KM 42 (1993)
pg.9-18

EDITORIAL

THE ANALYSIS OF DICHOTOMOUS PREFERENCE DATA: MODELS BASED ON CLYDE H. COOMBS' PARALLELOGRAM MODEL

HERBERT HOIJTINK

Abstract

This paper gives an introduction to parallelogram analysis according to Coombs. It will be argued that due to the deterministic nature of this model it is hard to apply in empirical research. Coombs suggested three kinds of models that circumvent this problem: latent class models, parallelogram models and latent structure models. A short introduction to these models will be given. Furthermore it will be indicated to which class of models each of the models to be presented in this issue of Kwantitatieve Methoden belongs.

Key words: parallelogram analysis, unfolding, latent class analysis, latent structure analysis, item response theory.

Requests for reprints should be send to Herbert Hoijtink, Department of Statistics and Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.

Note that the software associated with some of the models to be presented in this issue of Kwantitatieve Methoden is available from software-house iecProGAMMA, P.O. Box 841, 9700 AV Groningen, The Netherlands (UNFOLD, MUDFOLD, and PARELLA).

1. Introduction

In 1964 Clyde H. Coombs' book 'A Theory of Data' was published. In this book, Coombs distinguishes different types of data, proposes methods for the analysis of these data, and discusses the rationale behind these methods. In the past decennium several authors have contributed to the further development of one of Coombs' models: the parallelogram model for the analysis of pick any/n data (Coombs, 1964, Chapter 15).

This issue of Kwantitatieve Methoden provides an overview of these developments. Each paper in this issue will discuss: the relation between the model presented and Coombs' parallelogram model; the most important features of the model presented; and, the application of the model to two empirical data sets, one concerning the measurement of the attitude with respect to nuclear power stations (Formann, 1988), and one concerning the measurement of the attitude in the car-environment issue (Doosje and Siero, 1991). In the final paper by Wijbrandt van Schuur, the models presented will be compared, and similarities and differences will be discussed.

2. Coombs Parallelogram Model for the Analysis of Pick any/n Data

Parallelogram models can be used by researchers who need measurements of latent person characteristics such as attitudes and preferences. Since the characteristic to be measured is latent, the actual measure has to be inferred from a person's responses to a set of items that are indicative of the trait to be measured. For example, in Table 1 five items are presented that are indicative of the attitude towards nuclear power stations. Item 1 expresses a positive attitude towards nuclear energy, item 2 expresses a neutral attitude, items 3 and 4 a slightly negative attitude, and item 5 a negative attitude. I.e. the items can be ordered along the latent characteristic from positive to negative.

In order to be able to apply parallelogram analysis, the responses of persons to items have to be dichotomous, with 1/0 denoting a response that can be labelled positive/negative or agree/disagree. In Table 2 the responses of 600 persons to the five items indicative of the attitude towards nuclear energy are presented.

Table 1
Items Indicative of the Attitude towards Nuclear Power Stations, Keywords/Numbers Used in some of the Papers for Identification of these Items are in Brackets

Item	Phrasing
1.	In the near future alternative sources of energy will not be able to substitute nuclear energy (NOALT/1).
2.	It is difficult to decide between the different types of power stations if one carefully considers all their pros and cons (DIFFDE/2).
3.	Nuclear power stations should not be put into operations before the problems of radio-active waste have been solved (PROBSOL/3).
4.	Nuclear power stations should not be put in operation before it is proven that the radiation caused by them is harmless (SAFEPRO/4).
5.	The foreign power stations now in operation should be closed (CLOSFOR/5).

As can be seen, 6 persons indicated agreement only with the first item, 18 persons indicated agreement with the first three items, etc. The meaning of 'pick any/n data' is hereby explained: there are no restrictions on the number of items a person is allowed to agree with, a person is allowed to pick any of the five items.

In the same way as the items can be ordered from positive to negative, the persons can be ordered from having a positive attitude towards nuclear power stations to having a negative attitude. Persons responding 10000 have a positive attitude, they are certain that nuclear power stations are necessary and have no doubts whatsoever. Persons responding 00110 have a slightly negative attitude. They recognize that nuclear power stations may be useful, but should not be put into operation before certain problems have been solved. Persons responding 00001 are opposed to any use of nuclear power stations.

Applying Allen and Yen's (1979, p2.) definition of measurement to parallelogram analysis it would become: 'the assignment of locations to persons and items such that the relation between the locations reflects the relation between the persons and the items'. This can be interpreted as: the closer persons and items are located on the latent characteristic of interest, the smaller the psychological distance between the attitude/ preference of the person and the content of the item.

Not all persons are as easily ordered as those in the examples given above. A person responding 10001 is neither located on a small (psychological) distance from item 1 (such a person would not respond positively to item 5), nor at a small (psychological) distance from item 5 (such a person would not respond positively to item 1). In the sequel it will be illustrated that where Coombs' original parallelogram model is not able to model such deviant response patterns, the models to be presented in this issue have less difficulties in doing so.

Table 2
Response Patterns and Observed Frequency

Resp.	Freq.	Resp.	Freq.
10000	6	11000	3
11100	18	01100	22
01110	65	00111	118
00011	39	00001	11
00000	3	01000	6
00100	14	00010	3
10100	22	10010	3
01010	4	01001	2
00110	41	00101	1
11010	2	10110	37
10101	2	10011	5
01101	1	01011	16
11110	52	11101	2
11011	2	10111	15
01111	61	11111	24

As indicated above, parallelogram analysis can be of help whenever researchers need measurements of latent person characteristics. Measurements are obtained if the responses of persons to items, are related to the locations of persons and items on the latent characteristic of interest. In his parallelogram model Coombs specifies the following relation:

$$X_{ai}=1 \quad \text{if} \quad |\beta_a-\delta_i| \le \tau,$$

and, (1)

$$X_{ai}=0 \quad \text{if} \quad |\beta_a-\delta_i| > \tau,$$

where X_{ai} denotes the response of person a=1,...,N to item i=1,...,n, β_a denotes the location of person a, and δ_i the location of item i. The parameter τ is a threshold parameter. If the distance between person and item is smaller than τ, a person will give the positive response, otherwise the person will give the negative response. Coombs also considered models where τ is item specific, person specific, occasion specific or a combination of these. Since most of the models to be discussed in this issue are based on a fixed threshold, models with a variable threshold will not be discussed any further.

If Coombs' parallelogram model provides an adequate description of the data, i.e. if (1) provides an adequate description of the response proces, the positive responses will resemble a parallelogram structure if both persons and items can be ordered according to their location on the latent characteristic of interest (see Table 3 for an example with 6 persons and 5 items). With such a response structure, the origin of the name parallelogram is obvious.

Table 3
Example of Parallelogram Structure for a Five Item/Six Person Data Set

Response Pattern
10000
11100
01110
01111
00011
00001

3. New Developments with respect to Coombs' Parallelogram Model

A disadvantage of Coomb's parallelogram model is its deterministic nature: responses are completely dependent on the distance between person and item, random characteristics of either person or item are not allowed to interfere. Data collected in empirical research will never be so perfect that (1) applies. Looking at Table 2, it can be seen that there exists no order of items and persons for which the positive responses will resemble a parallelogram structure. There will always be response patterns with 0's located between 1's. This violates the basic principle expressed by (1): a person should respond positively to all (and not most) items within a certain threshold τ of his location.

Coombs himself was aware of this disadvantage (1964, p.70): "In any practical applications of these 'pick k/n' [or, 'pick any/n'] methods perfect parallelogram patterns are not usually obtained." The way out of this problem is to relax (1) and to replace it by

$$P(X_{ai}=1|\beta_a,\delta_i) = f(|\beta_a-\delta_i|), \tag{2}$$

i.e. the probability that person a responds positively to item i is a function of the distance between person and item. This function should be chosen such that the probability of a positive response increases if the distance between person a and item i decreases.

Coombs discusses three types of models that are based on (2) rather than on (1): parallelogram models, latent structure models, and latent class models. Note that the name 'parallelogram models' is used twice, once as the generic name of a class of models, and once to identify a specific class of models. Since computers were not as common and able to perform complex calculations in 1964 as they are nowadays, Coombs discussion of these three types of models remained superficial and at a rather theoretical level. However, in the past decennium several authors have shown that these models are viable and applicable.

3.1 Parallelogram Models

Quoting Coombs (1964, p. 311), "In general, then, parallelogram analysis of 'pick any/n' data will usually involve permuting columns in order to construct a parallelogram with as few gaps as possible." I.e. a structure resembling a structure like the one displayed in Table 3 as close as possible (as few as possible 0's located between 1's). The definition of 'as few gaps as possible' is not as clear cut as one might think at first thought. The models discussed by Rian van Blokland-Vogelesang and Wijbrandt van Schuur are both typical parallelogram models, although both the method to permute the columns, and the definition of 'as few gaps as possible' differs between both athors. The interested reader is refered to Davison (1980), Heiser (1981, Chapter 3), and to Cliff, Zatkin, Gallipeau and McCormick (1988), who also discuss parallelogram models.

3.2 Latent Structure Models

Although both the parallelogram and latent structure models are based on (2), the latent structure models make a more explicit use of (2) to determine the locations of persons and items on the latent characteristic of interest (which constitute the latent structure). Either the function f is completely specified. See the papers by Norman Verhelst and Huub Verstralen, and Herbert Hoijtink in this issue, and Andrich (1988), and DeSarbo and Hoffman (1986). Or a set of assumptions with respect to its behavior is specified. See the paper by Wendy Post and Tom Snijders in this issue.

The basic formula for the latent structure model is:

$$P(\mathbf{X}_a) = \int_{-\infty}^{\infty} P(\mathbf{X}_a|\beta)\, dG(\beta), \tag{3}$$

where,

$$P(\mathbf{X}_a|\beta) = \prod_{i=1}^{n} P(X_{ai}=1|\beta_a,\delta_i)^{X_{ai}} \, (1 - P(X_{ai}=1|\beta_a,\delta_i))^{(1-X_{ai})}, \tag{4}$$

$\mathbf{X}_a$ is the response vector of person a, and $G(\beta)$ is the density function of the person locations.

Either (3) constitutes the atomic element of a likelihood function that can be used to estimate the locations of the items, and the parameters of the density function of the person locations (see Norman Verhelst and Huub Verstralen, and Herbert Hoijtink in this issue), or, (3) is used to derive properties the data have, if the set of assumptions with respect to the behavior of (2) is correct (see the paper by Wendy Post and Tom Snijders).

3.3 Latent Class Models

Latent class models are very similar to latent structure models. However, where latent structure models assume that the latent characteristic is continuous, latent class models assume that it consists of a number of (ordered) classes. Two types of latent class models may be distinguished: either (2) is completely specified (see the contribution by Ulf Bockenholt in this issue); or a set of assumptions with respect to its behavior is specified (see the contributions by Marcel Croon, and Anton Formann in this issue).

The basic formula for the latent class models is

$$P(\mathbf{X}_a) = \sum_{t=1}^{T} P(\mathbf{X}_a | t)\, \pi_t, \tag{5}$$

where,

$$P(\mathbf{X}_a | t) = \prod_{i=1}^{n} P(X_{ai}=1 | t, \delta_i)^{X_{ai}} \, (1 - P(X_{ai}=1 | t, \delta_i))^{(1-X_{ai})}, \tag{6}$$

and π_t denotes the probability that an observation is sampled from class t. Each of the latent class models to be discussed in this issue uses (5) as the atomic element of a likelihood function that can be used to estimate the parameters of interest.

4. The Data

Two data sets will be analyzed with each of the models to be discussed in this issue of Kwantitatieve Methoden. The first data set is concerned with the measurement of the attitude with respect to nuclear power plants (Formann, 1988) and has been discussed above. Some authors will refer to these items using keywords/numbers which are in brackets in Table 1.

The second example is concerned with the measurement of the attitude in the car-environment issue (Doosje and Siero, 1991). Ten items that are ordered on the latent characteristic from favorable with respect to the environment, to favorable with respect to car use, are presented in Table 4. Some authors will refer to these items using keywords/numbers which are in brackets in Table 4.

Table 4
Items Indicative of the Attitude in the Car-Environment Issue
Keywords/Numbers Used in some of the Papers for Identification of these Items are in Brackets

Item	Phrasing
1.	People who keep driving a car, are not concerned with the future of our environment (NOCONC/9).
2.	The environmental problem justifies a tax burden on car driving so high that people quit using a car (MAXTAX/3).
3.	Putting a somewhat higher tax burden on car driving is a step in the direction of a healthier environment (HIGHTAX/4).
4.	Car users should have to pay taxes per mile driven (PAYTAX/10).
5.	A cleaner environment demands for sacrifices like a decreasing car usage (DECREA/2).
6.	Car use cannot be abandoned. Some pressure on the environment has to be accepted (CARSTAY/1).
7.	It is better to deal with other forms of environmental pollution then car driving (OTHPOLL/5).
8.	Considering the environmental problems, everybody should decide for themselves how often to use the car (SELFDE/8).
9.	Technically adapted cars do not constitute an environmental threat (NOTHREA/7).
10.	Instead of environmental protection measures with respect to car use, the road system should be extended (EXTROA/6).

Doosje and Siero (1991) wanted to measure the influence of an information campaign explaining the effect of car use on the environment, on the attitude of persons in the car environment issue. In order to do this, two samples of persons were taken: one before the information campaign and one after the information campaign. The original samples where approximately 1000 persons each. The examples in the papers that follow are based on two random subsamples of 300 persons each. Besides the application of their models to the data, most authors will discuss the effect of the information campaign on the location of the items and the persons on the latent characteristic of interest.

References

Allen, M.J., and Yen, W.M. (1979). Introduction to measurement theory. Belmonth: Wadsworth Inc.

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. Applied Psychological Measurement, 12, 33-51.

Cliff, N., Collins, L.M., Zatkin, J., Gallipeau, D., McCormick, D.J. (1988). An ordinal scaling method for questionnaire and other ordinal data. Applied Psychological Measurement, 12, 83-97.

Coombs, C.H. (1964). A theory of data. Ann Arbor: Mathesis Press.

Davison, M.L. (1980). A psychological scaling model for testing order hypotheses. British Journal of Mathematical and Statistical Psychology, 33, 123-141.

DeSarbo, W.S., and Hoffman, D.L. (1986). Simple and weighted unfolding threshold models for the spatial representation of binary choice data. Applied Psychological Measurement, 10, 247-264.

Doosje, B.J., and Siero, F.W. (1991). Invloed van discrepantie tussen boodschap en eigen standpunt, kwaliteit van argumenten en motivatie tot elaboratie op de attitude ten aanzien van de auto milieu problematiek [Influence of the difference between message and opinion, quality of arguments and motivation to change on the attitude towards the car environment issue]. In: R.W. Meertens, A.P. Buunk, and R. van der Vlist. Sociale psychologie: Voorlichting en maatschappelijke problemen [Information and social problems].

Formann, A.K. (1988). Latent class models for nonmonotone dichotomous items. Psychometrika, 53, 45-62.

Heiser, W.J. (1981). Unfolding analysis of proximity data. Dissertation, University of Leiden, The Netherlands.