

主成分分析法（Principal Component Analysis,略してPCA）、多変量解析法において最も基本的な手法である。互いに相関のある多種類の特性値を持つ情報を、互いに無相関な少数個の総合特性値に要約する。また、ある問題に対して、いくつかの要因が考えられるとき、それらの要因を一つ一つ独立に扱うのではなく、総合的に取り扱おうとするのがPCAである。このような説明では、利用する立場の我々にとっては、ますます手の届かない存在の手法と考えられるけれども、そうではないことを、いくつかの例題で知っていただきたいと願い拙文を作成した。

主成分分析法の解析は複雑であるが、パソコンが高度化し、無償の解析ソフト「R」を利用して、必要なすべてのデータが得られ、我々はこれら分析結果を、利用すればよいことになる。

パソコンは、身近で利用できるWindows10に「RとRcmdr」をインストールする必要がある。データは、Excelで管理して、分析が必要なときに呼び出し利用すればよい。

PCAに用いられるデータは、 n 個のサンプルについて測られた P 個の特性値である。これら P 個の特性値の単位は、g、cm、%などが混在していてもよい。 n 個のサンプルについて p 個の特性ごとに検査または計測して入手した数値（評価点）をまとめる。この表はサンプル数 n と P 個のデータで構成された原表（2元表）である。取り上げた特性が P 個であるから、第 P 主成分までの分散を得ることができる。分散は、第1主成分から第 P 主成分まで、徐々に減少している。次に個々の分散の寄与率を求める。寄与率は、 P 個の分散の合計値で各主成分の分散を除した値である。分散の寄与率が1以上の主成分まで、または、第1主成分から第 P 主成分に向かって累積寄与率が80～90%になるまでの主成分で、結果を判断する。この判断に用いなかったものは、情報の損失である。

次に、得られた「固有ベクトルの因子負荷量（Component loading）」と2元表（原表）のデータ値を用いて主成分スコア値をサンプルごとに求める。第 a 主成分の因子負荷量とデータから算出した主成分スコア値を $PC(a)$ としている。 $PC(a)$ には、第 a 主成分の因子負荷量で求めた n サンプルごとの主成分得点が含まれていることになる。 $PC(a)$ は、 p 個存在するが互いに独立（無相関）している。これらの nP 個の主成分スコア値はパソコン内で算出して保管されているから、これを利用すればよい。

第1主成分の主成分スコア値は総合的なもので、第2主成分以下からのスコア値は、特性の組み合わせからのスコア値である。2つのスコア値（ $PC1$ と $PC2$ ）を組み合わせた散布図を作成し、さらに $PC1=0, PC2=0$ で、散布図を4分割した（第1象限から第4象限まで）。この散布図から、サンプルのそれぞれの立ち位置を知ることができる。この散布図から、共通的にいえることは、第1主成分のスコア値は総合的なもので、いろいろな分野で利用されている（プロ野球選手の次年度の俸給額決定など）。この拙文で取り上げた少ない実施例では、散布図の第3象限にプロットされたサンプルは、「改善策が必要である。」ということであった。

主成分分析の手順を説明したが、次の4実施例で主成分分析のやっかいな分析はすべてパソコン内で処理されていて、効果的に結果を活用することだけが残されている。このことから、主成分分析は身近な存在になっていることを実感していただきたいと思う。

実施例1：中学3年生の生徒23名について行った体力テスト6項目の測定値を主成分分析した結

果である。筋力（握力・背筋力）と運動力に分けて考えることができる。さらに、運動力を持久力と瞬発力に分けて考えられる。

実施例 2：東京都内 20 ホテルの例である。サービスを人的サービスと物的サービスに分けて検討することができた。

実施例 3：中学 2 年生 23 人の期末テストについて主成分分析を行った結果、生徒一人一人の立ち位置を図示できた。

実施例 4：全国主要 13 大学の学生が、自身の大学の授業に満足しているどうか、調査した結果について主成分分析した結果である。難関大学に苦学して入学できたのに、自大学の授業に満足していない。先生方も大学側も努力して提供した授業を、必ずしも満足していないのである。

多くの企業や研究所などだけでなく、個人にとっても、新製品だけでなく、調査分析報告書等の成果物の評価を知り、さらにそれらの質を高めたいと常に努力している。そのようなとき、実施例 4 は、主成分分析が役立つことを教えてくれている。

新製品や調査報告書など成果物の利用対象者を層別して n 人を選びモニターとし、成果物の P 個の特性を設定して、モニターごとに P 個の評価値を得たならば、 n 行 P 列の 2 元表が得られる。この表を主成分分析すれば、散布図が得られ、その散布図からさらに優れた成果物の改良に必要な情報が得られる。それぞれの層の人たちは、どのような成果物（情報）を期待しているのか、などを知ることができる。

このようなことを考えると、主成分分析は身近な存在となる。主成分分析を繰り返し使用することで、いろいろなところで活用できるようになると思われる。

主成分分析を行うために、統計計算ソフト「R」をインストールする必要がある。ほとんどの理系を持つ大学では、「R」をインストールできるソフトを公開している。理系の卒業生なら、この「R」を利用した経験を持っていると思う。「R」をもちいて「Rcmdr」も簡単にインストールできる。

Windows10 のパソコンがあれば、無償で「R、Rcmdr と Excel」を利用して、主成分分析を利用できることになる。「R,Rcmdr,Excel」の活用法は、「近畿 pe 技術相談室」で解説しているから参照していただきたい。

完 2021年1月26日

公益社団法人日本技術士会近畿本部登録近畿PE技術相談室

<http://kinkipesodan.xsrv.jp/>

実施例 1

中学 3 年生の生徒 25 名について、体力テストを行った。テストの項目（特性値）は。反復とび、垂直とび、背筋、握力、50 メータ走、幅跳の 6 項目である。各項目は、データが大きいほど能力が高いことになるが、50 メータ走は、所要時間が小さいほど能力が高くなる。そのためデータは、逆に速度（50m / 所要時間）にして、他の 5 項目の評価と合わせている。その原データは、第 1 表のようになった。

各データの単位は、反復跳び（回/分）、垂直跳び(cm)、背筋(kg)、握力(kg)、x50m (m/s)。幅跳び(cm)で、6 項目ごとに異なっている。

id	反復	垂直	背筋	握力	X50m	幅跳び
1	49	65	150	49	7.1	480
2	50	57	167	45	6.8	490
3	42	54	174	54	6.1	410
4	47	75	205	54	7.1	580
5	45	65	155	60	6.8	520
6	52	67	140	42	7.1	520
7	50	66	150	42	6.9	460
8	50	65	156	52	7	535
9	47	73	180	50	6.8	480
10	50	60	135	36	6.7	420
11	50	61	163	48	6.6	430
12	47	65	105	42	6.2	430
13	46	60	143	46	6.6	445
14	57	64	173	52	6.6	480
15	51	59	140	48	6.4	440
16	50	71	162	49	7.4	520
17	49	60	140	43	6.9	490
18	48	58	160	50	6.9	550
19	48	60	125	43	6.2	440
20	48	60	125	43	6.2	440
21	53	65	165	41	7.2	530
22	47	57	156	40	6.2	435
23	44	68	137	45	6.4	505
24	53	58	172	49	6.9	510
25	48	67	145	43	7.1	520

平均値、標準偏差などの計算、主成分分析などは、統計計算ソフト「R」を用いて行った。

第2表 各測定値の平均値、標準偏差、各測定値間の相関係数表。

	mean	sd		X50m	握力	垂直	背筋	反復	幅跳
X50 m	6.72	0.36	X50 m	1.00	0.13	0.54	0.41	0.39	0.75
握力	46.64	5.40	握力	0.13	1.00	0.18	0.57	-0.18	0.37
垂直	63.200	5.25	垂直	0.54	0.18	1.00	0.24	0.03	0.55
背筋	152.920	20.78	背筋	0.41	0.57	0.24	1.00	0.11	0.47
反復	48.840	3.11	反復	0.39	-0.18	0.03	0.11	1.00	0.17
幅跳び	482.400	45.80	幅跳び	0.75	0.37	0.55	0.47	0.17	1.00

第1表、第2表の表を用いて、25名の生徒の体力能力の検討をしようとしたが、測定値ごとに単位が異なり、その上測定間には相関がある。その上、データの広がり(最大値-最小値)にも差がある。このような場合、主成分分析が適している。

主成分分析の出力結果

Component variances:主成分の分散

Comp.1	Comp.2	Comp.3	omp.4	Comp.5	Comp.6
2.7927399	1.3545324	0.8970691	0.4126367	0.3608662	0.1821556

6項目のデータを基準化した後、主成分分析したから分散の合計値は6

Importance of components:(分散の寄与率と累積寄与率)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Proportion of Variance	0.46545	0.22575	0.14951	0.06877	0.0601	0.03035
Cumulative Proportion	0.46545	0.69121	0.84072	0.90949	0.9696	1.00000

Comp.1 ; 第1主成分、Comp.2 : 第2主成分、・・・Comp,6 : 第6主成分

主成分の分散は、第1主成分から第6主成分にかけて順番に小さくなっている。

主成分分析は25人の生徒の6つの特性値(X50m走など)を用いて、(相関行列の)主成分分析を行ったので、分散の合計値は6である。

各主成分の寄与率は、主成分の分散 / (分散の合計値 = 6) である。

通常、主成分分析では、主成分の分散が1以上、または累積寄与率が80 ~ 90%の主成分で判断する。この場合は、第2主成分。約30%は、情報の損失である。

Proportion of Variance (分散の寄与率 Comp1の場合 $2.7927399 / 6 = 46.54\%$)

Cumulative Proportion (累積寄与率)

通常、主成分分析では、累積寄与率が80 ~ 90%または、主成分の分散が1以上の主成分で判断する。この場合は第2主成分で判断した。30%は、情報の損失

Component loadings: (固有ベクトルの因子負荷量)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
X50m	0.5044835	0.32301364	0.05901315	0.3371511	0.116594618	0.71443046
握力	0.3064130	-0.62873074	-0.22773210	-0.1139704	-0.622383824	0.24206605
垂直	0.4075799	0.07935996	0.62115828	-0.6617498	-0.009404733	-0.06117031
背筋	0.4206980	-0.31854140	-0.4507996	-0.1996461	0.678923684	-0.13239466
反復	0.1674513	0.62452835	-0.5799022	-0.3492137	-0.32479008-	-0.13490346
幅とび	0.5295842	0.05057238	0.13896405	0.5230853	-0.180359799	-0.62571882

主成分得点..

各主成分における固有ベクトルの因子負荷量 (Component loading) とデータ値 (第1表) を用いて主成分得点をサンプルごとに求める。第 a 主成分の因子負荷量とデータ表から算出した主成分得点を PC(a) としている。PC(a) には、第 a 主成分の因子負荷量で求めた 25 名の生徒の主成分得点が含まれていることになる。PC(a) は、6 個存在するが互いに独立 (無相関) している。

縦軸、横軸に pc(a), pc(b) を用いて散布図を作成できる。この散布図から 25 名の生徒の立ち位置を知り、新たな方針などを作成できる。

因子負荷量の表より

Comp1 から 6 測定項目の因子負荷量はすべて+で 0.17 ~ 0.53 である。全ての項目の因子負荷量の値が大きくなると、第 1 主成分のホテルの主成分得点が大きくなることから第 1 主成分は、「総合的な良さ」を示す変数であると考えられる。

このことから CP1 は、総合した得点である。

Comp2 から、握力と背筋は(-0.62,-0.31) で筋力であるが、その他 (X50m 走(0.32)、反復飛び (0.62)、垂直跳び(0.079)、幅跳び (0050) を運動力と考える。運動力が優れているほど主成分得点は高くなる。

また、X50m 走と反復飛びを持耐力と考えると、垂直跳びと幅跳びは瞬発力と考えられないか？

縦軸を PC1、横軸を PC2 とする散布図を作成できる。

この散布図に、PC1 = 0、PC2 = 0 で画面を 4 分割して検討する。

第 1 象限

総合運動能力は高い。どちらかという短時間で測定される垂直跳びや幅跳びよりも x50m 走と反復跳びのデータが高いと CP2 の主成分得点が高くなる。

生徒番号 ; 6, 21, 25, 14, 16. 1, 24

第 4 象限

総合運動能力は高い。握力と背筋のデータ値が高いと CP2 の得点が低くなる。

生徒番号 ; 8, 18, 9, 4, 5

第 2 象限

総合運動能力は低い。どちらかという短時間で測定される垂直跳びや幅跳び

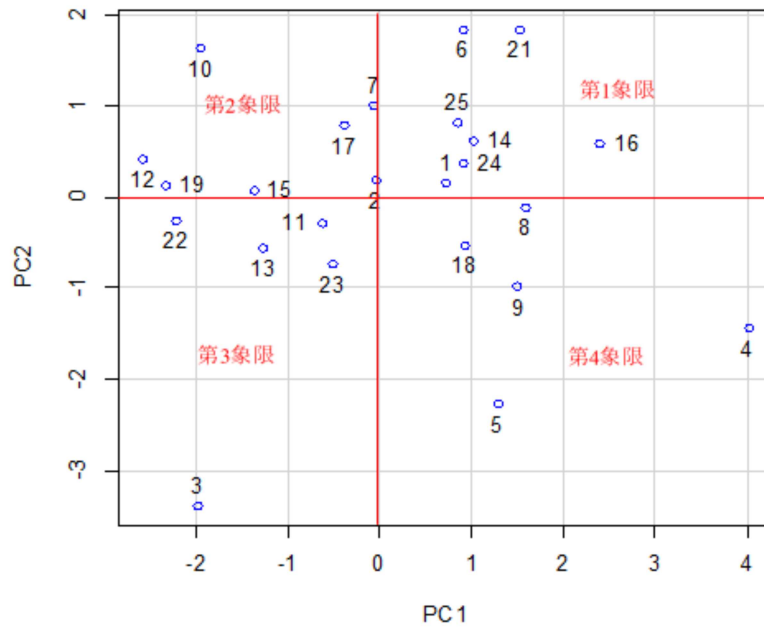
よりも x50m 走と反復跳びのデータが高いと主成分得点が高くなる。

生徒番号 ; 10, 7, 17, 12, 19, 15, 2

第3象限

総合運動能力は低い。全般的に体力向上に努める必要がある。

散布図で点の近くの数字は生徒番号



この散布図から 25 名の生徒のの立ち位置を知ることができ、新たな体力強化の方向付けが課題となる。

実施例 2

東京都内のホテルを 20 施設選び、6 つの項目（特性値：客室、施設、食事、フロント、サービス、予約対応）について、評価した結果が次のデータ表である。各項目 80 点満点で、宿泊客のアンケート結果をもとに点数付けしている。このデータを主成分分析により解析する。

（主成分分析の基本と活用 内田 治著 日科技連 P18 図 2.1 ホテルに関するデータ）

第 1 表 20 ホテルデータ表

ホテル番号	客室	施設	食事	フロント	サービス	予約対応
1	36	55	41	44	31	53
2	41	65	53	51	44	72
3	20	46	47	56	32	48
4	32	56	50	64	41	54
5	34	54	58	53	29	41
6	34	61	50	50	38	56
7	14	44	30	71	49	72
8	30	52	50	72	51	66
9	42	61	68	66	48	63
10	41	57	56	65	49	65
11	57	67	70	57	42	59
12	14	39	36	42	28	36
13	27	48	84	54	56	56
14	18	42	42	57	47	54
15	41	79	68	62	51	68
16	19	51	39	54	59	64
17	26	59	50	56	31	47
18	25	57	48	62	45	66
19	54	73	67	57	44	67
20	40	61	60	71	58	66

	mean	sd	min	max
サービス	43.65	9.59	28	59
フロント	58.20	8.42	42	72
客室	32.25	12.16	14	57
施設	56.35	10.11	39	79
食事	3.35	13.28	30	84
予約対応	58.85	10.06	36	72

	サービス	フロント	客室	施設	食事	予約対応
サービス	1.00	0.60	0.09	0.16	0.29	0.75
フロント	0.60	1.00	0.10	0.14	0.13	0.60
客室	0.09	0.10	1.00	0.83	0.68	0.34
施設	0.16	0.14	0.83	1.00	0.57	0.46
食事	0.29	0.13	0.68	0.57	1.00	0.16
予約対応	0.75	0.60	0.34	0.46	0.16	1.00

測定項目 6 項目間には互いに相関があり、データの広がりにも差があるので、各ホテルの評価を簡単に行うことはできない。そのため、主成分分析を行い比較検討する。(測定項目内でデータを基準化した後に主成分分析を行う。)

主成分分析の出力結果

Component variances (主成分の分散=固有値)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
固有値	3.0208863	1.7314366	0.5971670	0.4315039	0.1385216	0.0804843

Importance of components (分散の寄与率と累積寄与率)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Proportion of Variance	0.503481	0.288572	0.099527	0.071917	0.023086	0.013414
寄与率 (PV/6 %)	50.35%	28.85%	9.95%	7.19%	2.31%	1.34%
累積寄与率	50.35%	79.20%	89.16%	96.34%	98.66%	100%

Comp1 : 第 1 主成分、Comp2 : 第 2 主成分、・・・、Comp6 : 第 6 主成分

主成分の分散は、第 1 主成分から第 6 主成分にかけて順番に小さくなっている。

主成分分析は 20 ホテルの 6 つの特性値 (客室や施設など) を用いて、(相関行列の) 主成分分析を行ったので、分散の合計値は 6 である。

各主成分の寄与率は、主成分の分散 / (分散の合計値 = 6) である。

通常、主成分分析では、主成分の分散が 1 以上、または累積寄与率が場合によって異なるが 80 ~ 90% の主成分で判断する。この場合は、第 2 主成分。約 20% は、情報の損失である。

Component loading (固有ベクトルの因子負荷量)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
サービス	0.383	0.461	0.336	0.468	0.003	0.554
フロント	0.331	0.467	0.051	-0.814	-0.060	0.038
客 室	0.430	-0.443	-0.160	-0.128	0.707	0.272
施 設	0.449	-0.362	-0.397	0.011	-0.676	0.225
食 事	0.387	-0.346	0.733	-0.005	-0.127	-0.419
予約対応	0.451	0.346	-0.403	0.316	0.149	-0.624

主成分得点

各主成分における固有ベクトルの因子負荷量 (Component loading) とデータ値 (第1表) を用いて主成分得点をサンプル (20ホテル) ごとに求める。第 a 主成分の因子負荷量とデータ表から算出した主成分得点を PC(a) としている。PC(a) には、第 a 主成分の因子負荷量で求めた 20 ホテルの主成分得点が含まれていることになる。PC(a) は、6 個存在するが互いに独立 (無相関) している。(a=1,2,3,4,5,6)

縦軸、横軸に pc(a), pc(b) を用いて散布図を作成できる。

因子負荷量の表より

Comp1 から因子負荷量はフロント (0.331) から、予約対応 (0.451) まですべて (+) である。全ての項目の因子負荷量の値が大きくなると、第1主成分のホテルの主成分得点が大きくなることから第1主成分は、「総合的な良さ」を示す変数であると考えられる。

Comp2 から因子負荷量の符号は、フロント・サービス・予約対応の符号が (+) で、食事・施設・客室の符号が (-) である。

フロント・サービス・予約対応=人的サービスの値が大きくなると、主成分得点が大きくなる。食事・施設・客室 (=物的サービス) の値が大きくなると主成分得点は小さくなる。このことから、サービスの良さが人的か物的かを判断することができる。

x 軸を PC1, y 軸を PC2 として散布図を作成する。

散布図で PC1 = 0, PC2 = 0 の線を入れて4分割し、ホテルの評価を考える。

散布図で・の近くの数字はホテル番号

第1象限のホテルは、総合的な評価が高く、かつ 人的サービスも良好なホテル群で、この象限のホテルは、団体客の利用 (例えば修学旅行など) に適していると考えられる。

ホテル : 8, 20, 18, 10

第2象限のホテルは、規模は小さいが、人的サービスが良いホテル群。静かに落ち着いた旅館で、長期滞在して、思考するのに適していると考えられる。

ホテル : 7, 16, 14, 4

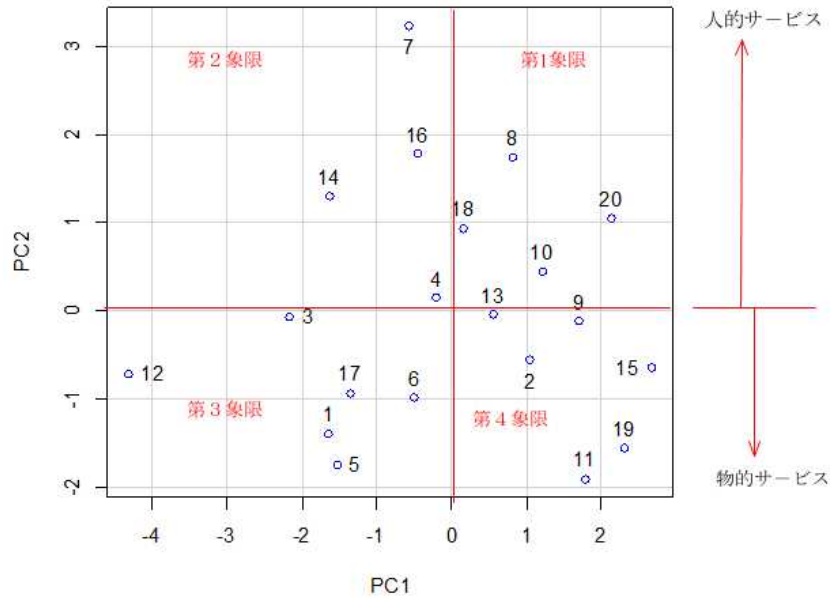
第4象限のホテル群は、総合的に充実しているが、人的なサービスはそれ程でもない。

同じ職場の多くの従業員達の人間関係を良好にするための宿泊・慰安旅行に適していると考えられる。

ホテル：9, 13, 15, 2, 19, 11

第3象限ホテル群は、総合的な評価が低く際だった特徴も無く、人々から忘れられる存在のホテルであると考えられ。

ホテル：3, 12, 17, 6, 1, 5



6 おわりに

この散布図から 20 ホテルの立ち位置を知ることができ、課題や方針などを作成できる。

実施例 3

中学 2 年生 23 人の国語、社会、数学、理科、英語の期末試験の成績を示したものである。主成分分析を行い検討する。

第 1 表

番号	国語	社会	数学	理科	英語
1	29	33	55	79	84
2	71	68	72	64	97
3	74	91	79	76	100
4	52	58	58	60	85
5	77	92	96	88	98
6	60	85	66	66	88
7	81	91	73	63	95
8	61	84	72	78	92
9	70	75	81	67	96
10	53	70	72	51	92
11	69	64	96	57	97
12	87	89	90	85	100
13	83	75	96	81	98
14	76	61	67	57	86
15	87	82	78	82	97
16	77	80	78	70	94
17	38	43	45	12	96
18	67	73	78	67	95
19	83	77	80	67	100
20	47	61	66	21	95
21	70	62	88	51	96
22	81	51	63	66	92
23	51	16	36	48	84

	mean	sd	最低	最高		英語	国語	社会	数学	理科
英語	93.78	5.10	84	100	英語	1.00	0.58	0.62	0.70	0.15
国語	67.13	15.83	29	87	国語	0.58	1.00	0.65	0.68	0.55
社会	68.73	19.36	16	92	社会	0.62	0.65	1.00	0.73	0.51
数学	73.26	15.46	36	96	数学	0.70	0.68	0.73	1.00	0.53
理科	63.30	18.48	12	88	理科	0.15	0.55	0.51	0.53	1.00

5 個の特性値を持つ主成分分析のため、主成分は、5 個

5 教科には、それぞれの間に関連があり、最低点と最高点間にも違いがあるので

相関行列の主成分分析を行う。

主成分分析の結果 ()

Component variances: (分散=固有値)

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
3.3409344	0.8539891	0.3482303	0.2801380	0.1767083

Importance of components: (分散の寄与率と累積寄与率)

	Comp.	Comp.2	omp.3	Comp.4	omp.5
Proportion of Varianc	0.668186	0.1707978	0.0696460	0.056027	0.03534166
Cumulative Proportio	.6681869	0.8389847	.90863074	0.9646583	1.0000000

Comp.1 ; 第1主成分、Comp.2 : 第2主成分、・・・Comp,5 : 第5主成分

主成分の分散は、第1主成分から第5主成分にかけて順番に小さくなっている。

主成分分析は23名の生徒の5つの特性値(英語や国語など)を用いて、(相関行列の)主成分分析を行ったので、分散の合計値は5である。

各主成分の寄与率は、主成分の分散 / (分散の合計値 = 5) である。

通常、主成分分析では、主成分の分散が1以上、または累積寄与率が80 ~ 90%の主成分で判断する。この場合は、第2主成分、約17%は、情報の損失である。

Component loading (固有ベクトルの因子負荷量)

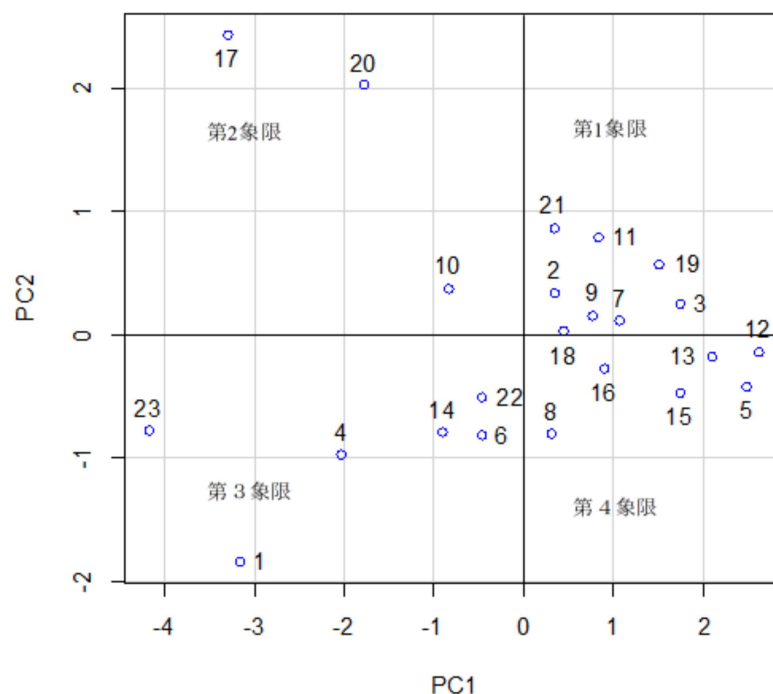
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
英語	0.4200817	0.61813004	0.04961150	0.2348448	0.61954281
国語	0.4694766	-0.0868734	0.80868076	-0.288177	0.18717412
社会	0.4783321	0.02631879	-0.5330848	-0.695976	-0.0440865
数学	0.4954799	0.09430782	-0.2238307	0.5549770	-0.6224999
理科	0.3589168	-0.7751010	-0.0964076	0.2634766	0.43781592

主成分得点

各主成分における固有ベクトルの因子負荷量 (Component loading) とデータ値 (第1表) を用いて主成分得点をサンプルごとに求める。第a主成分の因子負荷量とデータ表から算出した主成分得点を PC(a) としている。PC(a)には、第a主成分の因子負荷量で求めた23名の生徒の主成分得点が含まれていることになる。PC(a)は、5個存在するが互いに独立(無相関)している。

縦軸に PC2、横軸に PC1 を用いて散布図を作成できる。

さらに CP1=0、CP2=0 で、4つに分割した。



散布図の見方

因子負荷量の表より

Comp1 から 5 教科の因子負荷量はすべて+で 0.35 ~ 0.49 である。全ての項目の因子負荷量の値が大きくなると、第 1 主成分の主成分得点が大きくなることから第 1 主成分は、「総合的な良さ」を示す変数であると考えられる。このことから CP1 は、総合した得点である。

Comp2 から、因子負荷量は英語が 0.62、理科が-0.77 である。

英語の得点がよく、理科の得点がよくない方が主成分得点は高くなる。

他の教科（国語、社会、数学）の因子負荷量が 1 桁小さいがこの影響はないと考えられない場合も存在するから注意が必要である。

第 1 象限：主成分得点が高い。英語の得点がよく、理科の成績が平均より高いと考えられる。

生徒番号：21, 11, 19, 2, 3, 9, 7, 18.

第 2 象限：主成分得点が低い。英語の得点がよく、理科の成績が平均より高い場合が考えられる。(英語の成績差で、第 1 象限と第 2 象限にわかれた。)

生徒番号：17, 20, 10

第 3 象限：主成分得点が低い。英語の得点が低く、理科の成績が平均より低い場合が考えられる。

生徒番号：22, 6, 14, 4, 23, 1

第 4 象限：主成分得点が高い。英語の得点が低く、理科の成績が平均より低い場合が考えられる。

第 3 象限の生徒と第 4 象限の生徒の理科の成績を比較すると、第 4 象限の生徒の方が成績がよい。

生徒番号：12, 13, 16, 5, 15, 8

結びとして、第 3 象限の学力向上が必要

この散布図から 23 名の生徒のの立ち位置を知り、生徒一人一人はさらに勉学に励む結果になると思う。また、先生にとっても、指導の方向付けができた
と考える。

実施例 4 主要大学の授業満足度

国内主要 13 大学における学生の授業満足度に関する授業内容、選択自由度、授業一体感の三つの項目に関する 5 段階による回答でのアンケート調査により表のデータが得られた。(大学ランキング 1995. 朝日新聞より)

主成分分析を行い、学生達の自大学の評価を知る。

番号	大学番号	授業内容	選択自由度	授業一体感
1	1	5	2	5
2	2	1	1	5
3	3	4	3	4
4	4	1	1	1
5	5	5	5	4
6	6	1	1	3
7	7	1	1	2
8	8	1	5	3
9	9	1	1	4
10	10	5	5	4
11	11	1	3	1
12	12	4	5	1
13	13	2	4	1

第 1 表

平均値・標準偏差

	平均値	標準偏差
授業一体感	2.92	1.55
授業内容	2.46	1.80
選択自由度	2.84	1.77

特性値間の相関

	授業一体感	授業内容	選択自由度
授業一体感	1.00	0.36	-0.12
授業内容	0.36	1.00	0.57
選択自由度	-0.12	0.57	1.00

3 個の特性値間には、それぞれの間に関係があるので、評価は簡単ではないので相関行列の主成分分析を行う。

3 個の特性値を持つ主成分分析のため主成分は、3 個

主成分分析の結果

Component variances: (分散 = 固有値)

Comp.1	Comp.2	Comp.3
1.626955	1.114204	0.258841

$$1.626955 + 1.114204 + 0.258841 = 3$$

Importance of components: (分散の寄与率と累積寄与率)

:

	Comp.1	Comp.2	Comp.3
Proportion of Variance	0.5423183	0.3714014	0.08628034
Cumulative Proportion	0.5423183	0.9137197	1.00000000

Comp.1 ; 第 1 主成分、Comp.2 : 第 2 主成分、Comp,3 : 第 3 主成分
主成分の分散は、第 1 主成分から第 3 主成分にかけて順番に小さくなっている。

主成分分析は国内主要 13 大学の 3 つの特性値(授業一体感、授業内容、選択の自由度)を用いて、(相関行列の) 主成分分析を行ったので、分散の合計値は 3 である。

各主成分の寄与率は、主成分の分散 / (分散の合計値 = 3) である。

通常、主成分分析では、主成分の分散が 1 以上、または累積寄与率が 80 ~ 90%の主成分で判断する。この場合は、第 2 主成分、約 8%は、情報の損失である。

主成分得点

各主成分における固有ベクトルの因子負荷量 (Component loading) とデータ値 (第 1 表) を用いて主成分得点を 1 3 大学ごとに求める。第 a 主成分の因子負荷量とデータ表から算出した主成分得点を PC(a)としている。PC(a)には、第 a 主成分の因子負荷量で求めた国内主要 13 大学の主成分得点が含まれていることになる。PC(a)は、1 3 大学の主成分得点を含んでいるが、3 個の特性値ごとに存在する。3 個の PC は互いに独立である。

Component loadings: (固有ベクトルの因子負荷量)

	Comp.1	Comp.2	Comp.3
授業一体感	0.3114407	0.84341277	0.4377895
授業内容	0.7333269	0.07968056	-0.6751908
選択自由度	0.6043479	-0.53132467	0.5936815

すでに、Comp1、Comp2、Comp3 における各大学の主成分得点は、「Rcmdr」の「データセットの編集」をクリックすることで知ることができる。

ただし、このデータは、画像データであるので、そのまま一太郎にコピーできない。けれども、このデータを用いて散布図にすることが可能である。

縦軸に PC2、横軸に PC1 を用いて散布図を作成した。

散布図の見方 (因子負荷量より)

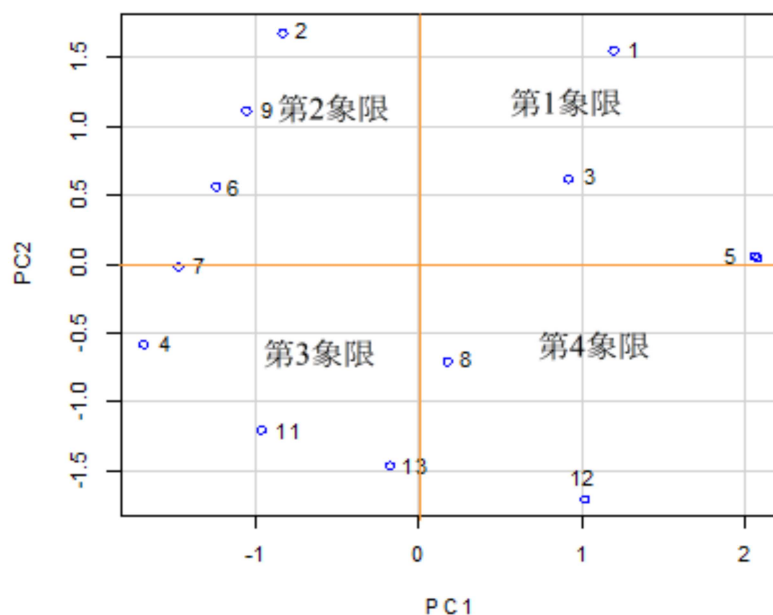
Comp1 から 3 特性値の因子負荷量はすべて+で 0.31 ~ 0.73 である。このことから

CP1 は、総合した主成分得点である。5 の大学が一番で、4 の大学が一番低い。

Comp2 では、授業一体感と授業内容が (+) である。授業に対する良さであると

考えられる。選択自由度は (-) であるから、選択自由度の高い大学の主成分得点は、低くなる。授業内容の影響は小さい。

打点横の数字は、大学の番号である。



第1象限の大学

授業総合の主成分得点が高い。選択自由度は平均よりも低い。

大学は 5、1、3 (5は、東京大学と慶応義塾大学が重なっている。)

第2象限の大学

授業総合の主成分得点は低い。選択自由度が平均よりも低い。

大学は、2、9、6、7

第3象限の大学

授業総合の主成分得点が低い。選択自由度が平均よりも高い。

大学は、4、11、13

第4象限の大学

授業内容の主成分得点が高い。選択自由度が平均よりも高い。

大学は、8、12

まとめとして、

希望の大学に入学できたので 授業満足度は高いと考えたが、そのようにはなっていない。大学側も、満足されていると考えているが、必ずしもそうではない。

近くの競合大学がある大学ほど満足感が低い傾向がある。

選択自由度が高いということは、講義を選べる度合いが高い (必須科目が少ないことか?)