A New Model for Visualizing Interactions in Analysis of Variance

Patrick J.F. Groenen   Alex J. Koning[*]

February 13, 2004

Abstract

In analysis of variance, there is usually little attention for interpreting the terms of the effects themselves, especially for interaction effects. One of the reasons is that the number of interaction-effect terms increases rapidly with the number of predictor variables and the number of categories. In this paper, we propose a new model, called the interaction decomposition model, that allows to visualize the interactions. We argue that with the help of the visualization, the interaction-effect terms are much easier to interpret. We apply our method to predict holiday spending[1] using seven categorical predictor variables.

1 Introduction

In many situations of empirical research, there is the need to predict a numerical variable by one or more categorical variables. In its most simple case, the question is posed whether two groups of persons or subjects have a different mean or not, which can be answered by a simple t-test. Here, we call the numerical variable of interest is called the dependent variable and the grouping variable a predictor variable. If the number of groups is larger than two, we would like to know whether the means between the groups are all the

---

[*]Econometric Institute, Erasmus University Rotterdam, The Netherlands groenen@few.eur.nl, koning@few.eur.nl.

same or not. In this situation analysis of variance can be used. As soon as there are two or more categorical predictor variables, interaction effects may turn up. If so, then different combinations of the categories of the predictor variables have different effects. The majority of the papers in the literature only report an ANOVA table showing which effect is significant, but do not present or interpret the terms belonging to the effects themselves.

Each effect is characterized by a number of terms depending on the categories involved. For example, a main effect is characterized by K_j terms, where K_j is the number of categories of predictor variable j. In this paper, we argue that it is worthwhile to consider the terms that constitute the effects directly. Doing so may be difficult if the number of categories and the number of categorical predictor variables grows, because the number of interaction-effect terms will also grow dramatically. Therefore, we describe a new interaction decomposition model that allows two-way interactions to be visualized in a reasonably simple manner. The interpretation of the interaction plot is similar to that of correspondence analysis. Although in principle, the method could be used to analyze higher-way interactions, we limit ourselves to two-way interactions only.

The remainder of this paper is organized as follows. First, we discuss the empirical data set on predicting holiday spending in some detail and apply a common ANOVA. Then, we explain the interaction decomposition model more formally. We continue by applying our model to the holiday spending data set. We end this paper with some concluding remarks.

2   Holiday Spending Data

In this section, we discuss the empirical data set that we use in this paper. The data concern the holiday spending of 708 respondents. The data were gathered by students of the Erasmus University Rotterdam in 2003. The purpose of this research is to predict the amount of holiday spending in euros out of seven categorical predictor variables, that is, number of children, income, destination, other big expenses, accommodation, transport, and holiday length. Table 1 gives the frequencies of each of the categories of the predictor variables. Most categories are reasonably filled, with the exception of number of children, where almost 80% of the respondents has no children, whereas the other 20% has one to five children.

For travel agencies it is important to understand the relation between amount of money spent on a holiday and the predictor variables. With this knowledge, they can provide better suited arrangements for their clients. It can be expected *a priori* that the variable holiday spending is heavily skewed.
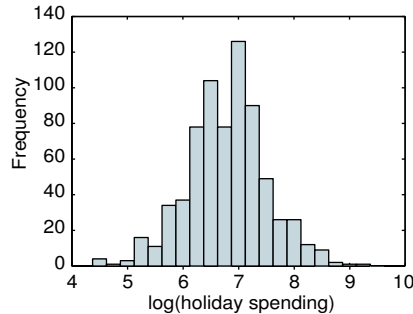
Figure 1: Histogram of the natural logarithm of the holiday spending.

The reason for this to happen is that there will always be a few people in a sample that are able to spend much more money on a holiday than the middle 50%. Such skewness is quite common in economics for variables like price, income, and spending in general. Cramér (1946, p. 220) remarks

> "Consider the distribution of incomes or property values in a certain population. The position of an individual on the property scale might be regarded as the effect of a large number of impulses, each of which causes a certain increase of his wealth. It might be argued that the effect of such an impulse would not unreasonably be expected to be proportional to the wealth already attained. If this argument is accepted, we should expect distributions of incomes or property values to be approximately log-normal."

To make the variable less skewed and more like a normal distribution, we take the logarithm of the holiday spending. The histogram of log holiday spending is given in Figure 1. Indeed, the logarithm transformation has made the variable behave much more like a normal distribution. Thus, throughout this paper, log holiday spending will be the dependent variable.

The most obvious method to investigate the relation between holiday spending and the predictor variables is analysis of variance (ANOVA). In essence, ANOVA compares the means of the dependent variable amongst the groups defined by predictor variables. ANOVA can be done in different ways. First, we can analyze the grouping for one predictor variable at a time. The disadvantage is that in our case there are seven different ANOVA's to be analyzed. Moreover, this approach imposes severe restrictions on the inferences to be drawn, especially so because the predictor variables form an unbalanced design. A better approach is to include all seven predictor variables simultaneously in the prediction, the so called main-effects model. In this manner, the predicted value of the dependent variable can be written

Table 1: Frequencies and percentages of categorical predictor variables for the holiday spending data.

Number of children	Freq	%
0 children	559	79.0
1 child	49	6.9
2 children	63	8.9
3 children	25	3.5
4 children	11	1.6
5 children	1	.1

Income	Freq	%
< 400 euro	75	10.6
400-800 euro	122	17.2
800-1600 euro	124	17.5
1600-3200 euro	237	33.5
3200-4000 euro	88	12.4
4000- euro	62	8.8

Destination	Freq	%
Within Europe	543	76.7
Outside Europe	165	23.3

Other big expenses	Freq	%
No big expenses	522	73.7
Other big expenses	186	26.3

Holiday length	Freq	%
< 7 days	65	9.2
7-14 days	277	39.1
14-21 days	223	31.5
21-28 days	88	12.4
> 28 days	55	7.8

Accommodation	Freq	
Camping	162	22.9
Apartment	189	26.7
Hotel	216	30.5
Other	141	19.9

Transport	Freq	%
By car	261	36.9
By airplane	377	53.2
Other transport	70	9.9

conveniently as a linear sum of terms for each of the categories of the predictor variables.

However, joint effects of two predictor variables are not taken into account. For our data, it may well be so that people with more children spend more and those who have longer holidays also spend more, but that the joint effect of having two or more children and have long holidays leads to less spending because a cheaper accommodation for example a camping site is chosen. Such effects are called *interaction* effects. In principle, it is possible to consider interaction effects between any number of predictor variables, but we limit ourselves to two-way interactions, that is, we only look joint effects of two predictor variables simultaneously. One important reason for doing so, is that interpreting three or higher way interactions gets increasingly more difficult.

Let us look at an ANOVA on the holiday spending of the main effects and all two way interactions of the predictor variables. Almost always, the results of the ANOVA are presented in an analysis of variance table (hence the name ANalysis Of VAriance) that shows how the sum of squares of the dependent variable can be decomposed into contributions by the main and interactions effects and whether these effects are significant or not. In Table 2, these results are presented for the holiday spending data. Note that the last column contains the partial η^2, a measure for the effect size. It measures the proportion of the variance of the dependent variable accounted for by the current factor.

From Table 2, we see that important contributors to the prediction of holiday spending are the main effects for 'Holiday length' and 'Number of children' and the interaction effects of 'Transport' by 'Holiday length', 'Transport' by 'Income', 'Holiday length' by 'Children', and 'Accommodation' by 'Income'. A few other main and interaction effects are significant and have reasonable effect size.

However, to understand *how* a certain main or interaction effect affects the prediction, one has to inspect the estimated terms of the effect. In this paper, we refer to a main effect as being the combination of terms belonging to the categories of a single predictor variable. A two-way interaction effect is the combination of terms belonging to all paired categories of two predictor variables. Considering our example, the number of parameters (and thus terms) to be considered depends on the number of categories per predictor variable. In total, there are $6 + 6 + 2 + 2 + 5 + 4 + 3 = 28$ terms for the main

Table 2: ANOVA table of all main effects and all interaction effects for the holiday spending data.

Source	Type III Sum of Squares	df	Mean Square	F	p	Partial η^2
Corrected Model	235.515(a)	172	1.369	5.584	.000	.642
Intercept	465.808	1	465.808	1899.622	.000	.780
Destination	1.451	1	1.451	5.917	.015	.011
Transport	.100	2	.050	.205	.815	.001
Holiday length	3.435	4	.859	3.502	.008	.026
Accommodation	.472	3	.157	.642	.588	.004
Big expenses	.850	1	.850	3.466	.063	.006
Children	3.084	5	.617	2.515	.029	.023
Income	4.510	5	.902	3.678	.003	.033
Destination * Transport	.361	2	.180	.736	.480	.003
Destination * Holiday length	.611	4	.153	.623	.646	.005
Destination * Accommodation	.972	3	.324	1.322	.266	.007
Destination * Big expenses	.535	1	.535	2.183	.140	.004
Destination * Children	1.151	3	.384	1.564	.197	.009
Destination * Income	1.333	5	.267	1.087	.366	.010
Transport * Holiday length	6.416	8	.802	3.271	.001	.047
Transport * Accommodation	2.554	6	.426	1.736	.110	.019
Transport * Big expenses	1.104	2	.552	2.251	.106	.008
Transport * Children	1.475	6	.246	1.002	.423	.011
Transport * Income	5.636	10	.564	2.299	.012	.041
Holiday length * Accommodation	1.492	12	.124	.507	.911	.011
Holiday length * Big expenses	1.471	4	.368	1.499	.201	.011
Holiday length * Children	7.340	13	.565	2.303	.006	.053
Holiday length * Income	9.693	20	.485	1.976	.007	.069
Accommodation * Big expenses	.201	3	.067	.273	.845	.002
Accommodation * Children	3.289	10	.329	1.341	.205	.024
Accommodation * Income	4.770	15	.318	1.297	.199	.035
Big expenses * Children	2.091	3	.697	2.842	.037	.016
Big expenses * Income	3.956	5	.791	3.227	.007	.029
Children * Income	2.928	12	.244	.995	.452	.022
Error	131.188	535	.245			
Total	33100.943	708				
Corrected Total	366.703	707				

effects and

$$
\begin{aligned}
6 \times 6 \; + \; 6 \times 2 \; + \; 6 \times 2 \; + \; 6 \times 5 \; + \; 6 \times 4 \; + \; 6 \times 3 \; + \\
+ \; 6 \times 2 \; + \; 6 \times 2 \; + \; 6 \times 5 \; + \; 6 \times 4 \; + \; 6 \times 3 \; + \\
+ \; 2 \times 2 \; + \; 2 \times 5 \; + \; 2 \times 4 \; + \; 2 \times 3 \; + \\
+ \; 2 \times 5 \; + \; 2 \times 4 \; + \; 2 \times 3 \; + \\
+ \; 5 \times 4 \; + \; 5 \times 3 \; + \\
+ \; 4 \times 3 \; = \; 327
\end{aligned}
$$

terms for the interaction effects summing to $28 + 327 = 355$ terms to be interpreted. Clearly, this amount of terms is too much to be interpreted. Certainly, one could choose to interpret only those main effects and interaction effects that are significant or have a high effect size, but still the number of terms to be interpreted can be quite large, especially if the number of categories of the predictor variables or the total number number of predictor variables increases.

Another problem with interactions terms may occur if the predictor variables form an unbalanced design, which generally is the case for nonexperimental data. Then, some of the interaction terms cannot be estimated due to the absence of relevant data.

Therefore, the reports in many studies are limited to an ANOVA table (as Table 2) only thereby ignoring the most important part of the analysis, that is, the terms of the effects themselves. The main purpose of this paper is discuss a new approach that allows to visualize the interaction effects directly. The main advantage is that the effects are easier to interpret. The next section discusses the new model more formally.

3 Decomposing Interactions

To express our interaction decomposition model more formally, we need to introduce some notation. Let y_i be the value of the dependent variable log 'Holiday spending' for subject i, where $i = 1, \ldots, n$. Suppose there are m categorical predictor variables. Each category can be presented as a dummy variable that equals one if subject i falls in that particular category and zero otherwise. The collection of all dummy variables for a single categorical predictor variable j can be gathered in the $n \times K_j$ indicator matrix $\mathbf{G}_j$, where K_j denotes the number of categories of variable j. For example, if $\mathbf{r}$ contains

the categories for 6 subjects, then $\mathbf{G}$ is given by

$$
\mathbf{r} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 3 \\ 1 \end{bmatrix} \implies \mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}
$$

Using the notation above, the main-effects model in ANOVA is given by

$$
y_i = c + \sum_{j=1}^{m} \sum_{k=1}^{K_j} g_{ijk} a_{jk} + e_i \tag{1}
$$

where c is the overall constant, a_{jk} is the main-effect term for category k of variable j, and e_i is the error in prediction for subject i. In matrix notation, (1) can be simplified by denoting row i of $\mathbf{G}_j$ by $\mathbf{g}'_{ij}$ and the main effects for variable j by $\mathbf{a}_j$, so that (1) becomes

$$
y_i = c + \sum_{j=1}^{m} \mathbf{g}'_{ij} \mathbf{a}_j + e_i \tag{2}
$$

The main-effects model in ANOVA determines the main effects $\mathbf{a}_j$ and the constant c in such a way that the sum-of-squares of the errors e_i is minimal, that is, it minimizes the loss function

$$
L_{\text{main}}(c, \mathbf{a}_1, \ldots, \mathbf{a}_m) = \sum_{i=1}^{n} \left(y_i - \left[c + \sum_{j=1}^{m} \mathbf{g}'_{ij} \mathbf{a}_j \right] \right)^2, \tag{3}
$$

where, for notational convenience, $\mathbf{a} = [\mathbf{a}'_1, \ldots, \mathbf{a}'_m]'$ contains all main-effects terms.

To specify an interaction effect between predictor variables j and l, consider the $K_j \times K_l$ matrix $\mathbf{B}_{jl}$ that contains all the terms of the interaction effect for all combinations of the categories of variables j and l. Suppose that subject i falls in category k of predictor variable j and in category s of predictor variable l. Then, the term of the interaction effect needed for this respondent is $b_{ks}^{(jl)}$, the element in row k and column s of $\mathbf{B}_{jl}$. For this person i, this element can also be picked by expression

$$
b_{ks}^{(jl)} = \mathbf{g}'_{ij} \mathbf{B}_{jl} \mathbf{g}_{il}.
$$

To make notation more compact, let all interaction effects be gathered in the symmetric partitioned block matrix

$$
\mathbf{B} = \begin{bmatrix}
\mathbf{0} & \mathbf{B}'_{12} & \dots & \mathbf{B}'_{1m} \\
\mathbf{B}_{12} & \mathbf{0} & \dots & \mathbf{B}'_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{B}_{1m} & \mathbf{B}_{2m} & \dots & \mathbf{0}
\end{bmatrix}.
$$

Note that the diagonal blocks are zero because $\mathbf{g}'_{ij}\mathbf{B}_{jj}\mathbf{g}_{ij}$ only selects the diagonal and thus estimates a main effect for variable j. Because main effects are already taken care of by $\mathbf{a}_j$, we choose the diagonal blocks $\mathbf{B}_{jj} = \mathbf{0}$, since it does not make sense to model a main effect twice.

Now, the ANOVA model with all main effects and all two-way interaction effects minimizes

$$
L_{\text{int}}(c, \mathbf{a}, \mathbf{B}) = \sum_{i=1}^{n} \left(y_i - \left[c + \sum_{j=1}^{m} \mathbf{g}'_{ij}\mathbf{a}_j + \sum_{j=1}^{m} \sum_{l=j+1}^{m} \mathbf{g}'_{ij}\mathbf{B}_{jl}\mathbf{g}_{il} \right] \right)^2. \tag{4}
$$

The number of parameters to be estimated is 1 for the constant, $\sum_{j=1}^{m} K_j$ for the main effects in $\mathbf{a}$, and $\sum_{j=1}^{m} \sum_{l=j+1}^{m} K_j \times K_l$ for the interaction effects in $\mathbf{B}_{jl}$. Note that some additional constraints are necessary to avoid that interaction effects pick up main effects and the main effects pick up the constant effect. The constraints often imposed are the sum of each of the main effects $\mathbf{a}_j$ to be equal to zero and the interaction effects $\mathbf{B}_{jl}$ must have row and column sums equal to zero.

We now turn to the interaction-decomposition model proposed in this paper. The key idea of this model is that the interaction terms $\mathbf{B}_{jl}$ are constrained such that an easy graphical representation is possible. The type of constrained used in the interaction-decomposition model is that of common rank-reduction, that is, we require that

$$
\mathbf{B}_{jl} = \mathbf{Y}_j\mathbf{Y}'_l, \tag{5}
$$

where the $K_j \times p$ matrix $\mathbf{Y}_j$ has rank not higher than $p > 0$. Equivalently, we may write that in the interaction-decomposition model an interaction term of category k of predictor variable j and category s of predictor variable l is given by

$$
b_{ks}^{(jl)} = \mathbf{y}'_{jk}\mathbf{y}_{ls},
$$

where $\mathbf{y}'_{jk}$ denotes row k of $\mathbf{Y}_j$. Using this kind of constraint, the interaction term is graphically represented by a projection of the vector $\mathbf{y}'_{jk}$ onto $\mathbf{y}'_{ls}$.

Thus, high projections indicate large interaction terms, and small projections indicate small interaction terms. Such a rank restriction is also used in bi-additive models such as correspondence analysis, multiple correspondence analysis, and joint correspondence analysis. The interaction-decomposition model has as its main advantage that there is only a single vector $\mathbf{y}'_{jk}$ to be estimated for each category of a predictor variable. In other words, the number of interaction parameters to be estimated only grows linearly with the number of categories, and not quadratically as for the unconstrained ANOVA interaction model in (4).

Because projections do not change under rotation, the vectors in $\mathbf{Y}_j$ are determined up to a common rotation that is the same for all $\mathbf{Y}_j$. Other than the rotation indeterminacy, the $\mathbf{Y}_j$'s and thus the interactions $\mathbf{B}_{jl} = \mathbf{Y}_j \mathbf{Y}'_l$ can be estimated uniquely if the number of dimensions is low enough. This contrasts with standard ANOVA that cannot estimate all interaction terms, if not all combinations of predictor categories are observed.

Obviously, because the interaction-decomposition model imposes constrains on the interactions, it will generally not fit as well as the unconstrained ANOVA interaction model in (4).

Note that we still have to require that $\mathbf{B}_{jl}$ has zero row and column sum to avoid confounding of the main and interaction effects. This restriction implies that each $\mathbf{Y}_j$ must have column sum zero so that indeed $\mathbf{B}_{jl} = \mathbf{Y}_j \mathbf{Y}'_l$ will have row and column sums equal to zero. To fit the interaction-decomposition model, we have developed a prototype in MatLab that minimizes $L_{\text{int}}(c, \mathbf{a}, \mathbf{B})$ subject to the constraints (5) by alternating least-squares. The details of this algorithm are beyond the scope of this paper and will be published in subsequent papers.

For two categorical predictor variables, similar decomposition models were proposed in the literature (Eeuwijk, 1995; De Falguerolles & Francis, 1992; Gabriel, 1996; Choulakian, 1996). In psychometrics, such a models has been known under the name FANOVA (Gollob, 1968). A different way for modelling three-way interactions among three categorical predictors was presented by Siciliano and Mooijaart (1997). The difference with our approach is that we limit ourselves to two-way interactions only. We believe that two way interaction convey the most important information while they are still reasonably easy to interpret. For three or higher way interactions, the interpretation becomes far more difficult. Another difference with previous models in the literature is that our model is that the interaction-decomposition model is not limited to two or three we categorical predictor variables, but we can handle any number of predictors.

The current model has some resemblance to joint correspondence analysis (Greenacre, 1988) and multiple correspondence analysis (see, for example,

Greenacre, 1984; Gifi, 1990). Similar to joint correspondence analysis, the diagonal effects of $\mathbf{B}_{jj}$ are simply discarded. However, the main difference lies in the fact that the main effects in joint correspondence analysis are not linearly modelled by separate terms $\mathbf{g}'_{ij}\mathbf{a}_j$, but are included as weights in the loss function. A minor difference consists in different normalization of the $\mathbf{Y}_j$.

4   Interaction Decomposition of Holiday Spending

Let us apply the interaction decomposition model to the holiday spending data. First, we consider the main effects for all the categories (see Figure 2). The most striking feature is the large main effect of '5 children'. From Table 1 we know that there was just a single family with 5 children in this data set. Apparently, this family has spent much more money on their holiday in comparison to the rest, because it has a large positive main effect that predicts more money to be spent during the holiday. Here we choose to interpret only effects larger than plus or minus .2. Since the dependent variable is the logarithm of holiday spending, effects larger than .2 imply that the holiday spending goes up by a factor $\exp(.2) = 1.22$ thus increases by 22%. Apart from the category '5 children', we see that more money is spent if the travel takes place by airplane, for holidays longer than 21 days, and if the income is 4000 euros or higher. On the other hand, holiday spending reduces for holidays shorter than 7 days and –to a lesser extent– from 7 to 14 days, made by car, if there are zero to four children, and if the income is lower than 800 euros. The other main effects are reasonably small, suggesting that they are not very important.

The interaction effects are more interesting, because we can investigate the joint effects of two predictor variables. Figure 3 shows the plot of the interaction effects by the interaction decomposition model. Panel (a) shows all effects simultaneously and Panel (b) zooms in on the center part in the box in Panel (a). The basic way to interpret the interaction solution is as follows. Condition on a single category of interest. Then project all categories of other variables onto this vector. High projections indicate a high interaction effect (thus more money spent during the holiday), high negative projections indicate a high negative interaction effect (thus less money spent during the holiday). In addition, vectors that are exactly orthogonal, have a zero interaction effect. Note that a reasonable interaction effect can still occur for vectors that are almost orthogonal if one or both of the vectors
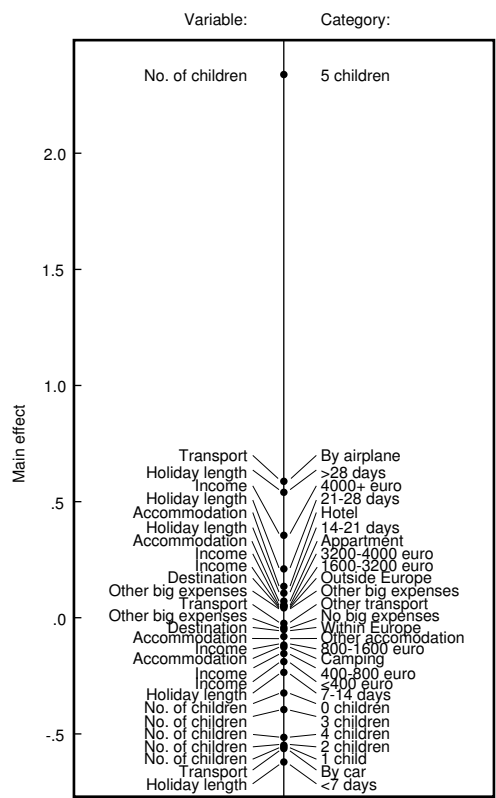
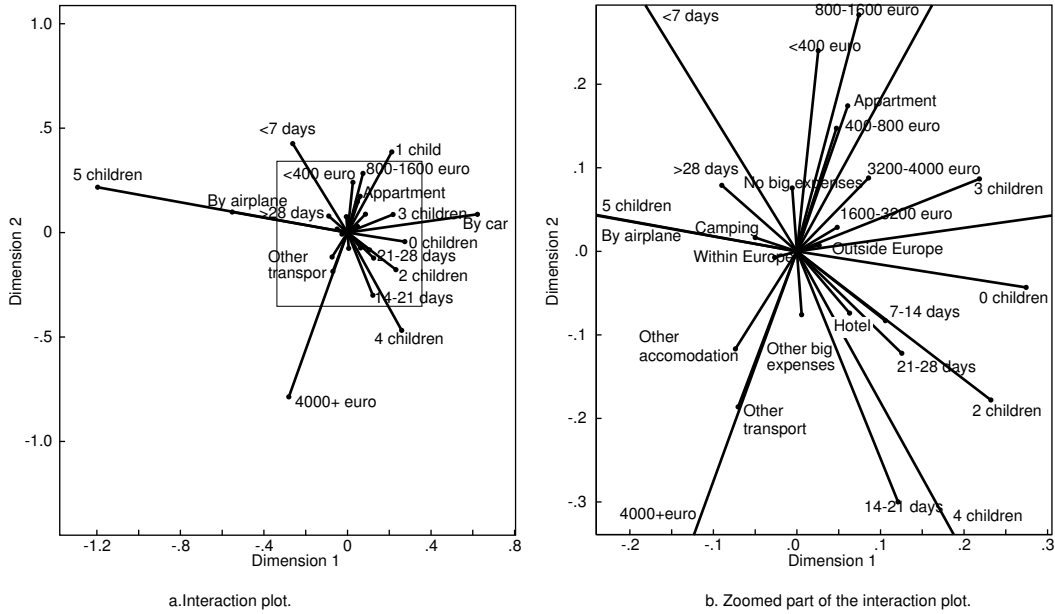Figure 2: Main effects with respect to the overal mean c.

Figure 3: Interaction plot for the interaction decomposition model on the holiday spending data. Panel (b) zooms in on the part of the box in Panel (a) to view the labels more clearly.

are long enough. An example of this case is the interaction effect between '5 children' and Income '4000+ euro', which is still reasonable. Also, long vectors generally have larger interaction with all other categories.

Note that variables with only two categories will have equal length vectors that are mirrored in the origin. In Figure 3b, we see an example of this case for variable 'Other big expenses' (with categories 'No big expenses' and 'Other big expenses'). The reason for two equal length and opposite vectors lies in the restriction that the coordinates have zero column sum per variable. For a two category variable, this restriction implies equal but mirrored vectors for the categories.

To see how high income families spend their holiday money, we have projected the categories of other variables onto the vector of category Income '4000+ euro' in Figure 4. Thus, we are considering interaction terms conditioned on the category Income '4000+ euro'. These conditional effects are also presented separately in Figure 5 together with the main-effect term of about .35 for category Income '4000+ euro'. The interaction decomposition model predicts that holiday spending increases for for this income group if the number of children is 4 or 5, the holiday length is 14 to 21 days, and the accommodation is 'Other'. Compared to the main effect of Income '4000+
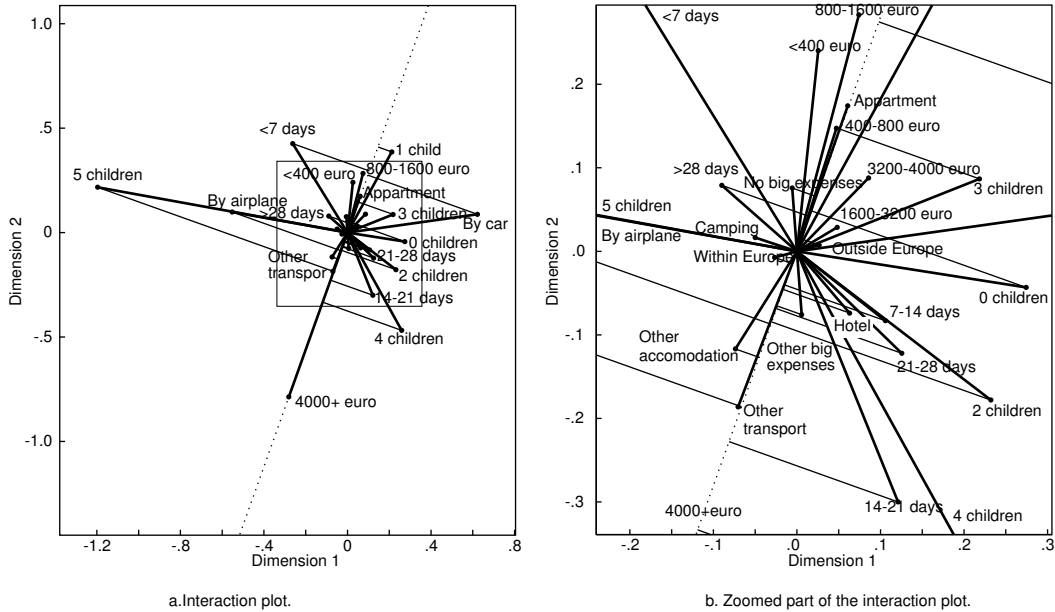
13

Figure 4: Interaction decomposition plot with projections onto the category 'Income 4000 or more'. Panel (b) shows the part in the box of Panel (a) to present the labels more clearly.

euro', the holiday spending decreases if there are only one or three children, the holiday length is less than seven days, the trip is made by car, and the accommodation is an appartement. In general, a plot as in Figure 5 is very helpful in interpreting the interaction terms conditioned on a certain category.

The most important interactions can also be derived directly from Figure 3 by looking at the largest vectors. The categories that matter most in the interactions are families of 0, 1, 4, and 5 children, holidays shorter than 7 days, having one child, using the car or the airplane as means of transportation. For example, for holidays shorter than 7 days, the model predicts higher spending if there is one child (because the vectors project highly) and lower spending if there are four children (because the vectors project negatively). Logically, holiday spending also increases for holidays shorter than 7 days made by air plane, but decreases if the holiday is done by car.

It is certainly possible to describe more interaction effects predicted by the interaction decomposition model. Of course, we have to keep in mind that the solution does not fit perfectly. Therefore, it is a compromise solution that tries to show as much as possible of the relations present in the data.
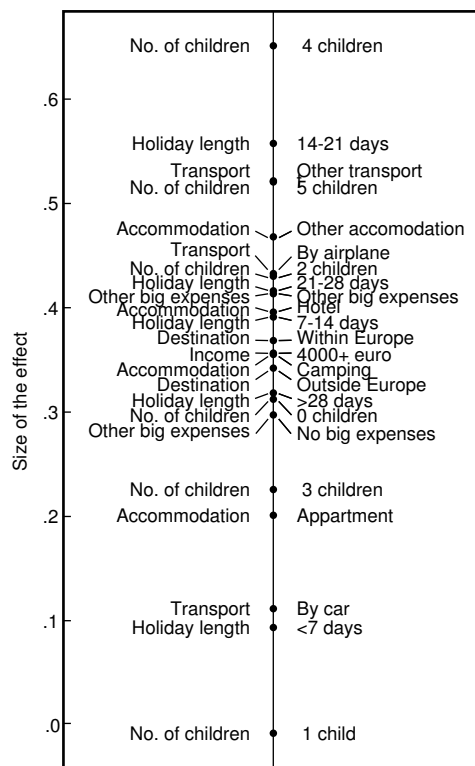
Figure 5: Resulting interaction effects for the category 'Income 4000 or more' with the categories of all other categories. The main effect for this category is also presented.

5   Conclusions

To investigate interactions in the traditional ANOVA framework, most researchers limit themselves to an ANOVA table. We have argued in this paper that it is worthwhile to study the interaction term themselves. Because the number of interaction terms increases rapidly with the number of categorical predictor variables and the number of categories per variable, we have proposed a new model, called the interaction decomposition model, that allows to visualize the interactions.

In principle, the current model can be extended to Generalized Linear Models (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). It remains to be seen if the present model needs to be adapted or not.

References

Choulakian, V. (1996). Generalized bilinear models. *Psychometrika, 61*, 271–283.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, N. J.: Princeton University Press.

De Falguerolles, A., & Francis, B. (1992). Algorithmic approaches for fitting bilinear models. In Y. Dodge & J. Whittaker (Eds.), *Compstat 1992* (p. 77-82). Heidelberg: Physica-Verlag.

Eeuwijk, F. A. (1995). Multiplicative interaction in generalized linear models. *Biometrics, 85*, 1017–1032.

Gabriel, K. R. (1996). Generalised bilinear regression. *Biometrika, 85*, 689–700.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester: Wiley.

Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika, 33*, 73-116.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* New York: Academic Press.

Greenacre, M. J. (1988). Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika, 75*, 457–467.

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

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A, 135,* 370–384.

Siciliano, R., & Mooijaart, A. (1997). Three-factor association models for three-way contingency tables. *Computational Statistics and Data Analysis, 24,* 337–356.