PRINCALS

Albert Gifi

Department of Data Theory
Leiden 1985

Table of contents

I.1. Introduction.

I.1.1.

PRINCALS is an acronym for PRINcipal Components analysis by means of Alternating Least Squares. In this User's Guide the term PRINCALS refers to the computer program with that name. It is a program with a number of options, with the effect that PRINCALS includes quite a family of solutions. One of them is the solution for PCA (Principal Components Analysis): this solution requires that all variables are treated as numerical. Another option requires that all variables are treated as multiple nominal: the PRINCALS solution then becomes the same as a HOMALS solution. There are specialized computer programs for PCA and HOMALS, and the user is advised to use those specialized programs if the interest is in a PCA or HOMALS solution only.

I.1.2.

PRINCALS can be introduced in different ways. In this User's Guide PRINCALS is introduced as an extension of PCA. A PCA computer program can handle numerical variables only; PRINCALS liberates such a program in the sense that PRINCALS also can handle ordinal variables and nominal variables, as well.

I.1.3.

This guide is non-technical: it is meant to be a practical outline which may help users to organize input and to understand output. There will be little attention for mathematical derivations. In particular, the rationale of the computer algorithms will not be discussed. A more sophisticated discussion of PRINCALS is given in Gifi (1981a, p. 163-196).

I.1.4.

The notation in this User's Guide is in agreement with the notation used in Gifi (1981a).

I.2. Terminology.

I.2.1. Data matrix.

The input for PRINCALS is a *data matrix* H, with n rows for objects, and m columns for variables. *Objects* are the units of observation, such as individual schoolchildren, or hospitals in The Netherlands, or countries in Europe. *Variables* refer to what is observed (schoolchildren: age, performance level, occupation of father - hospitals: number of beds, available specialistic treatments - countries: type of governement, number of inhabitants). Variables sort objects into *categories*. Countries with the same type of governement are in the same category. When a continuous variable is measured, such as age, measurements will be rounded-off to 'number of years'. In practice, therefore, this variable also sorts objects by a limited number of age categories.

I.2.2. Quantification.

PRINCALS quantifies the data matrix H, by assigning numerical values to the different categories of each variable. This produces a *quantified data matrix* Q (with n rows and m columns, just like H). PRINCALS also gives a quantification to objects in the form of a column of n *object scores* x. PRINCALS may give a number of different solutions for Q and x - they are called *dimensions*. We therefore need an index s to indicate the solutions Q_s and x_s for each dimension, where s runs from 1 to p (p is the total number of dimensions).

The criterion for quantification of H is that x_s should have large correlations with each of the variables in Q_s - a solution is "good" to the extent that this criterion is satisfied.

I.2.3. Treatment of variables.

PRINCALS quantifies variables in such a way that columns of Q_s have zero mean and unit variance (standard scores). Apart from that, there are four ways how variables can be quantified:

(i) numerical

(ii) ordinal

(iii) single nominal

(iv) multiple nominal

(i) Numerical. This assumes that the observed variable h_j already has numerical values for its categories. In the PRINCALS quantification of h_j these numerical values are respected in the sense that only a transformation on interval scale is permitted. But since we also require q_j to be standardized, there is only one possible solution for the quantified variable q_j. In fact, if all variables are treated as numerical, PRINCALS gives the same solution as classical PCA.

(ii) Ordinal. Ordinal quantification means that the categories of the quantified variable q_j have the same order as those of the observed variable h_j. More precisely: if h_{gj} and h_{ij} are the two observed values for objects g and i, quantified as q_{gj} and q_{ij}, then

if $h_{gj} = h_{ij}$, then $q_{gj} = q_{ij}$

if $h_{gj} > h_{ij}$, then $q_{gj} \geq q_{ij}$.

(iii) Single nominal. Now the only restriction on the quantification becomes

if $h_{gj} = h_{ij}$, then $q_{gj} = q_{ij}$.

In words: objects in the same category for h_j obtain the same quantification.

The three possibilities above have in common that h_j has the same quantification on all dimensions. The j'th column of Q_s is the same, irrespective of s. The three possibilities vary in the amount of restriction imposed on the quantification. Numerical quantification is very much restricted. Ordinal quantification gives more freedom, nominal quantification has the least amount of restriction.

(iv) Multiple nominal. This type of quantification differs from single nominal in that the quantification of h_j can be *different* for each dimension: the h'th column of Q_1 will not be the same as the h'th column of Q_2, etc.

I.2.4. PRINCALS options for treatment of variables.

PRINCALS has five options for treatment of variables. The first four options imply that all variables are treated in the same way: all of them numerical, or all of them ordinal, or all of them single nominal, or all of them multiple nominal. The fifth option is the *mixed* treatment: the user then has to specify for each separate variable how it must be treated.

PRINCALS does not yet include the possibility to treat variables as "multiple ordinal". This would mean that the quantification of h_j could be different at each dimension while obeying the ordinal restrictions. It follows that the ordinal option, described above, can be called the "single ordinal" treatment of variables. PRINCALS also has no option for "multiple numerical". However, the reason why this option is not available is quite different: the numerical restriction makes it impossible to quantify a variable in different ways. There is no freedom for it.

I.3. Remarks on treatment of variables.

I.3.1. Introduction.

This section contains some miscellaneous remarks about the organization, by the user, of PRINCALS input.

I.3.2. Binary variables.

A variable is said to be *binary* if it has only two categories. In this case it does not matter whether the variable is defined as nominal, ordinal, or numerical. Whatever quantification of the two categories is taken, it will automatically satisfy the requirement for numerical variables. (If *all* variables are binary, a different type of analysis will probably be better.)

I.3.3. The researcher must define.

Variables are not nominal, ordinal, or numerical on the basis of mysterious intrinsic properties, but because the researcher defines them as such.
Two extreme examples clarify this point.

(i) Suppose a variable sorts objects into age-groups. Age can very well be treated as a numerical variable. However, the reseacher may decide to treat age as ordinal. E.g., suppose one has a study on "driving safety", in wich one of the variables is "driver's age". It may very well be so that "safety" is related to age only for people younger then 25, and that above that age it does not matter how old the driver is. Ordinal treatment of age then would be more realistic than numerical treatment. Or, it may even be true that for people younger than 25, safety has positive relation with age; whereas for people older than 60 safety has negative relation with age. In that case age might better be treated as a nominal variable.

(ii) A variable that sorts persons by "political preference" looks as if it is essentially nominal. But the researcher may be interested in a "left-right" dimension, and on that basis might order political parties from left to right, and might require that quantification of parties to respect that order. The researcher might even have reasons to assume interval properties for the apriori quantification of parties (e.g., based on earlier research), and then might require that "political preference" is quantified as a numerical variable.

I.3.4. Number of categories.

Very often the number of categories is a natural one. Political preference is an example: if there are k political parties, then political preference is a variable with k categories. On the other hand, there is no natural way to classify persons into age groups: we can form as many categories as we want.

A rule of thumb is that there should be no more than 12 categories (each of them with not too small marginal frequency). For a variable like "age", this implies that one should round-off in such a way that about 12 reasonably well filled categories remain. For a variable like "political preference", if there are very many political parties, it implies that the researcher should consider whether some parties can be "taken together".

I.3.5. Merging of categories.

Researchers often worry whether some categories can be taken together, or merged. Essentially, their problem then is whether or not such categories will obtain the same quantification or not.

Instead of deciding on an apriori basis, one might decide after the analysis: if categories obtain the same, or very similar quantification, they may as well be merged, without affecting final results.

I.3.6. Ordinal variables.

It often happens that categories can be ordered on an a priori basis. An example is that objects are classified on an attitude scale in categories ranging from "in favor of something" to "against something", with "don't know" as an intermediate category. But it may very well happen that the category "don't know" defies the interpretation of being midway between "in favor" and "against". A nominal interpretation of the variable then might be more attractive. Also, it may happen that responses to an attitude scale are related to other variables not so much from "in favor" to "against", but from "extreme attitude" (either much in favor or much against) to "not outspoken attitude". A nominal treatment of the variable then might reveal better what is going on.

I.3.7. Judicious choice.

Examples above show that it is the researcher's responsibility to decide how variables must be treated. It sometimes might be useful to try out different possibilities (ordinal treatment of the attitude scale, versus nominal treatment; numerical treatment of age versus ordinal treat ment, etc.).

I.4. Some algebraic relations in PRINCALS.

I.4.1. Numerical illustration.

Table I.4.1. illustrates numerical relations implicit in a PRINCALS solution. The data in the table are the Guttman-Bell data (further discussed in section III.1). Table I.4.1. gives the original data matrix H with the quantified data matrix Q for the first dimension of a two-dimensional single ordinal analysis. The algebraic relations described in this section, however, are valid for any type of analysis: numerical, single ordinal, single nominal or multiple nominal.

(i) The column $\underline{a}$ gives correlations between object scores x and the m columns of Q.

(ii) The average of the squared correlations is the eigenvalue $\Phi = \Sigma a_j^2/m$ (j=1,..,m).

(iii) The column of object scores x is a weighted sum of the columns of Q, with weights equal to the values given in $\underline{a}$, divided by mΦ. So we may write $x = \Sigma q_j a_j / m\Phi$.

(iv) The table also gives the matrix Ŷ (with the same format as H and Q: n rows, m columns). Entries in Ŷ are averaged object scores x, averaged over the objects which are in the same category of a variable. E.g., in

Table I.4 1: First dimension of a two-dimensional single ordinal solution

H

1	1	1	2	2
2	2	2	2	2
1	1	2	1	1
4	2	4	2	3
4	4	4	2	3
3	3	3	1	2
2	3	3	2	2

Q

-1.525	-1.550	-2.186	.632	.017
.342	.504	-.325	.632	.017
-1.525	-1.550	-.325	-1.581	-2.183
.992	.504	.709	.632	1.057
.992	1.086	.709	.632	1.057
.381	.504	.709	-1.581	.017
.342	.504	.709	.632	.017

→ $\underline{a}$

-.986
-.967
-.737
-.485
-.845

↓ x

1.261
-0.259
1.720
-0.941
-1.109
-0.185
-0.487

← Ŷ

1.490	1.490	1.261	-.307	.082
-.373	-.600	.730	-.307	.082
1.490	1.490	.730	.767	1.720
-1.025	-.600	-1.025	-.307	-1.025
-1.025	-1.109	-1.025	-.307	-1.025
-.185	-.336	-.336	.767	.082
-.373	-.336	-.336	-.307	.082

variable 1 there are two objects in category 1 (objects 1 and 3) with object scores 1.2609 and 1.7197. The average is (1.2609 + 1.7197)/2=1.490, which is the value given to category 1 in the first column of $\hat{Y}$.

I.4.2. PRINCALS output.

Results above are in the following way related to PRINCALS output.

(i) For numerical, ordinal and single nominal variables: the elements to be used in Q are given under the heading *category quantifications*.

(ii) Values of $\underline{a}$ are given under the heading *component loadings*.

(iii) Values of a_j^2 are found under the heading *single fit*.

(iv) Values to be used in $\hat{Y}$ appear under the heading *multiple category coördinates*.

(v) PRINCALS also gives *single category coördinates*. They are equal to the category quantifications, multiplied by the corresponding value of a_j (first column of Q multiplied by a_1, second column of Q multiplied by a_2, etc.) As a consequence, if Q were replaced by a matrix of single coordinates, object scores x can be found as the sum of the columns of this revised matrix, divided by $m\Phi$.

(vi) PRINCALS gives these objectscores, with normalization N (i.e. Σx_{is}^2 =n). Objectscores for different dimensions are uncorrelated.

For *multiple nominal variables* PRINCALS output is different. First of all, it is characteristic of the quantification of multiple nominal variables that categories are quantified proportional to the average object score of objects in the corresponding category. In the terminology of table I.4.1.: a column $\hat{y}_j$ will be equal to $q_j a_j$: single category coördinates are identical to multiple category coördinates. PRINCALS gives these category coördinates under the heading *category quantifications*. One should be aware, though, of the difference with category quantification for single variables. For single variables the category quantification gives values to be used in Q (columns of Q have unit variance), whereas for multiple nominal variables the category quantification gives values to be used in $\hat{Y}$ (columns of $\hat{Y}$ have variance equal to a_j^2). Secondly, for multiple nominal variables PRINCALS does not give component loadings a_j. Squared component loadings are found under the heading *multiple fit*. The positive square root of multiple fit corresponds to the correlation between object scores and quantified variable.

I.4.3. Sign of component loadings.

Table I.4.1. shows all component loadings with negative sign: all correlations between x and columns of Q are negative. Interpretation of results would be more simple if these correlations were positive. This can be achieved as follows: reverse the sign of x (replace x by -x), and reverse the sign of all a_j (replace $\underline{a}$ by $-\underline{a}$). Or, one may want to reverse the sign of some a_j: this can be done by also reversing the sign of the category quantification of variable h_j.

I.4.4 Single versus multiple.

Figure I.4.1. gives a schematic illustration of the difference between single versus multiple treatment of variables. With all variables single, the quantified data matrix Q is the same for each dimension. Successive dimensions differ only in the solution for $\underline{a}$ and x. With all variables multiple nominal, the quantified data matrix will be different for each dimension: Q_1 will be different from Q_2, etc. In the mixed case (e.g., some variables multiple nominal, other variables single), columns of Q_s will remain the same for single variables, but become different for multiple variables.

I.5. Comparison of PRINCALS solutions.

I.5.1. All variables numerical.

This solution always is single: Q is the same at each dimension. In fact, Q just replaces H by columns of standard scores (zero mean, unit variance), and PRINCALS gives the same solution as classical PCA.

I.5.2. All variables multiple nominal.

This solution is the same as a HOMALS solution. We shall not discuss this solution further (consult Gifi A., HOMALS User's Guide 1981b).

I.5.3. Nested versus not-nested.

The PCA solution and the HOMALS solution are *nested*. It means that the first (p-i) dimensions of a p-dimensional solution are the same as the dimensions of a (p-i) dimensional solution. E.g., if p=4, the first dimension will be the same as the solution with p=1, the first two dimensions will be the same as the solution with p=2, the first three dimensions will be the same as with p=3.
PRINCALS solutions with single ordinal and single nominal variables (or with mixed variables) are *not nested*. The first two dimensions of a p-dimensional solution are not identical to the solution with p=2. The first two dimensions will be different, depending on the value of p.
Users educated in the tradition of factor analysis might surmise that these differences are a matter of "rotation"; they are not.

I.5.4. Two phases.

For the options single ordinal, single nominal, and mixed, PRINCALS proceeds in two phases.

(i) If all variables are single, in the first phase the numerical PCA solution is calculated. This solution is used as the initial configuration in the second phase which identifies the proper PRINCALS solution asked for.

(ii) In the mixed case, with some variables single and others multiple nominal, the first phase identifies a solution in which all single variables are treated as numerical, and all multiple nominal variables as multiple nominal. Again, this solution serves as the initial configuration for the second phase. Results of the first phase are printed on request. The user might find it interesting to compare these results with the final results. This comparison might be useful in particular for single variables (the comparison shows to what extent ordinal or nominal treatment improves upon numerical treatment).

I.5.5. <u>One-dimensional single nominal.</u>
Just for completeness: the one-dimensional solution with all variables single nominal is the same as the one-dimensional solution with all variables multiple nominal (and therefore is the same as one-dimensional HOMALS). As fig-

<u>Figure I.4.1:</u> Schematic representation of treatment with all variables single, or all variables multiple, and the mixed case

<u>Figure I.4.1a:</u> All variables single

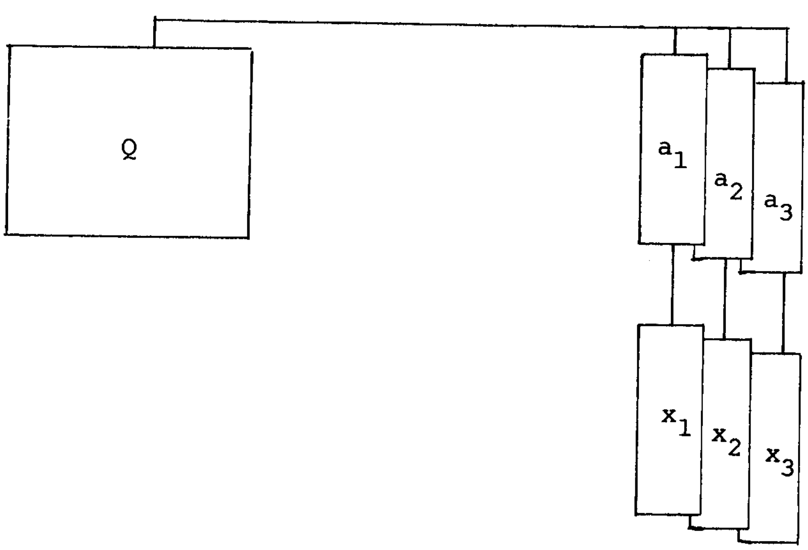


Figure I.4.1b: All variables multiple nominal

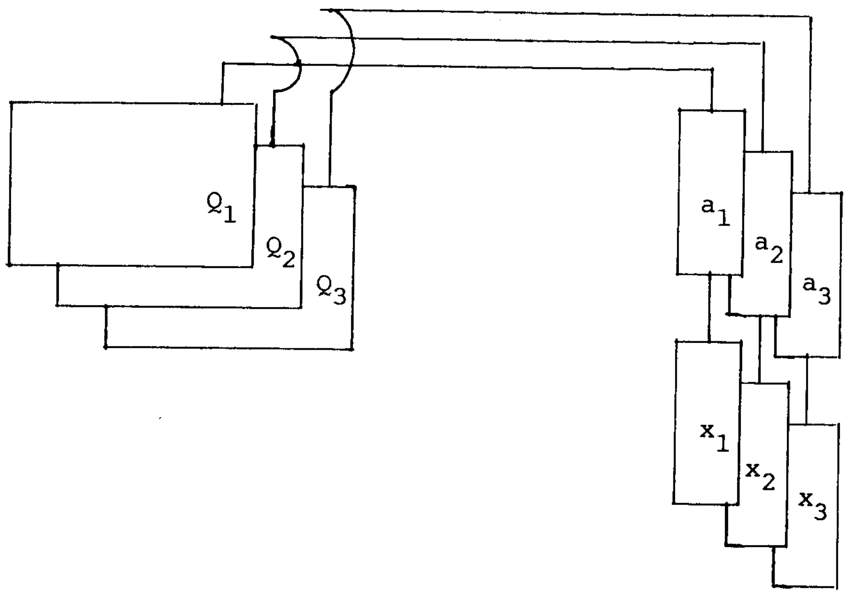
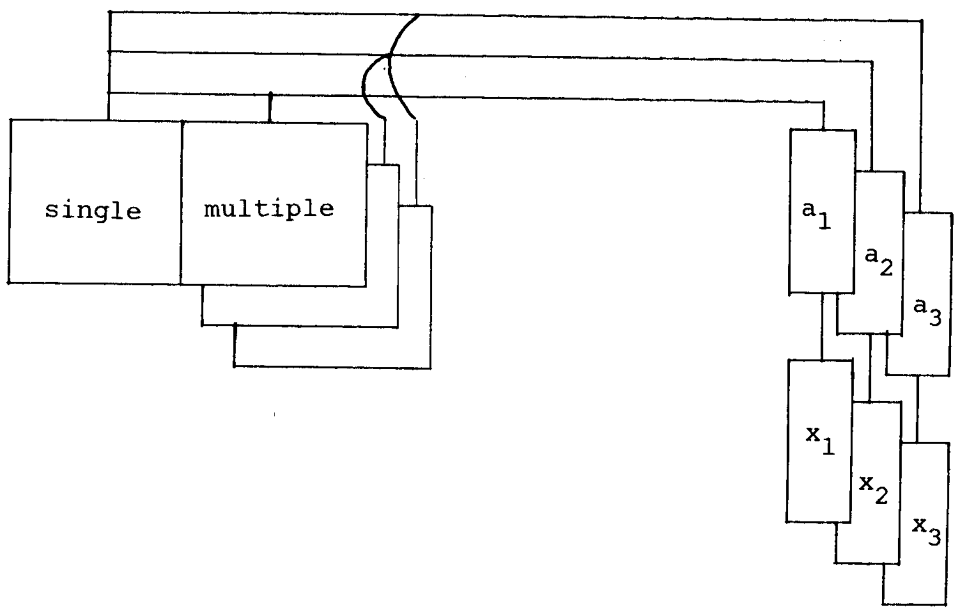


Figure I.4.1c: Mixed



ure I.4.1 shows, the difference between single and multiple becomes relevant only for solutions with two or more dimensions.

I.5.6. Multiple nominal variables.

When a variable is treated as multiple nominal, it obtains different quantifications at each dimension. The user should not make the error to think of these different quantifications as if it is the "same" variable h_j. E.g., suppose that h_j gives categories of "occupational status". The solution in the first dimension shows relations between "occupational status" and other variables. But the solution in the second dimension identifies a different sort of "occupational status": one should not interpret those different quantifications as if they are just "occupational status". In fact, these different quantifications may be very lowly correlated.

This problem is related to that of "component loadings". Component loadings are defined as the correlation between x_s and q_{js}. For single variables q_{js} is the same, whatever s. For multiple nominal variables q_{js} is different, depending on s. Component loadings for multiple nominal variables therefore should be calculated as the correlations between q_{js} and all $x_1,x_2,..,x_p$. PRINCALS does not give such correlations.

I.5.7. Number of solutions.

I.5.7.1. Maximum number of solutions.

If all variables are single, the maximum number of possible dimensions equals m (number of variables). If all variables are treated as multiple nominal (so that PRINCALS gives a HOMALS solution), the maximum number of dimensions equals $\Sigma k_j - m$, $(j=1,..,m)$, where k_j is the number of categories of variable h_j. In the mixed case with some variables single and other variables multiple nominal, we may agree, without loss of generality, that the first m_1 variables are treated as multiple nominal and the last m_2 as single $(m_1+m_2=m)$. Then the maximum number of dimensions equals $\Sigma k_j - m_1 + m_2$ $(j=1,..,m_1)$.

I.5.7.2. How many dimensions should be taken?

As a rough and general rule of thumb one could say that in the case with single variables only, one should retain only the first p dimensions for which the eigenvalue is larger than 1/m. This is the same rule of thumb as the one used in classical PCA. If a dimension has eigenvalue smaller than 1/m, it explains less variance than an individual variable; such a dimension has little or no generalizability.

In the numerical case (PCA solution) one can just retain the first p dimensions for which eigenvalues are larger than 1/m, and drop the m-p dimensions with eigenvalues smaller than 1/m. But with large m, the number of dimensions retained (p) will probably be too large. With single ordinal or single nominal variables, where the solution is not nested, a different approach is necesarry.

One might start with p=3, and (if the smallest eigenvalue of the p=3 solution is smaller than 1/m) go back to p=2. If for any choice of p the sum of the eigenvalues is close to (m-1)/m there is no need to try a solution with p+1 dimensions. If for some choice of p the sum is smaller than (m-1)/m, whereas the smallest eigenvalue still is far above 1/m, an increase of p might be considered.

If there are multiple nominal variables, no easy rule of thumb can be given. It remains true that if an eigenvalue is smaller than the reciprocal of the maximum number of dimensions (see above), the corresponding dimension has little generalizability and could better be dropped - this criterion alone, however, almost certainly will produce more dimensions than is acceptable. After all, remember that multiple nominal treatment of variables gives the greatest amount of freedom for optimal scaling, so that even with a set of uncorrelated random variables one will obtain a number of eigenvalues larger than the reciprocal of the maximum number.

In general we advice to keep p small - better too little dimensions than too much. Also we advice not to increase the number of dimensions beyond the point where their substantive interpretation becomes guess work.

I.5.7.3. Rotations.

In classical PCA there are common routines, such as VARIMAX, for rotation of the solution. If all variables are single, such rotations procedures remain feasible. Still, we do not recommend to apply rotation procedures routinely. It is true that there may be cases where some choice of rotation is helpful. But optimal quantification very often gives results with good interpretability, so that rotation is superfluous. In addition, when there also are multiple nominal variables, rotation problems become rather complicated. It is beyond the scope of this User's Guide to discuss such problems.

I.6. Loss and fit.

I.6.1. Introduction.

"Loss" and "fit" are measures of how bad or good a solution is. Large loss implies that the solution is bad; good fit implies that the solution is good. Loss and fit are complementary measures: good fit implies small loss, and large loss implies bad fit. In chapter 7 these concepts will be further illustrated in graphs - the following sections of this chapter are restricted to algebraic definitions.

I.6.2. Single fit.

PRINCALS gives measures for single fit of single variables. Per dimension (index: s) and per variable (index: j) single fit is defined as a_{js}^2 (the squared component loading). PRINCALS also gives single fit averaged over variables per dimension ($\Sigma a_{js}^2/m$, with j=1,..,m and s fixed). If all variables are single, this average is equal to the eigenvalue Φ_s. PRINCALS further gives "row sum single fit" for each single variable; it equals Σa_{js}^2 (s=1,..,p, with j fixed), and PRINCALS gives the mean of these row sums, equal to $\Sigma\Sigma a_{js}^2/m=\Sigma\Phi_s$ if all variables are single. In that case this mean also is equal to "total fit".
Single fit a_{js}^2 can be interpreted as the "explained variance" of q_j at dimension s. Row sum single fit Σa_{js}^2 corresponds to the "total explained variance" of q_j.

I.6.3. Multiple fit.

Multiple fit, per variable and per dimension, is defined as $\Sigma\hat{y}_{js}^2/n$ (the variance of a column of $\hat{Y}$). This measure of multiple fit never can be smaller than the corresponding measure of single fit a_{js}^2. The reason is that $\Sigma(x_s-\hat{y}_{js})^2/n=1-\Sigma\hat{y}_{js}^2/n$ stands for spread of object scores around their category means. This spread is always smaller than spread around other values than group means, such as $\Sigma(x_s-q_{js}a_{js})^2/n=1-a_{js}^2$. It follows that $\Sigma\hat{y}_{js}/n\geq a_{js}^2$.
Another standard output is the average multiple fit per dimension ($\Sigma\Sigma\hat{y}_{js}/mn$;j=1,..,m; s fixed). If all variables

are multiple nominal, this value will be equal to the eigenvalue Φ_s. PRINCALS further gives row sum multiple fit: multiple fit per variable summed over the p dimensions. The mean of row sum multiple fit also is given. If all variables are multiple nominal this mean will be equal to the sum of the eigenvalues (and therefore also be equal to "total fit").

I.6.4. Overall measures of fit and loss.

For all variables and dimensions together, the goodness of the solution is indicated in 'total fit'. 'Total fit' is equal to the sum $\Sigma\Phi_s$, of the eigenvalues.

'Total fit' also can be defined as the average row sum fit, where we should take row sum single fit for single variables, and row sum multiple fit for multiple variables.

'Total loss' is defined as p- $\Sigma\Phi_s$.

'Total multiple loss' is defined as p-(mean row sum multiple fit).

'Total single loss' is defined as ('total loss')-(multiple loss').

I.7. Plots.

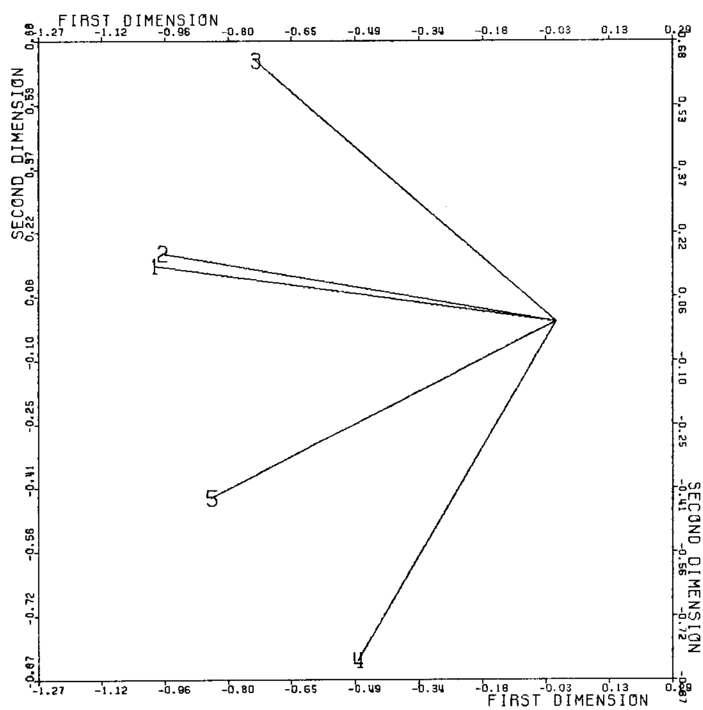
I.7.1. Objects unlabeled.

The plot of objects, unlabeled, graphs object scores x_1 and x_2, with x_1 in the horizontal direction and x_2 in the vertical direction. Comparable plots for other combinations of object scores are not provided by PRINCALS. The user who wants such plots should employ standard plot routines. The plot of object scores can be useful for detecting outliers, or for detecting typical subclouds of objects. The plot also may reveal some special pattern. An illustration is given in section III.2.

I.7.2. Component loadings.

For single variables PRINCALS gives a plot of component loadings for the first two dimensions. If a mixed case includes multiple nominal variables, such variables are

Figure I.7.1: Guttman-Bell data, single ordinal solution, component loadings



plotted in the origin - one should interpret this as that they are omitted from the plot. The plot of component loadings is illustrated in figure I.7.1. for the Guttman-Bell data with all variables treated as single ordinal (compare III.1.). The plot becomes more convincing by drawing arrows in it, as has been done in figure I.7.1.. The squared length of the arrow for q_j corresponds to the row sum single fit of q_j (amount of "explained variance"). To the extent that two arrows are long (with length approximating unity), the cosine of the angle between them reflects the value of the correlation coefficient between the two corresponding quantified variables. E.g., figure I.7.1. shows that the angle between the arrow for q_3 and q_4 is almost 90^o, reflecting the correlation r_{34}=.121. On the other hand, q_1 and q_2 have correlation r_{12}=.977, and the angle between the corresponding arrows is very small. However, when arrows are short, the angle between them will be an inadequate representation of the correlation between the corresponding quantified variables. In the example of figure I.7.1. all arrows are relatively long. This shows that the first two dimensions explain most of the variance of all quantified variables. In addition, the first dimension of the solution (object scores x_1) has negative correlation with all quantified variables (all arrows point to the left). This shows that, on the whole, objects with large positive value for their object score, will have low value in all quantified variables.
However, the last statement will be more true for quantified variables q_1,q_2, or q_5 (with large projections on the horizontal axis) than for variables q_3 or q_4.
The plot further shows that the second dimension of the solution is correlated mainly with quantified variables q_3 and q_4, in opposite direction. It means that objects with high score in x_2 will have high score in q_3 and low score in q_4. The second dimension (x_2) therefore reveals a contrast between quantified variables q_3 and q_4, whereas this second dimension has little relation with variables q_1 and q_2.

I.7.3. Multiple nominal variables.

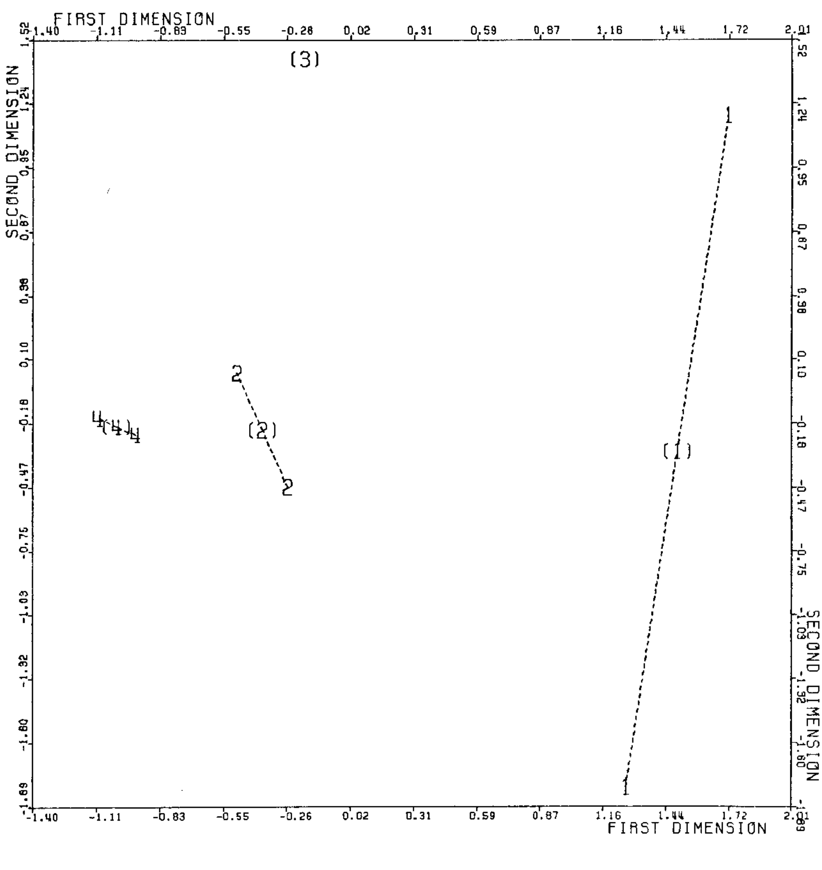
If all variables are multiple nominal - so that the PRINCALS solution is identical to the HOMALS solution - a plot of component loadings is not given.

In the mixed case with some variables multiple nominal and the others single, multiple variables are omitted from the plot of component loadings (they are plotted in the origin). Compare section I.5.7: PRINCALS does not compute the correlations (component loadings) between q_{js} (the s'th quantification of a multiple nominal variable h_j) and all object scores $x_1,..,x_p$.

I.7.4. Objects labeled by variable.

The plot of objects labeled by variable h_j shows the same object points as the plot with objects unlabeled. However,

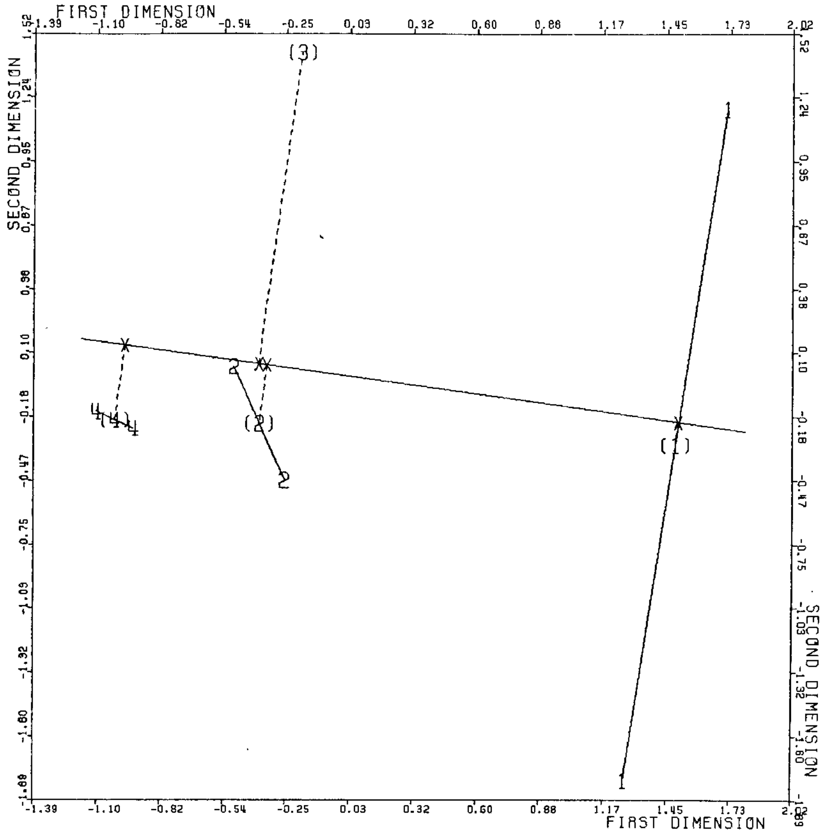
Figure I.7.2: Guttman-Bell data, single ordinal solution, object scores labeled by variable 1



this time the object points are labeled by the corresponding categories of h_j. Figure I.7.2 shows such a plot for the Guttman-Bell data, single ordinal treatment of variables, with objects labeled by the four categories of variable h_1.

The PRINCALS plot only shows the points for the objects. In figure I.7.2., however, we have added the four points corresponding to the "multiple category coordinates".
These "multiple category points" are the means (centers of gravity) of the objects within a category. E.g., the multiple category point for category 1 is half way between the two points for the two objects in category 1.
Also, the one object in category 3 coincides with the multiple category points for category 3.
Multiple fit can be visualized in the figure by the spread of the multiple category points. More precisely, multiple fit equals the average squared distance (weighted by marginal frequency) between multiple category points and the origin. Multiple loss can be visualized as the average squared distance between object points and their corresponding multiple category points (the dotted lines in figure I.7.2). The figure further illustrates that multiple fit for the second dimension is relatively bad, owing to the fact that the two objects in category 1 are far apart in the vertical direction.
Figure I.7.3 repeats figure I.7.2 with the points corresponding to the "single category coordinates" added. These single category points are located on a straight line through the origin. This line has the same slope as the arrow for variable q_1 in the plot of component loadings.
Single fit is visualized by the average squared distance of the single coordinate points to the origin: the farther these category points (on the line) are apart, the better the fit. Single loss appears in the figure as the average squared distance between object points and their corresponding single category point. The difference between multiple loss and single loss is visualized in the figure as the average squared distance between multiple category

Figure I.7.3: Guttman-Bell data, single ordinal solution, object scores labeled by variable 1, single and multiple category coordinates of variable 1

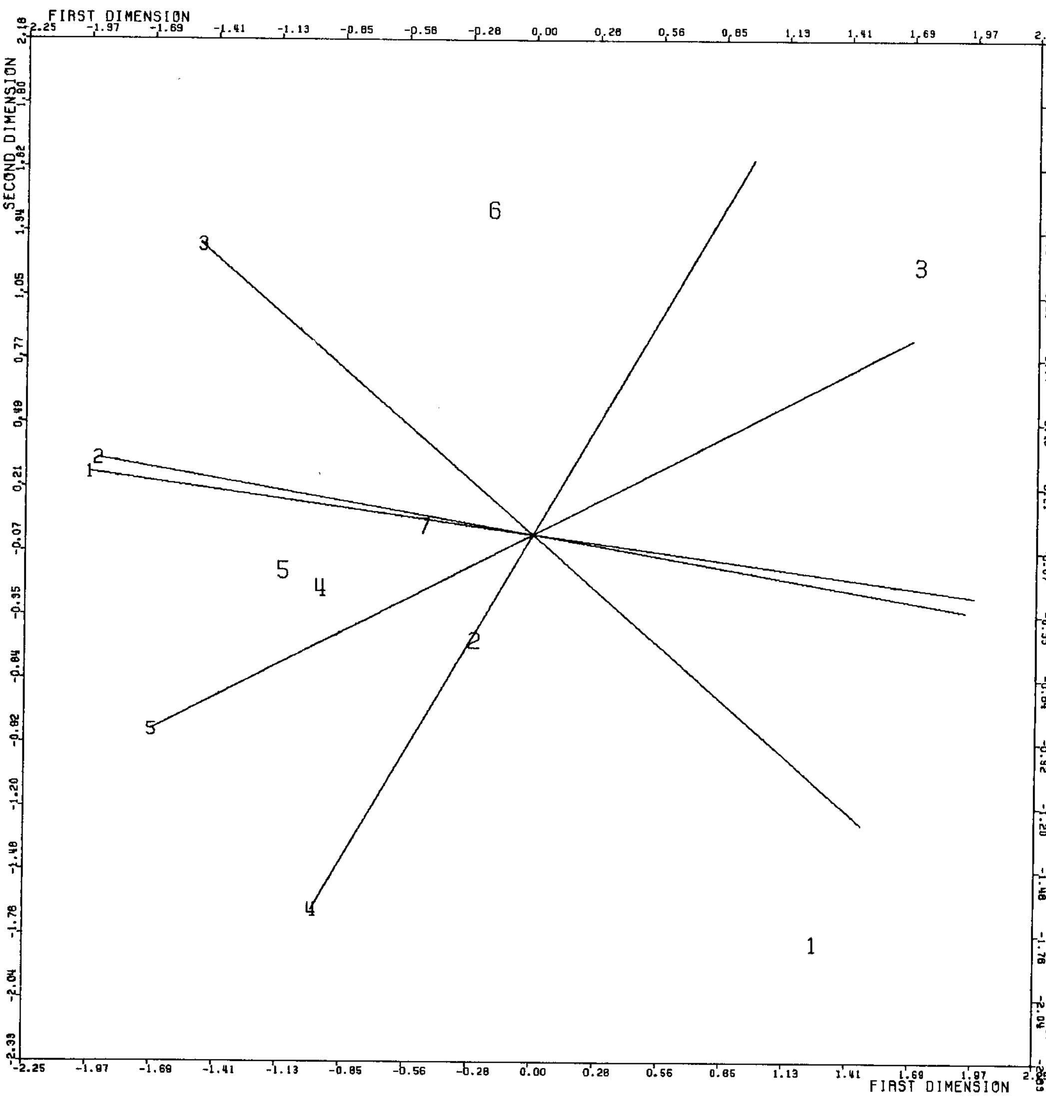


points and corresponding single category points.
In other words, multiple fit would be equal to single fit if the multiple category points themselves were located on a straight line through the origin. This will always be true for a binary variable (with only two categories)- one may verify this for variable h_4 of the Guttman-Bell data in III.1.. In figure I.7.3. single fit for variable 1 is equal to .991 which is very close to the maximum value of unity. Suppose that single fit were equal to its maximum value. We then would have found that the dotted lines in figure I.7.3 are orthogonal to the line on which the single category points are located. In other words,

category points are the projections of the multiple category points. Figure I.7.3 deviates from the ideal in that the dotted lines are not parallel, but approximates the ideal in that the projections on the line of the multiple category points are very close to the single category points.

The user might also make a graph of object points combined

Figure I.7.4: Guttman-Bell data, single ordinal solution, object scores and component loadings (drawn as lines through the origin)



with the "arrows" specified by the component loadings. Such a combined graph is shown in figure I.7.4.. In this figure the arrows of figure I.7.1. are drawn as lines through the origin. Object points are labeled by object number. The graph can be interpreted in two ways.

(i) Select some individual object. The projection of the corresponding object point on the directions representing variables, approximate the values of this object on the quantified variables.

Example. Take object 1. Its values on quantified variables are (see table I.4.1):

(-1.525, -1.550, -2.186, .632, .017).

Figure I.7.4. shows that object 1 in fact has negative projection on the lines for variables 1,2,3; positive projection on the line for variable 4; and almost zero projection on the line for variable 5.

(ii) Select some variable. Projections of object points on the line for this variable approximate the quantification of the categories of this variable. Example. Table I.4.1. shows the following quantification for variable 3: (-2.186 -.325 -.325 .709 .709 .709 .709).

Figure I.7.4. shows in fact that object 1 has very large negative projection on the line for variable 3; objects 2 and 3 have small negative projection; and objects 4,5,6,7 have positive projection. The interpretations suggested in (i) and (ii) above are approximately valid, depending on the values of single fit per variable. If for all variables single fit in the first two dimensions were equal to the maximum value of unity, the interpretations would be perfectly valid.

I.8. Miscellenaous subjects.

I.8.1. Passive and active variables.

PRINCALS has the option to designate some variables as "active" and others as "passive". The PRINCALS solution depends only on active variables; passive variables are ignored. However, one may ask for a plot of object scores labeled by the categories of a passive variable. E.g., if "sex" is a passive variable, such a plot might (or might not) show that males and females are well separated and have different means. Although the PRINCALS solution as such ignores a sex difference, results may show that PRINCALS dimensions are related to sex.

I.8.2. Missing data.

PRINCALS handles missing data in the sense that they have no effect on the final solution. It is important to note, however, that if there are missing data, the interpretation of PRINCALS results given in the previous sections is no longer strictly valid. In particular, component loadings no longer can be strictly interpreted as correlations between x_s and q_{js}. The reason is that normalization requirements are somewhat violated, due to different numbers of objects per variable. When there are not too many missing data, the effect will be limited, and the interpretation given in previous sections will hold true in an approximate way. But with many missing data, especially if they are not distributed randomly over variables, results may become more difficult to interpret.

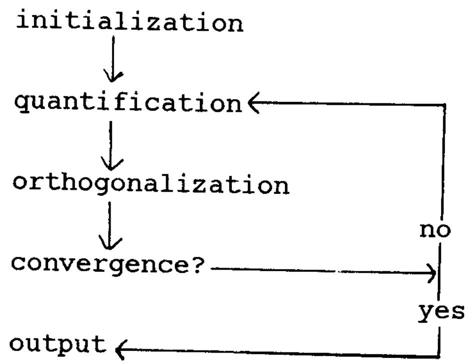
The user has two other possibilities for missing data. The first one is to treat missing data for variable h_j as if they form a separate category. The user then must adapt the input: missing data should now be recorded as a new category. This approach obviously implies that the user believes that the objects with missing data on variable h_j are in some sense "similar". In addition, this approach assumes that for ordinal treatment of a variable the category "missing" must be positioned somewhere in the order

of the other categories.
For numerical treatment, the category "missing" must even be given some numerical value. In general, therefore, the possibility to treat "missing" as a separate category will make sense only for nominal (single or multiple) treatment of a variable.
The second possibility is to define missing data on variable h_j as a separate category for each object with missing data. This implies that the number of categories of variable h_j is increased by as many new categories as there are objects with missing data. Again, this possibility makes sense only if h_j is treated as nominal. Also, if there are many missing data, this possibility implies the risk that the PRINCALS solution becomes dominated by objects with missing data. The general advice, therefore, is to treat missing data as really "missing": they then do not affect measures of stress (fit and loss), but the user takes the risk that the correlational interpretation of results becomes somewhat distorted.

II. The PRINCALS program.

II.1. Some general remarks.

PRINCALS is a portable ANS Fortran IV program, which works satisfactorily on IBM and CDC computers.
The IBM version has dynamical storage by means of an Assembler routine. For other installations this routine is replaced by a static Fortran array allocation routine. The main structure of the PRINCALS program is:



The term quantification is self-evident. The term orthogonalization concerns computation of object scores. These scores have to be orthogonal (i.e., independent) for each dimension. The term initialization means quantification (numerical for single variables and multiple nominal for multiple nominal variables) followed by orthogonalization.
The execution time of PRINCALS is approximately linear with the number of objects and quadratic with the number of dimensions. We have tried to reach a maximum of efficiency by optimizing the program by means of an Assembler timing routine. The optimization is done using the Fortran X compiler, by rewriting comparitivily slow parts of the program.

II.2. The input data.

a. The data are supposed to be on the unit, the number of which is the first parameter of the I/O options card; if the data matrix is on cards, it must follow the parameter cards.

b. In the data matrix the rows have to correspond with objects or individuals and the columns with variables. The data will be read according to the user's format.

c. The data should consist of positive integers starting with the number one (gaps are allowed). The number on the category card is regarded as the highest meaningful category number and thus as the total number of categories of that particular variable.

d. Any number in the data matrix less than one or larger than the highest number of categories of a variable, is considered as missing.

e. When there are more variables in the data matrix than should be analyzed, the analysis variables should come first. The other variables can be used for labeling the plot of object scores.

II.3. The plots.

There are four kinds of plots available in the program. One of the plots is unlabeled, and the other plots are labeled.

a. Plot of object scores (unlabeled).
   The printed symbols correspond with the object scores. The symbols 1,..,9, A,..,Z indicate the number of points that fall in a particular cell. A,..,Z stand for the numbers 10,..,35. When there are more than 35 points in one cell, a symbol '+' will be printed.
   After the plot the exact number of points for every '+' sign is given.

b. Plot of component loadings (labeled).
   The symbols printed correspond with the variable numbers. The symbols 1,..,9 are self evident, the symbols A,..,Z stand for the numbers 10,..,35. When there are more than 35 points the symbol attribution starts again with number 1. A '+' is printed when two points coincide; the labels for those points are specified after the plot.

c. Plot of object scores (labeled).
   The printed symbols correspond with the original category numbers of a variable. The symbolizing is as in b.

d. Plot of rescaled categories for each variable (labeled).
   The printed symbols correspond with the original category numbers of a variable. The symbolizing is as in b. For single variables the single and multiple category coordinates are printed, for multiple variables the category quantifications are printed.

II.4. Implementation.

a. The first two assignment statements of the subroutine PRIINM have the following meaning:
ICAR=5; 5 is the logical unit number for the card reader.
IPRI=6; 6 is the logical unit number for the printer.

b. The dynamic storage allocation in the IBM version of PRINCALS can easily be replaced by a static storage allocation. In this case not the Assembler subroutine DECLAR should be called, but the Fortran subroutine DECLAR. This routine allocates a superarry of a certain specified length. It has to be large enough to contain all the arrays used in the program. The size of the superarray should roughly be equal to:
4(N(T+4P+6)+M(10+3P)+T+K(3+3P)) bytes, with N the number of objects, M the number of active variables, T the total number of variables, K the total number of categories and P the number of dimensions. If the superarray is not large enough, the program will return from DECLAR with an error message and the correct size of the superarray.

c. The labeled common block VARBLS contains parameters and array sizes, all scored as integers. This common block is defined in the PRIINM subroutine. Three types of arrays are used in the program: integer arrays, which use 32 bits per element, real arrays, which also use 32 bits per element, and double precision arrays, which use 64 bits per element.

d. The random number generator in the subroutine RANDMA requires a largest available integer number of 348525375 ($<2^{29}$). If this number is too large another random generator must be impleted. This random generator should produce double precision numbers between -1 and +1.

e. If the plots are not square, or if they are too large or too small for the printer, some statements in subroutine PLOTTO should be adapted according to the notes in the source in the subroutine.

f. In subroutine IMTQL2 the parameter MACHEP is the machine dependent parameter specifying the relative precision of floating point arithmetic. The value for IBM-machines is 2^{-20}.

II.5. Job control.

We give an example of the job control cards for PRINCALS. The cards starting with //* are comment cards.

```
// TIME=(0,30),REGION=256K
/*JOBPARM,LINES=1,CARDS=0,DAG
/*ROUTE PRINT RMT3
//A EXEC PGM=PRINCALS
//STEPLIB DD DISP=SHR,DSN=DIENST
//FT06F001 DD SYSOUT=A
//* EXAMPLE OF AN INPUT DATA SET
//FT08F001 DD DSN=U.DATA,UNIT=3350,VOL=SER=USER01,
// DISP=(OLD,KEEP)
//* EXAMPLE OF AN OUTPUT DATA SET FOR OBJECT SCORES
//FT07F001 DD SYSOUT=B
//* EXAMPLE OF AN OUTPUT DATA SET FOR CATEGORY QUANTIFI-
//* CATIONS
//FT10F001 DD DSN=U.CATEGO,UNIT=3350,VOL=SER=USER02,
// DISP=(NEW,KEEP),SPACE=(TRK,(10,2)),
// DCB=(RECFM=FB,LRECL=80,BLKSIZE=3120)
//FT05F001 DD *
   parameters
/*
```

II.6. Output.

For the initial and final configuration identical output is possible. The user has to specify this by the coding of input parameters.

a. Print output.

1. Input parameters.
2. Raw data (optional), 10 objects are always printed.
3. Measurement levels.
4. Marginal frequencies.
5. History of iterations (optional).
6. Eigenvalues per dimension.
7. Category quantifications, single and multiple category coordinates for single variables; category quantifications for multiple variables.
8. Multiple fit.
9. Single fit (for single variables only).
10. Component loadings (for single variables only).
11. Total fit, total loss, multiple loss and single loss.
12. Correlation matrix (only if there are no missing data).
13. Object scores (optional).
14. Unlabeled plot of object scores (optional).
15. Labeled plot of component loadings (optional).
16. Labeled plot(s) of object scores (optional).
17. Labeled plot of single and multiple category coordinates for single variables, of category coordinates for multiple variables (optional).

b. Output on card, tape or disk.

1. Category quantifications (with variable and category-number) (only for single variables), format (I3,1X,I3,1X,F8.3) (optional).
2. Object scores (with variable and category-number), format (I5,9F8.3/(5X,9F8.3)) (optional).
3. Rescaled categories (with variable- and category-number) single category coordinates for single variables, category quantifications for multiple variables,

format (2I4,9F8.3/(8X,9F8.3)) (optional).

4. Component loadings (with variable number) (only for single variables), format (I5,9F8.3/(5X,9F8.3)).

II.7. Input parameters.

Card 1 : number of problems

Column	Format	Information
1-5	I5	number of problems

Card 2 : title

column	format	Information
1-80	20A4	any alphanumeric code to name the analysis

Card 3 : problem size

Column	Format	Information
1-5	I5	number of objects
6-10	I5	number of variables in the data matrix
11-15	I5	number of analysis-variables
16-20	I5	number of dimensions
21-25	I5	maximum number of categories over all variables
26-30	I5	total number of categories of the analysis-variables

Card 4 : analysis parameters

Column	Format	Information
1-5	I5	maximum number of iterations to compute the final configuration (default = 75)
6-15	F10.8	convergence criterion for the final configuration (default = 0.50E-04)
16-20	I5	maximum number of iterations to compute the initial configuration (default = 20)
21-30	F10.8	convergence criterion for the initial configuration (default: depends on kind and largness of problem)

Card 5 : I/0 options

Column	Format	Information
1-5	I5	unit number of the input medium for the data
6-10	I5	print of data matrix 0 : no 1 : yes
14-15	2I1	print options initial final 0 0 : no print 1 1 : print object-scores only 2 2 : print object-scores and category quantifications 3 3 : print category quantifications only
19-20	2I1	print history of iterations initial final 0 0 : no 1 1 : yes
24-25	2I1	plot options initial final 0 0 : no plot 1 1 : plot object-scores and component loadings 2 2 : like option 1 and in addition plots of objectscores labeled by the category numbers of selected variables specified on card 7 and/or plots of category quantifications of variables specified on card 7
29-30	2I1[1)]	unit number for output of object-scores to other media than line printer initial final 0 0 : no extra output required k j : output to unit number k and/or j

Column	Format	Information
34-35	2I1[1)]	unit number for output of category-coordinates to other media than line printer initial final 0 0 : no extra output required k j : output to unit number k and/or j
39-40	2I1[1)]	unit number for output of category quantifications to other media than line printer initial final 0 0 : no extra output required k j : output to unit number k and/or j
41-45	I5	measurement level of variables 0 : mixed levels which are specified on card 8 1 : only multiple nominal variables 2 : only single nominal variables 3 : only ordinal variables 4 : only numerical variables
46-50	I5	i/o options for initial configuration 0 : no output of initial configuration 1 : identical options as for final solution 2 : options specified in first column of the relevant parameters
51-55	I5	number of categories per variable 0 : variables have different numbers of categories specified on card 6 k : all variables have k categories

1) If column 46-50 (card 5) <u>not</u> equals 2, format I2 may be used to specify the options for the final solution or both the initial and the final solution.

Column	Format	Information
56-60	2I1 [1)]	unit number for output of component loadings to other media than line printer initial final 0 0 : no extra output k j : output to unit number k and/or j

Card 6 : number of categories

Only if the number of categories per variable=0 (column 51-55 of card 5)

Column	Format	Information
1-80	16I5	maximum number of categories of each variable in the datamatrix (16 variables per card; if necessary continue on following card(s))

Card 7 : plot specifications

Only if plot options = 2 (column 24/25 of card 5)

Column	Format	Information
1-80	80I1	The columns specify variables in the same order as in the data matrix (80 variables per card; if necessary continue on following card(s)) 0 : no extra plot options for this variable 1 : plot of object scores, labeled with category numbers of this variable 2 : like 1 and in addition the plot of category quantifications of this variable 3 : plot of the category quantifications of this variable

1) If column 46-50 (card 5) not equals 2, format I2 may be used to specify the options for the final solution or both the initial and the final solution.

Card 8 : measurement level

Only if measurement level = 0 (column 41-45, card 5)

Column	Format	Information
1-80	16I5	measurement level of each variable (16 variables per card; if necessary continue on following card(s) 0 : multiple nominal 1 : single nominal 2 : ordinal 3 : numerical

Card 9, 10 and 11 : data format (always three cards)

Column	Format	Information
1-80	20A4	Fortran integer format

Card 12,....: data

Only if column 1-5 of card 5 equals 5.

Depending on the value of the number of problems (card 1, column 1-5) more problems can be analysed in one job step. For every extra problem all cards, except the first one, have to be repeated.

III. Numerical examples.

III.1. Guttman-Bell data.

III.1.1. Initial data.

These data are discussed in Guttman (1968) and Lingoes (1968). The data were derived by Guttman from a sociological text. They characterize seven different groups in terms of five variables. The variables, with their categories are:

1. Intensity of interaction;
   a. slight,
   b. low,
   c. moderate,
   d. high.
2. Frequency of interaction;
   a. slight,
   b. non-recurring,
   c. infrequent,
   d. frequent.
3. Feeling of belonging;
   a. none,
   b. slight,
   c. variable,
   d. high.
4. Physical proximity;
   a. distant,
   b. close.
5. Formality of relationship;
   a. no relationship,
   b. formal,
   c. informal.

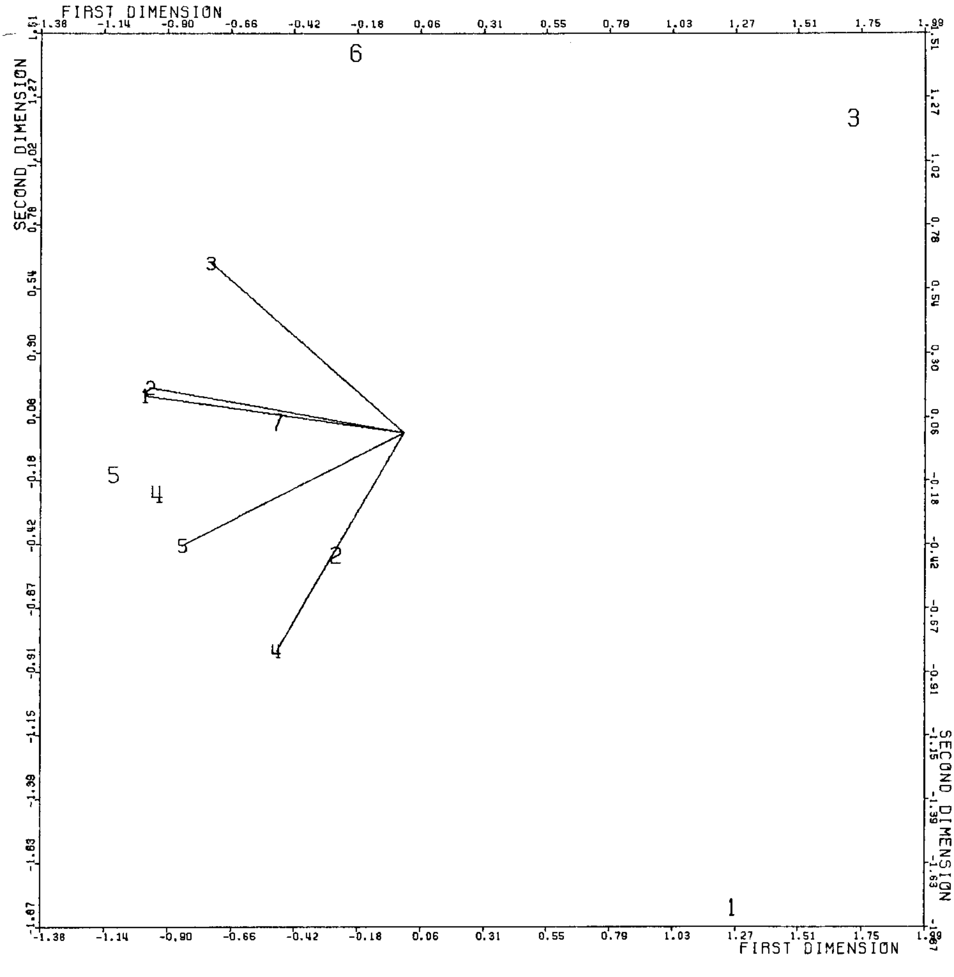
The seven objects were classified as follows,

- Crowd	a	a	a	b	b
- Audience	b	b	b	b	b
- Public	a	a	b	a	a
- Mob	d	b	d	b	c
- Primary group	d	d	d	b	c
- Secundary group	c	c	c	a	b
- Modern community	b	c	c	b	b

III.1.2. All variables single ordinal.

PRINCALS results with all variables single ordinal are given in the Appendix. A joint plot of variables (arrows based on component loadings) and objects is given in figure III.1.1. The figure does not need much comment. Obviously, the first dimension is correlated with all variables (but most with variables 1,2,5, less with variables 3,4), whereas the second dimension contrasts variables 3 and 4, and also contrasts objects Secondary Group (distant physical proximity, high feeling of belonging) and Crowd (close physical proximity, no feeling of belonging).

Figure III.1.1: Guttman-Bell data, single ordinal solution, object scores and component loadings



III.1.3. <u>All variables single nominal.</u>

Results of PRINCALS with all variables single nominal are not given; they differ little from the single ordinal solution. One typical result is that categories b and c of variable 2 have reversed order (note that the apriori order of the categories of variable 2 is not very clear). The single solution in fact merges categories b and c (they have the same quantification).

An interesting result is that the PRINCALS single nominal solution is not the optimal one. The fact that PCA is used as the initial configuration in this example leads to a single nominal solution which corresponds to a "local minimum". The best single nominal solution is given in Gifi (1981a), p.188. This best solution requires a much more drastic re-ordering of categories, far away from the PCA solution.

III.2. Suicide questionnaire.

III.2.1. Introduction.

In 1979/80 the Department of Clinical Psychology of the University of Leiden collected responses to a questionnaire about attitudes towards suicide (Speijer & Diekstra 1980; Diekstra & Kerkhof 1982). The 12 items of the questionnaire are described in table III.2.1.. In addition, data were obtained with respect to background variables of the 694 respondents - these variables are listed in table III.2.2..

For the following example only those respondents will be analyzed who have less than 7 missing values on the 12 suicide attitude items. This reduced the number of respondents to 580. The example will cover:

(a) analysis of relations between the 12 items (section III.2.2.). This includes a two-dimensional and a three-dimensional ordinal PRINCALS solution - both solutions will be compared with the numerical solution.

(b) analysis of relations between the 12 items and the background variables, with background variables active, or passive (section III.2.3.).

III.2.2. Analysis of responses on suicide items.

III.2.2.1. Number of dimensions.

We first performed a two-dimensional PRINCALS solution with all variables (m=12) ordinal. This resulted into eigenvalues of .362 and .165. Since the last eigenvalue is still larger than 1/m=.083, a three-dimensional ordinal solution was also considered. Its last eigenvalue appeared to be .098.

III.2.2.2. Two-dimensional ordinal solution.

Figure III.2.1.gives a plot of the 580 object scores, labeled by object number (this is not regular PRINCALS output). Clearly this plot shows a peculiar shape: a sort of blurred V with most objects along the left edge of the V.

Table III.2.1. Circumstances.

Under what circumstances would you commit suicide?
(1 = certainly not, 2 = probably not, 3 = perhaps, 4 = probably, 5 = certainly.) If you:

1. are old and decrepit,
2. were to suffer a lot,
3. were left by your partner,
4. became seriously handicapped,
5. became unemployed,
6. got a disabled child,
7. had to be taken into a mental hospital,
8. could not have children,
9. were to suffer from an incurable disease,
10. were to lose a loved one,
11. could not find a life-partner,
12. were responsible for someone's death.

Table III.2.2.: The background variables, their categories and frequencies.

1. SEX:	1=male	291	(MALE)
	2=female	285	(FEMALE)
	missing	4	
2. AGE:	1=15 - 24 Years old	86	(15-24Y)
	2=25 - 29 " "	132	(25-29Y)
	3=30 - 35 " "	132	(30-35Y)
	4=36 - 47 " "	111	(36-47Y)
	5=48 - 75 " "	113	(48-75Y)
	missing	6	
3. MARITAL STATUS:	1=married	380	(MARRIED)
	2=widowed	10	(WIDOWED)
	3=single	113	(SINGLE)
	4=divorced	28	(DIVORCED)
	5=living together	40	(LIV.TOGE)
	missing	9	

4. RELIGION:

1=no	294	(NO RELIG)
2=catholic	131	(CATHOLIC)
3=calvinistic	29	(CALVINIS)
4=protestant	99	(PROTEST.)
missing	27	

5. POLITICAL PREFERENCE:

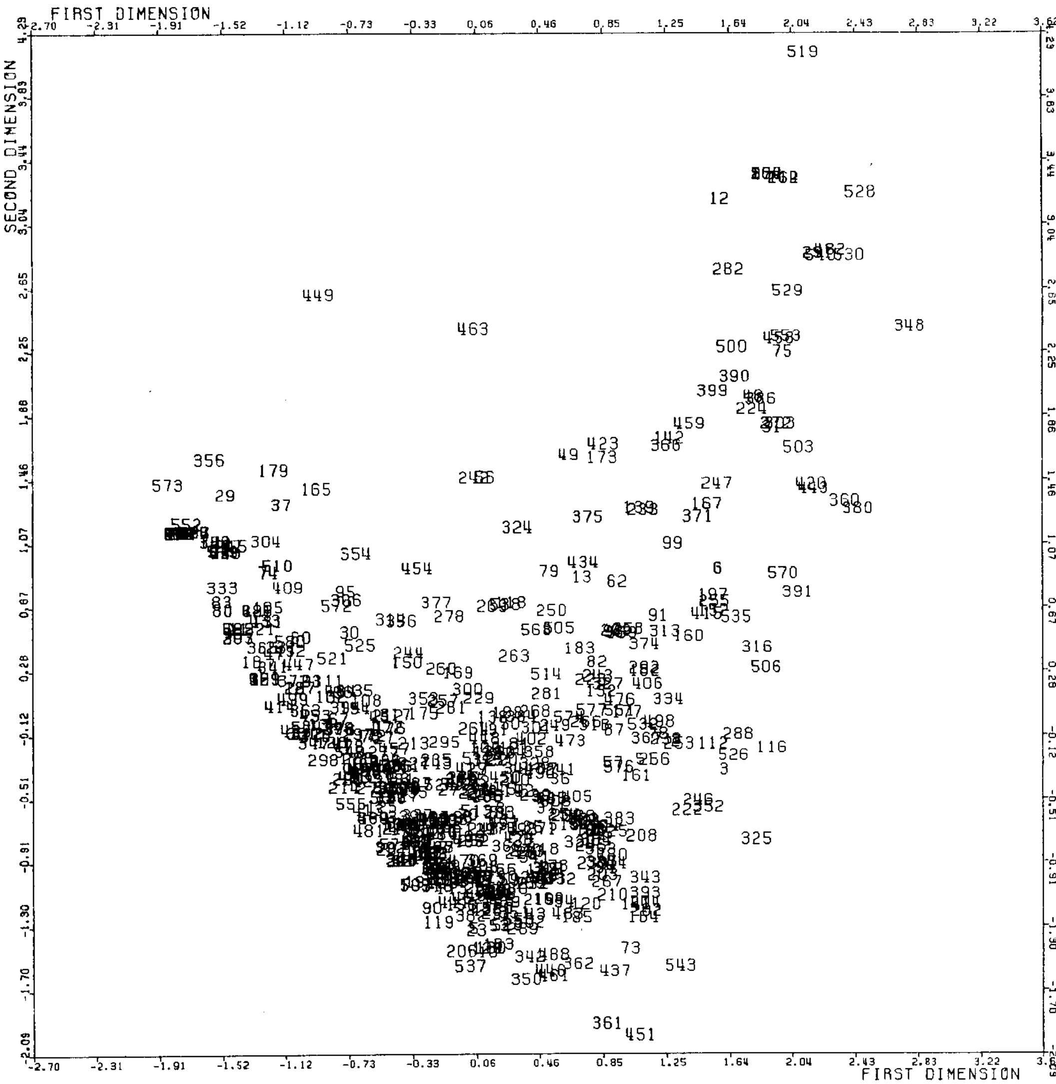
1=no	153	(NO POL.)
2=PVDA-socialists	148	(PVDA)
3=CDA-christian democrats	74	(CDA)
4=VVD-conservative liberals	55	(VVD)
5=PPR-radical socialists	15	(PPR)
6=CPN-communists	5	(CPN)
7=PSP-pacifist socialists	16	(PSP)
8=DS'70-conservative social democrats	4	(DS'70)
9=D'66-liberals	76	(D'66)
missing	34	

6. Broadcasting company:

1=no	180	(NO-BROAD)
2=KRO-catholic	37	(KRO)
3=VARA-socialist	48	(VARA)
4=AVRO-conservative	70	(AVRO)
5=NCRV-protestant	34	(NCRV)
6=VPRO-progressive	49	(VPRO)
7=EO-calvinistic	5	(EO)
8=TROS-light entertainment	78	(TROS)
9=VOO-light entertainment for young people	79	(VOO)
missing	0	

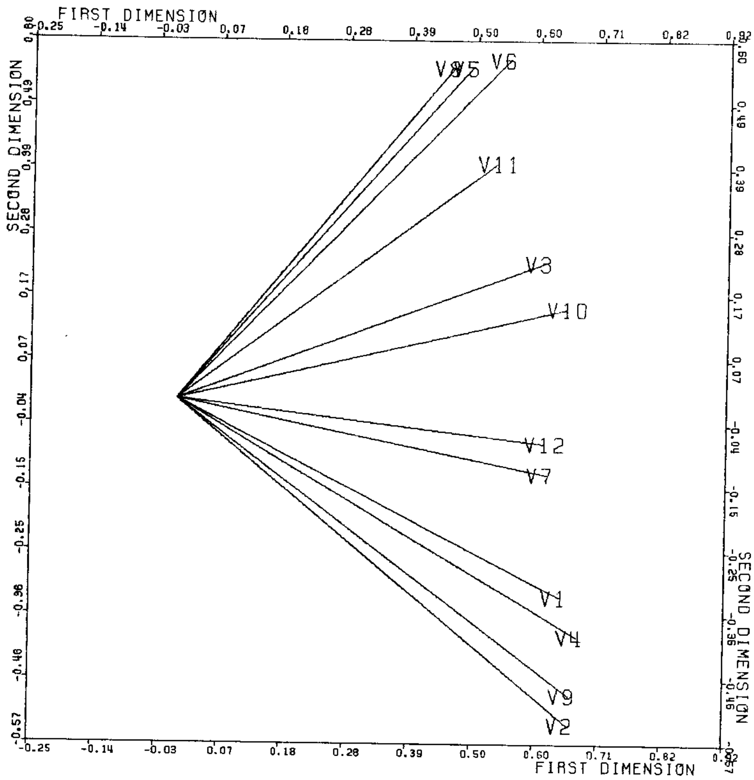
* In the Netherlands a broadcasting company is a non-profit organisation with an idealistic aim of which people can be a member.

<u>Figure III.2.1:</u> two-dimensional ordinal solution, object scores labeled by object number



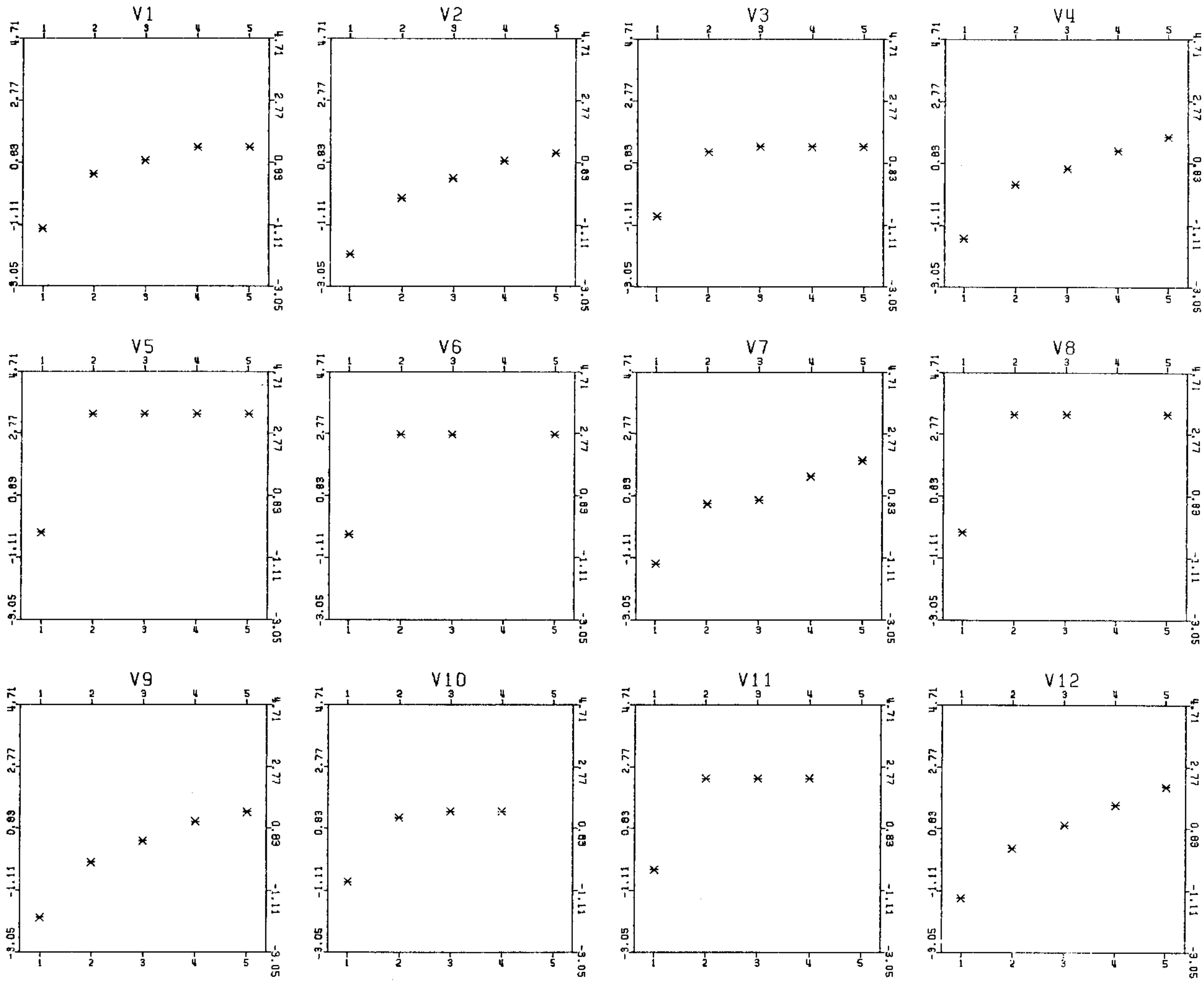
The second graph plots component loadings (figure III.2.2.). Squared lenghts of the arrows in this plot correspond to row sum single fit, and therefore to "explained variance" per item. Projections on the horizontal axis correspond to correlations between object scores on the first dimension and each of the quantified items; projections on the vertical axis correspond to correlations between quantified

<u>Figure III.2.2:</u> two-dimensional ordinal solution, component loadings



items and object scores on the second dimension.
The graph suggests a difference between items 1,2,4,9 on the one hand, and items 5,6,8,11 on the other. The interpretation is easy: the first set of items refers to physical circumstances under which suicide is considered; the second set gives social indications.
Figure III.2.3. plots original responses (from 1 to 5, as indicated in table III.2.1.) against their ordinal quantifications. Because all items are treated as ordinal, all plots show the quantification as an increasing function. There is the interesting result that for items 5,6,8,11 (the "social items") the quantification is dichotomous

<u>Figure III.2.3:</u> two-dimensional ordinal solution, categories, original responses against their quantifications



(all responses different from 1 are quantified in the same way), whereas for items 1,2,4,9 the quantification is more gradual. It is interesting to see that this result is related to the shape of the marginal distributions (table III.2.3.). This table shows that the "social" items are characterized by a reversed-J shape distribution, whereas the "physical" items have an almost symmetric distribution. Respondents clearly react to the physical items in a different way than to the social items.

<u>Table III.2.3:</u> Marginal frequencies

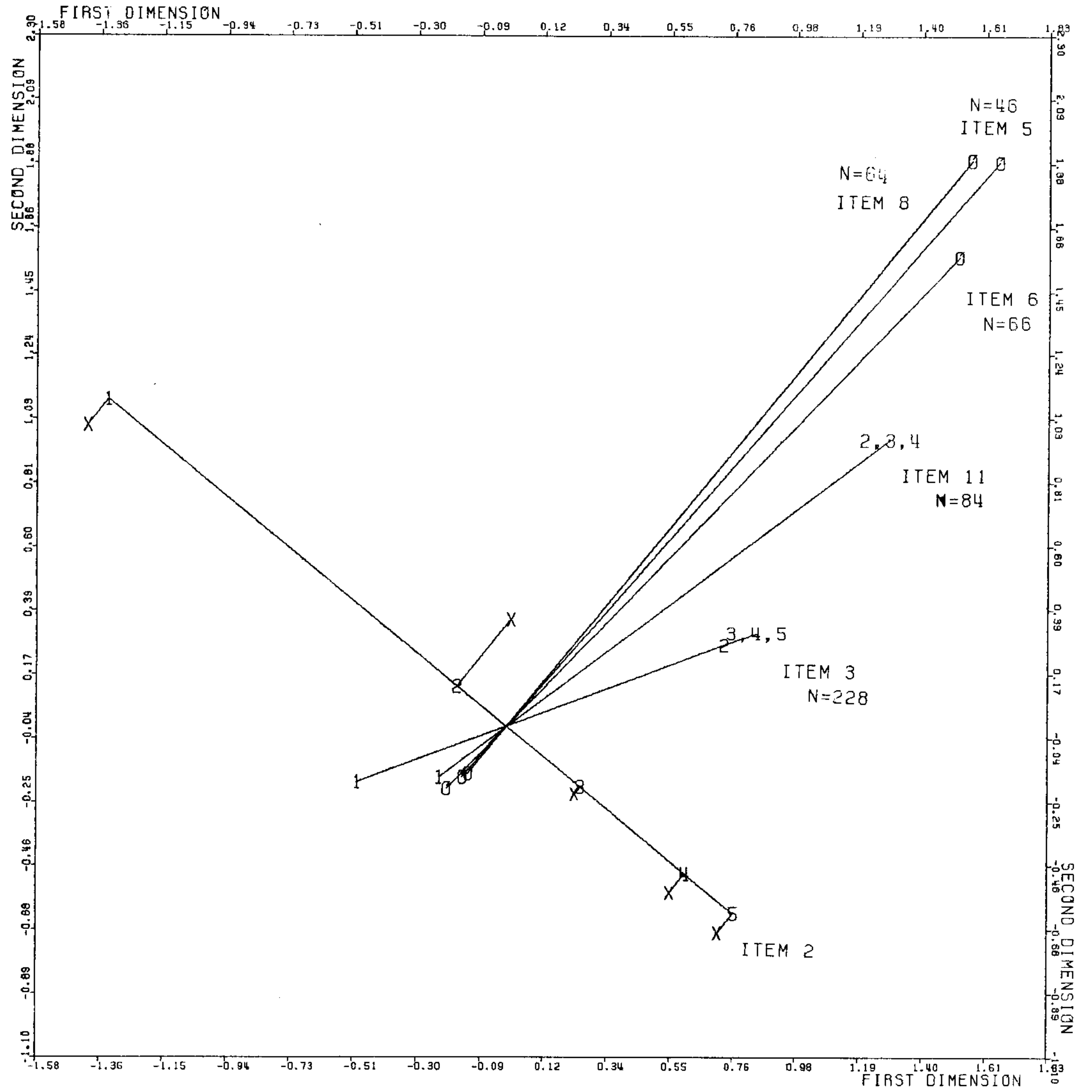
VARIABLES	MISSING	CATEGORIES 1	2	3	4	5
1.	1	230	156	147	37	9
2.	1	98	113	222	105	41
3.	7	345	163	54	9	2
4.	8	164	157	192	45	14
5.	14	520	42	2	1	1
6.	6	507	60	6	0	1
7.	7	213	202	123	23	12
8.	6	528	36	9	0	1
9.	9	102	135	224	83	27
10.	5	346	169	50	10	0
11.	17	479	61	19	4	0
12.	18	186	182	164	23	7

Figure III.2.4. further illustrates this point. This figure shows single and multiple coordinates for physical item 2. The single coordinate category points are on a straight line (this line has the same direction as the arrow for component loadings of variable 2). The figure shows that single and multiple coordinates are close together, for each of the five categories.

For each of the items 5,6,8 and 11, the figure shows the single category coordinates for the categories 2,3,4,5 of these variables - these categories have identical quantification. The figure also shows single category coordinates for category 2 of item 3, and for categories 3,4,5 of this same item. The figure makes it clear that the non-1 responses to items 5 or 8 (unemployment, or not being able to have children, respectively) are the most extreme in the "social direction", followed by respondents with non-1 response to item 6 (disabled child), item 11 (not being able to find a partner), and item 3 (left by partner). In addition, in this order the frequency of these responses increases.

It is interesting to see that in figure III.2.4. the non-1 category points for items 3,11,6,5 and 8 are on a curved line. This agrees with the general pattern of objects in figure III.2.1.. It means that objects high on the social dimension are less extreme on the physical dimension than objects moderate on the social dimension. It is good strategy to check such a suggestion in the actual data. In

Figure III.2.4: two-dimensional single ordinal solution, single and multiple coordinates for item 2, single coordinates for items 3, 5, 6, 8 and 11

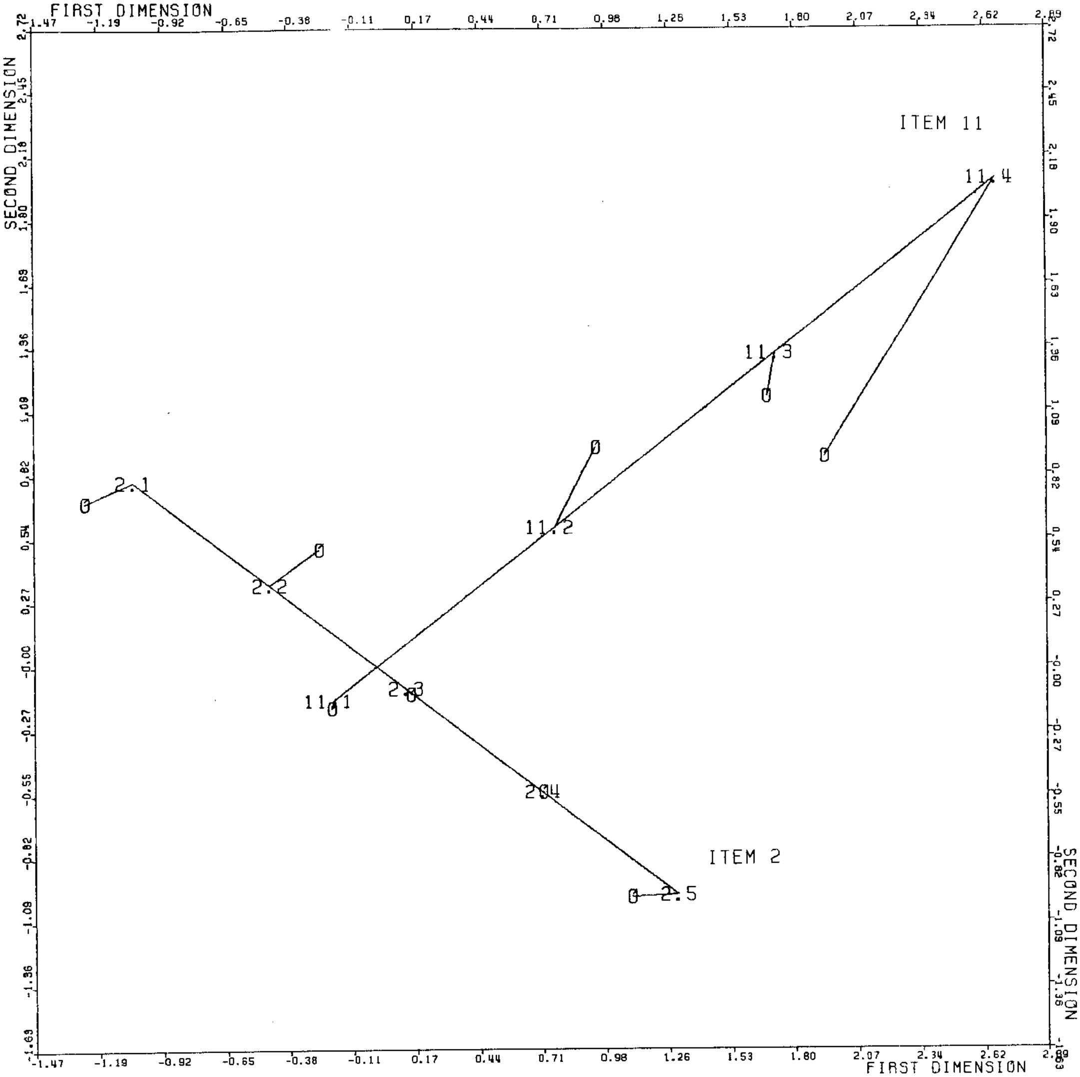


fact, the 22 objects in category 2,3,4,5 both of item 5 and item 8 have average quantified score of .198 on item 2. The respondents in category 2,3,4,5 of item 5 have average of .247; those in 2,3,4,5 of item 8 have average of .294; those in 2,3,4,5 of item 6 have average .312; those in 2,3,4,5 of item 11 have average .302; those in 3.4.5 of item 3 have average .360 on variable 2. These

results confirm that respondents with extreme position in the social direction have lower score in the physical direction than respondents with intermediate score in the social direction.
Possibly this rather unexpected result is related to age differences, in the sense that respondents with extreme score in the social direction are younger people (unemployment, or not being able to have children, are realistic threats, whereas physical illness still is a somewhat remote and abstract threat), whereas respondents with intermediate position in the social direction are of medium age (both types of threats are realistic to them). This hypothesis implies that elderly people in the whole would be low in the social direction, but have large variation in the physical direction. Such a hypothesis finds some confirmation from the analysis of background variables (section 10.3); it would be worth while, however, to make a direct check in the actual data. In fact, the methodological point here is that a PRINCALS solution might generate specific hypotheses: the PRINCALS solution may invite the researcher to have a closer look at the data with some specific idea what to look for.

III.2.2.3 Comparison with numerical solution.
The two first dimensions of a numerical solution (PCA) have eigenvalues .330 and .162, with sum (total fit) equal to .492, showing that these two dimensions explain 49.2% of the total variance. For ordinal PRINCALS the corresponding value is 52.7%, so that there is only a slight gain obtained from mildering restrictions from numerical to ordinal. In addition, component loadings for the numerical solution are almost the same as for the ordinal solution. Also, the plot of object scores looks very much the same as in figure III.2.1. and therefore is not reproduced here.
This result is not so surprising, because the ordinal quantification differs from the numerical solution primarily in the treatment of the social items (for which the re-

<u>Figure III.2.5:</u> two-dimensional numerical solution, single and multiple coordinates for items 2 and 11



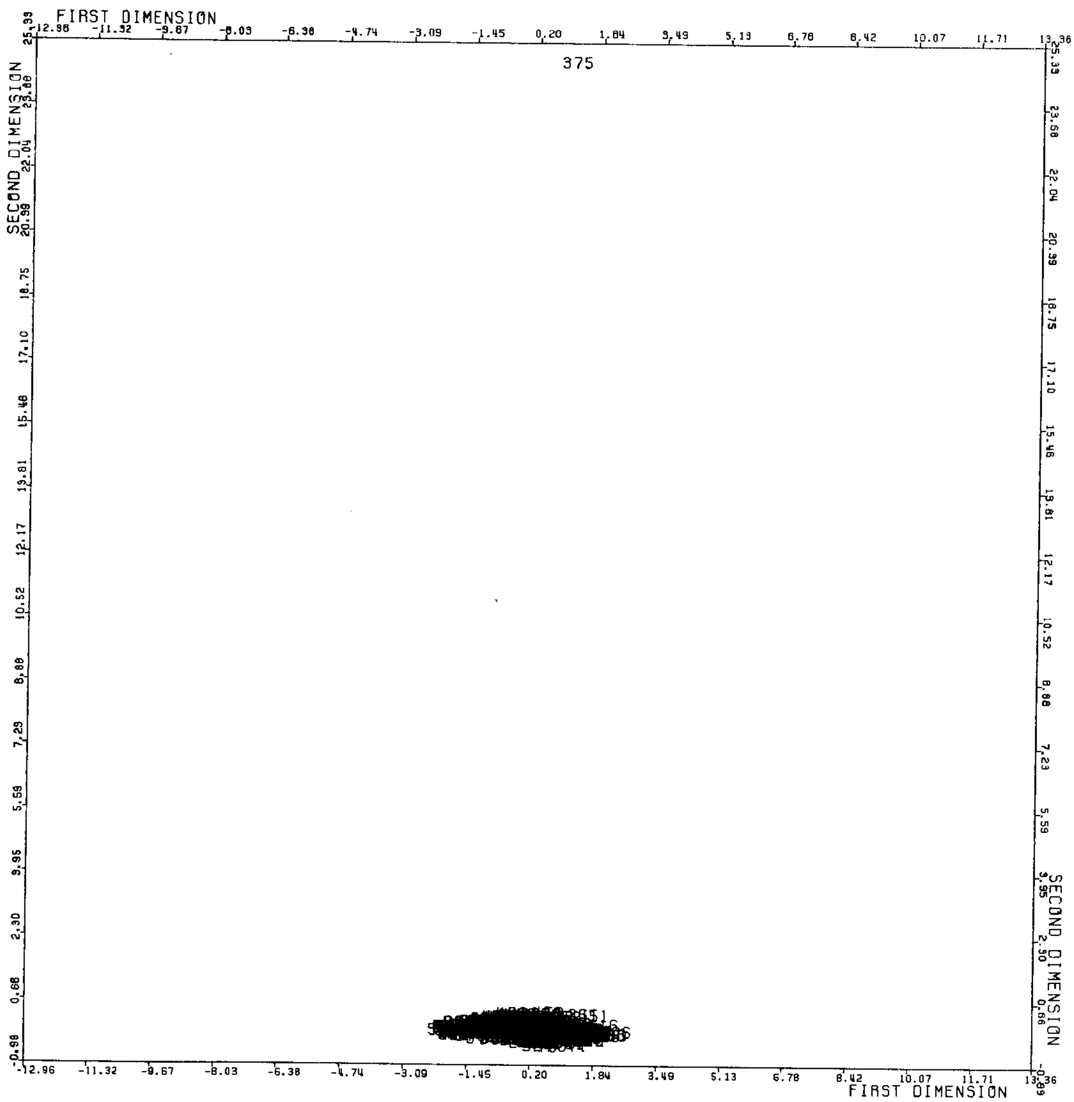
versed J-shape distibutions are dichotomized). This affects mainly the categories with low frequencies. The effect on correlations therefore is rather limited.

For comparison with figure III.2.4. figure III.2.5. has been drawn, showing single and multiple coordinates for the categories of variables 2 and 11 in the numerical solution. Clearly, for item 2 multiple category points remain close to their single category points (indicating that even a single multiple treatment of item 2 would not change much). For item 11, however, the multiple category points for categories 2,3,4 are much closer together than their single category points (the single ordinal solution will merge them into one category).

III.2.2.4. Three-dimensional solution.

As was explained in section III.2.2.1., also a three-dimen-

Figure III.2.6: three-dimensional ordinal solution, object scores labeled by object number



sional solution was explored. With all variables treated as single ordinal, this three-dimensional solution appears to degenerate. Figure III.2.6. (the plot of object scores in first and second dimension) shows what happens: the second dimension is entirely dominated by the single respondent numbered 375. Inspection of the data reveals that this respondent is the only one with response categories 5 to items 6 and 8 (see marginal frequencies in table III.2.3.). In addition, there are no respondents at all who use category 4 for these two items. In such a situation there are two possible options. The first one is to delete the peculair respondent from the data set, and repeat the analysis for the other respondents. Of course, this does not mean that in a final report of the analysis object 375 should be ignored as if this respondent never existed. To the contrary: it should be reported that PRINCALS spotted this unique respondent as a special "outlier", and that once this special respondent was recognized the analysis is restriced to the remaining respondents.

The other option is to re-code the extreme responses of the unique respondent. For example it means that category 5 on items 6 and 8 is merged with category 3 (category 4 does not occur). In fact, the two-dimensional solution already suggested this merging.

After re-coding object 375, the three dimensional solution has eigenvalues .360, .164 and .098, showing that there is about 10% gain in explained variance compared to the two-dimensional solution. In addition, the first two dimensions are roughly the same as those of the two-dimensional solution, for component loadings as well as for object scores. The third dimension distinguishes between items 6 (disabled child) and 8 (could not have children), both with negative loadings of about -.40 on one end, and on the other the items 3 (left by partner), 10 (lose loved one), 11 (no life-partner) and 12 (responsable for someone's death) with positive loadings of about .40.

III.2.3. Relations with background variables.

III.2.3.1. Introduction.

In the present example the 12 items referring to attitude towards suicide in given circumstances are the fore-ground variables - in addition there are six background variables listed in table III.2.2..

Relations between fore- and background variables can be investigated in several ways.

(i) One might investigate for each of the fore-ground variables whether there is a (significant) relation with each of the background variables. The disadvantage of this approach is that it ignores dependence between variables. E.g., if two fore-ground variables are highly correlated (such as items 2 and 9 in the present example), they will be found to be related to the same background variables. Conversely, if two background variables are highly interrelated, they will have similar relations with fore-ground variables.

(ii) Another approach is to treat background variables as active. PRINCALS results then will depend on all variables jointly. For the present example such an analysis will be illustrated in section III.2.3.2..

(iii) A third approach is to treat background variables as passive - this will be illustrated in section III.2.3.3..

(iv) A fourth possibility would be that the objectscores, obtained from PRINCALS on the fore-ground variables, are related to background variables by using canonical analysis, with background variables treated as single nominal. Such an analysis respects the PRINCALS results for the fore-ground variables (their quantification is not changed), whereas background variables obtain optimal quantification with respect to their relation to fore-ground variables.

(v) A fifth possibility would be that fore-ground and background variables are treated as two sets of variables, and are subjected to CANALS.

This approach falls outside the scope of this User's Guide (see van der Burg, CANALS User's Guide, 1983). In fact this approach assumes that the two sets of variables are, so to speak, on equal footing; whereas approaches (iii) and (iv) above analyse fore-ground variables on their own right, and make the connection with background variables afterwards, for fixed results of the fore-ground variables.

III.2.3.2. <u>Background variables active.</u>

Treating background variables as active implies that they are given the same importance as fore-ground variables. The risk is that the solution might become entirely dominated by relations between background variables (with low component loadings for fore-ground variables). In addition,

<u>Figure III.2.7:</u> background variables active, component loadings of 12 items

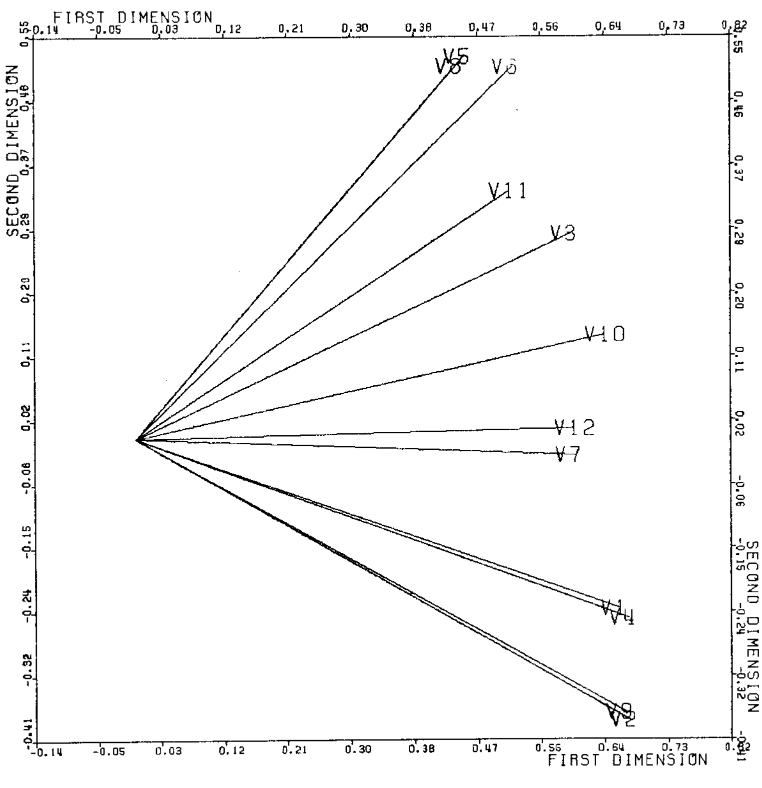
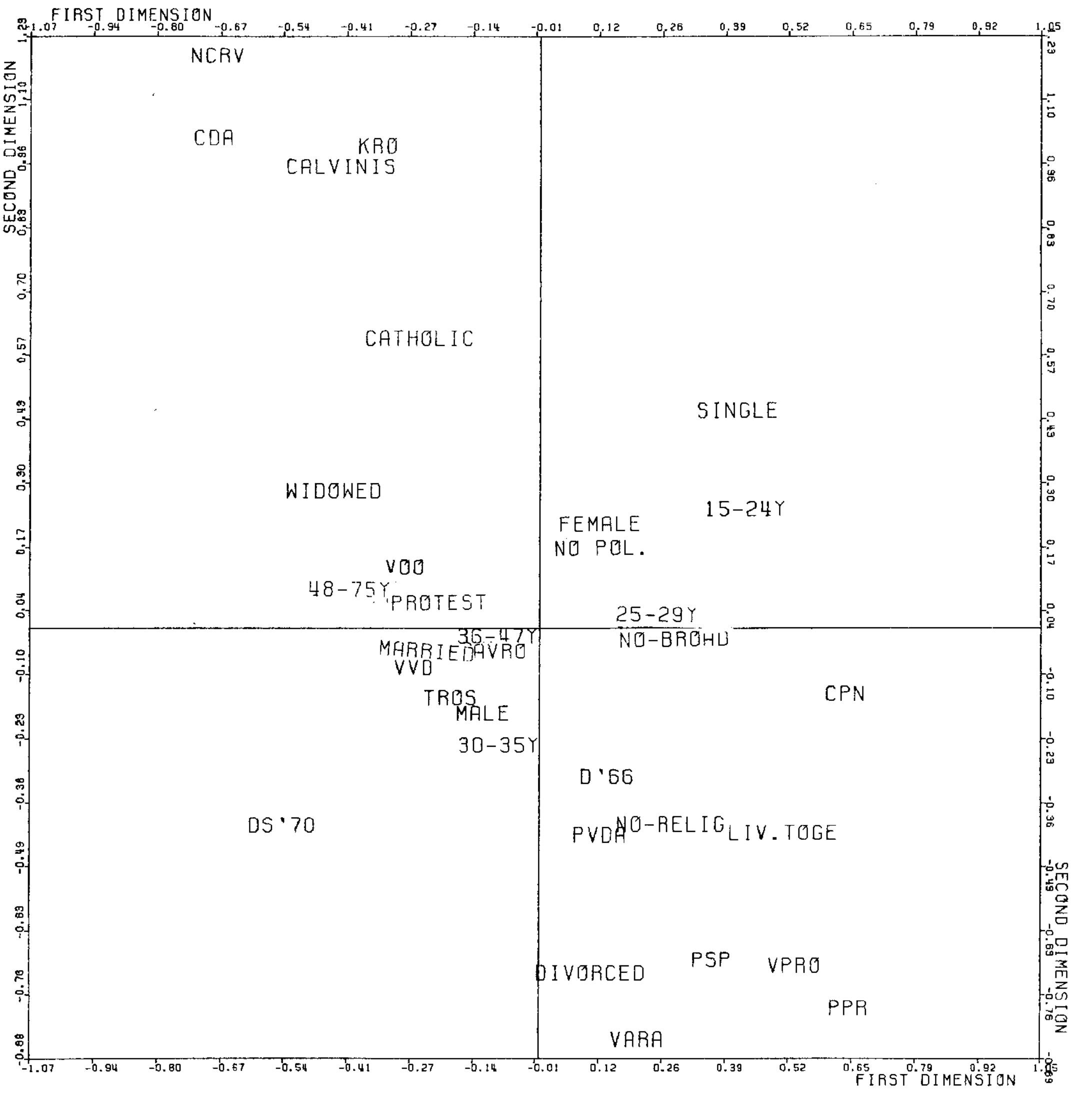


Figure III.2.8: background variables active, category coordinates of the background variables (EO has been omitted)



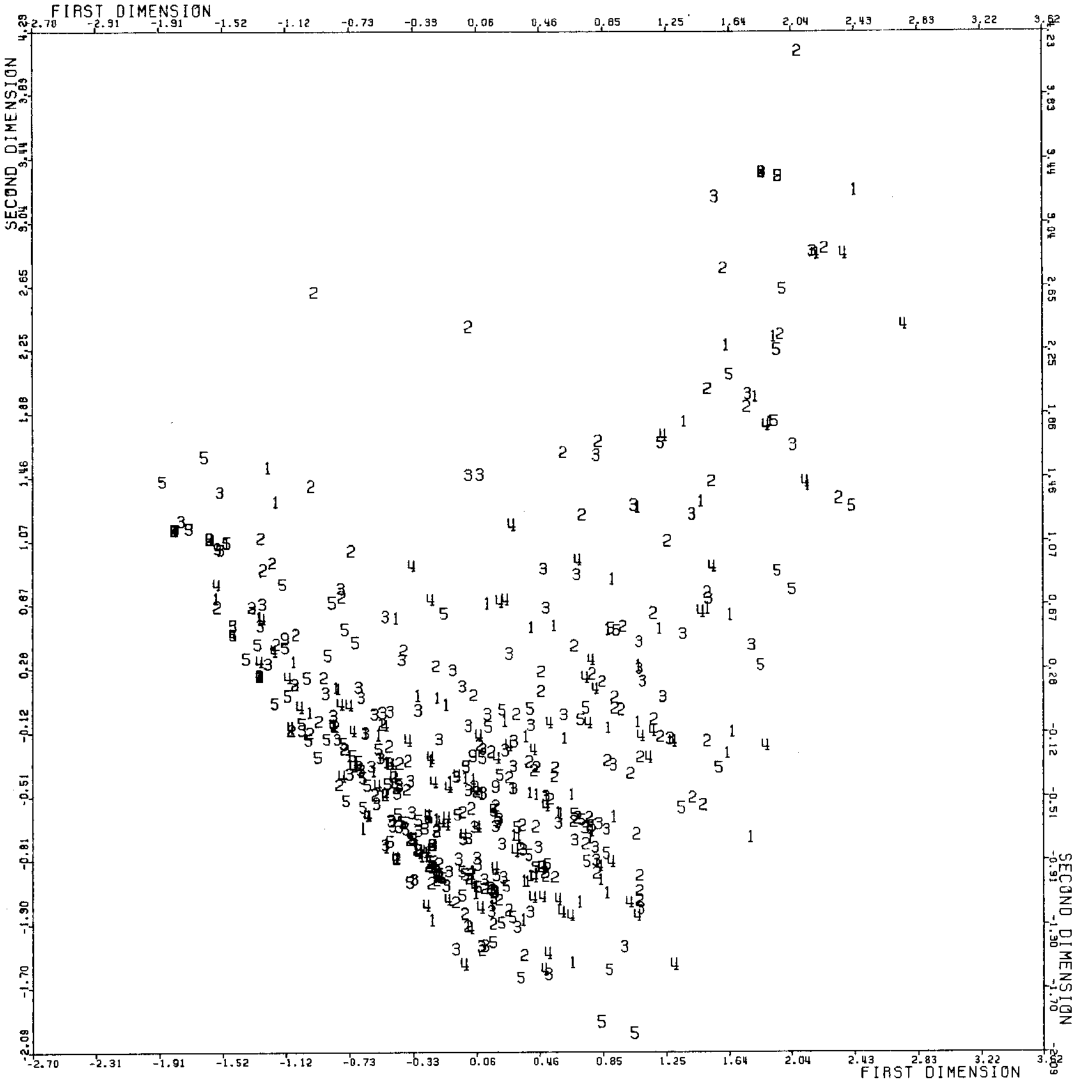
such relations between background variables might be quite trivial (such as showing that widowed persons are older, or that religious people more often prefer some denominational party). In the present example treatment of background variables as active (two dimensions, suicide items treated as ordinal, background variables as multiple nominal) does not change much with respect for results of the 12 items: we find the same distinction between "social" and "physical" items. This is shown in figure III.2.7. (component loadings for the 12 items). Figure III.2.8. shows multiple category coordinates for the categories of the background variables. In this figure the points for category EO of background variable 6 has been omitted; this point has coordinates (-1.685, 1.158), indicating that respondents in this category have extreme position both in the physical and the social direction. The figure further suggests that respondents low on the physical direction are older, more often widowed, more often calvinistic or Roman-Catholic, more often member of denominational broadcasting companies. In the opposite direction we find respondents who are divorced, have left wing political preference, and subscribe to "leftish" broadcasting companies such as VARA or VPRO. In the social direction we find respondents who are younger, more often female than male, and with either no political preference, or preference for CPN as contrasted with DS'70 (but note that both last mentioned categories have extremely low marginal frequencies).

Little is known about how political parties stand with respect to suicide. Present results suggest that their attitude towards physical motives for suicide is about the same as their attitude towards abortion.

In the present example the active treatment of background variables does not change much in results for fore-ground variables, possibly because there are 12 items in the foreground, and only 6 variables for background. But if there are many more background variables than fore-ground variables, the risk indicated in the beginning of this section becomes much larger.

Also, results in figure III.2.8. show the usual pattern for multiple nominal variables: categories with small marginal frequency tend to be farther away from the origin, whereas categories with high marginal frequency tend to be close to the origin.

<u>Figure III.2.9:</u> background variables passive, objects labeled by age categories: 1 = young, 5 = old



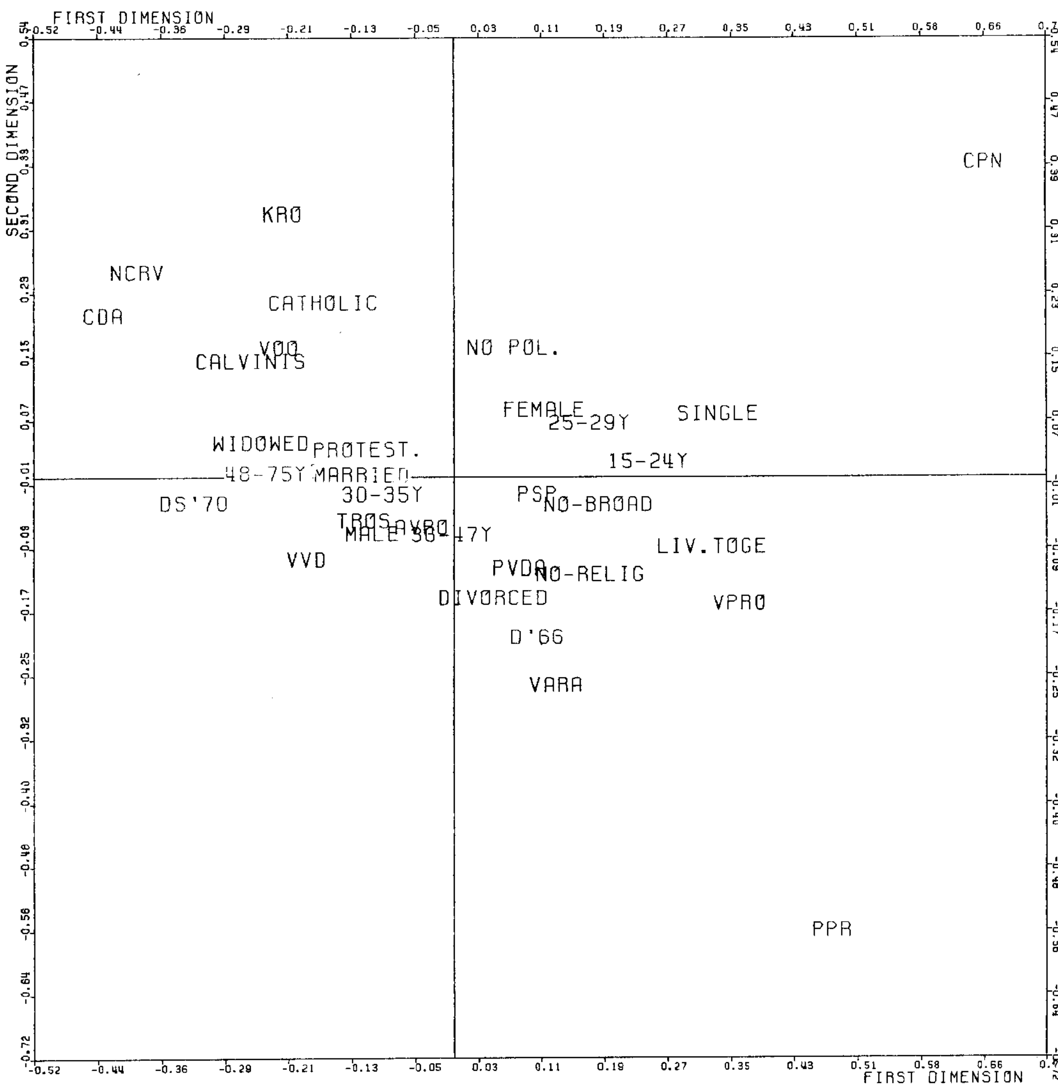
III.2.3.3. <u>Background variables passive.</u>

With background variables passive PRINCALS results depend only on the fore-ground variables (the 12 suicide items). What PRINCALS may add for a passive variable is just a plot of objects labeled by category numbers of a passive variable.

Such a plot is shown in figure III.2.9., with objects labeled by age categories.

Figure III.2.10. confirms most of the results of figure III.2.8.. In particular, figure III.2.10 shows that respondents high in the social direction are in the lower age groups, more often single, more often with preference for CPN. The figure also confirms that the physical dimension is related to religious affiliation, and corre-

<u>Figure III.2.10:</u> background variables passive, category coordinates of background variables (EO has been omitted)



sponding political preference.
In conclusion, our advice is to treat background variables as passive (unless one is prepared to enter into further detailed analysis, as was suggested in III.2.3.1. iv or v). It is true that PRINCALS output for passive variables is rather limited; the user should use a SPSS program (or other standard programs) if more detailed output is wanted.

III.3. Roskam's journal preference data.

III.3.1.

Table III.3.1. gives preference rank orders of 39 psychologists for ten psychological journals (from Roskam, 1968). Columns of the table refer to the following journals:

1: JEXP: Journal of experimental psychology,
2: JAPP: Journal of applied psychology,
3: JPSP: Journal of personality and social psychology,
4: MUBR: Multivariate behavioral research,
5: JCLP: Journal of consulting psychology,
6: JEDP: Journal of educational psychology,
7: PMEK: Psychometrika,
8: HURE: Human relations,
9: BULL: Psychological bulletin,
10: Hude: Human development.

In addition, table III.3.1. contains a final column that identifies each psychologist with respect to the department he or she is affiliated. The codes are:

S: social psychology (4),
D: educational and developmental psychology (7),
C: clinical psychology (4)
M: mathematical psychology and psychological statistics (3),
E: experimental psychology (10),
R: cultural psychology and psychology of religion (3),
T: industrial psychology (6),
A: physiological and animal psychology (2),

Numbers between parentheses in the list above give the number of psychologists from the department.
Table III.3.1. gives preferences in the usual way, from 1 (most preferred journal) to 10 (least preferred journal).

III.3.2.

A matrix of ranking such as given in table III.3.1. can be analyzed in two ways.

(a) columns (journals) as variables, rows (psychologists) as objects,
(b) rows (psychologists) as variables, columns (journals) as objects.

<u>Table III.3.1:</u> Preference rank orders of 39 psychologists for ten psychological journals

1:	7	4	1	8	1Ø	9	5	2	3	6	(S)
2:	7	6	2	9	3	8	1Ø	1	4	5	(S)
3:	1Ø	5	1	7	4	6	8	2	3	9	(S)
4:	6	5	3	7	4	8	9	2	1	1Ø	(S)
5:	6	3	5	1Ø	4	2	9	7	8	1	(D)
6:	8	7	4	9	2	5	1Ø	6	3	1	(D)
7:	5	9	4	8	6	2	1Ø	7	3	1	(D)
8:	6	7	4	9	5	3	1Ø	8	2	1	(D)
9:	2	3	6	4	5	8	9	7	1Ø	1	(D)
1Ø:	5	8	2	9	1	7	1Ø	6	4	3	(D)
11:	7	2	6	1Ø	5	1	9	8	4	3	(D)
12:	8	7	2	9	1	6	1Ø	5	3	4	(C)
13:	1Ø	7	1	9	4	6	8	2	3	5	(C)
14:	5	2	3	4	1	8	7	9	6	1Ø	(C)
15:	6	5	2	7	1	1Ø	9	8	4	3	(C)
16:	4	7	5	2	8	9	1	6	3	1Ø	(M)
17:	4	7	5	3	9	8	1	6	2	1Ø	(M)
18:	5	4	7	3	9	8	1	1Ø	2	6	(M)
19:	1	5	6	7	1Ø	9	3	8	2	4	(E)
2Ø:	1	5	8	7	9	3	6	1Ø	2	4	(E)
21:	3	7	6	2	8	4	5	9	1	1Ø	(E)
22:	1	3	8	6	9	7	4	1Ø	2	5	(E)
23:	1	4	6	5	9	1Ø	2	8	3	7	(E)
24:	1	7	5	4	1Ø	9	3	8	2	6	(E)
25:	1	8	6	5	9	4	3	1Ø	2	7	(E)
26:	1	2	5	6	1Ø	4	7	9	3	8	(E)
27:	1	5	6	4	8	7	2	9	3	1Ø	(E)
28:	4	6	5	1	7	1Ø	3	8	2	9	(E)
29:	8	7	1	2	9	1Ø	6	3	4	5	(R)
3Ø:	7	4	1	2	9	1Ø	8	6	3	5	(R)
31:	9	8	2	7	1	4	1Ø	5	6	3	(R)
32:	7	1	5	8	2	6	3	9	4	1Ø	(T)
33:	2	3	7	8	1Ø	9	1	6	4	5	(T)
34:	1Ø	4	2	9	3	5	6	8	1	7	(T)
35:	3	2	1Ø	6	8	4	7	9	1	5	(T)
36:	6	1	3	9	4	7	1Ø	2	5	8	(T)
37:	2	1	6	4	1Ø	9	5	7	3	8	(T)
38:	2	3	6	5	7	8	4	9	1	1Ø	(A)
39:	2	6	7	3	1Ø	8	4	9	1	5	(A)

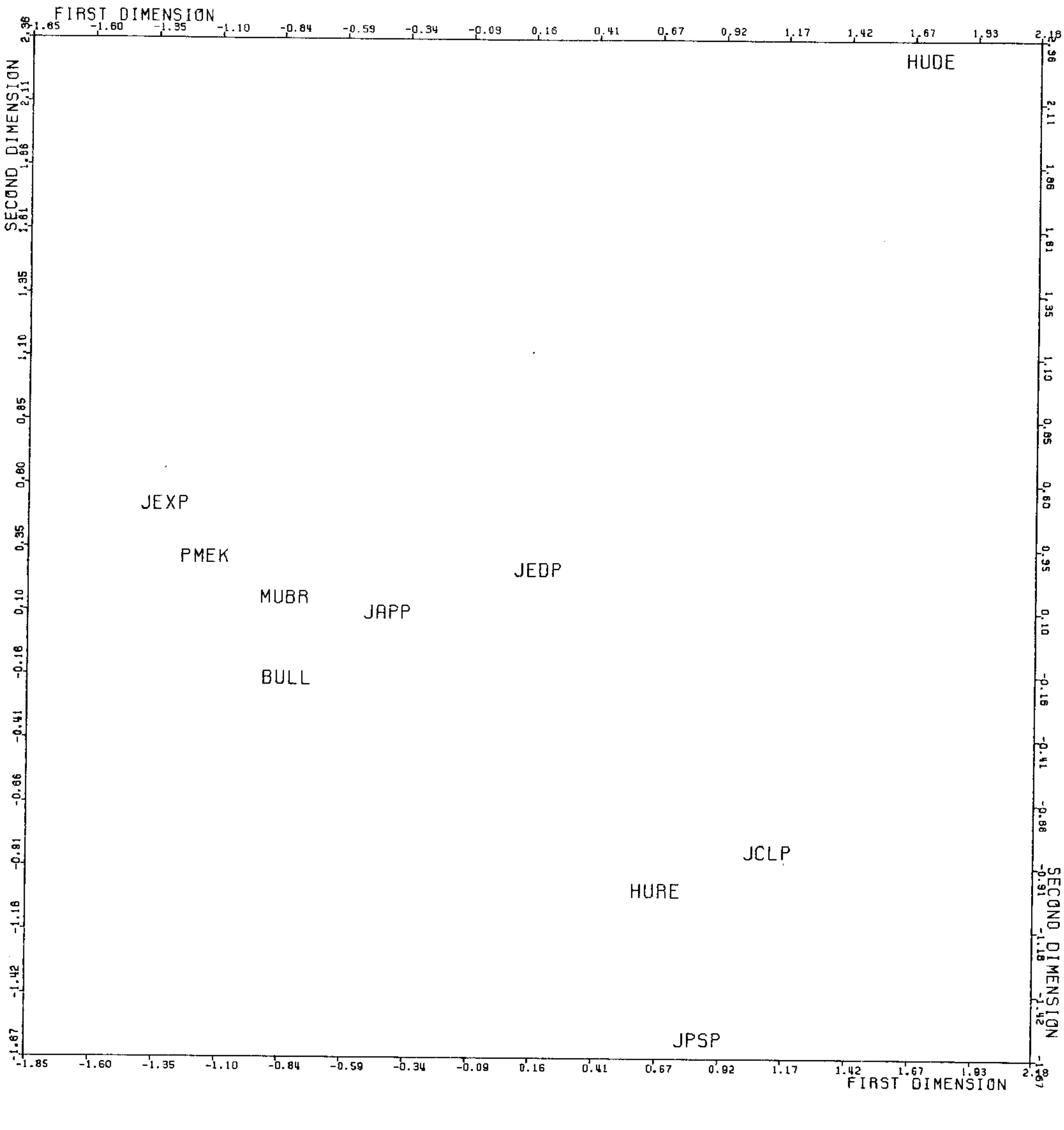
<u>ad a.</u> In the quantified data matrix variables have zero mean. The solution with journals as variables therefore implies that in the quantified data matrix journals no longer differ in average preference score. Note that in the raw data matrix journals are quite different with respect to average preference score: the journals 1,2,3 and 9 are on the average much more highly preferred than the journals 5,6,7 and 8.

<u>ad b.</u> In the quantified data matrix with psychologists as variables, columns will also have zero mean. Note, however, that if for each psychologist the 10 ratings are replaced by deviations from their mean, nothing much changes: all entries of table III.3.1. should be subtracted with the

value 5.5. This has no effect on the averages for journals: they maintain the same differences as before 5.5 was subtracted.
Also, in case (b), it makes no sense to try a solution with all variables (psychologists) nominal. In fact, for all variables, each category occurs only once - the categories are the rank numbers from 1 to 10. As a consequence, no matter how categories are quantified, they always coincide with the object quantification of the one object in the category , so that multiple fit, per variable and per dimension, always equals the maximum of 1.00. In case (a) a nominal solution (either single or multiple) will not necessarily degenerate. However, one should realize that case (a) implies that all psychologists who put journal i on the j'th place in their rank-order, are considered as if they are in the same category - an assumption that might be tenable in some cases, but may become too presumptuous in others. On the other hand, in case (b) such an assumption is not made. In general, therefore, it seems that with ranking data it will be better to take the rankers as variables, and the things ranked as objects. This approach is sometimes called the "vectormodel" for analysis of preferential choice data.

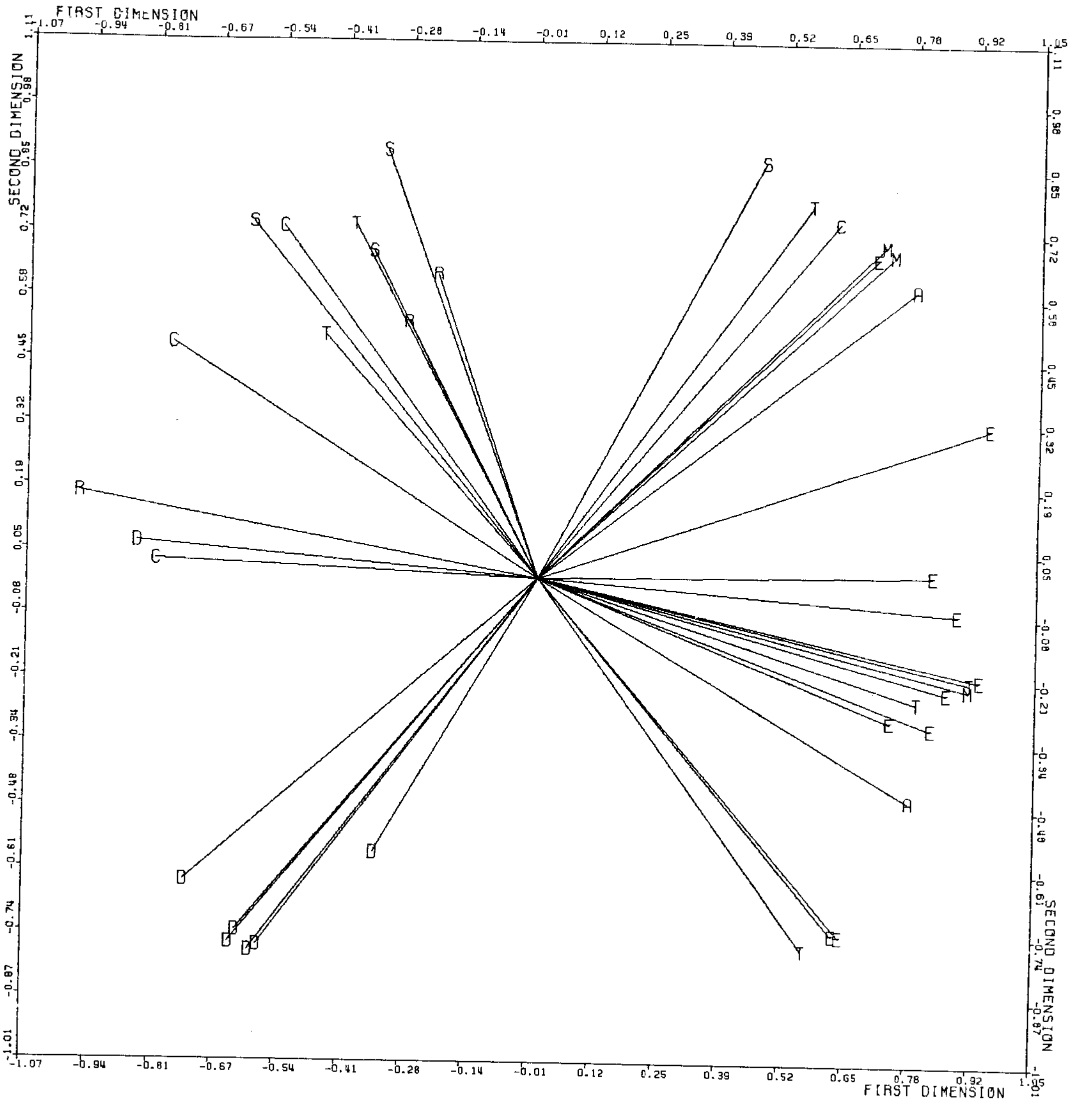
III.3.3. Ordinal solution with psychologists as variables.
The solution in two dimensions, for the present data, has eigenvalues .405 and .157. The plot of object scores (journals) is given in figure III.3.1., that of the component loadings (psychologists), labeled by department, in figure III.3.2.. The plot of psychologists clearly shows, on the first dimension, a contrast between experimental, animal, and mathematical psychologists on the positive side, versus social and developmental psychologists on the other. The second dimension distinguishes between social versus developmental psychologists. The psychologists from the other departments are somewhat scattered in both dimensions. The plot of the journals shows a similar division, with the "hard" journals (1,2,4,7) at the left, the more

Figure III.3.1: ordinal solution, object scores (journals)



"developmental" journals (6,10) to the NE, and the more "social" and "clinical" journals in the SE. Note that in figure III.3.1. the "hard" psychologists are on the right, whereas in figure III.3.2 the "hard" journals are on the left - the reason is that a high score reflects low preference, a low score reflects high preference.
The solution given here is somewhat different from that given in Gifi (1981a, p.191-192). This is probably due to

Figure III.3.2: ordinal solution, component loadings (psychologists) labeled by department



the fact that the solution in Gifi was not computed on the basis of the standard PRINCALS program, and differences in initial configuration may have had their effect on the final solutions. Nevertheless, the sum of the eigenvalues in both solutions is the same. It is quite probable with data of this sort, that there are different quantifications with about the same total fit; the choice of initial configuration then becomes decisive with respect to which of such equivalent solutions will turn up.

III.3.4. Ordinal solution with journals as variables.
The other approach is to take journals as variables, with psychologists as objects. The two-dimensional solution now has eigenvalues .460 and .258. Results are graphed in figures III.3.3./4.. The graph for psychologists is very similar to figure III.3.2.. The graph for journals is

Figure III.3.3: ordinal solution, object scores (psychologists) labeled by department

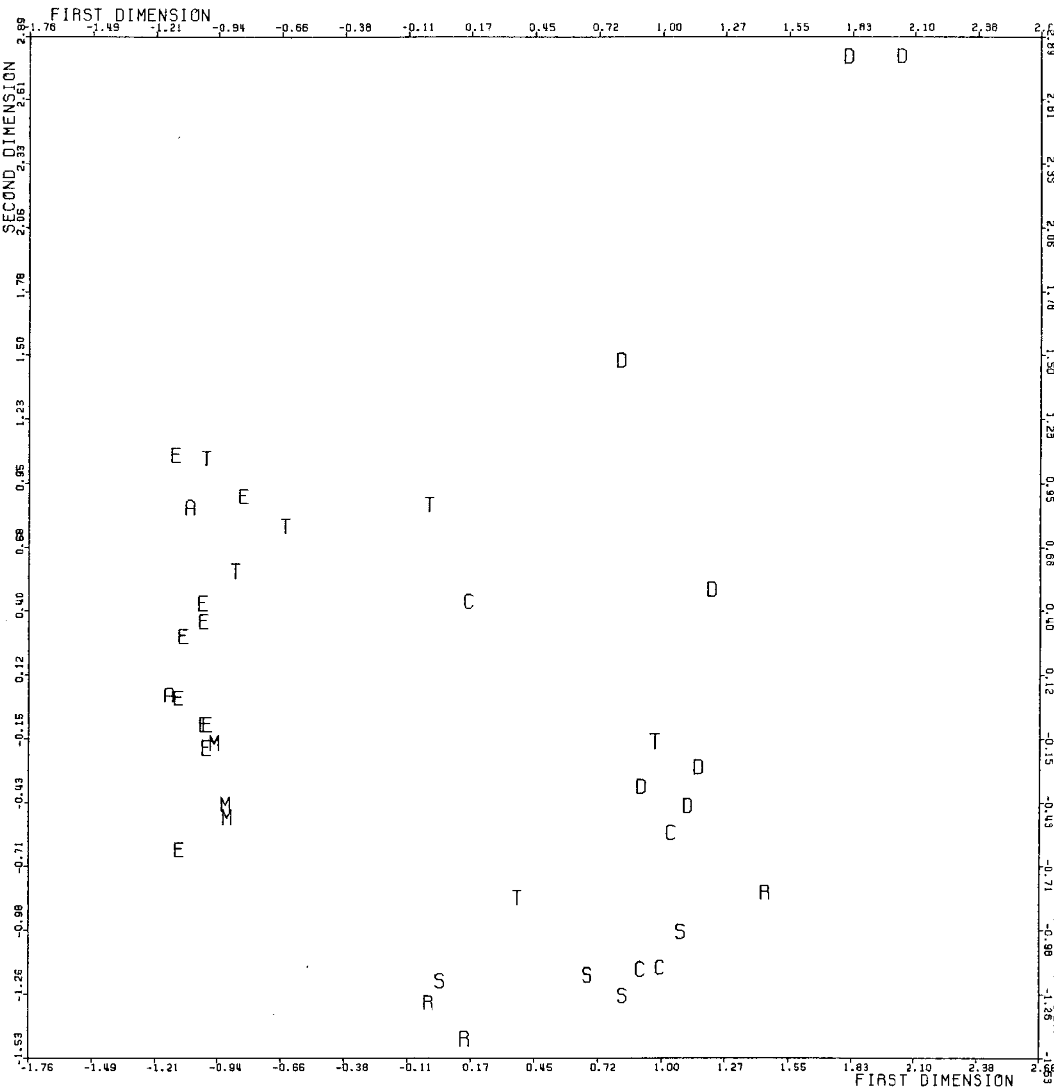
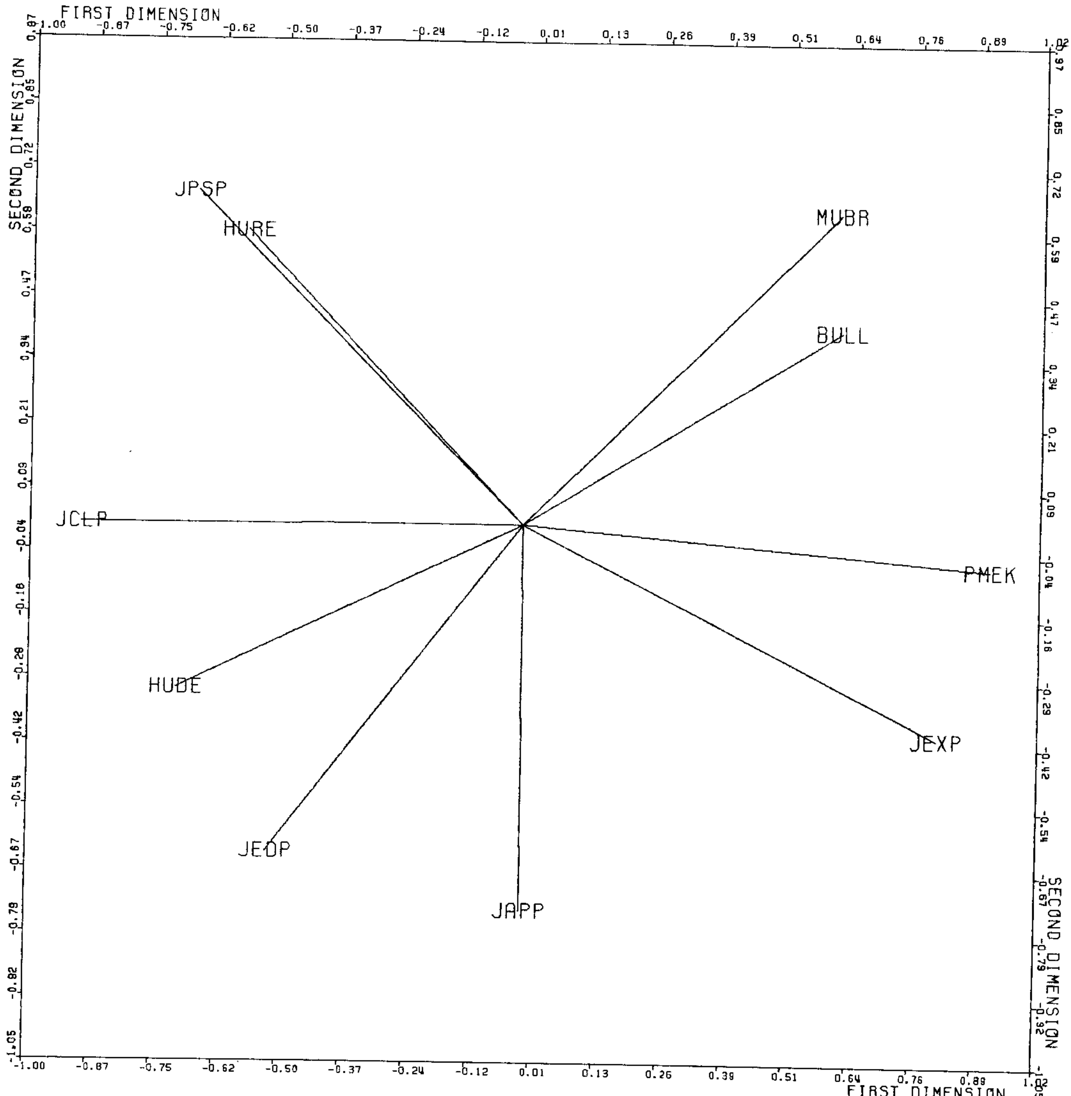


Figure III.3.4: ordinal solution, component loadings (journals)



somewhat different, as could be expected, because in the present analysis differences for average preference for journals - measured in terms of the quantified variables - are eliminated.

III.3.5. PRINCALS on rank-orders.

Summarizing: the present section III.3. has been included in this User's Guide in order to demonstrate that PRINCALS can also handle rankorder data. The example, however, has not been included to suggest that such an analysis should be recommended in all cases. The user should also consider application of one of the SMACOF programs (Heiser & de Leeuw, How to use SMACOF-3, 1979).

APPENDIX

```
P R I N C A L S               A PROGRAM FOR PRINCIPAL COMPONENTS ANALYSIS                ALBERT GIFI
                              OF DISCRETE DATA WITH MIXED MEASUREMENT LEVELS             DEPARTMENT OF DATATHEORY
VERSION    6.ØØ                                                                          THE UNIVERSITY OF LEYDEN
                                                                                         BREESTRAAT 7Ø
MARCH  1981                                                                              LEYDEN
                                                                                         THE NETHERLANDS

    =========
    JOB NR. 1
    =========

    INPUT-DATA SPECIFICATIONS:
    -------------------------

    NAME : PRINCALS: GUTTMAN-BELL DATA, ALL VARIABLES SINGLE ORDINAL, TWO DIMENSIONS

    * PROBLEM PARAMETERS *
      ------------------

    NUMBER OF OBJECTS                                                              7
    TOTAL NUMBER OF VARIABLES IN THE DATAMATRIX                                    5
    NUMBER OF ANALYSIS-VARIABLES                                                   5
    NUMBER OF DIMENSIONS                                                           2
    MAXIMUM NUMBER OF CATEGORIES OVER ALL VARIABLES                                4
    TOTAL NUMBER OF CATEGORIES OF THE ANALYSIS VARIABLES                          17


    * ANALYSIS PARAMETERS *
      -------------------

    MAXIMUM NUMBER OF ITERATIONS TO COMPUTE THE FINAL CONFIGURATION               75
    CONVERGENCE CRITERION FOR THE FINAL CONFIGURATION                      Ø.5ØE-Ø4
    MAXIMUM NUMBER OF ITERATIONS TO COMPUTE THE INITIAL CONFIGURATION             2Ø
    CONVERGENCE CRITERION FOR THE INITIAL CONFIGURATION                    Ø.71E-Ø2
```

```
* INPUT/OUTPUT PARAMETERS *
  -----------------------

UNIT NUMBER OF THE DATAMATRIX                                          5

PARAMETER INDICATING WHETHER OR NOT (1 AND Ø RESP.)                    1
TO PRINT THE DATAMATRIX

PARAMETER INDICATING WHETHER OR NOT TO PRINT                           2
OBJECT- AND CATEGORY INFORMATION
  =Ø: NO PRINT
  =1: OBJECT SCORES ONLY
  =2: OBJECT SCORES AND OPTIMALLY SCALED CATEGORIES
  =3: OPTIMALLY SCALED CATEGORIES ONLY

PARAMETER INDICATING WHETHER OR NOT (1 AND Ø RESP.)                    1
TO PRINT THE ITERATION HISTORY

PARAMETER INDICATING WHETHER OR NOT AND HOW TO PLOT                    2
  =Ø: NO PLOT
  =1: OBJECT SCORES, UNLABELED, AND COMPONENT LOADINGS
  =2: IN ADDITION, PLOT OF OBJECT SCORES AND/OR
      SEPARATE PLOTS OF THE CATEGORY QUANTIFICATIONS
      LABELED BY SELECTED VARIABLES

UNIT-NUMBER FOR OUTPUT OF THE OBJECT SCORES TO                         3
CARD, TAPE OR DISK (=Ø: NO OUTPUT OF THIS KIND)

LIKEWISE FOR THE CATEGORY COORDINATES                                  9

LIKEWISE FOR THE CATEGORY QUANTIFICATIONS                              7

MEASUREMENT LEVEL FOR ALL VARIABLES                                    3
  =Ø: MIXED MEASUREMENT LEVELS
  =1: ONLY MULTIPLE NOMINAL VARIABLES
  =2: ONLY SINGLE NOMINAL VARIABLES
  =3: ONLY SINGLE ORDINAL VARIABLES
  =4: ONLY SINGLE NUMERICAL VARIABLES

OUTPUT SUPPRESSING PARAMETER                                           Ø
  =Ø: NO OUTPUT OF THE INITIAL CONFIGURATION
  =1: IDENTICAL OUTPUT OPTIONS FOR BOTH THE INITIAL
      AND FINAL CONFIGURATION
  =2: OUTPUT OPTIONS FOR BOTH CONFIGURATIONS ARE
      FULLY SPECIFIED BY THE USER

NUMBER OF CATEGORIES FOR ALL VARIABLES                                 Ø
  =K: THE CATEGORY NUMBERS FOR ALL VARIABLES EQUAL K
  =Ø: THE CATEGORY NUMBERS FOR ALL VARIABLES ARE DIFFERENT

UNIT NUMBER FOR OUTPUT OF THE COMPONENT LOADINGS TO                    4
CARD, TAPE OR DISK (=Ø: NO OUTPUT OF THIS KIND)
```

* NUMBER OF CATEGORIES PER VARIABLE * :

VARIABLE	NUMBER OF CATEGORIES	VARIABLE	NUMBER OF CATEGORIES	VARIABLE	NUMBER OF CATEGORIES	VARIABLE	NUMBER OF CATEGORIES
1	4	3	4	5	3		
2	4	4	2				

THE FOLLOWING VARIABLES ACT AS LABELING-VARIABLES IN THE OBJECT SCORES PLOT, IF LABELED WITH A STAR, AND, IF LABELED WITH A SLASH, THE OPTIMALLY SCALED CATEGORIES OF THOSE VARIABLES ARE PLOTTED:

1 */

* FORMAT TO READ THE DATAMATRIX * : (5I2)

7 ROWS OF THE DATAMATRIX :

```
 *
  * VARIABLES
   *
    *
     *    1   2   3   4   5
OBJECTS  ***********************
      1 *   1   1   1   2   2
      2 *   2   2   2   2   2
      3 *   1   1   2   1   1
      4 *   4   2   4   2   3
      5 *   4   4   4   2   3
      6 *   3   3   3   1   2
      7 *   2   3   3   2   2
```

* MEASUREMENT LEVEL FOR ALL VARIABLES IS * : SINGLE ORDINAL

* MARGINAL FREQUENCIES *:

VARIABLES \ CATEGORIES	MISSING	1	2	3	4
1	Ø	2	2	1	2
2	Ø	2	2	2	1
3	Ø	1	2	2	2
4	Ø	2	5		
5	Ø	1	4	2	

* THE HISTORY OF ITERATIONS TO COMPUTE THE FINAL CONFIGURATION *

ITERATION NUMBER	TOTAL FIT	TOTAL LOSS	MULTIPLE LOSS	SINGLE LOSS	DIFFERENCE BETWEEN THE LAST TWO ITERATIONS
1	Ø.9278243	1.Ø721757	Ø.8148884	Ø.2572873	Ø.Ø247664
2	Ø.9348624	1.Ø651376	Ø.8192Ø89	Ø.2459287	Ø.ØØ7Ø381
3	Ø.9362593	1.Ø6374Ø7	Ø.8243ØØØ	Ø.23944Ø6	Ø.ØØ1397Ø
4	Ø.937Ø247	1.Ø629753	Ø.8261Ø39	Ø.2368714	Ø.ØØØ7654
5	Ø.9376295	1.Ø6237Ø5	Ø.8264288	Ø.2359417	Ø.ØØØ6Ø48
6	Ø.9381572	1.Ø618428	Ø.8266523	Ø.23519Ø5	Ø.ØØØ5277
7	Ø.9386715	1.Ø613285	Ø.8271581	Ø.23417Ø4	Ø.ØØØ5143
8	Ø.9391863	1.Ø6Ø8137	Ø.8279641	Ø.2328496	Ø.ØØØ5148
9	Ø.9396988	1.Ø6Ø3Ø12	Ø.8289738	Ø.2313273	Ø.ØØØ5125
1Ø	Ø.94Ø2Ø67	1.Ø597933	Ø.83ØØ946	Ø.2296988	Ø.ØØØ5Ø78
11	Ø.94Ø71Ø7	1.Ø592893	Ø.8312715	Ø.228Ø177	Ø.ØØØ5Ø41
12	Ø.9412Ø84	1.Ø587916	Ø.8324817	Ø.2263Ø99	Ø.ØØØ4977
13	Ø.9417ØØ7	1.Ø582993	Ø.8337177	Ø.2245816	Ø.ØØØ4923
14	Ø.942182Ø	1.Ø57818Ø	Ø.8349756	Ø.2228424	Ø.ØØØ4813
15	Ø.9426522	1.Ø573478	Ø.8362495	Ø.221Ø983	Ø.ØØØ47Ø2
16	Ø.9431Ø67	1.Ø568933	Ø.83753Ø9	Ø.2193623	Ø.ØØØ4546
17	Ø.9435427	1.Ø564573	Ø.83885Ø9Ø	Ø.2176482	Ø.ØØØ436Ø
18	Ø.9439584	1.Ø56Ø416	Ø.84ØØ727	Ø.2159689	Ø.ØØØ4157
19	Ø.94435Ø5	1.Ø556495	Ø.84131Ø8	Ø.2143388	Ø.ØØØ3921
2Ø	Ø.94469Ø2	1.Ø553Ø98	Ø.8425131	Ø.2127967	Ø.ØØØ3397
21	Ø.9448927	1.Ø551Ø73	Ø.844216Ø	Ø.21Ø8912	Ø.ØØØ2Ø26
22	Ø.945Ø137	1.Ø549863	Ø.8465169	Ø.2Ø84694	Ø.ØØØ121Ø
23	Ø.945Ø864	1.Ø549136	Ø.8484559	Ø.2Ø64577	Ø.ØØØØ727
24	Ø.94513Ø9	1.Ø548691	Ø.85ØØ138	Ø.2Ø48553	Ø.ØØØØ445

THE ITERATIVE PROCESS STOPS BECAUSE THE CONVERGENCE TEST VALUE IS REACHED

DIMENSION	EIGENVALUE
1	Ø.68Ø
2	Ø.265

VARIABLE 1. TYPE: SINGLE ORDINAL MISSING: Ø

DIMENSION : 1 2

CATEGORY	MARGINAL FREQUENCY	CATEGORY QUANTIF.	SINGLE CATEGORY COORDINATES 1	SINGLE CATEGORY COORDINATES 2
1	2	-1.525	1.5Ø4	-Ø.2Ø7
2	2	Ø.342	-Ø.337	Ø.Ø46
3	1	Ø.381	-Ø.376	Ø.Ø52
4	2	Ø.992	-Ø.979	Ø.135

MULTIPLE CATEGORY COORDINATES 1	MULTIPLE CATEGORY COORDINATES 2
1.49Ø	-Ø.3Ø8
-Ø.373	-Ø.213
-Ø.185	1.434
-1.Ø25	-Ø.197

VARIABLE 2. TYPE: SINGLE ORDINAL MISSING: Ø

DIMENSION : 1 2

CATEGORY	MARGINAL FREQUENCY	CATEGORY QUANTIF.	SINGLE CATEGORY COORDINATES (1)	SINGLE CATEGORY COORDINATES (2)
1	2	-1.55Ø	1.499	-Ø.259
2	2	Ø.5Ø4	-Ø.487	Ø.Ø84
3	2	Ø.5Ø4	-Ø.487	Ø.Ø84
4	1	1.Ø86	-1.Ø5Ø	Ø.182

MULTIPLE CATEGORY COORDINATES

1.49Ø	-Ø.3Ø8
-Ø.6ØØ	-Ø.35Ø
-Ø.336	Ø.737
-1.1Ø9	-Ø.159

VARIABLE 3. TYPE: SINGLE ORDINAL MISSING: Ø

DIMENSION : 1 2

CATEGORY	MARGINAL FREQUENCY	CATEGORY QUANTIF.	SINGLE CATEGORY COORDINATES (1)	SINGLE CATEGORY COORDINATES (2)
1	1	-2.186	1.611	-1.395
2	2	-Ø.325	Ø.24Ø	-Ø.2Ø7
3	2	Ø.7Ø9	-Ø.523	Ø.452
4	2	Ø.7Ø9	-Ø.523	Ø.452

MULTIPLE CATEGORY COORDINATES

1.261	-1.8ØØ
Ø.73Ø	Ø.36Ø
-Ø.336	Ø.737
-1.Ø25	-Ø.197

VARIABLE 4. TYPE: SINGLE ORDINAL MISSING: Ø

DIMENSION : 1 2

CATEGORY	MARGINAL FREQUENCY	CATEGORY QUANTIF.	SINGLE CATEGORY COORDINATES 1	SINGLE CATEGORY COORDINATES 2
1	2	-1.581	Ø.767	1.31Ø
2	5	Ø.632	-Ø.3Ø7	-Ø.524

MULTIPLE CATEGORY COORDINATES 1	MULTIPLE CATEGORY COORDINATES 2
Ø.767	1.31Ø
-Ø.3Ø7	-Ø.524

VARIABLE 5. TYPE: SINGLE ORDINAL MISSING: Ø

DIMENSION : 1 2

CATEGORY	MARGINAL FREQUENCY	CATEGORY QUANTIF.	SINGLE CATEGORY COORDINATES 1	SINGLE CATEGORY COORDINATES 2
1	1	-2.183	1.844	Ø.94Ø
2	4	Ø.Ø17	-Ø.Ø15	-Ø.ØØ7
3	2	1.Ø57	-Ø.893	-Ø.455

MULTIPLE CATEGORY COORDINATES 1	MULTIPLE CATEGORY COORDINATES 2
1.72Ø	1.185
Ø.Ø82	-Ø.198
-1.Ø25	-Ø.197

SUMMARY OF ANALYSIS

DIMENSION : 1 2

	ROW SUMS	MULTIPLE FIT (1)	MULTIPLE FIT (2)
1	1.324	Ø.979	Ø.345
2	1.166	Ø.945	Ø.221
3	1.378	Ø.712	Ø.666
4	Ø.922	Ø.236	Ø.686
5	Ø.96Ø	Ø.726	Ø.234
MEAN	1.15Ø	Ø.72Ø	Ø.43Ø

	ROW SUMS	SINGLE FIT (1)	SINGLE FIT (2)
1	Ø.991	Ø.973	Ø.Ø18
2	Ø.963	Ø.935	Ø.Ø28
3	Ø.951	Ø.544	Ø.4Ø7
4	Ø.922	Ø.236	Ø.686
5	Ø.899	Ø.714	Ø.186
MEAN	Ø.945	Ø.68Ø	Ø.265

	COMPONENT LOADINGS (1)	COMPONENT LOADINGS (2)
1	-Ø.986	Ø.136
2	-Ø.967	Ø.167
3	-Ø.737	Ø.638
4	-Ø.485	-Ø.828
5	-Ø.845	-Ø.431

ITERATION NUMBER	TOTAL FIT	TOTAL LOSS	MULTIPLE LOSS	SINGLE LOSS
24	Ø.9451	1.Ø549	Ø.85ØØ	Ø.2Ø49

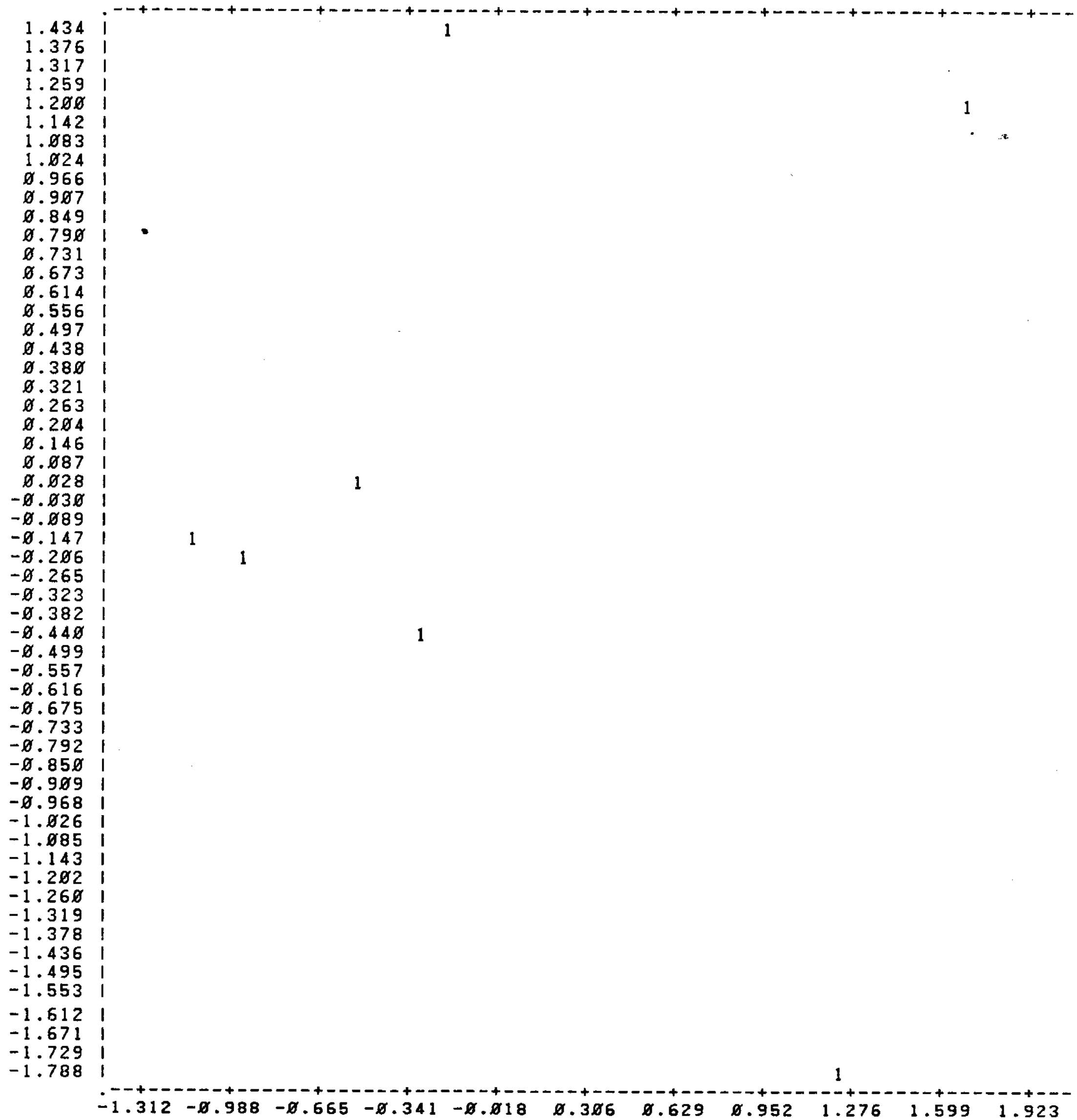
```
* CORRELATIONS BETWEEN OPTIMALLY SCALED VARIABLES *
  -----------------------------------------------
        1          2          3          4          5

1       *
2     Ø.977       *
3     Ø.8Ø5     Ø.796       *
4     Ø.362     Ø.331    -Ø.121       *
5     Ø.774     Ø.723     Ø.313     Ø.685       *
* OBJECT SCORES * :
  -------------
```

```
      *
       * DIMENSIONS
        *
         *
          *         1          2
OBJECTS    ***********************
         1 *     1.26Ø9    -1.8ØØ1
         2 *    -Ø.2591    -Ø.4654
         3 *     1.7197     1.1848
         4 *    -Ø.94Ø6    -Ø.2339
         5 *    -1.1Ø87    -Ø.1591
         6 *    -Ø.1851     1.4344
         7 *    -Ø.4872     Ø.Ø394
```

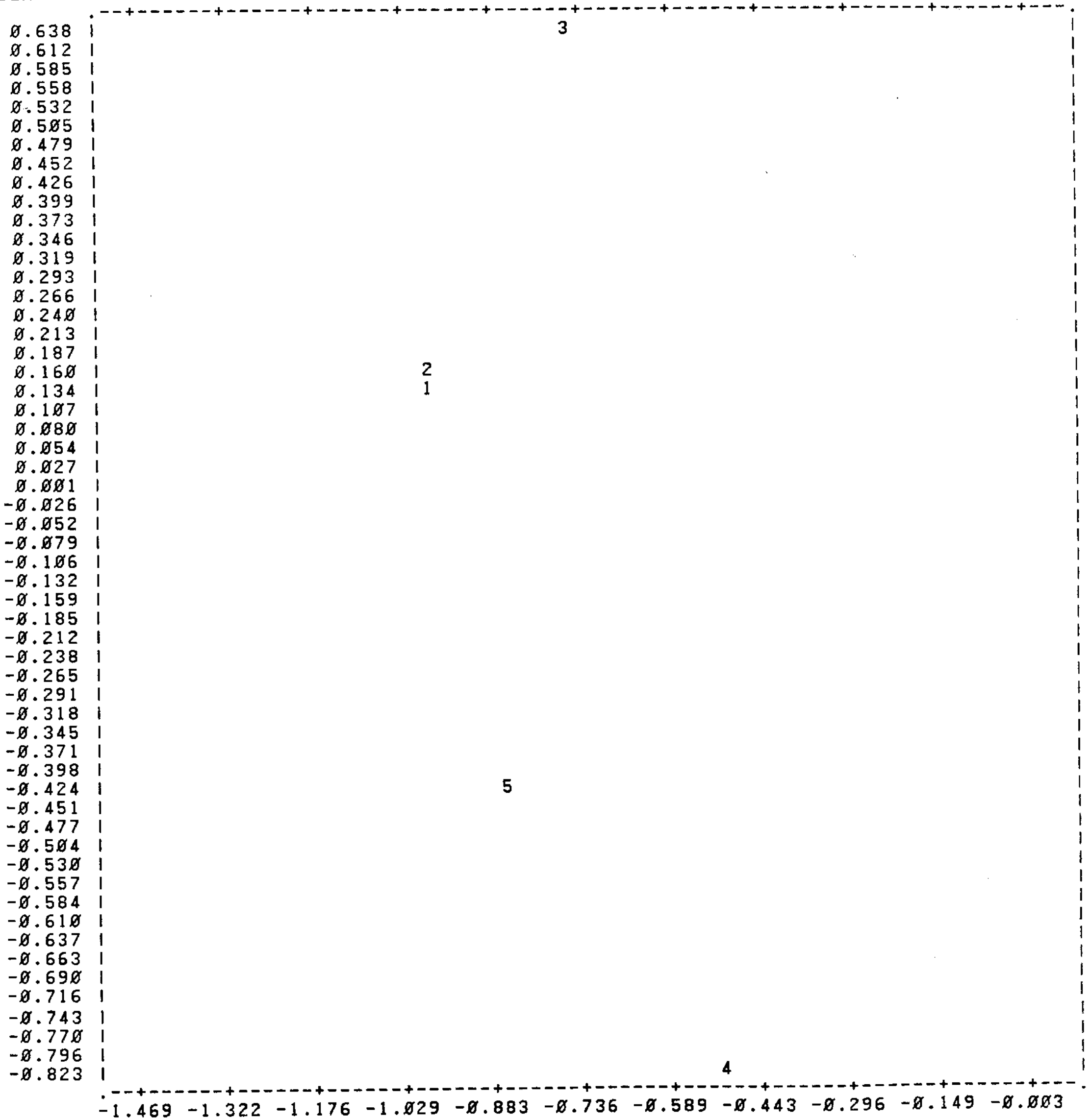
OBJECT SCORES, UNLABELED

PRINCALS: GUTTMAN-BELL DATA, ALL VARIABLES SINGLE ORDINAL, TWO DIMENSIONS



COMPONENT LOADINGS LABELED
BY THEIR VARIABLE NUMBER

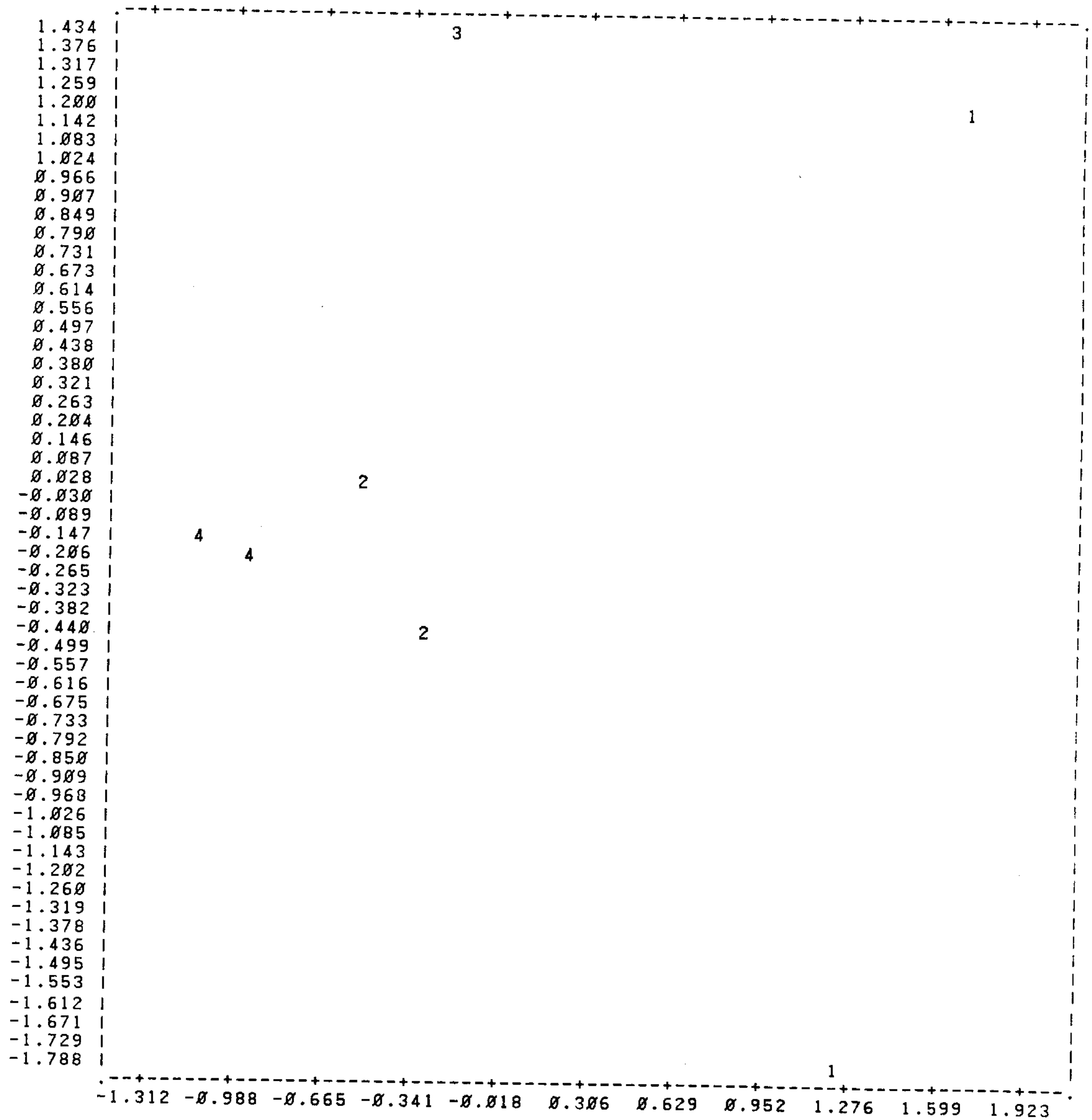
PRINCALS: GUTTMAN-BELL DATA, ALL VARIABLES SINGLE ORDINAL, TWO DIMENSIONS



```
RESCALED CATEGORIES FOR VARIABLE    1          PRINCALS: GUTTMAN-BELL DATA, ALL VARIABLES SINGLE ORDINAL, TWO DIMENSIONS
LABELED BY THEIR CATEGORY NUMBER
        .---+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+-----.
  1.828 I                                                                                                             I
  1.782 I                                                                                                             I
  1.736 I                                                                                                             I
  1.690 I                                                                                                             I
  1.645 I                                                                                                             I
  1.599 I                                                                                                             I
  1.553 I                                                                                                             I
  1.507 I                                                                                                             I
  1.461 I                                                                                                             I
  1.416 I                                    3                                                                        I
  1.370 I                                                                                                             I
  1.324 I                                                                                                             I
  1.278 I                                                                                                             I
  1.232 I                                                                                                             I
  1.187 I                                                                                                             I
  1.141 I                                                                                                             I
  1.095 I                                                                                                             I
  1.049 I                                                                                                             I
  1.003 I                                                                                                             I
  0.958 I                                                                                                             I
  0.912 I                                                                                                             I
  0.866 I                                                                                                             I
  0.820 I                                                                                                             I
  0.774 I                                                                                                             I
  0.729 I                                                                                                             I
  0.683 I                                                                                                             I
  0.637 I                                                                                                             I
  0.591 I                                                                                                             I
  0.545 I                                                                                                             I
  0.500 I                                                                                                             I
  0.454 I                                                                                                             I
  0.408 I                                                                                                             I
  0.362 I                                                                                                             I
  0.316 I                                                                                                             I
  0.270 I                                                                                                             I
  0.225 I                                                                                                             I
  0.179 I                                                                                                             I
  0.133 I     4                                                                                                       I
  0.087 I                                                                                                             I
  0.041 I                           3  2                                                                              I
 -0.004 I                                                                                                             I
 -0.050 I                                                                                                             I
 -0.096 I                                                                                                             I
 -0.142 I                                                                                                             I
 -0.188 I   4                                                                                                   1     I
 -0.233 I                             2                                                                               I
 -0.279 I                                                                                                             I
 -0.325 I                                                                                                      1      I
 -0.371 I                                                                                                             I
 -0.417 I                                                                                                             I
 -0.462 I                                                                                                             I
 -0.508 I                                                                                                             I
 -0.554 I                                                                                                             I
 -0.600 I                                                                                                             I
 -0.646 I                                                                                                             I
 -0.691 I                                                                                                             I
        .---+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+-----.
        -1.025 -0.772 -0.519 -0.266 -0.013 0.240 0.493 0.746 0.998 1.251 1.504
```

OBJECT SCORES, LABELED BY
VARIABLE 1

PRINCALS: GUTTMAN-BELL DATA, ALL VARIABLES SINGLE ORDINAL, TWO DIMENSIONS



REFERENCES

Burg, E. van der,

1983 CANALS User's guide. Dep. of Datatheory, Faculty of Social Sciences, University of Leiden, the Netherlands.

Diekstra, R.F.W. and A.J.F.M., Kerkhof,

1982 Suicide pogingen in de bevolking. Dep. of Clinical Psychology, University of Leiden, The Netherlands.

Gifi, A.

1981a Non-linear Multivariate Analysis, Dep. of Datatheory, Faculty of Social Sciences, University of Leiden, The Netherlands

Gifi, A.

1981b HOMALS User's Guide, Dep. of Datatheory, Faculty of Social Sciences, University of Leiden, The Netherlands.

Guttman, L.

1968 A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33, 469-506.

Heiser, W.J. and J. de Leeuw

1979 How to use SMACOF-3, Dep. of Datatheory, Faculty of Social Sciences, University of Leiden, The Netherlands.

Lingoes, J.C.

1968 The Multivariate analysis of qualitative data. Multivariate Behavioral Research, 3, 61-94.

Roskam, E.E.

1968 Metric analysis of ordinal data in psychology. Voorschoten, VAM.

Speyer, N. and R.F.W. Diekstra

1980 Hulp bij zelfdoding, een studie van problemen rond de zelfgewilde dood. Van Loghum Slaterus, Deventer.