RR 84·13

Research Report

ANALYSIS OF STUDENT SELECTION EXAMINATION
FOR
TURKISH UNIVERSITIES
OF
1984-85 ACADEMIC YEAR

Soner Gönen

November 1984

Department of Data Theory, University of Leiden, The Netherlands

ANALYSIS OF STUDENT SELECTION EXAMINATION FOR TURKISH UNIVERSITIES OF 1984-85 ACADEMIC YEAR

Soner Gönen

November 1984

This research was carried out at Leiden University while the author was a visiting researcher at the Department of Data Theory. I am greatful to prof. John P. van de Geer and his colleagues of Leiden University, Holland and prof. Altan Günalp at Ö.S.Y.M. in Turkey. Special thanks to Erhan Ögüt at Turkish Embassy in Holland for his valuable support.

1 INTRODUCTION

Student selection and placement for Turkish Universities is a very crucial and sensitive matter in Turkey. Two levels of SSP examination are used: First level for student selection and second level for student placement. The selection examination is composed of two tests, called 'science test' and 'social science test'. Student placement examination has five tests. These are mathematic test, science test, Turkish language test, social science test, foreign language test. Student selection examination was given on April 16, 1984 at 8.30 a.m. to 43000 examinees all over the country. Student placement examination was given on July 25, 1984 at 9.30 a.m. all over the country to the students who were succesfull in student selection examination (about 50% of the total number of examinees). According to the result of this examination and the preference of students, each student was admitted at only one department.

2 DATA

Data consist of a 1/100 systematic random sample from the 430.000 objects (examinees) in the student selection examination. Their answer are recorded on tape. There are 69 multiple choice variables (questions) in the science test, each with five alternatives. There are 73 multiple choice variables in social science test, also with five alternatives. (Ögrenci Seçme ve Yerlestirme Sinavi Birinci Basamak Soru Kitapçigi A, 1984 i.e. Student Selection Examination, 1984). Data are summarized as a table of marginal frequencies which has n rows and m columns. Here n refers to number of objects, and m to number of variables. A translation of some variables are given in the appendix.
It is assumed that all variables measure the 'same thing'.

3 ANALYSIS

Two kinds of analysis have been applied. They are HOMALS and PRIMALS.

HOMALS ANALYSIS

HOMALS is a computer program for homogeneity analysis which is an extension of principal component analysis to nominal variables. It uses an alternating least squares algorithm. Nominal variables are variables which sort objects into a limited number of categories. HOMALS gives optimal quantifications of objects in such a way that objects in the same category are closer to each other than objects in different categories. Dimensionality of HOMALS is chosen by the user (Gifi, 1981; Gif, 1983; Geer, J.P. van de e.a., 1984; Leeuw, J. de, 1984).

PRIMALS ANALYSIS

PRIMALS is a computer program also for optimal quantification of objects and categories for nominall variables. PRIMALS gives one dimensional solution. The quantification produced in the PRIMALS solution is the same as the first HOMALS solution. PRIMALS also gives a PCA (principal components analysis) of the quantified data (Gifi, 1981; Geer, J.P. van de and J. Meulman in press).

Both analyses assume that variables measure 'the same thing'. This assumption comes from the idea of homogeneity.

Two types of analyses were performed according to the data included in the analysis. Since marginal frequencies of missing data are very large, it is assumed in the first type of analysis that missing data on the same variable form a separate category. Therefore analysis of this type have six categories for each variable. The labels of these categories are missing data, alternative A, alternative B, alternative C, alternative D, alternative E. Corresponding category numbers are 1, 2, 3, 4, 5, 6. This type is called "analysis with missing data single category".

In the second type of analysis missing data are not included. Analysis of this type has five categories. The labels of these categories are alternative A, alternative B, alternative C, alternative D, alternative E. Their category numbers are 1, 2, 3, 4, 5 respectively. This type is called "analysis with missing data passive".

Actually there is one more category of illegal answers (two alternatives are chosen instead of one for a variable). This category is not included in this work since the frequency in this category is very low.

We have three HOMALS and two PRIMALS solutions with missing data single category. We also have three HOMALS and two PRIMALS solutions with missing data passive. All PRIMALS solutions include PCA solutions.

HOMALS solutions with missing data single category are considered to be the main part of this research because of they use all available data. Other solutions are employed to have some interesting information and also to make comparisons.

3.1 HOMALS analysis of science test with missing data single category

There are n=4300 object, m=69 variables and each variable assumes just k=6 categories. The HOMALS solution gives us the following matrices (tables):

Matrix of marginal frequencies of order mxk.

Matrix of discrimination measures of order mxp where p is a number of dimensionality, chosen at two in this analysis because of the difficulty of interpretation of higher dimensional solutions.

Matrices of category quantifications of order kxp, one for each variable.

Since we have so many matrices both in this section and in the following sections, we could not include complete matrices, instead we present the results for some selected variables.

Looking at the matrix of marginal frequencies one can see the followings.

Missing date frequency is the lowest for variable 2 and equal to 773
" " " " highest " " 30 " " " 3702

Correct alternative frequency is the lowest for variable 68 and equal to 194.

Correct alternative frequency is the highest for variable 2 and equal to 3099.

Wrong alternative frequency is the lowest for variable 22 and equal to 32.
Wrong alternative frequency is the highest for variable 26 and equal to 1336.

Marginal frequencies, category numbers, category quantifications and discrimination measures of some variables are given in table 1. Translations of these variables are given in the appendix.

The eigenvalues of the two dimensions are λ_1 = 0,2223 and λ_2 = 0,1291. Counterpart of ordinary PCA they become $m\lambda_1$ = 15,3387 and $m\lambda_2$ = 8,9079. But interpretation of eigenvalues of HOMALS is not the same as in ordinary PCA as an explained variance. Looking at the matrix of discrimation measures we can write the following:
The lowest discrimination measure on the first dimension is 0,0409 and belonging to variable 64.
The highest discrimination measure on the first dimension is 0,3340 and belonging to variable 36.
The lowest discrimination measure on the second dimension is 0,0288 and belonging to variable 37.
The highest discrimination measure on the second dimension is 0,2689 and belonging to variable 59.

Variables 36; 9; 57; 34; 59; 60; 20; 33; 37; 38; 22; ... have larger discrimination measures on the first dimension. These are well separated variables on the first dimension. In other words, examinees who skip on one of these variables, also tend to have missing data on other variables. See figure 1 for a plot of discrimination measure some variables.
Variables 59; 3; 22; 9; 63; 20; 25; 69; 68; 8; 29; ... have larger discrimination measures of the second dimension. These are well separated variables on the second dimension. In other words examinees who have correct answer on one of these variables also tend to have correct answers on many other variables.
Variables 9; 59; 60; 30; 20; 11; ... have larger discrimination measures on both dimensions. These are well separated variables on both dimensions.

Table 1: Category labels, c. numbers, c. quantifications, marginal frequencies, discrimination measures of some variables of science test with missing data single category.

		c. Quantifications on two dimensions							Disc. Meas.
Var	c. Label	M	A	B	C	D	E	IL	
	c. Number	1	2	3	4	5	6	-	
2	Mar. Freq.	773	95	157	83	3099+	91	2	-
	Dim.1	0.52	-0.12	-0.06	-0.59	-0.11	-0.38	-	0.0806
	Dim.2	-0.18	-1.10	-0.80	-1.42	0.18	-0.95	-	0.1386
22	Mar. Freq.	1182	272	100	609	32	2099+	6	-
	Dim.1	0.76	-0.16	-0.15	-0.00	-0.95	-0.39	-	0.2415
	Dim.2	-0.09	-0.92	-0.74	-0.70	-1.76	0.44	-	0.2559
26	Mar. Freq.	1591	1336	205	88	951+	122	7	
	Dim.1	0.57	-0.17	-0.38	-0.40	-0.58	-0.15	-	0.2141
	Dim.2	0.06	-0.22	-0.84	-0.77	0.52	-0.40	-	0.1272
30	Mar. Freq.	3702	80	110	99	94	213+	2	
	Dim.1	0.19	-0.92	-1.04	-1.18	-1.23	-1.36	-	0.2311
	Dim.2	0.06	-1.15	-0.35	-1.39	-1.36	0.79	-	0.1469
36	Mar. Freq.	2264	1110+	131	254	135	504	2	
	Dim.1	0.54	-0.75	-0.65	-0.50	-0.58	-0.34	-	0.3340
	Dim.2	-0.05	0.52	-0.62	-0.49	-0.88	-0.17	-	0.1180
37	Mar. Freq.	2075	581	498	267	414+	461	4	
	Dim.1	0.56	-0.43	-0.36	-0.49	-0.74	-0.65	-	0.3506
	Dim.2	-0.02	-0.10	-0.29	-0.07	0.30	0.28	-	0.0288
59	Mar. Freq.	2906	977+	117	112	90	94	4	
	Dim.1	0.39	-0.83	-0.79	-0.87	-1.09	-0.40	-	0.3506
	Dim.2	-0.11	0.78	-1.15	-1.48	-0.92	-0.69	-	0.2689
64	Mar. Freq.	1649	1287	449	208+	578	124	5	
	Dim.1	0.20	0.03	-0.15	-0.31	-0.34	-0.26	-	0.0409
	Dim.1	0.45	-0.39	-0.31	-0.23	0.10	-0.95	-	0.1630
68	Mar. Freq.	3363	264	157	200	194+	121	1	
	Dim.1	0.14	-0.35	-0.78	-0.59	-0.57	-0.37	-	0.0803
	Dim.2	0.23	-0.90	-0.79	-0.97	-0.54	-1.10	-	0.2059

+ = correct alternative,
IL: Illegal answer, not included in analysis

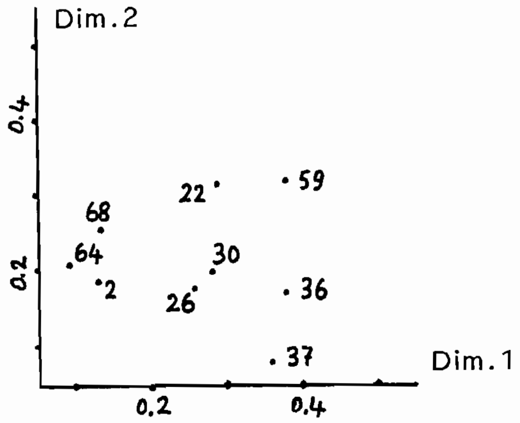


Figure 1. Plot of the Discrimination Measures of variables of table 1.

Discrimination measures of the last seven variables on the first dimension are very small. This may be due to the fact that some of the examinees did not ship these variables but simple did not have enough time.

Figure 2 plots two dimensional HOMALS category quantifications. There is one plot for each variable.

The first HOMALS dimension is scaled in such a way that positive quantifications go with missing data (table 1). This is also true for all 69 variables, therefore first HOMALS dimension reflects missing versus nonmissing answers.

In seven of the plots of figure 2 positive quantifications in the second dimension correspond to the correct alternative.

Plots of figure 1 also show that the correct alternative has highest quantification in eight plots of nine. This is true for almost all remaining 60 variables. Therefore the second HOMALS dimension reflects correct alternative versus wrong alternatives. Category quantifications of variable 64 on the second dimension suggest that alternative D (category number 5) seems more correct than the correct alternative C (category number 4). Category quantifications of variable 68 on the second dimension suggest that all alternatives seem incorrect. This is also true for six of the seven variables.

As in the case of discrimination measures it might be that some examinees could not attain those last seven variables, instead of skipping them. In any case one can say that answers for the last seven variables were influenced by some other factors. Category quantification of category 1 for variable 68 is the highest among the others, there-

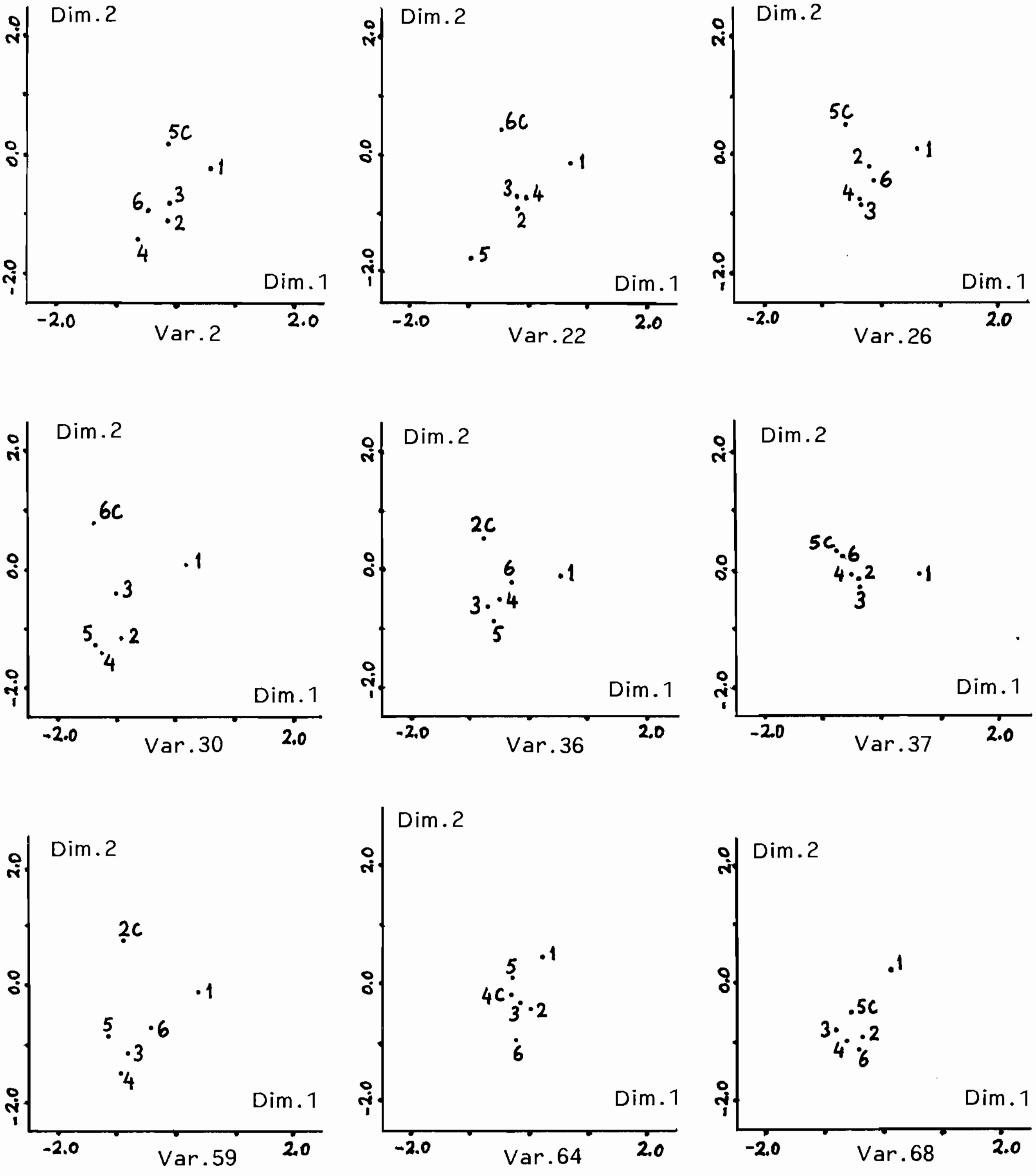


Figure 2: Category quantifications of variables labeled by original category Number in Table 1.
c: correct alternatives

fore we can also say that skipping a question seems more profitable than to chose one the alternatives. Since the Turkish examination system assigns 1.00 weight for correct alternative, -0,25 weight for wrong alternative, 0.00 for skipping a variable, this result might be logical.

Figure 2 also shows that category 1 is located between correct and wrong alternatives for each variable in the second dimension. So examinees who skip variables are neither good nor bad.

Patterns of signs and magnitude of category quantifications are almost the same for other 60 variables.

HOMALS analysis also gives a plot (not reproduced here) of the 4300 object scores in two dimensions. Since this plot does not look like a parabola ("horseshoe") the two dimensional solution is better than a one dimensional solution (Gifi, 1983).

3.2 HOMALS analysis of science test with missing data passive

We could not find a dominant one-dimensional solution in subsection 3.1. Moreover the dimension of missing versus nonmissing alternatives is not of primary importance. Therefore we merged the first category and the category of illegal data of subsection 3.1 under the label of missing to force the interesting information into the first two dimensions. Now we have k=5 categories which are alternative A, alternative B, alternative C, alternative D, alternative E with corresponding category numbers 1, 2, 3, 4, 5. Marginal frequencies for categories of objects in the analysis are different for each variable.

We have again p=2 dimensions with corresponding eigenvalues $\lambda_1 = 0{,}1959$ and $\lambda_2 = 0{,}0829$.

The second dimension of subsection 3.1 now becomes the first dimension of this solution as we shall see later. Bearing this in mind we see that most variables which have larger discrimination measure in the second imension of subsection 3.1 also have larger discrimination measure in the first dimension of this solution. We can write the following from matrix of discrimination measures of order 69x5.

The discrimination measure in the first dimension is the lowest for variable 18 and equal to 0,0214.

The discrimination measure in the second dimension is the lowest for variable 45 and equal to 0,0100.

The discrimination measure in the first dimension is the highest for variable 22 and equal to 0,5011.

The discrimination measure in the second dimension is the highest for variable 2 and equal to 0,2538.

Discrimination measures, marginal frequencies category numbers and category quantifications of these variables are given in table 2.

Table 2. Category labels, c. numbers, c. quantifications, marginal frequencies, discrimination measures of some variables of science test with missing data passive.

		c. qunatifications on two dimensions						Disc. Meas.
Var.	c. label	A	B	C	D	E	M	
	c. Number	1	2	3	4	5	-	
2	Mar. Frec.	95	157	83	3099+	91	775	
	Dim.1	-1.70	-1.19	-1.54	0.19	-1.19	-	0.2182
	Dim.2	0.64	0.18	1.40	-0.53	0.20	-	0.2538
18	Mar. Freq.	521	320	206	218	311+	2724	
	Dim.1	0.21	-0.43	-0.05	-0.21	-0.02	-	0.0214
	Dim.2	0.52	0.18	0.52	0.29	0.28	-	0.0577
22	Mar. Freq.	272	100	609	32	2099+	1188	
	Dim.1	-1.32	-0.96	-1.08	-1.60	0.61	-	0.5011
	Dim.2	-0.09	0.11	-0.67	-2.10	-0.19	-	0.1148
45	Mar. Freq.	256	347	796+	217	421	2263	
	Dim.1	0.07	-0.46	0.86	-0.74	0.02	-	0.1833
	Dim.2	0.20	-0.08	0.07	0.14	-0.23	-	0.0100

+ : correct alternative
M : missing value and illegal answers. Not included in analysis.

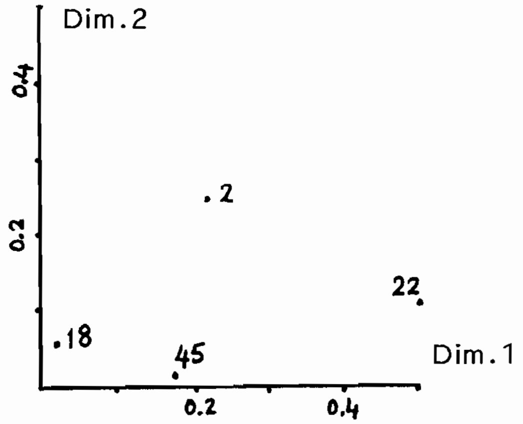


Figure 3. Plot of discrimination measures of variables of table 2.

The HOMALS solution gives m=69 category quantification tables of order 5x2. Sign pattern, rank order of magnitudes of category quantifications on the first dimension for each variable are almost the same as on the second dimension of subsection 3.1.

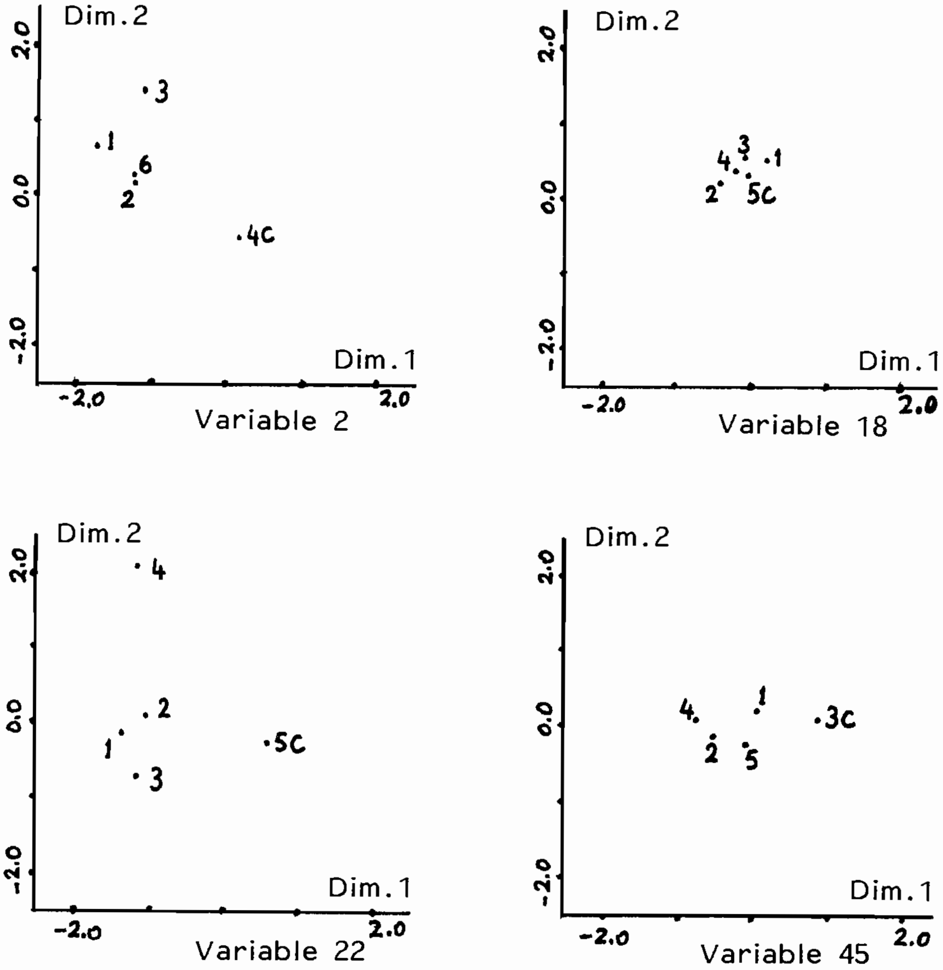


Figure 4. Plot of Category Quantitications of variables labeled by Original Category Numbers in table 2.
(C) Correct Alternatives.

Figure 4 gives some insight into the data. The right edge of the plots indicate correct alternatives, upper left of the plots indicate bluntly wrong alternatives, lower left of the plots indicate slightly wrong alternatives. This shows that the second dimension reflects bluntly wrong versus slightly wrong alternatives. Plot also shows that the two dimensions are acting together.

The plot of unlabeled object scores did not look like a parabola, so that again we may conclude that there is no dominant one dimensional solution.

3.3 HOMALS analysis of social science test with missing data single category

There are n=4300 objects, m=73 variables and k=6 categories. We have same type of matrices in this solution as we had in subsection 3.1.
We can see the following from the table of marginal frequencies:
The missing data frequency is the lowest for variable 4 and equal to 181.
The missing data frequency is the highest for variable 56 and equal to 2502.
The correct alternative frequency is the lowest for variable 32 and equal to 196.
The correct alternative frequency is the highest for variable 68 and equal to 3516.
The wrong alternative frequency is the lowest for alternative E of variable 68, and equal to 33.
The wrong alternative frequency is the highest for alternative A of variable 45, and equal to 2310.
A closer look at these variables shows the following:
Variables 4 and 68 seem very easy. Variable 56 seems to be a science question. Variable 32 looks like a philosophical question. Alternative A of variable 45 seems slightly wrong.

We have again two dimensions. The eigenvalues are λ_1 = 0,1739 and λ_2 = 0,1053.
We can write the following from the table of discrimination measures:
The discrimination measure on the first dimension is the lowest for variable 4 and equal to 0,0590.
The discrimination measure on the first dimension is the highest for variable 28 and equal to 0,3204.
The discrimination measure on the second dimension is the lowest for variable 20 and equal to 0,0160.
The discrimination measure on the second dimension is the highest for variable 25 and equal to 0,2509.

Table 3. Category labels, c. numbers, c. quantifications, marginal frequencies discrimination measures of some variables of social test with missing data single category.

Var.	c. Quantifications on two dimensions								Disc.Meas.
	c. Label	M	A	B	C	D	E	IL	
	c. Number	1	2	3	4	5	6	-	
4	Mar.Frec.	118	111	165	389	3312+	138	4	
	Dim.1	-1.11	0.24	0.10	0.23	0.01	0.15	-	0.0590
	Dim.2	-0.51	-1.09	-0.58	-0.33	0.16	-0.80	-	0.1058
20	Mar.Freq.	1589	572+	917	377	372	471	2	
	Dim.1	-0.72	0.43	0.45	0.31	0.47	0.42	-	0.3059
	Dim.2	0.00	0.19	0.09	-0.13	-0.29	-0.07	-	0.0160
25	Mar.Frec.	959	2666+	164	201	116	192	2	
	Dim.1	-0.90	0.19	0.50	0.49	0.65	0.57	-	0.2509
	Dim.2	-0.63	0.41	-0.52	-0.52	-0.85	-1.03	-	0.2804
28	Mar.Freq.	1824	977	634+	304	322	235	4	
	Dim.1	-0.65	0.45	0.37	0.62	0.62	0.55	-	0.3204
	Dim.2	-0.13	0.04	0.59	-0.21	-0.30	-0.09	-	0.0701
32	Mar.Freq.	1346	515	196+	819	861	560	3	
	Dim.1	-0.78	0.39	0.47	0.34	0.36	0.31	-	0.2808
	Dim.2	-0.26	0.05	-0.10	0.17	0.37	-0.19	-	0.0592
45	Mar.Freq.	699	2310	459	61	713+	51	7	
	Dim.1	-0.95	0.16	0.21	0.55	0.18	0.60	-	0.1792
	Dim.2	-0.04	0.16	-0.36	-1.24	-0.01	-1.78	-	0.0862
56	Mar.Freq.	2502	241	677+	309	328	241	2	
	Dim.1	-0.27	0.31	0.19	0.58	0.61	0.40	-	0.1149
	Dim.2	0.10	-0.35	0.32	-0.53	-0.29	-0.51	-	0.0703
68	Mar.Freq.	297	45	289	3516+	117	33	3	
	Dim.1	-0.99	0.35	0.26	0.05	0.20	0.19	-	0.0774
	Dim.2	-0.02	-1.29	-0.56	0.09	-0.48	-1.16	-	0.0621

+ = Correct alternative
IL = Illegal answer, not included in analysis

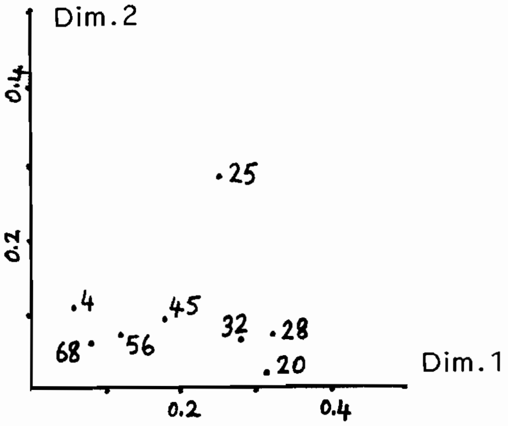


Figure 5. Plot of discrimination measures of variables of table 3.

Variables 28; 27; 32; 22; 27; 31; 18; 30; 33; ... have larger discrimination measures on the first dimension. Variables 25; 14; 33; 12; 24; have larger discrimination measures on the second dimensions.
Variables 33; 25; 14; ... have larger discrimination measures on both dimensions (see figure 5).

Category quantifications of table 3 show that negative quantifications go together with missing data and positive quantifications with alternative answers of the variables on the first dimension (see figure 6). This is also true for 65 remaining variables. Therefore dimension 1 reflects missing versus nonmissing variables.
In six variables out of eight in table 3, we see that positive quantification goes with correct alternative on the second dimension. This is also true for almost all 65 remaining variables. Therefore the second dimension reflects correct alternative versus wrong. There are some alternatives where an incorrect alternative has largest positive quantification on the second dimension; these are slightly wrong alternatives which are relatively often selected by good students, e.g. wrong alternative A of variable 45 and wrong alternatives A, C and D of variable 32.
Since plot of object scores did not have a parabolic shape, one can say the solution in two dimensions is better than a one-dimensional solution. In other words there is not a dominant one-dimensional solution for this data.

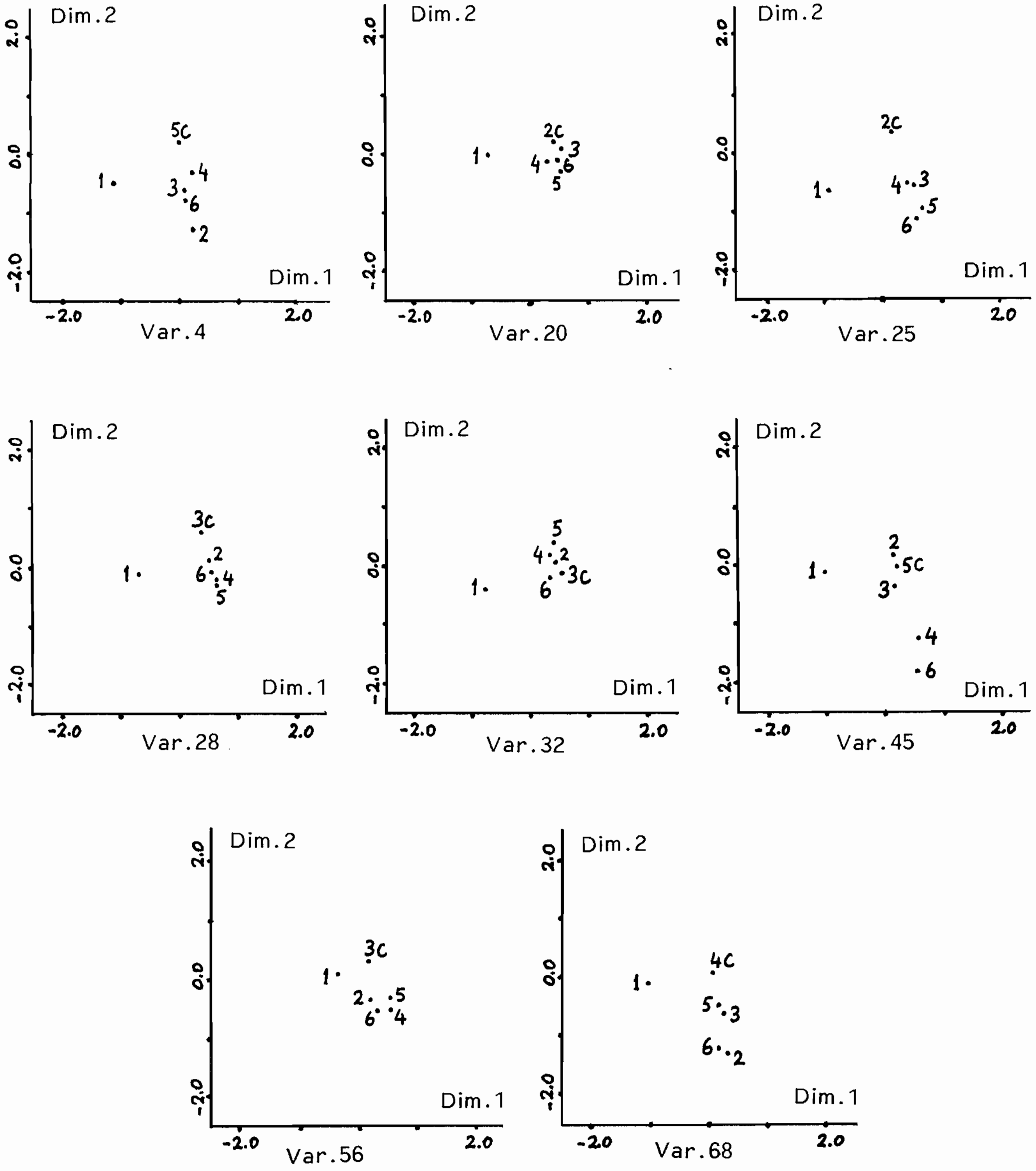


Figure 6. Category quantifications of variables labeled by original category number in table 3.
C : correct alternatives.

3.4 HOMALS analysis of social science test with missing data passive

Since there is no dominant one-dimensional solution, we merged first category and category of illegal data of subsection 3.3 under the label of missing to force the interesting information into the first two dimensions. There are five categories (k=5): alternative A, alternative B, alternative C, alternative D, alternative E with corresponding category numbers 1; 2; 3; 4; 5.
We have two dimensions (p=2) with eigenvalues $\lambda_1 = 0{,}1233$ and $\lambda_2 = 0{,}0424$.
The second dimension of subsection 3.3 now becomes the first dimension of this analysis and discriminates between good students versus bad students.
The discrimination meas. on the first dimension is the lowest for variable 31 and equal to 0,0143.
The discrimination meas. on the first dimension is the highest for variable 12 and equal to 0,2789.
The discrimination meas. on the second dimension is the lowest for variable 24 and equal to 0,0086.
The discrimination meas. on the second dimension is the highest for variable 62 and equal to 0,1483.
Marginal Frequency Discrimination Measures and Category Quantifications of selected Variables are given in table 4.

Variables 12; 40; 25; 24; 14; 7; 1; 19; 43; 34; 36; ... have larger discrimination measures on the first dimension.
Variables 62; 52; 63; ... have larger discrimination measures on the second dimension.

Since discrimination measures of variables 31; 59; 20; 55; 47; 32; ... are the lowest on the first dimension, they do not discriminate between examinees with many correct answers on other variables versus individuals with few correct answers. These variables are not adequate for selection of good examinees.

Table 4. Category labels, c. numbers, c. quantifications, marginal frequencies, discrimination measures of some variables of social science test without missing data.

Var.	c. Quantifications on two dimensions							Disc. Meas.
	c. Label	A	B	C	D	E	M	
	c. Number	1	2	3	4	5	-	
12	Mar.Freq.	591	945	326	269	1013+	1156	
	Dim.1	0.05	-0.26	-0.91	-0.22	.023	-	0.2789
	Dim.2	0.02	0.08	0.06	0.04	-0.16	-	0.0223
24	Mar.Freq.	324	317	2928+	99	330	302	
	Dim.1	-0.85	-0.67	0.33	-1.25	-0.72	-	0.2402
	Dim.2	0.07	-0.02	-0.10	0.10	0.12	-	0.0086
31	Mar.Freq.	227	235	594+	253	1709	1232	
	Dim.1	-0.06	-0.10	0.15	-0.41	-0.04	-	0.0143
	Dim.2	-0.05	0.01	-0.37	-0.21	-0.16	-	0.0326
62	Mar.Freq.	448	2321+	343	227	121	840	
	Dim.1	0.15	-0.05	-0.22	-0.69	-0.75	-	0.0846
	Dim.2	-0.15	0.46	-0.46	-0.47	-0.25	-	0.1483

\+ = correct alternative
M = missing value and illegal answer. Not included in analysis

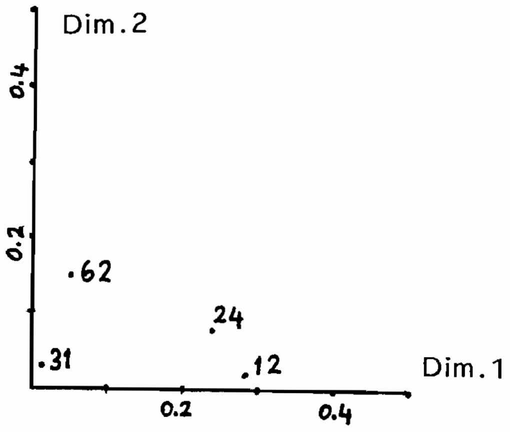


Figure 7. Plot of discrimination measures of variables of table 4.

There are 73 tables of category quantifications of order 5x2. Sign pattern and rank order of magnitudes of category quantifications on the first dimension for each variable are almost the same as on the second of subsection 3.3. We have plots of category quantifications of same variables in figure 8. Neither this figure nor plots of category quantifications labeled by the original category numbers give more insight to the data.

Plot of unlabeled object scores did not give a clue for a special nonlinear relation between the dimensions. Therefore two dimensions are better than one dimension in explaining the data.

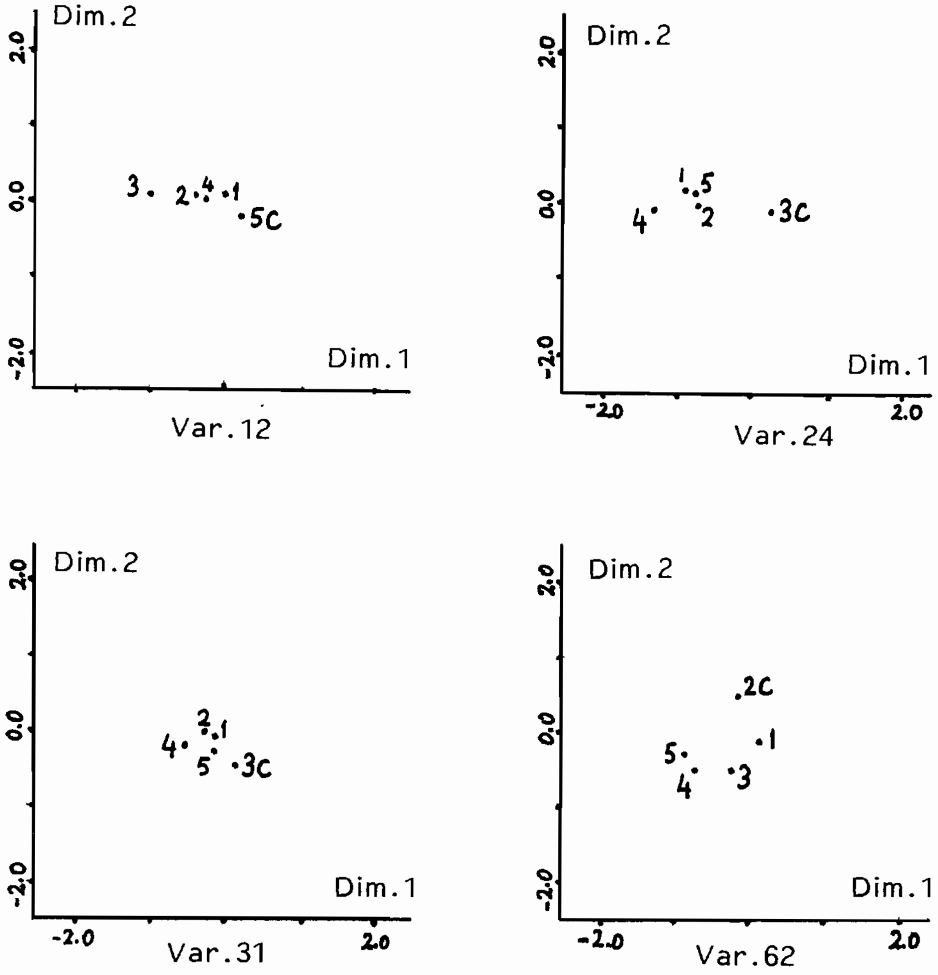


Figure 8. Category quantifications labeled by the original category numbers.
C : correct alternative.

3.5 HOMALS analysis of data in three dimension with missing data passive

We also calculated HOMALS solution for the science test in p=3 dimensions with missing data passive. Since HOMALS solutions are nested (Gifi, 1983) the first two dimensions are the same as in the

twodimensional counterparts. In the science test there is a slight evidence that the third dimension is related to variables at the end of the test, and which are not attained by a number of students.
Variables 64, 63, 67, 66, 62, 65, 68, 9, 16, 69, ... have larger discrimination measures on the third dimension. Eight of these ten variables are the last variables of two tests, since social science test is given to examinees before the science test. These variables can be denoted as nonattainable at least for many examinees due to lack of time at the end of two tests.

3.6 PRIMALS and principal component analysis of science test with missing data single category

The PRIMALS solution gives the matrix of marginal frequencies of order 69x6, the vector of discrimination measures of order 69x1 and a matrix of category quantifications of order 69x6. These are equal to their counterpart in the first dimension of the related matrices of subsection 3.1.
The program also provides for a matrix of component loadings of order 69x2 (p=2) of data matrix after transformation, with eigenvalues λ_1 = 15,438 and λ_2 = 4,022. The first eigenvalue is almost equal to the corresponding eigenvalue $m\lambda_1$ = 15,3387 in subsection 3.1. Squares of loadings on the first principal component are equal to the related discrimination measures in subsection 3.1. This property holds for full data. Since we did not include the category of illegal answers, there are very small discrepancies at at the third or fourth decimals.
We can write the following from the matrix of component loadings:
All component loadings on the first dimension are positive.
Component loadings are highest for variables 9 and 57 on the first dimension and equal to 0.577.
Component loading is lowest for variable 64 on the first dimension and equal to 0.212.
Component loading is highest for variable 66 on the second dimension and equal to 0.508.
Negative component loading is highest in absolute value for variable 8 on the second dimension and equal to -0.360.
Component loading is lowest in absolute value for variable 34 on the

second dimension and equal to 0.000.

Variables 3; 5; 7; 9; 11; 19; 20; 23; 25; 29; 32; 33; 34; 35; 36; 37; 38; 40; 41; 42; 45; 48; 53; 55; 56; 58; 59; 60; 61 have larger component loadings on the first dimension. These variables are of intermediate degree of difficulty (moderately difficult variables).

Variables 1; 2; 3; 4; 7; 8; 9; 21; 22; 29; ... have negative but bigger component loadings in absolute value on the second dimension. These are easy variables.

Variables 38; 39; 42; 50; 51; 52; 62; 63; 64; 65; 66; 67; 68; 69; ... have bigger component loadings on the second dimension. These are difficult variables.

Of course one may constract two more categories of "somewhat difficult" and "somewhat easy" variables.

Variables 3; 7; 9; ... are "halfway between" of "easy" and "moderately difficult" variables. These variables may be considered as "somewhat easy" variables. Similarly variables 38; 42; ... are "somewhat difficult" variables.

Component loadings after transformation provides for an ordering of the variables from easy to difficult (Geer, J.P. van de and Meulman, J., 1984).

We also have the PRIMALS analysis of science test with missing data passive. Since this PRIMALS solution has one dimension which is equal to first dimension of HOMALS solution we do not include it here. Component loadings of the two dimensions did not reflect the pattern of the loadings of the same solution with missing data single category. These results correspond to the findings in the HOMALS analysis.

3.7 PRIMALS and principal component analysis of social science test with missing data single category

PRIMALS solution of the social science test provides for a matrix of marginal frequencies of order 73x6, a vector of discrimination measures of order 73x1, a matrix of category quantifications of order 73x6. These are the same as their counterpart at the first dimension of the related matrices in subsections 3.3.

The PRIMALS solution also gives a matrix of component loadings of the data matrix after transformation of order 73x2 with eigenvalues

λ_1 = 12.786 and λ_2 = 4.849. The first eigenvalue almost equal to the eigenvalue $m\lambda_1$ = 12.7047 in subsection 3.3. Squares of component loadings on the first principal components are equal to the related discrimination measures in subsection 3.3. The first column of the table of discrimination measures of the related HOMALS solution is the same as the vector of discrimination measures in this subsection. Small discrepancies come from excluding illegal answers in the analysis. As could be predicted from all loadings on the first dimension are positive. We can write the following from the matrix of component loadings:

Component loading is the highest for variable 28 on the first dimension and equal to 0.566.

Component loading is the lowest for variable 4 on the first dimension and equal to 0.244.

Component loading is the highest for variable 23 on the second dimension and equal to 0.446.

Negative component loading in absolute value is the highest for variable 62 on the second dimension and equal to -0.391.

Component loading is the lowest in absolute value for variable 11 on the second dimension and equal to -0.043.

Variables 10; 13; 14; 15; 16; 17; 18; 20; 21; 22; 23; 25; 26; 27; 28; 29; 30; 31; 32; 33; 35; 38; 39; 41; 42; 43; 45; 47; 48; 49; 66; 67; 68; 72; ... have larger loadings on the first dimension. These variables are of intermediate degree of difficulty.

Variables 52; 54; 60; 61; 62; 63; 64; 65; 70; 71; 72; ... have negative but larger component loadings in absolute value on the second dimension. These variables may be considered more difficult than the others.

Variables 18; 19; 21; 22; 23; 24; 25; 26; 27; 29; 30; 32; 33; ... have larger component loadings on the second dimension. These variables may be considered more easy than the others.

Easy variables are all related to Turkish language. Difficult variables are related to geography, psychology, sociology and history. Moderately difficult variables are related to about every subject that social science test covers.

Component loadings after transformation may provide us with an ordering of variables from easy to difficult. This ordering not as obvious as the ordering in the science test.

3.8 Conclusions

There are no contradictions between the results of the different types of analysis if we consider full data (missing data single category). We may relay on the two dimensional HOMALS solution in this case, because this analysis uses all available data and satisfies all restrictions of HOMALS analysis.

HOMALS analysis with missing data single category indicates that the first HOMALS dimension is related to categories of missing versus non-missing. The second HOMALS dimension related to examinees success. If we delete the category of missing data, the first HOMALS dimension is related to examinees success and the second HOMALS dimension is related to slightly wrong alternatives versus bluntly wrong alternatives. These are the three main dimensions of each test which can be considered good.

There is a slight evidence of existence of another dimension in the three dimensional HOMALS solution of science test with missing data single category. This dimension is related to variables attainable versus nonattainable.

Using the category quantifications of variables in the second dimension of HOMALS solution with missing data single category as the weights, one can construct a scale which is better than the scale employed up till now for student selection examination, in the sense that the new scale makes a better solution of good versus bad examinees.

If it were the case that the same variables will be used in the test during the forthcoming years, than it would be logical to exclude the questions with low discrimination measures in the second dimension of HOMALS solution with missing data single category. However this remark is not pertinent since the Turkish system does not use the same variables in the following years. On the other hand it is advisable to draw the attention of those making the variables to the discriminatory property of the variables.

Using the loadings of principal components after transformation one can order the variables from "easy" to "difficult". This is also advisable to draw the attention of those preparing tests.

APPENDIX TRANSLATION OF SOME OF THE QUESTIONS

Since the 'question book' is written in Turkish, it will be good to translate some of the variables (questions) for better understanding.

Variable 2

??
8|94
..

Reminder ?

What will be the remainder of this division problem.

A)3, B)4, C)5, D)6, E)7.

Variable 18

Two objects are racing on a circle, the second one is finished the two laps race 5 minutes earlier than the first. In which one of the following cases, the result of race would change:

A) if the race was done on a straight road which was twice as long as the circumference of the circle.

B) if we doubled the radius and made a one lap race.

C) if the radius was halved and the speeds of the racers were also halved.

D) if the radius was halved and the race incresed to four laps

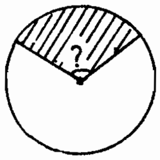
E) if the speeds of the racers were doubled and the race reduced to one lap.

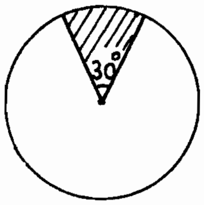
Variable 22

Ratio of sides of two squares is 1/6. What is the ratio of area of the same squares?

A) 1/12 B) 1/18 C) 1/24 D) 1/30 E) 1/36

Variable 26

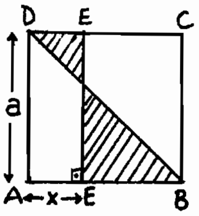




Radius of the smaller of the two circles is r, radius of the larger circle is 2r. Areas of the shaded sectors are the same. If measurement of <ABC is 30°, what is the measurement of <DEF?

A) 60°; B) 90°; C) 100°; D) 120°; E) 150°

Variable 30



ABCD is a square. $EF \perp AB$, $|AE| = x$, $|AD| = a$ are given. Which expression is correct for the shaded areas?

A) $x^2 + ax + a^2$ B) $2x^2 - ax - \frac{a^2}{2}$

C) $x^2 + 2ax + \frac{a^2}{4}$ D) $2x^2 + 2ax + a^2$

E) $x^2 - ax + \frac{a^2}{2}$

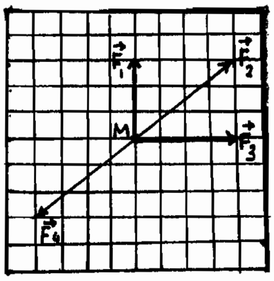
Variable 36

Four different force acting upon the object M on the plane. Which two of the forces would determine the resultant vector?

A) $\vec{F}_1$ and $\vec{F}_3$ B) $\vec{F}_1$ and $\vec{F}_2$

B) $\vec{F}_3$ and $\vec{F}_4$ D) $\vec{F}_1$ and $\vec{F}_4$

E) $\vec{F}_2$ and $\vec{F}_3$



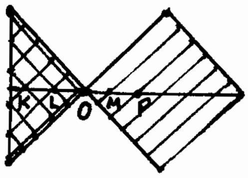
Variable 37

Two identical squares are cut of a same flat, homogenious and thin metal. One of them is folded along its diagonal and these two pieces are placed together as in the diagram. The parallel lines in the diagram are of equal distance.

Where is the center of gravity of these connected squares?

A) on point O, B) on point M, C) between the points L and O,

D) between the points O and M, E) between the points M and P.



Variable 45

Boiling and freezing points of X, Y and Z are as such:

	Freezing Pt. (°C)	Boiling Pt. (°C)
X	-88	57
Y	-12	118
Z	26	83

All are increased from -20°C to 100°C. How many of the X, Y and Z undergo change of phase or phases?

A) only X B) only Y C) only Z

D) X and Y E) X and Z

Variable 59

A mole of chemical compound with a molecular weight of 224 gr. is composed of 2 mol of X and 7 mol of Y atoms. Since 50% of the weight of this compound is element X. What are the atomic weights of X and Y?

	X	Y
A)	56	16
B)	28	32
C)	50	25
D)	64	16
E)	56	28

Variable 64

To get the identical plant as the first one which of the followings should be done?

A) to plant seeds

B) reverse crossing

C) to use runners

D) to let it polenate by itself

E) to let it polenate through insects

Variable 68

Malt with congo red is added as paremessium culture. In the beginning, nutrition cofuls of paremessium were red, then they turned to blue in a couple of minutes. Which one of the following is correct? (Congo red becomes blue in acidic mixtures and becomes red in basic mixtures).

A) Glucosen turns to glucose

B) Synthesis of fat

C) Sucrose dissolves to its parts

D) Digestion of proteins

E) Cellulose dissolves to its simple units.

REFERENCES

Gifi, A., Non-Linear Multivariate Analysis, Dept. of Data Theory, Univ. of Leiden, 1981.

Gifi, A., HOMALS Users Guide, Dept. of Data Theory, Univ. of Leiden, 1983.

Ögrenci Seçme ve Yerleştirme Sinavi, Birinci Basamak (ÖSS) Soru Kitapçigi A, Meteksan, Ankara, 1984.

Geer, J.P. van de, Meulman J., Heijden, P. van der, Heiser W., How to use HOMALS, (concept october 1984), Dept. of Data Theory, Univ. of Leiden, 1984.

Geer, J.P. van de and Meulman J., PRIMALS Users Guide, Dept. of Data Theory, Univ. of Leiden, (in press).

Leeuw, J. de, Beyond Homogenity Analysis, Department of Data Theory, Univ. of Leiden, 1984.