RN 007-68

CANONICAL DISCRIMINANT ANALYSIS OF RELATIONAL DATA

Jan de Leeuw

April 1968

PSYCHOLOGICAL INSTITUTE - UNIVERSITY OF LEIDEN- NETHERLANDS

CONTENTS

~~[illegible]~~ 9) multirelational systems

10) Statistical Aspects

~~[illegible]~~ Appendix I: Necessary Set theoretical Notation {Kelley, Tarski}

II: Necessary Matrix Algebra

SUMMARY

In the past few years there has been an increasing tendency in psychological scaling theory towards emphasizing only qualitative and/or ordinal features of data structures. As a matter of fact, most psychological data do indeed arise in a straightforward manner from simple classifications (right-wrong, yes-no, greater than-less than). These new measurement models have some remarkable properties in common: 1) they are all special cases of general conjoint measurement, 2) they can be formulated quite rigorously in terms of a partial order over the cartesian product of some finite number of sets, and 3) numerical representations can be derived by some general algorithms, which are applicable, at least in principle, to every member of this class of models. An important class of algorithms consists of measurement-oriented algorithms that seek to find an optimal fit to a nonmetric measurement model, which amounts in most cases to finding the maximum solvable subset of some system of inequalities. In this paper we discuss a fundamentally different class of algorithms the principal aim of which is data reduction. Elsewhere we have called them psychometric algorithms. The general approach, maximizing a generalized correlation coefficient, is applied to some particular cases like simple quantification (ch II), additive conjoint measurement and non-metric factor analysis (ch III), multiple relational systems and multiway matrices (ch IV). A short overview of a series of computer programs is given, as well as some examples of the various special cases.

++++++++++

Chapter I

INTRODUCTION

-1.1-

The current developments in measurement theory have casted doubt on the rigourous distinction between quantitative and qualitative data. In the terminology of Scott & Suppes (1958) or Suppes & Zinnes (1963), it is more useful to distinguish between fundamental and derived measurement. In derived measurement we operate on numbers that have been obtained by applying fundamental measurement procedures (in most cases extensive measurement) to the relevant relational systems. In fundamental measurement we operate on the relational sytems directly, and these relational systems constitute the so-called qualitative data. Long before these concepts from measurement theory were available, the distinction between quantitative and qualitative data had already been attacked by Guttman. "It is only that most of us have been exposed exclusively to certain algebraic manipulations that we conceive such manipulations to be the essence of mathematics. A more sophisticated view is to regard mathematics as unveiling necessary relationships that arise from classifications. Much useless discussion of mathematics as a 'tool' in social research could be saved by recognition of the fact that qualitative classifications lead to just as rigorous implications as do quantitative."(Guttman, 1944, p 139). In later papers Guttman continued criticizing the textbooks on psychological statistics, that treated qualitative data as inferior to quantitative data (Guttman 1950a, p 193), and he contrasted their treatment of qualitative data with that of Yule (1910). As Burt (1953, p 10) has pointed out, British psychometricians have been considerably more liberal in their treatment and interpretation of qualitative data, more in line with the practice of statisticians (Burt gives some familiar examples from statistics and probability theory, where quantitative and qualitative data are treated alike, using devices such as measure theory, Stieltjes integrals and generalized correlation coefficients). Despite this discussion, despite the continuing

cf Coombs: Inferential classification

contributions of statisticians (cf Kendall & Stuart, vol II, ch 33) and despite the elegant methods both Guttman and Burt proposed to quantify nominal data (Guttman 1941b, 1946, 1950, 1959a) or even to factorize them (Burt 1950), some psychologists continued to treat classificatory variables as non-scalable and useful only in combination with quantitative variables. Their usefulness in that context can hardly be denied, because analysis of variance, analysis of covariance, as well as discriminatory analysis, are based on the combination of classificatory and numerical variates. A very neat analysis of the interrelations of both types of data is given in the articles by Guttman (1941a) and Lubin (1950).

-1.2-

Thanks to the developments in measurement theory, the work of Coombs and his associates, and Kendall's contributions to ranking theory, there is at least one type of qualitative data, that has been universally accepted as scalable and useful. If the relation in our empirical relational system is a partial order, then there are a number of respectable scaling methods which can be used to analyze the system. The contributions of Coombs, summarized in Coombs (1964), have been very important from a historical and systematical point of view, but his analytic techniques were already out of date before they were even universally accepted and used. One of the reasons was the invention of a practical computing routine for nonmetric multidimensional scaling by Shepard (1962), and the perfection and generalisation of his method by Kruskal (1964a,b) and Guttman & Lingoes (Guttman 1967, 196-, Lingoes 1967, 196-). They constructed computer oriented routines, with which all ordinal problems can, at least in principle, be solved. For a discussion and critique of these algorithms, and a presentation of some alternatives, see De Leeuw (196- a). Their enormous generality is due to the fact that all these nonmetric problems can be shown to be particular cases of general conjoint measurement (Tversky 1967, Krantz 1967). The superiority of these computer algorithms to the techniques proposed by Coombs is proved by Shepard's

Monte Carlo studies on the determinateness of nonmetric scaling and factor analytic solutions (Shepard 1966). As Guttman, more than 25 years after his pioneering contributions to nonmetric methodology, observes: "In order to comprehend great complexities, it proves to be effective and powerful to focus only on most qualitative features; from these can be derived actual metric consequences, with no special assumptions." (Guttman 1966a, p 510).

-1.3-

A partial order is a binary relation. It has some special characteristics (transitivity, antisymmetry, reflexivity), which make it attractive for measurement theoreticians, mainly because the theory of partially ordered groups and rings is already a highly developed part of abstract algebra. The general quantification technique proposed by Guttman in 1941 can, however, be applied (with some adaptations) to ordered data that arise for example in the analysis of paired comparisons (Guttman 1946). It can also be applied, with very interesting results, to scalogram analysis (Guttman 1950b, 1954, Burt 1953). Scalogram analysis is the first example of a fully developed additive conjoint measurement model in psychometric literature, i.e. it is based on order relations between elements of two different sets. The feasability of Guttman's technique, and of the similar technique proposed in this paper, for analyzing partial orders, is based on the relationship between partial orders and a certain type of matrix. Suppose the data consist of the relational system $\langle A, B, \not\succ_o \rangle$, where A and B are finite sets with respectively n and m elements, and where $\not\succ_o$ is a partial order. Define the nxm dominance matrix Z, with

$$\begin{cases} z_{ij} = 1 \Leftrightarrow a_i \succ_o b_j \\ z_{ij} = 0 \Leftrightarrow a_i =_o b_j \Leftrightarrow a_i \not\succ_o b_j \;\; \& \;\; b_j \not\succ_o a_i \\ z_{ij} = -1 \Leftrightarrow b_j \succ_o a_i \end{cases}$$

Some familiar applications of this matrix are its use in the Coombs-Hays method of nonmetric euclidean scaling and in the theory of generalized correlation coefficients (Daniels 1944, Kendall 1962).

-1.4-

The kind of data for which scalogram analysis was developed, inspired other writers, for example Lingoes (1963a) and Coombs & Kao (1955). In the sequel these data will be called relational data, a more satisfactory term than either ordinal, categorical, nominal, or qualitative data. A formal description of relational data can be given in terms of cartesian products, ordered n-tuples, binary of more general relations, and relational systems. This conceptual system has already proved its usefulness in measurement theory (Suppes & Zinnes 1963), in facet theory (Guttman 1959a, 1961, 1966b, 1967, Foa 1965, Lingoes & Vandenbergh 1966, Wish 1965), and in data theory (Coombs 1964). In the following chapters some important relational systems, corresponding with the most common data matrices, will be considered.

+++++++++++

Chapter II

GENERAL CASE: SIMPLE QUANTIFICATION

-2.1-

Let A and B be two finite sets, i.e.

$$A = \{a_1, \dots, a_i, \dots, a_n\}$$

$$B = \{b_1, \dots, b_j, \dots, b_m\}$$

The set Δ is defined as a subset of the cartesian product A x B. Let ϕ denote the indicator of this subset, i.e.

$$\phi(a_i, b_j) = \begin{cases} 1 \Leftrightarrow (a_i, b_j) \in \Delta \\ 0 \Leftrightarrow (a_i, b_j) \notin \Delta \Leftrightarrow (a_i, b_j) \in A \times B - \Delta \end{cases}$$

The values of this function

$$\phi: A \times B \rightarrow \{0, 1\}$$

can be portrayed in an nxm matrix $E = \{e_{ij}\}$, in which $e_{ij} = \phi(a_i, b_j)$. This matrix E is called the attribute or trait matrix by Lingoes (196-). We shall call it the indicator-of-cartesian-product, or ICP-matrix. This ICP-matrix is thus defined in terms of the facets A and B, in conjunction with some binary relation $R \subseteq A \times B$. In summary

$$e_{ij} = 1 \Leftrightarrow \phi(a_i, b_j) = 1 \Leftrightarrow (a_i, b_j) \in \Delta \Leftrightarrow a_i R b_j$$
$$e_{ij} = 0 \Leftrightarrow \phi(a_i, b_j) = 0 \Leftrightarrow (a_i, b_j) \notin \Delta \Leftrightarrow \neg\, a_i R b_j$$

The technique discussed in this chapter quantifies the facets A and/or B by mapping them into Euclidean multidimensional space. More explicitly, we construct functions

$$f : A \rightarrow Re^p = Re \times \dots \times Re \quad (p \text{ times})$$

$$g : B \rightarrow Re^q = Re \times \dots \times Re \quad (q \text{ times})$$

in such a way that some rational criterium is optimized.

-2.2-

The matrix E contains all information in the relational system $\langle A, B, R \rangle$, and vice versa. The relation R has a complement with respect to A x B, which may be called not-R, or $\overline{R}$. Evidently

$$\overline{R} = A \times B - \Delta$$

or

$$a_i\overline{R}b_j \Leftrightarrow \neg\, a_iRb_j$$

In the same way as $\langle A,B,R\rangle$ defines E, the relational system $\langle A,B,\overline{R}\rangle$ defines an ICP-matrix H, which is the opposite of E, in the sense that

$$h_{ij} = 1 \Leftrightarrow e_{ij} = 0$$
$$h_{ij} = 0 \Leftrightarrow e_{ij} = 1$$

or, equivalently,

$$h_{ij} + e_{ij} = 1 \qquad \text{for all } i=1,\dots,n;\ j=1,\dots,m$$

We shall adopt the convention in this (and other) papers to write J_{nm} for an nxm matrix, with each element equal to unity. In the same way, j_n is an n-element vector, with $j_i=1$ for each i. When no confusion is liable to arise, subscripts will be dropped. For example $j_n'j_n = n$, $J_{nm}j_m = mj_n$, $j'J_{nn}j = n^2$. In this notation

$$H = J - E$$

The elements of the matrix $Q = E - H = 2E - J$ satisfy

$$q_{ij} = 1 \Leftrightarrow e_{ij} = 1 \Leftrightarrow h_{ij} = 0 \Leftrightarrow a_iRb_j \Leftrightarrow \neg\, a_i\overline{R}b_j$$
$$q_{ij} = -1 \Leftrightarrow e_{ij} = 0 \Leftrightarrow h_{ij} = 1 \Leftrightarrow a_i\overline{R}b_j \Leftrightarrow \neg\, a_iRb_j$$

Because Q is completely determined by E, it is just another scoring system for $\langle A,B,R\rangle$, but our discussion makes it possible to talk in terms of differences between relational systems. In some cases, however, it will be both useful and meaningful to use relations R and S, which do not necessarily exhaust A x B, and/or which are not mutually exclusive, i.e.

$$R \cup S \subset A \times B$$
$$R \cap S \neq \emptyset$$

A familiar example is the weak partial order $\succcurlyeq$. Define

$$e_{ij} = 1 \Leftrightarrow a_i \succcurlyeq b_j \text{ , otherwise } e_{ij} = 0$$
$$h_{ij} = 1 \Leftrightarrow b_j \succcurlyeq a_i \text{ , otherwise } h_{ij} = 0$$

Again, $q_{ij} = e_{ij} - h_{ij}$. We have four different possibilities:

1) $a_i > b_j \Leftrightarrow e_{ij} = 1,\ h_{ij} = 0 \Leftrightarrow q_{ij} = 1$

2) $a_i = b_j \Leftrightarrow e_{ij} = 1,\ h_{ij} = 1 \Rightarrow q_{ij} = 0$

3) $a_i < b_j \Leftrightarrow e_{ij} = 0,\ h_{ij} = 1 \Leftrightarrow q_{ij} = -1$

4) a_i and b_j incomparable $\Leftrightarrow e_{ij} = 0,\ h_{ij} = 0 \Rightarrow q_{ij} = 0$

These four mutually exclusive cases do indeed exhaust A x B. The relation $\geqslant$ is very important in psycholcgical measurement and scaling theory, because both preferential choice and similarity judgments can be considered as partial orders. Some applications of the generalized ICP-matrices will be presented in later chapters. Let D and G be diagonal nxn matrices, with $d_{ii} = \sum_j e_{ij}$ and $g_{ii} = \sum_j h_{ij}$. We assume that both d_{ii} and $g_{ii} \neq 0$ for all i. If $H + E = J$ then $G = mI - D$. Define $F = D^{-1}E - G^{-1}H$. Observe that $Fj = D^{-1}Ej - G^{-1}Hj = j - j = 0$, which implies that $F'F$ is a doubly centered matrix with all row and column totals equal to zero, and with determinant $\beta = 0$. Moreover, for all eigenvectors x of $F'F$, it is true that $x'j = 0$. The maximum rank of $F'F$ is $\min(n,m-1)$. If we postmultiply F by an m-element vector x, the element d_i of the resulting n-element vector $d = Fx$ can be interpreted as the difference between the mean of the x-values of those elements of B that are in the relation R to a_i, and the mean of those that are in the relation S to a_i.

-2.3-

The general procedure can be described as follows: maximize $d'd$, under the condition that $x'x = 1$. The Lagrangian function can be written as

$$V = x'F'Fx - \mu_1(x'x - 1)$$

Symbolic differentiation, and equating all partial derivatives to zero, gives the condition for an extreme value

$$(F'F - \mu_1 I)x = 0$$

By premultiplying this by x', we obtain

$$\mu_1 = x'F'Fx$$

This derivation shows, that x is the eigenvector associated with the dominant eigenvalue of the matrix $F'F$, and that the resulting value of V is indeed the maximum value of V. Because $F'F$ is a symmetric matrix of product sums, all characteristic roots are positive or zero ($F'F$ is singular, so at least one root actually equals zero). As was already demonstrated, $x'j = 0$. By the same reasoning a second solution y may be found, which maximizes $y'F'Fy$ under the conditions that $y'y = 1$, and $x'y = 0$. It turns out to be the eigenvector associated with the dominant eigenvalue of the matrix $F'F - \mu_1 xx'$, or, equivalently, the eigenvector of the matrix $F'F$, associated with the characteristic root which is second in magnitude. Proceeding in this way, we obtain a set of maximally $m-1$ linear independent, mutually orthogonal solutions, each one of them scaled in such a way that the sum of squares of the elements equals unity, and that the sum of the elements equals zero (which means that their variance equals $1/m$).

-2.4-

According to their definition, the eigenvalues may be considered as the distance between the relations R and S in the n dimensional space, defined by the elements of facet A. It is, however, possible to give a statistically more satisfactory interpretation. As was pointed out in section 2.2, the elements of the vector d are signed distances between sample means. Their significance can be tested by a host of parametric or non-parametric procedures. Because the variance of the eigenvectors is $1/m$, the value md_i^2 is a Mahalanobis-distance (if we make the appropriate distributional assumptions). This provides us with an alternative interpretation of μ_1, the quantity we maximize, because $m\mu_1/n$ turns out to be the average Mahalanobis-distance. If we include the second eigenvector, which defines a second d-vector, then the quantity $m(d_{1i}^2 + d_{2i}^2)$ is again a Mahalanobis-distance, and $m(\mu_1 + \mu_2)/n$ is the average of these distances over all elements of A. For each eigenvector the statistical hypotheses could be tested that the mean of the corresponding distribution of distances equals zero. If this hypothesis cannot be rejected, we stop factoring.

-2.5-

Some statistical tests will be discussed, that are appropriate on the assumption of (multi)normality. The first test assumes normality and equality of the variance-covariance matrices of the two distributions generated by each of the partitions defined by the rows of F. These assumptions will be approximately satisfied for all partitions, if there is sufficient agreement in the rows of F, and if there is a tendency to bimodality. The quantity

$$Q = \frac{d_{ii} + g_{ii} - k - 1}{k} \cdot \frac{d_{ii}g_{ii}}{d_{ii}+g_{ii}} \cdot \frac{D^2}{d_{ii}+g_{ii}-2}$$

is distributed as F, with $d_{ii}+g_{ii}-k-1$ and k degrees of freedom (Kendall & Stuart III, p 252,260). In this formula k is the number of dimensions (eigenvectors) under consideration. If k = 1 then

$$Q = \frac{d_{ii}g_{ii}}{d_{ii}+g_{ii}} D^2 = mr^2$$

where r is the product moment correlation coefficient of the relevant row of F with the eigenvector x.

-2.6-

On the assumption that the projections on the eigenvector have a normal distribution, we may test the hypothesis H_o: $\mu_i = \sigma$ by the likelihood ratio test described by Kendall & Stuart (II, p 231-32). The general criterion is given by

$$t_i = \tfrac{1}{2}m(\mu_i\sigma^{-1} - 1)^2$$

which is asymptotically distributed as χ^2 with one degree of freedom. In most cases the assumption in this section will be much too strong. As a matter of fact, it can be argued that an ideal eigenvector would show a two-point distribution (cf 2.14). From a psychometric point of view, an eigenvector is satisfactory to the degree in which the points are bimodally distributed on it. If the rows of F are very similar, then the distribution of the stimuli on the first eigenvector will tend to bimodality, while the remaining eigenvalues will be relatively low. Clearly, there is a reciprocal relationship between the appropriateness of the test discussed in the previous section and this test. If the χ^2-test is appropriate, the Q-test is not, and vice versa.

-2.7-

It is possible to find a test value σ for the χ^2-criterion in a more or less rational way. In the matrix F, the element f_{ij} equals $\frac{1}{d_{ii}}$ with probability $\frac{d_{ii}}{m}$, it equals $\frac{1}{g_{ii}}$ with probability $\frac{g_{ii}}{m}$, and it equals zero with probability $\frac{m-d_{ii}-g_{ii}}{m}$. This entails, that

$$\mathcal{E}(f_{ij}) = 0$$

$$\mathcal{E}(f_{ij}^2) = \frac{d_{ii}}{m}\cdot\frac{1}{d_{ii}^2} + \frac{g_{ii}}{m}\cdot\frac{1}{g_{ii}^2} = \frac{1}{m}\left(\frac{1}{d_{ii}} + \frac{1}{g_{ii}}\right) = \frac{d_{ii}+g_{ii}}{md_{ii}g_{ii}}$$

$$\mathcal{E}\left(\sum_i f_{ij}^2\right) = \sum_i \mathcal{E}(f_{ij}^2) = \sum_i \frac{d_{ii}+g_{ii}}{md_{ii}g_{ii}} =_{def} \sigma$$

$$\mathcal{E}\left(\sum_i f_{ij}f_{ik}\right) = \frac{1}{d_{ii}^2}\left(\frac{d_{ii}}{m}\right)^2 + \frac{1}{g_{ii}^2}\left(\frac{g_{ii}}{m}\right)^2 - 2\left(\frac{d_{ii}g_{ii}}{m^2}\right)\frac{1}{d_{ii}g_{ii}} = 0$$

If the elements of E are placed absolutely at random in the matrix, with only row totals fixed, it follows that

$$\mathcal{E}(F'F) = \sigma I$$

Moreover we have

$$\mathrm{Trace}(F'F) = \sum_i\sum_j f_{ij}^2 = \sum_i\left(\frac{d_{ii}+g_{ii}}{d_{ii}g_{ii}}\right)$$

which means that

$$\sigma = \frac{\mathrm{Trace}(F'F)}{m}$$

All eigenvalues of the diagonal matrix $\mathcal{E}(F'F)$ are, of course, equal to σ, and it seems a suitable test value to use in the χ^2-criterion of the previous section. Summarizing, on the assumption of random allocation of a fixed number of zeroes and ones in the rows of E, we have obtained

$$\sigma = \mathcal{E}(\mu_i) = \mathcal{E}\left(\sum_i f_{ij}^2\right) = \frac{\mathrm{Trace}(F'F)}{m} = \sum_i \frac{d_{ii}+g_{ii}}{md_{ii}g_{ii}}$$

Obviously

$$\sigma \leqslant \sum_i \frac{1}{d_{ii}g_{ii}} \leqslant \frac{4n}{m^2}$$

which implies that we always have the following inequality for the eigenvalues

μ_i of F'F:

$$0 \leqslant \mu_i \leqslant \text{Trace}(F'F) \leqslant \frac{4n}{m} \qquad \text{for all } i=1,\dots,m$$

-2.8-

On the assumption of normality, we may also test the significance of residuals with H_0: $\mu_k=\mu_{k+1}=\dots=\mu_m$ by a similar likelihood ratio test (Kendall & Stuart II, p 234-36). In this case the criterion is

$$t = \frac{P}{R}$$

with

$$P = ab \ln\left\{\frac{1}{ab} \sum_{i=k}^{m} \mu_i\right\} - \frac{1}{b} \sum_{i=k}^{m} \ln \frac{\mu_i}{b}$$

$$R = \frac{3am-2a+1}{3ab}$$

$$a = m-k+1$$
$$b = m-1$$

Again, t is distributed asymptotically as χ^2, in this case with m-k+1 degrees of freedom.

-2.9-

In this section we shall discuss an obvious generalization which does not complicate the algorithm. On the basis of a previous analysis, or a priori judgments of the investigator as to the relative importance of the elements of A, we may assign weights to the elements of d, i.e. we maximize $d'\Lambda d = x'F'\Lambda Fx$, where Λ is a diagonal matrix of weights ($\lambda_{ii} > 0$ for each $i=1,\dots,n$). For convenience we also require $j'\Lambda j = n$. Our interpretation of the vector d does not change, but the value $m\mu/n$ becomes a weighted mean of Mahalanobis-distances. It is easy to show that by taking

$$\lambda_{ii} = \frac{d_{ii}g_{ii}}{d_{ii}+g_{ii}}$$

we maximize the average squared product moment correlation coefficient between the rows of F and x (cf 2.5, 2.7, 2.14).

-2.10-

At this stage of the analysis we have a q-dimensional quantification of the set B, and a set of Mahalanobis-distances for each of the elements a_i in respectively 1,...,q dimensions. Because the transpose of an ICP-matrix is clearly another ICP-matrix, we may apply the same analysis to E', resulting in a p-dimensional quantification of A and a set of Mahalanobis-distances for each of the m elements of B. There seems to be no simple relation between the two sets of eigenvectors and eigenvalues, which is due to the fact that the relation between the F-matrices of E and E' is not at all obvious. In the other chapters some cases will be discussed in which it is desirable to quantify the set A in a completely different way.

-2.11-

The matrix F can be written as $F = K\Psi L$. Let us suppose that F is of rank r. Then K is the nxr matrix of left eigenvectors, satisfying $K'K = I$, L is the rxm matrix of right eigenvectors, satisfying $LL' = I$, and Ψ is the rxr diagonal matrix of latent roots. Evidently, $F'F = L'\Psi K'K\Psi L$. By finding the eigenvectors and eigenvalues of $F'F = L'\Psi^2 L$, we solve for L and Ψ^2, i.e. $\mu_i = \psi_{ii}^2$, or $M = \Psi^2$. The signed distances on the various dimensions can be written as the nxr matrix $D = FL' = K\Psi LL' = K\Psi$. It follows that $D'D = \Psi K'K\Psi = M$, and that $DD' = K\Psi^2 K' = K\Psi LL'\Psi K' = FF'$. It is clear that the columns of D are proportional to the left eigenvectors of F, i.e. proportional to the eigenvectors of FF'. More explicitly, $K = D\Psi^{-1} = DM^{-\frac{1}{2}}$. Moreover the columns of D are orthogonal by pairs, which means that in general if $\mu_j > 0$ then some of the elements of the j^{th} column of D will be negative and some will be positive. In the notation of the previous sections: for some a_i it will be the case that $a_i R b_j$ if x_j is relatively high, and for other a_i this will be the case if x_j is relatively low. In the context of preference judgments (cf chapter III), we may say that a high value on a dimension may imply popularity for one subject and impopularity for another subject. Another fact to be noted is that although we also find the left eigenvectors of F, they must not be interpreted as the optimal scores for the elements of A. Finding the eigenvectors of FF' is not equivalent to finding the eigenvectors of $\tilde{F}'\tilde{F}$, where $\tilde{F}$ is the F-matrix of the transposed ICP-matrix E'.

-2.12-

In our technique the first vector of signed distances can be positive, but this implies that the second vector has at least one negative element. In some situations it may be reasonable to insist on positive signed distances only. In these cases only the first dimension is meaningful, which entails that μ_1 must be sufficiently high. An alternative technique to obtain a one-dimensional positive quantification is to maximize $j'd = j'Fx$, under the condition that $x'x = 1$. The Lagrangian function can be written as

$$\Phi = j'Fx - \tfrac{1}{2}\mu(x'x - 1)$$

The maximizing solution is given by $x_{max} \doteq F'j$, or $x_{max} = F'j(j'FF'j)^{-\frac{1}{2}}$. Substitution gives the maximum value

$$\Phi(x_{max}) = \mu = j'FF'j(j'FF'j)^{-\frac{1}{2}} = (j'FF'j)^{\frac{1}{2}}$$

With this alternative technique we maximize the sum of the differences, which means that their sign is essential (cf the techniques of simple and weighted summation in factor analysis). It is easy to prove that there is no vector y, such that $j'Fy$ is a maximum, $x'y = 0$, and $y'y = 1$. This vector should satisfy $y \doteq F'j - \mu x$ (by the extremum and orthogonality conditions), but $F'j - \mu x = 0$, which entails that $y = 0$, contrary to the requirement that $y'y = 1$. Multi-dimensional quantification is impossible, at least if we impose the conventional orthogonality restraints. The same conclusion would have been reached, if we had maximized $(j'Fx)^2 = x'F'JFx$, because the matrix J is of rank one.

-2.13-

In this section the technique of Guttman, which also operates directly on the ICP-matrix E, will be discussed. An example of such an E-matrix is:

$$\begin{Vmatrix} 1 & 0 & 0 & 1 & 0 & 1 & . & . & . & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & . & . & . & 0 & 1 \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . \\ 1 & 1 & 1 & 0 & 0 & 0 & . & . & . & 1 & 0 \end{Vmatrix}$$

in which the columns correspond for example with a set B of subjects and the rows with a set A of items. The scores for the elements of facet A (which are to be found) can be collected in the diagonal matrix X. The matrix $XE = M_x$ can be written as

$$\left\|\begin{matrix} x_1 & 0 & 0 & x_1 & 0 & x_1 & . & . & . & 0 & x_1 \\ x_2 & 0 & x_2 & 0 & 0 & x_2 & . & . & . & 0 & x_2 \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . \\ x_n & x_n & x_n & 0 & 0 & 0 & . & . & . & x_n & 0 \end{matrix}\right\|$$

Guttman observes that we can think of a similar matrix $\overline{M}_x$, which is just the opposite of M_x, i.e.

$$\left\|\begin{matrix} 0 & x_1 & x_1 & 0 & x_1 & 0 & . & . & . & x_1 & 0 \\ 0 & x_2 & 0 & x_2 & x_2 & 0 & . & . & . & x_2 & 0 \\ . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & x_n & x_n & x_n & . & . & . & 0 & x_n \end{matrix}\right\|$$

"We can best differentiate between what an individual does and what he does not do by maximizing the difference between these two distributions. The difference will tend to be maximized for all individuals if the relative variability in the columns of M_x is minimized." (Guttman 1941b, p 326). This relative variability can be expressed as a correlation ratio

$$\eta_x^2 = \frac{x'EE'x}{mx'Dx}$$

The quantification of A is obtained by solving for x in such a way that this correlation ratio is maximized. The scores for the set B can be obtained by maximizing

$$\eta_y^2 = \frac{y'E'D^{-1}Ey}{my'y}$$

For the technical details we refer the reader to Guttman's paper. The maximization problem reduces to finding eigenvalues and eigenvectors of the matrices $D^{-\frac{1}{2}}EE'D^{-\frac{1}{2}}$ and $E'D^{-1}E$. Evidently, the eigenvalues of these matrices are the same. Guttman shows that the first eigenvector, with a corresponding correlation ratio of unity, is a pseudo-solution of the determinantal equation, which means that the quantification is obtained by taking the second eigenvector of these matrices. The value of $\eta_x^2 = \eta_y^2$ is an index of the scalability of the data structure under consideration. As an alternative, Guttman investigates the maximization of a product moment correlation coefficient

$$\rho = \frac{x'Ey}{\sqrt{mx'Dxy'y}}$$

and he shows that the solutions for x and y are identical with those that maximize η_x^2 and η_y^2. Moreover $\eta_x^2 = \eta_y^2 = \rho^2$. These considerations show, that Guttman's technique is closely related to a procedure proposed by Fisher (1954, 289-298) for analyzing contingency tables, which maximizes the same quantity ρ^2 An extensive analysis of this technique is given in Kendall & Stuart (II, p 568-575), where the relationships between this 'canonical analysis of contingency tables' and chi-square theory are discussed: the sum of the squared canonical correlations is equal to Pearson's chi-square for the contingency table under consideration. A theorem of Lancaster (1957) proves that the procedure is essentially equivalent to operating on the margins of the table in order to produce a bivariate normal distribution. Lingoes (1963b, 196-) adapted the Guttman technique for computer usage, called it multivariate analysis of contingencies or MAC, and gave some multidimensional examples. The technique of Burt (1950) is simply to factorize $D^{-\frac{1}{2}}EE'D^{-\frac{1}{2}}$. Burt does not give an interpretation of the method in terms of maximizing a correlation ratio or coefficient, but he argues that it is perfectly legitimate to factorize symmetric matrices, which have the property that all their principal minors are non-negative.

-2.14-

The principal difference between the methods of Guttman, Burt and Fisher on one side, and our method on the other, is that the former methods concentrate on minimizing the variability of the distributions, while our method maximizes the distance between their centroids. As a matter of fact, if our purpose is to maximize the difference between what a subject (c.q. an element of A) does, and what he (it) does not do, it seems more logical to maximize the distance.

Our method seems to give more information (notably the discrimination indices in d, which are also meaningful in the case of multidimensional quantifications). There is, however, an interesting relationship between maximum distance and minimum variability. By the Cauchy-Schwartz inequality we have

$$\left[\sum_j f_{ij}x_j\right]^2 \leqslant \sum_j f_{ij}^2 \sum_j x_j^2 = \sum_j f_{ij}^2 = \frac{d_{ii}+g_{ii}}{d_{ii}g_{ii}}$$

This implies that by maximizing the difference between the elements of B that are in the relation R to a particular element of A and those that are not, we minimize at the same time the variability in the two groups. For a particular row of F, the ideal eigenvector would have a two-point distribution with

$$x_i = \frac{1}{d_{kk}}\sqrt{\frac{d_{kk}g_{kk}}{d_{kk}+g_{kk}}} \qquad \text{for } e_{ki} = 1$$

$$x_i = \frac{-1}{g_{kk}}\sqrt{\frac{d_{kk}g_{kk}}{d_{kk}+g_{kk}}} \qquad \text{for } e_{ki} = 0$$

-2.15-

Recently, Guttman & Lingoes have developed a multidimensional generalisation of scalogram analysis, called MSA. The elements of B are mapped into euclidean p-dimensional space Re^p in such a way, that each element of A defines a partitioning of Re^p. If the quantification of b_j is the ordered p-tuple of real numbers (coordinate values) $x_j = (x_{j1},\dots,x_{jp})$, then there must be a partitioning of Re^p into two subsets Re^p_{i+} and Re^p_{i-} in such a way, that

$$b_j \in Re^p_{i+} \Longleftrightarrow a_iRb_j$$

$$b_j \in Re^p_{i-} \Longleftrightarrow a_i\bar{R}b_j$$

$$Re^p_{i+} \cup Re^p_{i-} = Re^p$$

$$Re^p_{i+} \cap Re^p_{i-} = \emptyset$$

In the case of manifold classifications, with more than two mutually exclusive categories, the elements of A must partition Re^p into more than two regions. This is accomplished by maximizing Guttman's coefficient of contiguity λ with an iterative average steepest descend procedure. For technical details, see Lingoes (196-). The output of the program is the p-dimensional quantification of B, and the boundaries, which define the partitioning, can be drawn in by hand. It is clear, that MAC is a 'metric' approximation to the MSA-requirements, because minimum variability in the regions will make it more probable that simple partitioning boundaries can be found. As a matter of fact, the output of MAC is used as the initial configuration for the MSA iterations. If this is done, in most cases only a few iterations are sufficient to maximize λ. Our technique may be considered as a more direct approximation to the MSA requirements, because maximum distance between the points in the regions also means that relatively simple partitioning boundaries can be drawn.

+++++++++++

Chapter III

NONMETRIC FACTOR ANALYSIS

-3.1-

In this chapter the primitives are two finite sets A and B, and a ternary relation R. They define a relational system $\langle A,A,B,R\rangle$, which is studied for example in the quantification of paired comparisons and preference rank orders. Define E and H as the mxmxn three-way matrices, with

$$e_{ijk} = 1 \Longleftrightarrow R(a_i,a_j,b_k) \quad \text{otherwise } e_{ijk} = 0$$
$$h_{ijk} = 1 \Longleftrightarrow S(a_i,a_j,b_k) \quad \text{otherwise } h_{ijk} = 0$$

Because three-way matrices are somewhat difficult to work with, and in order to preserve the continuity of this paper, it seems preferable to define C = A x A, and to study the relational system $\langle C,B,R\rangle$. Define $l = \frac{1}{2}(i+j-1)(i+j-2) + i$, and

$$e_{kl} = 1 \Longleftrightarrow b_k R(a_i,a_j) \quad \text{otherwise } e_{kl} = 0$$
$$h_{kl} = 1 \Longleftrightarrow b_k S(a_i,a_j) \quad \text{otherwise } h_{kl} = 0$$

Both E and H are nxm^2 matrices. The interpretation might run as follows: $e_{kl} = 1$ iff subject k prefers stimulus i to stimulus j. The relation with the dominance matrix, defined in 1.3, is evident. From the E and H matrices an nxm^2 F-matrix may be constructed in exactly the same way as in the previous chapter. By using this matrix in the familiar way, the set A x A can be quantified by an m^2-element vector x, that maximizes $x'F'Fx$. In most cases, however, we do not want a quantification of A x A, but a quantification of A. To obtain this, observe that the d_i-values in the previous chapter are related to the point biserial correlations between the i^{th} row of E and x. From the theory of generalized correlation coefficients, we know that the value

$$r_k = \sum_{l=1}^{m^2} (e_{kl}-h_{kl})(x_i-x_j)$$

if properly scaled, is the relevant correlation coefficient for this case. Translating back to the distance model: instead of maximizing the difference between means of scale values, maximize the difference between means of differences of scale values. Define the $m^2 xm$ matrix S as

$$s_{lv} = \delta^{iv} - \delta^{jv}$$

where δ is the Kronecker delta, and $c_l = (a_i, a_j)$. This means that the vector $w = Sx$ has as element w_l the difference between x_i and x_j. Obviously the n-element vector $d = FSx$ has the desired properties, Moreover, from the definition of S it follows that

$$j'Sx = mj'x - mj'x = 0$$

$$x'S'Sx = 2x'(mI - J_{mm})x = 2m^2\sigma_x^2$$

This means that the mean of the vector w of differences is zero, while the variance of the m^2 elements is two times the variance of x. Maximize $d'd = x'S'F'FSx$, under the condition that $x'x = 1$. Again the matrix of scalar products is doubly centered, which implies that the variance of x is equal to $\frac{1}{m}$. The value $\frac{1}{2}md_i^2$ is again a Mahalanobis distance between the means of the two distributions of differences. The quantification of facet A proceeds in exactly the same way as in the previous chapter, with $S'F'Fs$ substituted for $F'F$.

-3.2-

If R is a consistent partial order, it is possible to take $C \subset A \times A$, with

$$(a_i, a_j) \in C \Leftrightarrow i < j$$

This makes E a matrix with $\frac{1}{2}m(m-1)$ columns, and S a matrix with $\frac{1}{2}m(m-1)$ rows. It actually means, that only the upper diagonal triangle of the dominance matrix is used. The only effect on the analysis is, that the Mahalanobis distances are the values md_i^2, because the variance of the $\frac{1}{2}m(m-1)$ differences is equal to the variance of x. The matrix E - H must however be scaled, as if the analysis was performed over the complete set of m^2 differences. Although the method for quantifying A is more complicated in this case, we may use the method of the previous chapter to quantify the set B.

-3.3-

A related technique for the analysis of preference rank orders was proposed by Slater (1960). Suppose the rank orders are collected in the matrix Z, where z_{ij} is the rank number of stimulus j in the preference rank order of subject i. The row sums of this matrix are all equal to $\frac{1}{2}m(m+1)$ which means that the rows can be centered by subtracting $\frac{1}{2}(m+1)$ from all elements of Z. This defines a matrix Y. We want to quantify the set of stimuli in such a way, that the quantity

$$r_k = \sum_i\sum_j (y_{ki}-y_{kj})(x_i-x_j)$$

is maximized. If we scale the rows of Y and the score vector x, this quantity is equal to the product moment correlation coefficient between the rank numbers in the k^{th} row of Y and the lements of x. For obvious reasons, we may also call this procedure Spearman-weighting of differences, while our technique may be called Kendall-weighting (cf Kendall 1962, p 22). Our technique has some important advantages. It makes it possible to analyze general partial orders and even 'partial orders' with inconsistencies. Moreover, our distance coefficient (which is related to the point biserial correlation coefficient) is theoretically more satisfactory than the ordinary product moment correlation coefficient (which assumes interval scale properties of rank numbers). Slater's procedure is the maximization of $r'r = x'Y'Yx$, which means that his eigenvalues are averaged product moment correlation coefficients. The subjects are quantified by the standard vector model methods, i.e. $Y = K\Psi L$, and $K = YL'\Psi^{-1}$. This procedure for assigning subject scores is, of course, entirely different from the technique outlined in chapter II. Its only justification is that it makes it possible to reproduce the original matrix Y numerically, but this is relatively unimportant, because the elements of Y are assumed to have only ordinal properties. Our mathod also makes it possible to reproduce the generalized ICP-matrix of the system, but this procedure is essentially independent of the quantification for subjects (cf section 2.11). A weaker analogue of Slater's procedure is non-metric factor analysis in the sense of Shepard & Kruskal (1964), in which Kruskal's approach to nonmetric scaling problems is extended to the vector model for preference rank orders (Tucker 1955).

-3.4-

It is possible to prove even closer relationships between Slater's technique and the one discussed in this paper. For a consistent ordening represented in the mxm matrix $A = \{a_{ij}\} = \{\mathrm{sign}(y_i - y_j)\}$, we have

$$\sum_i\sum_j a_{ij}(x_i - x_j) = \sum_i\sum_j a_{ij}x_i - \sum_i\sum_j a_{ij}x_j = \sum_i x_i \sum_j a_{ij} - \sum_i x_j \sum_j a_{ij} =$$

$$\sum_i x_i \sum_j a_{ij} - \sum_i x_i \sum_j a_{ji} = 2\sum_i x_i \sum_j a_{ij}$$

The quantities $\sum_j a_{ij}$ are related in a very simple way to the rank numbers $r(y_i)$. In the case of a linear order, without ties, we have

$$2\, r(y_i) = m - \sum_j a_{ij} + 1$$

or

$$-\frac{1}{2}\sum_j a_{ij} = r(y_i) - \frac{1}{2}(m+1) = r(y_i) - \frac{1}{m}\sum_k r(y_k)$$

The close relationships with Slater's technique are obvious (cf also Kendall, 1962).

-3.5-

In derived measurement the most common relational system is $\langle A, B, Re, \psi \rangle$, where A and B are finite sets, Re is a subset of the reals, and ψ maps A x B into Re. An important example is the relational system in which A is a set of n tests, B a set of m subjects, and $\psi(a_i, b_j)$ is the score of subject j on test i. Define the nxm matrix Y, with $y_{ij} = \psi(a_i, b_j)$, and suppose the y_{ij} are scaled in such a way, that $\sum_j y_{ij} = 0$, and $\sum_j y_{ij}^2 = 1$. If x is a quantification of the subjects, with $\sum_j x_j = 0$, and $\sum_j x_j^2 = 1$, then the product moment correlations between the tests and the vector x are given by $R = Yx$. Maximizing $\mu_1 = r'r = x'Y'Yx$ is the common factor analytic technique, which gives us a quantification of the subjects in terms of maximally m-1 orthogonal principal components. Moreover we obtain the matrix with the values of r for each of the tests and each of the components (the factor loadings). In 3.3 we already discussed a particular case in which $Re = N_m^+$, the set of the first m positive natural numbers. This reconstruction of factor analysis does not even mention the correlation matrix. The communality dilemma does not seem to arise in a natural way, although we may of course maximize $\mu_1 = r'\Lambda r = x'Y'\Lambda Yx$, where Λ is a diagonal matrix of weights (for example reliability coefficients). Because the trace of the matrix $X'X$ is invariant under rotations, the rotational problem is an exact analogue of that in factor analysis. The principal components are, of course, optimal in the familiar mathematical sense. The generalized correlation coefficient in this case is obtained by what may be called Pearson-weighting, because it is evidently nothing more or less than Pearson's method of principal axes (Pearson 1901). Up to now we have discussed three different weighting methods. In all cases $\Sigma\, r_k^2$ is maximized. In the case of Pearson-weighting of differences, r_k is defined as

$$r_k \doteq \sum_i \sum_j (y_{ki}-y_{kj})(x_i-x_j)$$

where y_{ki} is the score of b_i on the scale defined by element a_k. In the case of Spearman-weighting, we have a similar formula, but y_{ki} is defined as the rank number of b_i on the scale defined by a_k. In the case of Kendall-weighting of differences, we have

$$r_k \doteq \sum_i \sum_j \operatorname{sign}(y_{ki}-y_{kj})(x_i-x_j) = \sum_i \sum_j f_{ijk}(x_i-x_j)$$

And this is the approach used throughout this chapter. Clearly, this approach is feasible for all data matrices which can be analyzed by the Eckart-Young-Tucker techniques, for example the Tucker & Messick (1963) program for quantifying distances between stimuli (cf 3.7), the Tucker (1966) program for analyzing learning curves (cf 5.7), and ordinary factor analytic procedures. The fundamental differences are the non-metric treatment of the data in our approach, and the approach to quantify the set A, which we do by the methods of chapter II. In the case of standardized variates in Y, the scores on different tests of one particular subject can be meaningfully compared, which means that the method outlined in this chapter may be applied to quantify the set of tests.

-3.6-

It is also possible to apply the technique of this chapter to the ICP-matrices of the previous chapter. The rows of those ICP-matrices can be interpreted as (incomplete) I-scales, that give ordinal information about the positions of the elements of B on the scales defined by the elements of A. In this section we shall discuss a remarkable finding, which will emphasize the close relationship of the two techniques. We shall apply the nonmetric factor analysis of this chapter to the matrix E, composed of zeroes and ones as in chapter II. In the first place, observe that $\sum_j \sum_k |\mathrm{sign}(e_{ij}-e_{ik})|$ can be evaluated by a simple combinatorial argument. The only two relevant cases are $e_{ij} = 1$ and $e_{ik} = 0$, which occurs $d_{ii}g_{ii}$ times, and the opposite case $e_{ij} = 0$ and $e_{ik} = 1$, which also occurs $d_{ii}g_{ii}$ times. It follows that $\sum_i \sum_j |\mathrm{sign}(e_{ij} - e_{ik})| = 2d_{ii}g_{ii}$. Another relevant quantity is $\sum_j \mathrm{sign}(e_{ij}-e_{ik})$. If $e_{ik} = 0$ then evidently this quantity is equal to d_{ii}. If $e_{ik} = 1$, it is equal to $-g_{ii}$, which implies that in general it equals $(1-e_{ik})d_{ii}-g_{ii}e_{ik}$. By a similar argument $\sum_k \mathrm{sign}(e_{ij}-e_{ik})$ $= g_{ii}e_{ij}-d_{ii}(1-e_{ij})$. Define H = FS, then clearly $2d_{ii}g_{ii}h_{il} =$ $\sum_j \sum_k \mathrm{sign}(e_{ij}-e_{ik})(\delta^{jl}-\delta^{kl})$, where the δ's are the Kronecker operators. This can be simplified in the following way:

$$
\begin{aligned}
2d_{ii}g_{ii}h_{il} &= \sum_j \sum_k \mathrm{sign}(e_{ij}-e_{ik})\delta^{jl} - \sum_j \sum_k \mathrm{sign}(e_{ij}-e_{ik})\delta^{kl} = \\
&= \sum_k \mathrm{sign}(e_{il}-e_{ik}) - \sum_j \mathrm{sign}(e_{ij}-e_{il}) = \\
&= g_{ii}e_{il} - d_{ii}(1-e_{il}) - (1-e_{il})d_{ii} + g_{ii}e_{il}
\end{aligned}
$$

Evidently

$$h_{il} = \frac{(g_{ii}+d_{ii})e_{il} - d_{ii}}{d_{ii}g_{ii}}$$

This is equal to the element f_{il} of the F-matrix constructed by the method of the previous chapter. In that case $f_{il} = d_{ii}^{-1}$ if $e_{il} = 1$, and $f_{il} = -g_{ii}^{-1}$ if $e_{il} = 0$. Combining, we obtain

$$f_{il} = \frac{e_{il}}{d_{ii}} - \frac{(1-e_{il})}{g_{ii}} = \frac{(g_{ii}+d_{ii})e_{il} - d_{ii}}{d_{ii}g_{ii}} = h_{il}$$

Evidently this entails that the matrices of scalar products of both methods (respectively $F'F$ and $H'H = S'F'FS$) are equal, that their eigenvalues and eigenvectors are identical, and that the signed distances Hx and Fx are identical as well. The two methods, when applied to data matrices of a particular type, give exactly identical results. From a computational point of view, it would not even have been necessary to discuss the methods of chapter II.

-3.7-

In the same way as in the previous sections, the relational system $\langle A,A,A,A,B,R\rangle$ can be portrayed in the E-matrix of $\langle C,B,R\rangle$, where $C = A \times A \times A \times A$, i.e. the elements of C are ordered quadruples of elements of A. The interpretation of $R(a_i,a_j,a_k,a_l,b_r)$ might run as follows: subject r responds, that the dissimilarity between stimulus a_i and stimulus a_j is greater than the dissimilarity between stimulus a_k and stimulus a_l. The E-matrix is an nxt matrix, with $t \leq m^4$. In exactly the same way as before the set $A \times A$ can be quantified. If R is a consistent partial order, it is possible to reduce the system to one with fewer elements, which may be useful from the point of view of a computer programmer (in most cases the number of elements in E will be 10^5 or more). Define $Y \subset A \times A$, with $(a_i,a_j) \in Y$ iff $i < j$, and define V as a subset of $Y \times Y$, with $(y_p,y_q) \in V$ iff $p < q$. The ICP-matrix of the system $\langle V,B,R\rangle$ will have only $\frac{1}{8}m(m-1)(m+1)(m-2)$ columns. This means a reduction of the required memory space of roughly 90% or more. By defining the matrix S as in the previous chapter, and by maximizing $x'S'F'FSx$, a quantification of $Y \subset A \times A$ is obtained, i.e. a vector of $\binom{n}{2}$ differences between stimuli. The resulting p orthogonal components (or linear combinations of them) may be used as input for standard multidimensional scaling programs. The set B can be quantified by the method discussed in chapter II.

-3.8-

As in the previous sections, it would be even more satisfactory, if we had a method to quantify A directly from the ICP-matrix of $\langle V,B,R\rangle$. A straightforward generalisation of the procedure already used, gives us the correlation coefficient

$$\rho_r = \sum_i \sum_j \sum_k \sum_l f_{ijklr} \left\{(x_i - x_j) - (x_k - x_l)\right\}$$

By defining a slightly more complicated S-matrix, this quantity can be maximized quite simply for all r. This means that we maximize the difference between means of differences of differences of scale values. Unfortunately, the differences $(x_i - x_j)$ and $(x_k - x_l)$ are signed distances, while the similarity judgments of the subjects are judgments about absolute distances, i.e. the quantity that must be maximized is really

$$\rho_r = \Sigma \; \Sigma \; \Sigma \; \Sigma \; f_{ijklr} \left\{ |x_i - x_j| - |x_k - x_l| \right\}$$

There seems to be no way to write this formula as a quadratic form. That the standard method gives quite undesirable results can be shown by a simple example. Suppose the dissimilarity orderings from three different subjects are respectively $\delta_{12} > \delta_{13} > \delta_{23}$; $\delta_{12} > \delta_{23} > \delta_{13}$; $\delta_{13} > \delta_{12} > \delta_{23}$. The matrix that should be factorized is

$$\begin{Vmatrix} 8 & -12 & 4 \\ -12 & 24 & -12 \\ 4 & -12 & 8 \end{Vmatrix}$$

and the two principal components are

	I	II
1	$\sqrt{6}$	$\sqrt{2}$
2	$-\sqrt{24}$	0
3	$\sqrt{6}$	$-\sqrt{2}$
μ	36	4

For the distances in this two-space, we have $d_{12} = d_{23} > d_{13}$, which is clearly unsatisfactory, if we compare it with the orderings of the dissimilarities. Nevertheless, the technique may work as a scaling model in other situations.

-3.9-

Suppose that the elements of the k^{th} row of F are written in the diagonal of the matrix Q, and that the non-diagonal elements of Q are zero. Define matrices S and T in such a way, that the vectors Sx and Tx contain the elements $(x_i-x_j) - (x_k-x_l)$ and $(x_i-x_j) + (x_k-x_l)$ on the appropriate places. Then

$$\rho_r = \sum_i \sum_j \sum_k \sum_l f_{ijklr} \left\{(x_i-x_j)^2 - (x_k-x_l)^2\right\} = x'T'QSx$$

This seems to provide a solution to the problem of quantifying the set A for subject r. Observe, however, that the square matrix T'QS is in general not symmetric. Let us consider a simple example. Suppose that subject r orders the dissimilarities as $\delta_{12} > \delta_{13} > \delta_{23}$. The matrix T'QS is seen to be

$$\begin{Vmatrix} 4 & -10 & 6 \\ 2 & 0 & -2 \\ -6 & 10 & -4 \end{Vmatrix}$$

The non-zero roots of the characteristic polynomial of this matrix are $\pm i\sqrt{60}$, which can hardly be called satisfactory. The only way out seems to be maximizing

$$\rho_r = \sum_i \sum_j \sum_k \sum_l f_{ijklr}(d_{ij}-d_{kl})$$

by using some iterative procedure (cf De Leeuw 196- a). Moreover, with such a procedure the solution can be generalized to a maximization of $\sum_r \rho_r^2$, and to non-euclidean distance measures.

-3.10-

There is, however, an important example of a relational system in which more complicated S-matrices can be defined, without any danger of confusion, and without any problems concerning absolute or signed distances. In additive conjoint measurement (ACM), we study the relational system $\langle A_1,\ldots,A_N,\succsim_o\rangle$, where $\succsim_o$ is a partial order defined on $A = \prod A_i$, i.e. $\succsim_o \subseteq A \times A$. The columns of F correspond with the elements of A x A, the rows with different orderings of the same set (for different subjects), and the rows of the S-matrix consist of arrangements of 0, -1 and +1 in such a way, that Sx is the appropriate column of sums and differences of the elements of x. By maximizing $\mu_1 = x'S'F'FSx$ the best additive representation in this particular least squares sense is found. In this case the vector contains the quantifications of $A_1,\ldots,A_N$ as subvectors $x_1,\ldots,x_N$. The goodness of fit of the model can bejudged by investigating the ordinal relationships between the derived elements of Y, the quantification of the cartesian product A, which can be obtained by summing the relevant elements of x. To the measurement theoretician this procedure may seem somewhat perverse, since it is not concerned with the explicit solvability of the relevant system of linear inequalities, because inconsistent partial orders can be analyzed with it equally well, because it does not pay any attention to the question whether the necessary and/or sufficient conditions for ACM are met, and because it analyzes several different data structures at the same time. For a discussion of these and related criticisms, see chapter VI.

-3.11-

Suppose U is a given nxm matrix in which u_{ij} is the value of stimulus i on dimension j, and suppose Y is an nxp matrix of dependent variables (for example responses of p subjects). In this case we have more information than in the ACM-case treated in the previous section and we can apply a more restricted model, called nonmetric multiple prediction (NMP). In NMP the transformations of the independent variables are dictated by the measurement model, but the dependent variables are transformed monotonically. In the linear case, we define

$$\rho_r = \sum_i \sum_j \mathrm{sign}(y_{ir}-y_{jr}) \left\{ \sum_k u_{ik}x_k - \sum_k u_{jk}x_k \right\}$$

and we maximize

$$\mu_1 = \sum_r \rho_r^2$$

under the condition that $x'U'Ux = 1$. By transforming the matrix U in such a way, that $U'U = I$ before computing $U'S'F'FSU$, the matrix to be factorized, this condition reduces to $x'x = 1$. An interesting application of this linear case of NMP is its use in simultaneous nonmetric multiple linear regression, in which x is the vector of optimal regression weights for all p subjects simultaneously (after an optimal monotone transformation of their responses).

Another application is the nonmetric analogue of the general Procrustus problem: rotate a matrix U in such a way that it maximally resembles a target matrix Y, if only the ordinal relations within each dimension of Y are given. Evidently, the optimal rotation matrix can be found by computing the eigenvectors of U'S'F'FSU. The approach can be generalized to ar additive combination of q arbitrary functions of the stimulus characteristics

$$\rho_r = \sum_i \sum_j \operatorname{sign}(y_i - y_j) \left[\left\{ \sum_k^q f_k(u_{i1}, \ldots, u_{im}) - f_k(u_{j1}, \ldots, u_{jm}) \right\} x_k \right]$$

An important special case is the nonmetric analogue of polynomial regression analysis. Suppose b is an n-element vector, and $u_{ij} = b_i^{j-1}$. By transforming U in such a way that $U'U = I$, we reduce the problem to one with the familiar orthogonal polynomials of least squares.

+++++++++++

Chapter IV

MULTIWAY MATRICES AND MULTIRELATIONAL SYSTEMS

-4.1-

The data structures analyzed in the previous chapter actually are special cases of multiway matrices, with the restriction that the cartesian products are constructed out of only two different sets. A more general, and analytically more simple, three-way situation arises, if we have N different nxm matrices E_k. Suppose, for example, that A is a set of subjects, B a set of attributes and C a set of photographs of facial expressions.

$$A = \{a_1, \dots, a_n\}$$

$$B = \{b_1, \dots, b_m\}$$

$$C = \{c_1, \dots, c_N\}$$

We may construct a three-way nxmxN matrix E, with $e_{ijK} = 1$ iff subject a_i thinks that photograph c_K has attribute b_j (i.e. gay, angry, anxious, etc). This means that R is a ternary relation, and that we are stuck with a rather akward three-way array. Suppose we want to quantify the set of subjects. Then our data may also be interpreted as N different nxm matrices E_K, with N different binary relations R_K. Each transposed ICP-matrix defines a F_K-matrix, maximize $\sum_K x'F_K'F_Kx$, or, equivalently, $x'F'Fx$, where F is the mNxn supermatrix

$$\begin{Vmatrix} F_1 \\ F_2 \\ \cdot \\ \cdot \\ \cdot \\ F_N \end{Vmatrix}$$

It is reasonable to require one single quantification for subjects over all photographs and attributes. But in the case of quantifying photographs or attributes a different cognitive space for each of the subjects may be desirable. In that case simply analyze each of the n different mxN matrices by the mathods outlined in chapter II. The number of possibilities increases with the number of facets, not unlike the interaction terms in the analysis of variance, the partitionings of chi square, and the number of components in multivariate information transmission analysis.

-4.2-

Another important, but even more complicated case is the semantic differential. The complications arise mainly from the fact, that the responses are given on a seven point rating scale, which means that they have ordinal properties (although they most certainly do not have interval scale properties, as is implicitly assumed in the factor analytic procedures of Osgood & Suci). Matters would be much more simple if the rating scale was a three point scale (i.e. yes - don't know - no). In that case an extended ICP-matrix could be defined, and the resulting three-way matrix could be analyzed by the method outlined in the previous section. As a matter of fact, there are other advantages of three-point scales, beside the practical considerations discussed above (Tversky 1964b). It is, however, possible to quantify data obtained from a set of N subjects that rate m concepts on a set of n different 1-point scales by the three-way analogue of nonmetric factor analysis. To quantify the concepts for example, we define

$$\left\Vert \begin{matrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_N \end{matrix} \right\Vert$$

In this nNxm supermatrix all relevant information is collected, and we can transform it into an nNx $\frac{1}{2}m(m-1)$ F-matrix, that can be treated by the methods of chapter III. Alternatively, we can define an mNxn or an mnxN supermatrix X for quantifying the facet 'scales' and the facet 'subjects' respectively. We can analyze N different mxn matrices or m different nxN matrices or n different mxN matrices. The interpretation of these cases is obvious. Observe that in the case of an analysis of supermatrices such as X or F in the previous section, we may obtain information about the relative scalability of the submatrices X_K and F_K by a partitioning of the eigenvalues: the sum of squares of the signed distances in d can be computed for each submatrix separately.

-4.3-

Suppose that the set A can be partitioned into N mutually exclusive and exhaustive subcategories:

$$A = \bigcup_{K=1}^{N} A_K$$

$$A_K \cap A_J = \emptyset \qquad \text{(for } K = J\text{)}$$

$$A_K = \left\{ a_{K1}, \ldots, a_{Kn_K} \right\}$$

Evidently $\sum_{K=1}^{N} n_K = n$. Define the set

$$T_{Kl} = \left\{ x \,\middle|\, x \in B \;\; \& \;\; a_{Kl} R x \right\}$$

or: T_{Kl} is the subset of B, whose elements are in the relation R to a_{Kl}. The sets T_{Kl} obey the relationships

$$T_{Kj} \cap T_{Kk} = \emptyset \qquad \text{(for } k = j\text{)}$$

$$\bigcup_{l=1}^{n_K} T_{Kl} = B$$

In the context of test theory the set A might be interpreted as a test with N items. Item A_K has n_K mutually exclusive response categories. T_{Kl} is the set of those subjects, whose response to item K was alternative l. It is, of course, possible to quantify the set B by operating on the ICP-matrix in exactly the same way as in chapter II. This means that all $n = \sum_K n_K$ elements a_{Kj} are treated in the same way, as if there was no partitioning of A into subsets A_K, representing items, i.e. this analysis ignores the characteristic structure of the system.

-4.4-

There is yet another way of looking at this structure. Let each item define a relational system $\langle A_K, B, R_{K1}, \ldots, R_{Kn_K} \rangle$, in which $A_K = \{a_K\}$ is a one-element set, B has m elements and the R_{Kl} are binary relations. The interpretation might run as follows: $b_j R_{Kl} a_K$ iff subject j responds to item K with alternative l. Maximize the difference between the subjects, that are in the relation R_{Kl} to a_K and those that are in the relation R_{Ks} to a_K, for all pairs (l,s) for which $l < s$ simultaneously, i.e. for $(l,s) = (1,2), (1,3), \ldots, (n_K - 1, n_K)$. Analytically, E is interpreted as a supermatrix, consisting of N different $n_K \times m$ submatrices E_K, one for each item. Define an $n_K \times n_K$ diagonal matrix D_K, in which the diagonal elements are the row sums of E_K, and an $\binom{n_K}{2} \times n_K$ matrix S_K, analogous to the matrix S used in the previous chapter. If $n_K = \frac{n}{N}$ for all K (i.e. if all items have the same number of alternatives), then all N matrices S_K are equal. The $\frac{1}{2} n_K (n_K - 1)$ different signed distances between the n_K alternatives of item K, given some score vector x, are given by

$$d_K = S_K D_K^{-1} E_K x$$

Maximize

$$\mu_1 = \sum_{K=1}^{N} d_K' d_K = \sum_{K=1}^{N} x' E_K' D_K^{-1} S_K' S_K D_K^{-1} E_K x =$$

$$= x' \left[\sum_{K=1}^{N} E_K' D_K^{-1} S_K' S_K D_K^{-1} E_K \right] x =_{\text{def}} x' T x$$

There is, of course, the possibility of assigning different a priori weights to both items and alternatives. Define the following supermatrices

$$S: \left\| \begin{matrix} S_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & S_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & S_N \end{matrix} \right\| \qquad D: \left\| \begin{matrix} D_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & D_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & D_N \end{matrix} \right\|$$

Then the symmetric, doubly centered $m \times m$ matrix T is given by

$$T = E' D^{-1} S' S D^{-1} E$$

Solutions for x are found in the usual way. This analysis permits us to draw inferences about discrimination power of items by using $d_K' d_K$, about discrimination power and interrelati ns of item alternatives, by using the elements of d_K, and finally about the discrmination power of the test, by using μ_1. From this point of view too, the procedure is superior to that outlined in chapter II, because it provides us with more information for item analysis and selection. The item alternatives them selves can be quantified by applying the method of chapter II to the transposed supermatrix E, i.e. maximizing the difference between those alternatives that are chosen by a particular subject and the ones he does not choose. More sophisticated techniques for quantifying both items and alternatives at the same time are discussed in De Leeuw (196- b).

-4.5-

The expression for the T-matrix of item K looks very complicated, but the T_K-matrix itself turns out to be surprisingly simple to construct. To see this, observe in the first place that $S_K'S_K = n_K I - J_{n_K n_K}$, which means that $T_K = E_K' D_K^{-1} S_K' S_K D_K^{-1} E_K = E_K' D_K^{-1} \left[n_K I - J_{n_K n_K} \right] D_K^{-1} E_K = n_K^2 C$, where C is the variance-covariance matrix of the columns of $F_K = D_K^{-1} E_K$. Our matrix T turns out to be a weighted sum of variance-covariance matrices, which is quite surprising, because we started out with a maximization of $\sum_{K=1}^{N} \binom{n_K}{2}$ distances. From the mutual exclusiveness of the item alternatives a very simple expression follows for the elements of $T_K = n_K^2 C$. Suppose subject k responded with alternative i, and subject l with alternative j. Then

$$t_{kl}^{(K)} = \frac{n_K \delta^{ij} - 1}{d_{ii}^{(K)} d_{jj}^{(K)}}$$

Although this procedure can be generalized quite simply to non-exclusive relations, in that case this simple expression for T_K clearly does not hold.

-4.6-

It should be clear by now that the investigator has a considerable amount of freedom in choosing the analytic technique for his data. This is especially obvious in the case of multiway matrices in which we may 'poole' some facets and eleminate others by analyzing the data for each element of the facet seperately. Another source of freedom is the model we choose. Suppose that m subjects respond to n questions with yes-no-don't know responses. The results can be analyzed by the methods of chapter II: maximize the difference between the scale values of the subjects that say 'yes' to question a_i and those that say 'no', for all $i=1,\dots,n$ simultaneously. The subjects can have two different relations to a particular question, these relations are mutually exclusive, although they are not necessarily exhaustive. 'Don't know'-responses are ignored (although they contribute to the weighting of the question in the data structure, but this effect can be eliminated). An entirely different interpretation assumes that the responses have ordinal properties. By scoring them -1, 0, and +1 they can be treated by nonmetric factor analysis. A particular row of E is interpreted as a (degenerate) I-scale. The only ordinal information that is used in the analysis is provided by the relations between elements of a row of E that have different values. A third interpretation is even weaker than the first. We may interpret the responses as three different relations and maximize all $\binom{3}{2}$ distances between means of scale values with the methods of sections 4 and 5 of this chapter. This method does not single out 'yes' and 'no', but treates 'don't know' in exactly the same way. Other intermediary methods are possible (although rather unreasonable), for example one that maximizes the difference between 'yes' and 'don't know' and the difference between 'no' and 'don't kno', but does not pay any attention to the difference between 'yes' and 'no'. The choice of the method depends only on our interpretation of the data.

+++++++++++

Chapter V

COMPUTER PROGRAMS & EXAMPLES

-5.1-

For the techniques discussed in this paper a number of PL/I programs were written for the IBM 360/50. The first one in the series is CDARD1, designed for the simple quantification cases of chapter II. Some options are added which make it possible that some of the three-way designs of the previous chapter can also be analyzed with this program. As a first example, consider the following experiment: two sets of stimuli were sampled from two binormal distributions, with equal dispersion matrices but with different vectors of means. The subject's task was learning to classify the stimuli in the correct way, i.e. they learn to discriminate between the two populations. By projecting the stimuli on the loglikelihood-axis defined by the two distributions, it is possible to compute a d'-type measure, called α', indicating how well the subject performs this task. The data (provided by De Klerk) have been taken from an experiment of this type in which 8 subjects categorized three times the same block of 50 stimuli. The ICP-matrix that we use to quantify the stimuli is a 24x50 matrix. Only the first eigenvector and eigenvalue were significant in this case. The most interesting results obtained from this quantification concern the relation between the α'-measure, computed by using the stimulus-characteristics of the two distributions defined by the subject's classifications, and the D^2-measure, computed with our technique. The D^2-measure does not require measurable stimulus dimensions, because its value only depends on the classificatory responses of the subjects. If a particular subject has a relatively high D^2-measure, it means that he fits very well in the classification pattern of the group of subjects. His classifications are close to those of the typical subject. The plot of α' versus D^2 is given in figure I.

-5.2-

A second example is provided by the sociogram of table I (provided by Jaspars). It is useful to compare the quantification of the sets A and B, and the signed distances that are the consequences of these quantifications. In table IIa the two-dimensional quantification of B is given (principal components), in table IIb the corresponding signed distances. In table IIc we have the optimal quantification of A and in IId the signed distances, that are found by analyzing the transpose of table I (it is evident that all four tables would have been identical if the ICP-matrix had been symmetric). In this sociogram $e_{ij} = 1$ iff subject i knew the name of the street in which subject j lived. Another sociogram (using the same group of subjects) is given in table III. There $e_{ij} = 1$ iff subject i knew the first name of subject j. In table IVa-IVd the same results are collected as in table II.

-5.3-

The next example is taken from Coombs (1964 p 481), where a confusion matrix P for 10 morse code signals is given. Define a symmetric, reflexive binery relation R over A x A, in such a way that $a_i R a_j \Longleftrightarrow p(i,j) > \beta$. This relation can be interpreted as a_i is similar to a_j. We have taken β as 20, which is of course an arbitrary choice (and which may be very important). The ICP-matrix of the system $\langle A,A,R \rangle$ is given in table V. Obviously these morse code signals constitute a circumplex. This is brought out by the CDARD1-results in table VI and figure II.

-5.4-

To illustrate the joint-space MSA-representation, in which one set is represented as points, and the other set as partitionings, we did the following experiment. Eleven subjects (the set C) said of the twelve most important Dutch political parties (B) whether they possessed a particular attribute or not. This was done for a list of 17 attributes (A). The data were analyzed for each c_i separately. In table VII we have collected the three-dimensional quantification of the 12 parties for one subject, in table VIII the corresponding signed distances for the 17 attributes, and in figure III a-m the partitionings of the two-space, defined by the first two dimensions, by these attributes.

-5.5-

The second program in the series, CDARD2, is designed for nonmetric factor analysis in the sense of chapter III. It quantifies the set A from the ICP-matrix of the system $\langle A, A, B, \geq_o \rangle$. The first example is the analysis of a set of preference rank orders obtained from 100 different subjects. The ranked objects were nine Dutch political parties (data provided by Van Der Kamp). Again only two factors were significant. The results are given in table IX and figure IV. The results are virtually the same as those obtained earlier with an analysis with Slater's method.

-5.6-

Eleven subjects were asked to give similarity rankings of nine Dutch political parties, Take one partu as a standard and rank all others as to their similarity with the standard. Pivoting the standard results in a set of nine I-scales, that can be analyzed with the CDARD2-program. The results for one of the subjects are given in table Xa and figure V. In table Xb the signed distances for that same subject are given.

-5.7-

The CDARD3-program is a rather specialized routine for the analysis of additive conjoint meausrement structures (3.10). CDARD4 is designed for nonmetric multiple prediction (3.11), it finds the regression weights that result in an optimal prediction of a variable monotone with the dependent variable. Examples of both CDARD3 and CDARD4 are given in an as yet unpublished paper of De Klerk, De Leeuw and Oppe on functional learning.

-5.8-

CDARD5 quantifies the set B from the ICP-matrix of the system $\langle A, A, B, \geq_o \rangle$. This is done by the methods of chapter II, i.e. by the methods of CDARD1. The CDARD5 program, however, needs much less storage for the input matrices because it can be used directly on the relational system $\langle A,B,Re,\psi\rangle$, for example the nxm matrix of scores of m subjects on n tests. The example is taken from the Dutch newspaper 'De Telegraaf', march 27, 1968. Voters from seven Dutch political parties participated in an experiment on political preferences and similarities. The cell (i,j) of the input matrix was the percentage of voters of party i, that were in favor of cooperation with party j in a governmental coalition. The matrix was analyzed with CDARD2, results are presented in table XII and figure VII. Only the first factor seems to be of any importance. A CDARD5 analysis of the same matrix reveals the two dimensions of table XIII and figure VIII. The corresponsing signed distances are collected in table XIV.

-5.9-

The final program in the series, CDARD6, quantifies the multiple relational systems discussed in chapter IV. We applied it to data collected by Van Hoorn. Subjects expressed their opinion about a certain lecture course by answering 'yes', 'no' or 'don't know' to twelve questions. The set of subjects was quantified by CDARD1, CDARD2 and CDARD6 respectively. Results are given in table XV a,b,c.

-5.10-

In a paired associate word learning experiment with 20 subjects, 25 word pairs and 10 trials the number of correct responses on each trial was recorded in an 20x10 matrix. This matrix was analyzed with the CDARD2 program to quantify the 10 trials in order to find generalized learning curves in the sense of Tucker (1966). Results are given in table XI a and in figures VIa and b. In these figures the scores on the two principal components are plottedas a function of the trial numbers. The results agree with those obtained with an orthodox Tucker analysis. Signed distances are given in table XIb.

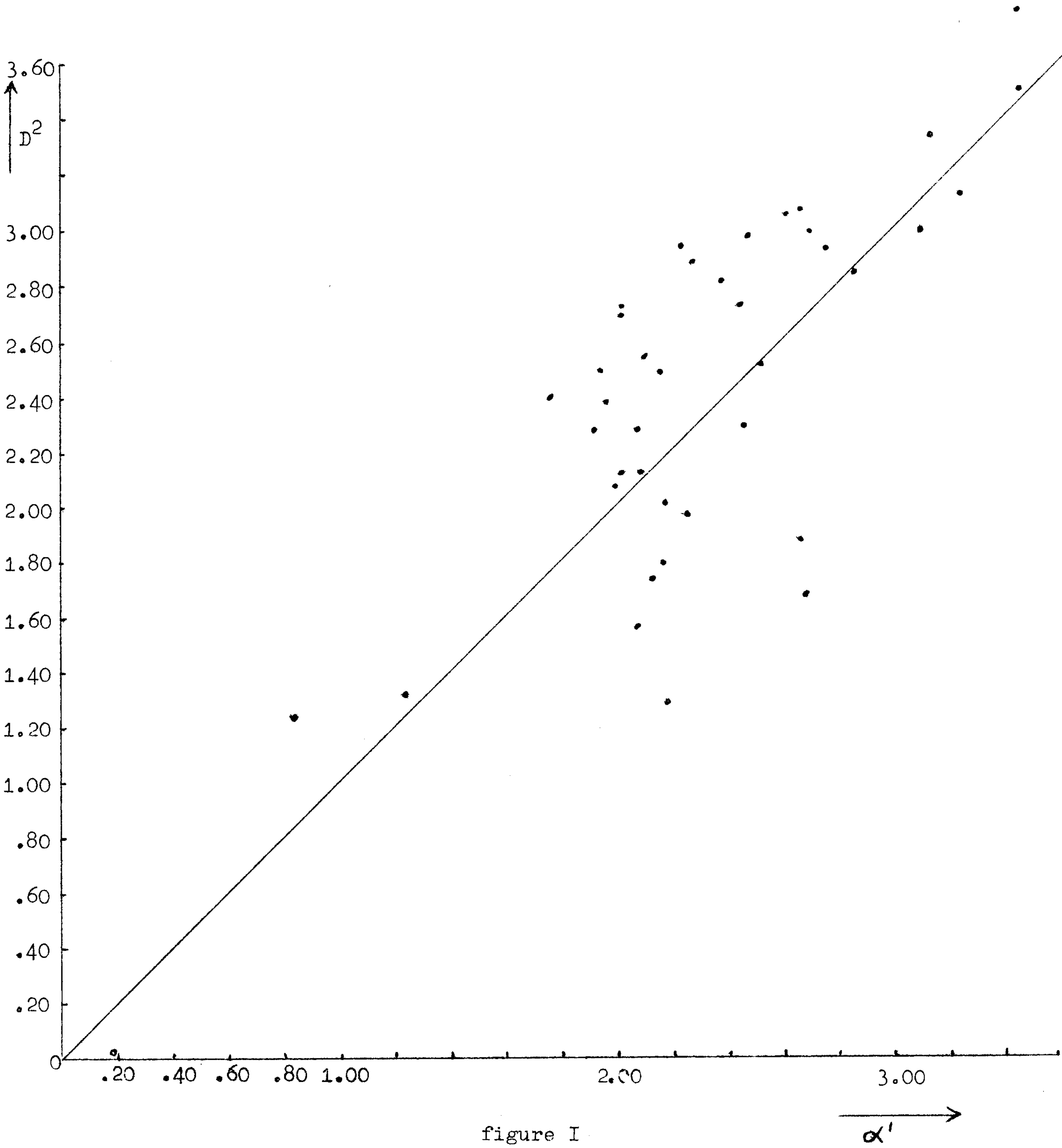

3.60
D^2
3.00
2.80
2.60
2.40
2.20
2.00
1.80
1.60
1.40
1.20
1.00
.80
.60
.40
.20
0
.20
.40
.60
.80
1.00
2.00
3.00
α'

figure I

1	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	1	0	0	1
1	0	0	0	1	0	0	0	1	0	0	1
0	1	0	0	0	1	1	0	0	0	1	0
0	1	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	1
1	0	0	1	1	0	0	0	1	1	0	1
0	0	0	0	0	0	0	0	1	1	0	0
0	1	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1	0	0	1

table I

1	0	0	0	1	1	1	0	1	0	0	0
1	1	1	1	1	1	1	0	1	1	1	1
1	0	1	0	0	0	0	0	0	0	0	0
0	1	0	1	1	1	0	0	1	1	0	0
1	0	0	1	1	1	0	0	1	0	0	1
1	1	0	0	1	1	1	0	1	0	1	0
1	1	0	0	1	1	1	0	0	0	0	0
~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
1	1	0	1	1	1	0	0	1	1	0	0
1	1	0	0	0	1	0	0	1	1	0	0
1	1	1	1	1	0	0	0	1	0	1	0
1	1	0	0	1	0	0	0	1	0	1	1

table III

I	II
-0.250	0.007
0.763	-0.322
0.142	1.029
-0.186	-0.002
-0.250	0.007
0.178	0.035
0.552	-0.192
-0.060	0.064
-0.638	-0.230
-0.193	-0.002
0.552	-0.192
-0.611	-0.203
33.37%	20.05%

table IIa

I	II
-0.201	0.008
0.556	-0.271
0.104	0.971
-0.428	-0.167
-0.440	-0.136
0.514	-0.217
0.529	-0.267
-0.270	-0.072
-0.476	-0.122
-0.334	-0.121
0.529	-0.267
-0.502	-0.225

table IIb

I	II
-0.390	0.149
-0.012	-0.476
0.092	0.623
-0.069	0.042
-0.352	-0.235
-0.298	-0.199
-0.028	0.026
1.203	-0.190
-0.330	-0.204
0.062	0.021
0.009	0.179
0.113	0.265
39.29%	17.00%

table IVa

I	II
-0.342	-0.167
-0.939	0.217
-0.128	0.484
-0.238	-0.366
-0.316	-0.064
-0.349	-0.273
-0.265	-0.264
0.512	-0.418
-0.341	-0.324
-0.171	-0.174
-0.362	-0.048
-0.230	-0.113

table IVb

I	II
-0.203	-0.101
0.462	-0.064
0.053	0.694
-0.322	-0.112
-0.385	-0.173
0.984	-0.391
0.308	-0.000
-0.136	0.694
-0.695	-0.471
-0.217	-0.110
0.308	-0.000
-0.159	0.037
30.65%	19.27%

table IIc

I	II
-0.381	-0.279
0.517	-0.144
0.039	0.638
-0.408	-0.295
-0.381	-0.279
0.717	-0.360
0.521	-0.171
-0.099	0.638
-0.407	-0.240
-0.365	-0.294
0.520	-0.171
-0.389	-0.008

table IId

I	II
-0.147	0.252
-0.420	-0.200
0.242	0.601
0.036	-0.540
-0.249	0.043
-0.203	0.024
0.012	0.188
1.248	-0.189
-0.189	-0.255
0.083	0.134
-0.269	-0.093
-0.145	0.036
35.90%	16.10%

table IVc

I	II
-0.538	0.455
0.043	-0.414
-0.139	0.142
-0.261	-0.373
-0.488	-0.252
-0.374	-0.227
-0.161	0.142
0.950	-0.215
-0.466	-0.278
-0.128	-0.336
-0.271	-0.092
-0.252	-0.056

table IVd

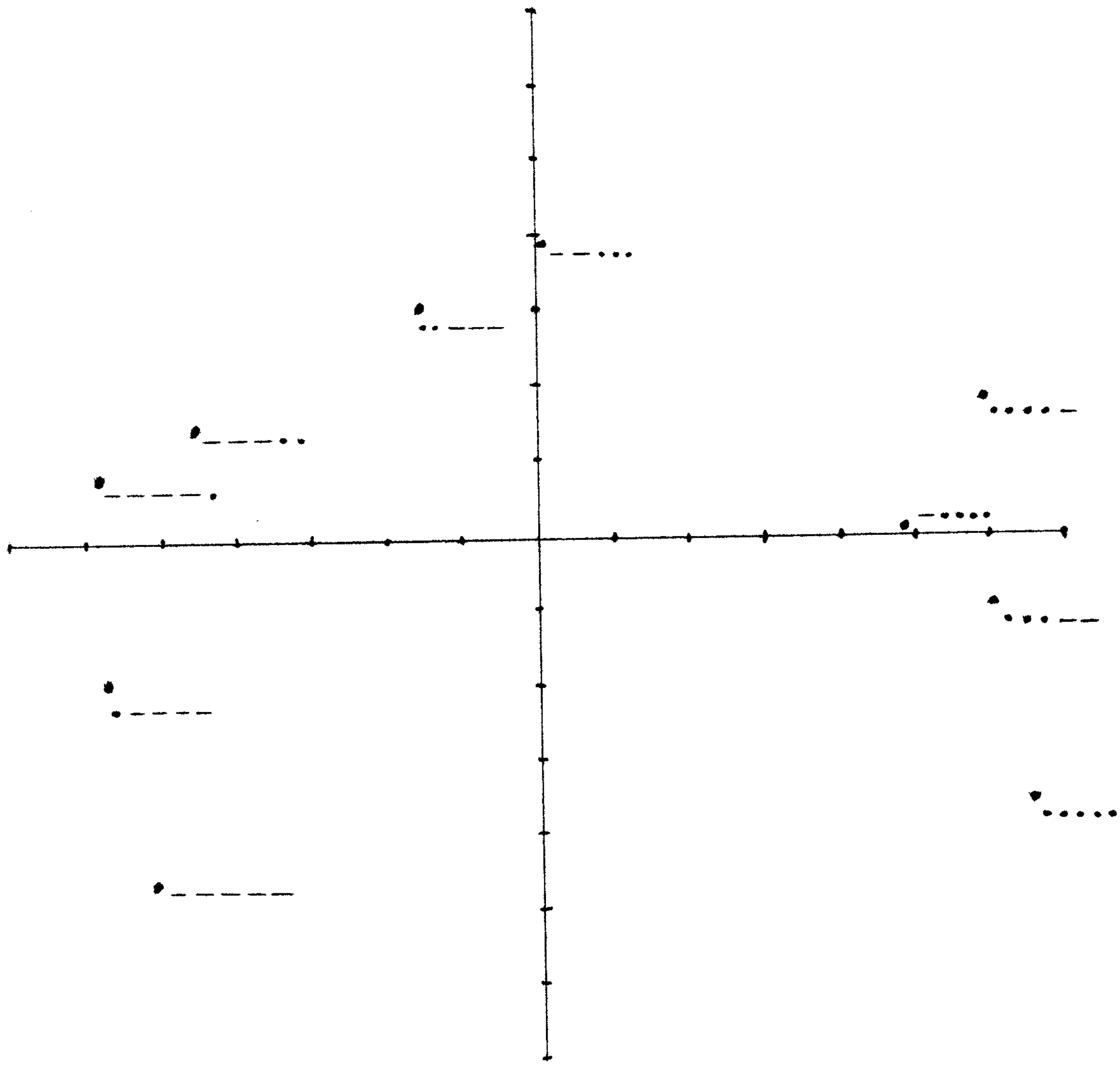


figure II

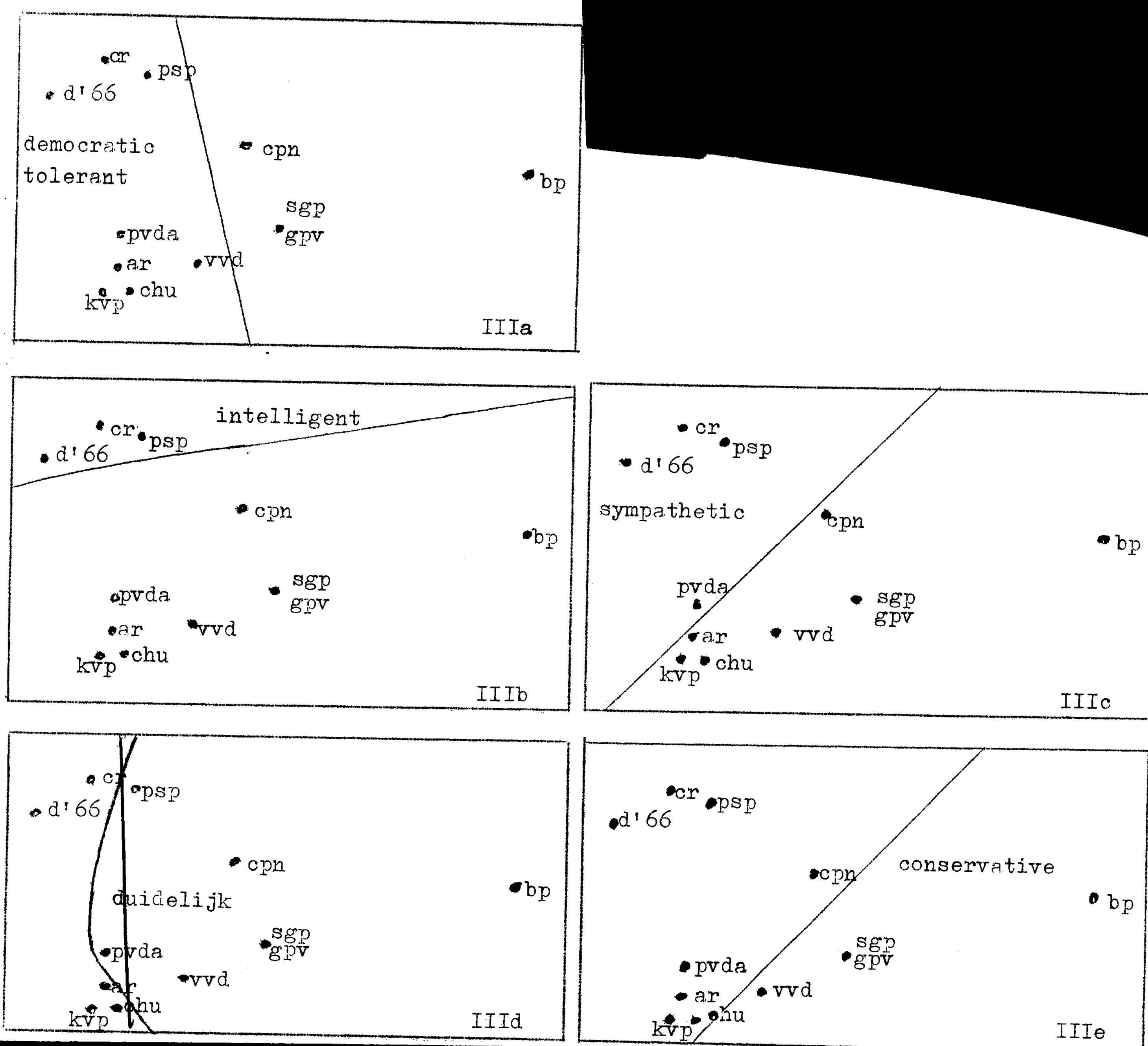
	I	II
•----	-0.575	-0.198
••---	-0.159	0.304
•••--	0.601	-0.094
••••-	0.590	0.182
•••••	0.656	-0.358
-••••	0.481	0.007
--•••	-0.026	0.393
---••	-0.461	0.147
----•	-0.586	0.079
-----	-0.520	-0.462
	60.10%	16.41%

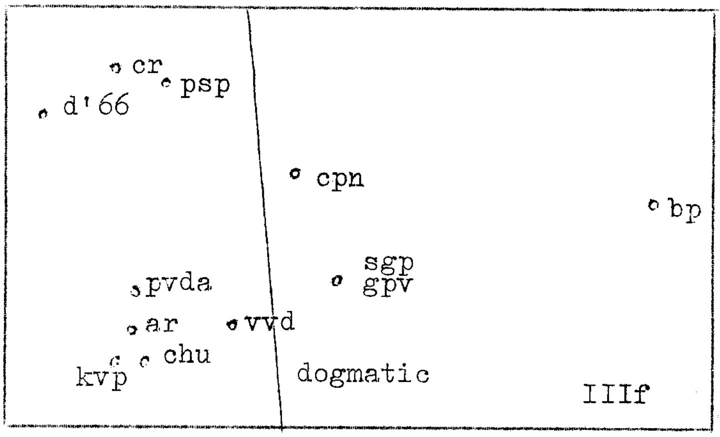
table VI

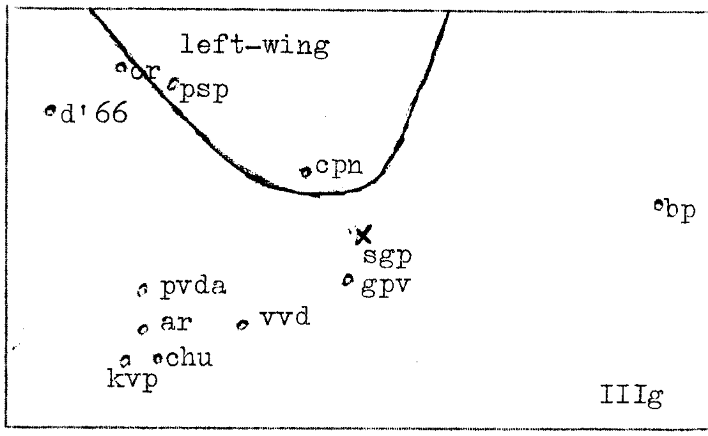
1	1	0	0	0	0	0	1	1	1
1	1	1	1	0	0	1	1	1	0
0	1	1	1	1	1	0	0	0	0
0	1	1	1	1	1	1	0	0	0
0	0	1	1	1	1	0	0	0	0
0	0	1	1	1	1	1	1	0	0
0	1	0	1	0	1	1	1	1	0
1	1	0	0	0	1	1	1	1	1
1	1	0	0	0	0	1	1	1	1
1	0	0	0	0	0	0	1	1	1

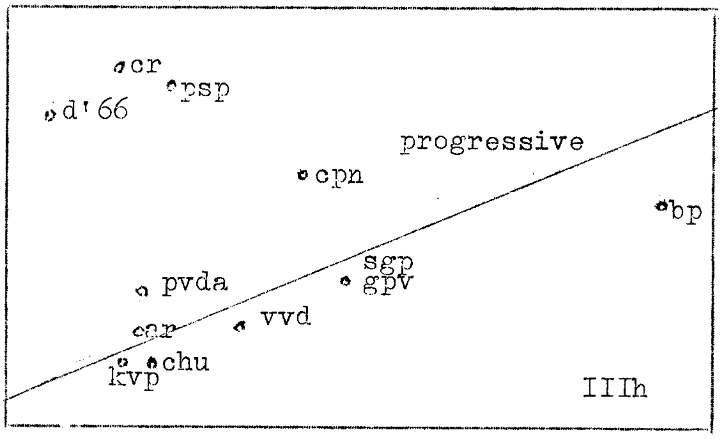
table V

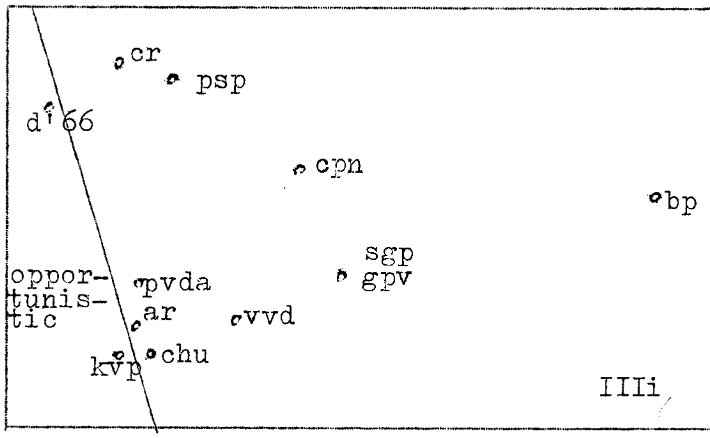
table VII

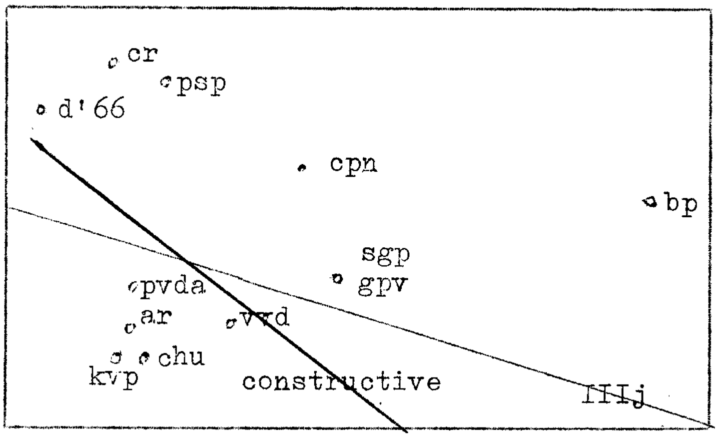

cr
psp
d'66
democratic
tolerant
cpn
bp
sgp
gpv
pvda
ar
vvd
kvp
chu
IIIa
intelligent
IIIb
sympathetic
IIIc
duidelijk
IIId
conservative
IIIe

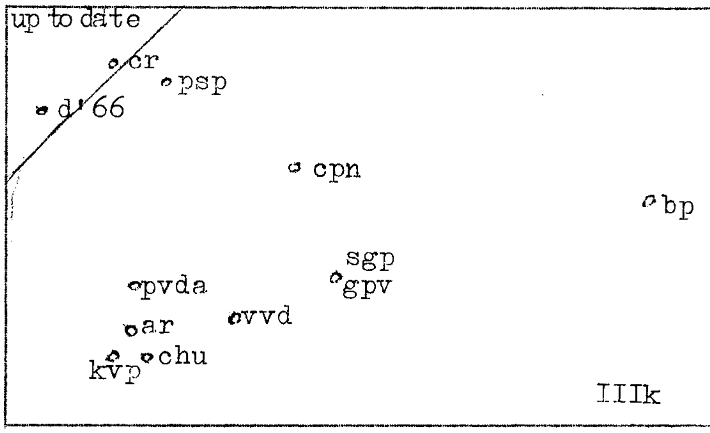

cr
psp
d'66
cpn
bp
sgp
gpv
pvda
ar
vvd
chu
kvp
dogmatic
IIIf

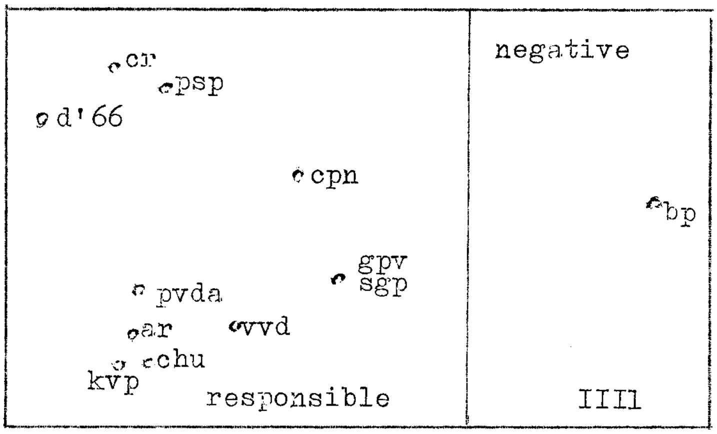

left-wing
cr
psp
d'66
cpn
bp
sgp
gpv
pvda
ar
vvd
chu
kvp
IIIg

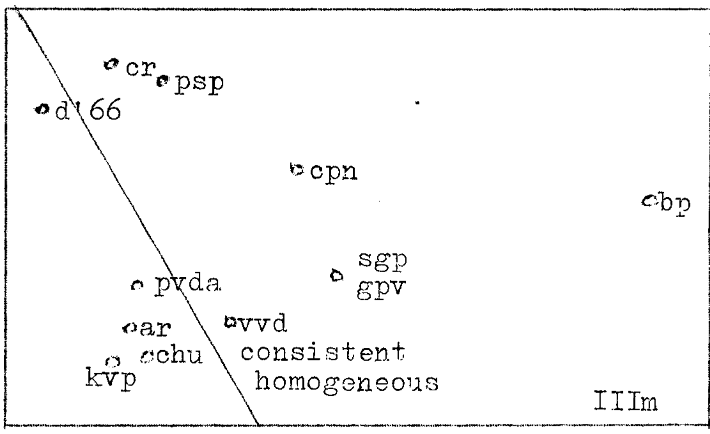

cr
psp
d'66
progressive
cpn
bp
sgp
gpv
pvda
ar
vvd
chu
kvp
IIIh


cr
psp
d'66
cpn
bp
sgp
gpv
oppor-
tunis-
tic
pvda
ar
vvd
chu
kvp
IIIi


cr
psp
d'66
cpn
bp
sgp
gpv
pvda
ar
vvd
chu
kvp
constructive
IIIj


up to date
cr
psp
d'66
cpn
bp
sgp
gpv
pvda
vvd
ar
chu
kvp
IIIk


negative
cr
psp
d'66
cpn
bp
gpv
sgp
pvda
ar
vvd
chu
kvp
responsible
IIIl


cr
psp
d'66
cpn
bp
sgp
gpv
pvda
ar
vvd
chu
kvp
consistent
homogeneous
IIIm

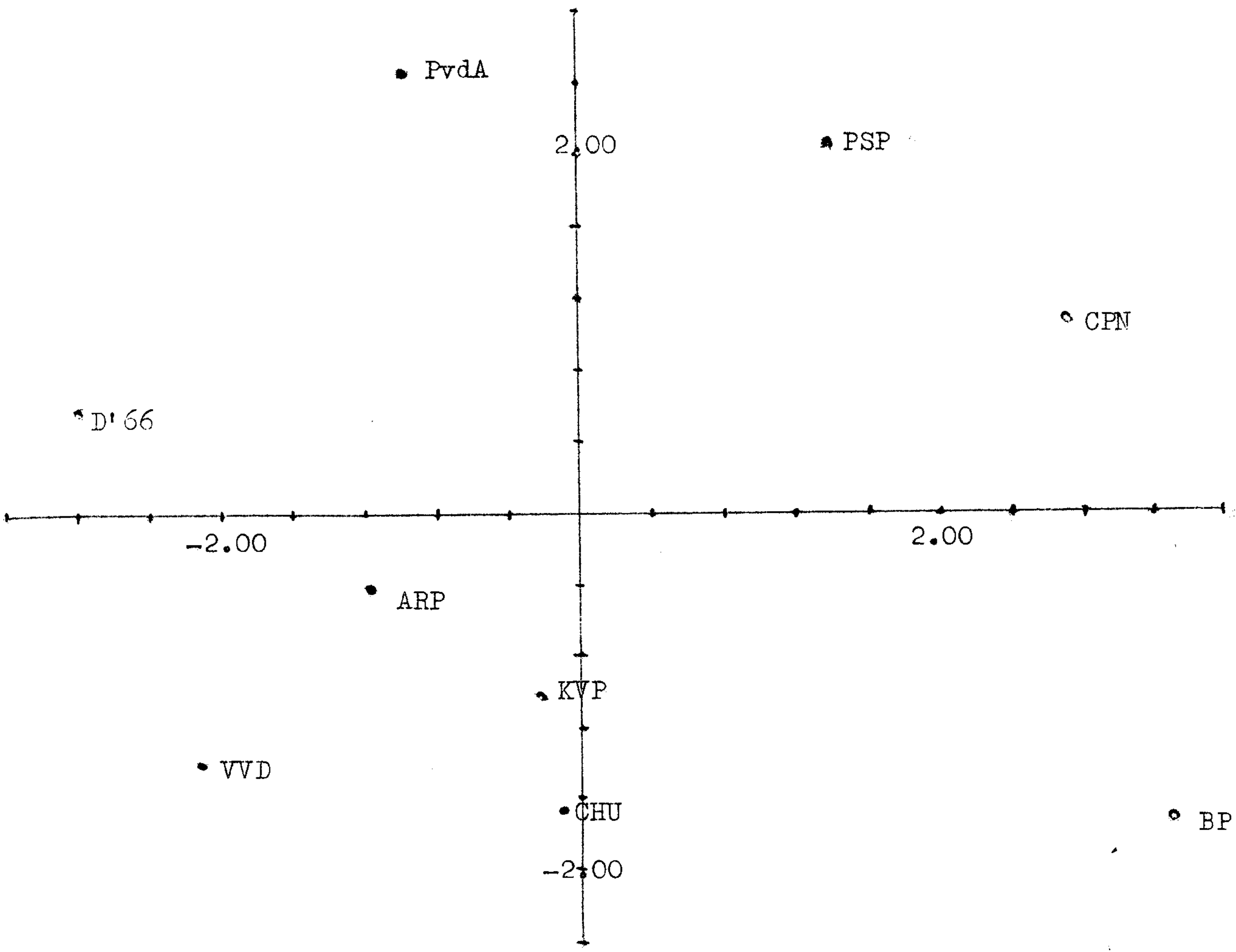


figure IV

	I	II
KVP	-0.225	-1.030
PvdA	-1.017	2.480
VVD	-2.115	-1.399
ARP	-1.180	-0.424
CHU	-0.108	-1.667
CPN	2.733	1.100
PSP	1.384	2.076
BP	3.318	-1.715
D'66	-2.789	0.577
	46.06%	27.43%

table IX

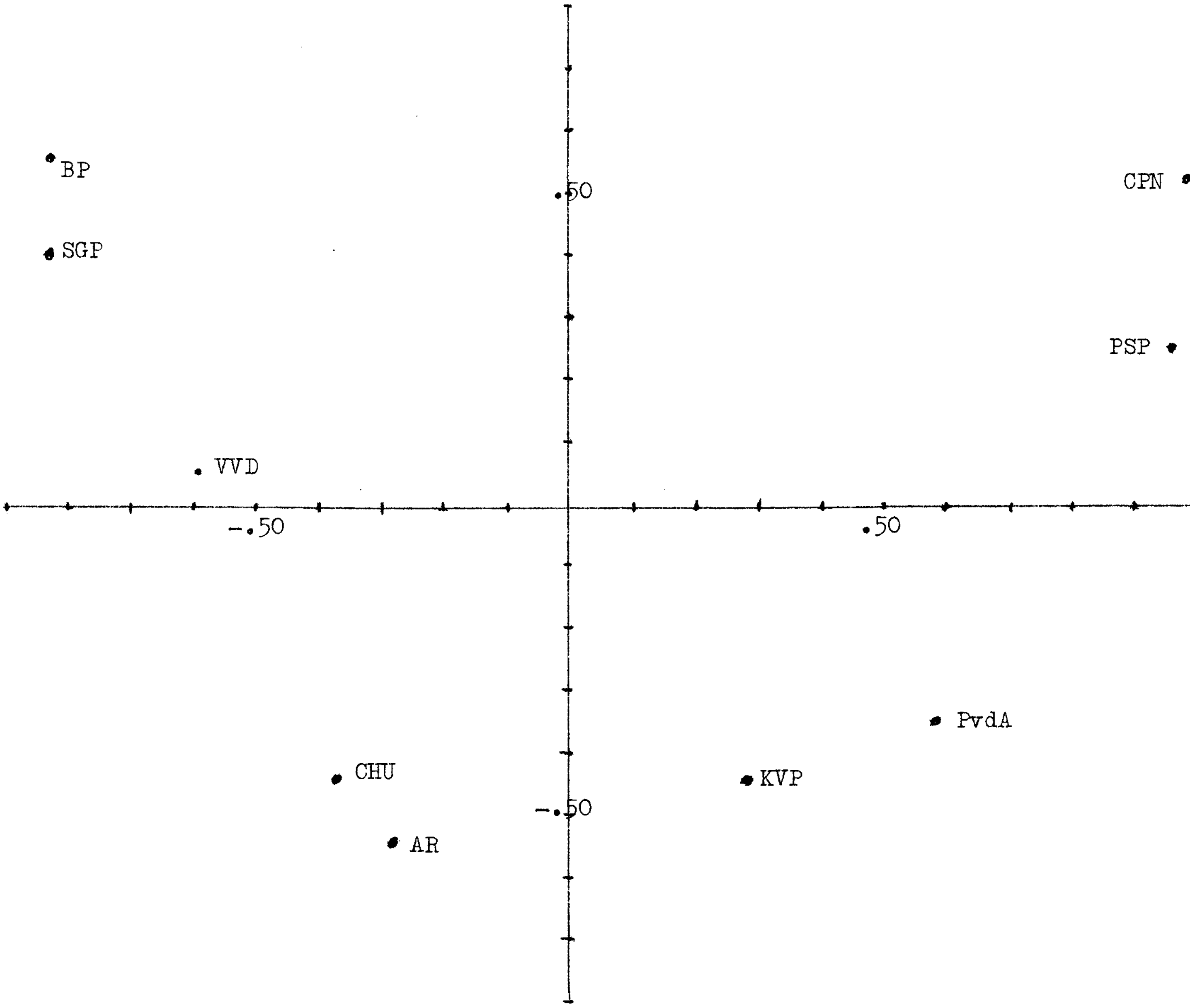


figure V

	I	II
PvdA	0.581	-0.354
PSP	0.955	0.251
BP	-0.830	0.555
SGP	-0.829	0.399
CPN	1.075	0.521
AR	-0.278	-0.536
CHU	-0.371	-0.436
KVP	0.283	-0.444
VVD	-0.585	0.043
	66.27%	24.05%

table Xa

	I	II
PvdA	-0.791	0.261
PSP	-0.837	0.145
BP	0.825	-0.081
SGP	0.795	-0.014
CPN	-0.844	0.121
AR	0.450	0.622
CHU	0.416	0.707
KVP	-0.145	0.778
VVD	0.818	-0.045

table Xb

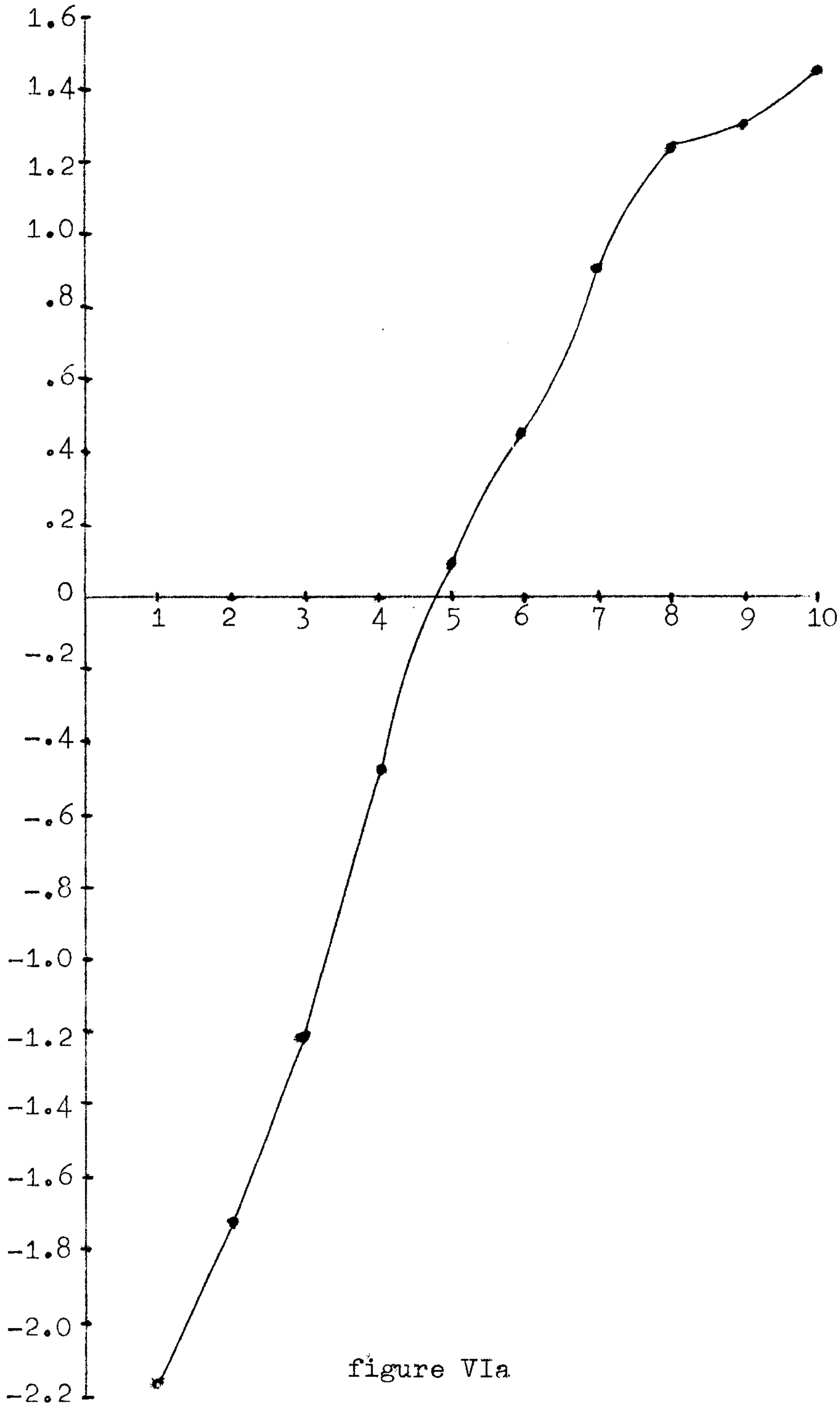


figure VIa

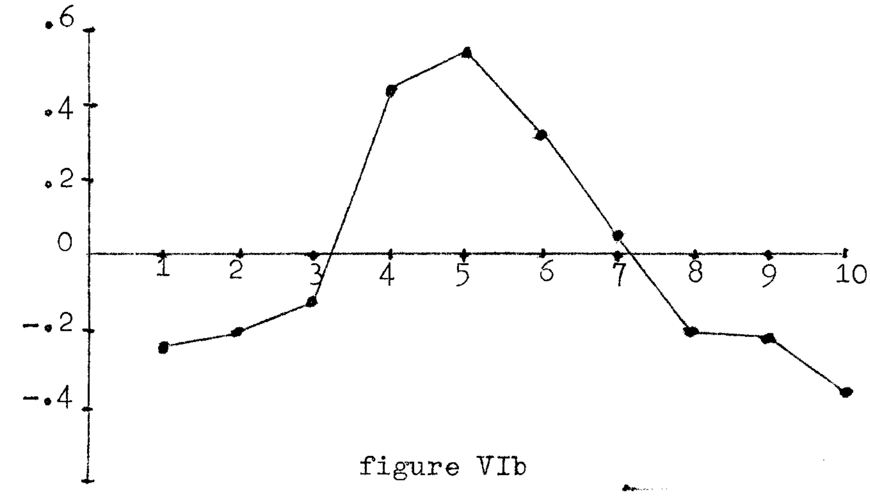


figure VIb

I	II
-2.158	-0.248
-1.696	-0.195
-1.196	-0.125
-0.458	0.440
0.096	0.544
0.467	0.319
0.912	0.051
1.253	-0.197
1.308	-0.223
1.470	-0.367
90.87%	5.41%

table XIa

I	II
0.791	-0.184
0.908	-0.083
0.802	-0.127
0.806	-0.136
0.783	-0.117
0.811	-0.094
0.858	-0.124
0.856	-0.139
0.794	-0.136
1.044	0.203
0.808	-0.141
0.966	0.010
0.873	-0.131
0.890	-0.093
1.103	0.511
1.022	0.235
0.821	-0.127
1.103	0.510
0.803	-0.127
0.735	-0.294

table XIb

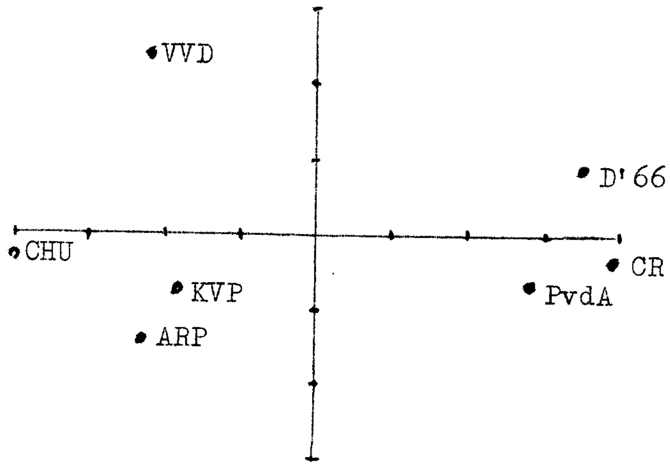


figure VII

	I	II
KVP	-0.539	-0.208
VVD	-0.651	0.732
ARP	-0.686	-0.417
CHU	-1.213	-0.072
PvdA	0.853	-0.216
D'66	1.056	0.278
CR	1.179	-0.098
	79.99%	12.09%

table XII

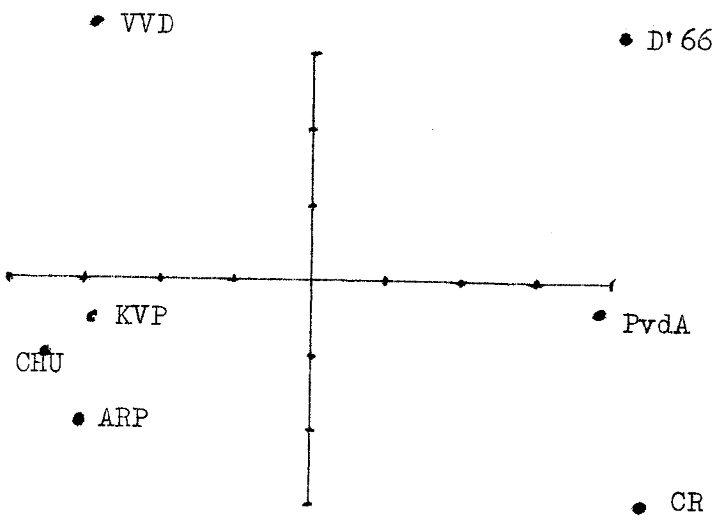


figure VIII

	I	II
KVP-VVD	-0.099	-0.781
KVP-ARP	0.255	0.457
KVP-CHU	0.582	-0.052
KVP-PvdA	-0.762	0.006
KVP-D'66	-0.762	0.006
KVP-CR	-0.762	0.006
VVD-ARP	0.099	0.781
VVD-CHU	0.205	0.819
VVD-PvdA	-0.762	0.006
VVD-D'66	-0.762	0.006
VVD-CR	-0.762	0.006
ARP-CHU	0.574	-0.190
ARP-PvdA	-0.762	0.006
ARP-D'66	-0.762	0.006
ARP-CR	-0.762	0.006
CHU-PvdA	-0.762	0.006
CHU-D'66	-0.762	0.006
CHU-CR	-0.762	0.006
PvdA-D'66	-0.022	-0.714
PvdA-CR	-0.540	0.574
D'66-CR	-0.099	0.565

table XIV

	I	II
KVP	-0.879	-0.143
VVD	-0.869	1.027
ARP	-0.914	-0.571
CHU	-1.057	-0.292
PvdA	1.130	-0.125
D'66	1.270	0.998
CR	1.319	-0.893
	59.63%	24.23%

table XIII

I	II
-0.117	-0.060
0.095	0.013
0.055	-0.049
-0.003	-0.011
-0.388	0.311
-0.148	0.296
-0.040	0.096
0.265	-0.212
0.030	-0.029
-0.078	-0.018
-0.584	-0.713
0.064	0.706
0.058	0.023
0.024	-0.033
0.019	0.067
-0.060	-0.020
0.995	-0.250
-0.074	-0.062
-0.015	-0.035
-0.098	-0.021
21.37%	17.32%

table XVa

I	II
-0.214	0.345
0.006	-0.249
-0.146	-0.212
-0.146	-0.219
-0.285	0.627
-0.322	0.486
0.371	0.700
0.261	-0.158
-0.196	-0.116
-0.203	0.079
-0.273	-0.102
0.914	0.106
0.005	-0.397
-0.128	-0.147
0.393	-0.302
-0.176	-0.222
0.846	-0.002
-0.278	-0.145
-0.215	0.013
-0.214	-0.087
26.57%	18.82%

table XVb

I	II
-0.075	0.218
0.006	-0.030
0.022	0.052
-0.048	-0.007
-0.143	-0.592
-0.023	1.385
0.327	0.246
0.052	-0.105
0.006	0.036
-0.032	-0.020
-0.222	-0.487
-1.026	-0.384
-0.038	0.014
-0.005	0.086
-0.275	-0.111
-0.053	0.021
1.625	-0.388
0.006	-0.071
-0.018	0.115
-0.087	0.021
19.67%	14.69%

table XVc

Chapter VI

DISCUSSION

-6.1-

The models developed in this paper can be criticized from a measurement theoretical point of view. We do not pay any attention to the nonmetric properties of the data matrix, before we analyze it with our computer programs. No check is made, whether the necessary and/or sufficient conditions for a particular measurement model are satisfied. This is due to the fact that , in the opinion of the present writer; it is rather useless to ask whether these conditions are satisfied or not (in almost all cases they will not be satisfied), and that the question should be 'how well are they satisfied ?' This is essentially a statistical question, for which the answer depends on error theorétical aspects and not directly on measurement theory. For most psychological measurement models (conjoint measurement, nonmetric multidimensional scaling, factor analysis, scalogram analysis) the error theory is relatively primitive, and the goodness of fit question is in most cases treated by inspeection (if it is treated at all). For some reflections on error theory in the context of measurement theory, see Krantz (1967 p 21-23) and Tversky (1964 p 26-27).

-6.2-

Our methods are not designed to satisfy an optimal number of the restraints imposed by the measurement model, but they optimize some coefficient. The numerical value of these coefficients is related to the number of violations but also to the size of the errors. They can be compared in this respect with the stress measures of Kruskal and Guttman & Lingoes for ordinal measurement models (cf De Leeuw 1968a). The procedure for ACM, outlined in section 10 of chapter III, is a good example. The ACM requirements can be translated into psychometric terms, i.e. define the matrices F and S and the vector x, and maximize x'S'F'FSx. It is rather unimportant, whether such a procedure is still called ACM. Clearly, it is closely related to it, and clearly it may produce interesting results from a practical point of view.

-6.3-

As was pointed out before, the success of our models in a particular situation can only be evaluated in a statistical framework. The famous when-to-stop-factoring problem can only be attacked with an appropriate statistical test for the significance of eigenvalues and/or residuals. The statistical techniques presented is this paper are rather unsatisfactory, for the very simple reason that the present writer is unable to devise better ones. However important an appropriate statistical test may be, and however incomplete any psychometric technique is without such a test, it is most certainly not a conditio sine qua non. As Burt points out:" ... if the statistician has not as yet developed an appropriate method which he can offer us, then the psychologist, however imperfectly, must set to work to devise his own " (Burt 1950, p 168). A more important argument is given by Guttman (1941b, p 341):"We cannot even begin to tackle sampling problems until we define what the 'best' answer would be for the population, for the case where there are no sampling errors." If our techniques are interesting enough to be accepted and used on a fairly large scale, then it is probable that someone more verged in statistics than the present writer will elaborate the statistical side of the problems.

-6.4-

Our maximization procedures lead to the computation of the principal components of a certain symmetrical matrix as the optimal quantification of a facet. In his fundamental paper Guttman emphasizes only the first principal component as a quantification vector. He objects to using multidimensional quantifications, because this 'factor analysis' does not seem to result in a natural chi square metric (1941b, p 331). As Kendall & Stuart (II p 574) point out, all eigenvactors give rise to orthogonal components of chi square, although a result of Lancester proves, that the corresponding eigenvalues are not independent chi square variates, which means that the familiar additive chi square analysis may not be used. Moreover, as Burt point out, Guttman's objections to factorizing qualitative data are one-sided and confusing, because they "limit the term factor analysis to very specific forms " (Burt, 1953, p 10), notably the classical Spearman-Thurstone approach. The two principal objections of Guttman (factor analysis is designed only for quantitative variates, and only covariances and correlations should be factored) are quite thoroughly demolished by Burt by showing that a factor analysis (in Pearson's sense) of ordinary product moment correlations yields exactly the same results as Guttman's technique in the mathematically very interesting case of a perfect scale. In our technique in addition, the results are evaluated in terms of Mahalanobis-type distance measures, which are evidently additive over the orthogonal dimensions, i.e. one-dimensional quantification is only a special case. In this paper principal components are used primarily in an algorithmic sense, although a well-defined and relatively simple statistical interpretation can indeed be given. What we have shown, is that an astonishing wide variety of non-metric measurement models, formulated originally in terms of relational systems, can be translated into matrix terminology. Moreover methods of quantifying the defining sets of the system can be found by computing the eigenvectors and eigenvalues of these matrices. In the worst case the results obtained by these methods can be used satisfactorily as starting configurations for more sophisticated iterative programs.

++++++++++

LITERATURE

Burt, C. The factorial analysis of qualitative data
British Journal of Statistical Psychology, 1950, 3, 166-185.

Burt, C. Scale analysis and factor analysis
British Journal of Statistical Psychology, 1953, 6, 5-24

Coombs, C.H. A theory of data
New York, Wiley, 1964

Coombs, C.H. & Kao, R.C. Nonmetric factor analysis
Engineering research bulletin, No 38, University of Michigan, Ann Arbor, 1955

Daniels, E. H. The relation between measures of correlation in the universe of sampling permutations
Biometrika, 1944, 33, 129-

De Leeuw, J. Algorithms for the numerical analysis of ordinal data structures
University of Leiden, 196- (a)

De Leeuw, J. Multivariate analysis of sets of contingency tables
University of Leiden, 196- (b)

Fisher, R.A. Statistical methods for research workers
Edinburgh, Oliver & Boyd, 1954

Foa, U.G. New developments in facet design and analysis
Psychological review, 1965, 72, 262-275

Guttman, L. An outline of the statistical theory of prediction
In: Horst, P. (ed): The prediction of personal adjustment. New York, Social Science Research Council, 1941, (a), 253-312

Guttman, L. The quantification of a class of attributes: a theory and method of scale construction
In: Horst, P. (ed): The prediction of personal adjustment. New York, Social Science Research Council, 1941, (b), 319-348

Guttman, L. A basis for scaling qualitative data
American Sociological Review, 1944, 9, 139-150

Guttman, L. An approach for quantifying paired comparisons and rank order
Annals of mathematical statistics, 1946, 17, 144-163

Guttman, L. The relation of scalogram analysis to other techniques
In: Stouffer, S.A. et al: Measurement and Prediction.
Princeton, Princeton University Press, 1950 (a) 172-213

Guttman, L. The principal components of scale analysis
In: Stouffer, S.A. et al: Measurement and Prediction
Princeton, Princeton University Press, 1950 (b) 312-361

Guttman, L. A note on Sir Cyril Burts 'Factorial analysis of qualitative data '
British Journal of Statistical Psychology, 1953, 6,1-4

Guttman, L. The principal components of scalable attitudes
In: Lazarsfeld, P.F. (ed): Mathematical Thinking in the Social Sciences, Glencoe, Free Press, 1954

Guttman, L. An introduction to facet design and analysis
In: Proceedings of the 15th International Congress of Psychology. Brussels, 1959 (a), p 130-132

Guttman, L. The structering of sociological spaces
In: Transactions of the fourth world congress of sociology. International Sociological Association, 1961, 3, 315-355

Guttman, L. The nonmetric breakthrough for the behavioural sciences
In: Proceedings of the second national conference on data processing. Information Processing Association of Israel, 1966 (a), 495-510

Guttman, L. Order analysis of correlation matrices
In: R.B. Cattell (ed): Handbook of multivariate experimental psychology. New York, Rand-McNally, 1966 (b), 438-458

Guttman, L. The development of nonmetric space analysis, letter to professor John Ross
Multivariate Behavioural Research, 3, 1967, 71-83

Guttman, L. A general nonmetric technique for finding the smallest Euclidean space for a configuration of points
Psychometrika, 196-

Kendall, M.G. Rank correlation methods
London, Griffin, 1962

Kendall, M.G. & Stuart, A. The advanced theory of statistics
Three volumes, London, Griffin, 1963, 1967, 1966

Krantz, D.H. A survey of measurement theory
Michigan Mathematical Psychology Program, MMPP 67-4

Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypotheses
Psychometrika, 29, 1964, 1-27

Kruskal, J.B. Multidimensional scaling: a numerical method
Psychometrika, 29, 1964, 115-129

Lingoes, J.C. Multiple Scalogram Analysis: a set theoretic model for analyzing dichotomous items
Educational and Psychological Measurement, 23, 1963 (a), 501-524

Lingoes, J.C. Multivariate analysis of contingencies
Computational Reports, University of Michigan Computing Center, Ann Arbor, 1963 (b)

Lingoes, J.C. New computer developments in pattern analysis and nonmetric techniques
In: Proccedings of the IBM-symposium 'Computers in psychological research', Gauthiers-Villars, Paris, 1967

Lingoes, J.C. The multivariate analysis of qualitative data
Multivariate Behavioural Research, 196-

Lingoes, J.C. & Vandenberg, S.G. A nonmetric analysis of twin data based on a multifaceted design
Report No 17, Louisvill Twin Study Development Unit, Department of pediatrics, University of Louisville, 1966

Lubin, A. Linear and non-linear discriminant functions
British Journal of Statistical Psychology, 1950, 3, 90-104

Pearson, K. On lines and planes of closest f it to points in space
Philosophical Magazine, 2, 1901, 557-572

Scott, D. & Suppes, P. Foundational aspects of theories of measurement
Journal of symbolic logic, 1958, 23, 113-128

Shepard, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function, I & II
Psychometrika, 1962, 27, 125-140, 219-246

Shepard, R.N. Metric structures in ordinal data
Journal of mathematical psychology, 1966, 3, 287-315

Shepard, R.N. & Kruskal, J.B. Nonmetric methods for scaling and factor analysis
American Psychologist, 1964, 19, 557-558 (abstract)

Slater, P. The analysis of personal preferences
British Journal of Statistical Psychology, 1960, 13, 119-135

Suppes, P. & Zinnes, J. Basic measurement theory
In Luce, R.D., Bush, R.R. & Galanter, E. (eds): Handbook of mathematical Psychology. volume I, New York, Wiley, 1963, 1-72

Tucker, L.R. Description of paired comparisons preference judgments by a multidimensional vector model
Princeton, Educational Testing Service, Research Memo 55-7

Tucker, L.R. Learning theory and multivariate experiment: illustration by determination of generalized learning curves
In Cattell, R.B. (ed): Handbook of multivariate experimental psychology. New York, Rand-McNally, 1966, 476-501

Tucker, L.R. & Messick, S. An individual differences model for multidimensional scaling
Psychometrika, 1963, 28, 333-367

Tversky, A. Finite additive structures
Michigan Mathematical Psychology Program, MMPP 64-6

Tversky, A. The optimum number of alternatives at a choice point
Journal of Mathematical psychology, 1964, 1, 386-391

Tversky, A. A general theory of conjoint measurement
Journal of Mathematical Psychology, 1967, 4, 1-20

Wish, M. A facet-theoretic approach for morse code
Michigan Mathematical Psychology Program, MMPP 65-6

Yule, G.U. An Introduction to the theory of statistics
London, Griffin, 1910