RR – 85-06

FIXED FACTOR SCORE MODELS WITH LINEAR RESTRICTIONS

Jan de Leeuw
Department of Data Theory

Ab Mooijaart
Rien van der Leeden
Department of Psychometrics and Research Methodology

University of Leiden

Introduction

Factor analysis, in the work of Spearman, Thomson, Burt, and Thurstone, was essentially an algebraic technique used to describe data matrices. Generalization aspects associated with statistical modelling were only introduced much later, and have not played a prominent role until quite recently. We think that a purely algebraic approach to factor analysis is perfectly respectable, and in some cases the only reasonable alternative. On the other hand embedding the equations into a statistical model has many advantages, at least in situations were the idea of replication and random sampling make sense. From the developments so far, two different statistical factor analysis models have received most of the attention. We present them in their strongest form, with the assumption of multivariate normality, although recent theory concentrates on studying weaker versions of both models (De Leeuw, 1983, Mooijaart, in press).

The first factor model, which is the most popular by far, is the random score model. It supposes that the observed scores z_{ij} (i=1,..., n; j=1,...,m) are realizations of random variables $\underline{z}_{ij}$ with the following structure.

A1: $\underline{z}_{ij} = \mu_j + \sum_{s=1}^{p} \underline{x}_{is}\alpha_{js} + \delta_j\underline{\varepsilon}_{ij}$,

A2: $\underline{x}_{is}$ and $\underline{\varepsilon}_{ij}$ are mutually independent standard normal.

Observe that we underline random variables, a convention which is particulary convenient in this context (Hemelrijk, 1966). The structural parameters of model A, i.e. the loadings α_{js} and the unique variances δ_j^2, can be estimated in various ways. The method of maximum likelihood (ML) is currently most popular. It gives consistent and efficient estimates of the structural parameters, with asymptotically a multivariate normal sampling distribution. Model A was introduced by Garnett (1919), and connected with modern statistical theory by Lawley (1940) and Rao (1955). The first practical algorithms were presented by Jöreskog (1967).

An alternative model was presented by Young (1940), and generalized by Lawley (1941). It is

B1: $\underline{z}_{ij} = \mu_j + \sum_{s=1}^{p} x_{is}\alpha_{js} + \delta_j\underline{\varepsilon}_{ij}$,

B2: $\underline{\varepsilon}_{ij}$ are mutually independent standard normal.

This is the fixed score model, because the factor scores x_{is} in this model are fixed quantities, i.e. additional parameters. Model B is more complicated from a statistical point of view than model A. Lawley (1941) derived the likelihood equations, but Anderson and Rubin (1956) showed that the likelihood function is unbounded and maximum likelihood estimates do not exist. Alternative estimates, based on the distribution of the covariances of the $\underline{z}_{ij}$, were also derived by Anderson and Rubin. They showed that these alternative estimates provided consistent asymptotically normal estimates of the structural parameters, with a dispersion that was equal to that of the ML estimates under model A. In fact they also showed that the ML estimates under A have the same asymptotic distribution under A as under B. All this material is reviewed in a recent very interesting paper by Anderson (1984).

The results of Anderson and Rubin have made it somewhat futile to look for alternative estimation methods for model B. Model A estimates can be used in all cases. This fact has also tended to make the model B somewhat unpopular, which is unfortunate. In fact Young (1940) and Whittle (1952) both argued that model B is more appropriate for most applications they could think of. And it is true, of course, that factor scores can be estimated in the usual sense only if they are parameters, i.e. only if model B is true. If we are interested in the scores of individuals, or if we are unwilling to make the assumption that individuals are a random sample from a single well-defined population, then we must use model B. It is not true, by the way, that there are only rather subtle statistical differences between models A and B. Their conceptual basis is quite different. This becomes more obvious if we introduce replications in model B. The basic assumption B1 becomes

$$\underline{z}_{ijk} = \mu_j + \sum_{s=1}^{p} x_{is}\alpha_{js} + \delta_j\underline{\varepsilon}_{ijk}.$$

This shows that the data matrix is interpreted in model B as arising from a factorial design, with individuals and variables as factors. The model imposes a particular bilinear structure on the interaction parameters. The problem is, that under B1 we only have a single replication in each cell. We have already seen that factor scores can be 'estimated' only under B. Under A we can construct a random variable as a (possibly randomized) function of the observed variables which is close to the latent variable $\underline{x}_s$. The identification situation is also quite different for models A and B, because of the additional parameters in B.

Ultimately the distinction between A and B, and the greater popularity of A, is perhaps of a philosophical nature. Factor analysis has been dominated for a very long time by the nomothetic approach that is familiar from psychophysics. Factor analysts looked for general structural laws, individual differences were just 'error'. Thus factor scores get almost no attention in the classical books of Thurstone (1947) and Harman (1960). Model B is much more idiographic, it can be used to describe individuals succinctly. Thus model B is much closer to the spirit of many factor analytic studies in applied social science, in which we want to describe individuals in terms of a small number of factors. It is clear that the distinction between A and B also has important implications for the recent factor score controversy, which can be studied much more clearly within B than within A.

The reason why the method of maximum likelihood fails if it is applied directly to model B is by now well understood (cf Anderson, 1984). If n, the number of individuals, tends to infinity, then the number of parameters tends to infinity too. Each new individual adds his p factor scores to the set of parameters. For this reason factor scores are called <u>incidental parameters</u>, and it is classical that incidental parameters may cause the method of maximum likelihood to fail, even in estimating the structural parameters. There is too much freedom in model B.

In this paper we investigate various ways to impose restrictions in model B, and we investigate the effect of these restrictions on the behaviour of the maximum likelihood estimates. If model B is preferable to model A, and if the restrictions make sense, then our restricted forms of fixed score factor analysis are useful additions to the literature.

Maximum likelihood estimation in the fixed score model

We first give a convenient reformulation of model B. It is

C1: $\underline{z}_i = Ax_i + \underline{u}_i$,
C2: $\underline{u}_i$ independent,
C3: $\underline{u}_i \sim N(0,\Sigma)$.

In model C the vector μ has disappeared. Either the variables are first centered, or μ is absorbed into the other structural parameters. The disturbances $\underline{u}_i$ are the unique factors of B. In B they have the diagonal dispersion Δ^2, but in C we do not necessarily make this additional assumption. The negative logarithm of the likelihood is, ignoring irrelevant constants,

$$f(A,X,\Sigma) = n \lg \Sigma + \mathrm{tr}\,(Z - XA')\Sigma^{-1}(Z - XA')'. \qquad (1)$$

In this paper we shall be interested in minimizing $f(A,X,\Sigma)$ under various restrictions on the parameters. But first we show why unrestricted minimization does not work. The problem arises because we can find X and A such that

$$S_n(X,A) = \frac{1}{n}(Z - XA')'(Z - XA')$$

is singular. Suppose u is a vector in its null space. Now choose Σ_0 of rank m - 1, with $\Sigma_0 u = 0$, and define $\Sigma_s = \Sigma_0 + \varepsilon_s uu'$, with $\varepsilon_s >$ 0 and $\varepsilon_s \rightarrow 0$ if $s \rightarrow \infty$. Then $f(A,X,\Sigma_s) \rightarrow -\infty$ if $s \rightarrow \infty$. The problem occurs because Σ and $S_n(X,A)$ have null-spaces with non-empty intersection.

The problem can be solved by restricting either Σ or X and A. But the restrictions must be chosen with some care. In the unrestricted fixed factor model, for instance, restricting Σ to be diagonal is not sufficient to remedy the problem. Because we can always choose X and A in such a way that one column of Z is fitted perfectly, the unit vectors e_j can all appear in the role of the null-space vector u. By letting the corresponding diagonal element of Σ_s tend to zero, we create intersecting null spaces, and the negative likelihood tends to infinity. One radical solution, advocated for instance by Whittle (1952), is to restrict Σ even further. If we require Σ to be scalar, i.e. $\Sigma = \sigma^2 I$, then (1) simplifies to

$$f(A,X,\sigma^2) = nm \ln \sigma^2 + (\sigma^2)^{-1} \operatorname{tr} (Z - XA')(Z - XA')'. \qquad (3).$$

Minimization of (3) over X and A and σ^2 only gives problems if we can find X and A such that Z = XA', which will never be the case in practice. In practice minimization of (3) defines, of course, principal component analysis or singular value decomposition. Although this certainly defines a respectable way out of the problem, many people do not feel comfortable with the assumption that all error variances are equal.

In 'confirmatory' factor analysis restrictions are imposed on A, while keeping Σ diagonal. This will generally not prevent the unbounded likelihood to occur. Even with the usual restrictions on A it will often be possible to find X and A such that at least one variable is fitted exactly. In the fixed factor score model confirmatory analysis is possible, but it usually does not solve the incidental parameter problem.

In this paper we shall investigate the possibility of deriving interesting results from another class of restrictions, more specifically from restrictions on X. Before we go into in more detail, we review another very interesting recent attempt to salvage the method of maximum likelihood estimation in unrestricted fixed score factor analysis. McDonald (1979) defines $f_1(A,X)$ as the infimum of (1) over all diagonal Σ, and $f_2(A,X)$ as the infimum over all Σ. For X and A such that $S_n(X,A)$ is nonsingular we have $f_2(X,A) = n \ln \; S_n(X,A) + m$ and $f_1(X,A) = n \ln \; dg(S_n(X,A)) + m$. McDonald suggests that we choose X and A to maximize the difference between $f_2(X,A)$ and $f_1(X,A)$, which is $f_2(X,A) - f_1(X,A) = n \ln \; R_n(X,A)$, with $R_n(X,A)$ the correlation matrix corresponding with $S_n(X,A)$. The resulting estimates are called <u>maximum likelihood ratio estimates</u>, because they maximize the ratio of two partial likelihoods. McDonald shows that the maximum likelihood ratio estimates of the <u>structural</u> parameters are identical with the maximum likelihood estimates in the random factor model. This makes them consistent, because of the results of Anderson and Rubin mentioned earlier. The estimates of the factor scores, the incidental parameters, are not consistent, but this is hardly surprising. It seems to us that, on the basis of these results, maximum likelihood ratio methods show some promise also in other situations in which incidental parameters make maximum likelihood behave badly. From a practical point of view, however, they do not give anything new in unrestricted factor analysis. They merely tell us to use the structural maximum likelihood estimates. Etezadi-Amoli and McDonald (1983) have some additional discussion on maximum likelihood ratio estimation.

<u>Using restrictions on the factor scores</u>

Imposing restrictions on the factor scores can be done by relating the factor analysis model to some other closely related models, which have been studied mainly in econometrics. There is a fairly extensive literature on <u>reduced rank regression models</u> and <u>linear functional errors-in-variables models</u> in which

C1: $\underline{Z} = XA' + \underline{U}$,

is used in combination with C2 and C3 and

C4: $X = YB$,

where Y is a known n x r matrix. If we combine C1 and C4 in a single likelihood function we find

$$f(A,B,\Sigma) = n \ln \Sigma + \mathrm{tr}\,(Z - YBA')\Sigma^{-1}(Z - YBA')'. \qquad (4).$$

The new data analysis problem becomes minimizing this loss function over the parameters. In practical situations it will usually be impossible to choose A and B such that Z - YBA' is singular. Thus the unbounded likelihood problem is not too serious in these situations. For interpretational purposes C4 can be very convenient. We can use design matrices for Y, or 'background' variables. Restriction C4 then says, that the factor scores are in the space spanned by the background or design variables.

We have interpreted C4 as a restriction on the factor scores, which is natural in our formulation. But we can also rewrite the assumptions as $\underline{z}_i \sim N(Cy_i,\Sigma)$, where C is of the form $C = AB'$, i.e. C is an m x r matrix with $\mathrm{rank}(C) \leqq p$. This explains the terminology of reduced rank regression model, which is more natural in other contexts. More precisely econometrists generally will prefer to think of the model as a special regression model, psychometrists will think of it as a special factor model. Anderson realized from the start that the two models are essentially the same, he brilliantly summarizes his work and that of others in Anderson (1984).

Before we give concrete examples of the types of linear restrictions that can be used we first deal with the estimation problem. The model is usually extended by making assumptions about the covariance of the errors. We have already discussed some possibilities informally, and Anderson (1984) also distinguishes approximately the same cases. For ease of reference we mention the most important ones.

C5a: $\Sigma = \sigma^2\Sigma_0$, with Σ_0 known, and σ^2 unknown.
C5b: Σ is diagonal.
C5c: Σ is completely known.
C5d: Σ is completely free.
C5e: $\Sigma = \sigma^2 I$, with σ^2 unknown.

Of course many other possibilities can be distinguished. In fact a very general model would allow for a general parametric covariance structure of the errors. Anderson points out that reduced rank regression models using C5c were used in econometrics by Tintner, Geary, Malinvaud, and others. Interested recent work on models which use C5a has been done by Gleser (1981) and Kelly (1984). The key paper on reduced rank regression using C5d is, undoubtedly, Anderson (1951), although Fisher already studied a special case (multiple group discriminant analysis) in 1938. Interesting recent work on C5d is by Tso (1981), who stresses the relationship with canonical analysis. Again many historical remarks and references can be obtained from Anderson (1984).

The problem of minimizing (4) is treated in two steps. We first solve the problem of minimizing over A and B, with fixed Σ, which is of course assumed to be nonsingular. There are no restrictions on A and B, but for convenience (or for identification purposes) we require $B'Y'YB = I$ (which means that we suppose that $p \leqq r$). The optimal solution for A, given B, is independent of Σ, and is equal to $A = Z'YB$. Substituting this in (4) gives the conditional minimum

$$f(*,B,\Sigma) = n \ln \ \Sigma \ + \mathrm{tr}\ Z\Sigma^{-1}Z' - \mathrm{tr}\ B'Y'Z\Sigma^{-1}Z'YB. \qquad (5)$$

This must be minimized over B with $B'Y'YB = I$. It follows that B are the p eigenvectors corresponding with the p largest eigenvalues of the eigen-problem $Y'Z\Sigma^{-1}Z'YB = Y'YB\Lambda$. The eigenvalues $\lambda_1,\ldots,\lambda_p$

can also be found as the eigenvalues of the matrix $(Y'Y)^{-\frac{1}{2}}Y'Z\Sigma^{-1}Z'Y(Y'Y)^{-\frac{1}{2}}$. Thus

$$f(*,*,\Sigma) = n \ln |\Sigma| + \text{tr } Z\Sigma^{-1}Z' - \sum_{s=1}^{p} \lambda_s((Y'Y)^{-\frac{1}{2}}Y'Z\Sigma^{-1}Z'Y(Y'Y)^{-\frac{1}{2}}). \quad (6)$$

These results provide a complete solution of the estimation problem for known Σ. In fact it is easy to see that they also solve the problem essentially if we assume C5a or C5e. The additional parameter σ^2 is simply

$$\hat{\sigma}^2 = (nm)^{-1} \{\text{tr } Z\Sigma_0^{-1}Z' - \sum_{s=1}^{p}\lambda_s ((Y'Y)^{-\frac{1}{2}}Y'Z\Sigma_0^{-1}Z'Y(Y'Y)^{-\frac{1}{2}})\}, \quad (7)$$

where $\Sigma_0 = I$ in case C5e.

In more complicated cases minimizing (6) over Σ may not be simple at all. In these cases it is usually preferable to go back to (5), and minimize this over Σ for given fixed B. This can be done by rewriting (5) as

$$f(*,B,\Sigma) = n \ln |\Sigma| + \text{tr } \Sigma^{-1}T(B), \quad (8)$$

with $T(B) = Z'Z - Z'YBB'Y'Z$.
The minimum is attained for $\hat{\Sigma} = n^{-1}T(B)$ and thus

$$f(*,B,*) = n \ln |T(B)| + nm. \quad (9)$$

For diagonal Σ the minimum in (8) is attained if $\hat{\Sigma} = \frac{1}{n} \text{dg}(T(B))$, and for more complicated constraints more complicated methods of minimizing (8) must be used. This immediately suggests a general algorithmic strategy. In the first substep of each iteration we minimize the function with respect to B for fixed Σ. This means that we must solve a generalized eigenproblem. In the second substep we minimize loss, for fixed B, over Σ. This means minimizing (8). A whole class of methods to perform this minimization is available from the analysis of covariance structures methodology. After solving the two subproblems, possibly only partially, we start a new major iteration, and we continue until convergence. It is important to observe that in some of these 'alternating maximum likelihood' methods the subproblem of solving for Σ can be done exactly, in other cases we need iterative methods (and perhaps we can sometimes profit by splitting the Σ-subproblem into subproblems of smaller order).

There is one other case, in which analytic solution is feasible. This is the very important one investigated by Anderson (1951) and Tso (1981). If all parameters are unconstrained, going from (4) to (5) to (9) shows that the problem we have to solve is minimizing (9) over B. Equivalently we can also try to minimize (6) over Σ, but this seems less easy. Now

$$T(B) = Z'Z \quad I - (Z'Z)^{\frac{1}{2}}Z'YBB'Y'Z(Z'Z)^{-\frac{1}{2}} . \qquad (10)$$

This must be minimized over all B such that $B'Y'YB = I$, and by familiar matrix results (discussed, for instance, in Theobald, 1975) the minimum is attained for B equal to the solution of the generalized eigenvalue problem $Y'Z(Z'Z)^{-1}Z'YB = Y'YB\Lambda$, where we need, again, the p largest eigenvalues. These are, of course, also the p largest eigenvalues of $(Y'Y)^{-\frac{1}{2}}Y'Z(Z'Z)^{-1}Z'Y(Y'Y)^{-\frac{1}{2}}$, and the squares of the p largest singular values of $(Y'Y)^{-\frac{1}{2}}Y'Z(Z'Z)^{-\frac{1}{2}}$. In short, these are the p largest squared canonical correlations between Y and Z. The case in which all parameters are unconstrained can be solved by solving a canonical correlation problem. If A and B are free, and Σ is required to be proportional to a known matrix, then we have already seen earlier that we have to solve another generalized eigen-problem. Compare equation (6). If $\Sigma_0 = I$ we want to compute the eigenvalues of $(Y'Y)^{-\frac{1}{2}}Y'ZZ'Y(Y'Y)^{-\frac{1}{2}}$, which is sometimes known as redundancy analysis of Y and Z (Van den Wollenberg, 1977).

Thus canonical analysis of two matrices Y and Z can be interpreted as constrained fixed factor score maximum likelihood estimation, with a free error covariance matrix. Redundancy analysis can be interpreted as constrained fixed factor score maximum likelihood estimation with covariance matrix equal to or proportional to the identity. Of course we do not want to identify the techniques of canonical analysis or redundancy analysis completely with maximum likelihood estimation in the constrained fixed factor model. The techniques have nice geometrical origins, which makes their interpretation completely independent from statistical assumptions such as multivariate normality. Nevertheless in some cases it may be useful to interprete the techniques in the framework of maximum likelihood estimation, because inferential statements can be made using the results of Anderson (1951) and others.

Choice of linear restrictions

We still have to discuss the problem of choosing linear restrictions, i.e. choosing an appropriate matrix Y, in factor analytic contexts. This will also be illustrated in our examples below. In principle a great many choices are possible, and seem interesting. We just mention some which can be used on a more or less routine basis. It is often the case that Y is composed from information on background variables. These background variables can be categorical, in which case Y is usually an ANOVA-type design matrix with columns for main effects, interactions, and so on. The background variables can also be numerical, in which case we can collect them in columns of Y, and add columns with products, powers, products of powers, and so on. If we have a single background variable, and use powers as columns of Y, then we require the factor scores to be a degree p polynomial function of the background. In stead of polynomials we can, of course, also use other types of bases such as trigonometric functions or polynomial splines. If Y is the design matrix derived from a single classification of the individuals, then we require all individuals in the same class to have the same factor score. If Σ is free, then the technique reduces to discriminant analysis. If Σ is restricted to be diagonal we could call the resulting technique discriminant factor analysis. In longitudinal situations we might observe the same variables and the same individuals on a number of occasions.

It would then be possible to require that the factor scores of the individuals are the same for the different occasions.

The above sampling of possibilities is hopefully illustrative for the great number of possibilities one has. This can be combined, of course, with many possible choices of Σ. We have mentioned the usual ones above, but in some cases more exotic choices are possible. We can require Σ to be block-diagonal, to be a simplex, to have compound symmetry, or multitrait-multimethod form, and so on. In fact we can use the LISREL-model to give even more complicated structure to Σ. We have to remember, however, that we are modelling the distribution of errors, so very complicated models may be somewhat far-fetched.

Example 1

The data for this example are the results of a survey, held in 1974, among 575 respondents (Veenhoven & Hentenaar, 1975). In this survey people were asked to give their opinion with respect to certain issues, such as abortion, capital punishment, euthanasia, etc. Also some background information about the respondents was recorded. The variables we choose for our analysis are described in Table 1.

We will analyse these variables with a fixed factor score model with restricted factor scores. These restrictions will be made by using a design matrix, constructed out of the background variables. The design matrix gives individuals the same factor score when they have the same pattern on the background variables.

Previous knowledge about the data made us decide just to analyse the variables described in Table 1. A practical reason is also that taking more background variables would have enlarged our problem too much. The first problem is to determine which model should be appropriate using this set of variables: one can think of models with and without interactions, using two instead of three background variables etc. By manipulating the design matrix we have investigated the various possibilities.

Table 1. Description of the variables in the analysis.

- Three statements on capital punishment (CP):
  - CP1. Taking hostages should be punishable by death.
  - CP2. Murder should be punished by death.
  - CP3. In times of war killing people is justifiable.

- Four statements about abortion (AB):
  - AB1. It is the woman's right to have an abortion if she wants it.
  - AB2. Medical practitioners who perform abortion are not better than murderers.
  - AB3. People who agree with abortion have little respect for life.
  - AB4. Abortion is justifiable under no circumstances.

These statements have the response from (1) = agree completely to (5) = disagree completely. The responses of AB1 are reordered, so as to get a positive correlation with AB2, AB3 and AB4.

- Three background variables:
  1. Religion (REL) with categories: (1) = Protestant (PRO), (2) = Reformed (REF), (3) = Roman Catholic (RC), (4) = none (NON).
  2. Political preference (POL) with categories: (1) = left (LEF), (2) = denominational (DEN), (3) = liberal (LIB), (4) = right (RI), (5) = none (NON).
  3. Educational level (EDU) with categories: (1) = LO, VGLO* (A), (2) = ULO (B), (3) = VHMO (C), (4) = professional training or university (D).

* LO, VGLO, ULO, VHMO are abbreviations of typical Dutch schooltypes. Therefore we have not tried to translate them. These schooltypes are ranging from elementary school to university and we will denote them bij A, B, C and D, with D indicating the highest level.

Table 2. Background variables, interactions, - 2 log likelihood-values and the total number of parameters to be estimated for the various models

variables, interactions	-2 log likelihood	total number of parameters to be estimated
REL	12,981.17	23
POL	12,830.39	25
EDU	13,133.65	23
REL, POL	12,731.49	31
REL, POL, REL*POL	12,679.92	53
REL, EDU	12,802.83	29
REL, EDU, REL*EDU	12,767.41	47
POL, EDU	12,659.64	31
POL, EDU, POL*EDU	12,606.20	53
REL, POL, EDU	12,563.99	37
REL, POL, EDU REL*POL, REL*EDU, POL*EDU	12,416.63	99
REL, POL, EDU REL*POL, REL*EDU, POL*EDU, REL*POL*EDU	12,306.95	141

Table 2 summarizes the results of this "model-hierarchy" by giving for each model the variable(s) and interactions involved, minus two times the log of the likelihood function and the total number of parameters that is estimated. We can decide which model is the best by looking at the differences between the - 2 log likelihood - values. These differences are chi-square distributed with degrees of freedom equal to the differences between the total number of parameters to be estimated (Anderson, 1951). Because there is no proper null-model it is impossible to establish the actual fit of a model in an absolute sense. However, by differencing we can test models against each other.

From Table 2 it becomes clear that all differences between the last model and the others are significant. So for our analysis we choose the model with all (background) variables and all interactions. Table 3 gives the results of the two factor solution.

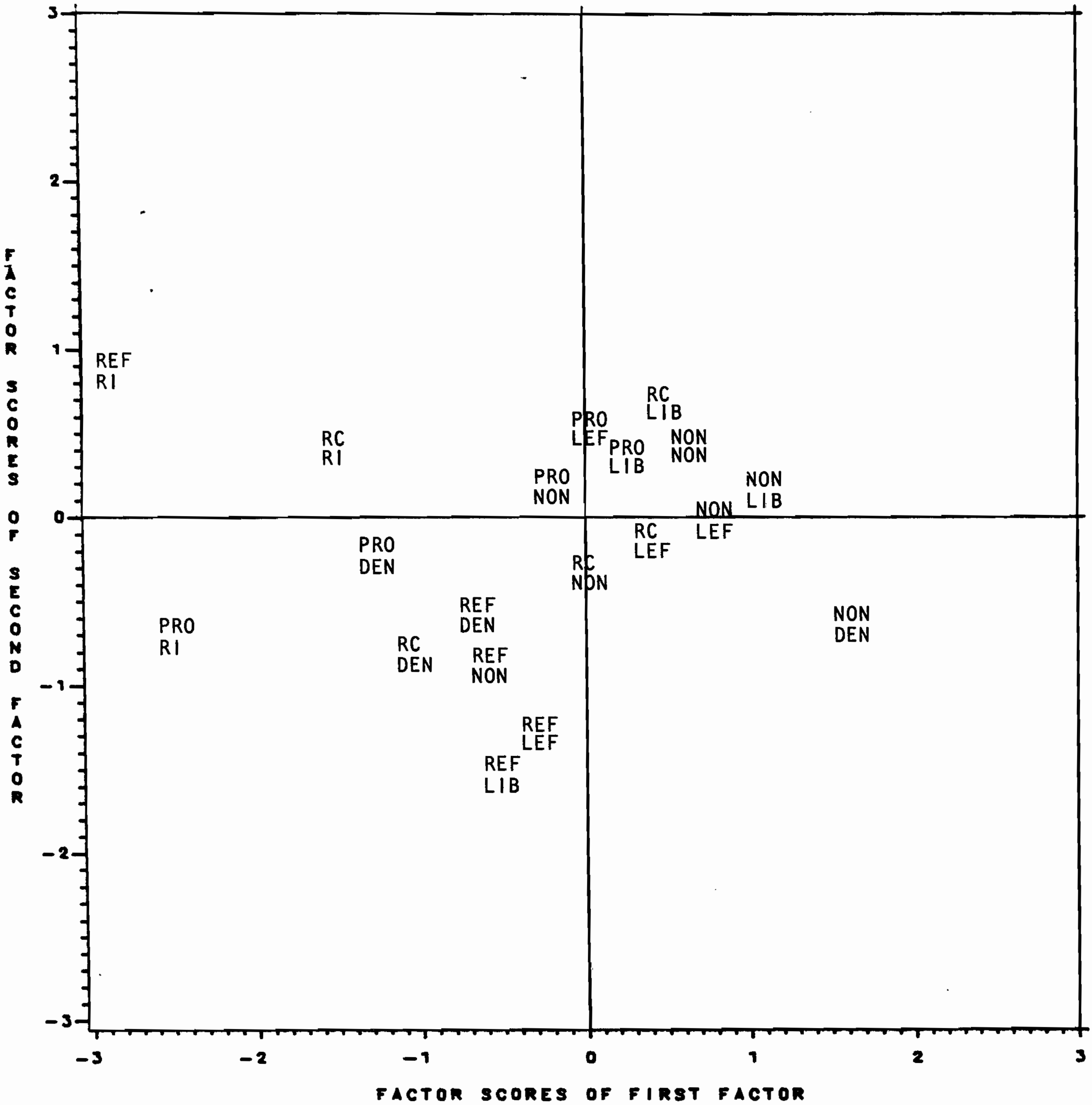
Table 3. Factor loadings for the two factor solution applied to the questions about capital punishment (CP) and abortion (AB)

variables	F_1	F_2
CP1	.223	-.627
CP2	.161	-.624
CP3	.249	-.173
AB1	.730	.281
AB2	.669	-.006
AB3	.885	.093
AB4	.737	.013

In Table 3 we see that the four abortion items have very much in common with the first factor, while the second factor deals with capital punishment. CP3 does not fit well in the model. A possible explanation for this could be that killing people in times of war is a somewhat different issue than punishing people by death for a committed crime.

Combining our three background variables yields eighty possibilities for grouping the individuals and for calculating the corresponding factor scores. Of course not all combinations really exist in the data and we will only consider those combinations with non-zero frequency. In our case 63 of such combinations are left for interpretation. Another reason for not getting as many combinations as one might expect is that with a design matrix of this size linear dependence among the dummy variables may occur. Those linearly dependent dummy's have to be removed.

Figure 1. Centroids over categories of the variable educational level: combinations of religion and political preference.



One way to interpret the estimated factor scores is to make plots. A plot of all 63 points, however, would be too complicated; i.e. one can find four different points for all educational levels with the combination Reformed/Right. This can make it a bit obscure to see what effects really exist. Therefore we have plotted the centroids of the scores of the categories of each background variable. This reduces the combinations just mentioned to one point, Reformed/Right in the plot where all categories of the variable educational level have been taken together. In fact, by doing so, one reduces the interaction effect of that variable. The centroids for each pair of background variables are plotted in Figure 1, 2 and 3.

As we have already mentioned, one can identify two factors: the first factor deals with abortion, the second with capital punishment. When we look at Figure 1, the plot of the centroids over educational categories on the first dimension, going from left to right, one finds points whose projections are ordered from very religious (Reformed, Protestant) and political "right" towards religious, and no political preference towards non-religious and political "left" or "liberal". This is also illustrated in Table 4, in which the mean scores on the abortion variables and the scores on the first factor are given for all combinations of "religion" and "political preference", a high mean score indicating a pro-abortion opinion.

From Table 4 it is evident that the mean scores for all combinations of both background variables and the corresponding scores on the first factor can be ordered almost perfectly in a double-monotone way. It is also possible to make an almost perfect monotone ordening of these scores over all cells. These facts support our conclusion that on the first factor at the left side one finds individuals who are anti-abortion, politically at the "right" and religious versus individuals who are pro-abortion, politically at the "left" or "liberal" and non-religious at the right side.

Table 4. Mean scores on the abortion variables, scores on the first factor and frequencies.

	R1	DEN	NON	LEF	LIB
	1.57	3.08	3.11	3.25	3.25
REF	-2.951	-.762	-.700	-.355	-.444
	11	33	7	4	5
	1.75	2.77	3.46	3.73	3.90
PRO	-2.468	-1.257	-.313	-.024	.185
	5	28	24	31	12
	2.75	2.76	3.64	3.95	4.02
RC	-1.540	-1.106	-.013	.369	.410
	1	60	46	31	26
	--	4.88	4.14	4.34	4.52
NON	--	1.555	.533	.834	1.016
	--	2	65	102	42

To interpret the second factor we look at Figure 2 and 3. From these plots it becomes clear that the educational level is most important for this factor. One can see that for both background variables, political preference and religion, the centroid points are roughly ordered from low educational level at the top towards high educational level at the bottom of the plot. On the contrary, the corresponding categories of the other variables have no clear ordering on the second dimension. It appears that people with a higher educational level are much more against capital punishment than those with a lower educational level. It also appears that the centroid points of both figures can only be ordered over all cells according to the mean score on the capital punishment variables.

Figure 2. Centroids over categories of the variable religion: combinations of political preference and educational level.

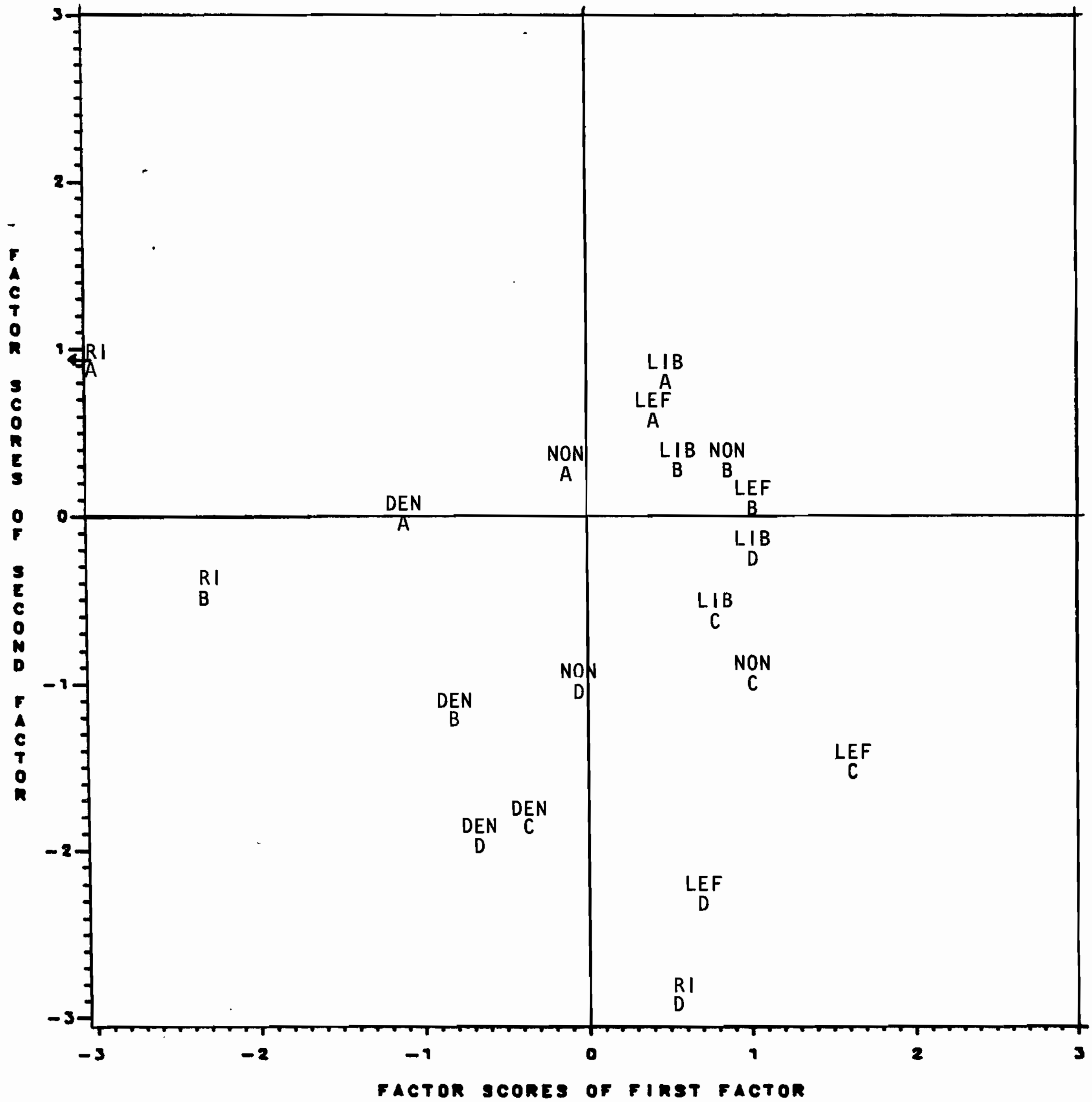
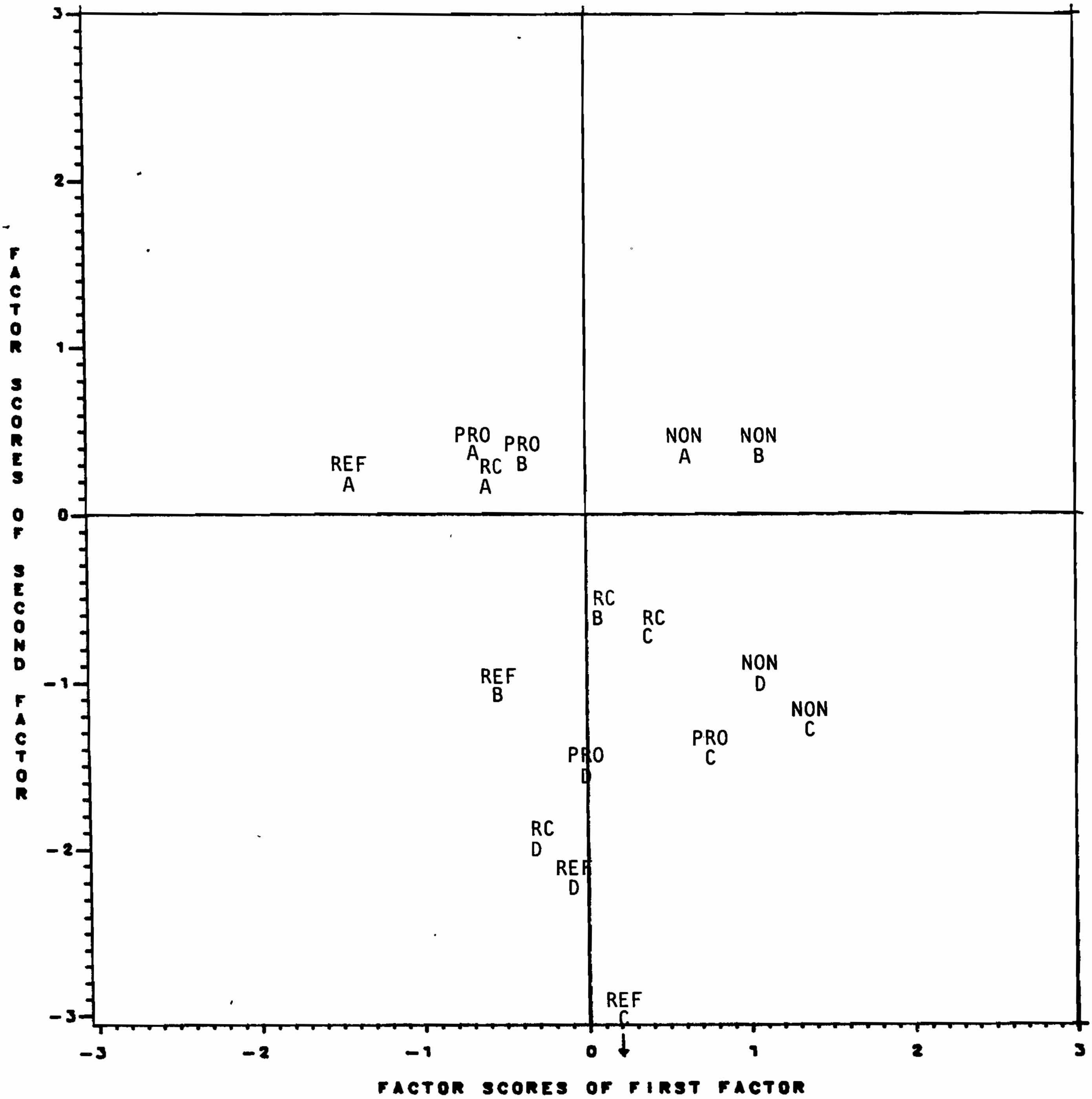


Figure 3. Centroids over categories of the variable political preference: combinations of religion and educational level.



Such an ordering will be somewhat less perfect than the one on the abortion factor, but a double monotone ordering cannot be accomplished. The horizontal spread of the points in Figure 2 and 3 can be explained by interaction with "religion" and "political preference". Roughly one finds again religious and political "right" versus non-religious and political "left" or "liberal". So concluding one can say that on the second dimension (at the top) one finds individuals with a low educational level who favour capital punishment versus (at the bottom) individuals with a high educational level who are strongly against capital punishment.

Example 2

For our second example we have analyzed data collected with a 63-item suicide-attitude questionnaire.
This questionnaire has been constructed and used by Diekstra and Kerkhof in a large-scale study on attitudes towards suicide (Diekstra and Kerkhof, 1985). The data we used are the results of the administration of this questionnaire, in 1975, to a sample of 712 subjects from the population of Nijmegen, a medium size town in the eastern part of the Netherlands.
We analyzed the same 19 scales that Diekstra and Kerkhof constructed out of the original 63 items.
These scales combine so called "referents" and "attitude-components". The term "referents" refers to the fact that people may differ in their opinions about suicide with respect to themselves, their most beloved, and people in general. As "attitude-components" one can distinguish between affective, cognitive and instrumental components.
Of course in this context it is not possible to go into further detail with regard to the content of the scales. In the following we will denote them by abbreviations such as "affective-beloved", "instrumental-self", etc.

For our designmatrix we used the background variables age, educational level and membership of broadcasting organization. In this case age is a numerical variable, collected in the first column of the designmatrix and we also added the second and third power of this variable.

In Table 5 a brief description of the scales and the background variables is given.

Table 5. Description of the variables in the analysis.

- Nineteen scales with respect to suicide:

AFFS. affective-self	AFFB. affective-beloved	AFFP. affective-people
ABNS. abnormality-self	ABNB. abnormality-beloved	ABNP. abnormality-people
CONS. consequences-self	-	CONP. consequences-people
RIS. right to -self	-	RIP right to-people
INSS. instrumental-self	INSB. instrumental-beloved	INSP. instrumental-people
FYSS-fysical-self	FYSB. fysical-beloved	FYSP. fysical-people
SOCS. social. self	SOCB. social-beloved	SOCP. social-people

- Three background variables:
  1. Age (AGE), numerical, ranging from 16 to 71 years.
  2. Educational level (EDU) with categories: (1) = LO* (A), (2)= LBO/MAVO (B),(3)=MBO/HAVO (C),(4)= HBO (D), (5)= University (E).
  3. Broadcasting organization (BO) with categories: (1) = none (NON), (2) = KRO**, (3)=VARA, (4)= AVRO, (5)= NCRV, (6) = VPRO, (7) = EO, (8) = TROS.

* As we have already mentioned we will not translate these categories. Although compared with EDU in example 1, there is a slightly different division of categories, E indicates the highest level.

** In the Dutch broadcasting system a number of broadcasting organizations or "unions" operate at the same time. People can obtain membership of these "unions" and each one is having its own identity. Roughly speaking KRO, NCRV and EO are religious and politically ranging from "the middle" (KRO) to "the right" (EO), AVRO and TROS are liberal and VARA and VPRO are non-religious and politically at "the left".

All scales are recoded to five-point scales, a high score indicating a tolerant attiude towards suicide. Note that all possible combinations are included. According to Diekstra and Kerkhof these missing combinations are not possible or meaningful.

After removing the respondents with missing observations on the scales and the background variables, 545 cases were left for analysis.

As in the previous example we applied various two-factor models to the data in order to establish which model would fit the best. Table 6 summarizes the results.

Table 6. Background variables, interactions, - 2 log likelihood-values and the total number of parameters to be estimated for the various models.

variables, interactions	- 2 log likelihood	total number of parameters to be estimated
AGE	33,364.03	55
AGE, AGE^2	33,316.38	57
AGE, AGE^2, AGE^3	33,304.01	59
EDU, BO, AGE		
EDU* BO, EDU*AGE, BO*AGE		
EDU*BO + AGE	32,52434	175
EDU, BO, AGE, AGE^2		
EDU*BO, EDU*AGE, EDU*AGE^2		
BO*AGE, BO*AGE^2,		
EDU*BO*AGE, EDU*BO*AGE^2	32,286.49	233
EDU, BO, AGE, AGE^2, AGE^3		
EDU*BO,EDU*AGE, EDU*AGE^2,EDU*AGE^3		
BO*AGE, BO*AGE^2, BO*AGE^3		
EDU*BO*AGE, EDU*BO*AGE^2, EDU*BO*AGE^3	33,099.91	287

Table 7. Factor loadings (covariances) for the two factor solution with regard to the "AGE, AGE^2, AGE^3"-model(1) and the most complex model (2).

Variables	(1) F_1	(1) F_2	(2) F_1	(2) F_2
AFFS	.35	-.03	.47	-.200
AFFB	.32	.01	.46	-.16
AFFP	.18	-.09	.30	-.16
ABNS	.34	.05	.60	.28
ABNB	.20	-.06	.34	-.05
ABNP	.27	.04	.65	.03
CONS	.26	.05	.49	.41
CONP	.24	.13	.47	.46
RIS	.17	.01	.30	-.04
RIP	.35	.06	.65	.06
INSS	.12	.08	.23	-.04
INSB	.06	.10	.28	.10
INSP	.24	.11	.42	.01
FYSS	.39	-.03	.61	-.22
FYSB	.24	-.03	.47	-.25
FYSP	.02	-.01	.26	-.05
SOCS	.39	-.23	.39	-.38
SOCB	.36	-.10	.36	-.38
SOCP	.15	.05	.19	-.02

Again, all differences between the -2 log likelihood-values of the most complex model and the other models are significant, so this model is the most appropriate.
In order to illustrate our technique we have included the models of which the designmatrix is constructed only out of the variable "AGE". So in the case of the "AGE, AGE^2"-model for instance, one can think of this matrix as having only two columns.
Table 7 gives the factor loadings for the "AGE, AGE^2, AGE^3"-model and the most complex one. Note that we have analyzed a covariance-matrix, so the loadings are also covariances.

In table 7 we see that the loadings of model (1) are much lower than those of model (2). In fact, one can hardly speak of a second factor in model (1).
For model (2) we can try to identify the factors. On the first factor all loadings are positive and most of them are moderately high. The scales dealing with "the right to commit suicide for people in general" and "the question whether people who commit suicide are abnormal" have the highest loadings.
The second factor roughly distinguishes between the affective, fysical and social components on the one hand and the abnormality-self and consequences components on the other hand.

Such an interpretation supports the conclusion of Diekstra and Kerkhof, that the first factor deals with general tolerance towards suicide, while the second factor makes a distinction between emotional and rational arguments.
In terms of factorscores this means that a high score on the first factor indicates tolerance towards suicide. A high score on the second factor means more emphasis on rational rather then on emotional aspects.
In order to illustrate the importance of the interaction of the backgroundvariables, in fig. 4 we have plotted the factorscores of model (1) against the variable AGE for the first factor. Also in this plot the centroids are marked for most combinations of the variables EDU and BO.

Figure 4. Factorscores of the first factor against the values of AGE for model (1). Centroids for some combinations of EDU and BO.

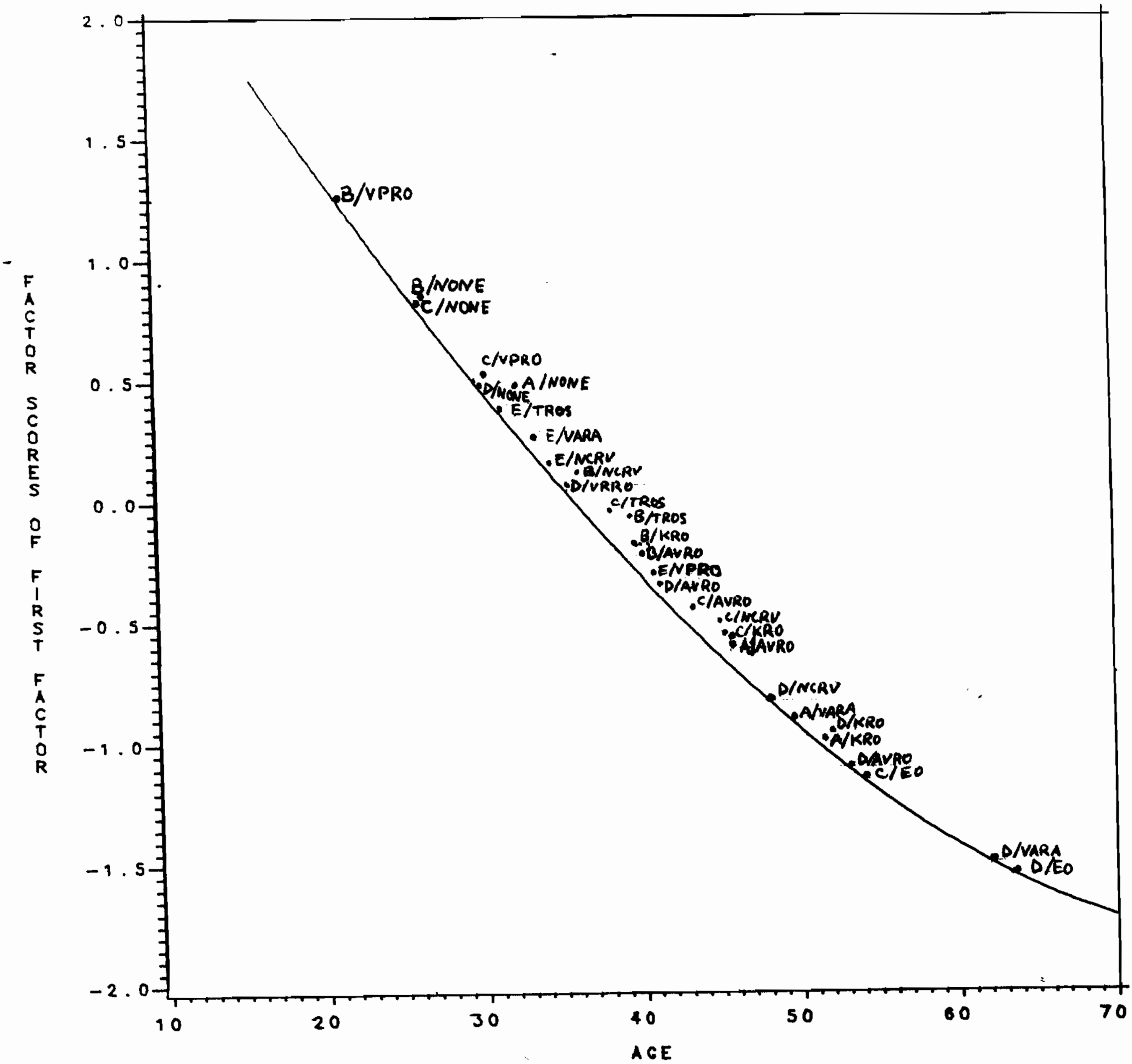


Table 8 shows the contingency table of these variables.

Table 8. Contingency table of the variables EDU and BO.

		BO								
		NONE	KRO	VARA	AVRO	NCRV	VPRO	EO	TROS	
	A	33	28	8	34	6	-	-	19	128
	B	79	43	17	46	10	1	-	18	214
EDU	C	49	31	6	16	13	8	2	10	135
	D	18	11	2	7	1	7	1	3	50
	E	6	3	3	1	1	3	-	1	18
		185	116	36	104	31	19	3	51	545

Fig. 4 reveals two arguments for the existence of interaction:

- Ordering the variable BO along the curve given one category of EDU and vice versa, gives a different ordering for each combination.
- AGE has a different distribution for each group of respondents determined by a combination of BO and EDU.

Although, as we can see from table 8, there exist a number of really small groups, which will be very unstable, a trend is still visible. In general, one can conclude that older individuals tend to be less tolerant towards suicide.
Subsequently we have plotted the factorscores of the most complex model but one (the model with AGE, AGE^2 and all interactions), against the values of AGE for the eight largest groups of respondents as determined by the combination of BO and EDU (see table 8). Although model (2) see table 7) is the most appropriate, for reasons of interpretation we decided not to plot the factor-scores of this model.

The reason for doing so is that the inspection of the third degree polynomials revealed that the extreme points of these curves were strongly determined by very few individuals and therefore very unstable.
So the factorscores will be on polynomials of second instead of third degree. For each factor the plots of the factorscores are given in figure 5 and 6.

As can be seen from figure 5 and 6 the polynomials indicate different relationships between factorscore and age for the various groups of respondents. These differences become visible while in drawing polynomials for specific combinations of EDU and BO one removes the interaction of these variables.
For the first factor (figure 5), some polynomials behave corresponding to the general trend: older people tend to have lower factorscores indicating less tolerance towards suicide. For some groups there is a slight increase in tolerance with increasing age, but only to the age of approximately 35 years. Finally for the other groups there is an increase in tolerance starting somewhere around the ages between 40 and 50 years.
Looking at figure 6 it becomes clear that for the second factor there is a general trend in older people having higher factorscores indicating a more rational attitude towards suicide. For some groups however there is a change towards a more emotional view, starting somewhere around the age of 45 years (except for the C/NON-group, probably due to a somewhat unfortunate distribution of AGE). There is also one group (A/AVRO) which shows more emotion with increasing age up to approximately 45 years and then becomes rational again.
When we want to have more information regarding factorscores and background variables, these plots do not give much insight. Therefore we calculated the mean factorscores for each possible group for both factors. Tables 9 and 10 show the results.

Figure 5. Factorscores of the first factor against the values of AGE for the eight largest groups of respondents.

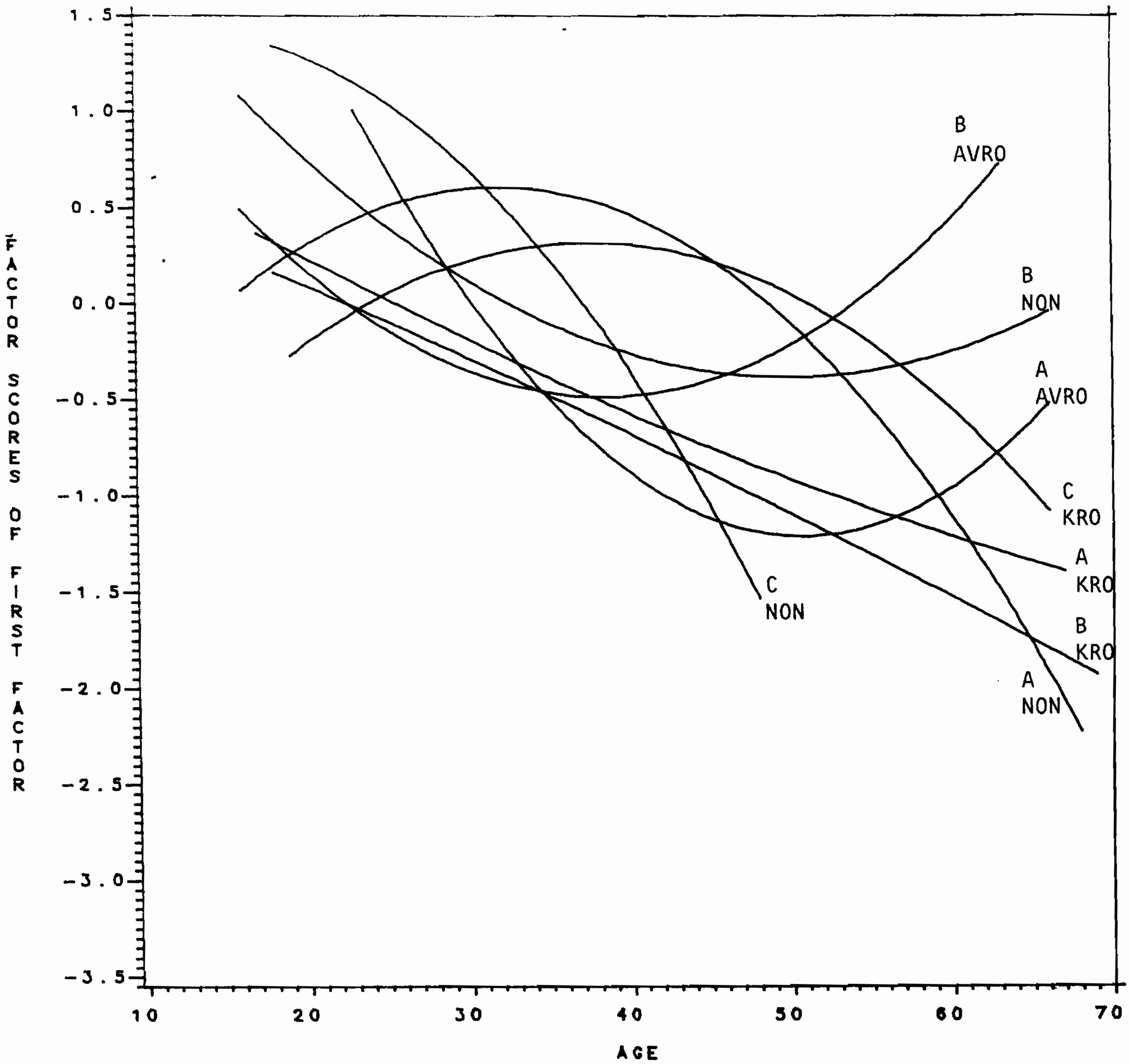


Figure 6. Factorscores of the second factor against the values of AGE for the eight largest groups of respondents.

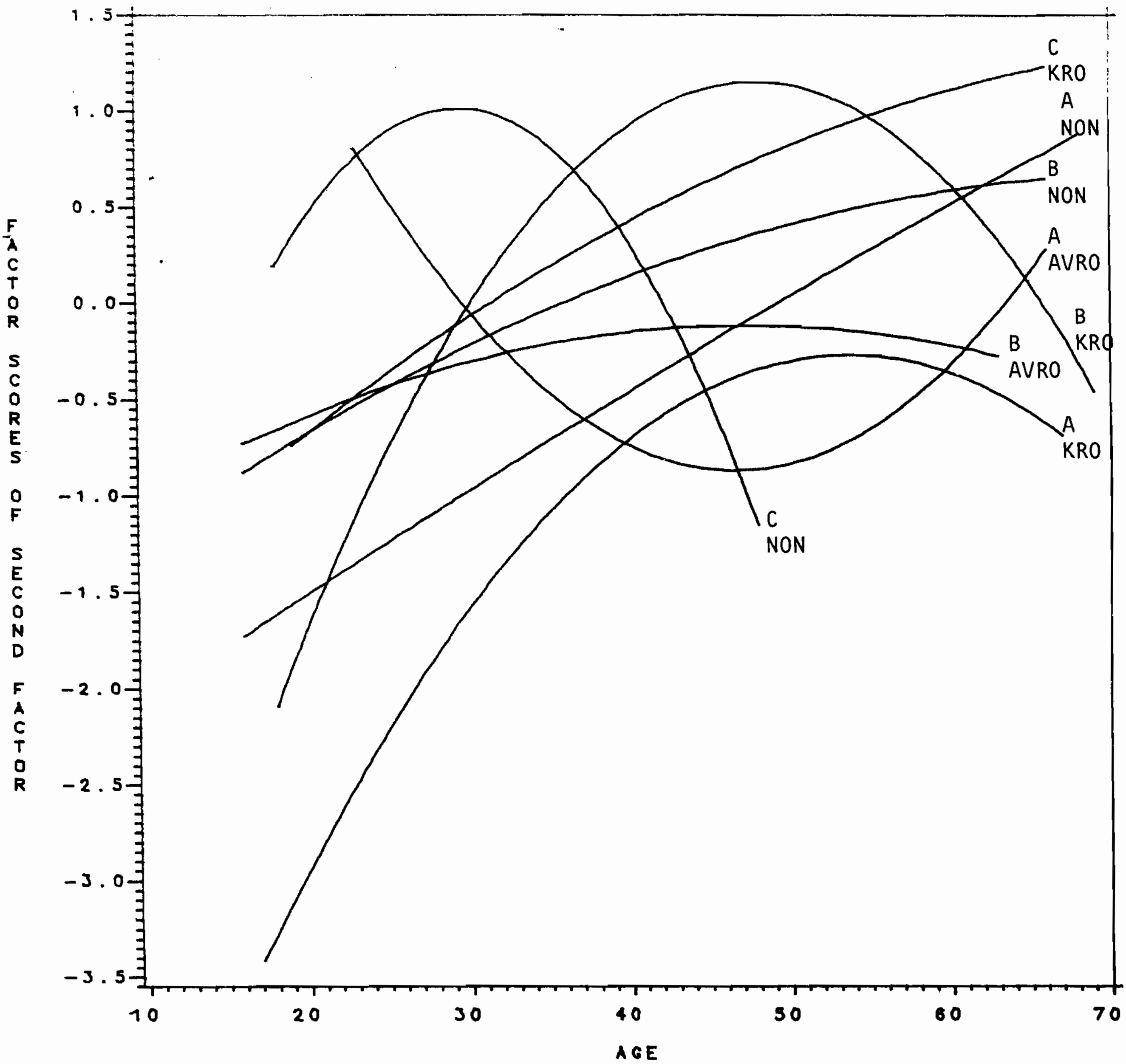


Table 9 Mean factorscores on first factor for all combinations of the variables EDU and BO.

		BO							
		NONE	KRO	VARA	AVRO	NCRV	VPRO	EO	TROS
	A	-.01	-.95	-.42	-.76	-.51	-	-	-.29
	B	.46	-.72	.38	-.23	-.1.11	1.25	-	-.37
EDU	C	.86	-.08	.72	-.23	-.94	1.94	-.50	-.95
	D	.87	-.50	-.40	-.21	.39	1.02	-.12	-.36
	E	2.20	.34	1.45	.46	1.87	3.64	-	.44

From table 9 it becomes clear that higher educated people show more tolerance towards suicide. Also people who are no member of any broadcasting organization or have a membership of progressive broadcasting organizations (VARA, VPRO) tend to be more tolerant. People with lower educational level and people who favour religious or political "at the right" broadcasting organizations tend to be more restrictive. Combinations of EDU and BO speak for themselves. Interpretation of table 10 is much less clear. It looks as if with increasing educational level there is a change from a rational towards a more emotional argumentation. A relation between mean factorscore and broadcasting organization does not seem to exist. We must keep in mind however that this second factor is not so important. Also because of the possibly unstable, small groups of respondents we have to be very careful with our interpretation.

Table 10 Mean factoscores on second factor for all combinations of the variables EDU and BO.

		BO							
		NONE	KRO	VARA	AVRO	NCRV	VPRO	EO	TROS
	A	.87	.52	2.23	.48	-.31	-	-	1.30
	B	.42	-.24	-1.17	.23	-1.0	-.52	-	-.12
EDU	C	-.63	-.58	-1.08	-.45	-.27	-.04	-.71	.38
	D	-.87	.20	-.14	-1.54	-2.49	-.48	3.0	.21
	E	-.71	.15	-1.31	-1.90	-1.76	-.01	-	1.04

Reference notes

Diekstra, R.F.W. & Kerkhof, A.J.F.M. (1985). Attitudes towards suicide. The development of a suicide-attitude questionnaire (Suiatt) (unpublished paper, to be presented to Journal of Clinical Psychology).

References

Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics, 22,* 327-351.

Anderson, T.W. (1984). The 1982 Wald Memorial Lectures. Estimating linear statistical relationships. *Annals of Statistics, 12,* 1-45.

Anderson, T.W. & Rubin, H. (1956). Statistical inference in factor analysis. In: Jerry Heyman (ed.), *Proceedings of the Third Berkely Symposium in Mathematical Statistics and Probability, 5,* 111-150.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22,* 113-137.

Etezadi-Amoli, J. & McDonald, R.P. (1983). A second generation nonlinear factor analysis. *Psychometrika, 48,* 315-342.

Fisher, R.A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenetics, 8,* 376-386.

Garnett, J.C.M. (1919). On certain independent factors in mental measurement. *Proceedings of the Royal Society of London, 46,* 91-111.

Gleser, L.J. (1981). Estimation in a multivariate 'errors in variables' regression model: large sample results. *Annals of Statistics, 9,* 24-44.

Harman, H.H. (1960). *Modern factor analysis.* Chicago: University of Chicago Press.

Hemelrijk, J. (1966). Underlining random variables. *Statistica Neerlandica, 20,* 1-8.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443-482.

Lawley, D.N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, Sect. A 60,* 64-82.

Lawley, D.N. (1941). Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh, Sect A 61,* 176-185.

McDonald, R.P. (1979). The simultaneous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology, 32,* 212-228.

Mooijaart, A. (1985). Factor analysis for non-normal variables. (in press).

Rao, C.R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika, 20,* 93-111.

Theobald, C.M. (1975). An inequality with application to multivariate analysis. *Biometrika, 62,* 461-466.

Thurstone, L.L. (1947). *Multiple-Factor Analysis.* Chicago: University of Chicago Press.

Van den Wollenberg, A.L. (1977). Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika, 42,* 207-219.

Veenhoven, R. & Hentenaar, F. (1975). Nederlanders over abortus. Stimezo-onderzoek, nr. 3.

Whittle, P. (1952). On principal components and least squares methods of factoranalysis. *Skandinavisk Aktuarietidskrift, 35,* 223-239.

Young, G. (1941). Maximum likelihood estimation and factor analysis. *Psychometrika, 6,* 49-53.