Japanese Psychological Research
1972, Vol. 14, No. 2, 87-102

ESTIMATION OF THE RECRUITMENT LATENT CLASS BY LEAST SQUARES METHODS

YOSHIO TAKANE[1]

Department of Psychology, University of Tokyo

First the formal relationship between factor analysis and latent class model of latent structure analysis was explicated and thereupon the recruitment latent class was estimated in the same manner that factor scores are estimated in factor analysis. Finally thus obtained results were compared with the one through direct recruitment probability procedure.

Introduction

One objective of the study is to show the formal relationship between two multivariate analysis models, both to be applied in the situation where appropriate external criteria are unknown. They are factor analysis and latent class model of latent structure analysis.

Conceptually factor analysis and latent structure analysis in general (not confined to latent class model) bear resemblance in that they both hypothesize latent (unobservable) variates thereby to elucidate correlational structures among manifest (observable) variates (Green, 1952). Latent variates are specifically called ' factors ' in factor analysis and ' latent traits ' in latent structure analysis. They differ, however, in the metric property of manifest variates they are intended to cover. Latent structure analysis was originally formulated for dichotomous manifest variates (Lazarsfeld, 1950), whereas factor analysis model assumes continuous manifest variates (it actually best fits to the data under the normal distribution in the sense that a variety of statistical inferences can be made which otherwise are unavailable). This difference, however, is not of great concern now that latent structure analysis model has been so extended that it can deal with any continuous data. First latent class model was generalized into latent profile model (Gibson, 1959) and then the most general formulation where both manifest variates and latent variates can be either discrete or continuous, was provided (McDonald, 1962a; Anderson, 1959). Latent class model will be referred hereafter as it implies general latent class model including latent profile model as well as original latent class model.

Still one more difference lies in their restriction upon the functional forms through which latent variates relate to manifest variates. In usual factor analysis, even if normality of the distribution is not assumed, linearity of regression of manifest variates on latent variates is assumed. On the other hand in latent structure analysis no such restraint is imposed; any curvilinear type of regression is within the scope of the model. This limitation of factor analysis, however, is not essential. Nonlinear Factor Analysis (McDonald, 1962b, 1967) and Transform Factor Analysis (Nishisato, 1971) are just two examples of the attempt to formulate a more general model of factor analysis which does not necessarily involve the linearity assumption.

Thus, one direction of methodological developments is in the generalization of

[1] This paper is based on master's thesis of the present author. He wishes to express his appreciation to Dr. Yoshihisa Tanaka of Univ. of Tokyo for his academic advice and to the members of multivariate analysis study group at the Educational Psychology Department.

models, and from a sufficiently general model, existing models are derived as its special cases, describing their common and peculiar characteristics at the same time. This paper is not along this line, however; arguments are rather confined to linear models—(linear) factor analysis and latent class model.

Now before going further into the main theme of the paper, let's briefly review some of the significant remarks which have been pointed out regarding the specific context of this study.

(1) Green (1951) has shown that the solution of latent class parameters is nothing but a particular solution of factor analysis, which directly implies that latent class model is a special type of factor analysis. He has not stated this fact explicitly in his 1951 thesis (it is aimed to provide with a general estimation procedure for latent class parameters), nevertheless it is shown that any solution of latent class parameters can be obtained through rotating the diagonal-centroid factor pattern matrix in a certain specified way.

(2) Anderson (1959) takes up a general model

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda\theta} + \boldsymbol{u}$$

and from this model derives various scaling models among which are factor analysis and latent structure analysis including latent class model. Here two models are distinguished by different additional assumptions.

(3) Lord & Novick (1968) characterize latent trait models including latent structure analysis as models involving the assumption of local independence, whereas factor analysis only requires linear local independence.

(4) Ikuzawa (1968) has applied discriminant analysis to the individual's identification problem in latent class model. Transforming individuals' response patterns into canonical variates is also a kind of factor measurements, say, in terms of discrimination.

Formal Relationship

Let's employ the algebraic model of factor analysis. That is,

$$Z = FA' + U \tag{1}$$

where Z is an $N \times n$ data matrix of standardized measure, F an $N \times r$ matrix of common factor scores, A an $n \times r$ matrix of factor pattern coefficients, and U designates unique portion of the score Z.

This model is reformulated as regression of Z on F, though F is unknown at first. (This may not be an accurate statement: Z may be considered as a set of n N-component vectors and F a set of r N-component vectors. Regression of Z on F means projection of each of n vectors in Z on the space spanned by r vectors in F.) Thus the vectors in Z are decomposed into two parts, images of F and anti-images of F.

This is denoted by

$$Z = F(F'F)^{-1}F'Z + (I - F(F'F)^{-1}F')Z. \tag{2}$$

If we put

$$(F'F)^{-1}F'Z = A'$$

and

$$(I - F(F'F)^{-1}F')Z = U$$

then it is evident that (1) and (2) are formally identical. Now consider an $N \times (r+1)$ matrix G of dummy variables indicating individuals' recruitment latent classes (unknown at first) and an $N \times n$ matrix $\mathcal{X}$ of data (either continuous or discrete; either raw score or standardized measure), and define a supermatrix,

$$X = [\mathbf{1}_N \mathcal{X}]$$

where $\mathbf{1}_N$ is an N-component vector whose elements are all unity. According to the notion of regression analysis mentioned above, let's decompose X with respect to G; that is, to get a projection of X on G and its residual.

$$X = G(G'G)^{-1}G'X + (I - G(G'G)^{-1}G')X \tag{4}$$

TABLE 1

A hypothesized set of latent class parameters (relative class sizes and within-class means): diagonal elements of D_c and matrix L'

Class	Class sizes	Variate No. 0	1	2	3	4
1	0.600	1.000	0.900	0.200	0.800	0.400
2	0.400	1.000	0.100	0.100	0.300	0.300

If we define

$$L' = (G'G)^{-1}G'X = \begin{bmatrix} 1 & l_{11} \cdots\cdots l_{1n} \\ \vdots & \vdots \qquad \vdots \\ 1 & l_{(r+1)1} \cdots l_{(r+1)n} \end{bmatrix} \quad (5)$$

L is a within-class mean matrix, which is the latent class parameters. (As a special case, if X is dichotomous variates, L indicates what portions of individual in each class respond positively to items.)

Also define $(I-G(G'G)^{-1}G')X=V$, and take the product moment of (4); that is, to premultiply the transpose of itself, we get

$$\frac{1}{N}X'X = \frac{1}{N}L(G'G)L' + \frac{1}{N}V'V. \quad (6)$$

Since $\frac{1}{N}G'G$ is a diagonal matrix whose diagonal elements are relative class sizes, we equate this to D_c. Namely,

$$D_c = \frac{1}{N}G'G \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_{r+1} \end{bmatrix} \quad (7)$$

$\frac{1}{N}V'V$ is a within-class variance-covariance matrix, which is also a diagonal matrix because all within-covariances vanish on account of the assumption of local independence. Its diagonal elements are within-class variances which are usually non-zero, and we denote this matrix by D_v^2. Then (4) is rewritten as follows.

$$\frac{1}{N}X'X = LD_cL' + D_v^2 \quad (8)$$

This agrees with Green's representation of the second moment equation of latent class model (Green, 1951).

If we regard G as F in factor analysis model (1), (8) is identical with

$$R = A\Phi A' + S^2 \quad (9)$$

which is the moment equation of (1), and where

$$R = \frac{1}{N}Z'Z, \qquad \Phi = \frac{1}{N}F'F$$

and

$$S^2 = \frac{1}{N}U'U.$$

Then

$$X = GL' + V \quad (10)$$

is supposed to be the basic equation for latent class model. Thus latent class model is a kind of factor analysis with G as factor score matrix. Differences are only in the metric properties of G and F. G's matric elements are either 0 or 1 measured on the nominal scale. G must also satisfy the following conditions.

$$G\mathbf{1}_{r+1} = \mathbf{1}_N \quad (11)$$

and

$$\frac{1}{N}G'\mathbf{1}_N = D_c\mathbf{1}_{r+1} = \begin{bmatrix} d_1 \\ \vdots \\ d_{r+1} \end{bmatrix} \quad (12)$$

Column means of G are relative class sizes which are usually non-zero.

On the contrary F consists of measures on the interval or ratio scale, and its column means are usually taken to be zero. That is,

$$\frac{1}{N}F'\mathbf{1}_N = \mathbf{0}_r \quad (13)$$

where $\mathbf{0}_r$ is an r-component zero vector.

Factor Score Estimation

Now there is good reason to estimate G

TABLE 2

Between-class moments: matrix LD_cL'

0	1	2	3	4
1.000	0.580	0.160	0.600	0.360
0.580	0.490	0.112	0.444	0.228
0.160	0.112	0.028	0.108	0.060
0.600	0.444	0.108	0.420	0.228
0.360	0.228	0.060	0.228	0.132

TABLE 3

Absolute class sizes and frequencies of positive response to each variate (of the generated data)

Class No.	Class sizes	Variate No.			
		1	2	3	4
1	605	547	139	498	251
2	385	41	36	114	118
Total	990	588	175	612	369

TABLE 4

Absolute frequencies of response patterns

No.	Response patterns	Class		Total
		I	II	
1	0000	6	160	166
2	0001	2	64	66
3	0010	16	53	69
4	0011	17	33	50
5	0100	1	17	18
6	0101	3	5	8
7	0110	8	8	16
8	0111	5	4	9
9	1000	38	17	55
10	1001	37	7	44
11	1010	208	10	218
12	1011	142	5	147
13	1100	9	1	10
14	1101	11	0	11
15	1110	68	1	69
16	1111	34	0	34
Total		605	385	990

TABLE 5

Estimates of relative class sizes and within-class means calculated from the generated data: diagonal elements of $\hat{D}_c$ and matrix $\hat{L}'$

Class	Class sizes	Variate No.				
		0	1	2	3	4
1	0.611	1.000	0.904	0.230	0.823	0.414
2	0.389	1.000	0.106	0.094	0.296	0.306

via usual least squares techniques which they often use in estimating factor scores. Computer experiment has been performed to test, with hypothetical data, the adequacy of estimating G on the principle of least squares. First an arbitrary set of latent class parameters are hypothesized (Table 1) and according to this set of parameters, dichotomous data are generated using uniform random numbers. Generated data are shown in tabulated forms in Tables 3 and 4.

Then within-class mean matrix $\hat{L}$ and estimated class size matrix $\hat{D}_c$ (Table 5) are calculated. And using these $\hat{L}$ and $\hat{D}_c$, as estimates of L and D_c respectively, 6 different estimates of G are calculated. These estimating methods are provided by Shiba (1969).

$$\hat{G}_1 = XB(B'B)^{-1}D_c^{1/2} \tag{14}$$

$$\hat{G}_2 = XB(B'PB)^{-1/2}D_c^{1/2} \tag{15}$$

$$\hat{G}_7 = XP^{-1}(P-D_v^2)B(B'B)^{-1}D_c^{1/2} \tag{16}$$

$$\hat{G}_8 = XP^{-1}(P-D_v^2)B(B'(P-D_v^2) \times P^{-1}(P-D_v^2)B)^{-1/2}D_c^{1/2} \tag{17}$$

$$\hat{G}_{13} = XP^{-1}BD_c^{1/2} \tag{18}$$

$$\hat{G}_{14} = XP^{-1}B(B'PB)^{-1/2}D_c^{1/2} \tag{19}$$

where $\quad B = LD_c^{1/2}$

TABLE 6

Matrix $\hat{B}' = \hat{D}_c^{1/2}\hat{L}$

0	1	2	3	4
0.781	0.706	0.179	0.643	0.323
0.623	0.066	0.058	0.184	0.190

TABLE 7

Estimates of between-class moments: matrix $\hat{L}\hat{D}_c\hat{L}'$

0	1	2	3	4
1.000	0.593	0.177	0.617	0.371
0.593	0.503	0.130	0.466	0.241
0.177	0.130	0.035	0.126	0.069
0.617	0.466	0.126	0.447	0.243
0.371	0.241	0.069	0.243	0.141

TABLE 8

Estimates of total moments: matrix $(1/N)X'X$

0	1	2	3	4
1.000	0.593	0.176	0.618	0.372
0.593	0.593	0.125	0.472	0.238
0.176	0.125	0.176	0.129	0.062
0.618	0.472	0.129	0.618	0.242
0.372	0.238	0.062	0.242	0.372

TABLE 9

A list of solutions of factor score estimates in term of different combinations of conditions (From Shiba, 1969)

Minimizing criteria \ Restriction (orthogonality of estimates)	Non	Yes
$X-\hat{G}L'$	$\hat{G}_1$	$\hat{G}_2$
$GL'-\hat{G}\hat{L}'$	$\hat{G}_7$	$\hat{G}_8$
$G-\hat{G}$	$\hat{G}_{13}$	$\hat{G}_{14}$

and
$$P = \frac{1}{N}X'X.$$

B is required in order to satisfy

$$\frac{1}{N}H'H = I_{r+1}$$

in the equation which is a modified version of (10). That is,

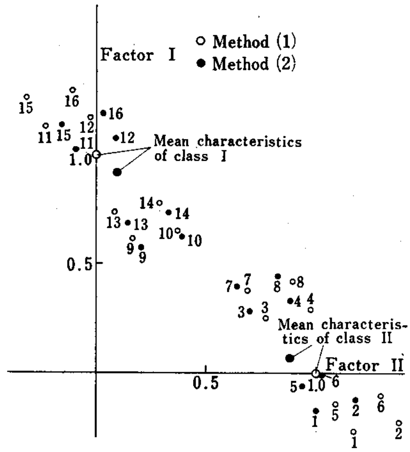

FIG. 1. Configurations of estimated factor scores by methods (1) and (2).

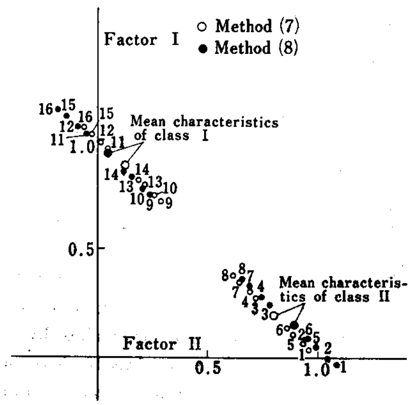

FIG. 2. Configurations of estimated factor scores by methods (7) and (8).

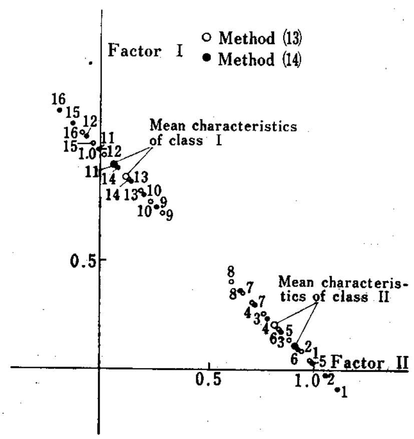

FIG. 3. Configurations of estimated factor scores by methods (13) and (14).

TABLE 10

Mean characteristics of class and estimated factor scores by method 1

Class	Factor I	Factor II
1	1.000	−0.000
2	−0.000	1.000

No. of response patterns	A*	B**	C***	Factor I	Factor II	$D_{(1)}^2$	$D_{(2)}^2$
1	2	2	2	−0.272	1.169	2.985	0.102
2	2	2	2	−0.232	1.374	3.407	0.194
3	2	2	2	0.248	0.770	1.158	0.114
4	2	2	2	0.288	0.975	1.458	0.083
5	2	2	2	−0.142	1.087	2.486	0.027
6	2	2	2	−0.102	1.292	2.885	0.095
7	2	2	2	0.378	0.688	0.860	0.240
8	1	2	2	0.418	0.893	1.136	0.186
9	1	1	1	0.609	0.165	0.179	1.067
10	1	1	1	0.649	0.370	0.260	0.817
11	1	1	1	1.130	−0.232	0.071	2.798
12	1	1	1	1.170	−0.027	0.029	2.426
13	1	1	1	0.739	0.083	0.074	1.386
14	1	1	1	0.779	0.288	0.132	1.113
15	1	1	1	1.260	−0.315	0.166	3.317
16	1	1	1	1.300	−0.109	0.102	2.922

* A: Most probable class by maximum likelihood method.
** B: Minimum squared error class by least squares method.
*** C: Minimum distance class by least squares method.
(Tables 11 through 15 follow the same notations.)

$$X = HB' + V$$

where

$$H = GD_c^{-1/2}$$

hence

$$G = HD_c^{1/2}.$$

Method numbers follow the notation by Shiba. Table 9 shows the characteristic of estimating methods in terms of difference in minimizing criteria and restriction on orthogonality of estimated factors.

In actual estimation, $\hat{L}$ and $\hat{D}_c$ are utilized in place of L and D_c respectively. $\hat{B} = \hat{L}\hat{D}_c^{1/2}$ is listed in Table 6. Tables 7 and 8 are the estimated matrices of between-class moments and total moments. Goodness of fit of the generated data evaluated in terms of chi square is 18.062 with 15 degrees of freedom. (McHugh, 1956) [$p(\chi^2 > 7.261) = 0.95$, $p(\chi^2 > 24.996) = 0.05$]

Tables 10–15 show the estimated factor scores. Figs. 1–3 show their respective configurations.

It is observed that different estimating methods yield more or less similar results except perhaps $\hat{G}_1$ and $\hat{G}_2$. In $\hat{G}_7$, $\hat{G}_8$, $\hat{G}_{13}$ and $\hat{G}_{14}$ configuration points are nearly in a line, whereas in $\hat{G}_1$ and $\hat{G}_2$ they are somewhat more scattered around a line. We prove the following equation;

$$\hat{G}_{13}\mathbf{1}_{r+1} = \mathbf{1}_N \qquad (20)$$

$\hat{G}_{13}$ is the so called regression method of

TABLE 11
Mean characteristics of class and estimated factor scores by method 2

Class	Factor I	Factor II
1	0.921	0.098
2	0.067	0.877

No. of response patterns	A*	B**	C***	Factor I	Factor II	$D_{(1)}^2$	$D_{(2)}^2$
1	2	2	2	−0.171	0.999	2.008	0.072
2	2	2	2	−0.121	1.183	2.265	0.129
3	2	2	2	0.281	0.700	0.773	0.076
4	2	2	2	0.331	0.884	0.966	0.069
5	2	2	2	−0.057	0.939	1.668	0.019
6	2	2	2	−0.007	1.123	1.915	0.066
7	2	2	2	0.395	0.641	0.572	0.163
8	1	2	2	0.445	0.825	0.755	0.145
9	1	1	1	0.573	0.205	0.132	0.707
10	1	1	1	0.623	0.389	0.173	0.548
11	1	1	1	1.026	−0.093	0.047	1.862
12	1	1	1	1.076	0.090	0.024	1.638
13	1	1	1	0.686	0.145	0.057	0.919
14	1	1	1	0.737	0.329	0.087	0.749
15	1	1	1	1.139	−0.153	0.110	2.212
16	1	1	1	1.190	0.030	0.076	1.978

TABLE 12
Mean characteristics of class and estimated factor scores by method 7

Class	Factor I	Factor II
1	0.876	0.122
2	0.194	0.807

No. of response patterns	A*	B**	C***	Factor I	Factor II	$D_{(1)}^2$	$D_{(2)}^2$
1	2	2	2	0.035	0.967	1.421	0.050
2	2	2	2	0.064	0.938	1.325	0.034
3	2	2	2	0.276	0.722	0.720	0.013
4	2	2	2	0.305	0.693	0.652	0.025
5	2	2	2	0.105	0.893	1.188	0.015
6	2	2	2	0.134	0.864	1.100	0.006
7	2	2	2	0.346	0.648	0.557	0.048
8	1	2	2	0.375	0.619	0.497	0.068
9	1	1	1	0.715	0.288	0.053	0.541
10	1	1	1	0.744	0.258	0.036	0.603
11	1	1	1	0.956	0.043	0.012	1.164
12	1	1	1	0.985	0.013	0.023	1.254
13	1	1	1	0.785	0.214	0.016	0.701
14	1	1	1	0.814	0.184	0.007	0.772
15	1	1	1	1.026	−0.030	0.046	1.395
16	1	1	1	1.055	−0.060	0.065	1.493

TABLE 13

Mean characteristics of class and estimated factor scores by method 8

Class	Factor I	Factor II
1	0.932	0.044
2	0.148	0.887

No. of response patterns	A*	B**	C***	Factor I	Factor II	$D_{(1)}^2$	$D_{(2)}^2$
1	2	2	2	−0.034	1.084	2.018	0.072
2	2	2	2	−0.001	1.048	1.882	0.048
3	2	2	2	0.242	0.783	1.023	0.019
4	2	2	2	0.275	0.747	0.926	0.035
5	2	2	2	0.046	0.994	1.687	0.021
6	2	2	2	0.079	0.957	1.562	0.009
7	2	2	2	0.323	0.692	0.791	0.068
8	1	2	2	0.356	0.656	0.706	0.096
9	1	1	1	0.747	0.247	0.075	0.768
10	1	1	1	0.780	0.211	0.051	0.856
11	1	1	1	1.024	−0.053	0.018	1.653
12	1	1	1	1.057	−0.090	0.033	1.782
13	1	1	1	0.828	0.156	0.023	0.996
14	1	1	1	0.861	0.120	0.010	1.096
15	1	1	1	1.105	−0.144	0.065	1.982
16	1	1	1	1.138	−0.180	0.092	2.122

TABLE 14

Mean characteristics of class and estimated factor scores by method 13

Class	Factor I	Factor II
1	0.873	0.126
2	0.198	0.802

No. of response patterns	A*	B**	C***	Factor I	Factor II	$D_{(1)}^2$	$D_{(2)}^2$
1	2	2	2	0.038	0.963	1.396	0.051
2	2	2	2	0.087	0.911	1.233	0.024
3	2	2	2	0.253	0.747	0.769	0.006
4	2	2	2	0.302	0.695	0.649	0.022
5	2	2	2	0.138	0.865	1.085	0.007
6	2	2	2	0.187	0.813	0.941	0.000
7	2	2	2	0.354	0.649	0.543	0.047
8	1	2	2	0.402	0.597	0.443	0.083
9	1	1	1	0.717	0.282	0.048	0.539
10	1	1	1	0.766	0.230	0.022	0.649
11	1	1	1	0.933	0.066	0.007	1.080
12	1	1	1	0.981	0.015	0.024	1.233
13	1	1	1	0.818	0.184	0.006	0.765
14	1	1	1	0.866	0.132	0.000	0.895
15	1	1	1	1.033	−0.031	0.050	1.391
16	1	1	1	1.082	−0.082	0.087	1.564

TABLE 15
Mean characteristics of class and estimated factor scores by method 14

Class	Factor I	Factor II
1	0.930	0.069
2	0.108	0.891

No. of response patterns	A*	B**	C***	Factor I	Factor II	$D_{(1)}^2$	$D_{(2)}^2$
1	2	2	2	−0.086	1.087	2.069	0.076
2	2	2	2	−0.026	1.024	1.827	0.036
3	2	2	2	0.176	0.825	1.139	0.008
4	2	2	2	0.235	0.762	0.962	0.032
5	2	2	2	0.035	0.967	1.607	0.011
6	2	2	2	0.095	0.905	1.395	0.000
7	2	2	2	0.298	0.705	0.804	0.070
8	1	2	2	0.357	0.642	0.656	0.123
9	1	1	1	0.741	0.259	0.071	0.799
10	1	1	1	0.800	0.196	0.032	0.961
11	1	1	1	1.003	−0.003	0.010	1.600
12	1	1	1	1.063	−0.065	0.035	1.826
13	1	1	1	0.863	0.139	0.009	1.133
14	1	1	1	0.922	0.077	0.000	1.325
15	1	1	1	1.125	−0.122	0.074	2.060
16	1	1	1	1.185	−0.185	0.129	2.317

factor score estimation; that is, regression of G on X.

$$\hat{G}_{18} = X(X'X)^{-1}X'G$$

where

$$\left(\frac{1}{N}X'X\right)^{-1} = P^{-1}$$

and

$$\left(\frac{1}{N}X'G\right) = BD_c^{1/2} = LD_c\,.$$

Then

$$\hat{G}_{18}\mathbf{1}_{r+1} = X(X'X)^{-1}X'\mathbf{1}_N\,.$$

Since X contains $\mathbf{1}_N$ as one of its vector elements, the relation

$$X(X'X)^{-1}X' = \Pi_N + \Pi_N^{\perp}\mathcal{X}(\mathcal{X}'\Pi_N^{\perp}\mathcal{X})^{-1}\mathcal{X}'\Pi_N^{\perp}$$

holds (Takeuchi & Yanai, 1972), where

$$\Pi_N = \frac{1}{N}\mathbf{1}_N\mathbf{1}'_N$$

and

$$\Pi_N^{\perp} = I_N - \Pi_N\,.$$

Since

$$\Pi_N\mathbf{1}_N = \mathbf{1}_N$$

and

$$\Pi_N^{\perp}\mathbf{1}_N = \mathbf{0}_N\,,$$

$X(XX')^{-1}X'\mathbf{1}_N = \mathbf{1}_N$ holds.
Because estimates $\hat{L}$ and $\hat{D}_c$ are used in place of parameters, there can be observed some fluctuations, but still the relation (20) holds in an approximate sense. (20) states that the configuration forms a line on the simple plane; that is, in case the number of latent classes is two.

And also if (8) is true, then

$$P - D_v^2 = LD_cL' = BB'\,.$$

This leads to the following equation (Shiba, 1972):

$$\hat{G}_7 = \hat{G}_{18}\,.$$

Again $\hat{L}$ and $\hat{D}_c$ cause the discrepancy between these two methods.

TABLE 16

Relative class sizes and within-class means calculated on the basis of assigned class by least squares methods: diagonal elements of $\tilde{D}_c$ and matrix $\tilde{L}'$

Class	Class sizes	Variate No.				
		0	1	2	3	4
1	0.593	1.000	1.000	0.210	0.795	0.401
2	0.406	1.000	0.000	0.126	0.358	0.330

It seems that orthogonality condition imposed upon estimated factors only stretch or contract configurations along a line, and yield a very close result to the corresponding method without this condition.

Now it is clear why the results are all alike except $\hat{G}_1$ and $\hat{G}_2$, and they are in the form of a ' quasi ' straight line.

Let's consider

$$G = X^*L(L'L)^{-1}$$

where $X^* = GL'$, and

$$\hat{G}_1 = XL(L'L)^{-1} = G+VL(L'L)^{-1} . \quad (21)$$

Thus $\hat{G}_1$ can be conceived as an approximate method for what is called ' ideal variables '. But the problem is that (21) counts uniqueness variation as part of $\hat{G}_1$. This may be why $\hat{G}_1$ and its orthogonal counterpart, $\hat{G}_2$ give somewhat different configurations from others.

The recruitment class of a response pattern is determined in two different ways.

In the first we find a dummy variable matrix which best fits to $\hat{G}$; that is $\tilde{G}$ which minimizes trace of $E'E$, where $E=\tilde{G}-\hat{G}$. This incidentally agrees simply to find $\hat{j}$ such that

$$\hat{j} = \max_j (\hat{g}_{kj})$$

where

$$[\hat{g}_{kj}] = \hat{G} .$$

In the second we transform within class means L into factor scores and measure the distances between these class means and each of response patterns. In the first way we have assumed that the characteristics of latent classes are represented as the points (1, 0) or (0, 1) on factor scores. But it is only in method (1) that these points also signify the mean characteristics of latent classes.

$$\hat{G}_1 = XB(B'B)^{-1}D_c^{1/2} = XLD_c^{1/2}(D_c^{1/2}L'LD_c^{1/2})^{-1}D_c^{1/2} = XL(L'L)^{-1}$$

If we replace X by L', we get

$$G_m = L'L(L'L)^{-1} = I_{r+1} .$$

No other estimation methods have this theoretical feature. Distances are calculated in the following way.

$$D_{k}^2{}_{(\alpha)} = \sum_j (g_{kj}-\mu_{j(\alpha)})^2$$

where

- i: response pattern index.
- j: factor index.
- α: class index.
- $\mu_{j(\alpha)}$: mean characteristic of class α on factor j.
- $D_{k(\alpha)}$: distance between class α and response pattern k.

The decision function is

$$\hat{\alpha} = \min_\alpha D_k^2{}_{(\alpha)} .$$

We call the first way ' minimum squared error method' and the second 'minimum distance method '.

In the present case, results of class assignment agree unanimously through different estimating methods of factor scores and different ways of determining classes, though this is not confirmed generally. (The author has found in another example the case in which different estimating meth-

TABLE 17

Total moments calculated on the basis of assigned class by least squares methods: matrix $\tilde{P}$

0	1	2	3	4
0.999	0.593	0.176	0.618	0.372
0·593	0.593	0.125	0.472	0.238
0.176	0.125	0.032	0.118	0.067
0.618	0.472	0.118	0.428	0.237
0.372	0.238	0.067	0.237	0.140

ods lead to different class assignments.)

Tables 16 and 17 are computed on the basis of assigned classes in the ways mentioned above to show how they can reproduce original matrices L and D_c, and how they satisfy the assumption of local independence.

Goodness of fit of $\tilde{L}$ and $\tilde{D}_c$ to L and D_c is evaluated in terms of chi square which amounts to 7.460 with 15 degrees of freedom. This is a fairly good approximation. It is interesting to note that this suggests a possibility of an easy, though rough, approximating method for latent class parameters by partitioning response patterns into classes through rotating principal or diagonal-centroid factor score matrix into simple structures.

Finally the results of least squares meth-

TABLE 18

Recruitment probabilities

No. of response patterns	A*	$P(k)$	$P(k, 1)$	$P(k, 2)$	$P(1/k)$	$P(2/k)$
1	2	.158623	.004684	.153938	.029532	.970467
2	2	.071184	.003309	.067874	.046493	.953506
3	2	.086506	.021782	.064724	.251798	.748201
4	2	.043927	.015388	.028538	.350324	.649675
5	2	.017370	.001399	.015971	.080554	.919445
6	2	.008030	.000988	.007042	.123099	.876900
7	2	.013221	.006506	.006715	.492097	.507902
8	1	.007557	.004596	.002960	.608216	.391783
9	1	.062365	.044113	.018252	.707335	.292664
10	1	.039165	.031165	.008047	.794768	.205231
11	1	.212789	.205115	.007674	.963935	.036064
12	1	.148294	.144910	.003383	.977182	.022817
13	1	.015070	.013176	.001893	.874342	.125657
14	1	.010144	.009309	.000834	.917688	.082311
15	1	.062064	.061268	.000796	.987171	.012828
16	1	.043636	.043285	.000351	.991954	.008045

A*: Most probable class.

k: Response pattern index.

j: Latent class index.

$P(k)$: Marginal probability of response pattern k.

$P(k,j)$: Joint probability of response pattern k and class j.

$P(j/k)$: Conditional probability of class j given response pattern k (Recruitment probability).

$$P(k,j) = d_j \prod_{i=1}^{n} l_{ij}^{x_{ik}}(1-l_{ij})^{1-x_{ik}}$$

x_{ik}: Response of pattern k to i^{th} variate.

$$P(k) = \sum_j P(k,j).$$

$$P(j/k) = P(k,j)/P(k).$$

ods are compared with the one via recruitment probability procedure, a sort of maximum likelihood method of estimation. Recruitment probability is the probability that an individual with a certain response pattern to items belongs to a certain latent class, and it can be calculated whenever the type of population distribution is known and estimates for distribution parameters are available (Lazarsfeld, 1954; Lazarsfeld & Henry, 1968). (Recruitment probability is the conditional probability of latent class given a response pattern.) Computa tional procedure is given at the bottom of Table 18.

It is found that least squares methods do not always give results which perfectly agree with the most probable class by maximum likelihood method. Here Response Pattern No. 8 is always assigned to Class II by least squares methods and to Class I by maximum likelihood method. Generally maximum likelihood method seems to have more acute discriminating power than least squares methods.

But a pretext here is that the marginal probability of Response Pattern No. 8 is so small (less than 8 out of 1000) that this error of discrimination would only raise the total error probability by approximately 0.164% (from 9.825% to 9.989%). In addition maximum likelihood methods cannot be applied unless the type of population distribution is known, while least squares methods facilitate much broader applications.

Concluding Remarks

It has been shown that latent class model is a special type of factor analysis with dummy variable matrix G as its factor scores. Then what are the significances of latent class solution over usual factor analytic solutions, say perhaps, varimax rotation? One is obviously the uniqueness of latent class solution, as contrasted to the fundamental indeterminate nature of factor decomposition; for the complete specification of factors some other criteria such as varimax should be introduced from outside the model (1) itself.

Another, still more important to note, is an 'intrinsic' linearity nature of latent class model (Gibson, 1959). (Linear) factor analysis has also this linear property, but their meanings are quite different.

Now let's take up the general regression model,

$$y_k = \alpha_{ok}+\sum_i \alpha_{ik}\varphi_i+\sum_{i\geqslant j}\alpha_{ijk}\varphi_i\varphi_j+\cdots+\varepsilon_k \quad (22)$$

$$(k = 1, \cdots, n)$$

$$(i, j = 1, \cdots, r)$$

where y_k is a k^{th} dependent variate, $\varphi_i(i=1, \ldots r)$ an i^{th} independent variable, and α_{ok}, α_{ik}, α_{ijk} regression coefficients. ε_k is a disturbance term.

(Sometimes coefficients are not uniquely determined. This problem involves the notion of functional independence. And in fact our second case to be mentioned below violates this condition. But for the sake of explanatory convenience we assume that coefficients are determined for the moment.)

It is expected that (22) is a good approximation to the more general regression formula,

$$\boldsymbol{y} = f(\boldsymbol{\varphi})+\boldsymbol{\varepsilon}$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_r \end{bmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

and f is some unspecified function. If we ignore all terms equal and higher than the second degree in (22), we get

$$y_k = \alpha_{ok}+\sum_i \alpha_{ik}\varphi_i+\varepsilon_k' \quad (23)$$

where

$$\varepsilon_k' = \varepsilon_k+\sum_{i\geqslant j}\alpha_{ijk}\varphi_i\varphi_j+\cdots.$$

If we further assume that

$$E(y_k) = 0 \quad (k = 1, \cdots, n)$$
$$E(\varphi_i) = 0 \quad (i = 1, \cdots, r)$$
$$E(\varepsilon_k) = 0 \quad (k = 1, \cdots, n)$$

hence $\alpha_{ok}=0$, we have factor analytic model.

On the other hand if we define φ_i, φ_j as

$$\varphi_i = \begin{cases} 1 \\ 0 \end{cases}$$

$$(i,j = 1, \ldots, r)$$

$$\varphi_i\varphi_j = \begin{cases} \varphi_i & i = j \\ 0 & i \neq j \end{cases}$$

we have from Eq. (22)

$$y_k = \alpha_{ok} + \sum_i \alpha'_{ik}\varphi_i + \varepsilon_k \tag{24}$$

because all higher cross terms vanish and since $\varphi_i^m = \varphi_i (i=1, \ldots, r)$ for any integer m, if we put

$$\alpha'_{ik} = \alpha_{ik} + \alpha_{iik} + \ldots \qquad (i=1, \ldots, r)$$

We have latent class model indicated by (24). Notice that (23) is only a linear approximation to (22), while (24) is a perfect equivalence to (22) except a restriction upon the φ_i's domain. This implication is of considerable interest in that in factor analysis strict linear relations between manifest and latent variates are required, whereas in latent class model manifest and latent variates are always related linearly no matter how manifest variates are distributed.

Finally we go back to the problem of factor score estimation to give more deliberations to some possible reasons why the disagreement we have observed occurs.

It must be noted that this disagreement cannot be attributed to the simple fact that $\hat{L}$ happens to be just not so good an estimate of L. There is an evidence to support this statement. The results of estimation in which parameters L is used in stead of estimates $\hat{L}$ reveal still one disagreeing pattern between maximum likelihood method and each of least squares methods. (There is no disagreement among least squares methods, or between $\hat{L}$ and L.) Response Pattern No. 8 is always assigned to Class I by maximum likelihood method and to Class II by least squares methods.

One possible explanation to this is that least squares methods are exclusively based on the distances and do not take class sizes, or prior probabilities, into account. This supposition is supported by the fact that without prior probabilities even in maximum likelihood method, the assigned class of disagreeing response pattern is so reversed as to yield a result which perfectly agrees to those through least squares methods.

One way to consider prior probabilities in the discrimination is to assume normality of the distribution upon factor scores and to apply maximum likelihood method to the estimated factor scores. The similar procedure has been taken up by Ikuzawa (1968) who has applied maximum likelihood method to canonical variates to discriminate latent classes. (Ikuzawa's example is somewhat unfair in the respect that in his example prior probabilities do not influence the results of discrimination in any way.) But even his approach is not a final answer. His procedure applied to our data has resulted in the same disagreement to maximum likelihood method; that is, Rosponse Pattern No. 8 is assigned to Class II even by discriminant analysis.

Further investigation is yet necessary to make use of informations about prior probabilities, when available, in the real intrinsic way in least squares methods.

References

Anderson, T. W. 1959 Some scaling models and estimation procedures in the latent class model. In U. Grenander (Ed.) *Probability and Statistics*. New York: Wiley.

Gibson, W. A. 1959 Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, **24**, 229–252.

Green, B. F. 1951 A general solution for the latent class model of latent structure analysis. *Psychometrika*, **16**, 151–160.

Green, B. F. 1952 Latent structure analysis and its relation to factor analysis. *J. Amer. Stat.*

Ass., **47**, 71–76.

IKUZAWA, M. 1968 Individual identification in latent class analysis: Application of canonical analysis and discriminant analysis. *Shinrigakuhyoron*, **11**, 171–198.

LAZARSFELD, P. F. 1950 Chapters 10 and 11. In S. Stouffer, et al. (Eds.), *Measurement and prediction*. Princeton: Princeton Univ. Press.

LAZARSFELD, P. F. 1954 A conceptual introduction to latent structure analysis, in P. Lazarsfeld (Ed.) *Mathematical thinking in the social sciences*. Glencoe: Free Press.

LAZARSFELD, P. F. & HENRY, N. W. 1968 *Latent Structure Analysis*. Boston: Houghton Mifflin.

LORD, F. M. & NOVICK, M. R. 1968 *Statistical theories of mental scores*. New York: Wiley.

MADANSKY, A. 1960 Determinantal methods in latent class analysis. *Psychometrika*, **25**, 183–198.

McDONALD, R. P. 1962a A note on the derivation of the general latent class model. *Psychometrika*, **27**, 203–206.

McDONALD, R. P. 1962b A general approach to nonlinear factor analysis. *Psychometrika*, **27**, 397–415.

McDONALD, R. P. 1967 Nonlinear factor analysis. *Psychometr. Monogr.*, **15**.

McHUGH, R. B. 1956 Efficient estimation and local identification in latent class analysis. *Psychometrika*, **21**, 331–347.

NISHISATO, S. 1971 Transform factor analysis: A skechy representation of a general approach. *Jap. psychol. Res.*, **13**, 155–166.

SHIBA, S. 1969 New estimates of factor scores. *Jap. psychol. Res.*, **11**, 129–133.

SHIBA, S. 1972 *Factor analysis*. Tokyo: Univ. of Tokyo Press.

TAKEUCHI, K. & YANAI, H. 1972 *Basis of multivariate analysis*. Tokyo: Toyo Keizai Shimpo.

(Received July 8, 1972)