

LONGITUDINAL K -SETS ANALYSIS USING A DUMMY TIME VARIABLE

EEKE VAN DER BURG and CATRIEN C.J.H. BIJLEVELD

Generalised Canonical Analysis (GCA) (Carroll, 1968) and nonlinear Generalised Canonical Analysis (Van der Burg, De Leeuw & Verdegaal, 1988) are shortly introduced. The latter technique is suited for categorical variables by means of optimal scaling and is implemented in the computer program OVERALS. Nonlinear GCA is illustrated using a longitudinal study on the dyadic interaction of mothers and children (Van den Boom, 1988). A dummy variable models the time dependence. The supposed relation between mother and child behaviour during the first six months and the type of attachment is confirmed.

Key words: Canonical correlation analysis, generalised canonical analysis, longitudinal research, nonlinear transformations, optimal scaling.

INTRODUCTION

Generalised canonical analysis (GCA) or K -sets canonical analysis relates several sets of variables simultaneously, searching for what is common between the sets (Carroll, 1968). Van der Burg, De Leeuw and Verdegaal (1988) presented an extension in which categorical variables are rescaled by means of optimal scaling, which was implemented in the computer program OVERALS.

Catrien C.J.H. Bijleveld and Eeke van der Burg. Department of Psychometrics & Research Methodology. Faculty of Social and Behavioural Sciences. Leiden University. P.O.Box 9555, 2300 RB Leiden, the Netherlands.

-Article presentat al Seventh International Conference on Multivariate Analysis, setembre 1992.

-Acceptat el maig de 1993.

Nonlinear GCA has not often been applied in longitudinal analysis; in the present study the application of nonlinear GCA to the dyadic interaction of mothers and children (Van den Boom, 1988) is illustrated, using a dummy variable to model the time dependence. During the first six months after birth, characteristic behaviour of mothers and children has been registered, that supposedly predicts children's future attachment; the attachment hypothesis could be confirmed in our analysis.

This paper is meant to illustrate the usefulness of OVERALS in longitudinal research. First a short introduction of (nonlinear) GCA is given. As longitudinal data correspond to measurements in time, we are dealing with a three dimensional data table or data box. We show how to rearrange the data box and how to construct a dummy time variable. Next, the longitudinal application on the attachment between mothers and their babies is discussed extensively.

GENERALISED CANONICAL ANALYSIS

Generalised Canonical Analysis (GCA) is a generalisation of Canonical Correlation Analysis (CCA) as described in many text books on multivariate analysis (e.g. Tatsuoka & Lohnes, 1988). The generalisation refers to the number of sets, which can be more than two, say K . In fact, canonical correlation analysis can be generalised in many ways (e.g. Kettenring, 1971; Van de Geer, 1984; Gifi, 1990), however, we are interested in a special type, namely the one defined by Carroll (1968).

CCA solves the problem of finding linear composites of variables for two sets in such a way that the correlation between the weighted sums is maximised. GCA solves the same problem for K sets. Thus both CCA and GCA can be applied in case there are sets of variables in the data, and in case the research question deals with the correspondence between the sets.

Carroll's (1968) formulation of GCA is as follows: maximise the weighted sum of squared correlations between unknown scores for subjects and linear composites of variables per set. The weights in this formulation are chosen constants. Taking the weights equal to one, Carroll's formulation can be shown to be equivalent to: minimise the squared difference between unknown scores for subjects, referred to as object scores, and linear composites of variables for K sets. A second solution for the object scores must be taken independently from the first solution, and these scores must be standardised. In addition, all subsequent solutions, say P , of the object scores must be independent from

each other. Let us denote the sets of (standardised) variables by X_k ($N \times M_k$) with N the number of subjects, and M_k the number of variables. In addition, denote object scores by Z ($N \times P$) and the weights for the linear composites by A_k ($M_k \times P$). According to the above formulation of GCA, we get the following expression, which is equivalent to Carroll (1968):

$$(1) \quad \text{minimise } \sigma(Z, A_k) = \sum_{k=1}^K \text{SSQ} (Z - X_k A_k),$$

subject to the condition that the columns of Z are standardised and independent.

The notation SSQ stands for Sum of Squares. If we allow the data matrices X_k to be rescaled by optimal scaling, a nonlinear version of GCA results (Van der Burg, De Leeuw & Verdegaal, 1988). In optimal scaling we deal with several types of measurement levels: nominal, ordinal and numerical. For nominal data only the classification (of the subjects) by the variable counts, for ordinal data the classification and the order are taken as fixed, and for numerical variables the interval level is assumed. Denoting the rescaled variables by Q_k ($N \times M_k$), and referring to the restrictions of each variable as measurement restrictions, we can reformulate expression (1) by adding optimal scaling as follows:

$$(2) \quad \text{minimise } \sigma(Z, Q_k, A_k) = \sum_{k=1}^K \text{SSQ} (Z - Q_k A_k),$$

subject to the conditions that the columns of Z are standardised and independent, and the columns of Q_k are standardised and satisfy the measurement restrictions.

The weighted sum of variables for set k , $Q_k A_k$, can be rewritten in several ways:

$$(3) \quad \sum_{j=1}^{M_k} G_{jk} B_{jk} = \sum_{j=1}^{M_k} G_{jk} c_{jk} a'_{jk} = \sum_{j=1}^{M_k} q_{jk} a'_{jk} = Q_k A_k,$$

with q_{jk} a column of Q_k , a'_{jk} a row of A_k , and $G_{jk} c_{jk} = q_{jk}$. The matrix G_{jk} corresponds to the indicator matrix of variable j of set k , the vector c_{jk} has to satisfy the measurement restrictions of the corresponding variable and matrix B_{jk} is equal to $c_{jk} a'_{jk}$, which means that this matrix has rank one. The matrix B_{jk} contains the category quantifications, i.e. the rescaled values of the category scores for a variable. As this matrix has rank one, the category quantifications are found on a line.

If we allow a nominal variable to be rescaled differently for each solution, this means that the restriction ' B_{jk} has rank one' is relaxed, that is ' B_{jk} is unrestricted'. In that case we say that the variable is considered as multiple

nominal. If we deal with one rescaling for a nominal variable, we say that the variable is considered as single nominal. In case of a multiple nominal variable the category quantifications are no longer on a line.

Expression (2), combined with the possibility of handling a nominal variable as multiple, is called nonlinear GCA (Van der Burg, De Leeuw & Verdegaal, 1988; Gifi, 1990). This technique is implemented in the computer program OVERALS, which is in the SPSS package (SPSS, 1990).

The OVERALS analysis results contain several tools for interpretation. In the first place the object scores form perpendicular axes, on which subjects as well as rescaled variables can be projected. The subjects form a swarm of points with variance equal to N in the direction of the axes. We can average the subjects (object scores) for special categories of a variable, to obtain an idea how these categories move around in the swarm of points. For the rescaled variables we can compute the correlations with the axes. These correlations, referred to as component loadings, provide the coordinates for the rescaled variables in the object scores space. The component loadings can be interpreted in a similar way as the component loadings in Principal Component Analysis (PCA). The fit of the solution is reflected in the eigenvalue, which is maximally one and minimally zero. The fit measure corresponds to the mean explained variance by the linear composites per set (and not by the variables like in PCA). As we deal with sets, we also deal with overlap within sets (multicollinearity), so that weights per set give rise to the same interpretative problems as in multiple regression. Therefore we use the component loadings for interpretation.

We realise that this description of nonlinear GCA is limited, however, it is beyond the scope of this paper to provide a full explanation of nonlinear GCA. For a more detailed description we refer to Van der Burg, De Leeuw and Verdegaal (1988), as well as to Gifi (1990).

FLATTENING THE DATA BOX AND TIME AS A DUMMY VARIABLE

In longitudinal research, we are usually confronted with a three dimensional data box as there are, in the majority of the cases, measurements for N individuals on M variables for T time points. Only few multivariate techniques are suited to analyse an $N \times T \times M$ data box. For instance LISREL can model the time dependence and analyse a data box, however, only few time points and numerical measurement levels are possible.

In many instances we are forced to flatten the data box (Visser, 1985, chapter 4). Also in case of GCA we need a two dimensional data matrix. The flattening of the data box is done in the following way. First we stack the individual $T \times M$ data matrices in a two-dimensional super-matrix, that is of size $(N \times T) \times M$. This matrix constitutes the first set, the so-called variables-set. Next, a dummy time-variable is constructed as an $(N \times T) \times 1$ vector, a super-vector, that contains the time points that have been numbered in such a way that each t -th time point gets a unique identification. For the values of the dummy time-variable, we could take the number of the time point. For T time points, we then get as subsequent values of the dummy time-variable: $(1, 2, \dots, T), (1, 2, \dots, T), \dots, (1, 2, \dots, T)$, where the values for each subject have been put within brackets. The dummy time-variable constitutes the second set, the so-called time-set. For a schematic representation see Figure 1.

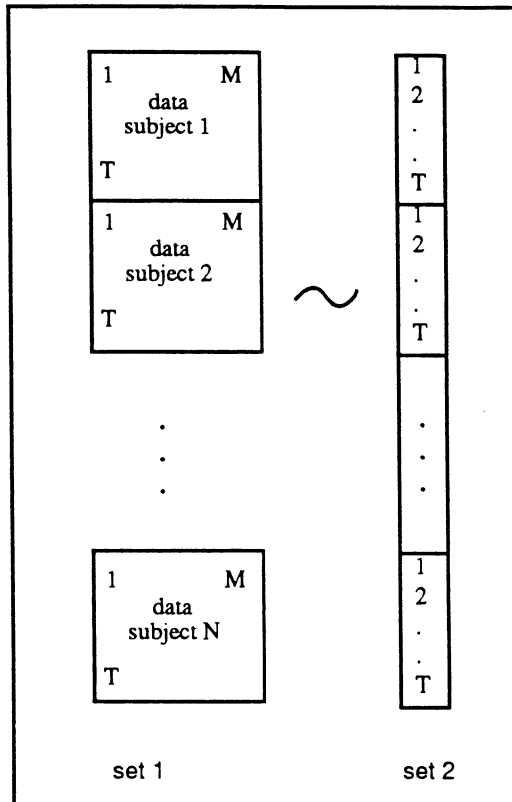


Figure 1.
Schematic Representation of the Dummy Time-Variable Set Up.

Next, the dummy time-variable will be treated as a single nominal variable; the variables in the first set will be treated according to their measurement level. Because CCA or GCA maximises the canonical correlation between the two sets, the relation between the variables-set and the time-set is optimised. Consequently, the category quantifications of the nominal dummy time-variable are the best possible summary of the scores of all subjects on the variables at the respective time points. The category quantifications of the dummy time-variable then describe the general development of subjects on the variables, over the consecutive time points.

Variations on this theme are possible. The dummy time-variable could be treated as multiple nominal, which gives several, independent, summaries of development over time. A similar result is obtained by adding more dummy time-variables to the time-set, labeled in different fashions (otherwise most algorithms collapse). All dummy time-variables can then be treated as nominal, and higher dimensional canonical spaces can be analysed, opening up the possibility to investigate more-dimensional development.

Another possibility would be to scale the dummy time-variables not nominally but ordinally, or even numerically, so that trends can be approximated. For instance, if a first dummy time-variable would be labeled as above with its values equal to the number of the time point, this dummy time-variable would serve to approximate linear development. Quadratic development could be approximated by adding a second dummy time-variable to the time-set, that could then be labeled for instance: $(1, 2, \dots, \frac{1}{2}T, \dots, 2, 1), (1, 2, \dots, \frac{1}{2}T, \dots, 2, 1), \dots, (1, 2, \dots, \frac{1}{2}T, \dots, 2, 1)$, with the values of the time-variable again put between brackets for each subject. And so on for more dummy codings.

The analysis of time-dependent data using the dummy-time variable set-up, is possible using OVERALS. However, the summary of time-development is not computed in the space of the time-set, but is projected into the space for the object scores. In addition, the contribution of the variables-set is expressed in terms of the space of object scores by means of component loadings. In two sets canonical correlation analysis, usually a difference is made between the space of the first set and the second set, and then we could choose the space of the set referring to the dummy time variable (compare Van der Burg & De Leeuw, 1984). In OVERALS the two spaces are averaged into the space of object scores. If we are dealing with more variables-sets and one time-set, OVERALS averages all spaces.

Using one dummy time-variable and one set of variables, CCA or GCA reduces to a multiple regression. In that case, other regression-type multivariate techniques can be used as well, such as multiple regression and redundancy analysis, on the understanding that a nonlinear version exists.

A LONGITUDINAL EXAMPLE: DEVELOPMENT OF ATTACHMENT IN YOUNG CHILDREN

The data used for this example stem from a study into the development of attachment in babies (Van den Boom, 1988). Broadly speaking, babies can develop three types of attachment relationships to their mothers: secure attachment, resistant attachment and avoidant attachment. Infants who are securely attached use the mother as a secure base from which to explore, they reduce their exploration and may be distressed in her absence, but greet her positively on her return, and soon return to exploring. This is the pattern shown by two-thirds of infants in normative samples. It has been associated with responsive care in the home during the first year. Infants whose attachment pattern is avoidant, explore with minimal reference to the mother, are minimally distressed by her departure, and seem to ignore or avoid her on return. In normative samples, this pattern characterises one in five infants. Prior maternal home behaviour in this group has been described as intrusive and reflecting discomfort with physical contact. Theoretically, avoidant attachment in infancy is associated with later anti-social and aggressive behaviour, but the subject is controversial. The third major pattern is described as insecure resistant. It is marked by minimal exploration, reflecting inability to move away from the mother. These infants are highly distressed by separations and are difficult to settle on reunions. In normative samples, approximately one in seven babies shows this pattern; it is considered to reflect a history of inconsistent maternal responsiveness, and subsequent social development vulnerable to social withdrawal. However, since this is also the least frequent pattern, it has been impossible to provide strong empirical tests of these propositions (for details and references see Van den Boom, 1988).

Table 1

Mother and Child Behaviour Variables

child variables	mother variables
positive sociable behaviour	observing baby
observing persons and objects	effective stimulation
vocalising	vocalising/offering objects
whining/crying	physical contact
exploration	comforting
sucking	uninvolved
	responsiveness to crying
	responsiveness to positive behaviour

The design of this study was as follows. Thirty mothers with their children were observed for 40 minutes on two occasions in each month during the six months after birth. During these observations, a number of variables were scored. Each time the mother or child exhibited a certain type of behaviour, this was scored, so that the measurements were actually frequencies of occurrence of types of behaviour in small (6 second) time intervals. The mother and child variables are in Table 1. After 12 months, it was assessed whether the child had developed a secure, resistant or avoidant attachment; this constituted the (non-time varying) criterion variable. It was assumed that mothers influenced children, and children influenced mothers, in addition a time-dependence was supposed.

To investigate the general development of mother's and children's behaviour, in relation to attachment, we chose the following design. The mother and child variables were put in separate sets. Attachment constituted the third set; in a fourth set we had a dummy time-variable, formed in the same way as in the previous section, except that now for each mother and child we had 6 time points, and we used a second, independent dummy time axis by taking the measurement level of the time variable as multiple nominal. Each observation unit (mother and child) is considered as six subjects (one new subject at each time point). All behaviour variables were treated ordinally (the original frequencies of behaviour were recoded into 7 or 8 categories), the attachment variable was treated as multiple nominal. One child had gone to hospital in the last two months, which, with lots of crying, caused rather deviant scores, so for this child we deleted the last two months from the data set. By choosing this set-up, OVERALS will maximise the relation between mother and child variables, the time axis and attachment variable. As attachment has only three categories, we perform a type of multiple group (longitudinal) discriminant analysis. It means that the groups are predicted as well as possible on the basis of what is common in child behaviour, maternal behaviour and the development of these.

ANALYSIS RESULTS

The eigenvalues of the first and second canonical dimension were .630 and .560 respectively. This is slightly on the low side, which can be explained by the fact that attachment does not vary in time as the other variables do, which thus more or less suppresses the relation between sets. In Table 2 the contribution of each set to the eigenvalues is shown. This contribution, or fit per set, corresponds to the squared multiple regression coefficient of the object scores

and the variables of a set. Indeed we see from Table 2 that attachment and time contribute mainly to different dimensions. Although time and attachment were completely independent in the raw data, the rescaling allows for a very small dependency to be introduced in the second dimension (c.f. Figure 3).

Table 2

Fit For Every Set per Dimension and Eigenvalues

	dim.1	dim.2
mother variables	.732	.821
child variables	.872	.812
time	.895	.101
attachment	.019	.505
eigenvalues	.630	.560

The component loadings are in Figure 2; maternal behaviour has been typed in plain face, children's behaviour in italic face. The object scores are not plotted in Figure 2, however we can interpret the component loadings as characterising those subjects who can be found in the direction or the neighbourhood of an arrow.

In the center top of the picture, crying occurs together with comforting, so children who cry a lot, are usually comforted much (although the component loadings are not identical). Thus we can expect in the top of the plot, mother and child pairs in observation periods during which the baby cries a lot, and the mother does quite a bit of comforting; in the bottom of the plot we can expect mother and child pairs in observation periods during which the baby cries little, and the mother does little comforting. Turning clockwise, we find physical contact of the mother closest to responsiveness to crying. Next, children's positive behaviour, stimulation by the mother, and less strongly so, vocalising by the baby and the mother's responsiveness to positive behaviour point in approximately the same direction. Vocalisation by the child is apparently more related to stimulation by the mother, than by her vocalising, as we find this variable at some angle, closest to exploration by the baby. Turning further, we find uninvolved by the mother. Nearby, the baby's watching of the mother can be found, and close to it, the mother's watching of the baby. Both watching variables and the sucking variable have moderate to low component loadings.

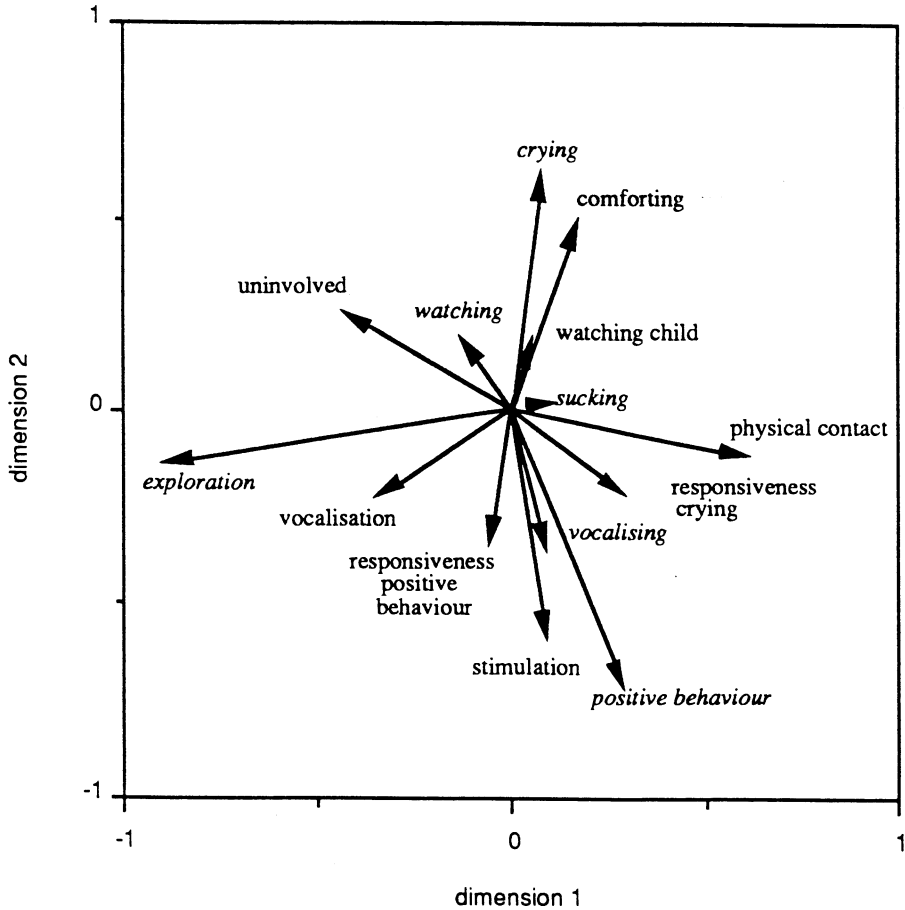


Figure 2.
Component Loadings of Rescaled Mother and Child Variables.

Looking at the children's development, we have drawn in Figure 3 the average monthly object scores for the mother-child pairs in month one to six: all start in month 1 at the right-hand side of the picture, and end in month 6 at the left-hand side. Figure 3 should be interpreted in connection with the former figure of component loadings, which means that for instance the resistantly attached cry a lot in month 1, the mothers have lots of physical contact with the securely attached in month 1, and so on.

In Figure 3 the first thing that catches the eye is, that the three attachment groups are markedly separated in the plot. The securely attached 'travel'

through the bottom of the plot, from month 1 to 6, and the avoidantly attached and those with resistant attachment both travel in the upper regions. On the average, children develop from right to left. Combining this with the information from Figure 2, the following emerges. The mothers of those children who will eventually become securely attached, start out with lots of physical contact with their babies; they are very responsive to crying, especially in the second month. In the second and third month, these children show lots of positive behaviour, and, though less markedly so, start vocalising, with their mother stimulating them. The mother's responsiveness to positive behaviour then increases towards the fifth month, as does her vocalising and offering of objects. In the last month, children start to explore to the full. Towards the end of the six months, mothers become less involved with their children. For the securely attached children, crying and comforting are more or less stable over the months.

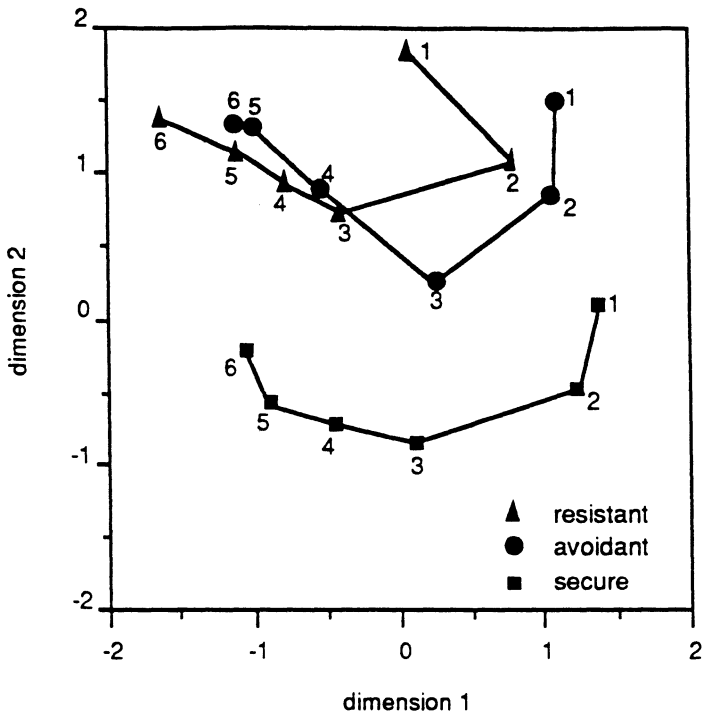


Figure 3.
Developments for Three Attachment Types, Numbers Correspond to Time Points.

Those children who eventually develop an insecure attachment pattern, interact with their mothers along different lines. The resistantly as well as the avoidantly attached are both located close to crying. Especially the resistantly attached, start with a lot of crying in the first month. These children (and their mothers) make a bad start. Mothers and children then make a swing to the right, as probably mothers try to cope with this behaviour by lots of comforting and more physical contact in the second month. In the third month, physical contact decreases already, and mothers become less and less involved with their children towards the last months, while children become more explorative. The children that will become avoidantly attached do not make such a bad start as the resistantly attached, although these children also cry a lot in the first month. The mother comforts the child more than in the resistant group, and in the second month there is also a movement towards more physical contact. In the third month, the mother and child do some stimulation, positive behaviour, vocalising, but this is counteracted by still lots of crying by the baby. In the fourth month, things go 'wrong' for this group, as mothers become less involved, a tendency which continues until the last month. Looking at differences between the avoidant and resistant group, it appears that the resistantly attached make an even worse start than the avoidantly attached, and that uninvolvement by the mother occurs earlier in the resistant group than in the avoidant group. In both groups, children are more left alone at quite an early stage, than in the secure group. It should be noted, that all babies develop into the direction of exploration, meaning that, independent of the development of attachment, explorative behaviour grows in time.

DISCUSSION

Previous experience with the research presented in this paper showed that it was hard to demonstrate the attachment hypothesis as it is not the behaviour of mothers and children that produces a special type of attachment, but the change of behaviour in time that is related to attachment. By using OVERALS we succeeded in a very nice way to demonstrate this phenomenon. Plotting the component loadings (Figure 2) and the mean object scores (Figure 3) gave a very clear impression of how the relation between attachment, time and different types of behaviours can be interpreted.

In this paper we used a dummy time variable to model the relation between the measurements and time. Another way of modelling time dependence in longitudinal research is by dealing with lagged versions of the data. In Bijleveld

and Van der Burg (1993), this method is explored, also in combination with OVERALS. The conclusion of this analysis is rather similar to our finding, although it is more difficult to interpret lagged versions of variables. The method presented in this paper shows an elegant and easy to interpret solution.

REFERENCES

- [1] **Bijleveld, C.C.J.H. & Van der Burg, E.** (1993). "Longitudinal K -sets analysis using lagged variables". *Qüestió*, **17.3**, 321-332.
- [2] **Carroll, J.D.** (1968). "Generalization of canonical correlation analysis to three or more sets of variables". *Proceedings of the 76th Annual Convention of the American Psychological Association*, 227-228.
- [3] **Gifi, A.** (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.
- [4] **Kettenring, J.R.** (1971). "Canonical analysis of several sets of variables". *Biometrika*, **56**, 433-451.
- [5] **SPSS** (1990). *SPSS Categories*, user's manual. Chicago: SPSS Inc..
- [6] **Tatsuoka, M.M. & Lohnes, P.R.** (1988). *Multivariate Analysis. Techniques for Educational and Psychological Research*. New York: Wiley.
- [7] **Van de Geer, J.P.** (1984). "Linear relations between k sets of variables". *Psychometrika*, **49**, 79-94.
- [8] **Van den Boom, D.C.** (1988). *Neonatal irritability and the development of attachment: Observation and intervention*. Unpublished PhD thesis, Leiden: Leiden University.
- [9] **Van der Burg, E. & De Leeuw, J.** (1983). "Non-linear canonical correlation". *British Journal of Mathematical and Statistical Psychology*, **36**, 54-80.
- [10] **Van der Burg, E., De Leeuw, J., & Verdegaal, R.** (1988). "Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features". *Psychometrika*, **53**, 177-197.
- [11] **Visser, R.A.** (1985). *The analysis of longitudinal data in behavioural and social research*. Leiden: DSWO Press.

