

# 主成分分析による入試成績の 順位づけシミュレーション

野 坂 弥 蔵\*

yazo NOZAKA

## Rank-Ordering Simulation by the Principal Component Analysis on Results of an Entrance Examination

**Abstract:** University examinees are rank ordered on scholastic ability evaluated by the sum of their scores of subjects in an entrance examination. These ranks exhibit a degree of disarray with respect to their ranks on their university scores.

In this paper, it comes to light that the ranks based on the first, second and third Principal Component calculated from the scores of examinees coincide fairly well with their rankings on the university scores. And it is also known that the first Principal Component is a synthetic value of the scholastic ability of an examinee and that the second and third Principal Component includes informations about his natural aptitude for liberal arts or science.

### 1. ま え が き

入学試験，就職試験などでは，受験者の資質，適性を代表すると思われる種々の特性値を測定し，終局的にはそれらを1つの総合特性値に要約し，その大小によって順位をつける場合が多い。大学入試においても，共通1次試験の各教科の得点，2次試験の得点および高校成績を特性値とし，その各々に，適当と思われる重みを乗じて合計した値を総合特性値として順位づけを行っていることが多い。このような手段によって選抜され入学した学生の中には，種々の理由で学習意欲を失い，留年したり退学したりする者がある。その原因の一つとして，適性不一致による学習内容への興味喪失が考えられる。文科系の素質を持つ学生が理工系の学部へ入学した場合などである。素質や適性を見分けることは一回の入試データからは困難で，高校での進路指導が望まれるが，入試成績や高校からの内申書の中にも適性についての情報がある程度含まれている筈であり，これを抽出して適正な選抜を行う方法の検討が望まれ，大学入試センターを

じめ各大学で研究が進められている。今回は，共通1次の各教科の得点，2次試験の得点および高校成績を主成分分析し，第1，第2，第3主成分の内容を調べ，主成分スコアの組み合わせによって順次づけを行い，どのような組み合わせを使えば大学での順位が予測できるかを検討した結果を報告する。

### 2. 標準得点と主成分スコアの算出法

$n$ 人の受験生中  $i$  番目の人の科目  $k$  の得点を  $x_{ik}$  とする。 $(k=1, 2, \dots, 12)$  は，それぞれ共通1次の国，社，数，理，英，および高校の国，社，数，理，英，ならびに2次試験の2教科に対応している。例えば  $x_{i6}$  は高校の国語の成績を表わす) また教科  $k$  の平均点を  $\bar{x}_k$ ，標準偏差を  $\sigma_k$  とすると  $i$  番目の人の教科  $k$  の標準得点  $y_{ik}$  は次式から算出される。

$$y_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k} \dots \dots (1)$$

$i$  番目の人の第1主成分のスコアを  $z_{i1}$ ，第2主成分のスコアを  $z_{i2}$  とし，以下同様に第12主成分のスコアを

$z_{i12}$  とすると、これらは次式から算出される。

$$\left. \begin{aligned} z_{i1} &= a_1 y_{i1} + a_2 y_{i2} + \dots + a_{12} y_{i12} \\ z_{i2} &= b_1 y_{i1} + b_2 y_{i2} + \dots + b_{12} y_{i12} \\ &\dots\dots\dots \\ z_{i12} &= l_1 y_{i1} + l_2 y_{i2} + \dots + l_{12} y_{i12} \end{aligned} \right\} \dots\dots(2)$$

係数  $a_k, b_k, \dots, l_k$  ( $k=1, 2, \dots, 12$ ) は、各教科の得点間の相関行列  $R$  の固有ベクトルに等しいことを利用して算出する。行列  $R$  の  $k$  列  $j$  行要素  $r_{kj}$  は

$$r_{kj} = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)}{\sigma_k \sigma_j} \dots\dots(3)$$

で求まる。

また行列  $R$  の固有値は各主成分の分散に等しい。尚、各主成分間には相関がない。

3. 計算結果とその解釈

(a) 文科系受験者534名についての計算結果

第1表に第1, 第2, 第3主成分(以下  $z_1, z_2, z_3$  と略記する)の固有値と寄与率を、第2表に各教科別の固有ベクトルと因子負荷量を示す。

また第1図には  $z_2$  と  $z_3$  の因子負荷量の関係を図示する。第1表の寄与率から分かるように  $z_1$  には38%の情報が含まれるが  $z_2, z_3$  はそれぞれ16.7%, 10.1%であ

第1表 固有値と寄与率(文系534名)

	$z_1$	$z_2$	$z_3$
固有値	4.56	2.00	1.22
寄与率	0.380	0.167	0.101

第2表 固有ベクトルと因子負荷量(文系534名)

	固有ベクトル(主成分の係数)			因子負荷量		
	$z_1$	$z_2$	$z_3$	$z_1$	$z_2$	$z_3$
国語	0.177	0.355	- 0.303	0.378	0.503	- 0.334
社会	0.233	0.316	- 0.294	0.498	0.447	- 0.324
数学	0.218	0.205	0.316	0.466	0.290	0.349
理科	0.250	0.318	- 0.170	0.534	0.450	- 0.187
英語	0.239	0.244	0.506	0.509	0.345	0.558
高国	0.355	- 0.246	- 0.205	0.757	- 0.348	- 0.227
高社	0.340	- 0.237	- 0.348	0.726	- 0.335	- 0.384
高数	0.324	- 0.342	0.178	0.692	- 0.484	0.196
高理	0.363	- 0.250	- 0.135	0.775	- 0.354	- 0.149
高英	0.328	- 0.333	0.249	0.699	- 0.472	0.275
2次国	0.255	0.320	- 0.156	0.545	0.453	- 0.172
2次英	0.310	0.251	0.372	0.663	0.355	0.411

り、 $z_3$ までの累積寄与率は0.648であるから、 $z_1$ から  $z_3$ までの中に約65%の情報が含まれていると考えられるので  $z_4$ 以下は考えないことにする。

第2表を見ると  $z_1$  の固有ベクトルと因子負荷量はいずれも正で、且つ高校の各教科に対する値が、入試教科に対する値より大きい。このことから  $z_1$  は受験生の学力全般についての特性値で、特に高校成績の重みが大きいことが分かる。また第1図を見て分かることは次の4つである。

①  $z_2$ に対する入試成績の因子負荷量( $z_2$ と各教科得点との相関係数)は正であるが、高校成績の因子負荷量は負であるから、入試成績と高校成績とは全く異質の量である。

② 高校成績は良くないが入試成績は良かった人の  $z_2$ スコアは正で大きく、逆のときは負となる。

③ 入試成績でも高校成績でも、 $z_3$ に対する英語と数学の因子負荷量は正であるが、国, 社, 理のそれは負であるから、英, 数の成績と国, 社, 理のそれとは異質の量である。

④ 入試および高校で、英, 数の成績は良いが国, 社, 理のそれは芳しくない人の  $z_3$ は正で大きく、逆の人の  $z_3$ は負で大きい。従って  $z_3$ は理工系の素質のある人と文科系の素質を持つ人とを区別する量であると考えられる。

(b) 理科系受験者49名についての計算結果

第3表に  $z_1$ から  $z_4$ までの固有値と寄与率を示す。これから分かるように  $z_1$  には39.4%,  $z_2$ には15.4%,  $z_3$ には12.6%の情報が含まれ、 $z_3$ までの累積寄与率は0.674であるから  $z_4$ 以下は考えないことにする。(この

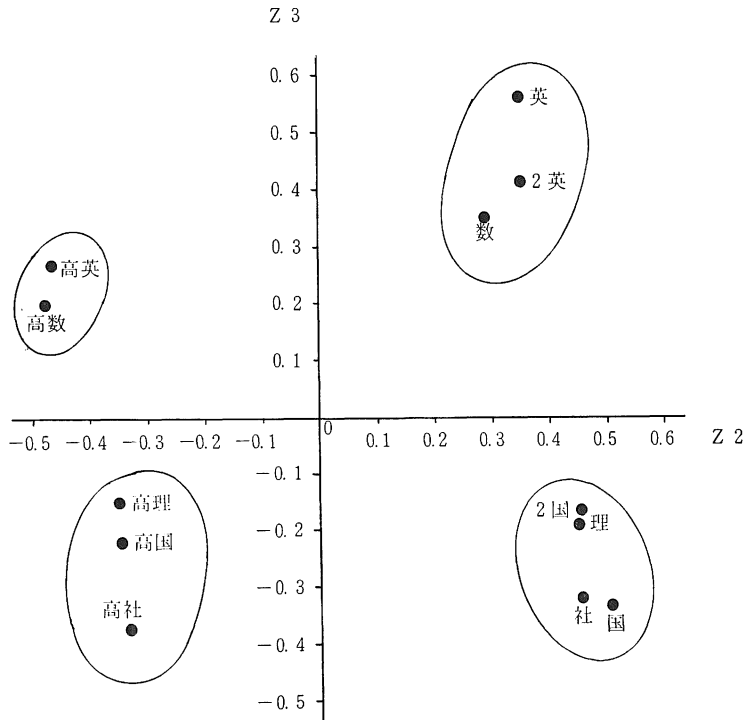


図1 z2とz3の因子負荷量(文系534名)

ための情報損失は約33%である)

z1に対する固有ベクトルと因子負荷量は(a)の場合と同様な傾向を示すので、z1は、高校成績に重みを置いた学力の総合評価値であると考えられる。

z2とz3の因子負荷量の関係を第2図に示す。これを見て分かることは次の4つである。

① z2について(a)の場合と異なる点は、国語と社会の因子負荷量が高校成績のそれと同様に負になることである。高校成績の因子負荷量はz3についても負になるので、入試成績と高校成績とが全く異質の量であると云える点は(a)の場合と同じである。

② 数学、理科、英語の入試成績が良く、国語、社会の高校成績が芳しくない理工系素質の人のz2は正で大きい。

③ z3についての因子負荷量は、入試関係は正で、高校関係は負である。この点は(a)の場合のz2と同じ傾向である。

④ 入試で、国語と社会が良くできた人のz3は正で大きい。従ってz3のスコアが大きい人は文系の素質を持つと考えられる。

第3表 固有値と寄与率(理系49名)

	z 1	z 2	z 3	z 4
固有値	4.73	1.84	1.51	1.05
寄与率	0.394	0.154	0.126	0.087

(c) 教育系受験者428名についての計算結果

第4表にz1からz3までの固有値と寄与率を示す。これから分かるようにz1には41.9%、z2には12.3%、z3には0.8%の情報が含まれ、z3までの累積寄与率は0.625である。

z1に対する固有ベクトルと因子負荷量は(a)、(b)の場合と同様な傾向を示すので、z1の意味も(a)、(b)と同様である。ただ、教育系受験者の中には文系、理工系、芸術系、体育系の人々が混在していることと、2次試験が学科ではなく小論文などであることが原因と思われるが、z1に約42%の情報が集中し、代わりにz2、z3の情報がやや減っている点が(a)、(b)の場合と異なっている。

z2とz3の因子負荷量の関係を第3図に示す。

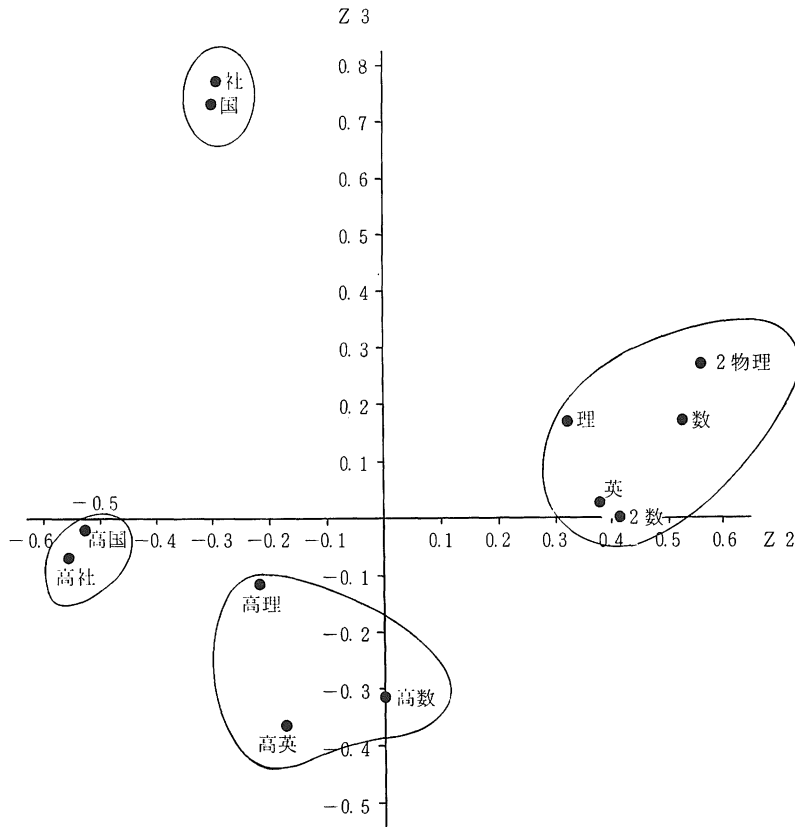


図2  $z_2$  と  $z_3$  の因子負荷量 (理系49名)

これを見て分かることは次の4つである。

- ①  $z_2$ については(a)の場合と同様に 高校成績の因子負荷量は負になっている。従って入試成績と高校成績とは全く異質のものであると云える。
- ② どちらかと云えば国語, 社会の得意な文系の人々の  $z_2$ スコアは正で大きい傾向はあるが, 余り明瞭ではない。
- ③  $z_3$ の因子負荷量から見ると, 共通1次の成績と高校成績と, 2次試験成績という3つの異質な群に分かれる。
- ④ 2次試験成績が良かった人の  $z_3$ スコアは負で値は大きい。

第4表 固有値と寄与率 (教育系428名)

	$z_1$	$z_2$	$z_3$
固有値	5.02	1.48	1.00
寄与率	0.419	0.123	0.083

#### (d) まとめ

上記(a), (b), (c)をまとめると次のようになる。

- ① 第1主成分  $z_1$ は入試成績 (1次と2次), および高校成績を総合した学力の特性値であるが, 高校成績に重点が置かれている。また  $z_1$ には入試および高校の内申から得られる情報の33~42%が含まれている。
- ② 学科目による入学試験成績と, 内申書より算出される高校成績および学科目以外による入学試験成績 (論文など) の3者は異質のものである。
- ③ 第2および第3主成分  $z_2$ と  $z_3$ は受験者の素質が文科的か理工科的かを見分けるのに役立つ量と考えられるが, その特徴は一定せず集団によって異なる。

## 4. 主成分による順位づけシミュレーション

主成分は素材となるデータに含まれる特性を鮮明にするため, それを先ず標準化することによって測定単位な

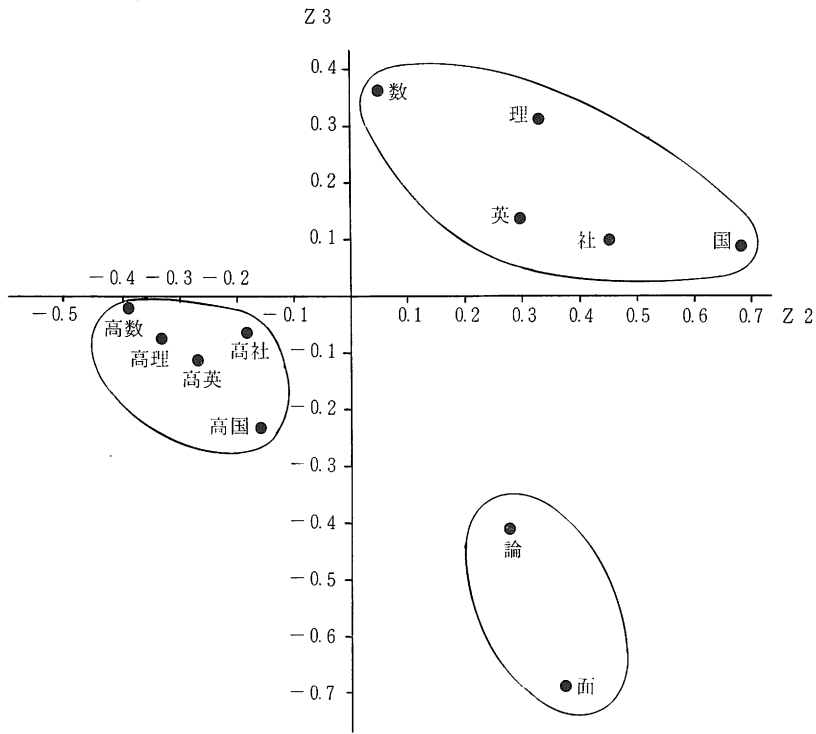


図3 z2 と z3 の因子負荷量 (教育系 428 名)

どの影響を除外したデータに変換した後、各データに乗ずる係数を、それらの係数の2乗の和が1になり、且つ、各主成分の間には相関がなく、分散が最大になると云う条件から算出したものであり、前記の例で分かるように第1主成分z1に可成りの情報が集中しているから、z1の大小によって順位をつけても良いが、受験者の素質が文系であるか理工系であることを示す情報を含んでいるz2やz3を加味した方が一層適切な順位が得られるであろうと考え、i番目の受験生が得たz1, z2, z3のスコア  $z_{i1}, z_{i2}, z_{i3}$  に、係数  $c_1, c_2, c_3$  を乗じて加え合わせて得られる値  $w_i$  の大小で順位をつけることにする。

$$w_i = c_1 z_{i1} + c_2 z_{i2} + c_3 z_{i3} \dots \dots \dots (4)$$

(a) 各種の順位間の相関係数

共通1次試験および2次試験の各教科得点の単純な合計による順位(入試順位と略称)、主成分の重みつき合計である  $w_i$  による順位(主成分順位と略称)および高校成績  $x_{ik} (k=6, \dots, 10)$  の合計による順位(高校順位と略称)との間のスピアマンの相関係数  $\rho$  を次式より計算した。

$$\rho = \frac{12 \sum_{i=1}^n R_i S_i}{n(n^2-1)} - \frac{3(n+1)}{n-1} \dots \dots \dots (5)$$

ここで  $n$  は受験生の総数であり  $R_i$  と  $S_i$  は入試順位、主成分順位または高校順位の中のいずれか2つである。

第5表には文系、理系、教育系について計算した  $\rho$  を示す。但し主成分順位を求める時に  $w_i$  中の  $c_1=1, c_2=c_3=0$  としてある。

第5表を見ると、入試順位と高校順位の相関は大きくないが、主成分順位は入試順位または高校順位との相関が大きく、特に高校順位との相関が大きい。

第5表 順位相関係数

	文系 (534名)	理系 (49名)	教育系 (428名)
主成分順位と高校順位	0.851	0.904	0.914
主成分順位と入試順位	0.701	0.729	0.758
入試順位と高校順位	0.314	0.439	0.509

第6表 大学順位と、入学試験時点での順位の関係

学 生	A	B	C	D	E	F	G	H	I	J
大学 順位	1	2	3	4	5	6	7	8	9	10
入 試 順位	9	2	7	10	8	1	6	4	3	5
高 校 順位	1	4	2	6	3	5	7	8	8	10
主成分順位1	1	2	3	6	4	5	7	8	9	10
主成分順位2	1	3	2	6	4	5	7	9	8	10

b) 入試順位, 主成分順位, 高校順位と大学順位との  
相関

教育系428名の受験者中, 大学に入学した10名につき大学での学習成績や人物評価を総合して定めた順位(大学順位と略称)と入学試験時の入試順位, 主成分順位, 高校順位との関係を示したのが第6表である。この表で主成分順位1, 主成分順位2としてあるのは,  $w_i$  の係数を

$c_1=1, c_2=c_3=0$  とした場合と,  $c_1=1, c_2=-0.3,$   
 $c_3=-0.2$  とした場合に対応している。

この表から, 大学順位と比較的よく一致するのは主成分順位であって, 入試順位は余り合わないことが分かる。尚, 10名の出身高校は異なるのが多いが, 格差に対する配慮はしていない。

## 5. む す び

入試および高校成績を主成分分析して第1～第3主成分の持つ情報を明らかにし, それを考慮に入れた総合特性値の大小から主成分順位を求めたところ, 主成分順位は高校順位との相関が高いことが分かった。また教育系428名の受験者中, 大学に入学した10名につき大学順位と入学試験時点での順位とを比べて見ると, 主成分順位が良く一致するという結果を得た。しかし, データが少数である上に大学順位の正しさにも問題があるから, 主成分順位が入試順位より妥当であると速断はできないが, 今後更に多くのデータを用いて検討して行く価値はあると思われる。

入試データの統計処理について種々御指導いただきました電気通信大学熊本 芳朗教授, 高知大学野町幸男教授, 鹿児島大学永田昭三教授, 山梨医科大学平野光昭教授を初めとする入試, 入研関係の多くの教官, 事務官の方々に厚く感謝いたします。

## 参 考 文 献

- 河口至商: 多変量解析入門 森北出版  
 奥野忠一他: 多変量解析法 日科技連  
 竹内 啓: 入学試験成績の分析における相関係数の意味について