figure 6.2: Test on the second order interactions (AB).

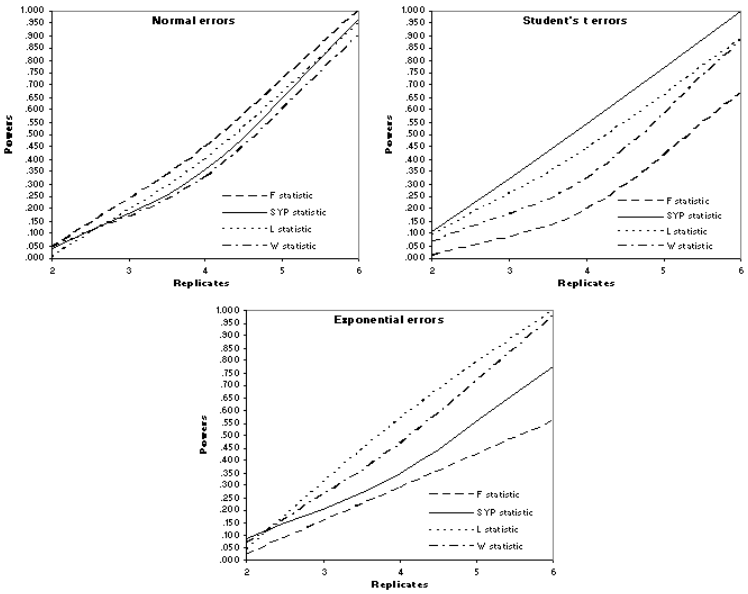

Figure 6.3: Test on the second order interactions (AC).

		H_1 **(Rejection Rate)**			H_0 **(Nominal Level)**		
		Error			**Error**		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.012	.078	.065	.001	.012	.048
SYP	4	.396	.459	.349	.028	.038	.036
	6	.980	1.000	.991	.035	.050	.056
	2	.026	.003	.030	.012	.035	.024
F	4	.467	.078	.189	.016	.045	.012
	6	.993	.678	.780	.044	.051	.036
	2	.014	.040	.067	.003	.004	.012
L	4	.354	.400	.561	.013	.012	.021
	6	.876	.893	1.000	.035	.044	.034
	2	.078	.070	.056	.007	.013	.017
W	4	.456	.432	.456	.022	.018	.023
	6	.901	.777	.1.000	.033	.042	.043

Table 6.5: Power simulation study in a SP Design: test on interaction between whole-plot and split-plot factor (ABC).

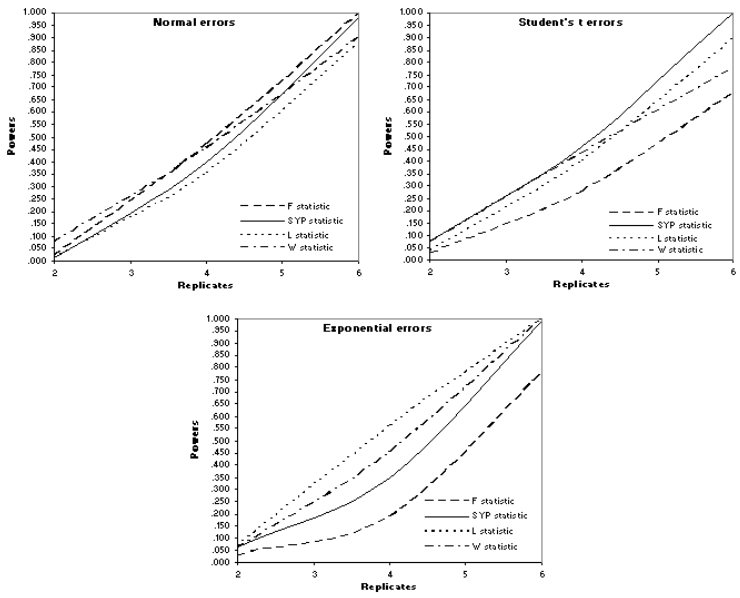

Figure 6.4: Test on the third order interactions (ABC).

		H_1 (Rejection Rate)			H_0 (Nominal Level)		
		Error			Error		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.087	.099	.089	.021	.019	.032
SYN	4	.660	.812	.779	.039	.024	.036
	6	.887	.999	.998	.020	.039	.066
	2	.099	.013	.041	.020	.012	.015
F	4	.706	.432	.667	.029	.035	.032
	6	.997	.786	.887	.035	.052	.045
	2	.079	.065	.077	.003	.009	.006
L	4	.349	.697	.303	.012	.014	.016
	6	.886	.806	1.000	.035	.046	.044
	2	.075	.077	.098	.007	.004	.006
W	4	.589	.703	.811	.019	.029	.037
	6	.901	.891	.997	.050	.048	.051

Table 6.6: Power simulation study in a SP Design: test on split-plot factor (B).

		H_1 (Rejection Rate)			H_0 (Nominal Level)		
		Error			Error		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.078	.089	.076	.024	.002	.011
SYN	4	.651	.711	.536	.045	.014	.056
	6	.941	1.000	.892	.012	.033	.028
	2	.070	.045	.056	.013	.011	.023
F	4	.713	.345	.498	.024	.029	.032
	6	.958	.568	.879	.024	.041	.032
	2	.035	.055	.091	.009	.014	.011
L	4	.443	.457	.600	.018	.023	.037
	6	.779	.778	.905	.039	.048	.040
	2	.066	.067	.085	.001	.017	.021
W	4	.567	.598	.679	.030	.031	.035
	6	.800	.886	.993	.040	.047	.054

Table 6.7: Power simulation study in a SP Design: test on split-plot factor (C).

		H_1 (Rejection Rate)			H_0 (Nominal Level)		
		Error			Error		
Statistic	**r**	**Normal**	**t-Student**	**Exp.**	**Normal**	**t-Student**	**Exp.**
	2	.098	.156	.076	.008	.036	.036
SYN	4	.604	.698	.345	.023	.041	.032
	6	.988	.998	.889	.031	.045	.036
	2	.101	.056	.040	.006	.001	.016
F	4	.712	.154	.256	.034	.013	.016
	6	.712	.154	.256	.034	.013	.016
	2	.086	.014	.057	.003	.001	.017
L	4	.435	.324	.456	.017	.019	.029
	6	.866	.895	.999	.033	.044	.050
	2	.034	.039	.065	.000	.001	.009
W	4	.456	.561	.555	.015	.021	.013
	6	.789	.876	.998	.029	.044	.047

Table 6.8: Power simulation study in a SP Design: test on interactions between split-plot factors (BC).

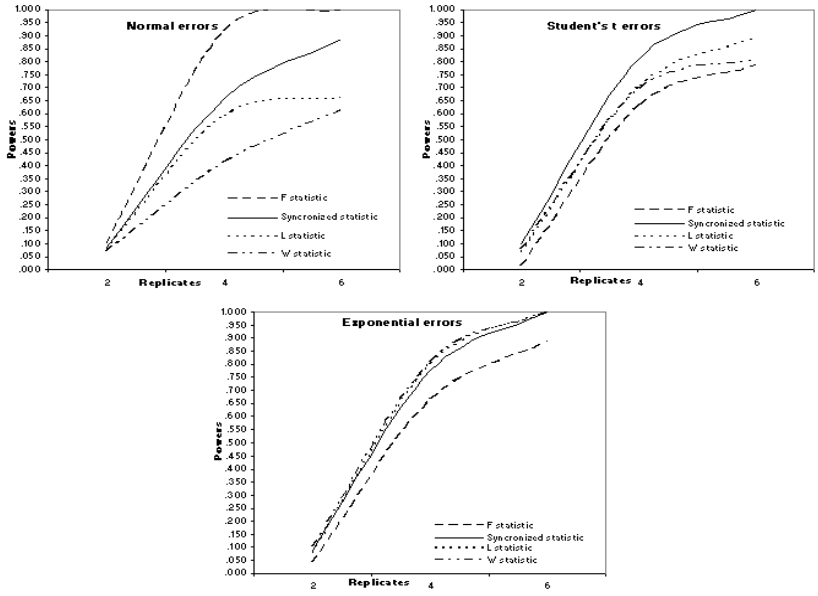

Figure 6.5: Test on sub-plot main effects (factor B).

We note also that, as expected, nonparametric L gives a good performance under asymmetric errors (such as exponential, see Figures 6.1, 6.2, 6.3 and 6.4). This result confirms that the SYP statistic is somewhat negatively affected in the case of non symmetric error distributions. We recall that we developed the SYP statistic under the assumption of symmetry for the random error distribution. Under exponential errors L represents the best alternative for both main effect and interactions; in any case, under Student's t distribution this nonparametric procedure is not more powerful than SYP, which is the best one under heavy-tailed distribution.

Even if the focus of the present work is concerned with testing on the whole-plot main effect and its interaction, it is interesting to consider simulation results on split-plot main effects and related interactions as well.

To this end we recall that we applied the synchronized permutation approach. As highlighted by Figures 6.5, it is clear that under non normal errors the SYN statistic performs much better than the parametric F test and is also better than the rank-based nonparametric counterparts, especially in the case of heavy-tailed errors. Finally, note that as expected the W statistic appears more powerful than the L statistic which is not appropriate here for split-plot effect testing.

6.6 An example

In this Section we face a real case study related to a study on wheat varieties grown in different fertility regimes. Actually, it is a slightly modified example exposed in Milliken

and Johnson (2009) p. 426. Let us consider the yields in pounds of two varieties of wheat ($B1$ and $B2$) grown in two different fertility regimes ($C1$ and $C2$). The field (whole-plot factor) was divided into two blocks ($A1$ and $A2$), each with two whole plots. Each of the two fertilizer levels was randomly assigned to one whole plot within each block and two replications were performed. The dataset, along with the data structure, is reported in Table 6.9.

	Block: $A1$		Block: $A2$	
Fertility	Variety		Variety	
regime	$B1$	$B2$	$B1$	$B2$
$C1$	35.4	34.8	41.6	40.3
	36.7	39.5	42.7	41.6
$C2$	37.9	36.4	43.6	42.8
	38.2	40.0	44.5	47.6

Table 6.9: Data for the variety by fertility regime split-plot example.

When applying all the considered parametric and nonparametric split-plot procedures to the wheat variety data, we can obtain p-values reported in Table 6.10, where we focus our attention on testing the whole-plot effect and the related interactions. Table 6.11 report p-values obtained applying parametric and nonparametric procedures adopted for testing main and interaction split-plot effects.

	Factor under testing			
Statistic	A	AB	AC	ABC
SYP (absolute sum)	0.0813	0.0750	0.0875	0.0875
SYP (simple sum)	0.0563	0.0437	0.0500	0.0437
F	0.0361	0.0722	0.0531	0.0528
L	0.0106	0.0217	0.0183	0.1025

Table 6.10: P-values of all the split-plot procedures for wheat variety data.

	Factor under testing	
Statistic	B	BC
SYN (absolute sum)	0.0530	0.0525
SYN (simple sum)	0.0380	0.0170
F	0.0460	0.0235
W	0.0350	0.0120

Table 6.11: P-values of all the split-plot procedures for wheat variety data.

Note that we applied the proposed permutation SYP statistic using not only the absolute value of the sum of the two within block partial statistics (as stated in equation

(29), p. 12) but also using as synthesis function the simple sum of partial statistics.

It is interesting to observe that not all procedures agree to reject the null hypotheses (Table 6.10) ($\alpha = 0.05$). Rank-based L statistic appears to be the more powerful, while the parametric F-test seems to be the more conservative. The SYP statistic behaves a little worse than the L statistic and a possible explanation of this point can be found in the relative small number of different p-levels which are attainable in this case (namely, 160) so that the p-values are actually multiple of $1/160 = 0.00625$. Testing split-plot main and interaction effects we observe that all the statistics (except SYN statistic with absolute value of the sum of the two within block partial statistics) agree to reject the null hypotheses (Table 6.11).

6.7 Final Remarks

In the framework of the combination-based permutation methodology we presented a novel solution for the whole set of hypotheses testing in split-plot experiments. The main idea underlying the proposed nonparametric method is concerned with the joint application of several synchronized and combination-based permutation tests.

Here we focus on situations with replicates, but our solution seems to be more general and it has also been shown to have good properties.

As suggested by the simulation study, we can confirm that the proposed solution is a good alternative to the traditional parametric F test.

Hence, in each experimental situation where normality is hard to justify, this nonparametric procedure can be considered a valid solution. Furthermore, in the case of errors with heavy-tailed distribution, such as Student's t, permutation tests can provide higher power than rank-based nonparametric counterparts. As far as further research development is concerned, the proposed approach seems promising for multivariate extensions. In fact, via nonparametric combination methodology (Pesarin (2001)), the dependence among univariate response variables can be handled in an effective manner.

References

1. Anderson T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, New York: Wiley and Sons.

2. Arboretti R., Corain L., Ragazzi S. (2010). A comparison among Combination-Based Permutation Statistics for Randomized Complete Block Design. To appear *Journal of Statistical Planning and Inference*

3. Arboretti R., Corain L., Ragazzi S. (2011). The multivariate randomized complete block design: a novel permutation solution in case of ordered categorical variables. To appear *Communication in Statistics - Theory and Methods*.

4. Arboretti R., Fontana R., Ragazzi S. (2011). Construction and nonparametric testing of orthogonal arrays through algebraic strata and inequivalent permutation matrices. To appear *Communication in Statistics - Theory and Methods*.

5. Basso D., Chiarandini M., Salmaso L. (2007). Synchronized permutation tests in replicated $I \times J$ designs, *Journal of Statistical Planning and Inference*, 137, 2564-2578.

6. Basso D., Evangelaras H., Koukouvinos C., Salmaso L. (2004). Nonparametric Testing for Main Effects on Inequivalent Designs. Proceedings of the 7th International Workshop on *Model-Oriented Design and Analysis* held in Heeze, Netherlands, June 14-18, 2004.

7. Bayley R.A. (1983). The decomposition of treatment degrees of freedom in quantitative factorial experiments, *J. Roy. Statist. Soc.* B 44(1), 6370.

8. Bierbrauer J. (1995). Bounds on orthogonal arrays and resilient functions, *Journal of Combinatorial Designs*, 3(3), 179- 183.

9. Brouwer A.E., Cohen A.M., Nguyen M.V.M (2006). Orthogonal arrays of strength 3 and small run sizes, *Journal of Statistical Planning and Inference*, 136, 3268-3280.

10. Cheng C.S. (1995). Some projection properties of orthogonal arrays, *Ann. Stat.*, 23, 1223-1233.

11. Collombier D. (1996). *Plans d'expérience factoriels: construction et propriété des fractions de plans*, Springer.

12. Commenges D. (2003). Transformations which preserve exchangeability and application to permutation tests, *Journal of Nonparametric Statistics*, 15(2), 171-185.

13. Corain L., Ragazzi S., Salmaso L. (2010). A Permutation Approach to Split-Plot Experiments. To appear *Journal of Statistical Planning and Inference*.

14. Dey A., Mukerjee R. (1999). *Fractional Factorial Plans. Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons Inc., New York.

15. Diestelkamp Wiebke S. (2004). Parameter Inequalities for Orthogonal Arrays with Mixed Levels, *Designs, Codes and Cryptography*, 33(3), 187-197.

16. Edgington E.S. (1995). *Randomization tests*, 3rd ed., Marcel Dekker, New York.

17. Evangelaras H., Koukouvinos C., Lappas E.(a)(2007). Further contributions to non-isomorphic two level orthogonal arrays, *Journal of Statistical Planning and Inference*, 137, 20802086.

18. Evangelaras H., Koukouvinos C., Lappas E.(b)(2007). 18-run nonisomorphic three level orthogonal arrays, *Metrika*, (66):31-37.

19. Evangelaras H., Koukouvinos C., Lappas E. (2006). An efficient algorithm for the identification of isomorphic orthogonal arrays, *Journal of Discrete Mathematical Sciences & Cryptography*, 9(1), 125132.

20. Fawcett R.F., Salter K.C. (1984). A monte carlo study of the f test and three tests based on ranks of treatment effects in randomized block designs, *Communications in Statistics - Simulation and Computation*, 13, 1532-4141, Volume 13, Issue 2, 1984, Pages 213 225.

21. Finos L., Pesarin F., Salmaso L., Solari A. (2008). Exact Inference for Multivariate Ordered Alternatives. *Statistical Methods and Applications* 17, 195-208.

22. Finos L., Salmaso L. (2004). Nonparametric multi-focus analysis for categorical variables, *Communications in Statistics - Theory and Methods*, 33, 1931-1941.

23. Finos L., Salmaso L., Solari A. (2007). Conditional inference under simultaneous stochastic ordering constraints. *Journal of Statistical Planning and Inference* 137, 2633-2641.

24. Fisher R.A. (1935). The fiducial argument in statistical inference, *Annals of Eugenics*, 46, 391-398.

25. Fontana R., Pistone G. (2010). Algebraic generation of orthogonal fractional factorial designs. Proceedings of the Statistical Conference SIS in Padua, Italy, June 29-July 1, 2010.

26. Fontana R., Pistone G., Rogantin M. P. (1997). Algebraic generation of two-level orthogonal designs, *Statistica Applicata*, 9, 15-29.

27. Fontana R., Pistone G., Rogantin M. P. (2000). Classification of two-level factorial fractions, *Journal of Statistical Planning and Inference*, 87, 149-172.

28. Friedman M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *J. Am. Stat. Assoc.*, 32, 675-701.

29. Groggel D.J. (1987). A Monte Carlo study of rank tests for block designs, *Communications in Statistics - Simulation and Computation*, 16 (3), 601-620.

30. Hamada M., Balakrishnan N. (1998). Analyzing unreplicated factorial experiments: a review with some new proposals, *Statist. Sinica*, 8(1), 1-41.

31. Hedayat A.S., Sloane N.J.A., Stufken J. (1999). *Orthogonal Arrays Theory and Applications*, Springer.

32. Hollander M., Wolfe D.A. (1999). *Nonparametric Statistical Methods, 2nd Edition*, Wiley Series in Probability and Statistics.

33. Kobilinsky, A., (1990). Complex linear model and cyclic designs, *Linear Algebra Appl.*, 127, 227282.

34. Kobilinsky, A., Monod, H., (1991). Experimental design generated by group morphism: an introduction. *Scand. J. Statist.* 18, 119134.

35. Koch G.G. (1969). Some Aspects of the Statistical Analysis of "Split-Plot" Experiments in Completely Randomized Layouts, *Journal of American Statistical Association*, 64, No 326, 485-505.

36. Kowalsky SM., Cornell JA., Vining GG., (2002). Split-plot designs and estimation methods for mixture experiments with process variables, *Technometrics*, 44(1),72–79.

37. Kreuzer M., Robbiano L. (2000). *Computational Commutative Algebra*, vol. I, Springer, Berlin, Heidelberg.

38. Kruskal W.H. and Wallis W.A. (1963), Use of ranks in one criterion variance analysis, *Journal of the American Statistical Association*, 46, 583-621.

39. Lamond E. (1970). *Methods for sensory evaluation of food*, Department of Agriculture, Otawa, Canada.

40. Larocque D., Bussières I. (2003) *Aligned rank test for the bivariate randomized block model.* In: M. Moore, S. Froda, and C. Léger, eds., Mathematical statistics and applications: Festschrift for Constance van Eeden, Beachwood, OH: Institute of Mathematical Statistics, 85-97.

41. Lehmann E., D'Abrera H. (2006). *Nonparametrics: Statistical Methods Based on Ranks*, Springer, Berlin.

42. Lenth R.V. (1989). Quick and easy analysis of unreplicated factorials, *Technometrics*, 31(4):469473.

43. Letsinger J.D., Meyers R.H., Lentner M. (1996). Response surface methods for bi-randomization structures, *Journal of Quality Technology*, 28(1), 381-397.

44. Loh W.Y. (1992). Identification of active contrasts in unreplicated factorial experiments, *Computational Statistics & Data Analysis*, (14):135-148.

45. Loughin T.M. and Noble, W. (1997). A permutation test for effects in an unreplicated factorial design, *Technometrics*, 39(2):180190.

46. Mahfoud Z.R., Randles R.H. (2005) *Practical tests for randomized complete block designs* . Journal of Multivariate Analysis, 96, 73-92.

47. Meilgaard M., Civille G.V., Carr B.T. (2006), *Sensory Evaluation Techniques, 4th ed. CRC Press*, Inc. Boca Raton, Florida.

48. Möttönen J., Hüsler J., Oja H. (2003). Multivariate nonparametric tests in a randomized complete block design, *Journal of Multivariate Analysis*, 85 (1), 106-129.

49. Möttönen J., Oja H.(1995). Multivariate spatial sign and rank methods, *J. Nonparam. Statist.*, 5, 201-213.

50. Möttönen J., Oja H., Tienari J. (1997). On the efficiency of multivariate spatial sign and rank tests, *Ann. Statist.*, 25, 542-552.

51. Montgomery D.C. (2001). *Design and Analysis of Experiments (5th edn)*, Wiley: New York.

52. Montgomery D.C. (2004). *Design and Analysis of Experiments*, 5th ed., Wiley, New York.

53. Mukerjee R. and C.F.J. Wu (1995): On the existence of saturated and nearly saturated asymmetrical orthogonal arrays. Ann. Statist. (to appear).

54. Naes T., Aastveit A.H., Sahni N.S. (2006). Analysis of Split-plot Designs: An Overview and Comparison of Methods, *Qual. Reliab. Engng. Int.*, 23, 801-820.

55. O'Gorman T.W. (2001). A Comparison of the F-Test, Friedman's Test, and Several Aligned Rank Tests for the Analysis of Randomized Complete Blocks, *Journal of Agricultural, Biological, and Environmental Statistics*, 6, 3, 367-378.

56. Oja H. (1999). Affine Invariant Multivariate Sign and Rank Tests and Corresponding Estimates: a Review, *Scandinavian Journal of Statistics*, 26, 319-343.

57. Pesarin F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*, Wiley, Chichester.

58. Pesarin F., Salmaso L. (2010). *Permutation tests for complex data: theory, applications and software*, Wiley, Chichester.

59. Peters D., Randles R.H. (1990). A multivariate signed-rank test for the one-sampled location problem, *J. Amer. Statist. Assoc.*, 85, 552-557.

60. Pfanzagl J. (1974). On the Behrens-Fisher problem, *Biometrika*, 61, 39 - 47.

61. Pistone, G., Rogantin, M.P. (2008)(a): Indicator Function and Complex Coding for Mixed Fractional Factorial Designs, *Journal of Statistical Planning and Inference*, 138, 3, 787-802.

62. Pistone G., Rogantin M.P. (2008)(b): Algebraic statistics of level codings for fractional factorial designs, *Journal of Statistical Planning and Inference*, 138, 234-244.

63. Pistone G., Wynn H.P. (1996). Generalised confounding with Gröbner bases, *Biometrika*, 83(3), 653666.

64. Quade D. (1979). Using weighted rankings in the analysis of complete blocks with additive block effects, *J. Amer. Statist. Assoc.* 74, 680-683.

65. Randles R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test, *J. Amer. Statist. Assoc.*, 95, 1263-1268.

66. Rao C. R. (1947). Factorial Experiments Derivable from Combinatorial Arrangements of Arrays, *Journal of the Royal Statistical Society*, Suppl. 9, 128-139.

67. Rao C. R: (1973). Some Combinatorial Problems of Arrays and Applications to Design of Experiments, in *A Survey of Combinatorial Theory*, eds. J. N. Srivastava et al., Amsterdam: North-Holland, pp. 349-359.

68. Robbiano L., Rogantin M.P. (1998): *Full Factorial Designs and Distracted Fractions. Gröebner Bases and Applications*, Proc. of the Conf. 33 Years of Gröebner Bases, London Mathematical Society Lecture Notes Series, 251, 473-482.

69. Sahni N.S., Isaksson T., Ns T. (2001). The use of experimental design methodology and multivariate analysis to determine critical control points in a process, *Chemometrics and Intelligent Laboratory Systems*, 56, 105-121.

70. Salmaso L. (2003). Synchronized permutation tests in factorial designs, *Communication in Statistics, Theory and Methods*, 32, 7, 1419-1437.

71. Schoen E.D., Eendebak P.T., Nguyen M.V.M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays, *Journal of Combinatorial Designs*, 18(2), 123140.

72. Schrijver, A. (1986). *Theory of linear and integer programming*, Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons Ltd., Chichester. A Wiley-Interscience Publication.

73. Sen P.k.(1968). On a class of aligned rank order tests in two-way layouts, *Annals of Mathematical Statistics*, 39.

74. Sen P.K, Puri M.L. (1967). On the theory of rank order tests for location in the multivariate one sample problem, *Ann. Math. Statist*, 38, 1216-28.

75. Sepansky J.H. (2007). A Modification on the Friedman Test Statistic, *Communications in Statistics-Simulation and Computation*, 36, 783–790.

76. Sloane N.J.A., Stufken J. (1996). A linear programming bound for orthogonal arrays with mixed levels, *Journal of Statistical Planning and inference*, 56, 295-305.

77. Stufken J., Tang B. (2007). Complete Enumeration of Two-Level Orthogonal Arrays of Strength d with $d + 2$ Constraints, *The Annals of Statistics*, 35(2), 793-814.

78. Tang, B., Deng, L. Y., (1999). Minimum G2-aberration for nonregular fractinal factorial designs, *The Annals of Statistics*, 27 (6), 1914-1926.

79. Tyler D.E. (1987). A distribution-free m-estimator of multivariate scatter, *Ann. Statist.*, 15, 234-251.

80. Van de Ven P.M, Di Bucchianico A. (2009). On the equivalence of definitions for regular fractions of mixed-level factorial designs, *Journal of Statistical Planning and Inference*, 139, 2351-2361.

81. Wang J.C. (1996). Mixed difference matrices and the construction of orthogonal arrays, *Statistics & Probability Letters*, 28, 121-126.

82. Wang J.C., Wu C.F.J. (1991). An Approach to the Construction of Asymmetrical Orthogonal Arrays, *Journal of the American Statistical Association*, 86, 450-456.

83. Westfall P.H., Tobias R.D., Rom D., Wolfinger R.D., Hochberg Y. (2000). *Multiple Comparisons and Multiple Tests using the SAS*, (SAS Books by Users).

84. Wilcoxon F. (1949). *Some rapid Approximate Statistical Procedures*, American Cynamid Co., Stamford, Connecticut.

85. Wu C.F.J., Hamada M. (2000). *Experiments: Planning, Analysis and Parameter Design Optimization*, Wiley, New York.

86. Wu C.F.J., Zhang R., and Wang R. (1992). Construction of Asymmetrical Orthogonal Arrays of the Type $OA(s^k, s^m(s(s^{r_1})^{n_1}, \ldots, s(s^{r_t})^{n_t})$, *Statistica Sinica*, 2, 203-219.

87. Xu H., Wu C.F.J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs, *Ann. Statist.*, 29(4), 1066-1077.

88. Ye K.Q. (2003). Indicator function and its application in two-level factorial designs, *The Annals of Statistics*, 31(3), 984-994.

Susanna Ragazzi

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4111
e-mail: susanna.ragazzi@unipd.it

Current Position

Since January 2007; (expected completion: June 2011)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Innovative approaches for the construction of orthogonal arrays and permutation tests in complex experimental designs.
Supervisor: Prof. Luigi Salmaso.
Co-supervisors: Prof. Fortunato Pesarin, Dott. Roberto Fontana.

Research interests

- Design and analysis of experiments.
- Algebraic construction of fractional factorial designs.
- Nonparametric procedures for testing effects in unreplicated multifactorial designs, Split-plot designs and Randomized Complete Block designs.

Education

March 2006
Master degree in Mathematics.
University of Ferrara, Faculty of Mathematics, Physics and Natural Sciences.
Title of dissertation: "Parametric and Nonparametric statistical methods with applications to a biomedical study".
Supervisor: Prof. Camillo Fucci.

Work experience and Collaborations

March 2011
Statistical analyst junior.
Ufficio studi Lapam S.p.A., Bologna, Italy.

March 2010 – March 2011
Temporary research fellow.
Title research project: *Nonparametric methods in the performance's analysis of products and services.*
Department of Territory and Agri-Forestal Systems, University of Padova, Legnaro (PD), Italy.

January 2009 – December 2009
Former – Course on "Statical analysis with Excel".
Centro di formazione professionale IRIDE S.p.A., Modena, Italy.

June 2008 – December 2008
Former – Course on "Statical analysis with Matlab ".
Centro di formazione professionale IRIDE S.p.A., Modena, Italy.

March 2006 – December 2006
Collaboration research project – Prof. L. Pareschi.
Center for Modeling, Computation and Statistics (CMCS), Department of Mathematics, University of Ferrara, Italy.

Publications

Articles in journals

Arboretti R., Corain L., Ragazzi S. (2011). The multivariate randomized complete block design: a novel permutation solution in case of ordered categorical response variables. Accepted for the publication on *Communication in Statistics - Theory and Methods.*

Arboretti R., Fontana R., Ragazzi S. (2011). Construction and nonparametric testing of orthogonal arrays through algebraic strata and inequivalent permutation matrices. Accepted for the publication on *Communication in Statistics - Theory and Methods.*

Arboretti R., Corain L., Ragazzi S. (2010). A comparison among Combination-Based Permutation Statistics for Randomized Complete Block Design. Accepted for the publication on *Journal of Statistical Planning and Inference.*

Corain L., Ragazzi S., Salmaso L. (2010). A Permutation Approach to Split-Plot Experiments. Accepted for the publication on *Journal of Statistical Planning and Inference.*

Chapters in books

Arboretti R., Finos L., Guarda Nardini L., Manfredini D., Ragazzi S. (2007). Nonparametric methods for a clinical trial with multiple endpoints and small sample size. In *Abstracts of SISMEC - IV International Conference in Statistics* (Palermo, Italia, 19-22 Settembre 2007).

Corain L., Ragazzi S., Salmaso L. (2009). Permutation Tests for the Multivariate Randomized Complete Block Design, *Proceedings of the 6th St. Petersburg Workshop on Simulation* (Edited by S. M. Ermakov, V.B. Melas and A.N. Pepelyshev), St. Petersburg, 2009, June 28 - July 4, Volume I 465-472,(ISBN: 978-5-9651-0354-6).

Corain L., Ragazzi S., Salmaso L. (2009). Extending the Synchronized Permutation Approach to Split-Plot Designs, *Book of Abstract of the Conference Multivariate Methods and Models for Evaluating Public Services*, Rimini, May 25-26 2009.

Working papers

Arboretti R., Corain L., Ragazzi S., Salmaso L. (2009). Descrizione ed analisi degli aspetti economico-produttivi del territorio della provincia di Rovigo in funzione del suo nuovo sistema infrastrutturale. *Technical Report 01/2009*, University of Padova.

Arboretti R., Ragazzi S., Salmaso L. (2006). Efficacy of Botulim toxin in the treatment of Myofascial Face Pain in Clinically diagnosed mbruxer: pilot data from a placebo controlled, randomized clinical trial. *Technical Report 05/2006*, Center for Modeling, Computation and Statistics, University of Ferrara.

Ragazzi S. (2005). Elementi di teoria degli stimatori. *Technical Report 05/2005*, Center for Modeling, Computation and Statistics, University of Ferrara.

Conference presentations

---

28 June - July 4, 2009
Talk titled "Permutation Tests for the Multivariate Randomized Complete Block Design". *6th St. Petersburg Workshop on Simulation*, Saint Petersburg, Republics of Russia.

May 25-26, 2009
Talk titled "Extending the Synchronized Permutation Approach to Split-Plot Designs". *Multivariate Methods and Models for Evaluating Public Services*, Rimini, Italy.

Teaching experience

March 2011
Statistic.
Exercises, total number of hours 60.
Department of Economics, University of Ferrara.
Instructor: Dott. Stefano Bonnini.