# Fractional Factorial Designs in the Analysis of Categorical Data

Alexander von Eye
Michigan State University

Author footnote:
Address correspondence concerning this article to Alexander von Eye, Michigan State University, 316 Psychology Building, East Lansing, MI 48864 - 1116; e-mail: voneye@msu.edu. I am indebted to Richard P. DeShon, G. Anne Bogat, and Johannes Ledolter for helpful comments on earlier versions of this article.

Fractional Factorial Designs in the Analysis of Categorical Data

abstract

In this article, first, fractional factorial designs are reviewed, with an emphasis on Box - Hunter designs. Based on the Sparsity of Effects Principle, it is argued that, when higher order effects are indeed unimportant in most hypotheses, fractional designs can be used without loss of information. Fractional factorial designs can be specified such that the desired level of interactions can be interpreted. Given a fixed amount of time and money, these designs allow one to include more factors in a study than completely crossed designs, or to increase the sample. It is then shown that, when the outcome variables are categorical, the same principles apply. It is also shown that higher order effects need to be specified when the outcome variables are categorical. Parameter interpretation is illustrated for a selection of fractional factorial designs. The application of fractional factorial designs is illustrated in both explanatory and exploratory contexts. In addition, a data set is analyzed using the fractional and the complete information. It is shown that distortion from only using the fractional information can be minimal.

Key words: fractional factorial design; categorical outcome variables; log-linear modeling; Configural Frequency Analysis; parameter interpretability

Fractional Factorial Designs in the Analysis of Categorical Data

Routine designs in experimental research with metric outcome data involve completely crossed factors. Similarly, the analysis of categorical outcome data typically uses a complete cross-classification of all factors or categorical independent variables. In either case, completely crossing all factors can considerably limit the number of factors that can be analyzed simultaneously. Consider, for example, 10 dichotomous factors. The complete crossing of these factors has 1,024 cells or treatment combinations. The analysis of these cells involves interactions up to the $10^{th}$ order.

There are three major issues with designs in which all factors or categorical variables are completely crossed. The first issue involves cost and effort. Creating large numbers of treatment combinations is complex and cost-intensive. The second issue concerns interpretation. Interactions of very high order are hard to interpret. In addition, theories in the social and behavioral sciences rarely imply hypotheses that can be tested only using interactions of high order. Third, interactions of high order often fail to be significant and, even if they are significant, explain only small portions of variance. Related to the third issue is the *sparsity of effects principle* that is discussed in the contexts of linear models and design (e.g., Hamada & Wu, 1992; Kutner, Neter, Nachtsheim, & Li, 2004; Wu & Hamada, 2000). According to this principle, responses in most systems are driven largely by a limited number of main effects and lower-order interactions. Higher-order interactions are, therefore, usually relatively unimportant.

Because of the above issues and the sparsity of effects principle, full factorial designs are often not only cost-intensive and wasteful when many factors are taken into account. They also yield little information above and beyond designs that allow one to only consider main effects and lower order interactions[1]. Consider, for example, the cross-classification of six dichotomous factors. The analysis of this design comes with 1 df for the intercept, 6 df for the main effects, 15 df for the two-way interactions, 20 df for the three-way interactions, 15 df for the four-way interactions, 6 df for the five-way interactions, and 1 df for the six-way interaction. Now, suppose that only the intercept, the main effects, and the first order interactions are of interest. In this case, two thirds of the degrees of freedom in this design are used to estimate parameters that are not of interest and will not be

[1]It is important to note that interaction effects can still be used for a better error estimate. So, they are not a complete loss.

interpreted.

*Fractional factorial designs* use only a subset of the treatment combinations, or cells, of a completely crossed design. This subset can be chosen based on the sparsity of effects principle. Specifically, fractional designs allow the data analyst to estimate the effects of interest. Based on the sparsity of effects principle, these effects are assumed to be of low order. In fractional factorial designs, higher order effects are either not estimable or confounded. More detail follows below.

In this article, we discuss the use of fractional factorial designs in the context of categorical data analysis. Compared to the growing body of literature of fractional factorial designs for metric outcome variables, the literature for such designs for categorical outcome variables is sparse, with a focus on the two parameter single logistic regression model (cf. Salem, Rekab, & Whittaker, 2004). In this article, we show how fractional factorial designs can be applied when the outcome variables are categorical. We consider both explanatory and exploratory research. We argue that the selection of a design is guided by (i) decisions concerning the effects that are of interest and (ii) the research strategy. In the next section, we review concepts of fractional factorial designs. We focus on Box - Hunter designs (Box, Hunter, & Hunter, 2005; Wu & Hamada, 2000), because they allow one to specify designs based on the order of interactions that are of importance in a study. We discuss these designs from the perspective of parameter interpretation. We then show that fractional factorial designs can fruitfully be applied when the outcome variables are categorical, specifically using explanatory log-linear modeling and exploratory Configural Frequency Analysis.

## 1. Fractional Factorial Designs

In this section, we first review the interpretation of parameters in the context of the Generalized Linear Model. We then proceed to discussing fractional factorial designs.

### 1.1 Parameter Interpretation in the Context of the Generalized Linear Model

In this article, we focus on designs that can be analyzed using methods from the family of Generalized Linear Models (Nelder & Wedderburn, 1972). The best known members of this family are the General Linear Model (GLM; with members ANOVA and regression analysis; see Kutner et al., 2004) and the General Log-Linear Model (GLLM; Bishop, Fienberg, & Holland, 1975; cf.

Agresti, 2002; Vermunt, 2005).

The Generalized Linear Model contains three model components (cf., Agresti, 2002; Skrondal & Rabe-Hesketh, 2004). The first component is a distribution function from the exponential family. The second component is a linear predictor. Specifically, the Generalized Linear Model relates, in its systematic component, a vector, $\eta$, to the explanatory variables through a linear relationship, or $\eta = X\beta$, where $X$ is the design matrix and $\beta$ is the parameter vector. The third element is the *link function*, $g$, such that the mean of the outcome variable, $Y$, $E(Y) = \mu$, is linked to $\eta$ by $\eta = g(\mu)$. We obtain

$$g(\mu) = X\beta .$$

In the GLM, the link function is the identity function, or $g(\mu) = \mu$. In the GLLM, the link function is the natural logarithm, $\ln(\mu)$. The mean in the GLM is typically a score on a metric scale. In the GLLM, the mean is a probability. The design matrix, $X$, contains coding vectors that represent the contrasts (effects) of interest. The design matrix can be created based on the same principles in both models (Evers & Namboodiri, 1979; Mair & von Eye, 2007; von Eye, 1988).

In the remainder of this article, we focus on the design matrix, $X$. In the GLM, the OLS estimator of the $\beta$ vector is known to be $\beta = (X'X)^{-1}X'\mu$. In the GLLM, the relationship between the parameters and the design matrix is $\lambda = (X'X)^{-1}X' \ln m$, where we use $\lambda$ instead of $\beta$ only to indicate that we are talking about the GLLM, and where $\ln m$ is the vector of the logarithms of the cell frequencies.

A first difference between the GLM and the GLLM is that, in the GLM, the $\beta$ vector is actually estimated using the above equation. In contrast, in the GLLM, this equation only represents the relationship between the design matrix, the vector $\ln m$, and the parameters (Bock, 1975; Mair & von Eye, 2007; Rindskopf, 1990; von Eye, Schuster, & Rogers, 1998). Estimation is typically performed using ML methods. A second difference between the two models is that, on the outcome side, the GLM processes observed, metric scores, whereas the GLLM typically processes frequencies. Still, parameter interpretation in the GLM is parallel to the interpretation in the GLLM. A third difference between the GLM and the GLLM is related to the second, and important for the application of fractional factorial designs in the analysis of categorical variables. Because the GLM analyzes metric outcome variables, main effects already explain the relationships in pairs of variables. These are predictor and outcome variables. Two-way interactions qualify these main effects with respect to a second predictor, and so forth. In contrast, main effects in the GLLM describe the marginal frequencies of individual variables, regardless of whether they are predictor or criterion variables. Two-way interactions describe relationships in pairs of variables. Therefore, the GLLM needs two-way interactions to examine predictor-criterion relationships that are discussed in the form of main effects in the GLM. This applies accordingly to higher order interactions. In general, to discuss the same level of effect, the GLLM needs interactions of one order higher than the GLM.

As was discussed by Mair and von Eye (2007), parameter interpretation involves inspection of the meaning and the magnitude of estimates. In the present article, we focus on the meaning of parameters. Specifically, we use the ANOVA-like, formal representation of parameters (Bishop et al., 1975), and we use effects coding (Evers & Namboodiri, 1979; for discussions of effects coding in comparison to dummy coding, see, e.g., Mair & von Eye, 2007; Rindskopf, 1990).

To illustrate the interpretation of parameters in a completely crossed factorial design in the context of the GLLM, consider the three dichotomous variables, A, B, and C. Crossed, these three variables span a contingency table with eight cells. Table 1 displays the design matrix for the saturated model for this cross-classification. This model is

$$\ln m = \lambda + \lambda^{A} + \lambda^{B} + \lambda^{C} + \lambda^{AB} + \lambda^{AC} + \lambda^{BC} + \lambda^{ABC},$$

where the first $\lambda$ represents the intercept, the single-superscripted $\lambda$s represent the main effects, the doubly-superscripted $\lambda$s represent the two-way interactions, and the triple-superscripted $\lambda$ represents the three-way interaction.

The design matrix given in Table 1 has three characteristics that are important for the following discussion. First, the matrix contains the effects for a completely crossed factorial design. In the following discussion, we will encounter fractional factorial designs. Second, this design matrix represents a saturated model. The fractional factorial designs discussed in the following sections can be more parsimonious. Third, this design matrix is orthogonal in the sense that the inner products of each pair of column vectors are zero. If this is the case, the product *X'X* will be a diagonal matrix in which the diagonal elements are all equal to the number of cells, and its inverse will exist. In addition, parameter interpretation will be easy because each parameter will reflect the strength of the effect specified in the corresponding vector of the design matrix. In the following discussion, we will encounter situations in which *X'X* cannot be inverted.

Table 1: Design Matrix for Saturated Model for the 2 x 2 x 2 Cross-classification of the Variables A, B, and C

| | Main Effects | | | Interactions | | | |
|---|---|---|---|---|---|---|---|
| Intercept | A | B | C | AB | AC | BC | ABC |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | -1 | 1 | -1 | -1 | -1 |
| 1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 |
| 1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 |
| 1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 |
| 1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 |
| 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 |
| 1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 |

For the design matrix in Table 1, parameter interpretation is illustrated in Table 2 which displays the formal representations for the eight parameters from the column vectors in Table 1.

The parameters of the model in Table 1 can be interpreted as intended (for the interpretation of log-linear model parameters in terms of odds ratios, see Agresti, 2002). The sign pattern of the log-transformed frequencies are the same as indicated in the design matrix, the cells carry equal weights, and none of the cells was excluded when the parameter estimators were calculated.

Table 2: Formal Representation of the Eight Parameters of the Model in Table 1

| Parameter | Representation |
|---|---|
| Intercept | $1/8(\ln m_{111} + \ln m_{112} + \ln m_{121} + \ln m_{122} + \ln m_{211} + \ln m_{212} + \ln m_{221} + \ln m_{222})$ |
| Main Effect A | $1/8(\ln m_{111} + \ln m_{112} + \ln m_{121} + \ln m_{122} - \ln m_{211} - \ln m_{212} - \ln m_{221} - \ln m_{222})$ |
| Main Effect B | $1/8(\ln m_{111} + \ln m_{112} - \ln m_{121} - \ln m_{122} + \ln m_{211} + \ln m_{212} - \ln m_{221} - \ln m_{222})$ |
| Main Effect C | $1/8(\ln m_{111} - \ln m_{112} + \ln m_{121} - \ln m_{122} + \ln m_{211} - \ln m_{212} + \ln m_{221} - \ln m_{222})$ |
| A x B Interaction | $1/8(\ln m_{111} + \ln m_{112} - \ln m_{121} - \ln m_{122} - \ln m_{211} - \ln m_{212} + \ln m_{221} + \ln m_{222})$ |
| A x C Interaction | $1/8(\ln m_{111} - \ln m_{112} + \ln m_{121} - \ln m_{122} - \ln m_{211} + \ln m_{212} - \ln m_{221} + \ln m_{222})$ |
| B x C Interaction | $1/8(\ln m_{111} - \ln m_{112} - \ln m_{121} + \ln m_{122} + \ln m_{211} - \ln m_{212} - \ln m_{221} + \ln m_{222})$ |
| A x B x C Interaction | $1/8(\ln m_{111} - \ln m_{112} - \ln m_{121} + \ln m_{122} - \ln m_{211} + \ln m_{212} + \ln m_{221} - \ln m_{222})$ |

## 1.2 Fractional Factorial Designs

*Optimal designs* are specified with the goal of obtaining efficient parameter estimates and maximum power of statistical tests, while minimizing cost (Berger, 2005; cf. Dodge, Fedorov, & Wynn, 1988; Liski, Mandal, Shah, & Sinha, 2002; Pukelsheim, 2006). For example, optimal designs have been devised to estimate kinetic model parameters in pharmacological research (e.g., Reverte, Dirion, & Cabassud, 2006), to improve the accuracy of parameter estimates in research on the brain physiology of rats (e.g., Verotta, Petrillo, La Regina, Rocchetti, & Tavani, 1988), to maximize the information content of measured data while observing safety and operability constraints in process control research (e.g., Bruwer & MacGregor, 2005), to discriminate between two or more rival regression models in applied statistics (e.g., Atkinson & Fedorov, 1975), or to compare the probabilities from binomial data with misclassifications (Zelen & Haitovsky, 1991).

In most instances, additional criteria are set for optimal designs. Most frequently, statistical criteria are used (Pukelsheim, 2006; Stigler, 1971) as well as parsimony. Using these criteria, researchers attempt to maximize the information content of data and the precision of parameter estimates while minimizing the effort or pecuniary cost of an experiment. These criteria are optimized when the number of treatment combinations (cells) of a design (in the context of fractional factorial designs, this number is called the *number of runs*) is minimized without compromising the interpretability of the parameters of interest. Clearly, here, the sparsity of effects principle comes into play again. If higher order effects are unimportant, the damage done by designing experiments that do not allow one to estimate or interpret higher order effects is minimal. Similarly, if effects that are unimportant are confounded, the damage that is caused by the confounds is unimportant also.

*Fractional factorial designs* are sample cases of optimal designs. They include only a fraction of the cells of a completely crossed design. That is, they contain only ½, 1/4, or an even smaller portion of the cells of a completely crossed design. The earliest fractional factorial design discussed in the literature is the well known *Latin Square* (Euler, 1782). This design allows one to estimate only the main effects of the factors. The outcome variable has to be metrical. In the context of categorical variables analysis, this type of design is of lesser importance, because, as was discussed above, main effects in categorical variables designs (i) are rarely interesting, and (ii) do not allow one to describe predictor - criterion relationships. Therefore, latin squares and other designs that focus on main effects (see Kutner et al., 2004) will not be discussed in more detail in this article.

The theory of fractional designs was developed originally by Finney (1945, 1946) and Kempthorne (1947). Recent treatments include the text by Mukerjee and Wu (2006). Statistical software packages such as Minitab, Statistica, and SYSTAT contain modules that allow one to create fractional factorial designs (see also Kessels, Goos, & Vandebroek, 2006).

*Resolution*. A key characteristic of a fractional factorial design is its *resolution*, that is, the degree to which main effects and interactions can be independently estimated and interpreted. In different words, the resolution of a design indicates the order of effects that can be estimated and are not confounded with each other. Box, Hunter, and Hunter (2005) describe the hierarchy of resolution of designs for metric outcome variables as follows. For designs with *Resolution I*, no effect is independently estimable. Therefore, designs with Resolution I are not interesting. Similarly, *Resolution II* is largely useless. Main effects would be confounded with other main effects. In the analysis of metric outcome variables, the most useful fractional factorial designs have Resolution III, IV, and V. At Resolution III, main effects can be estimated, but they are confounded with two-way interactions. More interesting is often Resolution IV. At this level, main effects can be estimated, and they are not confounded with any of the two-way interactions. Two-way interactions, however, are confounded with each other.

Thus, Resolution IV designs are of interest in particular when researchers seek to determine whether 2-way interactions are important at all, without specifying which interaction in particular. It is important to note that designs at Resolution level IV will leave some of two-factor interactions unconfounded. If the researchers are interested in these interactions in particular, Resolution IV can be viable.

Moving up the resolution ladder, designs with Resolution V allow one to estimate main effects and two-way interactions independently, and neither will be confounded with each other, but possibly with higher order interactions. Three-way interactions can be estimated also. However, they are confounded. Designs with Resolution V are needed to guarantee that two-factor interactions are not confounded. Accordingly, designs with Resolution VI allow one to estimate three-way interactions such that they are unconfounded with each other, but four-way interactions are confounded with each other.

When categorical dependent variables are analyzed, one has to take into account that, in order to estimate the same effect, interactions of one order higher are needed than in the analysis of metric outcome variables. Specifically, designs with

- Resolution I, II, and III are largely non-interesting;
- Resolution level IV allows one to estimate main effects (in the form of two-way interactions) that are confounded with three-way associations;
- Resolution level V allows one to estimate main effects and two-way interactions such that particular interactions are not confounded with each other;

In general, in the analysis of categorical outcome variables, beginning with Resolution level V, effects of increasingly higher order can be estimated without confounds. Therefore, designs with resolution levels V or higher are needed for the analysis of fractional factorial designs with categorical dependent variables when predictor - criterion relationships are of interest. Application examples of such designs are given in the following sections.

Clearly, as the resolution level increases, a design becomes more complex and requires more factors and more runs. However, designs with higher resolution levels carry more information. Fractional factorial designs allow researchers to balance the need for parsimony and the desire for information by making decisions concerning the point from which higher order interactions carry no additional useful, important, variance-explaining information.

Examples of fractional factorial designs. As one can imagine, the number of fractional factorial design types is large. Here, we list just a selection of design types (for more types, see, e.g., Box et al., 2005; Wu & Hamada, 2000). Practically all of the following design types can be generated using numerical algorithms. Therefore, they are also called *computer-aided designs*.

The first type listed here includes *homogeneous fractional factorial* designs. In these designs, all factors have the same number of levels. Accordingly, *mixed-level* fractional factorial designs include factors that can have different numbers of levels.

A subtype of homogeneous fractional factorial designs is known as *Box - Hunter designs* (Box, et al., 2005). As was indicated above, these designs use only a fraction of the completely crossed design, for example, ½, 1/4, or an even smaller fraction of the total number of runs. The number of factor levels in Box - Hunter designs is 2, and the number of runs is a power of 2. If each factor has three levels, *Box - Behnken designs* (Box & Behnken, 1960) can be considered. These designs do not use those treatment combinations for which all factors assume extreme values (e.g., treatment combinations 3-3-3 or 1-1-1). Whereas Box - Hunter designs can be considered for nominal-level factors, Box - Behnken designs require factors that are scaled at least at the ordinal scale level. The number of runs in Box - Behnken designs is a multiple of 3.

*Plackett - Burman designs* (1946; Ledolter & Swersey, 2007), also called *screening designs*, operate at resolution level III. They are very economical in that the number of runs can be very small, when the dependent variable is metrical. For example, up to 11 dichotomous factors can be studied using only 12 runs; up to 19 factors can be studied using 20 runs, and up to 23 factors can be studied using 24 runs. The number of runs in Plackett - Burman designs is a multiple of the number of factor levels. These designs are used to estimate main effects. However, one has to assume that two-way interactions are absent. These designs are also called *saturated main effect designs*, because all available degrees of freedom go into the estimation of main effects. These designs are used to determine the factors that may have effects on the outcome variable.

Plackett-Burman designs will simplify to two-level fractional factorial designs if the number of runs is $2^k$. For example, for 8, 16, or 32 runs, they are the same as two-level fractional factorial designs. They are unique for 12, 20, 24, etc. runs.

To increase the resolution of Plackett - Burman designs, the use of *foldover designs* has been proposed. These designs result from reversing the signs of all scores in the design matrix, and appending the thus mirrored design to the original one. The resulting design allows one to estimate all main effects such that they are no longer confounded with two-way interactions, at the expense of doubling the number of runs.

As was noted above, a main effect model in the context of the GLM relates a predictor to a criterion. In contrast, in the GLLM, main effects allow statements about the univariate marginal distribution of a variable. To describe the relationship between a predictor and a criterion variable, a two-way interaction is needed. Therefore, standard Plackett - Burman designs are of lesser importance in the context of categorical variable analysis. To create a screening design for categorical outcome variables, resolution at level IV is needed.

*Taguchi design*s (1987) are orthogonal. Thus, they yield independent estimates of effects and minimized variances. Two-, three-, and mixed-level fractional factorial designs can be specified. Standard Taguchi designs are often large screening designs that allow one to estimate a maximum of main effects from a minimum number of runs, for metrical outcome variables. The number of factor levels does not need to be a constant. More elaborate designs have been developed (see, e.g., Bisgaard & Steinberg, 1997).

Generating fractional designs. In this section, we present an algorithmic description of how fractional factorial designs can be generated. We focus on Box - Hunter designs (see Box et al., 2005). Consider the number of variables, $p$, and the number of runs, $2^{p-k}$, where $p - k$ is the number of factors whose main effects can be coded as usual, in a completely crossed design. The main effects of the remaining $k$ factors have to be coded differently, because, in fractional factorial designs, the number of rows in the design matrix is reduced by at least 50% when compared to a completely crossed design. Then, a Box - Hunter design and the corresponding design matrix can be generated as follows (remember that all factors have 2 levels):

1. For the first $p - k$ factors, create a design matrix with main effects as in a completely crossed design with $2^{p-k}$ cells (= rows in the design matrix).
2. For Factor $p - k + i$, create the main effect as if it were the interaction among the factors in the first of the $\binom{p-k}{p-k-1} = \binom{p-k}{1}$ combinations of the first $p - k$ factor

   words, the remaining $k$ main effects are expressed in terms of the $(p - k)$-way interactions of those factors that can be coded as in $(p - k)$-factorial design. Thus, confounds will exist at least at the level of the $(p - k)$-way interactions.
3. Repeat Step 2 a total of $k$ times, until main effects are created for all $p$ factors.
4. Generate two-way interactions as in a standard ANOVA design, that is, by element-wise multiplication of vector elements from two different factors.
5. Generate three-way interactions also as in a standard ANOVA design, that is, by element-wise multiplication of vector elements from three different factors.
6. Repeat generating interactions until either the design is saturated or all unconfounded and important interactions are included in the design matrix.

It is important to realize that the number of designs that can be created this way is $p!/(p - k)!$. This number results from selecting different factors that are coded as in a completely crossed design with $p - k$ cells, and changing their order. This process is also called *randomizing the runs*. In different words, for fractional designs, alternatives often exist at the same resolution level (this aspect will we taken up again in the discussion).

## 1.3 Examples of Designs and Parameter Interpretation

In this section, we present a number of sample fractional factorial designs, with an emphasis on Box - Hunter designs. For each of the designs, we discuss the savings in the number of runs, in comparison to the corresponding completely crossed design, the resolution, and, most importantly, the interpretability of parameters. Two perspectives will be taken. In the first, we seek to create parsimonious designs, mostly based on resolution levels (Examples 1 and 3). In the second, we seek to create parsimonious designs with a specific method of analysis (logistic regression) in mind (Example 4). We also show an example in which Box - Hunter and Plackett - Burman designs coincide (Example 2).

Example 1: Box - Hunter Design with 8 Runs from 4 Factors; Resolution = III. We begin with a Box - Hunter design in which we study the four dichotomous factors A, B, C, and D. The complete cross-classification of these factors has $2^4$ = 16 cells. We decide to create a design that has 50% fewer cells, that is, eight runs. This design appears in the Intercept and Main Effect panels of Table 3.

Table 3: Box - Hunter Design with 8 Runs from 4 Factors; Resolution = III; All Interactions Included

| Int. | Main Effects | | | | 2-Way Interactions | | | | | | 3-Way Interactions | | | | 4-Way |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | A | B | C | D | AB | AC | AD | BC | BD | CD | ABC | ABD | ACD | BCD | ABCD |
| 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 |
| 1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | 1 |
| 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | 1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 | 1 | 1 | 1 |
| 1 | -1 | -1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | -1 | 1 |
| 1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 |
| 1 | -1 | 1 | 1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 | -1 | 1 | 1 | -1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

The matrix in Table 3 has eight rows (runs) and 16 columns. Thus, it is bound to be non-orthogonal. The main effect vectors are pairwise orthogonal. However, the following confounds are in the matrix:

- Interaction AD is confounded with Interaction BC: AD = BC

- Interaction AC is confounded with Interaction BD: AC = BD
- Interaction AB is confounded with Interaction CD: AB = CD
- Main effect D is confounded with Interaction ABC: D = ABC
- Main effect C is confounded with Interaction ABD: C = ABD
- Main effect B is confounded with Interaction ACD: B = ACD
- Main effect A is confounded with Interaction BCD: A = BCD; and
- The Intercept is confounded with Interaction ABCD: I = ABCD.

Because of these confounds, the application of this design requires the assumption that all three- and four-way interactions are zero. In addition, the model cannot be fitted when all two-way interactions are included because *X* would not be orthogonal, and *X'X* would have no inverse. There would be more unknowns (parameters) than equations (rows in *X*). This design has a resolution of III, that is, two-way interactions are confounded with each other. When estimating parameters, only up to three of the six two-way interactions can be included. When the outcome variable is categorical, this design is saturated. The above list of confounds shows which of the two-way interactions can be estimated so that they are not confounded. If any of the two-way interactions turns out significant, all one knows is that either this or the corresponding confounded interaction, or both, are important. Which of the two-way interactions exists, remains unknown until a design with higher resolution is used.

Parameter interpretation seems straightforward if, for instance, only the first three of the six two-way interactions are included in the model (any combination of mutually unconfounded two-way interactions can be used). The formal representation of the seven effect parameters appears in Table 4 (intercept omitted).

Although parameter interpretation seems straightforward, it must not be forgotten that the three two-way interactions that are included in the design matrix are confounded with the remaining three two-way interactions, and that analysis requires the assumption that all higher-order interactions explain only unimportant portions of the distribution in the 8-run table.

Table 4: Formal Representation of the First Seven Effect Parameters of the Model in Table 3 (Intercept Omitted)

| Parameter | Representation |
|---|---|
| Main Effect A | $1/8\ (-\ln m_{111} + \ln m_{112} - \ln m_{121} + \ln m_{122} - \ln m_{211} + \ln m_{212} - \ln m_{221} + \ln m_{222})$ |
| Main Effect B | $1/8\ (-\ln m_{111} - \ln m_{112} + \ln m_{121} + \ln m_{122} - \ln m_{211} - \ln m_{212} + \ln m_{221} + \ln m_{222})$ |
| Main Effect C | $1/8\ (-\ln m_{111} - \ln m_{112} - \ln m_{121} - \ln m_{122} + \ln m_{211} + \ln m_{212} + \ln m_{221} + \ln m_{222})$ |
| Main Effect D | $1/8\ (-\ln m_{111} + \ln m_{112} + \ln m_{121} - \ln m_{122} + \ln m_{211} - \ln m_{212} - \ln m_{221} + \ln m_{222})$ |
| AB Interaction | $1/8\ (+\ln m_{111} - \ln m_{112} - \ln m_{121} + \ln m_{122} + \ln m_{211} - \ln m_{212} - \ln m_{221} + \ln m_{222})$ |
| AC Interaction | $1/8\ (+\ln m_{111} - \ln m_{112} + \ln m_{121} - \ln m_{122} - \ln m_{211} + \ln m_{212} - \ln m_{221} + \ln m_{222})$ |
| AD Interaction | $1/8\ (+\ln m_{111} + \ln m_{112} - \ln m_{121} - \ln m_{122} - \ln m_{211} - \ln m_{212} + \ln m_{221} + \ln m_{222})$ |

Example 2: Plackett - Burman Design with 4 Runs from 3 Factors; Resolution = III. In many cases, in particular when the number of runs is small and the resolution level is the same, designs that were created using different models coincide. Consider the Plackett - Burman design with 4 factors and 4 runs (Resolution = III) in the Intercept and Main Effect panels of Table 5. This design is identical to a Box - Hunter design with 3 factors and 4 runs.

Table 5: Plackett - Burman design with 4 Runs from 3 Factors; Resolution = III; All Interactions Included

| Intercept | Main Effects | | | 2-Way Interactions | | | 3-Way |
|---|---|---|---|---|---|---|---|
| I | A | B | C | AB | AC | BC | ABC |
| 1 | 1 | 1 | -1 | 1 | -1 | -1 | -1 |
| 1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 |
| 1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 |
| 1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 |

Here again, the savings, measured in the number of runs, over the completely crossed design is 50%. The main effect vectors are mutually orthogonal. However, there are confounds with the 2- and the 3-way interactions. Specifically,

- A = -BC
- B = -AC
- C = -AB, and
- Intercept = -ABC.

Parameters for this model cannot be estimated unless confounded vectors are eliminated from the design matrix. Typically, the vectors for the interactions are taken out, reflecting the assumption that only the main effects are of interest (which is rarely the case in categorical data analysis). One has to make the assumption that none of the interactions explains important aspects of the data.

Example 3: Box - Hunter Design with 16 Runs from 5 Factors; Resolution = V. Naturally, higher levels of resolution can be achieved only with more factors. The following example presents a Box - Hunter design in which 16 runs are realized for 5 factors. This design has a resolution level of V. Table 6 displays the design matrix for the main effects.

Table 6: Main Effects in Box - Hunter Design with 16 Runs from 5 Factors; Resolution = V; Intercept Omitted

| Factor | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| -1 | -1 | -1 | -1 | 1 |
| -1 | -1 | -1 | 1 | -1 |
| -1 | -1 | 1 | -1 | -1 |
| -1 | -1 | 1 | 1 | 1 |
| -1 | 1 | -1 | -1 | -1 |
| -1 | 1 | -1 | 1 | 1 |
| -1 | 1 | 1 | -1 | 1 |
| -1 | 1 | 1 | 1 | -1 |
| 1 | -1 | -1 | -1 | -1 |
| 1 | -1 | -1 | 1 | 1 |
| 1 | -1 | 1 | -1 | 1 |
| 1 | -1 | 1 | 1 | -1 |
| 1 | 1 | -1 | -1 | 1 |
| 1 | 1 | -1 | 1 | -1 |
| 1 | 1 | 1 | -1 | -1 |
| 1 | 1 | 1 | 1 | 1 |

The confounds in this design are:

- A = BCDE
- C = ABDE
- D = ABCE
- E = ABCD
- AB = CDE

- AC = BDE
- AD = BCE
- AE = BCD
- BC = ADE
- BD = ACE
- BE = ACE
- CD = ABD
- CE = ABE
- DE = ABC, and
- Intercept = ABCDE.

This confound pattern shows again how the sparsity of effects principle can be translated into a parsimonious design. If indeed three- and four-way effects are unimportant, then this design allows one to estimate main effects and two-way interactions that are mutually independent. In addition to the vector for the intercept, the design matrix will then include only the five vectors for the main effects and the 10 vectors for the two-way interactions. When the outcome variable is categorical, this model is saturated. Only if interactions are either set equal or taken out of the model, a non-saturated model will result.

Models with a resolution level of V are of interest, when the relationships between pairs of variables are targeted. Methods of factor analysis, latent variables analysis, multidimensional scaling, cluster analysis, or correspondence analysis often start from similarity matrices that only reflect the relationships in pairs of variables. Models with a resolution level of V can also be of interest in logistic regression. This is illustrated in Example 4, below.

To illustrate the confounds, consider the researcher who first estimates the model for the current design that includes all 10 two-way interactions. This model can be estimated, and the parameters can be interpreted as indicated in the design matrix. For example, the interaction between variables D and E is estimated using the vector {-1, -1, 1, 1, 1, 1, -1, -1, 1, 1, -1, -1, -1, -1, 1, 1}. The resulting parameter has the interpretation

$$\lambda^{DE} = 1/16\,(-\ln m_{22221} - \ln m_{22212} + \ln m_{22122} + \ln m_{22111} + \ln m_{21222} + \ln m_{21211} - \ln m_{21121} - \ln m_{21112} + \ln m_{12222} + \ln m_{12211} - \ln m_{12121} - \ln m_{12112} - \ln m_{11221} - \ln m_{11212} + \ln m_{11122} + \ln m_{11111})\ .$$

In this equation, a subscript of 1 corresponds to a score of 1 in the design matrix. A subscript of 2 corresponds to a score of -1. Now, in a follow-up step, the same researcher decides to estimate the hierarchical model that only includes the two-way interactions AB, AC, AD, AE, BC, BD, and BE. The interactions CD, CE, and DE are replaced by the three-way interactions ABC, ABD, and ABE. This model can also be estimated. However, because of DE = ABC, the three-way interaction, ABC, comes with exactly the same interpretation as the substituted two-way interaction DE, and $\lambda^{ABC} = \lambda^{DE}$. This applies accordingly to $\lambda^{CD}$ and $\lambda^{DE}$ because $\lambda^{DE} = \lambda^{ABE}$ and $\lambda^{CD} = \lambda^{ABD}$. Thus, because of these confounds, nothing is gained by replacing the two-way interactions by three-way interactions. In different words, substituting, in this type of design, a two-way interaction by its confounded three-way interaction makes sense only if the assumption is entertained that the two-way interaction is zero.

Designs with resolution level V are positioned, in the analysis of categorical outcome variables, one resolution level above Plackett - Burman designs in the analysis of metric variables. Therefore, higher order interactions can be examined than with screening designs. Specifically, at resolution level V, one is able to examine all pairwise predictor - criterion relationships. Interestingly, when the categorical variables in such a design are grouped into predictors and criteria, the model is not necessarily saturated. If one assumes that the $p$ predictors are independent of each other and the $q$ criterion variables are also independent of each other, the number of interactions that need to be part of the model is $pq$. This number is always less than or equal to the number $\binom{p+q}{2}$ of interactions for the model in which the distinction between predictors and criteria is not made, and all pairwise interactions are estimated. The remaining degrees of freedom can be used to make statements about model fit, or to include covariates.

Example 4: Box - Hunter Designs for Logistic Regression. Instead of creating designs based on resolution, we now create a design based on goal of analysis. This goal determines the required resolution. Consider a logistic regression model (Agresti, 2002; von Eye, Mair, & Bogat, 2005). Standard logistic regression models make no assumptions about predictor interactions. Therefore, these models are typically saturated in the predictors, and the standard design is completely crossed (if all predictors are categorical). The models typically focus on bivariate predictor - criterion relationships. To examine these relationships, two - way interactions are estimated. Higher order interactions are often deemed unimportant. In these cases, a fractional factorial design as the one shown in Example 3 will do the job, at a savings of 50% of the cells.

To illustrate, suppose that Variable A in Example 3 is the criterion variable in a logistic regression model, and variables B, C, D, and E are the predictors. If only the predictive power of individual predictors is of interest, the logistic regression model can be cast in the form of the following hierarchical log-linear model,

$$\ln m = \lambda + \lambda^{AB} + \lambda^{AC} + \lambda^{AD} + \lambda^{AE} + \lambda^{BCDE},$$

where $m$ is the array of model frequencies, the $\lambda$ are the model parameters, and the superscripts indicate the interacting variables. All lower order terms are implied. If the three-way interactions among pairs of predictors and the criterion are also of interest, the model becomes

$$\ln m = \lambda + \lambda^{ABC} + \lambda^{ABD} + \lambda^{ABE} + \lambda^{ACD} + \lambda^{ACE} + \lambda^{ADE} + \lambda^{BCDE};$$

and if the four-way interactions among predictors and the criterion are of interest, the model becomes

$$\ln m = \lambda + \lambda^{ABCD} + \lambda^{ABCE} + \lambda^{ABDE} + \lambda^{ACDE} + \lambda^{BCDE}.$$

If the five-way interaction is included, the model becomes saturated. Based on the sparsity of effects principle, interactions become less and less interesting as their order increases. If this applies to the interactions among the predictors also, the first of these logistic regression models can be made more parsimonious by setting the four- and the three-way interactions among the predictors to zero. If (1) only the two-way interactions between predictors and the criterion are considered, and (2) the three and the four-way interactions among predictors are set to zero, we obtain the hierarchical model

$$\ln m = \lambda + \lambda^{AB} + \lambda^{AC} + \lambda^{AD} + \lambda^{AE} + \lambda^{BC} + \lambda^{BD} + \lambda^{BE} + \lambda^{CD} + \lambda^{CE} + \lambda^{DE}.$$

If (1) only the two-way interactions among predictors and the criterion are considered, and (2) only the four-way interaction is set to zero, we obtain the hierarchical model

$$\ln m = \lambda + \lambda^{AB} + \lambda^{AC} + \lambda^{AD} + \lambda^{AE} + \lambda^{BCD} + \lambda^{BCE} + \lambda^{BDE} + \lambda^{CDE} .$$

Other models can be specified in which both the order of interactions that involve predictors and criteria and the interactions among predictors are varied.

From the perspective of creating parsimonious designs, we now ask whether logistic regression parameters can be estimated using fractional factorial designs. The model in Table 6 operates at resolution level V. It thus allows one to estimate main effects and all two-way interactions such that they are not confounded with each other. Thus, if we set all three-, four-, and five-way interactions to zero, the second last of the above logistic regression models, that is, the one with only two-way interactions can be estimated using the Box - Hunter design in Table 6. To estimate the logistic regression model that includes three-way interactions, a resolution level of VI is needed. A Box - Hunter design that allows one to estimate such a model requires 6 variables and 32 runs. The completely crossed factorial design for the same six variables would require 64 runs. For this number of runs, a Box - Hunter design for seven variables with a resolution level of VII can be created, or a screening design with 11 factors with a resolution level of IV. This last design would represent a savings of 96.88% over the completely crossed design which has 2028 cells.

## 2. Fractional Factorial Designs in Explanatory and Exploratory Research

Explanatory and exploratory research differ in the degree to which explicit hypotheses exist. These two research strategies can be seen as the two poles of a spectrum. In purely explanatory research, hypotheses are derived from theory or prior results. In the multivariate case, these hypotheses typically link variables in the form of associations, predictions, or cause - effect relationships. Similarly, hypotheses may exist that group variables, occasions, or individuals. On the other pole, no hypotheses are specified that link variables, occasions, or individuals. The implications of these two research strategies for the selection of designs are quite different.

In explanatory research, a solid knowledge base exists that allows researchers to precisely specify the hypotheses under study. These hypotheses concern parameters that can be estimated using a particular study design. By implication, the remaining parameters are either not interesting, not important, or part of the exploratory component of the study. To cover the explanatory component, fractional factorial designs can be specified that allow the researchers to estimate the parameters of interest. In the typical case, interactions are interesting up to a pre-specified order. This order corresponds to a resolution level $J$ for metric outcome variables, and $J + 1$ for categorical outcome variables.

The first benefit from specifying designs based on resolution is that the data collection part of the study becomes less costly, financially as well as in units of effort and time. The second benefit is that more complex studies can be undertaken, studies that include far more factors, at costs that are at the level of far simpler studies that use completely crossed designs instead of fractional factorial designs. A third benefit is that the precision of parameter estimation will not suffer.

In *exploratory research*, various scenarios are conceivable. If researchers seek to determine whether factors have, from a bivariate perspective, effects on some outcome variable at all, screening designs can be considered. These designs require resolution levels III for metric and IV for categorical outcome variables. In different contexts, researchers aim at determining where, in a table the action is, or which variables interact. In these cases, the sparsity of effects principle is less capable of guiding the decision concerning the resolution of a design. Therefore, researchers will strive for the highest possible resolution which still may exclude the interactions of the highest order, if they are deemed unimportant a priori. However, if phenomena such as Meehl's paradox[2] (1950) cannot be excluded, fractional designs cannot be recommended. In these cases, setting higher order interactions to zero would make one miss the action in a table.

## 3. Data Examples

In this section, we present two data examples. In the first example, a model is estimated in which a cross-time association structure of Medicaid reception is estimated, and it is asked whether Medicaid reception is related to depression (see von Eye & Bogat, 2006). In the second example, an exploratory analysis is performed on the same data. In each example, we compare results from

---

[2] *Meehl's paradox* is the term for the phenomenon that the information in a table is entirely carried by the highest possible interaction. For examples, see von Eye (2002).

fractional and from completely crossed designs.

Example 1: Medicaid and Depression. In a study on the effects of social welfare on mental health in battered women, von Eye and Bogat (2006; cf. Levendosky, Bogat, Davidson, & von Eye, 2000) asked whether, longitudinally, depression is linked to social welfare as measured by food stamp and Medicaid reception. Data from 6 observation points are available. The first observation point was in the last trimester of the women's pregnancy, and the second was 3 months after birth. The second observation was performed to collect information about the child. In the following analyses, we focus on the data from the third and the following three observation points. For the following illustration of the application of fractional factorial designs in the analysis of categorical outcome variables, we use the following measures:

- Medicaid received at observation points 3, 4, 5, and 6 (M3, M4, M5, and M6; all scored as 1 = did not receive and 2 = did receive); and
- Depression at observation point 6 (D6; scored as 1 = below the cutoff for clinical-level depression and 2 = above cutoff; depression was measured using the BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961).

The data were collected in one-year intervals. Crossed, these 5 variables span a contingency table with $2^5 = 32$ cells. For the following analyses, we hypothesize that

(1) Medicaid reception predicts itself over time; and

(2) At Time 6, Medicaid reception predicts depression.

This model is depicted in Figure 1.

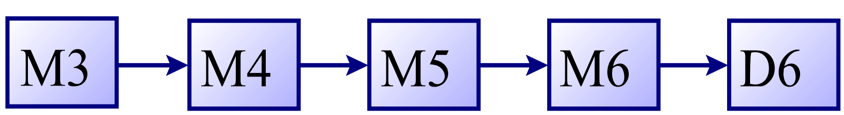


Figure 1: Medicaid and Depression

The model in Figure 1 shows that only two-way interactions are needed to test the hypothesized relationships. The hierarchical log-linear model that corresponds to this graphical model is

$$\ln m = \lambda + \lambda^{M3,\, M4} + \lambda^{M4,\, M5} + \lambda^{M5,\, M6} + \lambda^{M6,\, D6}.$$

This model can be enriched by also testing whether Medicaid reception at observation points 3, 4, and 5 are predictive of depression at Time 6. The enriched model is

$$\ln m = \lambda + \lambda^{M3,\, M4} + \lambda^{M4,\, M5} + \lambda^{M5,\, M6} + \lambda^{M6,\, D6} \\ + \lambda^{M3,\, D6} + \lambda^{M4,\, D6} + \lambda^{M5,\, D6},$$

where the new interactions are listed in the second row of the equation. The interactions tested in the enriched model will also be of an order not higher than two-way. An analysis of this model based on all 32 cells of the complete cross-classification comes with 32 degrees of freedom. Of these, 16, that is, 50%, are needed for the 3-, 5-, and 5-way interactions that are not of interest when the model depicted in Figure 1 or the enriched model are estimated. Therefore, there is no need to screen the women in all 64 cells of the design. Instead, a more parsimonious model will allow us to make a decision concerning the parameters in these models as well as about overall model fit.

A fractional factorial model that allows one to estimate all 2-way interactions so that they are not confounded with main effects or each other requires level V resolution. The Box - Hunter design given in Table 6 has these characteristics. Therefore, we employ, for the following analyses, this design. Table 7 shows the design matrix with all bivariate interactions that are part of the model in Figure 1 and the enriched model. The design has 16 runs. This represents a savings of 50% over the completely crossed design.

Based on the design matrix in Table 7, we now estimate three models. The first is the main effect model. We call it Model 1. It is used as a reference. The second, called Model 2, is the model depicted in Figure 1, and the third, called Model 3, is the enriched model. Table 8 shows the overall goodness-of-fit LR-$X^2$ values for these models, and the results of the model comparisons.

Table 7: Box - Hunter Design with 16 Runs from 5 Factors; Resolution = V; Intercept Omitted; Interactions Included for the Model in Figure 1 and the Enriched Model; Cell Frequencies in Last Column

| Main Effects | | | | | Two-Way Interactions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M3 | M4 | M5 | M6 | D6 | M3 x M4 | M4 x M5 | M5 x M6 | M6 x D6 | M3 x D6 | M4 x D6 | M5 x D6 | Freq. |
| -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | 45 |
| -1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | 1 | 0 |
| -1 | -1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | 1 | 1 | -1 | 0 |
| -1 | -1 | 1 | 1 | 1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 | 2 |
| -1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | 1 | 0 |
| -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 | 1 | -1 | 1 | -1 | 0 |
| -1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 | 1 | 1 | 1 |
| -1 | 1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | 1 |
| 1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | 1 | 1 | 0 |
| 1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 |
| 1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 2 |
| 1 | -1 | 1 | 1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 | 1 |
| 1 | 1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | 7 |
| 1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 | 0 |
| 1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | -1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 55 |

The results in Table 8 suggest that the main effect model (Model 1) does a poor job describing the data. In contrast, the model depicted in Figure 1 (Model 2) is not only significantly better than Model 1, it also describes the data very well. In fact, its LR-$X^2$ is so small that it is impossible to improve this model significantly. This is reflected by the results for the enriched model (Model 3). We thus retain Model 2. In Model 3, none of the additional parameters is significant.

These effects would have suggested that depression at Time 6 can also be predicted from Medicaid reception at the earlier years.

Table 8: Estimation and Comparison of Three Models of the Effects of Medicaid Reception on Depression; Results Based on the Fractional Factorial Design in Table 7

| Model | LR-$X^2$; *df; p* | $\Delta_1$ LR-$X^2$; *df; p* | $\Delta_2$ LR-$X^2$; *df; p* |
|---|---|---|---|
| 1 | 218.16; 10; < 0.01 | | |
| 2 | 2.37; 6; 0.88 | 215.79; 4; < 0.01 | |
| 3 | 0.08; 3; 0.99 | 218.08; 7; < 0.01 | 2.29; 3; 0.51 |

Table 9: Main Effect and Interaction Parameters for the Model in Figure 1 (Model 2)

| Parameter | Estimate | *z* | CI |
|---|---|---|---|
| M3 | -1.87 | -3.23 | -3.01 to -0.73 |
| M3 x M4 | 0.97 | 2.40 | 0.18 to 1.77 |
| M4 | 1.13 | 2.28 | 0.16 to 2.11 |
| M4 x M5 | 1.30 | 6.29 | 0.89 to 1.70 |
| M5 | -1.65 | -2.45 | -2.97 to -0.33 |
| M5 x M6 | 1.03 | 2.79 | 0.31 to 1.75 |
| M6 | 0.16 | 0.21 | -1.33 to 1.65 |
| M6 x D6 | -1.00 | -2.44 | -1.81 to -0.20 |
| D6 | 1.33 | 3.88 | 0.66 to 2.00 |

Table 9 shows that each of the interaction parameters is significant. We conclude that the model in Figure 1 describes the data well, both overall and in its individual parameters.

This conclusion is based on a fractional factorial design. Clearly, there is the temptation to check whether this design has led to a distortion of the relationships in the data. In the present example, we are unable resist this temptation because we possess the complete data matrix. This matrix is given in the Appendix. We now perform the same analyses as in Table 8 with the complete cross-classification. Table 10 shows the overall goodness-of-fit LR-$X^2$ and the results of the model comparisons.

Table 10: Estimation and Comparison of Three Models of the Effects of Medicaid Reception on Depression; Results Based on Completely Crossed Factors

| Model | LR-$X^2$; *df; p* | $\Delta_1$ LR-$X^2$; *df; p* | $\Delta_2$ LR-$X^2$; *df; p* |
|---|---|---|---|
| 1 | 292.16; 26; < 0.01 | | |
| 2 | 34.81; 22; 0.04 | 257.35; 4; < 0.01 | |
| 3 | 29.96; 19; 0.05 | 262.20; 7; < 0.01 | 4.85; 3; 0.43 |

The results in Table 10 are similar to the ones in Table 8. The main effect model does not describe the data well. The model depicted in Figure 1 is a significant improvement, and Model 3 does not improve Model 2 significantly. There is one notable difference between the solutions. Model 2 is not as close to the data when the complete cross-classification is used as when the fractional design is used. However, as before, all interaction parameters in Model 2 are significant (not shown here). In addition, the parameters for the associations between Medicaid Reception during the earlier years and depression at Time 6 are, again, not significant in Model 3. We thus conclude that the fractional design reflects the data structure well. There are no interactions higher than first order. Specifically, the null hypothesis that the three-way effects in the completely crossed table are zero comes with a LR-$X^2$ = 6.84 ($df$ = 10; $p$ = 0.74). For the four-way effects, we calculate LR-$X^2$ = 2.06 ($df$ = 5; $p$ = 0.84), and for the five-way effect, we calculate LR-$X^2$ = 0.001 ($df$ = 1; $p$ = 0.99). These results can be viewed as a sample case of the sparsity of effects principle.

Example 2: Configural Frequency Analysis of the Medicaid Data. Configural Frequency Analysis (CFA; Lienert & Krauth, 1975; von Eye, 2002; von Eye & Gutiérrez Peña, 2004) allows researchers to ask whether patterns of categorical variables, also called *configurations*, were observed more often than, less often than, or as often as expected with reference to some chance model, also called *base model*. In brief, consider the observed frequency of Cell $i$, $m_i$, and the corresponding expected frequency $\hat{m}_i$, estimated under some chance model where $i$ goes over all cells in the table. Then, CFA tests, for each cell, the null hypothesis that $E[m_i] = \hat{m}_i$. If Cell $i$ constitutes a *CFA type*, this null hypothesis is rejected because $E[m_i] > \hat{m}_i$. If Cell $i$ constitutes a *CFA antitype*, the null hypothesis is rejected because $E[m_i] < \hat{m}_i$. If $E[m_i] = \hat{m}_i$, the null hypothesis prevails, and Cell $i$ is said to constitute neither a type nor an antitype. In different words, *types* occur more frequently than expected by chance, and *antitypes* occur less frequently than expected by chance.

CFA types and antitypes can occur only if the base model is rejected. Only then, discrepancies between observed and estimated expected cell frequencies can be large enough to be significant even when the significance threshold α is protected. CFA examines either all or a selection of individuals cells in a table. It is the goal of CFA to identify and interpret those cells that contradict the base model. Therefore, model fit is not aimed at. However, the selection of a base model is of importance, because types and antitypes are interpreted with reference to a particular base model. Different base models lead to different interpretations of types and antitypes.

Several groups of base models have been described (von Eye, 2004). Examples include base models that reflect distributional assumptions, models that use a priori probabilities, and log-linear base models. In the present context, we use log-linear base models. These models reflect all effects that a researcher is *not* interested in. If types and antitypes emerge, those effects that the researcher is interested in, must exist. For example, if the base model takes only main effects into account (first order CFA), types and antitypes reflect interactions. Specifically, types and antitypes indicate where, in the table, interactions lead to the observation of more or fewer cases than expected. Another example is Prediction CFA. The base model for this variant of CFA is (1) saturated in the predictors, (2) saturated in the criteria, but (3) proposes independence of the predictors from the criteria. If, under this base model, types and antitypes emerge, they reflect, by necessity, predictor - criteria relationships.

In the following paragraphs, we report four applications of CFA to the Medicaid data. These include the first applications of CFA to fractional factorial designs reported in the literature.

1. First order CFA of the Medicaid data created for the fractional factorial design application in Table 7. Only main effects are included in the base model. If types and antitypes emerge,

they must be caused by two-way interactions. The model is

$$\ln m = \lambda + \lambda^{M3} + \lambda^{M4} + \lambda^{M5} + \lambda^{M6} + \lambda^{D6}.$$

2. First order CFA of the Medicaid data created for the completely crossed design in the appendix. Again, only main effects are included in the model. However, because higher order effects can, in principle exist, types and antitypes can be caused by interactions of any order. From the results in Table 10, we concluded that there are no higher order interactions in this data set. Therefore, the type/antitype patterns are expected to be very similar in both analyses. The base model for this analysis is the same as the base model for the first analysis.

3. The first two applications use straight main effect base models. None of the interactions is included. Therefore, in principle, types and antitypes can emerge from those two-way interactions that link earlier Medicaid reception with depression at Time 6, or from those two-way interactions among the Medicaid scores that were not included in the model. Specifically, these are the interactions M3 x M5, M3 x M6, M4 x M6, M3 x D6, M4 x D6, and M5 x D6. The good fit of Model 2 in Table 8 suggests that none of these terms is needed to explain the Medicaid data in the fractional design. Here, we focus on the long-term prediction of depression, and we ask whether including the interactions M3 x D6, M4 x D6, and M5 x D6 in the base model alters the pattern of types and antitypes that may result from the first CFA. Therefore, we now include these interactions in the base model. If types and antitypes still emerge, they indicate the relationships among the variables that are depicted in Figure 1, and, possibly, additional cross-time relationships among the Medicaid scores (cf. von Eye & Mair's, 2008, approach to explaining types and antitypes). The model thus becomes

$$\ln m = \lambda + \lambda^{M3} + \lambda^{M4} + \lambda^{M5} + \lambda^{M6} + \lambda^{D6} + \lambda^{M3,\,D6} + \lambda^{M4,\,D6} + \lambda^{M5,\,D6}.$$

4. Same as CFA 3, just from the completely crossed design.

CFA 1: First Order CFA from Fractional Factorial Design. For the base model in this application, we use the design matrix that is displayed in the main effects panel of Table 7. To perform the cell-wise tests, we use the $z$-test, and we protect $\alpha$ using the Holland-Copenhaver procedure (1987). Table 11 displays the results of this CFA.

Table 11: First Order CFA of Fractional Factorial Design for Medicaid Data

| Configuration[a] M3M4M5M6D6 | $m$ | $\hat{m}$ | $z$ | $p$ | Type/Antitype? |
|---|---|---|---|---|---|
| 11111 | 45 | 9.285 | 11.7205 | .000000 | Type |
| 11122 | 0 | .263 | -.5132 | .303913 | |
| 11212 | 0 | .293 | -.5410 | .294242 | |
| 11221 | 2 | 11.776 | -2.8489 | .002194 | Antitype |
| 12112 | 0 | .314 | -.5603 | .287631 | |
| 12121 | 0 | 12.631 | -3.5540 | .000190 | Antitype |
| 12211 | 1 | 14.039 | -3.4800 | .000251 | Antitype |
| 12222 | 1 | .398 | .9537 | .170110 | |
| 21112 | 0 | .337 | -.5804 | .280815 | |
| 21121 | 1 | 13.553 | -3.4099 | .000325 | Antitype |
| 21211 | 2 | 15.065 | -3.3660 | .000381 | Antitype |
| 21222 | 1 | .427 | .8762 | .190459 | |
| 22111 | 7 | 16.158 | -2.2782 | .011356 | |
| 22122 | 0 | .458 | -.6770 | .249217 | |
| 22212 | 1 | .509 | .6875 | .245899 | |
| 22221 | 55 | 20.492 | 7.6229 | .000000 | Type |

[a]A 1 in the configuration labels corresponds to a 1 in the design matrix; a 2 corresponds to a -1.

The results in Table 11 show that CFA identified 2 types and 5 antitypes. The first type, constituted by Configuration 11111 indicates that more women than expected with reference to the base model exhibit below-threshold depression when they never received Medicaid, over the entire observation period. The second type, constituted by Configuration 22221, shows that more women than expected who did receive Medicaid over the entire observation period show also below-threshold depression. These two types seem to suggest that Medicaid reception is unrelated to depression (cf. von Eye & Bogat, 2006).

However, the 5 antitypes show that there is a relationship. The first antitype, constituted by Configuration 11221, suggests that fewer women than expected show below-threshold, subclinical depression when they were placed on Medicaid reception between Time 4 and Time 5. Similarly, fewer women than expected show subclinical level depression when they received Medicaid only at Times 4 and 6 (Antitype 12121). The same applies when Medicaid was received only at Times 4 and 5 (Antitype 12211). The remaining antitypes (21121 and21211) also show that fewer than expected women who experience an unstable pattern of Medicaid reception are able to remain at below-threshold depression. We thus conclude that stability in Medicaid reception seems to be linked to subclinical levels of depression. In contrast, unstable patterns of Medicaid reception tend to be linked to a reduced probability of subclinical depression.

CFA 2: First Order CFA from Completely Crossed Design. First order CFA of the table from the completely crossed factors of the Medicaid data also used the *z*-test and the Holland-Copenhaver procedure. In addition, the log-linear base model was the same as before. However, the design matrix was that of a completely crossed design instead of a fractional design. If the fractional design does not lead to a distortion of the relationships in the table, results of the two analyses should be largely the same, even at the level of individual cells. Indeed, the results (not shown here) show strong overlap with the results from the fractional design. Specifically, Types 11111 and 22221 surfaced again. In addition, each of the configurations that constituted antitypes in the first analysis was, again, observed less frequently than expected.

However, the results from the fully crossed design differ from the ones for the fractional design in interesting ways. Specifically, none of the configurations that constituted antitypes differed from expectation strongly enough to constitute antitypes again, under the stricter levels of α protection which result for the larger table. Therefore, we asked whether they constitute a composite antitype (see von Eye, 2002). The Stouffer $Z = -3.59$ ($p = 0.0002$) suggests that this is the case. We, therefore, conclude that this difference between the two analyses may be due to the stricter α levels that result from α protection in larger tables. More important is that a new type emerged. It is constituted by Configuration 11112, indicating that more women than expected show clinical level depression in the absence of Medicaid reception over the entire observation period.

Types 11111 and 11112 differ only in the last digit. Therefore, they can be aggregated by

11111
11112

1111.,

to form the aggregated type 1111., where the dot indicates the variable aggregated over. The aggregated type suggests that more women than expected under the assumption of variable independence did not receive Medicaid at all. Interestingly, this pattern seems to contradict the conclusion that Pattern 1111 is associated with sub-threshold depression. Clearly, this result is an example that shows that CFA can lead to a more detailed and differing description of the relationships in data than log-linear modeling.

The same result was not found for Type 22221. Thus, the statement concerning this relationship does not need to be qualified based on the results from the completely crossed design.

CFA 3: Long-term Prediction of Depression: CFA from Fractional Factorial Design. In the following sample application, we ask whether those of the six interactions that would suggest long-term predictability of depression from Medicaid reception play indeed no role in the detection of types and antitypes in the Medicaid data. These are the interactions M3 x D6, M4 x D6, and M5 x D6. None of these are the interactions that were needed to explain the data (see Model 2 in Table 8). Another way of presenting this analysis is that we ask which types and antitypes emerge when we no longer consider the associations among the Medicaid scores M3, M4, and M5 and depression at Time 6. The log-linear base model for this CFA is, as was indicated above,

$$\ln m = \lambda + \lambda^{M3} + \lambda^{M4} + \lambda^{M5} + \lambda^{M6} + \lambda^{D6} + \lambda^{M3,\, D6} + \lambda^{M4,\, D6} + \lambda^{M5,\, D6}.$$

Considering that (1) the design for this analysis is fractional factorial with resolution at level V (three- and higher-way interactions can either not be estimated or are confounded), and (2), the two-way interactions M3 x M4, M4 x M5, M5 x M6, and M6 x D6 are not part of the base model, types and antitypes from this CFA base model reflect, by necessity, the cross time associations among the Medicaid reception variables, and the association between Medicaid reception and depression, at Time 6. If the three interactions M3 x D6, M4 x D6, and M5 x D6 play indeed no role in the detection of types and antitypes in the Medicaid data, the same types and antitypes will emerge from this analysis as in the first data example. The design matrix for this design appears in the first and the third panels of Table 7. For the CFA that uses this design matrix in its base model, we again employ the *z*-test and the Holland - Copenhaver procedure. The overall goodness-of-fit LR-$X^2$ for this model is 213.3 ($df$ = 7; $p$ < 0.01). This indicates major discrepancies between model and data. We thus can expect types and antitypes to emerge. Table 12 displays the results of this CFA.

Table 12: CFA of Fractional Factorial Design for Medicaid Data; Design Matrix Given in the First and Third Panels of Table 7

| Configuration[a] M3M4M5M6D6 | $m$ | $\hat{m}$ | $z$ | $p$ | Type/Antitype? |
|---|---|---|---|---|---|
| 11111 | 45 | 9.544 | 11.4765 | .000000 | Type |
| 11122 | 0 | 0 | .0000 | .499996 | |
| 11212 | 0 | .330 | -.5748 | .282716 | |
| 11221 | 2 | 13.542 | -3.1364 | .000855 | Antitype |
| 12112 | 0 | 0 | .0000 | .499995 | |
| 12121 | 0 | 11.211 | -3.3483 | .000407 | Antitype |
| 12211 | 1 | 13.702 | -3.4315 | .000300 | Antitype |
| 12222 | 1 | .670 | .4037 | .343199 | |
| 21112 | 0 | 0 | .0000 | .499995 | |
| 21121 | 1 | 13.844 | -3.4520 | .000278 | Antitype |
| 21211 | 2 | 12.881 | -3.0317 | .001216 | Antitype |
| 21222 | 1 | .859 | .1525 | .439396 | |
| 22111 | 7 | 18.400 | -2.6577 | .003934 | Antitype |
| 22122 | 0 | 0 | .0000 | .499994 | |
| 22212 | 1 | 1.141 | -.1323 | .447382 | |
| 22221 | 55 | 19.875 | 7.8789 | .000000 | Type |

The CFA types and antitypes in Table 13 are, with only one exception, identical to the ones in Table 11. The exception is that one additional antitype emerged. It is constituted by Configuration 22111 and suggests that fewer women than expected under the base model that was used for this run show subclinical depression when Medicaid is revoked between Time 4 and Time 5. We thus conclude that the long-term associations between Medicaid reception and depression have no effect on the pattern of types and antitypes. The types and antitypes in Tables 11 and 13 thus result solely from the cross time associations among the Medicaid scores and the association of Medicaid reception and depression at Time 6.

Additional base models are conceivable. For example, instead of asking which effects do not cause the types and antitypes in Table 11, one can ask which effects do cause these types and antitypes. To answer this question, a base model is needed that includes the cross-time associations among the Medicaid scores and the association between Medicaid and depression at T6. This is the base model

$$\ln m = \lambda + \lambda^{M3} + \lambda^{M4} + \lambda^{M5} + \lambda^{M6} + \lambda^{D6} + \lambda^{M3,\,M4} + \lambda^{M4,\,M5} + \lambda^{M5,\,M6} + \lambda^{M6,\,D6}.$$

If the interactions in this model are the causes for the types and antitypes in Table 11, they all will disappear under this base model (they do).

CFA 4: Long-Term Prediction of Depression: CFA from Completely Crossed Design. Using the table from the completely crossed variables and the hierarchical log-linear base model that includes the three interactions M3 x D6, M4 x D6, and M5 x D6, we obtain CFA results that mirror the ones from the second CFA, above (details not shown here). Configurations 11111 and 22221 constitute types. Also as before, each of the configurations that constituted antitypes in the first analysis was observed less frequently than expected, and the composite antitype exists. The additional type was not observed again. Thus, these results are even closer to the ones from Example 3 than the results from Example 2 were to the ones from Example 1.

We conclude again that, using the fractional factorial design can lead to appraisals of data structures that differ only minimally from those found using the complete design.

## 4. Discussion

It was the goal of this article to show that fractional factorial designs can fruitfully be applied when the outcome variables are categorical. Examples were given using explanatory log-linear modeling and exploratory Configural Frequency Analysis. One of the main arguments used in this article concerns the resolution of a design. Resolution is defined as the order of interactions that can be estimated so that they are not confounded with each other. Box - Hunter designs are particularly useful because they allow one, for an a priori determined desired resolution, to specify a design that has exactly this resolution.

It was also shown in this article, that the application of fractional factorial designs in the analysis of categorical outcome variables mirrors the application in the analysis of metric outcome variables only in part. Specifically, when categorical outcome variables are analyzed, resolution has to be selected one level higher than for metric outcome variables. The reason for this difference is that main effects in the context of the GLM already describe the relationships between one independent and one outcome variable. In contrast, two-way interactions are needed to do the same when the outcome variable is categorical. Therefore, fractional designs can be more parsimonious when metric outcome variables than when categorical outcome variables are analyzed.

Implications of this methodology are major. First, designs and data collection become far more parsimonious than when routinely completely crossed designs are used. Second, data analysis and interpretation of results are simplified because only those interactions are discussed that were of interest based on theory and prior results. Third, all this can be conducted within the context of methods of analysis from the Generalized Linear Model. Thus, methods of ANOVA, the GLLM, and CFA can be used without significant adaptation. The popular general purpose statistical software packages, for example, R, SAS, SYSTAT, or Minitab can be used for data analysis. In addition, some of these packages, for example, SYSTAT, Statistica or Minitab offer modules that allow the researcher to specify the design.

The data examples given in this article showed first that fractional factorial designs can be applied without any problems when log-linear and configural methods are used for analysis. In addition, the examples showed that the loss of information can be minimal when fractional designs are used. Loss will always be minimal when the sparsity of effects principle applies. In different words, loss will be minimal when the assumptions that researchers make about the existence of effects are correct. For example, when researchers assume that three- and higher-way interactions do not exist, there is no need to increase the resolution of a design beyond IV, when metric, and beyond V, when categorical outcome variables are observed.

An interesting characteristic of fractional factorial designs is that, in most cases, for the same resolution, more than one design exists. These alternative designs, called *fractions*, result from reversing the signs of those factors that are not coded as in a completely crossed design. Let the number of factors included in the design be $p$. Then, the completely crossed design will have $2^p$ runs (cells). The number of runs in a fractional design of the Box - Hunter type is $2^{(p - k)}$ where $k$ is the number of factors whose main effects are not coded as in a completely crossed design. Then, the number of fractions is $2^k$. Consider, for example, the design given in Table 5. The last of the main effect vector, the one for Variable C, is $c_1' = \{-1, 1, 1, -1\}$. The second fraction for this design comes with $c_2' = \{1, -1, -1, 1\}$. To give another example, consider the Box - Hunter design with 7 factors and 16 runs. For 16 runs, 4 factors can be coded as in a completely crossed design. The remaining 3 are coded in analogy to interactions. Thus, the number of fractions is 8. This number can be increased by changing the order of variables in the table. Thus, the researchers have considerable flexibility as to how to set up their experiments or data collection.

One important application of log-linear modeling involves the examination of the association structure of manifest variables (Goodman, 1984). In these applications, all variables have the same status. That is, variables are not considered dependent or independent. Here, fractional factorial designs are still applicable. Decisions about the complexity of designs can be made based on cost and effort considerations, the sparsity of effects principle, and knowledge concerning the order of interactions that were observed in prior studies. Similarly, screening studies can be designed in which researchers explore whether interactions of typically low order exist. In studies with no distinction of dependent and independent variables, the concept of resolution can be applied with no change from studies in which this distinction is made. For example, resolution level V is needed for the interpretation of two-way associations that are unconfounded with each other.

Designs with resolution V can also be of importance when such latent variable models are considered as correspondence analysis or structural modeling based on categorical variables. As was mentioned above, these methods typically start estimation from a matrix of bivariate relationships. To create such a matrix from a completely crossed design implies collecting and paying for information that is not used. Therefore, fractional factorial designs should be considered. These designs are not only far more parsimonious, they also allow one to include far more categorical variables (or factors). For example, 10 categorical variables can be observed at a resolution level of

V in 128 runs. The completely crossed design would require 1024 runs. The fractional design thus comes with a savings of 87.5%.

Another interesting option in the analysis of categorical outcome variables is that multiple outcome variables can be studied simultaneously, with the dependency structure of the outcome variables being part of the model. Here again, for a given number of runs that can realistically be conducted, designs with a resolution level of V or higher will allow the researchers to include far more variables than a completely crossed design. For example, for 8 categorical variables, only 128 runs are needed to obtain a resolution level of VIII, a savings of 50%. To obtain a resolution level of V with 8 variables, only 64 runs are needed, a savings of 75%.

Now, to turn the arguments around, suppose cost, effort, and time are not issues of concern. Would one always run the fully crossed design? Interestingly, the answer is no and, thus, counter the current common practice in Psychological research. The argument concerning interpretability raised at the beginning of this article holds even if one is flush. This argument involves the statement that interaction of very high order are virtually impossible to interpret. Consider a design with only eight fully crossed, binary factors. This is a design with $2^8 = 256$ treatment combinations (cells). Interpreting the interaction among all eight factors is close to impossible. The Box - Hunter design with 128 runs for these eight factors has a resolution of VIII. That is, at a savings of 50%, we can create a design that allows us to estimate up to five-way interactions. This is sufficient for the testing of a very large percentage of social science theory-driven hypotheses. In addition, this will be hard enough.

Where, then, in the process of research, will fractional factorial designs be discussed? The answer is clear: in the planning phase. When data are collected under a design that is completely crossed, it does not make much sense to exclude data from analysis just because fractional designs are attractive. Similarly, when sampling is multinomial or Poisson, fractional factorial designs may lead to the exclusion of too many cases. A third case is the above-mentioned Meehl paradox. When a hypothesis states that an interaction of a particular order carries the action in a table, the resolution must be selected accordingly. In contrast, when sampling proceeds after some screening is performed to select possible respondents, that is when the focus is on respondents with particular profiles, or when the experimenter screens and, based on screening results, assigns cases to treatments, fractional factorial designs can be very helpful.

References

Agresti, A. (2002). *Categorical data analysis*, 2nd ed. New York: Wiley.

Atkinson, A.C., & Fedorov, V.V. (1975). The designs of experiments for discriminating between two rival models. *Biometrika*, *62*, 57 - 70.

Berger, M.P.F. (2005). Optimal design for categorical variables. In B.S. Everitt, & D.C Howell (eds.), *Encyclopedia of statistics in behavioral science* (pp. 1474 - 1479). Chichester, England: Wiley.

Bisgaard, S., & Steinberg, D. M. (1997). The Design and Analysis of 2k-p Prototype Experiments. *Technometrics*, *39*, 52 - 62.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

Bock, R.D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw Hill.

Box, G.E.P., & Behnken, D.W. (1960). Some new three level designs for the study of quantitative variables. *Technometrics*, *2*, 455 - 475.

Box, G.E.P., Hunter, J.S., & Hunter, W.G. (2005). *Statistics for experimenters: Design innovation, and discovery*, 2nd ed. Hoboken, NJ: John Wiley & Sons.

Bruwer, M.-J., & MacGregor, J.F. (2005). Robust multi-variable identification: Optimal experimental design with constraints. *Journal of Process Control*, *16*, 581-600.

Dodge, Y., Fedorov, V.V., & Wynn, H.P. (Eds.)(1988). *Optimal Design and Analysis of Experiments*. Amsterdam: North-Holland.

Euler, L. (1782). Recherches sur une nouvelle espèce de quarrés magiques. *Verhandelingen uitgegeven door het zeeuwsch Genootschap der Wetenschappen te Vlissingen*, *9*, 85 - 239.

Evers, M., & Namboodiri, K.N. (1979). On the design matrix strategy in the analysis of categorical data. *Sociological Methodology*, *10*, 86 - 111.

Finney, D.J. (1945). The fractional replication of factorial arrangements. *Annals of Eugenics*, *12*, 291 - 301.

Finney, D.J. (1946). Recent developments in the design of field experiments III. Fractional replication. *Journal of Agricaultural Science*, *36*, 184 - 191.

Goodman, L.A. (1984). *The analysis of cross-classified data having ordered categories*. Cambridge, MA: Harvard University Press.

Hamada, M., & Wu, C.F.J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, *24*, 130 - 137.

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics, 43*, 417 - 423.

Kempthorne, O. (1947). A simple approach to confounding and fractional replication in factorial experiments. *Biometrika*, *34*, 255 - 272.

Kessels, R., Goos, P., Vandebroek, M. (2006) A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, *43*, 409 - 419.

Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2004). *Applied linear statistical models*, 4th ed. Boston: McGraw-Hill.

Ledolter, J., & Swersey, A.J. (2007). *Testing 1-2-3: Experimental design with applications in marketing and service operations*. Stanford, CA: Stanford University Press.

Levendosky, A.A., Bogat, G.A., Davidson, W.S., & von Eye, A. (2000). *Risk and protective factors for domestic violence*. Centers for Disease Control, Injury Prevention Center, RO1/CCR518519-01.

Lienert, G.A., & Krauth, J. (1975). Configural Frequency Analysis as a statistical tool for defining types. *Educational and Psychological Measurement*, *35*, 231 - 238.

Liski, E.P., Mandal, N.K., Shah, K.R., & Sinha, B.K. (2002). *Topics in optimal design*. New York: Springer.

Mair, P., & von Eye, A. (2007). Application scenarios for nonstandard log-linear models. *Psychological Methods*, *12*, 139 - 156.

Meehl, P.E. (1950). Configural scoring. *Journal of Consulting Psychology*, *14*, 165 - 171.

Mukerjee, R., & Wu, C.F.J. (2006). *A modern theory of factorial design*. New York: Springer.

Nelder, J.A., & Wedderburn, R.W.A. (1972). Generalized linear models. *Journal of the Royal Statstical Society, Series A (Statistics in Society)*, *135*, 370 - 384.

Plackett, R.L., & Burman, J.P. (1946).The Design of Optimum Multifactorial Experiments. *Biometrika, 33*, 305 - 325.

Pukelsheim, F. (2006). *Optimal design of experiments*. New York: Wiley.

Reverte, C., Dirion, J.-L., & Cabassud, M. (2006). Kinetic model identification and parameter estimation from TGA experiments. *Journal of Analytical and Applied Pyrolysis*, *79*, 297 - 305.

Rindskopf, D. (1990). Nonstandard log-linear models. *Psychological Bulletin*, *108*, 150 - 162.

Ryan, T. (2007). *Modern experimental design*. Hoboken, NJ: John Wiley & Sons.

Salem, A.M., Rekab, K., & Whittaker, J.A. (2004). Prediction of software failure through logistic regression. *Information and Software Technology*, *46*, 781 - 789.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. Boca Raton, FL: Chapman & Hall.

Stigler, S. (1971. Optimal experimental designs for polynomial regression. *Journal of the American Statistical Association*, *66*, 311 - 318.

Taguchi, G. (1987). *The System of Experimental Design*. White Plains, NY: Unipub/Kraus.

Vermunt, J.K. (2005). Log-linear models. In B.S. Everitt, & D.C Howell (eds.), *Encyclopedia of statistics in behavioral science* (pp. 1092 - 1093). Chichester, England: Wiley.

Verotta D., Petrillo P., La Regina A., Rocchetti M. & Tavani, A. (1988). D-Optimal design applied to binding saturation curves of an enkephalin analog in rat brain. *Life Sciences*, *42*, 735 - 743.

von Eye, A. (2002). *Configural Frequency Analysis - Methods, Models, and Applications*. Mahwah, NJ: Lawrence Erlbaum.

von Eye, A. (2004). Base models for Configural Frequency Analysis. *Psychology Science*, *46*, 150 - 170.

von Eye, A., & Bogat, G.A. (2006). Mental health in women experiencing intimate partner violence as the efficiency goal of social welfare functions. *International Journal of Social Welfare*, *15*, 31 - 40.

von Eye, A., & Gutiérrez Peña, E. (2004). Configural Frequency Analysis - the search for extreme cells. *Journal of Applied Statistics*, *31*, 981 - 997.

von Eye, A., & Mair, P. (2008). Functional Configural Frequency Analysis (under editorial review).

von Eye, A., Mair, P., & Bogat, G.A. (2005). Prediction models for Configural Frequency Analysis. *Psychology Science*, *47*, 342 - 355.

von Eye, A., Schuster, C., & Rogers, W.M. (1998). Modeling synergy using manifest categorical variables. *International Journal of Behavioral Development*, *22*, 537 - 557.

Wu, C.F.J., & Hamada, M. (2000). *Experiments: Planning, analysis and parameter design optimization.* New York: John Wiley & Sons.

Zelen, M., & Haitovsku, Y. (1991). Testing hypotheses with binary data subject to misclassification errors: Analysis and experimental design. *Biometrika*, *78*, 857 - 865.

Appendix: Complete Cross-Classification for the Medicaid Data Example

| T3Medicaid | T4Medicaid | T5Medicaid | 0 T6Medicaid | T6 Depression | |
|---|---|---|---|---|---|
| | | | | 1 | 2 |
| 1 | 1 | 1 | 1 | 55.000 | 9.000 |
| | | | -1 | 9.000 | 1.000 |
| | | | 1 | 4.000 | 0.000 |
| | | | -1 | 7.000 | 0.000 |
| | -1 | 1 | 1 | 4.000 | 1.000 |
| | | | -1 | 2.000 | 1.000 |
| | | -1 | 1 | 1.000 | 0.000 |
| | | | -1 | 7.000 | 0.000 |
| -1 | 1 | 1 | 1 | 5.000 | 1.000 |
| | | | -1 | 1.000 | 0.000 |
| | | -1 | 1 | 0.000 | 0.000 |
| | | | -1 | 2.000 | 0.000 |
| | -1 | 1 | 1 | 2.000 | 0.000 |
| | | | -1 | 4.000 | 0.000 |
| | | -1 | 1 | 5.000 | 0.000 |
| | | | -1 | 45.000 | 2.000 |