Methodika
1992, Vol. VI, pp. 76-85

On the Relation Between Interactions Obtained with Alternative Codings of Discrete Variables

Nanny Wermuth, and D. R. Cox

University of Mainz and Nuffield College, Oxford

Abstract: The definition of an interaction parameter in a statistical model depends on the chosen coding for the discrete variables, i.e. on the design matrix. A design matrix can be built up as Kronecker product of small matrices containing codes for an overall effect and for main effects. As a consequence the relations between different sets of interaction terms can be understood by relating just main effect definitions under different codings. An application is described in which the excellent fit of a model defined with orthogonal polynomial effects can be deduced directly from estimation results in terms of effect coding, but not in terms of indicator coding.

Key words: Effect coding, indicator coding, log linear model, orthogonal polynomial coding

1. Introduction

In statistical models containing discrete variables or factors, suitable constraints are needed to obtain unique definitions of effect parameters. Examples are exponential family models (Barndorff-Nielsen, 1978) linear in moment parameters, like a two-way analysis of variance model for a normally distributed response:

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}, \tag{1}$$

or linear in canonical parameters like a log-linear model, i.e. a linear model in the logarithm of probabilities $\pi_{ij} > 0$ of a two-way table:

$$\log \pi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}. \tag{2}$$

In both situations the levels of the discrete variables A and B are denoted by $0, 1, \ldots, (I-1)$ and $0, 1, \ldots, (J-1)$, where I and J are the number of categories of A and of B, respectively. There are $I \cdot J$ independent parameters in (1), $I \cdot J - 1$ independent parameters in (2) since in the latter case $\sum_{ij} \pi_{ij} = 1$. A common treatment of the two models is possible by writing h_{ij} for the observed counterparts of the lefthand sides of (1) and (2). We

Author's address: Prof. Nanny Wermuth, Ph.D., Psychologisches Institut, Universität Mainz, Saarstr. 21, Bau 413, D-6500 Mainz 1, Germany.

assume for (2) that there are no empty cells in the contingency table. The models can be regarded as special cases of generalized linear models (McCullagh & Nelder, 1989) and as building blocks of special graphical chain models (Wermuth & Lauritzen, 1990).

Two types of constraints on α_i, β_j and γ_{ij} are in common use: symmetric and baseline constraints. The symmetric constraints are

$$0 = \sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij},$$

and the baseline constraints with levels zero as base are

$$0 = \alpha_0 = \beta_0 = \gamma_{00} = \gamma_{01} = \gamma_{10}.$$

Other constraints, e.g. ones involving weighted sums, are possible but will not be considered here. We denote interaction parameters resulting from symmetric constraints by λ's, those from baseline constraints by δ's. We derive their relations to each other and to coding in terms of orthogonal polynomials denoted by ϱ's. In particular, we show that symmetric constraints correspond to effect coding and baseline constraints to indicator or dummy coding in the sense used in the statistical literature for social scientists (Kerlinger & Pedhazur, 1973).

For many purposes the individual interaction terms, i.e. components of an interaction, are not of direct interest. Usually, appreciable interaction will imply that summarization via main effects is inadequate and will lead to abandoning representations like (1) and (2) as a base for interpretation. Nevertheless, for some specific purposes it is desirable to know simple explicit expressions relating different sets of interactions: one might want

- to relate estimation results from one computer program like BMDP, employing symmetric constraints, i.e. effect coding, to those of another like GLIM, employing baseline constraints, i.e. indicator coding;
- to relate estimation results from different studies on the same set of variables, each reported for different coding systems;
- to decide whether a reduced model defined by zero restrictions on a subset of interaction terms in one coding system is equivalent to a comparable reduced model defined in terms of another coding system;
- to use results of a given computer program to obtain point and interval estimates of interactions defined in another coding system appropriate for a particular subject matter purpose.

2. The Relation of Common Constraint and Coding Systems for Binary Variables

Under the saturated model for two binary variables A and B estimated interactions $\hat{\lambda}$'s and $\hat{\delta}$'s are obtained after one-to-one transformations on h_{ij}:

$$\begin{aligned}
\hat{\lambda}_{-} &= \bar{h}_{..} &&= (h_{00} + h_{10} + h_{01} + h_{11})/4,\\
\hat{\lambda}_0^A &= \bar{h}_{0.} - \bar{h}_{..} &&= (h_{00} - h_{10} + h_{01} - h_{11})/4,\\
\hat{\lambda}_0^B &= \bar{h}_{.0} - \bar{h}_{..} &&= (h_{00} + h_{10} - h_{01} - h_{11})/4,\\
\hat{\lambda}_{00}^{AB} &= \bar{h}_{00} - \bar{h}_{0.} - \bar{h}_{.0} + \bar{h}_{..} &&= (h_{00} - h_{10} - h_{01} + h_{11})/4,
\end{aligned} \tag{3}$$

$$\begin{aligned}
\hat{\delta}_{-} &= h_{00},\\
\hat{\delta}_1^A &= h_{10} - h_{00},\\
\hat{\delta}_1^B &= h_{01} - h_{00},\\
\hat{\delta}_{11}^{AB} &= h_{11} - h_{10} - h_{01} + h_{00}.
\end{aligned} \tag{4}$$

Written in matrix notation with e.g.

$$h^T = (h_{00}, h_{10}, h_{01}, h_{11}), \qquad \hat{\lambda}^T = (\hat{\lambda}_{-}, \hat{\lambda}_0^A, \hat{\lambda}_0^B, \hat{\lambda}_{00}^{AB}),$$

the relation between $\hat{\lambda}$'s and $\hat{\delta}$'s can be expressed from $\hat{\lambda} = T_e^{-1} h$ and $\hat{\delta} = T_i^{-1} h$ as

$$\hat{\delta} = T_i^{-1} T_e \hat{\lambda} \tag{5}$$

where

$$T_e = \begin{pmatrix} 1 & 1 & 1 & 1\\ 1 & -1 & 1 & -1\\ 1 & 1 & -1 & -1\\ 1 & -1 & -1 & 1 \end{pmatrix}, \qquad T_i = \begin{pmatrix} 1 & 0 & 0 & 0\\ 1 & 1 & 0 & 0\\ 1 & 0 & 1 & 0\\ 1 & 1 & 1 & 1 \end{pmatrix}$$

are design matrices resulting from effect coding and from indicator coding. It may be checked directly that T_e and T_i have as inverses the matrices implicitly defined by (3) and (4). Similarly, the upper triangular form of $T_e T_i^{-1}$ can be deduced from the proportionality of T_e to an orthogonal matrix and the particular triangular form of the given T_i. However, it is instructive and generalizes more easily to exploit properties of Kronecker products for this purpose. It follows from

$$T_e = \begin{pmatrix} 1 & 1\\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1\\ 1 & -1 \end{pmatrix}, \qquad T_i = \begin{pmatrix} 1 & 0\\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0\\ 1 & 1 \end{pmatrix},$$

and from the inverse of a Kronecker product of matrices being the Kronecker product of the inverses that

$$T_e^{-1} = \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad T_i^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

This matrix formulation of the Yates algorithm (Yates, 1937) was given by Good (1958).

Furthermore, since the product of two matrices defined by Kronecker products is the Kronecker product of the matrix products, we have

$$\begin{aligned} T_i^{-1} T_e &= \left[\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right] \otimes \left[\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right] \\ &= \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix}, \end{aligned}$$

which is – as a Kronecker product of upper triangular matrices – of upper triangular form.

This result extends directly to all 2^k designs and implies the equivalence of reduced models defined by successively setting single interaction terms to zero starting in either of the two systems from the highest-order interaction term. This property is in general not retained for nonbinary data nor does it hold for nonhierarchical models. Recall that in a hierarchical model any zero main effect or any zero interaction implies the vanishing of all higher order interactions containing it.

For instance, the following vector of counts $n\hat{\pi}$ in a 2^3 table can be concisely described by a nonhierarchical model in effect coding but not in indicator coding. The counts were obtained after median dichotomizing 500 values of three variables simulated from a standardized trivariate normal distribution with two variable pairs having 0.8 and one pair having -0.4 as correlation coefficient

$$\begin{aligned} n\hat{\pi} &= n(\hat{\pi}_{000}, \hat{\pi}_{100}, \hat{\pi}_{010}, \hat{\pi}_{110}, \hat{\pi}_{001}, \hat{\pi}_{101}, \hat{\pi}_{011}, \hat{\pi}_{111}) \\ &= (87, 3, 78, 82, 82, 78, 3, 87). \end{aligned}$$

These counts follow a nonhierarchical model in effect coding with restrictions

$$0 = \lambda_i^A = \lambda_j^B = \lambda_k^C = \lambda_{ijk}^{ABC},$$

but with two-factor interactions nonzero. To put it differently, the counts are reproduced by $\hat{\lambda}_{-} = 3.582$, $\hat{\lambda}_{00}^{AB} = 0.8543$, $\hat{\lambda}_{00}^{AC} = 0.8293$, $\hat{\lambda}_{00}^{BC} = -0.7997$. The unusual structure with main effects zero but two-factor interactions nonzero is a consequence of dichotomizing at the median and the particular correlation structure having two positive and one negative correlation.

This model does not imply

$$0 = \delta_i^A = \delta_j^B = \delta_k^C = \delta_{ijk}^{ABC},$$

i.e. these two nonhierarchical models defined in terms of λ's and δ's are distinct.

3. Main Effects Under Different Codings for Few Variables

Many applications of statistical models with discrete variables concern mixed factorial systems and involve a small number of variables each having a small number of categories but they are not all binary. In these situations the design matrices may be built up in a way similar to that used in section 2. The needed building blocks are design matrices containing for each variable with K categories one column vector of ones and $K-1$ column vectors, corresponding to one of the available $K-1$ degrees of freedom associated with this variable. Equivalently, there is the K by K inverse design matrix defining the overall and main effects under the chosen coding system. In this section we present such matrices for effect, indicator and orthogonal polynomial coding for up to four variables.

Under *effect coding* the design matrices for variables with two, three and four levels denoted by L_2, L_3, L_4, are

$$L_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad L_3 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix}, \quad L_4 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix},$$

and their inverses, which define overall and main effects are

$$L_2^{-1} = \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad L_3^{-1} = \tfrac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix}, \quad L_4^{-1} = \tfrac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \end{pmatrix}.$$

Under *indicator coding* with levels zero as baseline the corresponding design matrices, denoted by D_2, D_3, D_4 are

$$D_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad D_3 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad D_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

and their inverses are

$$D_2^{-1}=\begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix},\quad D_3^{-1}=\begin{pmatrix}1 & 0 & 0\\ -1 & 1 & 0\\ -1 & 0 & 1\end{pmatrix},\quad D_4^{-1}=\begin{pmatrix}1 & 0 & 0 & 0\\ -1 & 1 & 0 & 0\\ -1 & 0 & 1 & 0\\ -1 & 0 & 0 & 1\end{pmatrix}.$$

Under coding in terms of *orthogonal polynomials* the corresponding design matrices denoted here by P_2, P_3, P_4 can be obtained from published tables (Fisher & Yates, 1963; Snedecor & Cochran, 1967), as

$$P_2=\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix},\quad P_3=\begin{pmatrix}1 & -1 & 1\\ 1 & 0 & -2\\ 1 & 1 & 1\end{pmatrix},\quad P_4=\begin{pmatrix}1 & -3 & 1 & -1\\ 1 & -1 & -1 & 3\\ 1 & 1 & -1 & -3\\ 1 & 3 & 1 & 1\end{pmatrix},$$

with inverses which may be obtained after transposing and applying suitable factors to columns:

$$P_2^{-1}=\tfrac{1}{2}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix},\quad P_3^{-1}=\tfrac{1}{6}\begin{pmatrix}2 & 2 & 2\\ -3 & 0 & 3\\ 1 & -2 & 1\end{pmatrix},\quad P_4^{-1}=\tfrac{1}{20}\begin{pmatrix}5 & 5 & 5 & 5\\ -3 & -1 & 1 & 3\\ 5 & -5 & -5 & 5\\ -1 & 3 & -3 & 1\end{pmatrix}.$$

While the form of the relevant matrices for variables with more than four categories under effect and indicator coding can be directly derived from extending the given examples, the codes of orthogonal polynomials based on equally spaced levels may be obtained by successive orthogonalization as suggested by Dempster (1969, p. 218) and as described, for the sake of completeness, in an example in the Appendix.

By using again properties of Kronecker products of matrices the relations among interactions for a $4\times 3\times 2$ contingency table are

$$\hat{\delta}=T_i^{-1}T_e\hat{\lambda},\quad \hat{\delta}=T_i^{-1}T_p\hat{\varrho},\quad \hat{\lambda}=T_e^{-1}T_p\hat{\varrho},$$

where

$$\begin{aligned}T_i^{-1}T_e &= D_2^{-1}L_2\otimes D_3^{-1}L_3\otimes D_4^{-1}L_4\\ &=\begin{pmatrix}1 & 1\\ 0 & -2\end{pmatrix}\otimes\begin{pmatrix}1 & 1 & 0\\ 0 & -1 & 1\\ 0 & -2 & -1\end{pmatrix}\otimes\begin{pmatrix}1 & 1 & 0 & 0\\ 0 & -1 & 1 & 0\\ 0 & -1 & 0 & 1\\ 0 & -2 & -1 & -1\end{pmatrix},\end{aligned}$$

$$T_i^{-1} T_p = D_2^{-1} P_2 \otimes D_3^{-1} P_3 \otimes D_4^{-1} P_4$$

$$= \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -3 \\ 0 & 2 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & -3 & 1 & -1 \\ 0 & 2 & -2 & 4 \\ 0 & 4 & -2 & 2 \\ 0 & 6 & 0 & 2 \end{pmatrix},$$

$$T_e^{-1} T_p = L_2^{-1} P_2 \otimes L_3^{-1} P_3 \otimes L_4^{-1} P_4$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -2 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 1 & -1 \\ 0 & -1 & -1 & 3 \\ 0 & 1 & -1 & -3 \end{pmatrix}.$$

Several conclusions may be drawn.

- The matrices resulting from these Kronecker products are all upper block-triangular, because of the special form of the first column of all matrices in the product. As a consequence any two interaction models are equivalent which are defined by setting all interaction terms within a block successively to zero, i.e. by starting from the highest-order interaction.
- Effect coding relates in a simpler way to orthogonal polynomials than indicator coding since definitions of the overall terms coincide for the former ($\lambda_- = \varrho_-$), but not for the latter.
- In 3^k-tables the relation between interactions obtained with effect and orthogonal polynomial coding is particularly simple, since then $T_e^{-1} T_p$ is of upper triangular – not just block-triangular – form.

One application is to interrelations of two ordinal scales with just three values: the plausibility of the hypothesis that their pairwise associations are just linear may be checked directly from studentized interactions under effect coding, i.e. from interaction terms with symmetric constraints divided by their standard error.

To see this note first that for counts ordered so that the index of the first variable changes fastest a lexicographical ordering of the interaction terms is produced. In a 3^2-table we get the following correspondences in the ordering

$$\hat{\pi}^T = (\hat{\pi}_{00}, \hat{\pi}_{10}, \hat{\pi}_{20}, \hat{\pi}_{01}, \hat{\pi}_{11}, \hat{\pi}_{21}, \hat{\pi}_{02}, \hat{\pi}_{12}, \hat{\pi}_{22}),$$

$$\hat{\lambda}^T = (\hat{\lambda}_-, \hat{\lambda}_0^A, \hat{\lambda}_1^A, \hat{\lambda}_0^B, \hat{\lambda}_{00}^{AB}, \hat{\lambda}_{10}^{AB}, \hat{\lambda}_1^B, \hat{\lambda}_{01}^{AB}, \hat{\lambda}_{11}^{AB}),$$

$$\hat{\varrho}^T = (\hat{\varrho}_-, \hat{\varrho}_l^A, \hat{\varrho}_q^A, \hat{\varrho}_l^B, \hat{\varrho}_{ll}^{AB}, \hat{\varrho}_{ql}^{AB}, \hat{\varrho}_q^B, \hat{\varrho}_{lq}^{AB}, \hat{\varrho}_{qq}^{AB}).$$

Then, the upper triangular form of the matrix relating ϱ to λ implies that the hypothesis

$$0 = \varrho_{ql}^{AB} = \varrho_q^B = \varrho_{lq}^{AB} = \varrho_{qq}^{AB}$$

is equivalent to the hypothesis

$$0 = \lambda_{10}^{AB} = \lambda_1^B = \lambda_{01}^{AB} = \lambda_{11}^{AB}.$$

By a symmetry argument note further that given $0 = \lambda_{10}^{AB} = \lambda_{11}^{AB}$ the hypothesis $0 = \lambda_1^A$ is equivalent to $0 = \varrho_q^A$. Thus, if standard computer output for interactions with symmetric constraints like BMDP shows studentized interactions of λ_1^A, λ_1^B, λ_{10}^{AB}, λ_{01}^{AB}, λ_{11}^{AB} under the saturated model to be all small, one can conclude that the hypotheses of just linear interrelations among the two ordinal scales and of no quadratic main effects are well compatible with the observations. If desired, approximate maximum-likelihood estimates under this model may then be obtained as described by Cox and Wermuth (1990). The variance of $\hat{\varrho}$ under the saturated model needed for these computations is readily computed from e.g. the given variance matrix of $\hat{\lambda}$ since $\hat{\varrho} = A\hat{\lambda}$ implies $var(\hat{\varrho}) = A var(\hat{\lambda}) A^T$.

4. An Example with a 3^2-Table

In a study on patients from a pain clinic (Schmitt, 1990) several ordinal scales are reported having three values which describe different aspects of chronic pain. Increasing duration and intensity of treatment are reflected in the three ordered values of the variables:

A = pain induced rehabilitative treatments,
B = pain induced stationary treatments.

Table 1: Counts and other data summaries for variables A = pain induced rehabilitative treatments and B = pain induced stationary treatments ($n = 149$)

Levels of			Studentized interactions from coding with					
			orth. polyn.		effects		indicators	
A	B	count	type	value	type	value	type	value
0	0	57	$\hat{\varrho}_-$	–	$\hat{\lambda}_-$	–	$\hat{\delta}_-$	−9.23
1	0	21	$\hat{\varrho}_l^A$	−3.02	$\hat{\lambda}_0^A$	2.79	$\hat{\delta}_0^A$	−3.91
2	0	6	$\hat{\varrho}_q^A$	−0.01	$\hat{\lambda}_1^A$	0.01	$\hat{\delta}_1^A$	−5.25
0	1	18	$\hat{\varrho}_l^B$	−3.49	$\hat{\lambda}_0^B$	3.18	$\hat{\delta}_0^B$	−4.26
1	1	15	$\hat{\varrho}_{ll}^{AB}$	4.47	$\hat{\lambda}_{00}^{AB}$	3.95	$\hat{\delta}_{00}^{AB}$	1.89
2	1	8	$\hat{\varrho}_{ql}^{AB}$	0.83	$\hat{\lambda}_{10}^{AB}$	0.45	$\hat{\delta}_{10}^{AB}$	2.38
0	2	6	$\hat{\varrho}_q^B$	−0.32	$\hat{\lambda}_1^B$	0.32	$\hat{\delta}_1^B$	−5.25
1	2	6	$\hat{\varrho}_{lq}^{AB}$	0.06	$\hat{\lambda}_{01}^{AB}$	−0.34	$\hat{\delta}_{01}^{AB}$	1.58
2	2	12	$\hat{\varrho}_{qq}^{AB}$	0.76	$\hat{\lambda}_{11}^{AB}$	0.76	$\hat{\delta}_{11}^{AB}$	4.47

These two variables are considered to be on an equal footing, i.e. none is regarded as response to the other; rather, they are both potential explanatory variables for the stage of chronic pain a patient has reached or for success of pain treatment. Although the three ordered values of A and B are not on an interval scale, it is a convenient help for subsequent analysis and interpretation to score them linearly, like $-1, 0, 1$, in the hope that the association between A and B is essentially linear, i.e. captured in the linear $\times$ linear term.

A likelihood ratio chi-square test clearly rejects independence of the two scales with a value of 51.88 on 4 degrees of freedom. Nevertheless, the anticipated simplified description of the observations can be given: the counts are well described by just linear main effects and a linear by linear interaction, i.e. by a model on 5 degrees of freedom. Table 1 shows that this result obtained by using orthogonal polynomial coding could have been deduced from the studentized interactions computed in terms of effect coding but not from those computed in terms of indicator coding. The structure is an example of a model being hierarchical in one coding system, i.e. for coding in orthogonal polynomials, but being nonhierarchical in the other two coding systems.

Appendix

For a variable with K categories the design matrix for main effects can be computed by successively orthogonalising the columns in the following matrix

$$\begin{pmatrix} 1 & 1 & 1^2 & 1^3 & \dots & 1^{(K-1)} \\ 1 & 2 & 2^2 & 2^3 & \dots & 2^{(K-1)} \\ 1 & 3 & 3^2 & 3^3 & \dots & 3^{(K-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & K & K^2 & K^3 & \dots & K^{(K-1)} \end{pmatrix} .$$

For example with $K=3$ this means to replace the matrix with three column vectors (a, b, c) by orthogonal column vectors (a', b', c') defined as

$$a' = a, \; b' = b - \beta_{ba'} a', \; c' = c - \beta_{ca'} a' - \beta_{cb'} b',$$

where β_{uv} denotes a simple linear regression coefficient obtained by regressing u on v. This gives

$$a' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad b' = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \tfrac{6}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

$$c' = \tfrac{1}{3}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 9 \end{pmatrix} - \tfrac{14}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \tfrac{8}{2}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

In books on linear algebra this procedure can be found under the name of Gram-Schmidt-orthogonalisation and can be applied directly to deal with unequally spaced levels.

References

Barndorff-Nielsen, O. E. (1978). *Information and exponential families in statistical theory.* Chichester: Wiley.

Cox, D. R. & Wermuth, N. (1990). An approximation to maximum-likelihood estimates in reduced models. *Biometrika, 77,* 747–761.

Dempster, A. P. (1969). *Elements of continuous multivariate analysis.* Reading: Addison-Wesley.

Fisher, R. A. & Yates, F. (1963). *Statistical tables for biological, agricultural and medical research. 6th ed.* London: Longman.

Good, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. Royal Statistical Society B, 20,* 361–372.

Kerlinger, F. N. & Pedhazur, E. J. (1973). *Multiple regression in behavioral research,* New York: Rinehart and Winston.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models, 2nd ed.* London: Chapman and Hall.

Schmitt, N. (1990). Stadieneinteilung chronischer Schmerzen. Medical Dissertation, University of Mainz.

Snedecor, G. W. & Cochran, W. G. (1967). *Statistical methods, 6th ed.* Iowa: The Iowa State University Press.

Wermuth, N. & Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *Journal of the Royal Statistical Society B, 52,* 21–75.

Yates, F. (1937). *The design and analysis of factorial experiments.* Harpenden: Imperial Bureau of Soil Science.