An all configurations approach for detecting hidden-additivity in two-way unreplicated experiments.

Christopher Franck

LISA (Laboratory for Interdiscipinary Statistical Analysis)

Department of Statistics

Virginia Tech, Blacksburg, VA 24061

email: chfranck@vt.edu

Jason Osborne

Department of Statistics

North Carolina State University, Raleigh, NC 27695

email: jason.osborne@ncsu.edu

Dahlia Nielsen

Department of Genetics

North Carolina State University, Raleigh, NC 27695

email: nielsen@stat.ncsu.edu

Summary

Assessment of interaction in unreplicated two-factor experiments is a challenging problem that has received considerable attention in the literature. Because fitting the full factorial model with main effects and interaction effects leaves no information to estimate variability, the assumption of additivity is frequently made excluding the interaction terms from the model. Many authors have developed models and tests with restrictions applied to the interaction terms in order to address non-additivity in unreplicated experiments. We propose a model in which the levels of one factor belong in two or more groups. Within each group the effects of the two factors are additive but the groups may interact with the ungrouped factor. We call this structure "hidden additivity." We develop a procedure which searches the space of all possible configurations, or placement of units into two (or more) groups. The method is illustrated using two datasets. the first describes copy number variation between normal and cancer tissues in dogs, and the second involves yield of fifteen wheat cultivars in 9 locations. Simulation is used to investigate the power of the proposed method and to compare it with other well-known tests for non-additivity under a variety of forms of interaction. A clustering method which uses a statistic computed from the all configurations approach is also developed.

KEYWORDS: non-additivity, interaction effects, latent variables.

Acknowledgements

We would like to thank Drs. Matthew Breen, Rachael Thomas, David Dickey, Thomas Severini, and Chris Smith for their contributions to this work.

1 Introduction

In factorial experiments factors are said to be additive if the effects of one factor on the response do not depend on the other factor. The assumption of additivity is commonly made when analyzing data from unreplicated two-factor experiments. Normal theory tests for interaction are unavailable here for lack of an appropriate error term. Any analysis which assumes additivity will produce biased results if non-additivity is present, so many authors have posited models with restricted forms of interaction which can be applied in the unreplicated case.

Consider a crossed two-way unreplicated experiment with factors A and B. Assume that the levels of factor B form a smaller number of groups and that there are interactions between group and factor A. Within groups, however, the effect of factor A is constant across levels of factor B leading to a structure we refer to as "hidden additivity." Group membership may be regarded as a latent variable, so we propose a test to detect non-additivity based on latent-grouping in Section 3. For a review of models involving latent variables see Guo et al. (2006) and the references therein.

Consider the balanced two-way analysis of variance (ANOVA) model for an independent response variable y_{ijk} observed in a crossed two-factor randomized experiment:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \ . \tag{1}$$

where $i = 1, .., a$ is an index for the levels of one of the two factors, $j = 1, ...b$ indexes the levels of the other factor, and $k = 1, ..., n$ is an index for the number of replicates for experimental units. Also we assume $\epsilon_{ijk} \sim N(0, \sigma^2)$. Note α_i, β_j, and $(\alpha\beta)_{ij}$ may be either fixed or random effects depending on the nature of the study. In the fixed effect case we adopt the sum-to-zero parameterization (e.g., $\sum_{i=1}^{a} \alpha_i = 0$). For random effects we assume normality (e.g., $\alpha_i \overset{iid}{\sim} N(0, \gamma^2)$ for $i = 1, ..., a$. In this scenario

all parameters are estimable and statistical inference can be performed using two-way ANOVA. If $n = 1$ the interaction mean square term is usually used to estimate σ^2, so there is no error-variance estimate to use in the denominator for the usual F tests for interaction effects.

The randomized complete block design (RCBD) is a well known example of a two-way unreplicated factorial design. The following additive model is commonly considered for RCBDs:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \ . \tag{2}$$

where $i = 1, .., a; j = 1, ...b; \sum_{i=1}^{a} \alpha_i = 0; \beta_j \sim N(0, \sigma_B^2); \epsilon_{ij} \sim N(0, \sigma^2)$. Frequently the treatment effects (denoted α_i) are considered fixed and the block effects (denoted β_j) are considered random in the RCBD experiment. Model (2) will be referred to as the "additive model."

In Section 2 we review current models and methods for addressing non-additivity in unreplicated factorial experiments. In Section 3 we develop a method to detect non-additivity using a latent variables approach. In Section 4 we apply the proposed method to two real world data sets and devise a related clustering algorithm. In Section 5 we compare the power of the proposed method with some well known methods under several forms of hidden additivity and non-additivity. Section 6 includes closing remarks.

2 Past Research

The one degree of freedom test for non-additivity was proposed by Tukey (1949). It has been pointed out (Alin and Kurt 2006) that the test is based on the following

model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \nu\alpha_i\beta_j + \epsilon_{ij} \ . \tag{3}$$

This model includes a single ν parameter in addition to those in (2). A test of non-additivity is formed by hypotheses $H_0 : \nu = 0$ versus $H_1 : \nu \neq 0$. The ANOVA table for this model includes a source of variability with one associated degree of freedom for the ν entry. A plot of the means from one version of (3) is shown in the third row and first column of Figure 3. Model (3) will henceforth be referred to as the "Tukey model." Tukey (1949) recommends transformation of the response to a scale where additivity holds in cases when the one degree of freedom test favors model (3) over model (2). Tukey (1955) describes how to apply this method to larger experiments.

Ghosh and Sharma (1963) derive the power function of Tukey's one degree of freedom test based on model (3). Anscombe and Tukey (1963) provide a graphical approach for detecting non-additivity of the form in model (3), and Cressie (1978) proposes a graphical technique to choose which Box-Cox (Box and Cox 1964) transformation to apply to the response to elicit additivity. Koziol (1989) proposes a bivariate generalization of Tukey's one degree of freedom test. McDonald (1972) provides a generalization of Tukey's one degree of freedom test for the multivariate case.

Mandel (1961) developed a generalization of Tukey's model.

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta_j\alpha_i + \epsilon_{ij} \ . \tag{4}$$

where $i = 1, ..., a > 2$; $j = 1, ..., b$. This model has been called the rows-linear model (Alin and Kurt 2006). A test for non-additivity with null hypothesis $H_0 : \theta_j = 0$ for all $j = 1, ..., b$ versus $H_1 : \theta_j \neq 0$ for some $j = 1, .., b$ is developed. A columns-linear model can be similarly defined. A plot of the means from one version of (4) is shown in the third row and second column of Figure 3. Snee (1982) reviews this model

and discusses the need for subject related knowledge to determine whether apparent non-additivity is due to a systematic interaction between factors, or nonhomogeneous variance.

Mandel (1971) put forth another model for non-additive effects:

$$y_{ij} = \mu + \alpha_i + \beta_j + \eta_{ij} \ . \tag{5}$$

where $\eta_{ij} = \sum_{q=1}^{Q} \theta_q U_{qi} \nu_{qj.}$, $Q \leq min(a-1, b-1)$. Both the value of Q and the estimates of parameters $\theta_q, U_{qi}, \nu_{qj.}$ are determined according to features of the spectral decomposition of $\boldsymbol{DD^T}$, where $\boldsymbol{D}$ is the matrix of residuals from the additive model defined by $d_{ij} = y_{ij} - \widehat{\mu} - \widehat{\alpha}_i - \widehat{\beta}_j$. The hat terms represent least squares estimators.

The parameters α_i and β_j account for all additive variability in the response, and the term η_{ij} is some combination of interaction effects plus random variability (to distinguish from ϵ_{ij} which denotes only random error). The parameters $\theta_q, U_{qi}, \nu_{qj.}$ are estimated using the matrix $\boldsymbol{DD^T}$, which has $min(a, b)$ real eigenvalues with orthogonal eigenvectors. If the full decomposition is performed there will be no terms left for residual error which leads to the restriction $Q \leq min(a-1, b-1)$. Mandel recommends choosing small Q so that only a few multiplicative interaction terms are included, combining the rest of the decomposition terms into residual error. Note that the estimates $\hat{\theta}_q$ for $q = 1, ..., Q$ are the eigenvalues of $\boldsymbol{DD^T}$. If there are no positive eigenvalues, then (5) reduces to the additive case. So testing the hypothesis $\theta_q = 0 \ \forall q = 1, ..., Q$ is analogous to testing for the presence of non-additivity by determining if any of the eigenvalues of $\boldsymbol{DD^T}$ differ significantly from zero. This eigenvalue decomposition strategy is similar to the Factor Analysis of Variance (FANOVA) technique illustrated by Gollob (1968).

Johnson and Graybill (1972) introduced a model to assess interaction in unreplicated

experiments. Their model is as follows:

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\tau_i\gamma_j + \epsilon_{ij} \ . \tag{6}$$

where $\sum_i^a \alpha_i = \sum_j^b \beta_j = \sum_i^a \tau_i = \sum_j^b \gamma_j = 0$, $\sum_i^a \tau_i^2 = \sum_j^b \gamma_j^2 = 1$, and $\epsilon_{ijm} \sim N(0, \sigma^2)$. Notice the second assumption need not constrain the structure of the model since λ can absorb any needed normalizing constants. The paper develops likelihood theory for this model including a likelihood ratio (LR) test of $\lambda = 0$. In this model the main effect for treatment and block have the same maximum likelihood estimators (MLEs) as in the additive model. The MLEs for parameters τ_i and γ_j are the components of the normalized eigenvector corresponding to the largest eigenvalues of the outer and inner products of the $a \times b$ matrix of residuals respectively. The parameter λ is estimated by the largest eigenvalue of the inner product of the $a \times b$ matrix of residuals. Johnson and Graybill develop a corresponding LR test of $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$ based on their model, and suggest that it may perform well in cases where the true form of the interaction is not a function of the treatment and block effects. This model is a specific case of (5) with $Q = 1$. Hegemann and Johnson (1976b) compare the power of the LRT above with Tukey's one degree of freedom test.

Hegemann and Johnson (1976a) discuss the following model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta_1\tau_{1i}\gamma_{1j} + \theta_2\tau_{2i}\gamma_{2j} + \epsilon_{ij} \tag{7}$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$, $b < a$, $\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a \tau_{1i} = \sum_{i=1}^a \tau_{2i} = \sum_{j=1}^b \gamma_{1j} = \sum_{j=1}^b \gamma_{2j} = \sum_{i=1}^a \tau_{1i}\tau_{2i} = \sum_{j=1}^b \gamma_{1j}\gamma_{2j} = 0$; $\sum_{i=1}^a \tau_{1i}^2 = \sum_{i=1}^a \tau_{2i}^2 = \sum_{j=1}^b \gamma_{1j}^2 = \sum_{j=1}^b \gamma_{2j}^2 = 1$, and $\epsilon_{ij} \sim N(0, \sigma^2)$. An approximate upper bound for σ^2 is provided along with tables to test treatment effects and the null hypothesis $\theta_2 = 0$.

Milliken and Rasmuson (1977) propose a strategy to detect interaction for k-factor unreplicated experiments where $k > 2$. The technique is based on choosing some mth factor among the k, and partitioning the data into t_i $k-1$ factor data sets by the levels of the mth factor, where t_i is the number of levels of the mth factor, $i-1, ..., k$. The interaction mean square S_j^2 is generated for each of the $j = 1, ...k-1$ data sets, and the equality of the expectation of these mean squares is tested to determine if the interaction of the $k-1$ factors change among various levels of the mth factor. The authors suggest max-min variance ratio and Bartlett's test for homogeneity of variance. Piepho (1994) suggests replacing the S_j^2 estimator with one proposed by Grubbs (1948) in order to better satisfy test assumptions. Since m can be chosen as any of the factors, up to k tests can be performed.

Tusell (1990) demonstrates that a sphericity test can be used to detect non-additivity when independence, homoscedasticity, and normality hold.

Boik (1993b) develops a locally best invariant (LBI) test for non-additivity in the unreplicated case. This method is similar to the Johnson and Graybill (1972) LR test since the form of interaction assumed is not a function of the main effects. The LBI test relies on the largest $p = min(a-1, b-1)$ normalized eigenvalues of $\boldsymbol{Z}^T\boldsymbol{Z}$, where$\boldsymbol{Z}$ is a matrix of residuals from the main effects model. Boik (1993a) compares the power of the LBI, the LR test from Johnson and Graybill (1972), and the Tusell (1990) sphericity test for non-additivity.

Zafar-Yab (1993) extends the Tukey one degree of freedom test to a p way layout. This method allocates $2^p - p - 1$ degrees of freedom for non-additivity, and each of these degrees of freedom is associated with a sum of squares for the linear component of interactions between each of the factors. For example in a $p = 3$ factor experiment,

$2^p - p - 1 = 4$ degrees of freedom and sums of squares terms exist for the linear×linear components of $A \times B$, $A \times C$, $B \times C$, and the linear×linear×linear components of $A \times B \times C$. The model for the $p = 3$ case is as follows:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \nu_k + \delta_1 \alpha_i \beta_j + \delta_2 \alpha_i \nu_k + \delta_3 \beta_j \nu_k + \delta_4 \alpha_i \beta_j \nu_k \ . \tag{8}$$

where $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{k=1}^{d} \nu_k = \sum_{i=1}^{a} \sum_{j=1}^{b} \alpha_i \beta_j = \sum_{i=1}^{a} \sum_{k=1}^{d} \alpha_i \nu_k = \sum_{j=1}^{b} \sum_{k=1}^{d} \nu_k \beta_j = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{d} \alpha_i \beta_j \nu_k = 0$. Notice that if $p = 2$ this is identical to the Tukey Model.

Speed (1994) develop a graphical approach and provide a test to detect non-additivity. The approach is based on computing the mean square error (MSE) for a sequence of models adding no interaction constraints sequentially. MSE for each of these models is plotted by sequence step, and any sudden drop in MSE indicates that the current model better satisfies the no interaction asusmption than any previous model. At any step the remaining matrix of constraints can be used as a hypothesis contract matrix to test for "no interaction" among the remaining steps.

Courcoux and Chavanne (2001) describe an Expectation Maximization (EM) algorithm based method of clustering H consumers who have 1 measurement on each of N products into T groups, where T is chosen *a priori*.

Pardo and Pardo (2005) address the analogous case of non-additivity in log-linear models for count data, including multinomial, product multinomial, and Poisson sampling schemes. A test statistic is developed to assess non-additivity using ϕ-divergences, which include Kullback-Leibler divergence as a special case.

Barker et al. (2009) develop the orthogonal interaction model (OIM) for unreplicated experiments. Interactions and main effects are constrained to be orthogonal.

This model may be used in two-way factorial designs whether or not replication is present. Starting with the original two-way model (1), the additional constraints $\sum_{i=1}^{a} \alpha_i(\alpha\beta)_{ij} = 0 \forall i$, and $\sum_{j=1}^{b} \beta_j(\alpha\beta)_{ij} = 0 \forall j$ are imposed which implies that main effects and interactions are orthogonal. Barker develops a remeasurement process in order to assign degrees of freedom for the purpose of testing for effects.

In practice the researcher may not know which of the available methods most accurately captures interaction present for a given application. In the literature, competing methods are generally compared in terms of statistical power for various forms of non-additivity.

3 Proposed method

We restrict our attention to two-way designs, and we adopt RCBD language by referring to the factor whose levels are grouped as "blocks," and the ungrouped factor as the "treatment." Our method is applicable to unreplicated factorial experiments regardless of the randomization used in the design.

If block membership in groups is known in advance, the following model is appropriate:

$$y_{ij(k)} = \mu + \alpha_i + \xi_k + (\alpha\xi)_{ik} + \beta_{j(k)} + \epsilon_{ij(k)} \tag{9}$$

where $i = 1, .., a; k = 1, ..., g; j = 1, ...b_k; g < b$. The treatment effects are denoted α_i, the group effects are denoted ξ_k, the treatment by group interaction parameters are denoted $(\alpha\xi)_{ik}$, the block effects are denoted $\beta_{j(k)}$, and the error terms are denoted $\epsilon_{ij(k)}$. The following constraints are adopted: $\sum_{i=1}^{a} \alpha_i = \sum_{k=1}^{g} \xi_k = \sum_{i=1}^{a} (\alpha\xi)_{ik} = \sum_{k=1}^{g} (\alpha\xi)_{ik} = 0$. Also $\beta_{j(k)} \sim N(0, \sigma^2_{B(G)})$, and $\epsilon_{ij(k)} \sim N(0, \sigma^2)$. Model (9) accommodates the full set of $(a-1)(g-1)$ treatment by group interaction parameters.

In Model (9) the block effects $\beta_{j(k)}$ are nested in the group effects. Within each group the treatment effects are constant across blocks. The treatment effects may vary across groups, so that there is group-by-treatment interaction. Within groups, the effects of treatments and blocks are additive.

Result If $(\alpha\xi)_{ik} = 0 \forall i, k$ then Model (9) reduces to Model (2) with treatment effects α_i and block effects $\beta_j^* = \xi_k + \beta_{j(k)}$. As a corollary, if Model (2) holds, then treatment by group interaction parameters are zero regardless of block assignment to groups. Since sum of squares for group by treatment interaction and error in Model (9) partition the the error sum of squares from Model (2), the standard F test for treatment by group interaction is equivalent to the F test for comparing models (9) and (2).

$$F = \frac{(SS(E)_{(2)} - SS(E)_{(9)})/(df_{E(9)} - df_{E(2)})}{MS(E)_{(9)}} \ . \tag{10}$$

where $F \overset{H_0}{\sim} F_{df_{E(9)}-df_{E(2)}, df_{E(9)}}$, $SS(E)$ represent the error sum of squares, $MS(E)$ represent the error mean square, df_E is the error degrees of freedom, and the subscript denotes the model.

If block membership in groups is unknown it can be treated as a latent variable. Of interest are the following two hypotheses:

$H_0 : g = 1$, i.e. blocks all belong to the same group. Additive model holds.
$H_1 : g = 2$, i.e. blocks belong to two groups. Treatment effects vary across the two groups.

We develop a test statistic for the above hypotheses by considering every configuration of blocks in two groups. Define a configuration as one possible assignment of b

blocks into g groups such that no group is empty and each block belongs to exactly one group. In the $g = 2$ case there are $c = 2^{(b-1)} - 1$ configurations.

Statistics such as sum of squares, p-values, etc. which arise from the c configurations share a dependency which has eluded generalization. The following theorem provides a level α test for the above hypotheses.

Theorem (Maximum interaction F test): Consider Model (9) for all c configurations. Rejecting $H_0 : (\alpha\xi)_{ik} = 0 \ \forall i, k$ based on the largest interaction F statistic at level $\frac{\alpha}{c}$ implies rejection of $H_0 : g = 1$ at level α.

Proof Let $(\alpha\xi)^l_{ik}$ denote treatment by group interaction parameters based on configuration l with $i = 1, \ldots, a,\ k = 1, \ldots, g$. If $g = 1$ then Model (2) holds and $(\alpha\xi)^l_{ik} = 0 \ \forall i, k, l$. Let F^* be the $(1 - \frac{\alpha}{c})$ percentile of the F distribution with $(a-1)(g-1)$ numerator and $(a-1)(b-g)$ denominator degrees of freedom. Let event A_l denote $F_l > F^*$. Let $P(A_l)$ denote the probability of event A_l with respect to the $F_{(a-1)(g-1),(a-1)(b-g)}$ distribution. Let $F_{(c)}$ be the largest interaction F statistic among the c configurations.

$P(A_l|H_0 \text{ true}) = \frac{\alpha}{c}$ for $l = 1, \ldots, c$. By the Bonferroni inequality $P(\bigcup_{l=1}^{c} A_l|H_0 \text{ true}) \leq \sum_{l=1}^{c} P(A_l|H_0 \text{ true}) = c \times \frac{\alpha}{c} = \alpha$. Hence rejecting $H_0 : g = 1$ if $P(F_l > F^*|H_0 \text{ true}) < \frac{\alpha}{c}$ for any l provides a level α test. Since $P(A_{(c)}|H_0 \text{ true}) \leq P(A_l|H_0 \text{ true}) \ \forall l$, rejection of $H_0 : (\alpha\xi)^{(c)}_{ik} = 0$ with $\frac{\alpha}{c}$ probability implies rejection of $H_0 : g = 1$ with α probability if the null hypothesis is true.

...

Intuition suggests that the configuration with the highest interaction F statistic provides the strongest evidence against additivity, and the above theorem shows that the

maximum of the interaction F values is a convenient statistic for testing additivity. The simulation study in Section 5 compares the power of this test to that of other methods for several latent group-based interaction scenarios.

4 Applications

Dog Lymphoma

It is well known that the genomes of cancer cells tend to become very different from those of normal cells, including in copy number. Copy number variation (CNV) in tumor cells can give tumors the ability to grow faster and overcome chemotherapy. One simple example of CNV can be seen by examining the X chromosome in humans: males have one copy of the X chromosome, while females have two copies. By examining a random sample containing both males and females and counting the number of copies per person, variation in X chromosome number will be observed. In addition to examining whole chromosomes, it may be of interest to search for smaller regions of the genome that display CNV, and specialized methods have been developed to do this. Comparative genomic hybridization (CGH) is a method designed to assess variation in copy number (or copy number variation, CNV) of regions of the genome across individuals. One active area of research is identifying genomic regions that tend to display CNV within given tumor types when compared with normal cells (and often whether identified variants can be related to treatment response or life expectancy).

In this example normal tissue samples from dogs are compared to tumor samples. The response variable for this technique is a light intensity reading which is used as a proxy for DNA increases or losses in tumor cells relative to normal cells. The

technique uses arrays on which both normal and tumor measurements are made. The arrays are variable and are considered blocks for our application. Probes correspond to a gene located on the chromosome which is positively related to lymphoma. Non-additivity exhibited in copy number variation could suggest that different arrays show discrepancies between normal and tumor cell copy number differently. This could be due to systematic differences between arrays and/or individuals.

These data have been made available courtesy of Drs. Matthew Breen and Rachael Thomas, College of Vet. Medicine, NCSU and the variables 1) gene 2) intensity 3) array id and 4) sample type (normal or tumor) are included as a supplement. Each of the probes is regarded as an individual data set to which a test for non-additivity is applied. The block term is the pairing of normal and test samples. There are $b = 6$ blocks per probe. The $a = 2$ treatment levels are the normal sample and the test sample. There are 5,899 candidate probes for which we test for non-additivity.

Table 1 compares the rejection rates of Tukey's test and ACMIF among all of the probes for the chromosome of interest. A significance level of $\alpha = 0.05$ was chosen for each test. The first entry in each cell is the count and the entry in parentheses is the proportion of counts in that cell. About sixty-six percent of the probes show no significant non-additivity for either method. Roughly twenty-two percent of the data sets show non-additivity according to the Tukey test but not ACMIF. About seven percent of the data sets show non-additivity according to ACMIF but not Tukey. Notice that the smallest percentage (4.64) of sets indicate non-additivity according to both Tukey and ACMIF methods. The non-zero off-diagonal counts suggest that there are probes which show non-additivity of a form that can only be detected by one of these two methods. The presence of these different classes of non-additivity emphasize the need for multiple methods to assess non-additivity in unreplicated ex-

Table 1: Comparison of rejection rates for Tukey's test and ACMIF for the dog lymphoma copy number data. Table counts refer to frequency of rejection of null hypothesis at $\alpha = 0.05$. Cell percentages are in parentheses. Non-zero off-diagonal counts suggest that non-additivity is present which is unique to each method.

	Tukey Fail to reject H_0	Tukey Reject H_0	Total
ACMIF Fail to reject H_0	3919 (66.43)	1296 (21.97)	5215 (88.40)
ACMIF Reject H_0	410 (6.95)	274 (4.64)	684 (11.60)
Total	4329 (73.38)	1570 (26.61)	5899 (100.00)

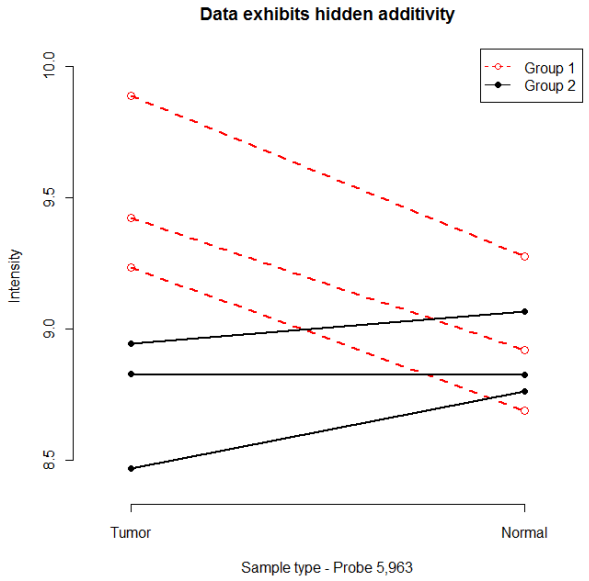
periments.

Figure 1 is an interaction plot of the signal intensities from probe 5,963. This plot suggests that the arrays may fall into two groups, and the effect of sample type (normal versus tumor) may be different in each of the groups. Perhaps there are CNV in these data, but a technical failure causes a subset of the arrays to not show the effect of CNV.

Wheat yields

Ramey and Rosielle (1983) provide data on wheat yields which includes mean yields (kg/ha) of 15 wheat cultivars grown in 9 locations in Western Australia in 1975. Regard the different cultivars as levels of the treatment, and consider the locations as blocks. The ACMIF test for latent group-based nonadditivity has a maximum

Figure 1: Data from probe 5,963 shows non-additivity according to the ACMIF test but not the Tukey test. Notice the similarity between this pattern and the Inert versus Active pattern in Figure 3.



F statistic of 10.26 on 14 numerator and 98 denominator degrees of freedom. The unadjusted ACMIF p-value for this test is 8.7×10^{-14} and the Bonferroni adjusted p-value (based on $c = 255$) is 2.2×10^{-11}. The highest F statistic corresponds to the configuration with the second location in one group and all other locations in the other group. Figure 2 shows this grouping using solid and dashed lines. For this data Tukey's one degree of freedom test has a p-value of 0.0019, and Mandel's rows linear test has a p-value of 0.05769.

ACMIF as a clustering algorithm

The ACMIF test can be implemented as a clustering algorithm. Lin (1982) and Ramey and Rosielle (1983) provide agglomerative clustering algorithms which use a minimum interaction MS criterion to place similar blocks into clusters. The ACMIF method can be used as a divisive clustering method where all blocks are initially in the same cluster. The algorithm proceeds in the following steps.

Step 1: Apply ACMIF test to data.

Step 2: If $H_0 : g = 1$ is rejected then place blocks into two groups according to the configuration with the maximum interaction F statistic. Otherwise end clustering algorithm.

Step 3: Repeat steps 1 and 2 within each established subgroup, and split any subgroup that has a significant ACMIF test according to the configuration with the highest F statistic for interaction. Subgroups must have more than two blocks to be eligible for the ACMIF test. Continue until no further splitting is done, or until all remaining subgroups have only one or two blocks. Final subgroups form the clusters.

The stopping rule described above is an advantage of the ACMIF clustering algorithm since a specific clustering is recommended. The methods of Ramey and Rosielle (1983)

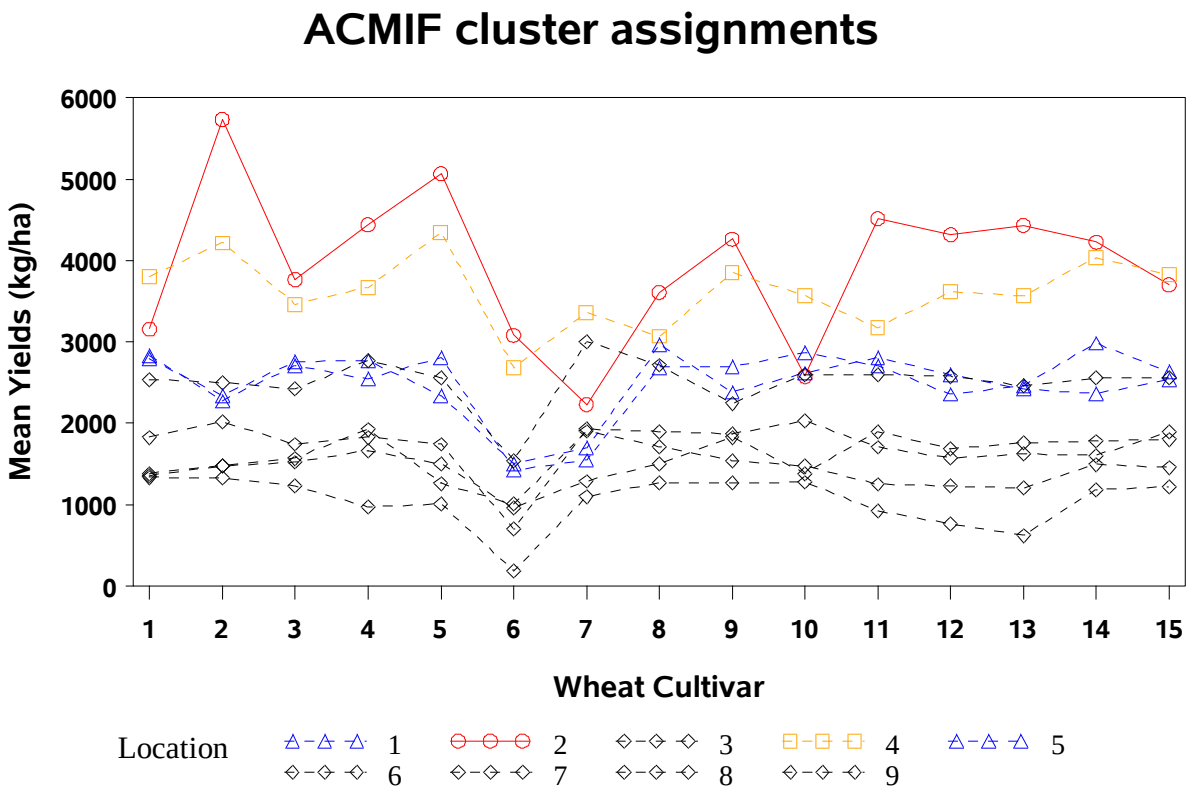
and Lin (1982) terminate when all blocks are placed in the same cluster, and different clusterings are presented at every iteration of the algorithms. If data is truly cluster-free then the ACMIF method stands a greater than $(1 - \alpha)$ chance of correctly concluding that blocks should not be split into clusters, while the Ramey and Rosielle (1983) and Lin (1982) methods provide spurious cluster assignments for lack of a stopping rule.

To illustrate the ACMIF clustering algorithm consider again the wheat data from Ramey and Rosielle (1983). We set $\alpha = 0.10$. Typically larger values of α will yield more splits (and hence clusters), while smaller values of α will result in fewer clusters. Figure 2 shows the result of the ACMIF clustering approach. This cluster assignment agrees with the clustering at the fifth iteration of both Lin's method and the average linkage method (Ramey and Rosielle 1983). Hence ACMIF clustering agrees with these other methods but also has the advantage of recommending a specific configuration among the 8 possibilities presented in the agglomerative methods.

5 Simulation

A simulation was conducted using SAS version 9.2 (SAS Inc., Cary, NC) to compare the performance of ACMIF, Tukey's one degree of freedom test, and Mandel's rows linear test in terms of statistical power. In all simulations $N = 1,000$ Monte Carlo (MC) data sets were generated with $a = 3$ and $b = 7$. Error variance was investigated at levels $\sigma^2 = 1, 5, 10$. For the purpose of computational simplicity the block effect was modeled as a fixed effect for the three methods presented. We used $\alpha = 0.05$ as a significance level for rejecting additivity in all cases.

Figure 2: Interaction plot of the wheat data. Dashed and solid lines correspond to ACMIF test for latent group-based interaction. Symbols show clustering results.



Seven specific data generating models were investigated in this study. Four of the models were of the latent group-based type (Model (9)), one was in the Tukey class (Model (3)), one was a rows linear model (Model (4)), and the last was a full factorial model (Model (1)). These models are visually represented in Figure 3, and the parameter values used are listed in Table 3.

The top four panels in Figure 3 are latent group type models, and the two lines in each plot represent the behavior of each group across the treatment levels. For these scenarios one group contained three blocks and the other four. Data were generated according to Model (9) for these four cases and $\sigma^2_{B(G)} = 25$ was used for block variance in each case. We have chosen the labels "Inert versus Active," "Average versus Active," "Cancellatory," and "Accordion" to describe these four latent group structures. These examples do not encompass all possible forms of latent group-based non-additivity.

The next three panels represent data generated according to non-latent group models (3) and (4) and (1). Each of these panels includes $b = 7$ lines since these three models prescribe a unique behavior for each block. This full factorial setting can be well represented using the two latent group model even though the data generating model includes the full set of standard interaction parameters, so we refer to this setting as 'Groupable.' The model does not exhibit exact hidden additivity exactly, but the treatment effects are groupable.

Power results for the three methods are presented in Table 2. Tukey's test and the rows linear test are more powerful than ACMIF for data generated from models (3) and (4). The ACMIF approach is more powerful than Tukey's test for Cancellatory and Accordion types of non-additivity. As σ^2 increases, Tukey's test is more

competitive with ACMIF for Inert versus Active and Average versus Active scenarios.

The rows linear test for non-additivity is competitive with ACMIF for a number of latent group-based forms of non-additivity, and it would appear to enjoy higher power than ACMIF in Inert versus Active and Average versus Active cases as σ^2 increases. Conversely, ACMIF is not competitive with the rows linear test for data which arises from the rows linear model.

The Groupable setting favors the ACMIF test over the rows linear and Tukey tests. In this setting the true form of interaction cannot be tested for due to lack of replication. Since the data here are better represented by Model (9) than either Model (3) or (4) ACMIF shows the highest power. Other full-factorial interaction settings might be better detected by either the rows linear test or the Tukey test.

ACMIF is competitive or superior in terms of power when the true form of non-additivity is latent group-based or can be well represented by a latent group structure. ACMIF has lower power when other forms of non-additivity are present. The rows linear test rejects additivity with high power for a variety of forms of non-additivity. Hence rejection of additivity on the basis of the rows linear does not necessarily suggest that the form of non-additivity present conforms to Model (3).

6 Discussion

ACMIF for large experiments

One feature of the ACMIF approach is that the number of possible configurations c can be large for even moderate sized experiments. While our simulation study sug-

Figure 3: Simulation settings.

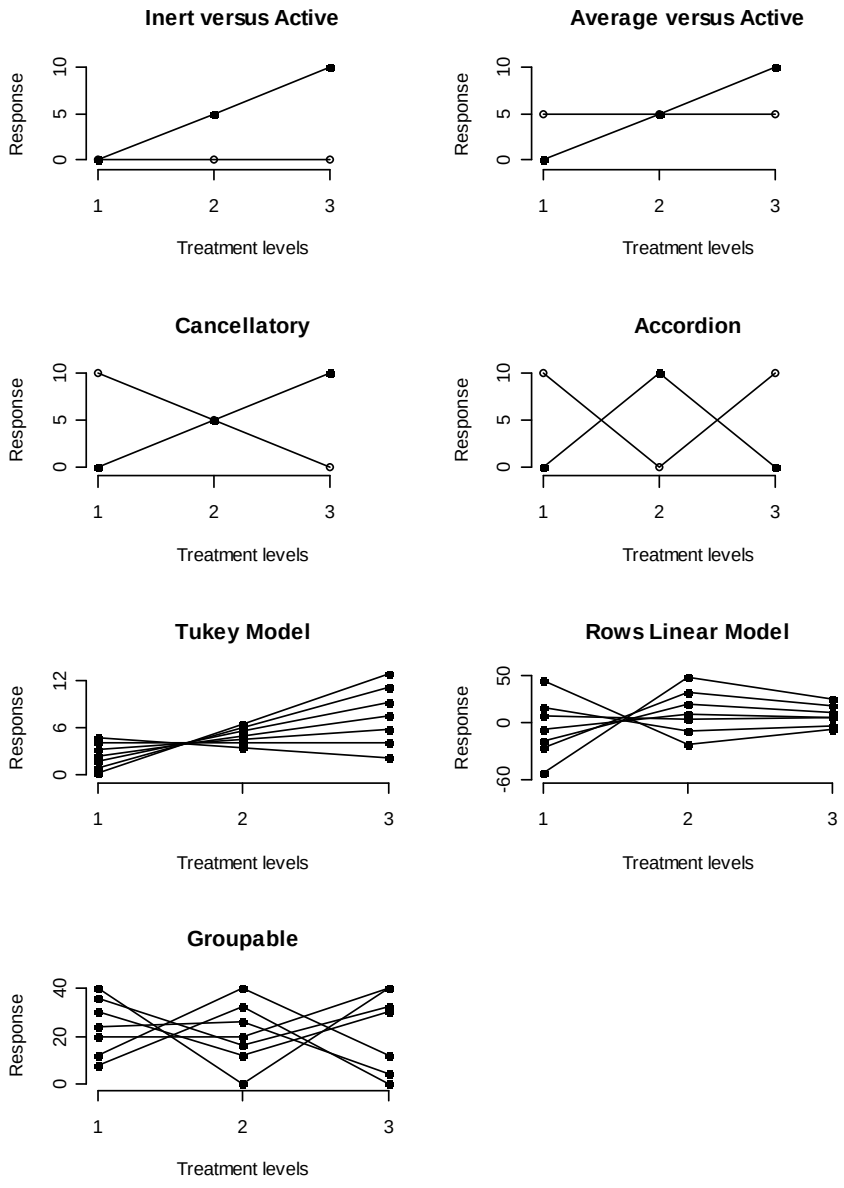


Table 2: Empirical Power simulation results. First two columns describe experimental setup. Third, fourth, and fifth columns provide empirical power for ACMIF, Mandel's and Tukey's test at $\alpha = 0.05$. Standard error for all proportions presented less than 0.0159.

$a = 3$; $b = 7$; MC reps:1000		Proportion of H_0 rejections at $\alpha = 0.05$		
σ^2	Data type	ACMIF	Mandel	Tukey
1	Inert vs. active	0.992	0.985	0.424
5		0.273	0.380	0.302
10		0.095	0.189	0.170
1	Average vs. active	0.992	0.992	0.200
5		0.249	0.380	0.118
10		0.109	0.225	0.108
1	Cancellatory	1.000	0.743	0.147
5		0.961	0.380	0.081
10		0.626	0.231	0.060
1	Accordion	1.000	0.804	0.256
5		0.992	0.415	0.184
10		0.794	0.305	0.144
1	Tukey	0.202	0.981	0.993
5		0.125	0.417	0.455
10		0.073	0.211	0.192
1	Rows Linear	0.007	1.000	0.312
5		0.050	0.982	0.261
10		0.091	0.858	0.205
1	Groupable	0.632	0.011	0.134
5		0.459	0.135	0.240
10		0.398	0.173	0.224

gests that our procedure which uses the Bonferroni adjustment has reasonable power for moderate sized experiments, computational expense can be an issue for larger experiments.

One benefit of the Bonferroni adjustment based on $c = 2^{b-1} - 1$ configurations is that any subset of configurations may be tested to screen for non-additivity. If certain configurations seem likely based on *a priori* knowledge of the experiment or from inspection of an interaction plot then these configurations can be tested in advance, and if any of these initial tests exceed the $\frac{\alpha}{c}$ critical value then additivity can be rejected without the need to test all c configurations. Let $F_\diamond$ be the treatment by group interaction F ratio for the configuration of interest. Denote $p_\diamond$ as the corresponding p-value. Notice $F_{(C)} \geq F_\diamond$ which implies $p_{(1)} \leq p_\diamond$. If $p_\diamond < \frac{\alpha}{c}$ then $p_{(1)} < \frac{\alpha}{c}$ and we can reject $H_0 : g = 1$ and conclude that the additive model is not sufficient. On the other hand, if $p_\diamond \geq \frac{\alpha}{c}$ then one would not necessarily favor the null hypothesis without considering the remaining configurations.

Conclusion

Hidden additivity is a form of non-additivity that appears when the levels of one factor fall into a smaller number of latent groups, and the groups interact with the other factor. A method for detecting hidden additivity in unreplicated experiments has been proposed. The ACMIF method is based on a model which accounts for latent-group driven interaction. The ACMIF method is shown to perform competitively for latent group-based interaction and is applied to real-world data. The method can be implemented as a clustering algorithm, and the significance based decisions in this algorithm avoid spurious clustering with greater than $(1 - \alpha)$ probability. The ACMIF method is applicable in all fields of study which utilize unreplicated experiments.

References

Alin, A. and S. Kurt (2006). Testing non-additivity (interaction) in two-way anova tables with no replication. *Statistical Methods in Medical Research 15*, 63–85.

Anscombe, F. and J. Tukey (1963). The examination and analysis of residuals. *Technometrics 5*, 141–160.

Barker, C., L. Stefanski, and J. Osborne (2009). The orthogonal interactions model for unreplicated factorial experiments, unpublished manuscript.

Boik, R. (1993a). A comparison of three invariant tests of additivity in two-way classifications with no replications. *Computational Statistics and Data Analysis 15*, 411–424.

Boik, R. (1993b). Testing additivity in two-way classifications with no replications: the locally best invariant test. *Journal of Applied Statistics 20*, 41–55.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B 26*, 211–252.

Courcoux, P. and P. Chavanne (2001). Preference mapping using a latent class vector model. *Food Quality and Preference 12*, 369–372.

Cressie, N. (1978). Removing nonadditivity from two-way tables with one observation per cell. *Biometrics 34*, 505–513.

Ghosh, M. and D. Sharma (1963). Power of Tukey's test for non-additivity. *Journal of the Royal Statistical Society. Series B 1*, 213–219.

Gollob, H. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika 33*, 73–115.

Grubbs, F. (1948). On estimation of precision of measuring instruments and product variability. *Journal of the American Statistical Association 43*, 243–264.

Guo, J., M. Wall, and Y. Amemiya (2006). Latent class regressors on latent factors. *Biostatistics 7,1*, 145–163.

Hegemann, V. and D. Johnson (1976a). On analyzing two-way aov data with interaction. *Technometrics 18*, 273–281.

Hegemann, V. and D. Johnson (1976b). The power of two tests for nonadditivity. *Journal of the American Statistical Association 71*, 945–948.

Johnson, D. and F. Graybill (1972). An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association 67*, 862–868.

Koziol, J. (1989). Multivariate tests for non-additivity. *Statistical Papers 30*, 27–37.

Lin, C. (1982). Grouping genotypes by a cluster method directly related to genotype-environment interaction mean square. *Theoretical and Applied Genetics*, 277–280.

Mandel, J. (1961). Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association 56*, 878–888.

Mandel, J. (1971). A new analysis of variance model for non-additive data. *Technometrics 13*, 1–18.

McDonald, L. (1972). A multivariate extension of tukey's one degree of freedom for non-additivity. *Journal of the American Statistical Association 675*, 674–675.

Milliken, G. and D. Rasmuson (1977). A heuristic technique for testing for the presence of ineraction in nonreplicated factorial experiments. *Australian Journal of Statistics 19*, 32–38.

Pardo, L. and M. Pardo (2005). Nonadditivity in loglinear models using ϕ-divergences and MLEs. *Journal of Statistical Planning and Inference 127*, 237–252.

Piepho, H. (1994). On tests for interaction in a nonreplicated two-way layout. *Australian Journal of Statistics 36*, 363–369.

Ramey, T. and A. Rosielle (1983). Hass cluster analysis: a new method of grouping genotypes or environments in plant breeding. *Theoretical and Applied Genetics*, 131–133.

Snee, R. (1982). Nonadditivity in a two-way classification: Is it interaction or nonhomogeneous variance. *Journal of the American Statistical Association 77*, 515–519.

Speed, F.M., S. F. (1994). An ad-hoc diagnostic tool for checking for interaction in a nonreplicated experiment. *Communications in Statistics - Theory and Methods 23*, 1365–1374.

Tukey, J. (1949). One degree of freedom for non-additivity. *Biometrics 5*, 232–242.

Tukey, J. (1955). Answers to query 113. *Biometrics 11*, 111.

Tusell, F. (1990). Testing for interaction in two-way anova tables with no replication. *Computational Statistics and Data Analysis 10*, 29–45.

Zafar-Yab (1993). On testing for non-additivity in factorial experiments. *Biometrical Journal 35*, 925–932.

7 Appendix

Table 3: Non-additivity labels and parameter values for each setting in simulation.

Non-additivity	Parameter Values
Inert vs. active	$\mu = \alpha_3 = \xi_2 = (\alpha\xi)_{11} = (\alpha\xi)_{32} = 2.5$; $\alpha_2 = (\alpha\xi)_{21} = (\alpha\xi)_{22} = 0$; $\alpha_1 = \xi_1 = (\alpha\xi)_{12} = (\alpha\xi)_{31} = -2.5$, $\sigma^2_{B(G)} = 25$.
Average vs. active	$\mu = 5$, $\alpha_1 = (\alpha\xi)_{12} = (\alpha\xi)_{31} = -2.5$, $\alpha_2 = \xi_1 = \xi_2 = (\alpha\xi)_{21} = (\alpha\xi)_{22} = 0$, $\alpha_3 = (\alpha\xi)_{11} = (\alpha\xi)_{32} = 2.5$, $\sigma^2_{B(G)} = 25$.
Cancellatory	$\mu = 5$, $\alpha_1 = \alpha_2 = \alpha_3 = \xi_1 = \xi_2 = (\alpha\xi)_{21} = (\alpha\xi)_{22} = 0$, $(\alpha\xi)_{12} = (\alpha\xi)_{31} = -5$, $(\alpha\xi)_{11} = (\alpha\xi)_{32} = 5$, $\sigma^2_{B(G)} = 25$.
Accordion	$\mu = 5$, $\alpha_1 = \alpha_2 = \alpha_3 = 0$, $\xi_1 = -1.67$, $\xi_2 = 1.67$, $(\alpha\xi)_{11} = (\alpha\xi)_{31} = 3.33$, $(\alpha\xi)_{12} = (\alpha\xi)_{32} = -3.33$, $(\alpha\xi)_{21} = -6.67$, $(\alpha\xi)_{22} = 6.67$, $\sigma^2_{B(G)} = 25$.
Tukey Model	$\mu = 5$, $\alpha_1 = -2.5$, $\alpha_2 = \beta_4 = 0$, $\alpha_3 = 2.5$, $\beta_1 = -1.5$, $\beta_2 = -1$, $\beta_3 = -0.5$, $\beta_5 = 0.5$, $\beta_6 = 1$, $\beta_7 = 1.5$, $\nu = 1$.
Rows Linear Model	$\mu = 5$, $\alpha_1 = -5$, $\alpha_2 = 3.5$, $\alpha_3 = 1.5$, $\beta_1 = -1.5$, $\beta_2 = -1$, $\beta_3 = -.5$, $\beta_4 = 0$, $\beta_5 = .5$,$\beta_6 = 1$,$\beta_7 = 1.5$, $\theta_1 = -1.25$, $\theta_2 = 0$, $\theta_3 = .625$, $\theta_4 = -2.5$,$\theta_5 = -.625$, $\theta_6 = 2.5$, $\theta_7 = 1.25$.
Groupable	$\mu = 22.57$, $\alpha_1 = 1.71$, $\alpha_2 = -1.71$, $\alpha_3 = 0$, $\beta_1 = -4.57$, $\beta_2 = -1.24$, $\beta_3 = -9.25$, $\beta_4 = 4.10$, $\beta_5 = 1.43$, $\beta_6 = 5.43$, $\beta_7 = 4.10$, $(\alpha\beta)_{11} = 1.43$, $(\alpha\beta)_{21} = 3.43$, $(\alpha\beta)_{31} = -18.57$, $(\alpha\beta)_{12} = -10.57$, $(\alpha\beta)_{22} = 17.43$, $(\alpha\beta)_{32} = -10.57$, $(\alpha\beta)_{13} = -14.57$, $(\alpha\beta)_{23} = 9.43$, $(\alpha\beta)_{33} = -22.58$ $(\alpha\beta)_{14} = -2.58$, $(\alpha\beta)_{24} = -2.58$, $(\alpha\beta)_{34} = 17.43$, $(\alpha\beta)_{15} = 7.43$, $(\alpha\beta)_{14} = -2.58$, $(\alpha\beta)_{24} = -2.58$, $(\alpha\beta)_{34} = 17.43$, $(\alpha\beta)_{14} = -2.58$, $(\alpha\beta)_{24} = -2.58$, $(\alpha\beta)_{34} = 17.43$, $(\alpha\beta)_{15} = 7.43$, $(\alpha\beta)_{25} = -10.57$, $(\alpha\beta)_{15} = 7.43$, $(\alpha\beta)_{25} = -10.57$, $(\alpha\beta)_{35} = 7.43$, $(\alpha\beta)_{16} = 13.43$, $(\alpha\beta)_{26} = -6.57$, $(\alpha\beta)_{36} = 9.43$,$(\alpha\beta)_{17} = 17.43$, $(\alpha\beta)_{27} = -22.57$, $(\alpha\beta)_{37} = 17.43$.

Table 4: Degrees of freedom for model (2)

Source	df
Treatment	a-1
Block	b-1
Error	(a-1)(b-1)
Total	ab - 1

Table 5: Degrees of freedom for model (9)

Source	df
Treatment	a-1
Group	g-1
Treatment×Group	(a-1)(g-1)
Block(Group)	b-g
Error	(a-1)(b-g)
Total	ab - 1