Exploring Interaction Effects in Two-Factor Studies using the hiddenf Package in R.

by Christopher T. Franck and Jason A. Osborne

Abstract In crossed, two-factor studies with one observation per factor-level combination, interaction effects between factors can be hard to detect and can make the choice of a suitable statistical model difficult. This article describes **hiddenf**, an R package that enables users to quantify and characterize a certain form of interaction in two-factor layouts. When effects of one factor (a) fall into two groups depending on the level of another factor, and (b) are constant within these groups, the interaction pattern is deemed "hidden additivity" because within groups, the effects of the two factors are additive, while between groups the factors are allowed to interact. The **hiddenf** software can be used to estimate, test, and report an appropriate factorial effects model corresponding to hidden additivity, which is intermediate between the unavailable full factorial model and the overly-simplistic additive model. Further, the software also conducts five statistical tests for interaction proposed between 1949 and 2014. A collection of 17 datasets is used for illustration.

Introduction

The crossed, two-factor layout with one observation per factor-level combination is a simple and widely used plan. These designs arise in the context of (a) completely randomized two-factor experiments, (b) randomized complete block experiments that block on a source of nuisance variability, and (c) observational studies with two factors. If the effects of the two factors do not interact, then this design permits inference for both factors simultaneously. The following model is commonly used for data collected using these designs:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \tag{1}$$

where $i = 1, \ldots, c$ and $j = 1, \ldots, r$ are indices for levels of the two factors C and R, μ represents the overall mean of the outcome y_{ij}, and the τ_i and β_j terms quantify the effects of factors C and R. The errors are frequently assumed independent and identically normally distributed, with constant error variance σ^2. Since data from these designs are easy to display using two-way tables, frequently the levels of factors R and C are called "rows" and "columns," respectively. In this report (and the **hiddenf** manual), Factor R is referred to as the "row" or "grouped" factor, while factor C is called the "column" factor.

Model (1) assumes that the effect of the row factor on the response y_{ij} does not depend on the level of the column factor, and vice versa. In statistical parlance this is known as the assumption of "additivity." This assumption is sometimes made out of necessity as the alternative of including the interaction term from the full factorial effects model precludes statistical inference. After computing the interaction mean square, zero degrees of freedom remain with which to estimate the error variance, σ^2. Data arising from (1) are shown in Figure 1 panel A.

If the effect of factor C varies across levels of the factor R, then there is statistical interaction. In this case (1) is misspecified. Graphical assessment of additivity is commonly made using an interaction plot. Figure 1 panel B is an interaction plot for the `cjejuni.mtx` data set included in the **hiddenf** package (Qiu, 2013).

Since the data in Figure 1 panel B do not exhibit parallelism, there is graphical evidence to suggest that (1) may not be a suitable model. This is problematic because inferences conducted under (1) will be incorrect if (1) fails to describe the true nature of the variables under investigation. Fortunately, many methods have been devised to assess non-additivity in these designs, including those in Tukey (1949), Anscombe and Tukey (1963), Mandel (1961), Mandel (1971), Johnson and Graybill (1972), Hirotsu (1982), Kharrati-Kopaei and Sadooghi-Alvandi (2007), Tusell (1990), Boik (1993a), Franck et al. (2013), and Malik et al. (2014). Alin and Kurt (2006) provide a review of methods available up to 2006. We return to the analysis of these and other data in subsequent sections of the paper.

Despite the wide usage of unreplicated two-way designs, computational tools to assess non-additivity
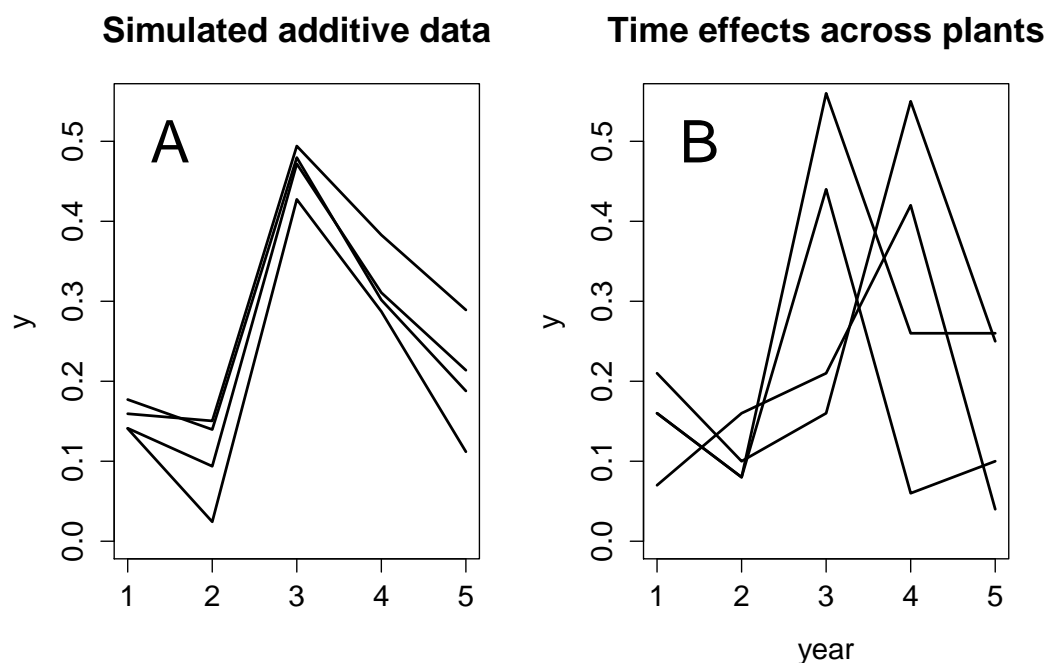
Simulated additive data **Time effects across plants**

Figure 1: Interaction plots. Lines correspond to levels of factor *R* and tick marks on the horizontal axis correspond to levels of factor *C*. Panel A is simulated data from (1). These data do not exhibit statistical interaction so departure from parallel lines is due to random noise. Panel B is `cjejuni.mtx` data, which visually exhibit non-additivity.

are lacking, particularly for more recent methods. Currently, the **additivityTests** package (Simecek and Simeckova, 2012) provides test statistics, critical thresholds, and binary reject/fail to reject decisions for tests proposed in Tukey (1949), Mandel (1961), Boik (1993a), Tusell (1990), Johnson and Graybill (1972), and Simecek and Simeckova (2012). The **additivityTests** package does not produce p-values.

The hiddenf package

This paper describes the R package **hiddenf**, which makes available several tools for the analysis of non-additivity in unreplicated two-way layouts. This package contributes to available computational resources by (a) providing statistical and graphical diagnostic tools within the context of hidden additivity that enable the user to go beyond overall tests of additivity towards the inferential goal of characterizing and quantifying the magnitude of interaction, and (b) providing p-value computations for five methods to detect non-additivity. Supported tests include those proposed in Tukey (1949), Mandel (1961), Kharrati-Kopaei and Sadooghi-Alvandi (2007), Franck et al. (2013), and Malik et al. (2014). The latter three are newly available in an open-source repository via the **hiddenf** package, which is available from the Comprehensive R Archive Network (CRAN).

Hidden additivity occurs when the levels of factor *R* fall into a smaller number of groups, such that within these groups the effect of factor *C* is constant across levels of factor *R*, but there is $C \times group$ interaction (Franck et al., 2013). Group membership of levels of *R* can be regarded as latent, unobservable random variables. Inspection of Figure 1 panel B provides an example. If one group is formed from the two lines (processing plants) that reach their peak in year 3, and another from the two lines that reach their peak in year 4, then the year and plant effects are roughly additive within these groups.

The `HiddenF()` function accepts an $r \times c$ data matrix as input and returns an object of the *HiddenF* class. Objects of this class store the p-value from the all-configurations maximum interaction F (ACMIF) test (Franck et al., 2013) for hidden additivity, the number of configurations under consideration, the configuration that achieves the maximum interaction F score (or maximum hidden additivity), and the data as a list. Several generic functions, including `print()`, `anova()`, `summary()` and `plot()`, can be applied to objects of the *HiddenF* class to produce output useful for the quantification and characterization of interaction. By default, the `HiddenF()` function considers the row factor as the

variable to group. Submitting a transposed matrix executes the grouping on the column variable. Deciding which factor to group is subjective, and can be approached using domain knowledge or a trial-and-error approach. Figure 2 summarizes the functionality of the **hiddenf** package.
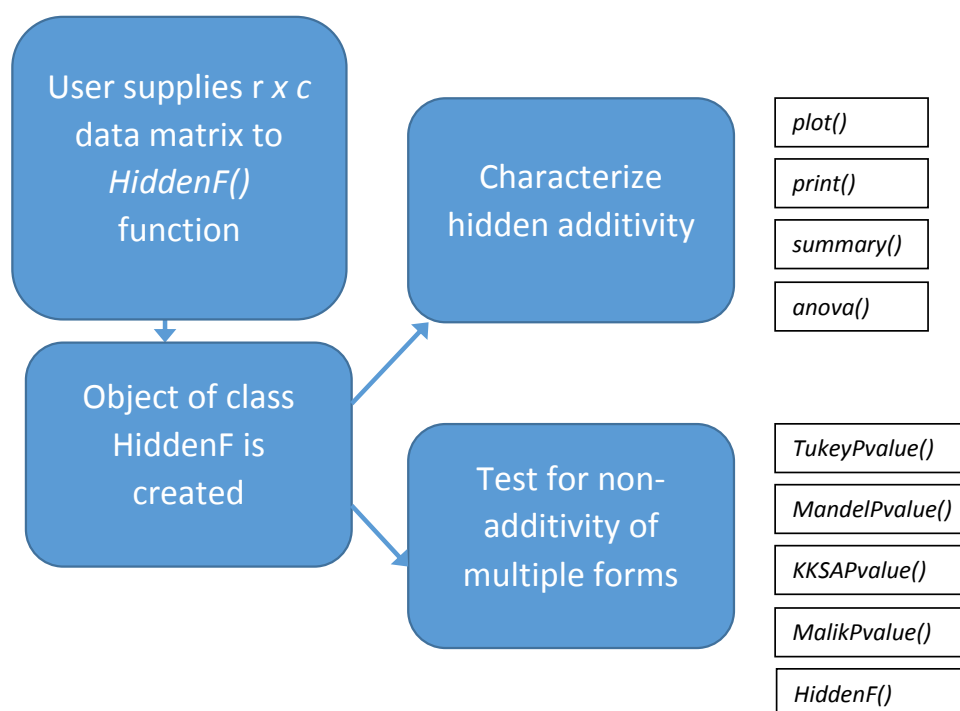

Figure 2: Flowchart of **hiddenf** functionality. The user supplies a $r \times c$ data matrix to the `HiddenF()` function, which returns an object of the *HiddenF* class. The user can then characterize hidden additivity and obtain p-values for tests of non-additivity.

The remainder of the paper is organized as follows. The *C. jejuni* data are first described as a relevant example. The following section reviews the testing procedure for hidden additivity. Functionality to detect hidden additivity using **hiddenf** is then detailed. Four other supported methods to detect non-additivity are then reviewed. Seventeen data sets are described and explored using the methods supported by **hiddenf**. The article concludes with a summary.

Example

We now return to the data from Figure 1 panel B. The vertical axis in this plot is the proportion of disease-resistant bacteria samples of the *C. jejuni* strain taken from turkey processing facilities. These data were collected over a five year period across four processing plants in North Carolina. Each line corresponds to a plant, and the year labels correspond to 2008-2012. The matrix `cjejuni.mtx` contains the fractions of bacteria that are classified as the strain *C. jejuni*.

```
> library(hiddenf)
> data(cjejuni.mtx)
> cjejuni.mtx

     [,1] [,2] [,3] [,4] [,5]
[1,] 0.16 0.08 0.44 0.06 0.10
[2,] 0.21 0.10 0.16 0.55 0.25
[3,] 0.16 0.08 0.56 0.26 0.26
[4,] 0.07 0.16 0.21 0.42 0.04
```

Testing for hidden additivity

Hidden additivity is a useful concept for the analysis of many types of data because the structure is easy to visualize and interpret. The ACMIF test (Franck et al., 2013) to detect hidden additivity assigns classical significance testing-based p-values. A combined testing and plotting approach allows the researcher to interpret the way in which column effects change across groups of the rows.

The ACMIF test works by (1) considering each placement of factor R levels into two nonempty groups, (2) testing factor $C \times group$ interaction, and (3) considering the test that provides the most evidence of interaction (and greatest additivity of C and R within groups) after correcting for multiplicity using a Bonferroni adjustment. Simulation suggests that the Bonferroni correction is not overly conservative for $r \leq 7$ despite the magnitude of the correction factor $2^{r-1} - 1$, because the Bonferroni adjusted thresholds are remarkably similar to simulated critical thresholds for a variety of c and r values (Franck et al., 2013). If g is an index for the latent group membership such that $g = 1, 2$, then the factorial effects model corresponding to the hidden additivity structure is given by

$$y_{ij}^g = \mu + \tau_i + \gamma_g + (\tau\gamma)_{ig} + \beta_{j(g)} + \epsilon_{ijg} \tag{2}$$

where μ is the overall mean, τ_i and γ_g are the main effects of factor C and the grouping variable, $(\tau\gamma)_{ig}$ represents factor $C \times group$ interaction, and $\beta_{j(g)}$ is the effect of factor R, nested within group. The ACMIF test is based on the largest F-ratio corresponding to $H_0 : (\tau\gamma)_{ig} = 0$ among all assignments of R factor levels into two non-empty groups.

Characterizing hidden additivity

Objects in the *HiddenF* class can be used to characterize hidden additivity further with application of the generic `plot()`, `print()`, `summary()`, and `anova()` functions. The following code produces an enhanced interaction plot to help visualize hidden additivity (Figure 3):

```
> cjejuni.out <-HiddenF(cjejuni.mtx)
> plot(cjejuni.out, main="Hidden Additivity of time effects across plants",
+ rfactor="plant", cfactor="year", colorvec=c("blue", "red"), lwd=3, legendx=TRUE)
```

The enhanced interaction plot in Figure 3 displays levels of factor C (year) on the horizontal axis, levels of factor R (plant) using lines that are visually distinguished by line type, and the outcome values on the vertical axis. Color is used to display which levels of factor R are assigned to which groups based on the maximal interaction F statistic among all possible configurations and (2). Note that the grouping colors appear whether or not the ACMIF p-value (test for $C \times group$) to detect hidden additivity is significant. The default colors are black and red for the two groups, but they can be customized by supplying an argument called `colorvec` which is a vector of length two giving color names. Another optional argument, `legendx`, may be set to `TRUE` in order to provide a legend whose location will be determined according to where on the plot the user clicks.

Because of the ellipsis(...), arguments such as `main` (for a title) or `lwd` (for specifying line widths) can be passed to `matplot` or `legend`. The arguments `lty`, `type`, `ylab`, and `xlab` are utilized by the `plot.HiddenF()` function to produce the enhanced interaction plot and therefore are not available for customization by the user.

The `print()` function returns results from the ACMIF test.

```
> print(cjejuni.out)
The ACMIF test for the hidden additivity form of interaction
F=8.965 p-value =0.03309 df=4,8
(Bonferroni-adjusted for all 7 possible configurations)
```

This output can also be obtained by typing the name of an object of class *HiddenF* into the console. The p-value above has been adjusted for multiplicity using the Bonferroni multiplier of $2^{4-1} - 1 = 7$. Additionally, the `method` argument in the `print.HiddenF()` function supports all of the methods supported by the **hiddenf** package discussed later. To see which of these seven possible configurations of rows in two non-empty groups leads to the greatest additivity within groups, use the generic `summary()` function, which returns information useful for analysis based on (2):
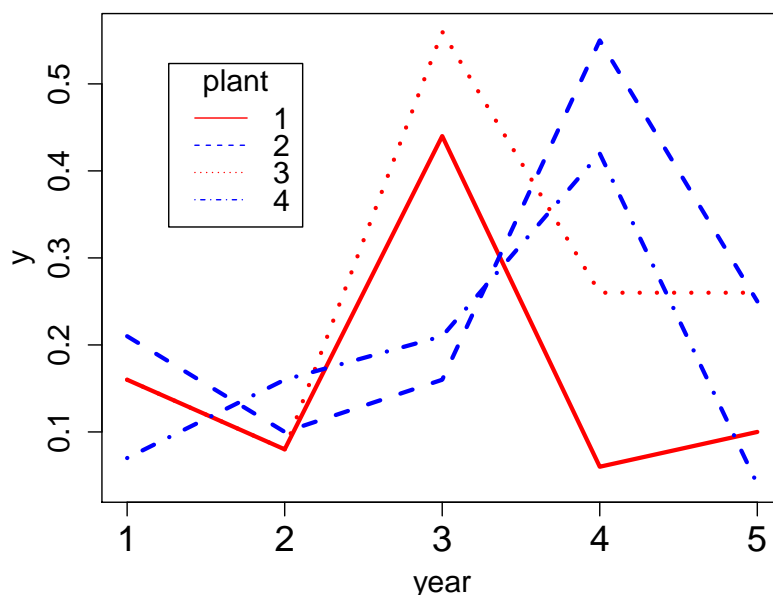
Hidden additivity – time effects across plants

Figure 3: Hidden additivity plot for the `cjejuni.mtx` data. Lines are color-coded to correspond to the configuration that achieves the maximal $C \times group$ interaction F statistic. Individual rows are distinguished by line type to allow the user to easily identify specific contributions of rows to hidden additivity .

```
> summary(cjejuni.out)
Number of configurations: 7
Minimum adjusted pvalue: 0.03308869

Rows in group 1: 1 3
Rows in group 2: 2 4

Column means for grp 1: 0.16 0.08 0.5 0.16 0.18
Column means for grp 2: 0.14 0.13 0.185 0.485 0.145
```

The output above includes the number of configurations under consideration, the smallest Bonferroni adjusted p-value among these, the identity of the rows that fall into each group and the means across levels of C for both groups. Each of these items is returned in a list (not shown). To produce an ANOVA table:

```
> anova(cjejuni.out)
The ACMIF test for the hidden additivity form of interaction
Analysis of Variance Table

Response: y
          Df  Sum Sq  Mean Sq F value   Pr(>F)
group      1 0.00000 0.000005  0.0009 0.977350
col        4 0.18753 0.046882  8.0450 0.006606
row        2 0.03673 0.018365  3.1514 0.097874
group:col  4 0.20897 0.052243  8.9648 0.004727
Residuals  8 0.04662 0.005828
C.Total   19 0.47986
(Pvalues in ANOVA table are NOT corrected for multiplicity.)
```

For these data, the interaction between time (`col`) and plant group (`group`) accounts for more variability (43%) than any other term in the hidden additivity model. This partial coefficient of determination is computed by dividing the group $\times$ column sums of squares by the total sum of squares displayed above. The anova() function can also accommodate other tests for non-additivity by specifying the method argument as "KKSA", "Mandel" or "Tukey" (see Other Tests for Non-additivity section).

Levels of factor *C* can be grouped instead by transposing the input data matrix before applying the `HiddenF()` function. The output below suggests no statistical evidence for hidden additivity in the levels of factor *C*.

```
> cjejuni.trans.mtx<-t(cjejuni.mtx)
> cjejuni.trans.out<-HiddenF(cjejuni.trans.mtx)
> print(cjejuni.trans.out)
The ACMIF test for the hidden additivity form of interaction
F=3.63 p-value =0.8671 df=3,9
(Bonferroni-adjusted for all 15 possible configurations)
```

Note that support for more than 20 rows is not yet included due to the exponential increase in computational demand as the number of grouped levels increases. Factors with more levels than this will be accommodated in future versions of the software.

Centering

The center option in the plot command allows the user to graph the hidden additivity plot with data centered at the row means to better see the concordance of rows with regard to column effects. This is a way of de-noising the plot. Figure 4 demonstrates this feature using data from Graybill (1954). These data will be further analyzed later. Inspection of the right panel of Figure 4 suggests that the third genotype appears to give inferior yields for rows in the red group (relative to the performance of other genotypes in these blocks), but does better for rows colored black. To produce this plot:

```
> data(Graybill.mtx)
> Graybill.out <- HiddenF(Graybill.mtx)
> par(mfrow=c(1,2))
> plot(Graybill.out)
> plot(Graybill.out, center=TRUE, main="Hidden Additivity plot\ncenter=TRUE")
```


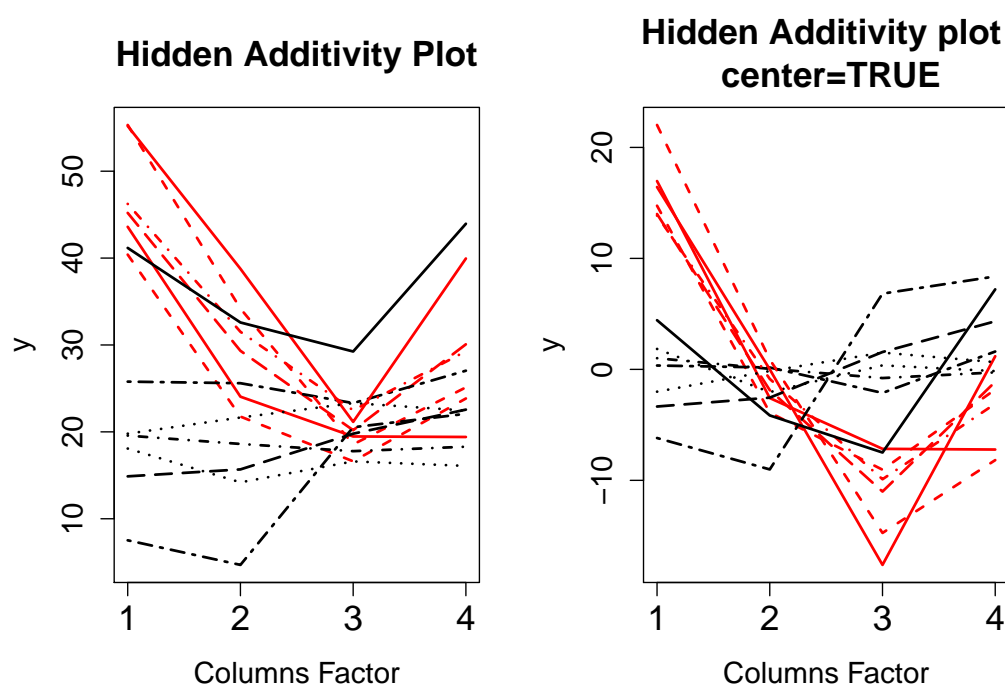**Figure 4:** Hidden additivity plots for the wheat yield data (Graybill, 1954). The left and right panels exhibit the uncentered and centered graphical options, respectively.

Other tests for non-additivity

The ACMIF approach Franck et al. (2013) to detect hidden additivity has been described above. The subsections below include code examples and briefly review the other four supported approaches to

detect non-additivity in unreplicated studies.

One degree of freedom approach

The one degree of freedom test (Tukey, 1949) is the earliest and most widely-known approach. The following model for non-additivity corresponds to Tukey's procedure:

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

A hypothesis test for this approach is based on the null hypothesis $H_0 : \nu = 0$. Under the null, (3) reduces to (1). The one degree of freedom test arises by fitting the squared fitted values from the additive model as a model term, then assessing the partial F-test for statistical significance corresponding to this term.

The `TukeyPvalue()` function accepts an object of class *HiddenF* and returns a list that contains the p-value for the one degree of freedom test and an *lm* object that contains the model that includes the squared predictions from (1) as described above. The `method` argument in the `anova.HiddenF()` function can be used to produce an ANOVA table for this test. Since the p-value is large, the output below does not provide statistical evidence for non-additivity of the form suggested in (3).

```
> TukeyPvalue(cjejuni.out)
$pvalue
[1] 0.7077391

$singledf.out

Call:
lm(formula = y ~ rows + cols + psq, data = hfobj$tall)

Coefficients:
(Intercept)         rows2         rows3         rows4         cols2         cols3
   0.095454      0.029036      0.030905      0.005445     -0.026989      0.043692
       cols4         cols5           psq
   0.044568      0.006369      1.569601
```

Rows-linear approach

An extension of the one degree of freedom test is based on the following model (Mandel, 1961):

$$y_{ij} = \mu + \tau_i + \beta_j + \theta_i \beta_j + \epsilon_{ij}, \tag{4}$$

where ϵ_{ij} are i.i.d. $N(0, \sigma^2)$. The test for rows-linear non-additivity is based on the null hypothesis $H_0 : \theta_i = 0$ for $i = 1, \ldots, t$. Under this null, (4) reduces to (1).

By setting $\mu_i = \mu + \tau_i$ and $\theta_i = (\phi_i - 1)$, the response variable y_{ij} in (4) can be represented as an intercept μ_i and slope ϕ_i that depend on effect β_j. Letting i and j index "rows" and "columns" respectively leads to the phrase "rows-linear" model, which was established later (see Alin and Kurt, 2006).

The ANOVA table sum of squares associated with the θ_i term is

$$SS_{rowlin} = \sum_i \left(\frac{\sum_j y_{ij} (\bar{y}_{.j} - \bar{y}_{..})}{\sum_j (\bar{y}_{.j} - \bar{y}_{..})^2} - 1 \right)^2 \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2. \tag{5}$$

The residual sum of squares has the following form:

$$SS_{res} = \sum_i \sum_j \left[(y_{ij} - \bar{y}_{i.}) - \frac{\sum_j y_{ij} (\bar{y}_{.j} - \bar{y}_{..})}{\sum_j (\bar{y}_{.j} - \bar{y}_{..})^2} (\bar{y}_{.j} - \bar{y}_{..}) \right]^2. \tag{6}$$

The test statistic is:

$$F_{rowlin} = \frac{SS_{rowlin} / (a - 1)}{SS_{res} / (a - 1)(b - 2)}. \tag{7}$$

The statistic F_{rowlin} has an F distribution with $(a-1)$ numerator and $(a-1)(b-2)$ denominator degrees of freedom under the null hypothesis. Construction of a columns-linear test is accomplished by defining a *HiddenF* object on the transposed data matrix and calling the `MandelPvalue()` in a similar fashion.

The `MandelPvalue()` function accepts an object of class *HiddenF* and returns a list that includes a p-value, sums of squares, F ratio, and degrees of freedom for the rows-linear test. The output below fails to detect interaction using Mandel's rows-linear test:

```
> MandelPvalue(cjejuni.out)
$pvalue
[1] 0.9458807

$SumSq
      SSRow        SSCol    SSMandel          SSE        SSTot
0.036735000 0.187530000 0.009849526 0.245740474 0.479855000

$Fratio
[1] 0.120243

$df
[1] 3 9
```

Error mean square subtable comparison approach

The error mean square (MSE) subtable comparison approach (Kharrati-Kopaei and Sadooghi-Alvandi, 2007) forms an F statistic that is the ratio of residual sums of squares that arise from rows being placed into two groups, with each group containing at least two rows,

$$F_{KKSA} = \frac{(r_2 - 1)SS_{E1}}{(r_1 - 1)SS_{E2}}. \tag{8}$$

Under the assumptions of additivity and homogeneous variance, F_{KKSA} has an F distribution with $(r_1 - 1)(c - 1)$ numerator and $(r_2 - 1)(c - 1)$ denominator degrees of freedom under the null hypothesis, no matter which rows are placed into which group. This approach suggests non-additivity when error mean square values within the subtables are discrepant.

As with the ACMIF procedure, grouping is subjective and is accomplished by choosing one of the two factors upon which to form the groups, and then placing that factor's levels into two groups. Different groupings result in different p-values. Groups are typically chosen based on *a priori* knowledge, inspection of an interaction plot, or by screening several candidate groupings and applying multiplicity adjustment to the resulting p-values. The **hiddenf** package adopts the authors' (Kharrati-Kopaei and Sadooghi-Alvandi, 2007) suggestion of assessing the $2^{r-1} - r - 1$ possible groupings for factor R and Bonferroni adjusting the resulting p-values to achieve an α level test.

The `KKSAPvalue()` function accepts an object of class *HiddenF* and returns a list that contains the maximal F statistic among configurations, a Bonferroni adjusted p-value, a length r vector that indicates the groupings of levels of factor R that correspond to the calculated p-value, and numerator and denominator degrees of freedom for the test statistic. Data may be transposed in order to form subtables according to factor C. At significance level $\alpha = .05$ the output below does not reject the null hypothesis of additivity according to the MSE subtable comparison approach when grouping levels of factor R.

```
> t(KKSAPvalue(cjejuni.out))
          fmax   pvalue  grp.vector   NumDf DenomDf
[1,]  1.748821       1    Numeric,4       4       4
```

Residual clustering approach

A recent approach (Malik et al., 2014) considers non-additivity elicited due to cell values which are remote relative to the additive model. Not all cells are assumed to contribute to the interaction pattern. The method detects non-additivity by exploiting the large residuals that arise under this structure.

Execution of the residual clustering method proceeds as follows. First, residuals from (1) $\hat{r}_{ij(1)} = y_{ij} - \hat{y}_{ij(1)}$ are placed into 3 groups using k-means clustering (MacQueen, 1967) in one dimension. Then, a two degrees of freedom cluster effect is incorporated into the additive model:

$$y_{ij} = \mu + \tau_i + \beta_j + \kappa_{k(ij)} + \epsilon_{ij} \tag{9}$$

where $\kappa_{k(ij)}$ represents the effect of the *kth* cluster, $k = 1, 2, 3$. The subscript $k(ij)$ indicates that the *ijth* residual is assigned to exactly one of the k clusters for $i = 1, \ldots, c$ and $j = 1, \ldots, r$. Since the cluster term is suggested by the data, the partial *F* test corresponding to the cluster term does *not* exhibit a central *F* distribution under the null. Rather, the null distribution for this statistic may be approximated via Monte Carlo simulation for the purpose of conducting statistical inference. Critical thresholds for a variety of c and r are available (Malik et al., 2014). The **hiddenf** package computes p-values and critical thresholds for user-supplied c and r using the functions MalikPvalue and MalikTab, respectively. By default $N = 500$ Monte Carlo replicates are used to approximate the p-value, which guarantees the standard error of the p-value $SE(p - value) \leq 0.023$.

The MalikPvalue function accepts an object of class *HiddenF* and returns a Monte Carlo p-value, observed test statistic, and Monte Carlo sample size used in the computation. The k-means approach uses the method of Hartigan and Wong (1979) with 100 starting values per Monte Carlo iteration. The output below suggests that the *C. jejuni* data do not exhibit non-additivity of the form suggested by this method. This approach is invariant to data transposition.

```
> t(MalikPvalue(cjejuni.out, N=10000))
(Pvalue from Malik's test estimated with N=10000 Monte Carlo datasets)
      pvalue    Tc      N
[1,]  0.8498 40.97983 10000
```

The *MalikTab* function accepts as input a user-specified r and c, and returns a Monte Carlo estimate of the 90%, 95%, and 99% for the null distribution of the test statistic. A variety of such thresholds are provided in Malik et al. (2014). The default number of Monte Carlo replicates is $N = 1,000$. An example is given below.

```
> Mtab.24.6<-MalikTab(r=24, c=6, N=10000)
> ls(Mtab.24.6)
[1] "q"      "Tcsim"
> Mtab.24.6\$q
        r        c       99%      95%      90%
 24.0000   6.0000 445.5725 404.3399 384.0591
```

The additivityPvalues **function**

The additivityPvalues function accepts an object of class *HiddenF* and returns p-values for each of the supported methods.

```
> t(additivityPvalues(cjejuni.out))
(Pvalue from Malik's test estimated with N=500 Monte Carlo datasets)
      Malik.pvalue Mandel.pvalue Tukey.pvalue KKSA.pvalue ACMIF.pvalue
[1,]    0.834        0.9459        0.7077          1         0.0331
```

Data examples

Table 1 summarizes an investigation of 17 data sets using the methods supported by **hiddenf**. These examples are taken from bioinformatic, agricultural, industrial, and medical settings, and are all publicly available. To our knowledge, this is the largest collection and description of data that has been used to study non-additivity in the literature to date. Table 1 includes columns for references, labels, and p-values. Note that ACMIF c, columns-linear, and KKSA c are obtained by submitting the transposed data matrix to HiddenF(). A brief description of each data set follows.

The "Liming" data arises from an experiment that explores the utility of seven types of blast furnace slags as liming material in agriculture on three types of soil (Carter et al., 1951). The data were analyzed in the context of non-additivity in Johnson and Graybill (1972) and Kharrati-Kopaei and Sadooghi-Alvandi (2007). The outcome variable is yields of corn in bushels per acre.

The "Penicillin" data set (Davies and Goldsmith, 1972) assesses variability between six samples of penicillin on 24 plates. The outcome variable is the diameter of the zone of inhibition. The data were analyzed for non-additivity in Malik et al. (2014).

The "Osmotic" data set (Kharrati-Kopaei and Sadooghi-Alvandi, 2007) explores five varieties of safflower grown in solutions with six osmotic potentials. Safflowers are an important crop due to their oil which can be used for cooking and their flowers which can be used as a substitute for saffron. The outcome variable is average root weight.

The "Fertilizer" example originally appeared in Ostle (1963) and was re-analyzed to detect non-additivity in Kharrati-Kopaei and Sadooghi-Alvandi (2007). This example explores the ratio of dry to wet wheat in four blocks for four fertilizers.

The "Bottles" example was originally analyzed in Ott and Snee (1973) and also explored in Boik (1993b) and Kharrati-Kopaei and Sadooghi-Alvandi (2007). The study assesses the performance of a six-headed machine on five occasions.

The "Grain Yields" data measures the yield of a grain crop in bushels per acre for five levels of fertilizer on five blocks. The data come from Ostle (1963) and are re-analyzed in Kharrati-Kopaei and Sadooghi-Alvandi (2007).

The "Absorbance" data (Mandel, 1991) measures absorbancy of wood pulp of nine polysachharide concentrations from seven different laboratories. The data are also analyzed in Alin and Kurt (2006).

The "Permeability" example examines permeability of three sheets of building material on nine different days. The data appear in Hald (1952) and Giesbrecht and Gumpertz (2004).

The "Wool" data (Lentner and Bishop, 1993) examines four cleaning processes for wool from five different batches. The outcome is losses in weight in *mg* of the sample after cleaning.

The "Red Blood Cells" data measures the number of red blood cells counted by five doctors in ten counting chambers. The data appear in Biggs and Macmillan (1948) and were also analyzed in Boik (1993a).

The "Tukey" data set is an illustrative example that includes three rows and four columns from Tukey (1949).

The "Ethyl Alcohol" data (Osborne et al., 1913) measures the density of aqueous solutions of ethyl alcohol at six concentrations and seven temperatures. These data were also analyzed in Mandel (1971).

The "Insecticide" data (Ott and Longnecker, 2001) comes from a complete block design that measures the impact of four plots and three insecticides on the number of string bean seedlings that emerge.

The "Rubber" data (Mandel, 1961) contains data on stress measurements in Kg/cm^2 on seven types of natural rubber vulcanizates from 11 laboratories.

The "*C. jejuni*" data measures the fraction of bacteria found on disease-resistant turkeys (Qiu, 2013). These data were collected over a five year period across four turkey plants in North Carolina. These data are displayed in Figures 1 and 3.

The "Wheat" data (Graybill, 1954) measures wheat yields in bushels per acre in four varieties in 13 locations. These data are plotted in Figure 4.

The eight Copy number variation data sets "CNV1-CNV8" compare discrepancies in the copy number signal between normal and tumor tissue samples from a comparative genomic hybridization array. Each data set corresponds to a separate location in the genome that is tested with the assay. Six dogs each had two tissue samples (one normal and one tumor) upon which the assay was conducted. The eight specific sets were selected from 5899 sets that were analyzed in a previous study and because they showed evidence of hidden additivity (Franck et al., 2013). The hidden additivity exhibited in these copy number responses might suggest the existence of multiple tumor sub-types for lymphoma in

dogs. The full data for this example may be accessed here: http://www.sciencedirect.com/science/article/pii/S0167947313001618.

The results of non-additivity tests using methods supported by the **hiddenf** package are reported in Table 1. The data come from 17 sources, and there are 24 total sets with the copy number variation contributing eight individual sets. This collection is not intended as a representative sample of the larger class of all unreplicated two-factor data, since many sets were included due to their apparent non-additivity. This exercise is intended to show the breadth of applications for which the study of non-additivity is important. To facilitate discussion, the standard p-value less than $\alpha = 0.05$ criterion is used to declare statistical significance without further multiplicity correction, although the reader is invited to interpret the raw p-values however they like. The ACMIF test for hidden additivity was significant for 16 of 24 sets when grouping levels of factor R. The test was significant for eight of 16 sets when grouping levels of factor C. The Tukey test was significant for eight of 24 sets. The rows-linear and columns-linear tests showed significant non-additivity for seven and six sets (out of 16 possible), respectively. The KKSA test grouping levels of R and C revealed non-additivity for 13 of 22 and five of 13 sets, respectively. The Malik approach yielded significant non-additivity for nine of 24 sets. Since all tests assume different restrictions on the form of interaction, no single test is optimal for every conceivable pattern. Hence, differential performance among methods is expected for different types of data. Each test excels at detecting different patterns of non-additivity.

Summary

This paper describes the main functionality of the **hiddenf** package. **hiddenf** provides descriptive and inferential tools to visualize and characterize hidden additivity. Further, **hiddenf** includes the ability to compute p-values for five tests for non-additivity. The package is illustrated using seventeen data sets spanning studies in industrial applications, agriculture, and the medical sciences. Non-additivity is evident in many of these data sets, motivating the importance of statistical interaction in unreplicated studies across a variety of scientific domains. The **hiddenf** package contributes to existing software resources specifically by making three recent tests for non-additivity available in an open-source repository for the first time (in addition to two historical methods) and by providing visualization and descriptive tools to further explore hidden additivity.

Citation	Label	r	c	ACMIF r	ACMIF c	Tukey	row-linear	col-linear	KKSA r	KKSA c	Malik
Carter et al. (1951)	Liming	7	3	0.1993	0.0110	0.9106	0.9045	0.8604	0.0068	‡	0.2471
Davies and Goldsmith (1972)	Penicillin	6	24	0.0387	◇	0.5382	0.9374	◇	1.000	◇	0.2231
Kharrati-Kopaei and Sadooghi-Alvandi (2007)	Osmotic Bars	6	5	0.2075	0.0388	0.4450	0.5300	0.1848	0.1859	0.4155	0.7453
Ostle (1963)	Fertilizer	4	4	0.0107	0.0043	0.0012	0.0146	0.0077	0.0339	0.0493	0.4373
Ott and Snee (1973)	Bottles	6	5	0.0001	0.0031	0.0003	0.0025	0.0006	0.0178	0.4925	0.8839
Ostle (1963)	Grain yields	5	5	0.6556	0.1485	0.1139	0.0103	0.4425	0.0528	1.0000	0.0403
Mandel (1991)	Absorbance	7	9	0.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.9972
Hald (1952)	Permeability	9	3	0.4066	0.2629	0.7844	0.7761	0.9544	0.3908	‡	0.8674
Lentner and Bishop (1993)	Wool	4	5	0.2355	0.4087	0.0359	0.0519	0.0410	0.7155	0.1998	0.4182
Biggs and Macmillan (1948)	Red blood cells	5	10	0.3291	0.0825	0.2322	0.6790	0.4903	0.0668	1.0000	0.8702
Tukey (1949)	Tukey	3	4	0.2213	0.5773	0.8566	0.8743	0.9298	‡	0.5806	0.4769
Osborne et al. (1913)	Ethyl Alcohol	6	7	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.0469	0.9954
Ott and Longnecker (2001)	Insecticide	3	4	0.4038	0.5569	0.6875	0.3846	0.8125	‡	0.3000	0.7603
Mandel (1961)	Rubber	11	7	<.0001	<.0001	0.3465	<.0001	0.5403	0.0021	0.0058	0.9787
Qiu (2013)	*C. jejuni*	4	5	0.0331	0.8671	0.7077	0.9459	0.8842	1.0000	0.2862	0.8443
Graybill (1954)	Wheat	13	4	<.0001	0.0070	<.0001	0.0003	<.0001	0.0320	0.0143	0.1335
Franck et al. (2013)	CNV1	6	2	0.0007	‡	0.0138	‡	‡	0.0310	‡	0.0009
	CNV2	6	2	0.0005	‡	0.0659	‡	‡	0.0010	‡	0.0164
	CNV3	6	2	<.0001	‡	0.4674	‡	‡	0.0017	‡	0.0018
	CNV4	6	2	0.0005	‡	0.1659	‡	‡	0.1249	‡	0.0182
	CNV5	6	2	0.0005	‡	0.2388	‡	‡	0.0259	‡	0.0173
	CNV6	6	2	0.0007	‡	0.6187	‡	‡	0.2799	‡	0.0224
	CNV7	6	2	0.0006	‡	0.0277	‡	‡	0.0939	‡	0.0193
	CNV8	6	2	0.0007	‡	0.1391	‡	‡	0.0340	‡	0.0216

Table 1: P-values for various tests of non-additivity supported by **hiddenf**. ‡: r and/or c insufficient for analysis. ◇: grouping on $c > 20$ for Penicillin not currently available in **hiddenf**. The Malik p-values employ N=100,000 Monte Carlo replicates.

Acknowledgements

Bibliography

A. Alin and S. Kurt. Testing non-additivity (interaction) in two-way ANOVA tables with no replication. *Statistical Methods in Medical Research*, 15:63–85, 2006. [p159, 165, 168]

F. Anscombe and J. Tukey. The examination and analysis of residuals. *Technometrics*, 5:141–160, 1963. [p159]

R. Biggs and R. Macmillan. The error of the red cell count. *Journal of Clinical Pathology*, 1:288–291, 1948. [p168, 170]

R. Boik. Testing additivity in two-way classifications with no replications: the locally best invariant test. *Journal of Applied Statistics*, 20:41–55, 1993a. [p159, 160, 168]

R. Boik. A comparison of three invariant tests of additivity in two-way classifications with no replications. *Computational Statistics and Data Analysis*, 15:411–424, 1993b. [p168]

O. Carter, B. Collier, and F. Davis. Blast furnace slags as agricultural liming materials. *Agronomy*, 43: 430–433, 1951. [p167, 170]

O. Davies and P. Goldsmith. *Statistical Methods in Research and Production*. Longman, 4th edition, 1972. [p168, 170]

C. Franck, D. Nielsen, and J. Osborne. A method for detecting hidden additivity in two-factor unreplicated experiments. *Computational Statistics & Data Analysis*, 67:95–104, 2013. [p159, 160, 162, 164, 168, 170]

F. Giesbrecht and M. Gumpertz. *Planning, Construction, and Statistical Analysis of Comparative Experiments*. Probability and Statistics. Wiley, 2004. [p168]

F. Graybill. Variance heterogeneity in a randomized block design. *Biometrics*, 10:516–520, 1954. [p164, 168, 170]

A. Hald. *Statistical Theory with Engineering Applications*. Statistics. Wiley Publications, 1952. [p168, 170]

J. Hartigan and M. Wong. A k-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979. [p167]

C. Hirotsu. An approach to defining the pattern of interaction effects in a two-way layout. *Annals of the Institute of Statistical Mathematics*, 35:77–90, 1982. [p159]

D. Johnson and F. Graybill. An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association*, 67:862–868, 1972. [p159, 160, 167]

M. Kharrati-Kopaei and S. Sadooghi-Alvandi. A new method for testing interaction in unreplicated two-way analysis of variance. *Communications in Statistics-Theory and Methods*, 36:2787–2803, 2007. [p159, 160, 166, 167, 168, 170]

M. Lentner and T. Bishop. *Experimental Design and Analysis*. Valley Book Company, second edition, 1993. [p168, 170]

J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967. [p167]

W. Malik, J. Mohring, and H. Piepho. A clustering-based test for non-additivity in an unreplicated two-way layout. *Communications in Statistics - Simulation and Computation*, 1:1–24, 2014. [p159, 160, 166, 167, 168]

J. Mandel. Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association*, 56:878–888, 1961. [p159, 160, 165, 168, 170]

J. Mandel. A new analysis of variance model for non-additive data. *Technometrics*, 13:1–18, 1971. [p159, 168]

J. Mandel. *Evaluation and Control of Measurements*. Quality and Reliability Series. Marcel Dekker, 1991. [p168, 170]

N. Osborne, E. McKelvy, and H. Bearce. Density and thermal expansion of ethyl alcohol and its mixtures with water. *Bulletin of the Bureau of Standards*, 9, 1913. [p168, 170]

B. Ostle. *Statistics in Research, Basic Concepts and Techniques for Research Works*. The Iowa State University Press, 2nd edition, 1963. [p168, 170]

E. Ott and R. Snee. Identifying useful differences in a multiple-head machine. *JOURNAL OF QUALITY TECHNOLOGY*, 5:47–57, 1973. [p168, 170]

R. Ott and M. Longnecker. *An Introduction to Statistical Methods and Data Analysis*. Brooks/Cole, check edition, 2001. [p168, 170]

Y. Qiu. Antimicrobial susceptibility and clonal population structure of multidrug resistant *campylobacter jejuni* isolates from commercial turkeys in North Carolina. Master's thesis, North Carolina State University, 2013. [p159, 168, 170]

P. Simecek and M. Simeckova. Modification of tukey's additivity test. *Journal of Statistical Planning and Inference*, 143:1, 2012. [p160]

J. Tukey. One degree of freedom for non-additivity. *Biometrics*, 5:232–242, 1949. [p159, 160, 165, 168, 170]

F. Tusell. Testing for interaction in two-way ANOVA tables with no replication. *Computational Statistics and Data Analysis*, 10:29–45, 1990. [p159, 160]

Christopher Franck
Virginia Tech Department of Statistics
403 E Hutcheson Hall Blacksburg, VA 24061
United States of America
chfranck@vt.edu

Jason Osborne
North Carolina State University Department of Statistics
5238 SAS Hall Raleigh, NC 27695
United States of America
jason.osborne@ncsu.edu