Computational Statistics and Data Analysis 67 (2013) 95–104

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A method for detecting hidden additivity in two-factor unreplicated experiments[☆]

Christopher T. Franck[a,*], Dahlia M. Nielsen[b], Jason A. Osborne[c]

[a] Department of Statistics, 213 A Hutcheson Hall, Virginia Tech, Blacksburg, VA 24061-0439, United States
[b] Department of Genetics, North Carolina State University, Raleigh, NC, United States
[c] Department of Statistics, North Carolina State University, Raleigh, NC, United States

ARTICLE INFO

Article history:
Received 19 June 2012
Received in revised form 4 May 2013
Accepted 5 May 2013
Available online 23 May 2013

Keywords:
Interaction effects
Latent variables
Non-additivity

ABSTRACT

Assessment of interaction in unreplicated two-factor experiments is a challenging problem that has received considerable attention in the literature. A model is proposed 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. This structure is called "hidden additivity" if group membership is latent. To identify plausible groupings a search is performed over the space of all possible configurations, or placement of units into two or more groups. A multiplicity-adjusted all-configurations maximum interaction F (ACMIF) test to detect hidden additivity is developed. The method is illustrated using two data sets taken from the literature and a third taken from a recent study of copy number variation due to lymphoma. A simulation study demonstrates the power of the test for hidden additivity and compares it with other well-known tests from the literature.

Published by Elsevier B.V.

1. Introduction

Consider a crossed two-way replicated experiment with factors A and B. The analysis of variance (ANOVA) model for response y_{ijk} is

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

where $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n$ are indices for levels of A, levels of B, and experimental units, or replicates, respectively. The experimental errors are typically assumed $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$. The α_i and β_j terms quantify the main effects of each factor. The $(\alpha\beta)_{ij}$ terms represent interaction parameters, which allow for the possibility that the effects of one factor on the response depend on the level of the other factor. The concept of interaction is fundamental in statistical analysis and determining its presence or absence is important for correct interpretation of data (Berrington de González and Cox, 2007).

In the unreplicated case where $n = 1$, the interaction mean square term is usually used to estimate σ^2. This leaves no information to assess interaction.

Cox (1984) provides a general overview of interaction effects. Many authors have posited restricted forms of interaction which are applicable to the unreplicated case. Alin and Kurt (2006) provides an overview of many existing methods to detect interaction in the unreplicated case, including reviews of Tukey (1949), Mandel (1961, 1971), Snee (1982), Johnson and Graybill (1972), Tusell (1987, 1990) (also see Tusell, 1992), and Boik (1993).

[☆] The data set "CNVdata.csv" is available as an online supplement.

[*] Corresponding author. Tel.: +1 540 231 4375; fax: +1 540 231 3863.
E-mail address: chfranck@vt.edu (C.T. Franck).

0167-9473/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.csda.2013.05.002

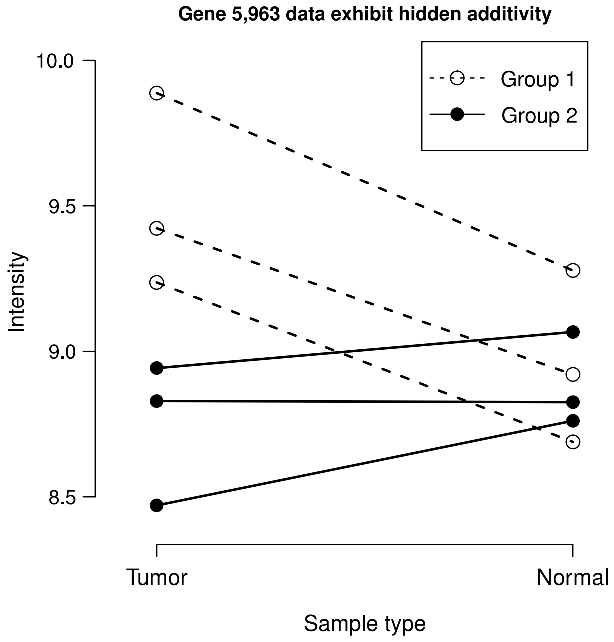

Fig. 1. Data from gene 5,963 shows non-additivity according to the ACMIF test but not the Tukey test.

Ghosh and Sharma (1963), Anscombe and Tukey (1963), Cressie (1978), Koziol (1989), McDonald (1972), and Zafar-Yab (1993) discuss and extend the one degree of freedom test proposed by Tukey (1949). Other techniques which detect non-additivity include those proposed by Gollob (1968), Hegemann and Johnson (1976a,b), Milliken and Rasmuson (1977), Piepho (1994), Speed and Speed (1994), Pardo and Pardo (2005), Barker et al. (2009), and Post and Bondell (2011). These approaches generally assume a specific form of non-additivity and they differ in statistical power to detect various non-additive forms.

A common approach when analyzing unreplicated data is to assume additivity, or that $(\alpha\beta)_{ij} \equiv 0$. An example is the model commonly used to analyze data from the randomized complete block design (RCBD):

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

If the two factors in (2) genuinely interact then this model is misspecified, which will result in biased estimation. As an alternative to the additivity assumption, suppose that the levels of factor B fall into a smaller number of groups such that within these groups the effect of factor A is constant across levels of factor B, but interaction between group and factor A is possible. Fig. 1 illustrates this idea. We refer to this structure as "hidden additivity" and regard group membership as a latent variable. For a review of models involving latent variables, see Guo et al. (2006).

In Section 2, a test for interaction is developed based on the idea of hidden additivity and latent variables. All possible group assignments are considered and the configuration which provides the most evidence of hidden additivity is identified. A formal test of the null hypothesis of no interaction is developed by comparing the maximum A by Group interaction F ratio to an appropriate critical threshold. In an approach closely related to the test developed in Section 2, Kharrati-Kopaei and Sadooghi-Alvandi (2007) develops a procedure based on all of the possible configurations in which the levels of factor B in (2) can be split into two sub-tables. This approach is discussed in Section 2.

In Section 3, two data sets taken from the literature are shown to exhibit the hidden additivity form of interaction. The third example is a genomic study of copy number variation in dogs with lymphoma. This study serves as a rich source of two-factor unreplicated data including thousands of response variables (gene signal intensities). In Section 4, the power of the proposed method is compared with some well-known methods under several forms of non-additivity. Section 5 includes closing remarks.

2. Proposed method

Consider the class of two-way unreplicated designs, i.e. a single observation y_{ij} is available for each $i = 1, \ldots, a$ and $j = 1, \ldots, b$. For convenience we refer to the levels of factor B as the "blocks", and factor A as the "treatment", although the method applies generally to two-way unreplicated experiments. Let $m_{ij} = E[y_{ij}]$ and $\boldsymbol{m}^{(j)} = (m_{1j}, \ldots, m_{aj})^T$, i.e. the jth column vector of the array of means. Under the additive model (2), $\boldsymbol{m}^{(j)} = \boldsymbol{\alpha} + (\mu + \beta_j)\mathbf{1}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_a)^T$ and $\mathbf{1}$ is an a-dimensional vector of ones. The $\boldsymbol{m}^{(j)}$ vectors differ from $\boldsymbol{\alpha}$ by a constant. This produces an "additive shift" which appears in interaction plots as parallel treatment means between blocks.