Preface

The technique of correspondence analysis is becoming increasingly popular in many areas of data analysis. In saying this, we have to realize that correspondence analysis is already quite old and that, curiously enough, the popularity of the technique differs greatly from country to country. It has been pointed out by de Leeuw (1983) that Pearson already discovered some of the basic facts connected with correspondence analysis in 1907, while Fisher was the first person who applied correspondence analysis (1940), and who advised others to apply it (Maung, 1941). At about the same time the technique of multiple correspondence analysis was proposed by Guttman (1941). Neither Fisher nor Guttman used the term correspondence analysis, of course, and in fact they did not refer to each other or to earlier work on the same subject. This state of affairs continued when correspondence analysis was rediscovered in the fifties by Burt (1950). Hayashi (1952) introduced the technique in Japan, where it became fairly popular, again under a new name, and with only a few references to the earlier work of Fisher and Guttman. It is not difficult to find at least ten references in the English-speaking applied statistics literature since 1955 in which correspondence analysis is rediscovered, and presented as something completely new. This has led to a great deal of confusion, and a generally quite unsatisfactory state of affairs. The major reason for the amount of disarray in the older literature is perhaps that the technique did not really catch on. Although everybody who studied the problems for which the technique was designed seemed to agree that it provided a very reasonable solution, the computations were simply not feasible until mainframe computing came along in the sixties. Just as factor analysis had its centroid method, whose only justification was its computational simplicity, the scaling of categorical variables had the Likert method, which also could be applied by hand.

The breakthrough came, surprisingly enough, in France. In the mid-sixties Benzécri reinvented the technique, and baptized it 'correspondence analysis'. Because of the peculiar situation in French statistics it rapidly became very popular in France, and in the mid-seventies data analysis and statistics were virtually identical there with correspondence analysis. There are still many French publications in which log-linear analysis or survival analysis is explained by showing how it relates to correspondence analysis. The French language, the highly idiosyncratic style of Benzécri's publications, and the somewhat sectarian and intolerant approach of French Benzécrists to the rest of statistics, did not help very much in spreading the technique to other countries.

This changed in the seventies, for various reasons. Guttman's weighting technique (also known as multiple correspondence analysis) was unearthed by psychometricians such as Lingoes, McDonald, and McKeon. It was now feasible computationally, and it consequently became more interesting. More recently the French approach to correspondence analysis began to attract attention. It has been discussed very completely and competently in the English literature by Greenacre (1984). The books of Lebart, Morineau, and Warwick (1984) and of Jambu (1983) have been translated into English, and there are a large number of papers in statistical journals which compare the French approach to data analysis with the Anglo-American approach (this usually means log-linear analysis). The Japanese contributions were reviewed very ably by Nishisato (1980). By now, classical correspondence analysis and multiple correspondence analysis are quite well known, and anybody who is interested can easily find the relevant literature. A very complete historical review is available in van Rijckevorsel (1987), and the Anglo-French comparisons are summarized and extended in van der Heijden (1987).

But what about optimal scaling? Because these techniques are even more linked with mainframe computing, they are less ancient. In fact one could say that they were hardly thinkable before the computer existed, because they are defined mainly in algorithmic terms. The first optimal scaling techniques were introduced by Shepard and Kruskal in the early sixties. These were actually multidimensional scaling techniques, in which similarity data were related to distances, and least squares loss functions indicating fit were minimized by gradient techniques. It became apparent rather soon that the process of fitting a non-linear model to the data and the process of transforming the data were two components which could be dealt with quite separately from each other. Kruskal, Roskam, and Lingoes wrote families of programs which combined different linear and non-linear models with the idea of optimal scaling (also called optimal quantification or optimal transformation). The separation of the fitting and scaling components of the non-linear techniques was carried out to its logical consequences by de Leeuw, Young, and Takane in the mid-seventies. They used alternating least squares methods to minimize the least squares loss functions, which means that the two algorithmic components were actually completely separate in the program. An example may clarify this. Take a least squares loss function, for example the one used in multiple regression. We choose a particular scaling or transformation of the variables and minimize the function over the model parameters, i.e. the regression coefficients. In the next step we choose transformations from the class of feasible transformations to minimize this very same loss function, keeping the regression coefficients fixed at their current values for the moment. And we alternate these two different minimization steps until convergence. This approach, reviewed by Young (1981), has produced a number of computer programs which can deal with many of the classical multivariate analysis models, and which allow for optimal transformation of each the variables at the same time (where the class of feasible transformation can be different for different variables, i.e. some variables can be transformed monotonically, some polynomially, some arbitrarily, and so on). More recently similar methods have been developed by Breiman and Friedman (1985) and their students, using a particular class of smooth transformations.

The first one to systematically exploit the relationship between correspondence analysis and optimal scaling was de Leeuw (1973). He reviewed almost all of the older literature, and he started a large research program with the explicit purpose of combining these two classes of techniques into a number of useful computer programs, that could be applied profitably in many of the sciences. This resulted in the book by Gifi (1981a, re-issued in 1988) and in a whole series of publications in various books and journals. In this approach multiple correspondence analysis is taken as the starting point to develop a whole class of non-linear multivariate analysis methods, of which the classical methods such as principal component analysis, canonical correlation analysis, multiple regression analysis, discriminant analysis, and path analysis are special (linear) cases. All these special cases are defined by imposing restrictions on the parameters of the multiple correspondence analysis problem. In Gifi the basic idea is to scale the individuals or objects (i.e. to map them into low dimensional Euclidean space) in such a way that individuals with similar response profiles are relatively close together, while individuals with different response profiles are relatively far apart. This is called homogeneity, and it is defined explicitly in terms of Euclidean distances. The emphasis in Gifi (1981a, 1988) is, as in the French literature, on the geometry of the problem, but the ideas are also close to the similarity representations of multidimensional scaling. The restrictions that define principal component analysis as a special case of multiple correspondence analysis, for example, also have a clear geometrical meaning in the space that scales the individuals. This is in contrast with the approach taken by Young (1981) or Breiman and Friedman (1985), who take optimal scaling as their starting point, and emphasize algebraic and analytic properties of solutions.

The authors of the various chapters of this book work in the tradition of the Gifi approach, but they extend the earlier results in many directions. This is most apparent in the first chapters, which are introductory. Bekker and de Leeuw discuss both multiple correspondence analysis (or homogeneity analysis, as it is called by Gifi) and the optimal scaling technique called non-linear principal component analysis. They then proceed to show that in various important special cases these two techniques are closely related, and actually produce the same solutions, although in a quite different form. These special cases are idealized, and not likely to occur exactly in practice, but it turns out that many actual empirical data sets do behave in approximately the same way as these mathematical gauges. This chapter clearly explains the occurrence of the famous horseshoe or Guttman effect in multiple correspondence analysis, and it shows that at least in some applications non-linear principal component analysis is more directly interpretable in terms of the parameters of the original problem.

One of the reasons for going beyond homogeneity analysis is that it is sometimes necessary to analyse continuous variables. In the book by Gifi continuous variables are often discretized in a small number of categories, and this can entail a substantial loss of information. One way of dealing with continuous variables is to use coding schemes that are directly related to classes of feasible transformations defined by splines. This theme is taken up by van Rijckevorsel in Chapter 2. He discusses the various forms of fuzzy coding proposed mainly in the French literature, both in their analytical and geometrical aspects. Splines are a particular form of fuzzy coding.

In the third chapter de Leeuw and van Rijckevorsel take the geometrical approach of Gifi a few steps further, still working in the space in which the individuals are represented, but modifying the definition of homogeneity in various ways. The geometry of using spline functions is explained in this chapter.

In the first French contribution, Besse explains why classical multiple correspondence analysis and principal component analysis are not appropriate for a very common format of continuous data: replicated time series data. Using ideas from functional analysis it is possible to define various alternative metrics which do justice to the fact that the variables are ordered in time. This can be interpreted as yet another variation on the theme of how homogeneity should be defined, and which restrictions are appropriate for which problems. Again splines play an important role in this treatment of dynamic component analysis. Because of the availability of a large variety of transformation functions and/or metrics, the classical question in PCA of how many components should be retained, is modified into which metric, transformation function, filter, . . . etc. leads to the smallest number of well-separated components? Besse illustrates various aspects of this problem.

The last two chapters have a more statistical content. The statistical interpretation of random errors, smoothing or density estimation is one way of incorporating prior information in data analysis. Martin is one of the French statisticians who introduced fuzzy coding. He gives a related probabilistic interpretation, called probability coding or P-coding, in Chapter 5. And analogously to Besse and Winsberg this interpretation of fuzzy coding can also lead to kernel estimation.

In Chapter 6 Winsberg discusses two different approaches to component analysis. First, optimal scaling, which uses the monotone splines introduced by Ramsay and Winsberg into the data analysis literature, is embedded in a more classical statistical framework. This can be seen as bridging another gap: the one between explorative multivariate data analysis and between confirmatory analysis based on statistical modelling, or between confirmatory analysis based on representing the actual data at hand in a pleasing way and the statistical approach of using the likelihood to arrive at inductive generalizations. Besse's procedure for incorporating functional relationships through metrics can also be implemented by using filters on sampled smooth functions. This procedure, originally introduced by Kruskal and Winsberg, is contrasted by Winsberg in the second part of Chapter 6 with the more elaborate approach of Besse and Ramsay.

We do not think, by the way, that it is necessary for us to choose an explicit location on the infamous exploratory–confirmatory continuum. Our point of view is simply that if there is trustworthy prior information available, then it should be incorporated in the data analysis technique, because it will reduce sampling variability of the representation. This prior information can be in the form of a statistical model, but also in the form of restrictions on the correspondence analysis representation. We do not think that the model should always precede the technique. In many situations with which we are familiar there is very little prior information. Imposing restrictions in such situations should give the data analyst a bad conscience, and can lead him rather far astray because he pays for the reduced variability with a large bias, and because he is making confirmatory statements which are at least partly based on figments of his imagination. Choosing between overparametrization and chance capitalization on one side, and between underparametrization and spurious precision on the other side, is a painful process. On the continuum of the various statistical techniques there seems to be an important range where correspondence and component analysis are suitable techniques, which show researchers results in a form in which they want to see them.

Acknowledgements

Since the *initial* planning for this monograph a number of individuals and organizations took part. The essential idea for a book was conceived at a workshop on PCA and splines at the Erasmus University Rotterdam in 1984. The participants were Besse, de Leeuw, Martin and van Rijckevorsel. Valuable assistance in organizing this workshop was given by Dr H. Coolen, and facilities and some funds were provided for by the Erasmus University.

Furthermore The Netherlands Service for Occupational Health and Safety in the Hague offered administrative facilities from 1984–1987 while getting the book organized. Finally the cooperation of the secretarial staff of the Department of Datatheory in typing some of the drafts has been extremely gratifying. We would like to thank especially Marie Verdel for shepherding the manuscript through. Dr P. Besse generously provided his expertise in a particular area, apart from his own contribution. Prof. H. Caussinus smoothed some communication problems.

THE EDITORS