入試成績への主成分分析法の適用について

金　田　　　峣*

On the Application of Principal Component Analysis to the Result of the Entrance Examination

Takashi KANETA

要　旨

多変量解析の手法の一つである主成分分析法の入試成績への適用を試み，分析結果の解釈及び有効な利用の可能性について述べた。

Synopsis

This article tries to apply the principal component analysis, which is one of data analytic methods in multivariate analysis, to the result of the entrance examination, and states the interpretation of the analyzed results and also examines the possibilities of the effective use of the result for the entrance examination.

1．は　じ　め　に

通常，入学者選抜試験（以後入試とする）の成績結果は実施された科目の得点の単純合計点を用い，これに中学校の調査書による内申点，その他を何らかの方法で加味して総合的に合格者を判定している。中学校の内申点をどのような重みで加味するのが適当であるかについては，本校教官立花[1]や松尾・中野[2]，春本・高尾[3]等の研究がある。また各試験科目の平均点や得点分布は異なるのが普通で，これを全く同等の重みで単純合計したものを総合点とすることの合理性については疑問視されるところである。この事に関しては各試験科目の得点を標準化したものを基礎資料として長谷川[4]，立花[5]等が研究している。本稿は総合特性値を得る統計的手法として十分な合理性をもつ主成分分析法を本校のある年度の入試成績に適用してみたもので，基礎資料として標準化した得点を用い，得られた主成分の意味や合格者判定に有効な資料の提供が可能かどうか等について言及する。

＊助教授　一般教科数学

2．主成分分析法の概要

主成分分析法は多変量の統計資料のもつ情報をより低次元の新変量に縮約する統計的手法である。今，同一のサンプルについてのｐ種類の観測値の組がｎ個得られたとする。このｐ変量のデータは同一サンプルについて測定されたものであるから変量相互に関連のある変動を示していると考えられる。これを説明する関数として，ｐ個の変量（$x_1, x_2, \cdots, x_p$）（$p \geqq 2$）の一次結合で

$$Z = l_1 x_1 + l_2 x_2 + \cdots + l_p x_p \qquad (2.1)$$

を仮定する。（ここでは x_p は標準化された標準変量とする。）(2.1) で $\sum_{i=1}^{p} l_i^2 = 1$ の条件のもとでＺの分散が最大になるときのＺが第１主成分と呼ばれているものである。これを Z_1，その係数を l_{1i}（$i = 1, 2, \cdots, p$）とすると，第１主成分 Z_1 は次式で表わされる。

$$Z_1 = l_{11} x_1 + l_{12} x_2 + \cdots + l_{1p} x_p \qquad (2.2)$$

次に同じ条件のもとで Z_1 とは無相関な最大分散をもつＺを Z_2 とする。これが第２主成分で，その係数を l_{2i} とする，以下同様にして Z_m まででｐ個の変量のもつ情報の大部分が説明されていれば主成分を求めることを中止すればよい。Z_m は次式で表わされる。

$$Z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \qquad (2.3)$$

ただし，$\sum_{i=1}^{p} l_{\alpha i}^2 = 1$ ($\alpha = 1, 2, \cdots, m$) とする。

ここで，主成分を求める手順の概略を示すために，p 変量データからの分散共分散行列を求める。ここでは標準変量を基礎データとしているので相関行列になる。それを

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \cdots r_{1p} \\ r_{12} & r_{22} \cdots r_{2p} \\ \vdots & \vdots \quad \vdots \\ r_{1p} & r_{2p} \cdots r_{pp} \end{bmatrix} \quad (r_{ii} = 1,\ i = 1, 2, \cdots, p)$$

で表わす。ベクトル変量を

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \text{そのとり得る値を} \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{bmatrix}$$

$$(j = 1, 2, \cdots, n)$$

また x_i の係数ベクトルを

$$\mathbf{l}_\alpha = \begin{bmatrix} l_{\alpha 1} \\ l_{\alpha 2} \\ \vdots \\ l_{\alpha p} \end{bmatrix} \quad (\alpha = 1, 2, \cdots, m)$$

と表わすならば，主成分は次式で表わされる。

$$Z\alpha = \mathbf{l}'_\alpha \mathbf{X} \quad (\alpha = 1, 2, \cdots, m) \qquad (2.4)$$

ただし，$\mathbf{l}'_\alpha \mathbf{l}_\alpha = 1$ である。($\mathbf{l}'$は転置ベクトル)

第1主成分は $\mathbf{l}'_1\mathbf{l}_1 = 1$ の条件のもとで $Z_1 = \mathbf{l}_1'\mathbf{X}$ の分散

$$V\{Z_1\} = \mathbf{l}'_1\mathbf{R}\mathbf{l}_1 \qquad (2.5)$$

を最大にする$\mathbf{l}_1$を決定することによって求められる。(2.5) にラグランジュの未定乗数法を適用して$\mathbf{l}_1$の要素に関する連立方程式

$$(\mathbf{R} - \lambda\mathbf{I})\ \mathbf{l}_1 = 0 \quad (\mathbf{I}\text{ は p 次の単位行列}) \qquad (2.6)$$

が，0以外の解を持つことに帰着される。このことより非負対称行列 $\mathbf{R}$ は

$$|\mathbf{R} - \lambda\mathbf{I}| = 0 \qquad (2.7)$$

を満たさねばならず，固有値 λ は非負の実数で存在し，$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p \geqq 0$ とするとき $\mathbf{l}_1$は最大固有値 λ_1に対応して求められ Z_1が決まる。

ここに，$V\{Z_1\} = \lambda_1$である。

第2主成分は更に

$$\mathrm{cov}\{Z_1, Z_2\} = 0 \qquad (2.8)$$

の条件が加わり，$\mathbf{l}'_2\mathbf{l}_2 = 1$ のもとで $V\{Z_2\}$ を最大にする$\mathbf{l}_2$を求めればよい。第1主成分のときと同様な方法により (2.6) と同じ連立方程式になり $\lambda_1 \neq \lambda_2$なら λ_2に対応する固有ベクトルとして $\mathbf{l}_2$ が求まる。ここに $V\{Z_2\} = \lambda_2$ で (2.8) より，$\mathbf{l}_1\mathbf{l}_2 = 0$ となり直交している。以下同様に Z_mまで求められる。この他，標準変量の分散が1であり，$\mathbf{R}$ の固有値 λ は主成分の分散であったことから，λ の総和は tr$\mathbf{R}$ に一致し，

$$\mathrm{tr}\mathbf{R} = p \qquad (2.9)$$

である。p 変量の分散和に対する主成分 Z_α の分散の割合は寄与率と呼ばれ，

$$\lambda_\alpha / p \qquad (2.10)$$

で示される。また，第m主成分までの累積寄与率は，

$$\sum_{\alpha=1}^{m} \lambda_\alpha / p \qquad (2.11)$$

で与えられm＝pのときは1である。

主成分分析法の目的は情報の縮約化にあるわけだから，目的に応じて累積寄与率が60〜80％を目途にmを決定すればよい。

また，主成分 Z_α と変量 x_iの相関係数 $r_{\alpha;i}$ は因子負荷量と呼ばれ，

$$r_{\alpha;i} = \sqrt{\lambda_\alpha}\ l_{\alpha i} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2.12)$$

で求められる。これにより主成分 Z_αと類似な関連の仕方をする変量に注目し考察を加えることになる。このとき，採り上げた主成分が2個，3個の場合には，各主成分を軸とする直交座標系の上に主成分のスコアーの組を座標とする点をプロットするなどして，縮約された情報の理解を助けるための処理の仕方が定着している。

(理論的なことや実証例の詳細については河口[7]，(8)，(9)，(10)，(11) 等にある)。

3．入試成績への適用

入学試験，入社試験等の選抜に際して，単に試験成績の単純合計点を合否の判定資料とするのではなく，適性なども考慮できる総合特性値が得られれば，より合理的な判定資料となり得る。この点，主成分分析法によって算出される特性値はこれらの目的にかなったものと思われる。高専の入学者を対象として入試成績，入学後学習成績について主成分分析した実証例が伊藤[6]によって示されているが，本稿では，本高専におけるある年度の全受験者の入試成績への適用を試み，得られた主成分の解釈や処理の方法を通じて，全受験者の成績における入学者の成績等も考えたい。

3.1　対象資料について

調査対象はある年度の全受験者476名についての入試成績を用いる。本高専は4学科各1学級で構成されている。入試5科目の成績は各科目100点満点であるが，素得点において平均，分散ともに異なるので標準化したものを各科目の成績得点とする。中学校の調査書による内申成績は5段階評定による3年次の9科目合計点を2倍したもの

を用い，これの標準化したものを内申点とする。実際の入試判定資料には内申成績を何倍かして入試成績と独立させて取扱う場合と，入試成績に加えて取扱う場合などがあると思われるが，ここで使われる手法では1次結合として標準化して使用するので何倍しても相関係数が変らない。したがって分析結果も不変である。

ここで使用する記号，変数は次の通りである。

学科名；　C_1，C_2，C_3，C_4

変　数；　X_1，X_2，X_3，X_4，X_5，X_6

各学科の人数は順に 120，127，102，127　計476で，変数には順に国語，社会，数学，理科，英語，内申が対応している。

3.2　分析結果について

分析処理は本校の電子計算機HITAC 8250を使用した。表1は全受験者の相関行列である。表2，表3はそれぞれC_1,C_2,C_3,C_4の学科別の相関行列で表2は対角線の右上がC_1，左下がC_2であり，表3も同様にC_3が右上，C_4が左下に書かれている。

表1では国語と数学の相関係数が他に比べて小さく，理科と数学，英語と内申，理科と社会の相関係数が表中では大きい。表2，3ではC_1については国語と数学が最も小さく，国語と内申が最も大きい。C_2では理科と内申，社会と内申が大きく，全般に数値も大きい。C_3では国語と理科，数学と理科が大きい方で極端に小さいものはない。C_4は国語と数学が小さく，他学科からみると全般に数値が小さいのが見られる。

分析の結果を示そう。表4，表5は**2**で述べたことより表1をもとに計算されたもので，全受験者の成績についての主成分である。**2**の説明におけるp，mは6であるが，ここではm＝4まで示した。表4は固有値とその固有ベクトル，表5は各主成分での因子負荷量である。表5から第1主成分の固有値は最大固有値で3.015その寄与率は52.25％である。第2主成分までの累積寄与率は63.75％，第3主成分までのそれは73.76％である。第1，第2主成分を式で示すと（2.2）から次のようになる。

$$Z_1=0.389X_1+0.409X_2+0.364X_3+0.433X_4+0.425X_5+0.425X_6 \qquad (3.1)$$

$$Z_2=-0.514X_1-0.255X_2+0.762X_3+0.155X_4+0.127X_5-0.222X_6 \qquad (3.2)$$

6変量の持つ情報を（3.1）で示される第1主成分1つに縮約したとき，この式が全情報の約半分を持ち込んでいることになる。第2主成分まで採り

表－1　相関行列（全受験者）

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.000	0.393	0.236	0.417	0.400	0.437
x_2	0.393	1.000	0.294	0.450	0.429	0.442
x_3	0.236	0.294	1.000	0.454	0.431	0.333
x_4	0.417	0.450	0.454	1.000	0.408	0.443
x_5	0.400	0.429	0.431	0.408	1.000	0.453
x_6	0.437	0.442	0.333	0.443	0.453	1.000

表－2　相関行列　$(C_2 \diagdown C_1)$

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.000	0.332	0.096	0.339	0.388	0.497
x_2	0.443	1.000	0.360	0.429	0.369	0.370
x_3	0.305	0.309	1.000	0.424	0.388	0.220
x_4	0.520	0.541	0.532	1.000	0.414	0.478
x_5	0.395	0.515	0.521	0.441	1.000	0.468
x_6	0.493	0.558	0.417	0.561	0.536	1.000

表－3　相関行列　$(C_4 \diagdown C_3)$

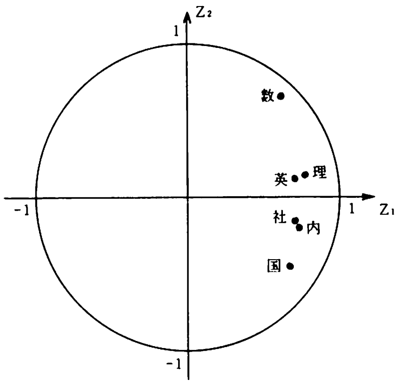
	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.000	0.393	0.340	0.471	0.471	0.400
x_2	0.393	1.000	0.331	0.386	0.448	0.444
x_3	0.161	0.174	1.000	0.466	0.405	0.360
x_4	0.295	0.405	0.325	1.000	0.463	0.377
x_5	0.353	0.384	0.400	0.323	1.000	0.445
x_6	0.353	0.363	0.269	0.286	0.344	1.000

表－4　固有値・固有ベクトル

主成分		Z_1	Z_2	Z_3	Z_4
固有値		3.015	0.810	0.601	0.597
寄与率		50.25	13.49	10.02	9.95
固有ベクトル	ℓ_1	0.389	−0.514	−0.466	−0.410
	ℓ_2	0.409	−0.255	0.723	0.280
	ℓ_3	0.364	0.762	−0.110	−0.112
	ℓ_4	0.433	0.155	0.311	−0.581
	ℓ_5	0.425	0.127	−0.367	0.608
	ℓ_6	0.425	−0.222	−0.126	0.184

表－5　因子負荷量

主成分		Z_1	Z_2	Z_3	Z_4
因子負荷量	x_1	0.675	−0.462	−0.361	−0.317
	x_2	0.711	−0.230	0.561	0.217
	x_3	0.632	0.686	−0.086	−0.087
	x_4	0.752	0.140	0.241	−0.449
	x_5	0.738	0.114	−0.285	0.470
	x_6	0.738	−0.200	−0.097	0.143



図－1　Z_1とZ_2の因子負荷量

表－6　学科別固有値・寄与率

科＼主成分	固有値	Z_1	Z_2	Z_3	Z_4
C_1	固有値	2.879	0.995	0.664	0.577
	寄与率	47.97	16.58	10.73	9.62
C_2	固有値	3.375	0.773	0.618	0.504
	寄与率	56.25	12.89	10.30	8.41
C_3	固有値	3.071	0.734	0.643	0.558
	寄与率	51.18	12.23	10.72	9.03
C_4	固有値	2.621	0.926	0.735	0.649
	寄与率	43.68	15.44	12.24	10.82

表－7　学科別固有ベクトル

科		C_1		C_2		C_3		C_4	
主成分		Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2
固有ベクトル	ℓ_1	0.375	－0.588	0.379	－0.474	0.410	－0.045	0.400	－0.456
	ℓ_2	0.405	0.154	0.410	－0.344	0.397	－0.496	0.433	－0.375
	ℓ_3	0.339	0.693	0.367	0.740	0.381	0.638	0.346	0.759
	ℓ_4	0.445	0.162	0.438	－0.006	0.423	0.411	0.411	0.096
	ℓ_5	0.435	0.016	0.413	0.291	0.434	－0.083	0.444	0.219
	ℓ_6	0.440	－0.352	0.436	－0.157	0.401	－0.412	0.408	－0.134

表－8　学科別因子負荷量

科		C_1		C_2		C_3		C_4	
主成分		Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2
因子負荷量	x_1	0.637	－0.586	0.696	－0.417	0.719	－0.039	0.647	－0.438
	x_2	0.687	0.154	0.754	－0.302	0.696	－0.425	0.702	－0.361
	x_3	0.574	0.691	0.675	0.651	0.667	0.547	0.560	0.731
	x_4	0.754	0.161	0.805	－0.005	0.742	0.352	0.666	0.093
	x_5	0.737	0.015	0.758	0.256	0.761	－0.071	0.718	0.211
	x_6	0.747	－0.351	0.802	－0.138	0.703	－0.353	0.661	－0.129

上げれば，63.75％の情報を吸収していることになり，この二つの新変量でデータのもつ特徴などの概略が捕えられる。各個人の新変量（主成分）による評点は，各個人のもつ数値を（3.1），（3.2）に代入して得られる。各個人のZ_1，Z_2の評点をデータとしたとき**2**での説明から

$$V\{Z_1\}=3.015,\qquad V\{Z_2\}=0.810,$$

$$\mathrm{cov}\{Z_1,Z_2\}=0$$

である。

また表5での因子負荷量については，主成分と変量の相関係数であったことから

$$r_{1,1}=\sqrt{3.015}\times0.389=0.675$$

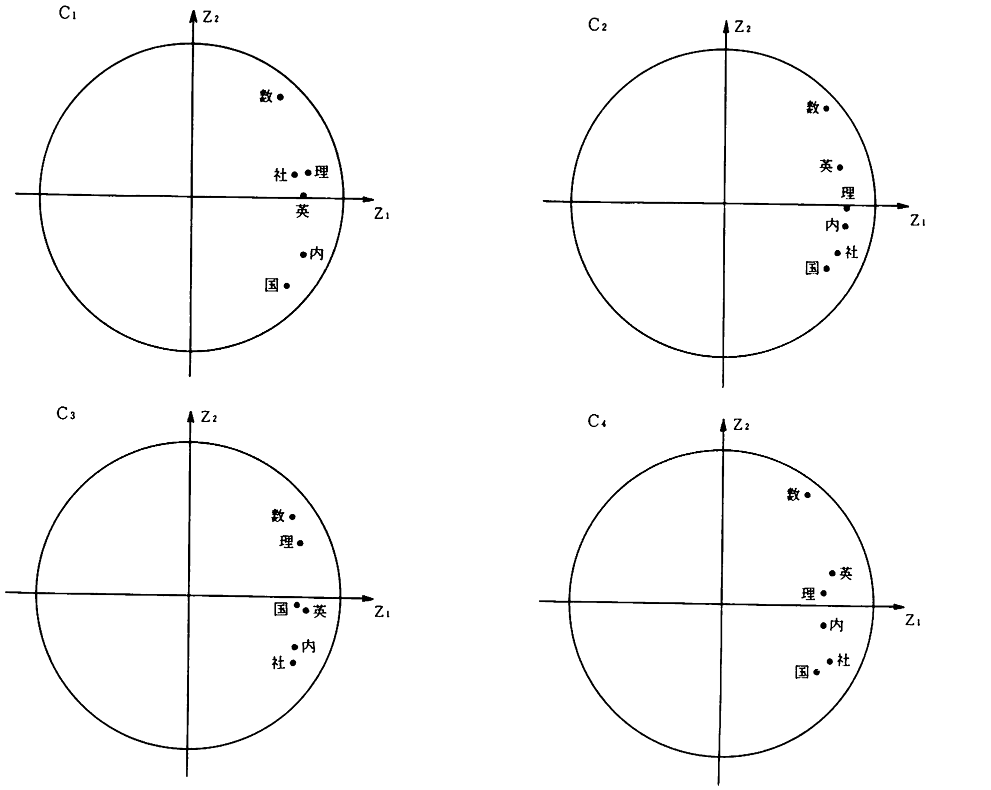
となっている。

これらを考えあわせて（3.1）（3.2）での意味を今少し検討してみると，Z_1については各係数がすべて正で，数値がほぼ一定していて，Zとの相関係数もほぼ0.7で大きく異なるものはない。このことより各変量は同じような寄与をしていると考えられ，変量全体として大きな得点をしたものはZ_1の評点が大きく，成績の良さを表わす総合特性値と考えてよい。Z_2については，Z_1とは無相関であることからZ_1とは異なる意味を含むと考えられる。Z_1が一定のサンプルにおいて，負の係数をもつ変量の値が大きければZ_2の評点は小さく，正の係数をもつ変量の値が大きければZ_2の評点は大きくなる。各係数の絶対値の大きい変量の値すなわち因子負荷量の絶対値の大きい変量の値がZ_2の値の大小に大きく作用することになる。

（3.2）では国語の得点が大きいときZ_2の評点は小さく，数学の得点が大きいときはZ_2の評点は大きくなる。このことよりZ_2のもつ意味は国語に関する能力と数学に関する能力に深く関係したものを分離する特性値と考えられる。このことは因子負荷量を図1のようにZ_1を横軸，Z_2を縦軸とする因子平面上に目盛るとこの様子が視覚的に理解される。

次に表2，表3でみられたように学科によって数値に多少の違いが表われていて，6次元の成績空間の分布にいくらかの差異が感じられるが，学科毎に分析した場合どうであろうか。結果は表6，表7，表8の通りである。固有ベクトル，因子負荷量は第2主成分まで示した。

表6から4学科中C_2の第1主成分の固有値が最大でC_4の第1主成分の固有値が最小である。これは表2，表3でみられたようにC_2は全般に数値が大きく，C_4は全般に数値が小さかったことに対応していると考えられる。一般に各変量間の



図－2　学科別 Z_1 と Z_2 の因子負荷量

相関係数が小さければ多次元空間内のデータは拡散傾向を示し主成分の固有値も小さく各固有値の大きさの差が小さくなる。また逆に各変量間の相関係数が大きければデータは一方向へ凝集傾向にあり，第1主成分の寄与率が大きくなると云われていることからもうなずける。第2主成分までの累積寄与率はほぼ60%以上で各科毎にデータの特徴の概観は出来そうである。表7，表8からはいくつかの共通点がみられる。第1主成分においては各科とも数学の負荷量が一番小さいがウエィトとして大差のない数値である。第2主成分についてはこれも各科共数学の負荷量が最大で C_1,C_2,C_4 で数学と国語がそれぞれ逆方向の影響をもっている。C_3 ではこの対比が数学と社会にみられる。総じて第2主成分は数学と国語，社会の分離を示していると考えられるので理科系文科系の分離に関連する特性値とみなして差支えないようである。この様子を図1と同様に各科毎に図2に示しておく。

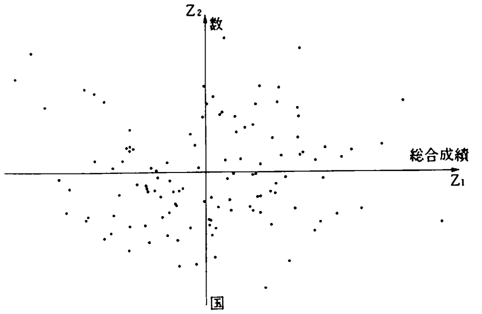
ここで，全受験者と学科別分析の結果を振り返って第1主成分に注目すると，係数ベクトルである固有ベクトルの各要素の大きさはほぼ一様とみる事ができた。このことは変量が標準化されている場合には単純合計点をもって総合特性値として成績順位を定めても差支えないと思われる。また，この年度における本校の判定規準によって順位づけられた資料と主成分分析法によって得られた第1主成分の評点の大きさに従って順位づけられた資料について，少し粗いがスピアマンの順位相関係数を C_1 に対して求めてみた。順位差 d_i は表9に示す。スピアマンの順位相関係数 r_s は次式で与えられ，結果は0.99であった。

表－9　順位差の度数表

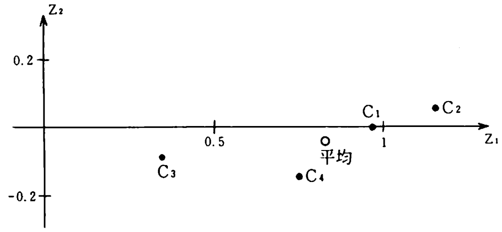
d_i	0	1	2	3	4	5	6	7	8	9	10	11	12
度数	29	25	15	22	8	7	2	4	3	1	1	2	1

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

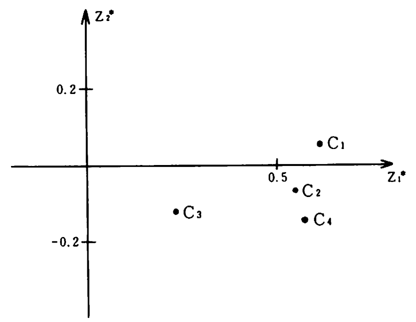
実際，40位で前後したものは1，50位では3，60位では1と変動は少なく，順位差の大きいものは分散の大きい科目で極端な得点をしているのが認められ，Z_1 で大きく変動するのは当然のことである。これらの事から実際的には判定資料が得点



図－3　Z_1, Z_2 平面上での各人の位置



図－4　学科の平均評点



図－5　入学者の学科毎標準平均評点

表－10　入学者の平均評点

科 / 主成分	C_1	C_2	C_3	C_4	全体
Z_1	0.965	1.153	0.357	0.759	0.823
Z_2	− 0.009	0.041	− 0.086	− 0.157	− 0.052

表－11　学科毎の入学者の標準平均評点

科 / 主成分	C_1	C_2	C_3	C_4
Z_1^*	0.619	0.553	0.229	0.570
Z_2^*	0.054	− 0.068	− 0.132	− 0.144

をそのまま主材とする総合特性値でもあまり気にかけることはないのかもしれない。

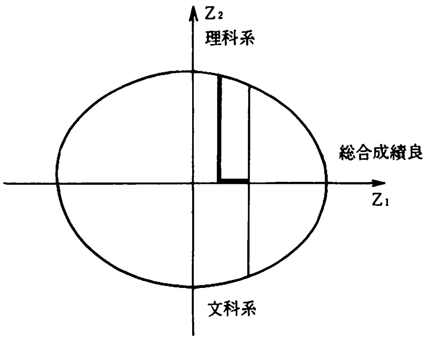
さて，Z_1, Z_2 の評点について少しふれる。Z_1, Z_2 を両軸とする座標平面上にそれらの値の組をプロットしてみると，その位置により，総合成績の良さ，文科系，理科系の成績とかが理解しやすい。C_1全員についてプロットしたものが図３である。Z_1, Z_2 の無相関の様子も確認される。更に実際に本校に入学したものの評点について考えてみる。全受験者を対称として主成分分析した第１主成分，第２主成分の評点を各学科ごとに平均値を算出し，表 10 に示す。これを座標平面上にプロットしてみると図４のようになった。学科の実態や性格が浮き彫りにされて興味深い。

学科別に分析された結果について次のような処理を考えてみた。同一集団について得られた主成分の評点については図３や図４にみられるように同一方向での数値の大小によって比較が可能であるが，異なる分布の集団それぞれに同一の分析手法を適用させた結果については，得られた評点の大小は質的な差を意味しない。いま，各主成分の平均は０で分散が固有値であったから，各評点を固有値の平方根すなわち標準偏差で割ってやると標準化された評点が得られる。これは異なる集団での評点を同一の尺度で表わすことであるからこの標準化された評点は相対的評点として捕えられる。固有値を λ として $Z^*=Z/\sqrt{\lambda}$ とおき，各学科毎に算出された主成分による評点について，各学科毎の入学者の平均値を Z^*に変換したものが表 11 である。これを Z_1^* Z_2^*平面上にプロットしたものが図５である。これは図４と異なるものであり，各学科毎の入学者の平均評点の相対的位置を示している。Z_1^*について C_3 が他より小さい評点をもつことが見られるが，辞退者の数などが影響しているものと思われる。

４．選抜資料としての有効性

ここでは，主成分分析法が，入試選抜に際して有用な資料を提供出来るかどうか考察してみたい。先に，この手法からの第１主成分の評点による順位と実際の順位には粗い見方ではあるが，あまり差のないことを見た。この１例をもって決定的な結論を下すことはできないし更に慎重な検討と実証例の積み重ねが必要であろう。

さて，今回使用した資料は同年度のものであるが一応５個の分析例をみた。それらを比較して，第１主成分への各変量の寄与は，ほぼ等しく，全情報の約半分以上をも含むことがわかった。この



図－6　(Z_1, Z_2) の分布図

ことより変量が標準化されていれば，単純合計点もほぼ同じ効果をもつことも裏付けた。更に第2主成分は第1主成分と無関係な方向で文科系科目と理科系科目の成績を分離させる特性値とみれた。また，第2主成分までとり上げると全情報の約60%以上を含むこともみられた。ただこの手法では，中学校の内申成績は他の科目並の扱いであり重みをつけても不変であったから，内申成績をより重要視する場合には分離して独立的に扱うか他の方法によらねばならない。

以上のことから各科目の成績を標準化して扱うことを前提として，単純合計点を資料とする場合は全情報のほぼ50%位をもとに判定することになる。第1主成分を用いる場合も同様である。選抜に際してはデータのもつ情報が100%生かされた形で行われるのが最善であろうが，提供されたデータを全て用いたからと云って，それらのもつ情報が100%生かされた形で取扱われるわけでないことはこのことより明らかであろう。データを縮約的に取扱う方法は種々あろうが，データのもつ情報が出来るだけ多く含まれることが望ましい。その意味では，入試選抜に関するデータの取扱いとして，主成分分析法による第2主成分までを採り上げるのがより有効であると考えられる。

実際的な利用方法としては，全受験者の順位づけなら全員を対象に，学科独自の順位づけなら学科毎にそれぞれ主成分分析を行い，第2主成分までの評点を用意すればよい。

主成分 Z_1, Z_2 による座標平面上で，各個の評点を座標とする点はほぼ図6のようになるから，定員やその他の事情を考慮して，先ず第1に実線のように第1主成分によって境界を設ける。次に第2主成分の内容を検討して，適当な分離が示されないときは第1主成分のみで考え，文科系，理科系の分離がみられれば高専が工学系の学校であることを考慮して，図6の太線のような境界を設けることなどが考えられる。

5. お わ り に

以上，入試成績への主成分分析法の適用を試みるにあたり，この分析手法の概要，適用結果とその解釈，入試選抜への利用可能性等について述べて来た。適用対象の例が単年度のものでありいささか確信に欠ける面も感じられるが，学科別適用において，いくらかのデータの違いが見られる中で，ほぼ同様な分析結果が得られ，解釈が容易であった。更に多くの実証例の累積が必要なことは論を待たないが，概ね次のようにまとめられる。

(1) 第1主成分は全情報のほぼ50%以上を，第2主成分までを採り上げるとほぼ60%以上の情報を含む。

(2) 第1主成分への各変量の寄与はほぼ等しく，標準化された変量の場合，変量の単純合計点と第1主成分はほぼ同等の総合特性値と考えられる。

(3) 第2主成分は文科系，理科系の成績を分離する特性値と考えられる。

(4) 入試選抜における，成績データの縮約的処理については，第2主成分までの評点を利用することにより，より有効な判定資料が得られると思われる。

4 では主成分分析法による判定資料の取扱いについても意見を述べてみた。

最後に日頃指導を受けている北海道大学大学院情報数理工学講座の河口至商教授，同佐藤義治助手および分析処理に当り多忙な中多大な助力と有益な助言を与えられた本校電気工学科今田孝保助教授に対し感謝の意を表するものである。

参 考 文 献

〔1〕立花敏之：「相関分析による入学者選抜についての考察—調査書の成績の重み—」苫小牧高専紀要第9号 (1974)

〔2〕松尾春雄・中野　昭：「入学者選抜についての研究」大分高専紀要第10号 (1973)

〔3〕春本　繁・高尾幸男：「入学生選抜に際しての一資料について」高松高専紀要8号 (1972)

〔4〕長谷川俊勝：「入学試験合格判定基準についての考察」函館高専紀要第9号 (1974)

〔5〕立花敏之：「相関分析による入学者選抜についての考察—標準得点—」苫小牧高専紀要第12号（1977）
〔6〕伊藤俊彦：「入試の足跡」現代数学 Vol. 10 No. 7 (1977)
〔7〕河口至商：「多変量解析入門」森北出版（1973）
〔8〕奥野忠一他：「多変量解析法」日科技連（1971）
〔9〕伊藤孝一：「多変量解析の理論」培風館（1969）
〔10〕Kendall, M. G. :「A Course in Multivariate Analysis.」Griffin, (1957)（浦 昭二・竹並輝之共訳「多変量解析の基礎」サイエンス社（1972））
〔11〕Anderson, T. W. :「Introduction to Multivariate Analysis」John Wiley & Sons, (1958)

（昭和52年11月30日受理）