埼玉大学紀要　教育学部，66(2)：595－607（2017）

A Statistical Method Based on Multivariate Analysis in The Comprehensive Evaluation of Achievement Test: From The Standpoint of Principal Components Analysis

DÔKU, Isamu
Faculty of Education, Saitama University

Summary

In this article we introduce a statistical method based upon Multivariate Analysis in order to evaluate the achievement test in a comprehensive manner. Here the so-called principal components analysis does play an important role. Some examples are given as well to illustrate great usefulness of the method.

Key Words: statistical method, multivariate analysis, comprehensive evaluation, achievement test, principal components analysis.

1. Introduction

In this article we shall introduce a statistical method based upon Multivariate Analysis in order to evaluate the achievement test in a comprehensive manner, where the so-called principal components analysis does play an important role. Lastly, we will provide with some examples, which illustrate great usefulness of the method. Generally, data obtained from the investigation via questionnaire consist in plenty of variates, terms or characteristics, where lots of primary factors are recorded usually (Table 1). If that is the case, although it is important to analyze each primary factor one by one, it is much more important to search for comparatively more influential primary factors and their better combination, grasping mutual relations. The multivariate analysis is one of the strongest methods to summarize the characteristic features of the given data by taking the strength of correlation between several variates into consideration. It is the principal components analysis method that we make use of in this article.

2. Fundamental statistics and vector representation

Better understanding of mathematical scheme in the theory of multivariate analysis would be achieved by considering those given observables as components of the vector. The mean (or average), variance, standard deviation, correlation coefficient, and variance of the composite of variables are fundamental concepts, which are required to analyze the object in the theory of multivariate analysis. Let us now consider representing those fundamental quantities or statistics in terms of vectors and matrices.

表 1 **multivariate data**

Objects No.	X_1	$\cdots$	Variates X_{i}	$\cdots$	X_p
1	x_{11}	$\cdots$	x_{1i}	$\cdots$	x_{1p}
.	.		.		.
.	.		.		.
.	.		.		.
j	x_{j1}	$\cdots$	$x_{j\mathrm{i}}$	$\cdots$	x_{jp}
.	.		.		.
.	.		.		.
.	.		.		.
n	x_{n1}	$\cdots$	x_{ni}	$\cdots$	x_{np}
mean	$\overline{x}_1$	$\cdots$	$\overline{x}_i$	$\cdots$	$\overline{x}_p$
S.D.	s_{X1}	$\cdots$	$s_{X\mathrm{i}}$	$\cdots$	s_{Xp}

When $x_{1i}, \ldots, x_{ni}(i = 1, 2, \ldots, p)$ denote p-pieces of inspected values that were obtained from the experiment executed to n testees, then p pieces of n-dimensional vector whose components are given by its multiplication by $1/\sqrt{n}$, can be expressed as

$$\mathbf{x}_i = \frac{1}{\sqrt{n}}(x_{1i}, \ldots, x_{ni}), \qquad \text{for} \quad i = 1, 2, \ldots, p. \tag{1}$$

For the unit vector **e** in the n-dimensional space, namely, $\mathbf{e} = (1, 1, \ldots, 1)/\sqrt{n}$, when we make an inner product for that with $\overline{\mathbf{x}}_i$, it will be given by

$$\langle \mathbf{x}_i, \mathbf{e} \rangle = \frac{1}{n}\sum_{j=1}^{n} x_{ji} = \bar{x}_i, \tag{2}$$

and this is nothing but the mean (or average) of the variate X_i. If we set

$$\bar{\mathbf{x}}_i = \frac{1}{\sqrt{n}}(\bar{x}_{1i}, \ldots, \bar{x}_{ni}), \qquad \text{for} \quad i = 1, 2, \ldots, p, \tag{3}$$

it follows therefore that

$$\bar{\mathbf{x}}_i = \bar{x}_i \mathbf{e} = \langle \mathbf{x}_i, \mathbf{e} \rangle \mathbf{e} = P\mathbf{x}_i, \tag{4}$$

where P is the projection from an n-dimensional space to $\{\mathbf{Ce}\}$. Hence, $\overline{\mathbf{x}}_i$ is nothing but the orthogonal projection of $\mathbf{x}_i$ onto **e**. Moreover, we put for convenience

$$\begin{aligned} \tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}_i &= \frac{1}{\sqrt{n}}(x_{1i} - \bar{x}_i, \ldots, x_{ni} - \bar{x}_i) \\ &= (I - P)\mathbf{x}_i. \end{aligned} \tag{5}$$

Next we shall make a square of the norm for the vector $\tilde{\mathbf{x}}_i$, then it follows that

$$\|\tilde{\mathbf{x}}_i\|^2 = \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i \rangle = \frac{1}{n}\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)^2 = s_{X_i}^2. \tag{6}$$

This quantity is the variance of the variate X_i. Since we have

$$\langle \bar{\mathbf{x}}_i, \tilde{\mathbf{x}}_i \rangle = \langle P\mathbf{x}_i, (I-P)\mathbf{x}_i \rangle = 0, \tag{7}$$

we can easily get

$$\|\mathbf{x}_i\|^2 = \|\tilde{\mathbf{x}}_i\|^2 + \|\bar{\mathbf{x}}_i\|^2 = s_{X_i}^2 + \bar{x}_i^2. \tag{8}$$

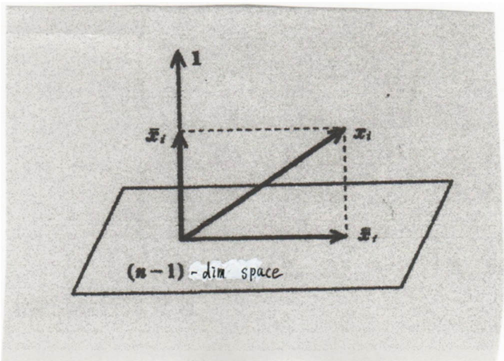

図 1 **Projection figure**

3. Correlation coefficient and covariance

In this section we shall introduce the concepts of covariance and correlation coefficient, and shall define the correlation matrix by employing those notions. The correlation coefficient $r \equiv r_{XiXj}$ is defined as the covariance $s_{XiXj} \equiv s\,(X_i,\, X_j)$ divided by aproduct $s_{Xi}s_{Xj}$ of S.D.'s (i.e., standard deviations) of each variate. For convenience, it would be much better to standardize it in advance as the quantity obtained from variates with variance 1. So that, let us now put

$$\mathbf{z}_i = \frac{\mathbf{x}_i}{s_{X_i}} = \frac{\mathbf{x}_i}{\|\tilde{\mathbf{x}}_i\|}. \tag{9}$$

Then we may consider the inner product for those vectors $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$to get

$$\langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle = \frac{1}{n}\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)\cdot(x_{kj} - \bar{x}_j) = s_{X_iX_j}, \tag{10}$$

which is the covariance of variates X_i and X_j. Therefore, its correlation coefficient is given by

$$r_{X_iX_j} = \frac{s_{X_iX_j}}{s_{X_i}s_{X_j}} = \frac{\langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle}{\|\tilde{\mathbf{x}}_i\| \cdot \|\tilde{\mathbf{x}}_j\|} = \langle \tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j \rangle. \tag{11}$$

Consequently, the covariance matrix S for p pieces of variates X_i $(i = 1, 2, \ldots, p)$ is given by

$$S = \begin{pmatrix} s_{X_1}^2 & s_{X_1X_2} & \cdots & s_{X_1X_p} \\ s_{X_2X_1} & s_{X_2}^2 & \cdots & s_{X_2X_p} \\ \cdots & \cdots & \cdots & \cdots \\ s_{X_pX_1} & s_{X_pX_2} & \cdots & s_{X_p}^2 \end{pmatrix}$$

$$= (\langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle)_{i=1,2,\ldots,p; j=1,2,\ldots,p} \in M(p \times p) \tag{12}$$

and the correlation coefficient matrix R is written as

$$R = \begin{pmatrix} 1 & r_{X_1X_2} & \cdots & r_{X_1X_p} \\ r_{X_2X_1} & 1 & \cdots & r_{X_2X_p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{X_pX_1} & r_{X_pX_2} & \cdots & 1 \end{pmatrix}$$

$$= (\langle \tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j \rangle)_{i=1,2,\ldots,p; j=1,2,\ldots,p} \in M(p \times p). \tag{13}$$

4. Composite variate in multivariate analysis

It is of extreme importance to make up a composite variate in multivariate analysis. In this section we shall investigate how the variance of composite variate which is obtained by summing variates with a proper weight, is related to the variance of the original variate. In addition to that, we shall try to express their relation with vectors.

For p pieces of vectors $\mathbf{x}_i$ $(i = 1, 2, \ldots, p)$, and α_i's with

$$\sum_{i=1}^{p} \alpha_i^2 = 1 \quad \text{and} \quad \alpha_i > 0,$$

the composite variate vector is given by

$$\mathbf{F} = \sum_{i=1}^{p} \alpha_i \mathbf{x}_i, \tag{14}$$

and its vector $\tilde{\mathbf{F}}$ is also given by

$$\tilde{\mathbf{F}} = \sum_{i=1}^{p} \alpha_i \tilde{\mathbf{x}}_i \qquad \text{with} \quad \sum_{i=1}^{p} \alpha_i^2 = 1. \tag{15}$$

Consequently, the variance of the composite variate F is computed and given by

$$s_F^2 = \|\tilde{\mathbf{F}}\|^2 = \langle \tilde{\mathbf{F}}, \tilde{\mathbf{F}} \rangle = \left\langle \sum_{i=1}^{p} \alpha_i \tilde{\mathbf{x}}_i, \sum_{j=1}^{p} \alpha_j \tilde{\mathbf{x}}_j \right\rangle$$

$$= \sum_{i=1}^{p} \alpha_i \sum_{j=1}^{p} \alpha_j \langle \tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j \rangle = \langle S\alpha, \alpha \rangle, \tag{16}$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)$. For the composite variate vector $\mathbf{G}$, i.e.

$$\mathbf{G} = \sum_{i=1}^{p} \alpha_i \mathbf{z}_i \qquad \text{with} \quad \sum_{i=1}^{p} \alpha_i^2 = 1,$$

we can have

$$\tilde{\mathbf{G}} = \sum_{i=1}^{p} \alpha_i \tilde{\mathbf{z}}_i. \tag{17}$$

While, the variance of the composite variate G is given by

$$\begin{aligned} s_G^2 &= \|\tilde{\mathbf{G}}\|^2 = \langle \tilde{\mathbf{G}}, \tilde{\mathbf{G}} \rangle = \left\langle \sum_{i=1}^{p} \alpha_i \tilde{\mathbf{z}}_i, \sum_{j=1}^{p} \alpha_j \tilde{\mathbf{z}}_j \right\rangle \\ &= \sum_{i=1}^{p} \alpha_i \sum_{j=1}^{p} \alpha_j \langle \tilde{\mathbf{z}}_i, \tilde{\mathbf{z}}_j \rangle = \langle R\alpha, \alpha \rangle \quad \text{with} \quad \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p). \end{aligned} \tag{18}$$

Since R is a symmetric matrix, we are able to get the following representation

$$R = \bigoplus_{i=1}^{p} \lambda_i \mathbb{E}_i \qquad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p, \tag{19}$$

where $\mathbb{E}_i$ is a projection onto a one-dimensional eigenspace (or characteristic space) mutually orthogonal, peculiar to the eigenvalue λ_i of R. Hence it follows immediately that

$$\begin{aligned} s_G^2 &= \langle R\alpha, \alpha \rangle = \sum_{i=1}^{p} \lambda_i \langle \mathbb{E}_i \alpha, \alpha \rangle = \sum_{i=1}^{p} \lambda_i \|\mathbb{E}_i \alpha\|^2 \\ &\leqslant \lambda_i \sum_{i=1}^{p} \|\mathbb{E}_i \alpha\|^2 = \lambda_i \left\| \bigoplus_{i=1}^{p} \mathbb{E}_i \alpha \right\|^2 = \lambda_1 \|\alpha\|^2 = \lambda_1 \sum_{i=1}^{p} \alpha_i^2 = \lambda_1. \end{aligned} \tag{20}$$

Furthermore, it is easy to see that

$$s_G^2 = \sum_{i=1}^{p} \lambda_i \|\mathbb{E}_i \alpha\|^2 \geq \lambda_p \sum_{i=1}^{p} \|\mathbb{E}_i \alpha\|^2 = \lambda_p, \tag{21}$$

so that, we can obtain

$$\lambda_p \leqslant s_G^2 \leqslant \lambda_1. \tag{22}$$

Here, λ_1 is the maximum eigenvalue of R and λ_p is the minimum eigenvalue of R.

5. Cumulative coefficient of determination

In this section we shall introduce the notion of coefficient of determination, and define the cumulative coefficient of determination. Since we can regard each $\tilde{\mathbf{z}}_i$ as a series of vectors being linearly independent, it follows that $\|\tilde{\mathbf{G}}\| > 0$, in other words,

$$\langle R\alpha, \alpha\rangle = s_G^2 = \|\tilde{\mathbf{G}}\|^2 > 0. \tag{23}$$

On this account, we can deduce naturally that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0. \tag{24}$$

When we denote by α_i $(i = 1, 2, \ldots, p)$ the unit eigenvectors subordinated to the eigenvalues λ_i of R, the composite variate vector

$$\mathbf{G}_i = \sum_{j=1}^{p} \alpha_{ij} z_j \qquad \text{with} \quad \sum_{j=1}^{p} \alpha_{ij}^2 = 1 \tag{25}$$

weighted by each component α_{ij} is called the i-th principal component. Since $\langle \alpha_i, \alpha_j\rangle = 0$ for $i \neq j$, we observe with ease that

$$\begin{aligned} s_{G_iG_j} &= \langle \tilde{\mathbf{G}}_i, \tilde{\mathbf{G}}_j\rangle = \left\langle \sum_{k=1}^{p} \alpha_{ik}\tilde{\mathbf{z}}_k, \sum_{\ell=1}^{p} \alpha_{j\ell}\tilde{\mathbf{z}}_\ell \right\rangle \\ &= \sum_{k=1}^{p} \alpha_{ik} \sum_{\ell=1}^{p} \alpha_{j\ell} \langle \tilde{\mathbf{z}}_k, \tilde{\mathbf{z}}_\ell\rangle = \langle R\alpha_j, \alpha_i\rangle = \lambda_j \langle \alpha_j, \alpha_i\rangle = 0. \end{aligned} \tag{26}$$

Moreover, it proves to be that

$$s_{G_i}^2 = \|\tilde{\mathbf{G}}_i\|^2 = \lambda_i \langle \alpha_i, \alpha_i\rangle = \lambda_i. \tag{27}$$

On this account, the total sum of eigenvalues of the correlation coefficient matrix R is equivalent to that of variances of all the principal components. Therefore, we call

$$p_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \tag{28}$$

the coefficient of determination of the i-th principal component, and the total sum of coefficients of determination up to the i-th principal component is also called the cumulative coefficient of determination up to the i-th principal component. Notice that $\lambda_1 + \lambda_2 + \cdots + \lambda_p = p$.

6. Principal components analysis: example 1

Let us now consider the situation where some achievement test consists of five subjects: namely, mother tongue (X_1), mathematics (X_2), social studies (X_3), science (X_4), English (X_5). The comprehensive evaluation of the achievement test is mainly due to the total sum S of the scores (or marks): that is to say,

$$S = X_1 + X_2 + X_3 + X_4 + X_5,$$

then obviously the obtained result is not rational if that is the case where the average and variance of each subject are not equal. Especially, it is clear that the specific subject with large variance does have decisive influence over the total sum, namely, the results of achievement test.

On this account, when the inspected values of distinct kinds are given, then it suffices to find the total score for weighted inspected values of several kinds in order to analyze, compare and evaluate them in a comprehensive manner. It is the principal components analysis that can realize concretely the comprehensive evaluation of such an achievement test, by taking correlation of an aggregate of variates given into consideration as well as by expressing exactly the variance that plenty of variates possess and by computing the weighted total sum of scores in a reasonable manner.

The next table is the results of the achievement test executed at a junior high school in Kanto area. Let us now analyze the principal factors for pupils with good marks. Incidentally, the full scores of X_i's $(i = 1, 2, 3, 4, 5)$ are 250, 200, 200, 200, 200 in numerical order. The size of samples is twenty-eight, namely, n=28, that just corresponds to a class.

表 2 achievement test result no.1

No.	1	2	3	4	5	6	7	8	9	10
X_1	210	133	154	150	139	116	140	121	128	138
X_2	13	11	33	8	4	3	26	4	35	17
X_3	54	62	111	55	53	50	97	45	106	75
X_4	53	60	87	73	70	39	91	51	153	90
X_5	51	77	74	60	44	49	40	35	39	36

表 3 achievement test result no.2

No.	11	12	13	14	15	16	17	18	19	20
X_1	148	150	123	109	131	127	127	101	113	131
X_2	27	20	19	5	2	26	22	11	19	15
X_3	87	84	55	28	29	93	80	46	73	62
X_4	88	85	70	43	66	117	84	79	74	101
X_5	40	31	44	21	63	100	42	41	47	39

Let mother tongue (X_1), mathematics (X_2), social studies (X_3), science (X_4), and English (X_5) be our analytic variates. The averages $(\overline{x}_i)$ and SD (s_{Xi}) are computed and given below.

The correlation coefficient matrix R is computed and given below.

表 4 achievement test result no.3

No.	21	22	23	24	25	26	27	28
X_1	104	121	86	120	110	107	87	113
X_2	6	22	8	24	2	23	13	14
X_3	38	81	53	73	28	69	46	45
X_4	70	112	48	125	58	135	86	87
X_5	30	89	38	83	68	36	56	47

表 5 **Table of averages and SD's**

	X_1	X_2	X_3	X_4	X_5
$\overline{x}_i$	126.321	15.4286	63.5	81.9643	50.7143
s_{Xi}	23.7217	9.3596	22.5587	27.1734	18.9659

$$R = \begin{pmatrix} 1 & 0.235678 & 0.34314 & -0.020316 & 0.104512 \\ 0.235678 & 1 & 0.919158 & 0.773093 & 0.202687 \\ 0.34314 & 0.919158 & 1 & 0.649357 & 0.230223 \\ -0.020316 & 0.773093 & 0.649357 & 1 & 0.244674 \\ 0.104512 & 0.202687 & 0.230223 & 0.244674 & 1 \end{pmatrix}$$

We compute the i-th principal components (G_i) $(i = 1, 2, 3, 4, 5)$, eigenvalues, eigenvectors, coefficients of determination, and cumulative coefficients of determination. For simplicity, we use the abbreviations, and e.g.v. stands for the eigenvalue of R, c.d. stands for the coefficient of determination, and c.c.d. stands for the cumulative coefficient of determination.

表 6 **Table of principal components**

	G_1	G_2	G_3	G_4	G_5
e.g.v. of R	2.72824	1.02635	0.910796	0.275144	0.0594524
c.d. (%)	54.5648	20.527	18.21592	5.50288	1.189048
c.c.d.	54.5648	75.0918	93.30772	98.8106	100

Next we shall present the values of the eigenvectors. We shall use the simplified notation for presentation only here. For instance, we use a_1 for a_{i1} for the i-th principal component. For the first principal component, the eigenvector is given by

$$a_1 = 0.199344,\ a_2 = 0.579625,\ a_3 = 0.566614,\ a_4 = 0.501909,\ a_5 = 0.226569.$$

For the second one,

$$a_1 = 0.910652,\ a_2 = -0.077446,\ a_3 = 0.0793738,\ a_4 = -0.392279,\ a_5 = 0.0673457.$$

For the third one,

$$a_1 = -0.0847083,\ a_2 = -0.177542,\ a_3 = -0.152475,\ a_4 = -0.0262406,\ a_5 = 0.968177.$$

For the 4th one,

$$a_1 = -0.350839,\ a_2 = 0.228842,\ a_3 = 0.517846,\ a_4 = -0.742356,\ a_5 = 0.0727026.$$

For the 5th one,

$$a_1 = -0.0266714,\ a_2 = -0.757728,\ a_3 = 0.617443,\ a_4 = 0.205968,\ a_5 = -0.0384624.$$

According to our analysis for the principal components analysis method, we conclude that the principal factors for G_1 consist of the high scores of mathematics, social studies and science; while the the principal factors for G_2 consist of the high scores of mother tongue and science.

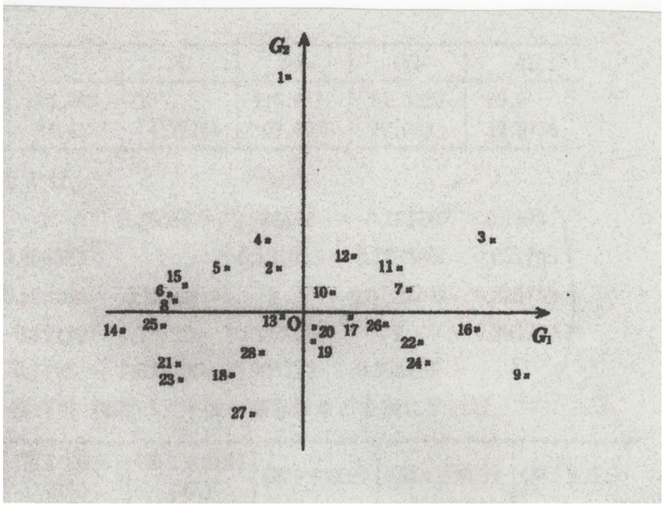

図 2 **G-1, G-2 graph No.1**

Lastly we shall compute the values of G_1 and G_2, and then plot the position of (G_1, G_2) in the graph sheet. The conclusion is clear. Form the graph No.1, we can observe that pupils of no.1, 3, 16, 9 are quite remarkable existence with high scores. We shall show the data of G_1, $G_2)$ just for reference. For example, $Dk = \eta, \zeta$ indicates that the position for the pupil no.k is (η, ζ), in other words, (G_1,G_2) for No.k pupil is equal to (η, ζ).

Here are the data. $D1$=−0.04,0.68. $D3$ =0.55,0.21. $D16$ =0.50,−0.05. $D9$=0.65, −0.19.

7. Principal components analysis : example 2

The next table is the results of the achievement test executed at another junior high school in different Kanto area. Let us now analyze the principal factors for pupils with good marks. Incidentally, the full scores of X_i's $(i = 1, 2, 3, 4, 5)$ are 250, 100, 200, 100, 100 in numerical order. The size of samples is eighteen, namely, $n = 18$, which is a little bit fewer than the previous example.

表 7 **achievement test result no.4**

No.	1	2	3	4	5	6	7	8	9	10
X_1	125	175	183	169	166	161	114	175	127	174
X_2	9	16	16	12	18	13	15	5	13	12
X_3	85	185	144	153	156	95	109	169	104	110
X_4	32	51	32	49	42	31	65	69	75	42
X_5	38	58	58	58	63	57	44	63	49	64

表 8 **achievement test result no.5**

No.	11	12	13	14	15	16	17	18
X_1	140	129	150	203	128	178	142	176
X_2	11	15	14	19	16	28	18	20
X_3	58	93	57	148	88	149	96	73
X_4	52	27	40	46	52	56	93	70
X_5	48	48	49	83	54	80	73	84

Let mother tongue (X_1), mathematics (X_2), social studies (X_3), science (X_4), and English (X_5) be our analytic variates. The averages $(\overline{x}_i)$ and SD (s_{Xi}) are computed and given below.

表 9 **Table of averages and SD's**

	X_1	X_2	X_3	X_4	X_5
$\overline{x}_i$	156.389	15	115.111	51.3333	59.5
s_{Xi}	24.4112	4.77261	37.6089	17.088	12.9754

The correlation coefficient matrix R is computed and given below.

$$R = \begin{pmatrix} 1 & 0.286587 & 0.586262 & -0.121107 & 0.7492 \\ 0.286587 & 1 & 0.12102 & 0.127386 & 0.621705 \\ 0.586262 & 0.12102 & 1 & 0.0115262 & 0.325712 \\ -0.121107 & 0.127386 & 0.0115262 & 1 & 0.351789 \\ 0.7492 & 0.621705 & 0.325712 & 0.351789 & 1 \end{pmatrix}$$

We compute the i-th principal components (G_i) $(i = 1, 2, 3, 4, 5)$, eigenvalues, eigenvectors, coefficients of determination, and cumulative coefficients of determination. For simplicity, we use the abbreviations, and e.g.v. stands for the eigenvalue of R, c.d. stands for the coefficient of determination, and c.c.d. stands for the cumulative coefficient of determination.

表 10 **Table of principal components**

	G_1	G_2	G_3	G_4	G_5
e.g.v. of R	2.42554	1.21882	0.838328	0.47426	0.043048
c.d. (%)	48.5108	24.3765	16.7666	9.4852	0.860959
c.c.d.	48.5108	72.8873	89.6539	99.1391	100

Next we shall present the values of the eigenvectors. We shall use the simplified notation for presentation only here. For instance, we use a_1 for a_{i1} for the i-th principal component. For the first principal component, the eigenvector is given by

$$a_1 = -0.546508,\ a_2 = -0.414885,\ a_3 = -0.396698,\ a_4 = -0.140295,\ a_5 = -0.59342.$$

For the second one,

$a_1 = 0.364592$, $a_2 = -0.331086$, $a_3 = 0.428744$, $a_4 = -0.724871$, $a_5 = -0.219535$.

For the third one,

$a_1 = 0.0124341$, $a_2 = -0.606354$, $a_3 = 0.509504$, $a_4 = 0.606203$, $a_5 = -0.0714419$.

For the 4th one,

$a_1 = 0.454184$, $a_2 = -0.539636$, $a_3 = -0.60807$, $a_4 = 0.00492321$, $a_5 = 0.36433$.

For the 5th one,

$a_1 = 0.601634$, $a_2 = 0.243682$, $a_3 = -0.171658$, $a_4 = 0.295589$, $a_5 = -0.679571$.

According to our analysis for the principal components analysis method, we conclude that the principal factors for G_1 consist of the high scores of mother tongue, mathematics, and English; while the the principal factors for G_2 consist of the high scores of social studies and science.

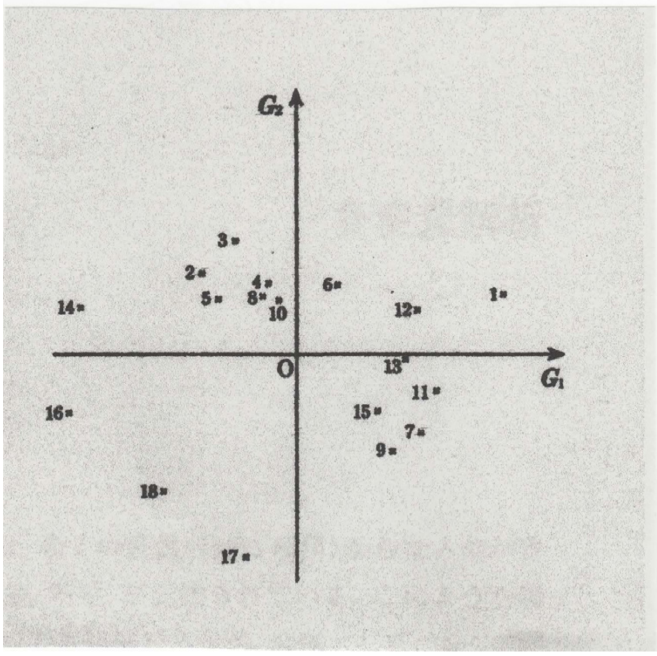

図 3 **G-1, G-2 graph No.2**

Lastly we shall compute the values of G_1 and G_2, and then plot the position of (G_1, G_2) in the graph sheet. The conclusion is clear. Form the graph No.1, we can observe that pupils of no.1, 3, 6, 12 are quite remarkable existence with high scores. We shall show the data of $G_1, G_2)$ just for reference. For example, $Dk = \eta, \zeta$ indicates that the position for the pupil no.k is (η, ζ),in other words, $(G_1, G_2)$for No.k pupil is equal to (η, ζ).

Here are the data. $D1$ =0.63,0.18. $D3$ =−0.18,0.35. $D6$ =0.13,0.21. $D12$=0.37, 0.13.

Acknowledgements

This work is supported in part by Japan MEXT Grant-in-Aids SR (C) 24540114 and also by ISM Coop.

Res. Program: 2016-ISM-CRP-5011.

References

1. Dôku, I. : Exponential moments of solutions for nonlinear equations with catalytic noise and large deviation. Acta Appl. Math. **63** (2000), 101–117.
2. Dôku, I. : Weighted additive functionals and a class of measure-valued Markov processes with singular branching rate. Far East J. Theo. Stat.**9** (2003), 1–80.
3. Dôku, I. : A certain class of immigration superprocesses and its limit theorem. Adv. Appl. Stat. **6** (2006), 145–205.
4. Dôku, I. : A limit theorem of superprocesses with non-vanishing deterministic immigration. Sci. Math. Japn. **64** (2006), 563–579.
5. Dôku, I. : Statistical methods related to educational evaluation. J. Japan Soc. Math. Edu. **89** (2007), 21–36.
6. Dôku, I. : Limit theorems for rescaled immigration superprocesses. RIMS Kôkyûroku Bessatsu, **B6** (2008), 56–69.
7. Dôku, I. : Statistical analysis in educational evaluation. J. Japan Soc. Math. Edu. **91** (2009), 41–50.
8. Dôku, I. : A limit theorem of homogeneous superprocesses with spatially dependent parameters. Far East J. Math. Sci. **38** (2010), 1–38.
9. Dôku, I. : *Probability and Statistics*. Series in Textbooks on Mathematics, Volume 9, Sugaku-Shobo, Tokyo, 2012.
10. Dôku, I. : Star-product functional and unbiased estimator of solutions to nonlinear integral equations. Far East J. Math. Sci. **89** (2014), 69–128.
11. Dôku, I. : An example for convergence of environment-dependent spatial models. J. Saitama Univ. Fac. Educ. (Math. Nat. Sci.) **65** (2016), no.1, 179–186.
12. Dôku, I. : On a limit theorem for environment-dependent models. ISM Res. Rept. **352** (2016), 103–111.
13. Dôku, I. : Applications of environment-dependent models to tumor immunity. RIMS KôKyûroku (Kyoto Univ.), **1994** (2016), 68–74.
14. Dôku, I. : Tumour immunoreaction and environment-dependent models. Trans. Japn. Soc. Indu. Appl. Math. **26** (2016), no.2, 213–252.
15. Dôku, I. : A recursive inequality of empirical measures associated with EDM. J. Saitama Univ. Fac. Educ. (Math. Nat. Sci.) **65** (2016), no.2, 253–259.
16. Dôku, I. : An estimate of survival probability for superprocesses. J. Saitama Univ. Fac. Educ. (Math. Nat. Sci.) **66** (2017), no.1, 259–263.
17. Dôku, I. : On compactness for superprocesses. To appear in ISM Res. Rept. (2017), 8p.
18. Dôku, I. : A support problem for superprocesses in terms of random measure. To appear in RIMS Kôkyûroku (Kyoto Univ.) (2017), 8p.
19. Dôku, I. : Existence of a class of superprocesses by locally finite random measure and pseudodifferential operator. Preprint
20. Dôku, I. : A compact support of superprocesses in terms of locally finite random measure. Preprint

Isamu Dôku
Department of Mathematics
Faculty of Education, Saitama University Saitama, 338-8570 Japan
e-mail : idoku@mail.saitama-u.ac.jp

(Received March 30, 2017)
(Accepted April 17, 2017)