RB - 002 - 71

Jan de Leeuw

CANONICAL ANALYSIS OF CONTINGENCY TABLES

RB - 002 - 71

1.1 Theory

Suppose X and Y are ordered sets of real numbers, and $F(x,y)$ is a probability distribution function on $X \times Y$. If X and Y are discrete we mean by $p(x,y)$ the probability measure of the pair (x,y), if $F(x,y)$ has continuous derivatives of first and second orders almost everywhere then $p(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$. The symbol $\int$(expression) d(variable) is used both for integration and summation. Let

$$p_1(x) = \int p(x,y)dy, \tag{1a}$$

$$p_2(x) = \int p(x,y)dx. \tag{1b}$$

We look for measurable real valued transformations ϕ (on X) and ψ (on Y) such that the PM-correlation between the transformed variates is as large as possible. Consequently we want to find

$$\sup_{\phi,\psi\in\Omega} \iint \phi(x)\,\psi(y)\; p(x,y)\; dxdy, \tag{2}$$

under the conditions that

$$\int \phi^2(x)\; p_1(x)\; dx = \int \psi^2(y)\; p_2(y)\; dy = 1, \tag{3a}$$

$$\int \phi\,(x)\; p_1(x)\; dx = \int \psi\,(y)\; p_2(y)\; dy = 0. \tag{3b}$$

This is a simple calculus of variations problem for which the stationary equations are (using only the conditions 3a)

$$\int p(x,y)\;\phi\,(x)\; dx = \mu\,\psi(y)\; p_2(y), \tag{4a}$$

$$\int p(x,y)\;\psi\,(y)\; dy = \mu\,\phi(x)\; p_1(x), \tag{4b}$$

where μ is an Euler-multiplier.

We can rewrite these equations, alternatively, as operator equations on the (complete, separable, real, Hilbert) space of all measurable functions.

$$B(\phi) = \mu D_2(\psi), \tag{5a}$$

$$B^*(\psi) = \mu D_1(\phi). \tag{5b}$$

Here B^* is the adjoint of B, D_1 and D_2 are strictly positive and self-adjoint, all operators are linear and bounded. Suppose η satisfies

$$D_1^{-\frac{1}{2}} \circ B^* \circ D_2^{-1} \circ B \circ D_1^{-\frac{1}{2}}\,(\eta) = \mu^2\,\eta, \tag{6}$$

then $\phi = D_1^{-\frac{1}{2}}\,(\eta)$ satisfies

$$B^{*} \circ D_2^{-1} \circ B(\phi) = \mu^2 D_1(\phi), \tag{7a}$$

and $\psi = \mu^{-1} D_2^{-1} \circ B(\phi)$ satisfies

$$B \circ D_1^{-1} \circ B^{*}(\psi) = \mu^2 D_2(\psi). \tag{7b}$$

The triple (ϕ, ψ, μ) satisfies (5a) and (5b). There are, in general, a denumerably infinite number of solutions of (6), denoted by η_i with corresponding characteristic values μ_i^2. Suppose $\mu_i^2 \neq \mu_j^2$ for $i \neq j$. Then the ϕ_i corresponding with η_i are D_1-orthogonal, the corresponding ψ_i are D_2-orthogonal. It follows from (4) and (5) that $\phi_0 \equiv 1$ and $\psi_0 \equiv 1$ give a solution of (4) with $\mu_0 = 1$, satisfying (3a) but not (3b). All other solutions to (4) satisfy $\langle D_1(\phi_i), \phi_0 \rangle = \langle D_2(\psi_i), \psi_0 \rangle = 0$, which is precisely (3b). Consequently (4) has one 'improper' solution, all other solutions satisfy (3b) automatically. Obviously $0 \leq \mu_i \leq 1$ for all $i = 1, 2, 3, \ldots$ Moreover, from (6),

$$\sum_{i=1}^{\infty} \mu_i^2 = \iint \frac{p^2(x,y)}{p_1(x)p_2(y)}\, dx\, dy - 1 = \varphi^2(x,y), \tag{8a}$$

where φ^2 is Pearsons contingency measure. Moreover

$$p(x,y) = p_1(x)p_2(y)\,(1 + \sum_{i=1}^{\infty} \mu_i \phi_i(x)\psi_i(y)). \tag{8b}$$

The regression of $\phi_i(x)$ on $\psi_i(y)$ (and of $\psi_i(y)$ on $\phi_i(x)$) is linear, with regression lines

$$\phi_i(x) = \mu_i \psi_i(y), \tag{8c}$$

$$\psi_i(y) = \mu_i \phi_i(x). \tag{8d}$$

In practical situations we are dealing with samples in which the values of X and Y are either discrete or grouped into intervals. B is an $n \times m$ matrix, and we may suppose without loss of generality that $m \leq n$. The marginal frequencies are collected in the diagonal matrices D_1 and D_2, the total number of observations is N. Again we suppose that none of the marginal frequencies is zero. The quantifications ϕ and ψ are simply an n-element vector x and an m-element vector y. The stationary equations are

$$By = \mu D_1 x, \tag{9a}$$

$$B'x = \mu D_2 y. \tag{9b}$$

Again we find any m-element vector z such that

$$D_2^{-\frac{1}{2}}B'D_1^{-1}BD_2^{-\frac{1}{2}}\, z = \mu^2 z, \tag{10}$$

and set $y = D_2^{-\frac{1}{2}} z$, $x = \mu^{-1}D_1^{-1}B\, y$. If we let $z'z = \mu$, then $x'D_1x = y'D_2y = \mu$. The m stationary values μ_i^2 $(i = 0,\ldots,m-1)$, ordered in decreasing order of magnitude, satisfy $\mu_0^2 = 1$ (the improper solution), and $N\sum_{i=1}^{m-1}\mu_i^2$ equals Pearson's X^2 for the contingency table B. Asymptotically X^2 is distributed as χ^2 with $K = (n-1)(m-1)$ df on the hypotheses of independence. We are not intested in independence, and we can use the following approximate test. Compute

$$X_p^2 = N\sum_{i=p}^{m-1}\mu_i^2, \tag{11}$$

for $p=1,\ldots,m-1$, and compare it with a χ^2 with K df. As soon as prob$(\chi_K^2 > X_p^2)$ π, for some critical value π, we can 'stop factoring'. [redacted]

There are several alternative approaches to the problem which are also useful. The first one uses the indicator matrices S and T of orders n x N, and m x N respectively. Both S and T are binary matrices with $s_{ik} = 1$ iff the X-value of sample element k is x_i ($t_{jk} = 1$ iff the Y-value of sample element k is y_j). By assigning numerical weights (a vector x) to the elements of X we get an induced quantification u of the sample, by assigning numerical weights (a vector y) to the elements of Y we get another induced quantification v, by

$$u = S'x, \tag{12a}$$

$$v = T'y. \tag{12b}$$

We now interpret the N x 2 matrix Z, containing the two columns u and v, as a one-way classification with two levels. The ANOVA-table is

Source	SSQ	df
Between	$2\sum_{k=1}^{N} u_k v_k$	1
Within	$\sum_{k=1}^{N}(u_k - v_k)^2$	2N-2
Total	$\sum_{k=1}^{N}u_k^2 + \sum_{k=1}^{N}v_k^2$	2N-1

(13)

(provided, of course, that $\sum u_k = \sum v_k = 0$). In matrix notation

$$SSQ_B = 2\, x'ST'y, \tag{14a}$$

$$SSQ_W = x'SS'x + y'TT'y - 2\, x'ST'y, \tag{14b}$$

$$SSQ_T = x'SS'x + y'TT'y. \tag{14c}$$

We maximize SSQ_B under the condition that SSQ_T equals some positive constant. The stationary equations are

$$ST'y = \mu SS'x, \tag{15a}$$

$$TS'x = \mu TT'y. \tag{15b}$$

Because both SS' and TT' are diagonal, this is exactly identical to our previous problem (9). There also is an improper solution, and we are also dealing with orthogonal decomposition of the X^2 computed for ST'. We also obtain induced quantifications u and v. This is nothing new. The U-score for a sample element is simply the corresponding X-score, the V-score the corresponding Y-score. If we want to characterize a sample element by one single score the ANOVA-context demands an additive combination of u_i and v_i. This can also be seen by considering the supermatrix $\begin{bmatrix} S \\ \cdots \\ T \end{bmatrix}$ itself as an $(n + m) \times N$ contingency table E. We quantify the sample facet by an N-element vector w, and the facets X and Y by a supervector $(\; x \;\vdots\; y \;)$. Call the maximum correlation λ. The stationary equations (9a) and (9b) reduce to

$$w = \tfrac{1}{2}\lambda^{-1}(S'x + T'y), \tag{16a}$$

$$ST'y = (2\lambda^2 - 1)\, SS'x, \tag{16b}$$

$$TS'x = (2\lambda^2 - 1)\, TT'y. \tag{16c}$$

From this it follows that again x and y are the same as previously,, the vector w is proportional to the average of u and v.

It is also quite instructive to consider a geometrical approach to the same problem. We start with a contingency table B (or, more generally, a bivariate probability density), and define a pseudo-metric between the elements of Y by

$$d^2_{Y(X)} = \int p_1(x) \left[\frac{p(y/x)}{p_2(y)} - \frac{p(y'/x)}{p_2(y')} \right]^2 dx. \tag{17}$$

This is the distance between y and y', as measured by X. It is easy to see that

$d^2_{Y(X)}(y,y') = 0$ iff $p(x/y) = p(x/y')$ for all $x \in X$ (assuming again that $p_1(x) \neq 0$ for all $x \in X$). If $p(x/y) = p(x/y')$ for all $x \in X$, then we can construct a new point y'', define $p(x,y'') = p(x,y) + p(x,y')$, and find $d^2_{Y(X)}(y'',z) = d^2_{Y(X)}(y,z) + d^2_{Y(X)}(y',z)$ for all $z \in Y$. Thus we can bring Y in 'canonical form', such that $d^2_{Y(X)}(y,y') = 0$ iff $y = y'$. That the triangle inequality is satisfied can be seen most easily if we define $Q_{Y(X)}$ as

$$Q_{Y(X)} = D_2^{-1}B'D_1^{-1}BD_2^{-1}. \tag{18}$$

Introduce the m-element unit vectors e_i and e_j. Then

$$d^2_{Y(X)}(y_i,y_j) = (e_i - e_j)'Q_{Y(X)}(e_i - e_j). \tag{19}$$

Because $Q_{Y(X)}$ is positive semi-definite, $d_{Y(X)}$ is a metric. Equations (18) and (19) are written in terms of matrices, we can also use general linear operators, of course. Let $K\Lambda K'$ be the canonical form of $Q_{Y(X)}$, and define the vector y_i by $y_i' = e_i'K\Lambda^{\frac{1}{2}}$. If δ^2_{ij} denotes the squared Euclidean distance between the endpoints of the y_i, then $\delta^2_{ij} = d^2_{Y(X)}(y_i,y_j)$. Again we can disregard the small roots, and so on. A more satisfactory result may be reached if we give each point e_i the mass (weight) $p_2(y_i)$, and then reduce to principal axes. We need an inertia-matrix M_Y defined by

$$M_Y = D_2 - D_2ee'D_2, \tag{20}$$

(e all elements equal to unity), and we look for the right eigenvectors of $W_{Y(X)} = Q_{Y(X)}M_Y$. But

$$\begin{aligned} W_{Y(X)} &= D_2^{-1}B'D_1^{-1}BD_2^{-1}\ (D_2 - D_2ee'D_2) = \\ &= D_2^{-1}\ (B'D_1^{-1}B - D_2ee'D_2). \end{aligned} \tag{21}$$

And consequently $W_{Y(X)}y = \lambda y$ iff

$$(B'D_1^{-1}B - D_2ee'D_2)y = \lambda D_2\, y, \tag{22}$$

iff

$$B'D_1^{-1}By = \hat{\lambda}\, D_2\, y. \tag{23}$$

The equations (22) and (23) differ only in that (23) has an improper solution, which is removed by simple Hotelling-type deflation in (22). The same way of reasoning gives a distance $d^2_{X(Y)}$, with quadratic form $Q_{X(Y)}$, an inertia matrix M_X, and the stationary equations

$$BD_2^{-1}B'x = \hat{\lambda} D_1 x \ . \tag{24}$$

Again the solutions are obviously the same as those of (9), (15), and (16). It can be proved (along the lines of equation (16)) that it is perfectly legitimate to draw the plots of X and Y in a joint space.

1.2 Historical

The history of our problem is quite complicated. We have to start with Karl Pearson, who investigated the relation between correlation and contingency as early as 1900. His investigations are contained in (17, 18, 19). In 1935 Hirschfeld (7) studied the linearizing approach for the regressions in a contingency table (generalized in our formulas 8c and 8d). This resulted in the idea of the maximal correlation in a bivariate distribution, which was generalized to the continuous case by Gebelein (5), compare also Renyi (20) and Richter (21). Yates (23) investigated the assignment of scores to the marginals and gives a significance test for the maximised correlation when one set of scores (say y) is known. In this case, of course,

$$x \propto D_1^{-1}By \tag{25}$$

gives the unique maximum. The approach using the homogeneity ideas (and the ANOVA-terminology) is due to Fisher (4), who also found the identity (8b) for the finite case. Fisher's approach was further extended by Maung (15), Bartlett (1), and Williams (22). It was proved by Maung and, with more satisfactory methods, by Lancester (11), that for the binormal distribution with correlation parameter ρ

$$\mu_i = |\rho|^{2i} , \tag{26}$$

while the corresponding eigenelements are the Hermite-Chebyshev polynomials. As a corrolary

$$\varphi^2 = \sum_{i=1}^{\infty} \mu_i^2 = \frac{\rho^2}{1 - \rho^2} , \tag{27}$$

a result already proved by Pearson (18). The ANOVA-rationale was discovered, independently, by Guttman (6) using correlation ratio terminology, and extended and used by him in many subsequent papers. He also discussed the canonical partitioning of X^2, which has been developed more extensively by Lancaster

(9, 10, 13). A generalization and summary of the previous results is given in Lancester (12), an easily accessible summary of most of the work in this field can be found in Kendall & Stuart (8). Burt (2), Lingoes (14), and McDonald (16) are also useful review papers. The geometrical approach to the problem is due to Jean Paul Benzécri, and his methods have been published most extensively in the thesis of Cordier (3). The equivalence of this 'analyse de correspondences' with our problem was pointed out to me by Matthijs Koornstra. The proof is mine. Computer programs for the finite case have been written by Lancester, Lingoes (under the name of MAC), and Cordier.

1.3 Applications

The 'cross-table' and the associated X^2 are used very frequently in social science research. For some reason or another the partitioning of the total X^2 in additive components (which can give a lot of extra information at a relatively cheap price) is hardly practiced at all. This is a regrettable situation. The canonical partitioning of X^2 discussed in this chapter is very nice in the sense that we have a satisfactory geometrical interpretation, and consequently we can make nice plots of two dimensional projections. It is unsatisfactory because the canonical components of X^2 are not (asymptotically) independent χ^2-variates (in other types of partitions they are), and significance tests are difficult to obtain. Personally, I am quite sceptical about the usefulness of classical hypothesis testing in typical multivariate data reduction procedures, and consequently I do not believe that this is a very serious objection. Moreover the exact partitionings of Lancester-Irwin or of Kullback-McGill have unsatisfactory aspects of their own.

The procedure can be applied to any contingency table, and will always give supplementary information. Only in some rare cases (for example $p(x,y) = p_1(x) = p_2(y)$ for all x,y or p(x,y) is constant on X x Y) the outcome is not interesting from a practical point of view. Observe that in this method we do not have to be afraid for departures from normality. In fact it can be argued that these departures actually make the method worthwhile. In the binormal case we can always reproduce our data - within sampling and grouping errors- from one component. It may not even be such a bad idea to treat (grouped) quantitative data in this way (we partition the correlation coefficient in orthogonal components).

In our theoretical treatment we did not aime at maximal generality. Our use of the integral sign indicates that, in an informal manner, we have used Stieltjes integrals without going into the notational complications this implies, for example in defining $\varphi^2(x,y)$. It is possible to generalize even further, forget about the probabilistic interpretation, and deal with general product measures on X x Y. In the finite case we can handle nonnegative correlation matrices (our method reduces to principal component analysis of the first centroid residual matrix), we can handle symmetric similarity matrices, ratio-estimation data, confusion matrices, m rankings of n objects, and so on. In the examples we shall try to give some indication of the applicability of the method.

1.4 References

1 Bartlett, M.S. The goodness of fit of a single hypothetical discriminant function in the case of several groups. Ann. Eugen. London, 16, 1951, 199-

2 Burt, C. The factorial analysis of qualitative data. Brit. J. Statist. Psychol., 1950, 3, 166-

3 Cordier, B. L'analyse factorielle des correspondences Thèse de troisième cycle, Faculté des sciences de l'universite de Rennes, mai 1965.

4 Fisher, R.A. The precision of discriminant functions Ann. Eugen. London, 10, 1940, 422-

5 Gebelein, H. Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit Ausgleichsrechnung. Z. angew. Math. Mech., 21, 1941, 364-

6) 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.

7 Hirschfeld, H.O. A connection between correlation and contingency.
Proc. Camb. Phil. Soc., 31, 1935, 520-

8 Kendall, M.G. & Stuart, A. The advanced theory of statistics.
Volume II, second edition, Ch. 33, London, Griffin 1967.

9 Lancester, H.O. The derivation and partition of X^2 in certain discrete distributions
Biometrika, 31, 1949, 370-

10 —— A reconcillation of χ^2 considered from metrical and enumerative aspects.
Sankhyā, 13, 1953, 1-

11 —— Some properties of the bivariate normal distribution considered in the form of a contingency table.
Biometrika, 44, 1957, 289-

12 —— The structure of bivariate distributions.
Ann. Math. Statist., 29, 1958, 719-

13 —— Canonical correlations and partitions of X^2.
Quart. J. Math., Oxford, 14, 1960, 220-

14 Lingoes, J.C. The multivariate analysis of qualitative data.
Multivariate Behavioural Research, 3, 1968, 61-

15 Maung, K. Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children.
Ann. Eugen. London, 11, 1941, 189-

16 McDonald, R.P. A unified treatment of the weighting problem.
Psychmetrika, 33, 1968, 351-

17 Pearson, K. Mathematical contributions to the theory of evolution VII. On the correlation of characters not quantitatively measurable.
Philos. Trans. Roy. Soc. (London), 195A, 1900, 1-

18 —— Mathematical contributions to the theory of evolution XIII. On the theory of contingency and its relation to association and normal correlation.
Drapers co. Research memoirs, Biometric series, I, 1904.

19 —— On the measurement of the influence of broad categories on correlation.
Biometrika 9, 1913, 116-

20 Rényi, A. On measures of dependence.
Acta Math. Acad. Sci. Hung., 10, 1959, 441-

21 Richter, H. Zur Maximalcorrelation.
Z. angew. Math. Mech., 29, 1949, 127-

22 Williams, E.J. Use of scores for the analysis of association in contingency tables.
Biometrika, 39, 1952, 274-

23 Yates, F. The analysis of contingency tables with grouping based on quantitative characters.
Biometrika, 35, 1948, 176-

1.5 Examples

We shall first give some examples of our principle used as a one dimensional scaling method. Consider the data on page 203 of Guilford's Psychometric Methods (spot patterns placed in nine succesive categories spaced at equal-appearing intervals). The squared correlations given by our method are

$$\mu_1^2 = .861 \qquad \mu_5^2 = .086$$
$$\mu_2^2 = .638 \qquad \mu_6^2 = .029$$
$$\mu_3^2 = .384 \qquad \mu_7^2 = .010$$

$$\mu_4^2 = .135 \qquad \mu_8^2 = .007$$

At least for the first four roots the equations $\mu_i^2 = |\rho|^{2i}$ are satisfied quite nicely for $\rho \approx .93$. Consequently we may assume that we are dealing with a sample from a normal distribution, and the most informative transformation we can apply to the marginals is our first principal component. In figure Ia we have plotted the values of the first left component as a function of the number of spots in the pattern, in Ib we have a plot of the right component vs the category number. Observe the relative underestimation of the intervals between extreme stimuli, observe that the regression in figure Ia is linear, and not logarithmic.

As a second example we have analyzed Guilford's data one preferences for movie actors obtained by the method of rank order (lc p 180). The squared correlations were (first six only)

$$\mu_1^2 = .182 \qquad \mu_4^2 = .019$$
$$\mu_2^2 = .062 \qquad \mu_5^2 = .017$$
$$\mu_3^2 = .025 \qquad \mu_6^2 = .012$$

The maximal correlation is about .43. There are serious departures from normality A joint two-dimensional plot of actors and rank numbers is shown as figure IIa. The second degree polynomial is there (especially for the ranks), but obviously actor KD deviates strongly from this pattern. Inspection of the data shows that his conditional distribution of rank numbers is clearly bimodal, even somewhat U-shaped (modi at 15 and 1,2,3). The rank numbers and their scale values are plotted separately in IIb (first dimension only). There are only two violations of the expected monotonic order out of a possible 105 ones. Tentatively we may conclude that there is an acelleration at the extreme of the scale, which justifies, more or less, Guilford's normalized rank method. In IIc we compare the results: they are virtually identical. Nevertheless our method seems better. Guilford's method assumes a binormal distribution, and then applies a fixed transformation to the rank numbers in order to find the scale values (essentially by using equation 25). We optimize normality by scaling the marginals, and we find additional information about the nature of the deviations (and possibly quite sensible multidimensional solutions). The correlation coefficient in this case is simply an index of agreement between subjects (for a particular dimension). Consequently our scale is the dimension on which the subjects agree most. There is a way to quantify subjects which is consistent with this model. Each subject k defines a permutation matrix P_k, our data matrix B is the sum of these permutation matrices. Define (for a particular dimension) $s_k = z'P_k r$, then $\sum s_k = \sum z'P_k r = z'Br = \mu$, with z the scale values and r the rank number scores from that dimension. Because D_1 and D_2 are scalar matrices, s_k is proportional to the correlation between $P_k r$ and z.

In our next example, formally identical, the usefulness of our method is even clearer. The data (collected by Dr. van der Kamp) consist of the rankings of nine Dutch political parties by 100 students. The data matrix B is

KVP	07	05	14	11	21	15	15	11	01
PVDA	19	19	16	10	06	13	11	04	02
VVD	31	12	06	14	16	11	07	03	00
ARP	05	18	18	16	21	16	05	01	00
CHU	04	05	11	17	22	17	12	10	02
CPN	02	00	05	06	08	09	14	29	27
PSP	08	06	09	09	06	10	23	23	06
BP	00	01	00	01	04	07	12	15	60
D'66	23	33	19	15	02	03	02	03	00

It is interesting to look at the conditional distributions defined by the rows of this matrix. The three denominational parties (KVP, CHU, ARP) have flat, symmetric distributions. The liberals (VVD) and socialists (PVDA) have ~~kinumink~~ bimodal distributions, indicating that our subjects are not homogeneous, there are leftist students who dislike the VVD and rightist who dislike the PvdA. Both groups do not like the denominational parties very much, while D'66 is popular with both groups. Finally everybody strongly dislikes the farmer's union (extreme right, BP), and the communists (CPN). The pacifistic socialist have some sympathy in the PVDA group. Our analysis of this table (which has a total X^2 of 676.8) yields two (possibly three) significant components with

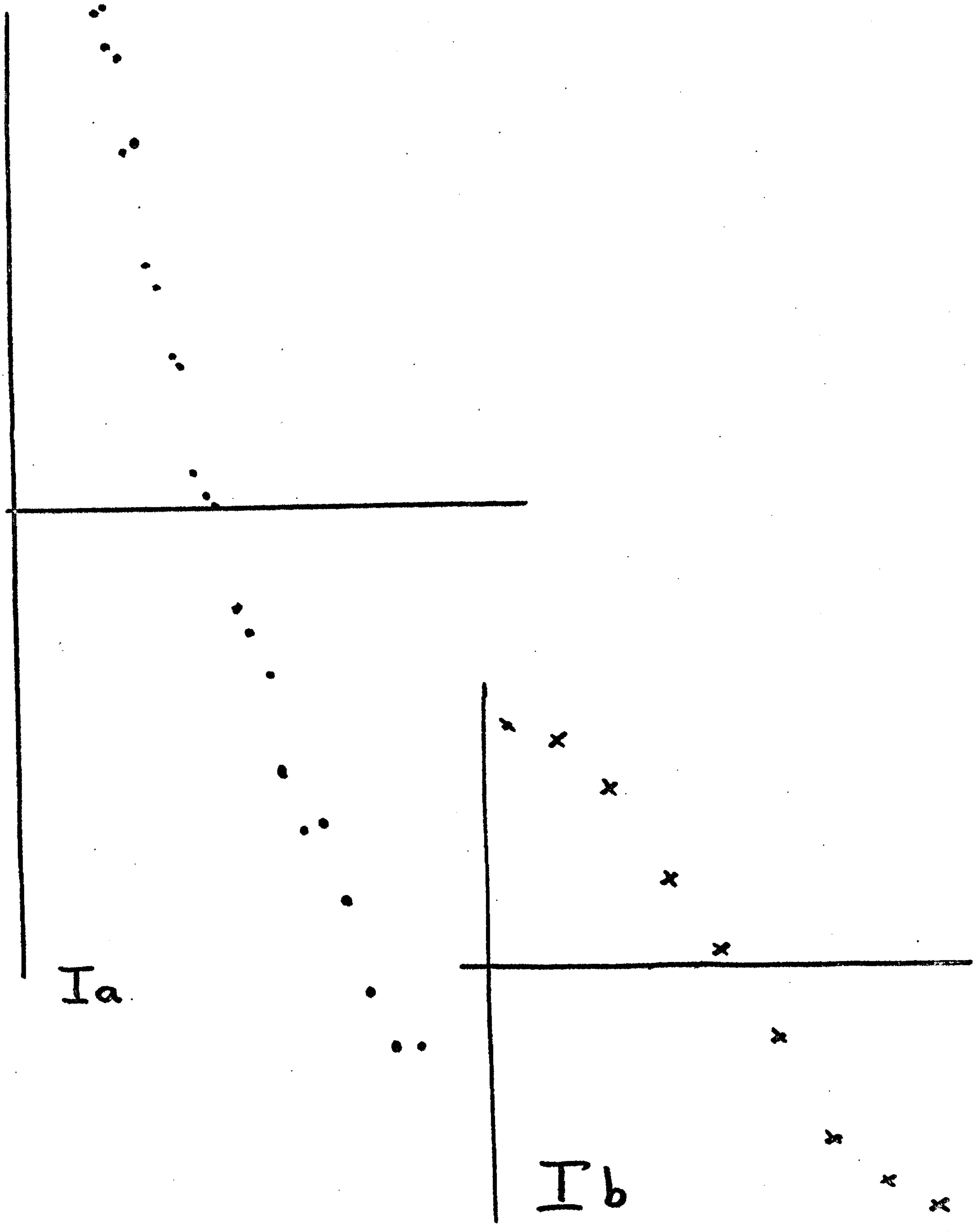
$$\mu_1^2 = .490$$
$$\mu_2^2 = .140$$
$$\mu_3^2 = .064$$

The plots comparable to IIa and IIb are given in IIIa and IIIb. Observe that the transformation applied to the rank numbers is nonlinear (but monotonic) in this case (this is due tothe 'crowding' of stimuli on the most preferred end of the scale), and observe that first-choice behaviour is not very representative for the scale in this case (which is a conclusion of some political interest). As the best preference scale for this example I would take the projections on the parabola of figure IIIa, not the projections on the first axis. Tentatively we may also conclude that the preference behaviour of our subjects towards the governmental coalition (KVP, CHU, ARP, VVD) is somewhat diffeeent from their behaviour towards the opposition. The two groups mentioned previously may result in two different curved scales.

For our final example we use data collected by the Dutch student council (NSR) in 1969. The data were made available to me by Prof. Lammers. Essentially they consist of a three-way contingency table having the facets universities, political parties, faculties. The two marginal tables which interest us here are given below. For the first table (X^2 = 177.7) we find two significant components (μ_1^2 = .067, μ_2^2 = .030). Joint plot in figure IVa. The first dimension is simply right-left. The order of the projections corresponds closely with the one computed by Lammers using preassigned scores for parties (only theology moves to the left). The second dimension is denominational (veterinarians and agricultural scientists are more inclined to vote for KVP/ARP/CHU, perhaps bacause they come from rural areas). This explains immediately why the scale value of theology is changed, compared with the one computed by Lammers. The analysis of the other table (X^2 = 202.9) gives μ_1^2 = .070, μ_2^2= .041. In figure IVb we see that the interpretation of the two dimensions is the same, only their order of importance is interchanged. Because a precise interpretation of the figures requires a knowledge of Dutch universities, politics, and so on, we skip it. This example mainly shows how useful the technique is when there is virtually no trace of normality, and when the μ_i^2 are low.

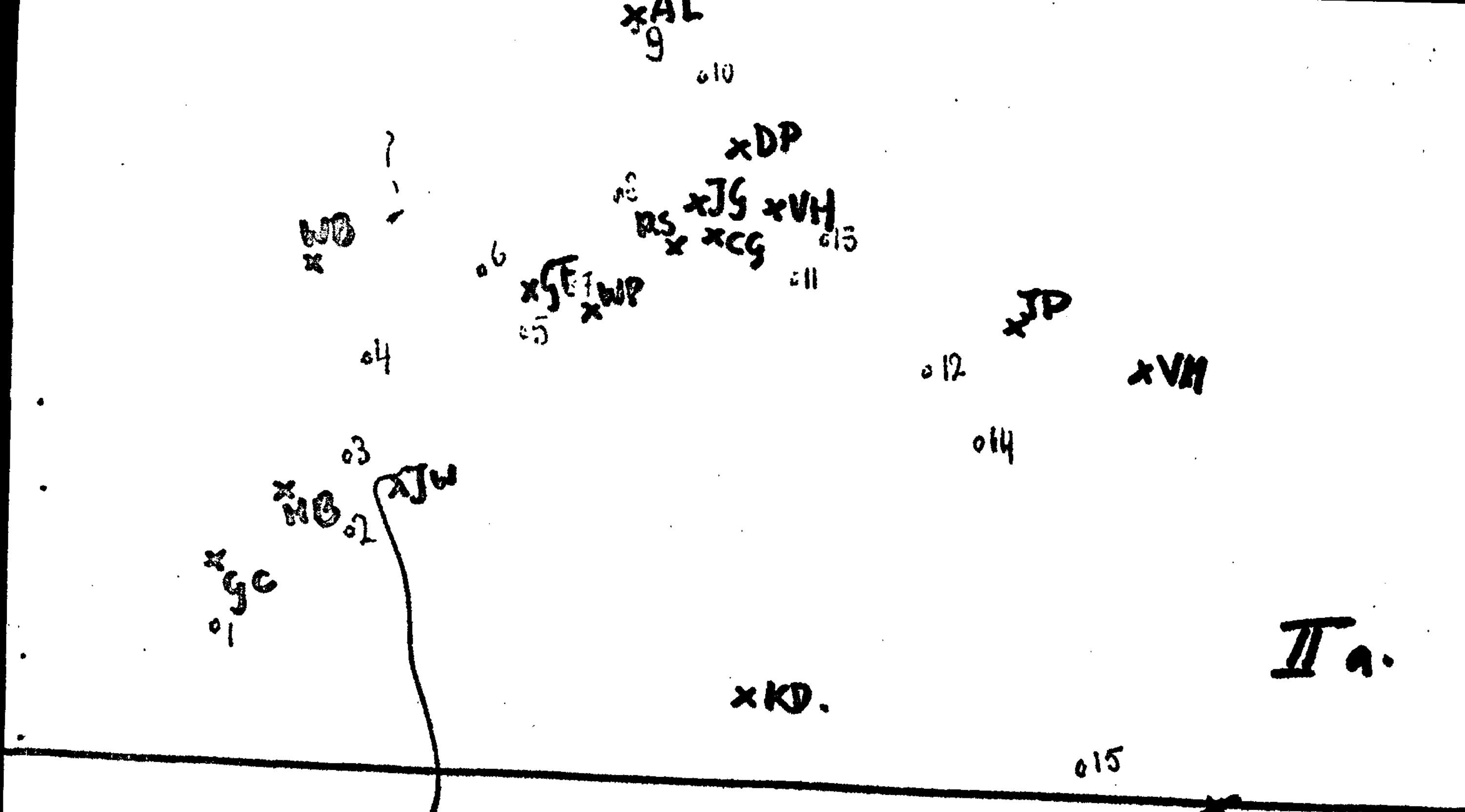
	conf	vvd	pvda	cpn psp	d'66	tot
DELFT	24	66	22	20	50	182
EINDH	12	07	03	00	13	035
RODAM	20	43	20	04	24	111
TBURG	08	12	03	03	17	043
NYGEN	33	22	22	13	50	140
DRIEN	00	07	03	04	05	019
GRONI	20	46	43	14	55	178
WAGEN	11	15	08	06	15	055
UTREC	40	65	36	27	72	240
ADAMG	27	59	92	43	113	334
ADAMV	27	01	05	03	09	045
LEIDE	29	88	34	21	62	234
TOTAAL	251	431	291	158	485	1616

	conf	vvd	pvda	cpn psp	d'66	tot
JUR	33	79	33	06	49	200
MED	37	61	30	19	64	211
W&N	33	40	45	22	74	214
SOC	28	34	58	38	85	243
LET	22	31	40	20	44	157
TEC	36	80	28	24	68	236
PSW	01	01	05	00	04	011
VEE	03	06	01	01	05	016
TND	02	07	00	01	04	014
THE	19	01	05	04	07	036
LBW	11	14	07	06	14	052
CIF	02	04	09	09	12	036
ECO	24	73	30	08	55	190
TOT	251	431	291	158	485	1616

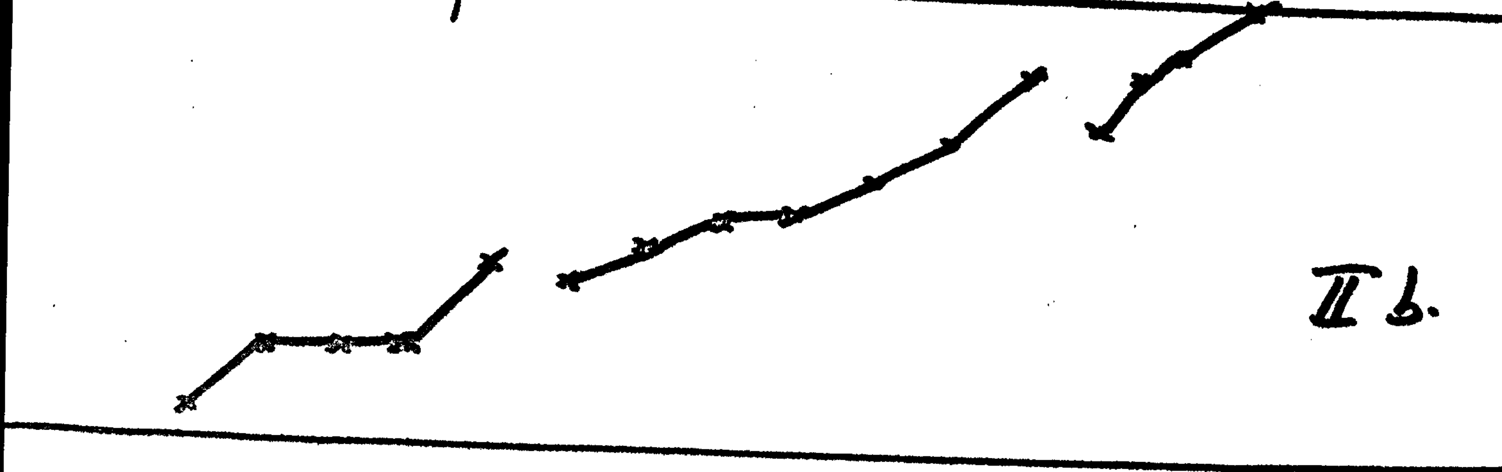


Ia

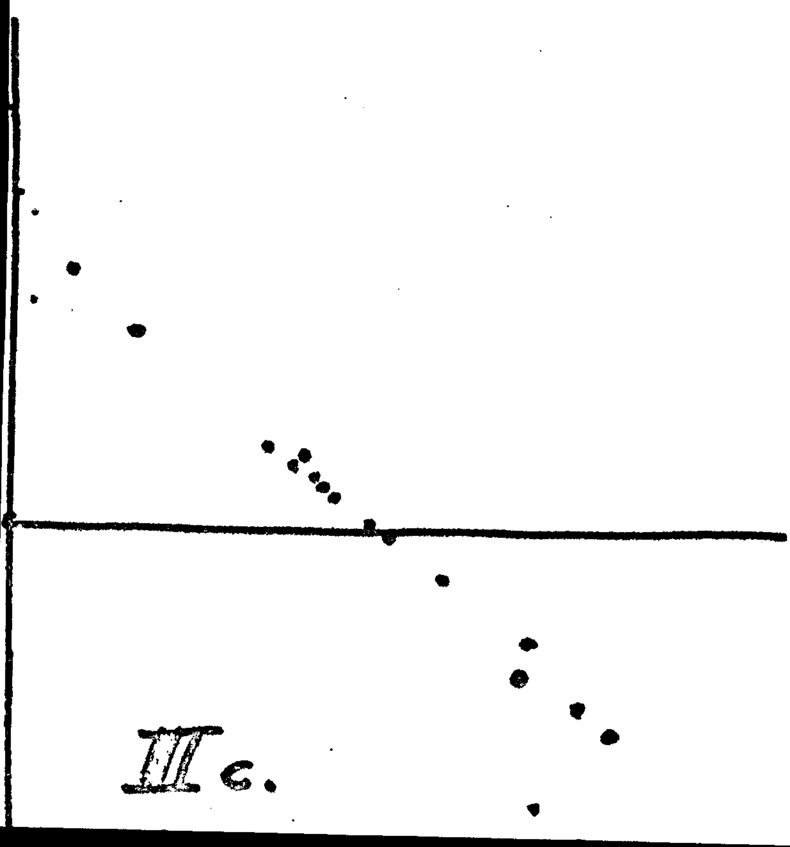
Ib



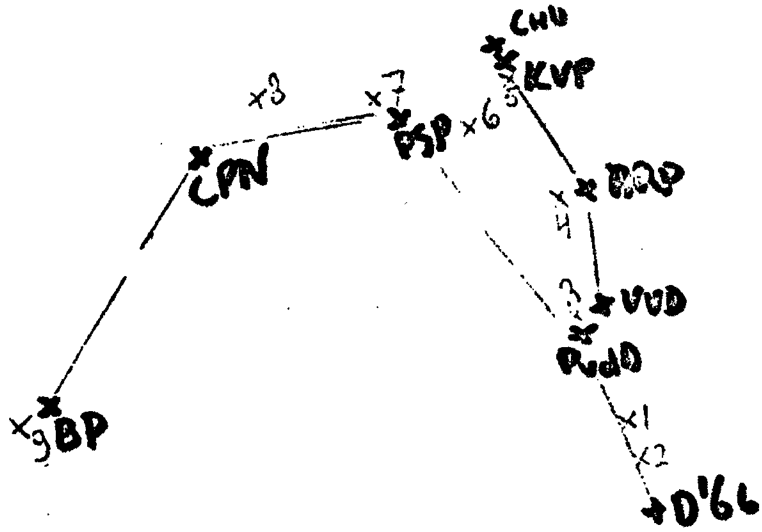
II a.



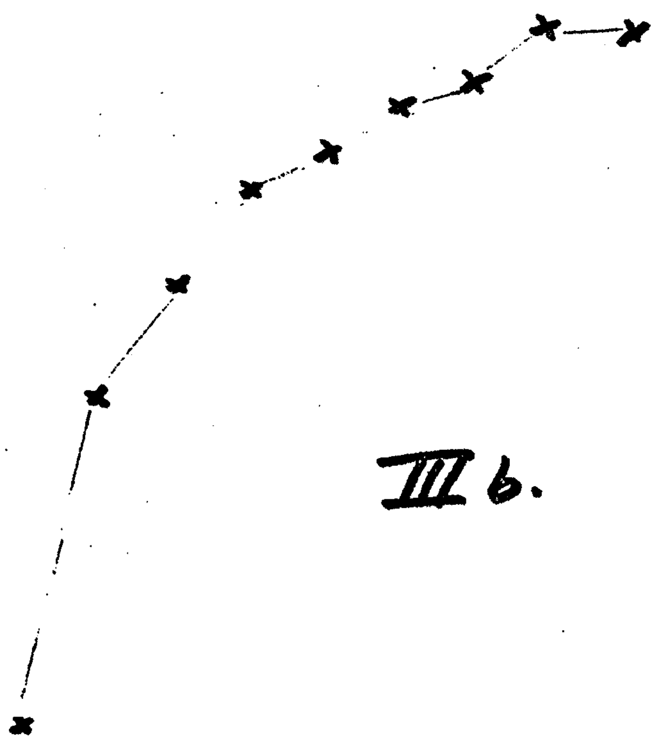
II b.



III c.

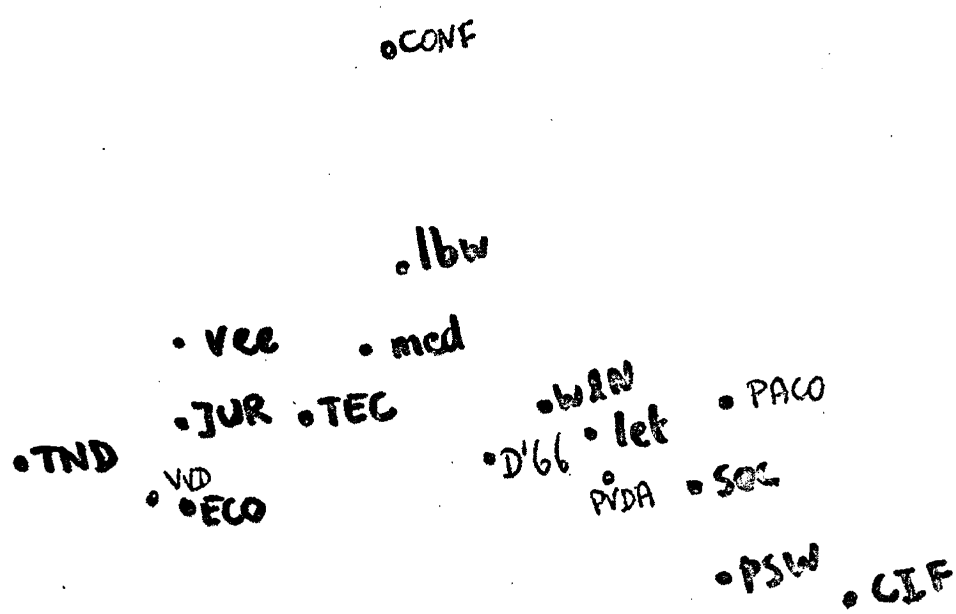


III a.

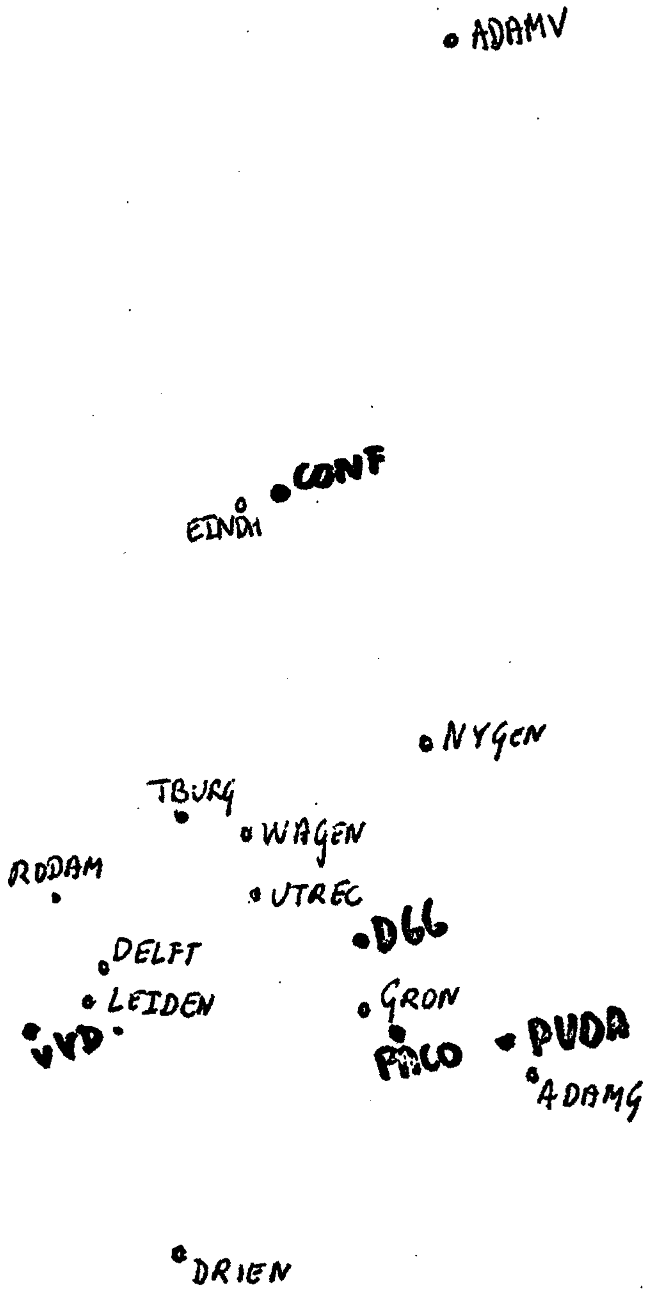


III b.

the - theology
lbw - agriculture
med - medicin
W&N - math & natural sciences
let - Literature
Soc - social sciences
PSW - political science
CIF - philosophy
tec - technical sciences
vee - veterinarians
Jur - law
eco - economists
tnd - dentists



IV a.


ADAMV
CONF
EINDH
NYGEN
TBURG
WAGEN
RODAM
UTREC
D66
DELFT
LEIDEN
VVD
GRON
PACO
PVDA
ADAMG
DRIEN

IV b