CORRESPONDENCE ANALYSIS USED COMPLEMENTARY TO
LOGLINEAR ANALYSIS

PETER G. M. VAN DER HEIJDEN
DEPARTMENT OF METHODOLOGY
UNIVERSITY OF LEIDEN

JAN DE LEEUW
DEPARTMENT OF DATA THEORY
UNIVERSITY OF LEIDEN

Loglinear analysis and correspondence analysis provide us with two different methods for the decomposition of contingency tables. In this paper we will show that there are cases in which these two techniques can be used complementary to each other. More specifically, we will show that often correspondence analysis can be viewed as providing a decomposition of the difference between two matrices, each following a specific loglinear model. Therefore, in these cases the correspondence analysis solution can be interpreted in terms of the difference between these loglinear models. A generalization of correspondence analysis, recently proposed by Escofier, will also be discussed. With this decomposition, which includes classical correspondence analysis as a special case, it is possible to use correspondence analysis complementary to loglinear analysis in more instances than those described for classical correspondence analysis. In this context correspondence analysis is used for the decomposition of the residuals of specific restricted loglinear models.

Key words: correspondence analysis, data analysis, loglinear analysis, multi-dimensional scaling.

1. Introduction

For the analysis of contingency tables, loglinear analysis is already a very popular technique in the English speaking countries. Standard references are Andersen (1980), Bishop, Fienberg and Holland (1975), and Fienberg (1980). In the last few years there is a growing interest in correspondence analysis, which has been the most important data analytic technique in France for many years. When one has full command of the French language, the basic works are those of the group around Benzécri (1973, 1980). In the English speaking world the growing interest is apparent from works written by de Leeuw (1984), Gifi (1981), Greenacre (1984), and Nishisato (1980). Apart from these books, the number of articles and contributions at conferences is growing rapidly.

Strangely enough, correspondence analysis was already known in the English literature for a long time be it under several other names. Nishisato (1980) gives a full survey of all these names and references which have appeared in the history of correspondence analysis. Greenacre (1984) accentuates that these other approaches have a different rationale and interpretation. He discusses this for the approaches “reciprocal averaging,” “dual (or optimal) scaling,” “canonical correlation analysis,” and “simultaneous linear regressions.” The recent flourishing of correspondence analysis as a data analytic technique is probably due to the heavy emphasis on the geometrical aspects of the method. Canonical correlation analysis of categorical data (Kendall & Stuart, 1973, pp. 588–598), which is

Requests for reprints should be sent to P. G. M. van der Heijden, Department of Methodology, University of Leiden, Hooigraacht 15, 2312 KM Leiden, THE NETHERLANDS.
proved by de Leeuw (1971) to be formally identical to correspondence analysis, emphasizes the quantification aspect.

With the growing interest in correspondence analysis, there is a natural interest in relations between correspondence analysis and loglinear analysis. Until now, the literature on this topic can be divided into two approaches. The first approach deals with the RC-model (RC for "row, column") of Goodman (1979, 1981a, 1981b) and generalizations of this model (Agresti, 1983; de Leeuw, 1983). The RC-model is a model for a two-way contingency table in which the loglinear interaction parameter has a multiplicative form. Goodman (1981b) shows that, when the frequencies in a two-way table are generated from an underlying discretized bivariate normal distribution (or a bivariate distribution which is bivariate normal after a proper transformation), the estimates of the multiplicative row and column parameters are approximately the same as the row and column scores found for the first dimension in correspondence analysis. See also Fienberg and Meyer (1983) and Israëls and Sikkel (1982) for other relations between the RC-model and correspondence analysis.

Contributions in the second approach discuss the complementary use of correspondence analysis and loglinear analysis. Examples are Daudin and Trécourt (1980) and Israëls and Sikkel (1982). They conclude that loglinear analysis is the method most apt to trace higher-order interactions, although it is recognized that the interpretation of these interactions is often difficult. When the number of variables is small, loglinear analysis can be complemented by correspondence analysis. In these two papers complementary use of the techniques is advocated only in a general way. It is not made explicit how exactly these methods can be used in a complementary way. It is our purpose to propose more specific ways of combination in this paper.

First we give a short introduction to correspondence analysis. Following the French tradition, heavy emphasis will be placed upon the geometrical aspects of the technique. Correspondence analysis as such is a technique for the analysis of two-way tables. We will discuss the most usual way to analyse higher-way tables with correspondence analysis. We will not deal with correspondence analysis of higher-way contingency tables by means of the so-called Burt matrix (see Gifi, 1981, pp. 134–162; Greenacre, 1984, chap. 5). In section 3 we will discuss loglinear analysis. Special attention is paid to the case of the three-way contingency table. In section 4 we will present our main results on the complementary relation of correspondence analysis and loglinear analysis. It will be shown that the correspondence analysis solution is based on the decomposition of the difference between two loglinear models, which will be specified there. The results will be presented for two-way, three-way and higher-way contingency tables respectively. Our results will be illustrated by an example. A three-way table on suicide behavior is analyzed. In section 5 a generalization of correspondence analysis is discussed. The generalization is used to make it possible to use correspondence analysis complementary to loglinear analysis in more cases than these discussed in section 4. We will end with some conclusions.

2. Correspondence Analysis of Contingency Tables

We will treat correspondence analysis briefly here, with an emphasis on the geometrical aspects. For details and proofs we refer to Gifi (1981) or Greenacre (1984), and the references mentioned there.

Correspondence analysis is a technique with which it is possible to find a multi-dimensional representation of the dependence between the row and column variable of a two-way contingency table. This representation can be constructed using the scores found for row and column categories as coordinates for the category points. These scores can be normalized in such a way that distances between row points or between column points in euclidian space are equal to chi-squared distances. By emphasizing the chi-squared dis-