

例題 17: [SampleData] の「DB 型 打撃成績.SFD6」は、プロ野球の個人別打撃成績を 13 の要因 (変数) について表したのものである。主成分分析により、個人の打撃成績の特徴を少数の因子で要約せよ。



選手名	打率	得点	安打	二塁打	三塁打	本塁打	打点	盗塁	犠打	犠飛	四球	三振	併殺打
坂本 (巨)	0.3455	91	160	27	3	22	73	13	1	6	77	62	6
鈴木 (広)	0.3333	72	151	26	7	27	93	16	3	3	50	77	10
筒香 (D)	0.3185	85	144	27	4	40	102	0	0	2	86	102	5
菊池 (広)	0.3172	92	178	22	2	13	55	13	23	3	40	101	3
山田 (ヤ)	0.3145	102	145	25	3	38	100	30	0	4	94	93	15
福留 (神)	0.3089	51	135	24	3	11	56	0	0	5	59	75	6
坂口 (ヤ)	0.3003	73	152	14	5	0	39	6	5	5	60	65	5
倉本 (D)	0.2998	38	155	19	2	1	38	2	5	3	21	95	13
新井 (広)	0.2997	65	134	22	2	18	98	0	0	4	54	98	12
...	...	n=54	以下省略

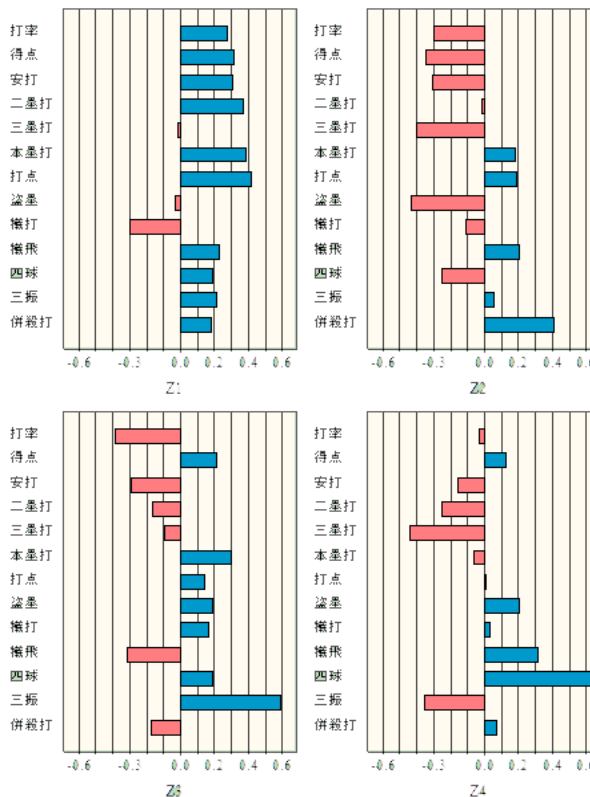
データファイルを読み込み、[統計] メニューの [多変量解析] の [主成分分析] を実行します。このデータでは、そのまま 13 変数を**対象変数**として指定します。そして、主成分の数を 4、各変数の単位が全く異なるため計算法に「相関行列」を指定します。その後、**実行** ボタンを押すと次の結果が表示されます。

計算出力例

<< 主成分分析 >> [第1頁: 群]
有効データ数 = 54

固有ベクトルと寄与率

変数名	Z1	Z2	Z3	Z4
打率	0.2762	-0.2976	-0.3802	-0.0263
得点	0.3178	-0.3432	0.2158	0.1310
安打	0.3074	-0.2992	-0.2874	-0.1574
二塁打	0.3746	-0.0141	-0.1602	-0.2509
三塁打	-0.0099	-0.3994	-0.0923	-0.4355
本塁打	0.3899	0.1837	0.3023	-0.0608
打点	0.4213	0.1940	0.1440	0.0151
盗塁	-0.0281	-0.4302	0.1939	0.2045
犠打	-0.2963	-0.1037	0.1707	0.0357
犠飛	0.2290	0.2088	-0.3093	0.3147
四球	0.1934	-0.2491	0.1931	0.6543
三振	0.2123	0.0582	0.5936	-0.3538
併殺打	0.1864	0.4101	-0.1669	0.0741
固有値	4.1427	2.9634	1.4878	0.9397
寄与率	0.3187	0.2280	0.1144	0.0723
累積寄与率	0.3187	0.5466	0.6611	0.7333



左側の出力は**固有ベクトル**の行列で、求まった 4 つの主成分と元の変数値との相関係数が示されています。各主成分にどの程度の情報が要約されているかは、表下端の**固有値**の大きさと、**寄与率**から推定できます。

この結果から、第1主成分 (z1) の固有値は4.14と大きく、寄与率より全体の31.8%の情報が集約されています。また、z1は打撃に関する変数値のほとんど(打率、得点、安打、二塁打、本塁打、打点など)と正の相関があることより、総合的な打撃力の強さを表すと考えられます。また第2主成分 (z2) の固有値は2.96で、全体の22.8%の情報が集約され、主に盗塁、三塁打、打率、安打と逆相関、併殺打と正相関などから、俊足でな傾向を表していると考えられます。

一方、第3および第4主成分 (z3, z4) は固有値や寄与率も小さく、その解釈はやや困難です。従って、第1と第2主成分に限定 (2つで累積寄与率54.7%) して、主成分をスコアをStatFlex形式で出力した結果は下記の通りで、左側は、z1の大きさと、z2の大きさと降順に並べ替えた結果です。

統計データ2		1	2
群1		第1主成分	第2主成分
1	山田(ヤ)	3.838	-1.333
2	筒香(D)	3.276	-0.948
3	浅村(西)	2.840	1.183
4	デスパイネ(2.743	2.196
5	丸(広)	2.738	-2.572
6	坂本(巨)	2.717	-1.894
7	メヒア(西)	2.553	2.466
8	鈴木(広)	2.156	-1.452
9	秋山(西)	2.034	-2.262
10	バレンティン	2.029	2.315
11	レアード(日)	1.901	2.246
12	柳田(ノ)	1.798	-2.149
13	中田(日)	1.726	2.171
14	角中(口)	1.670	-2.209
15	松田(ノ)	1.601	0.659
16	新井(広)	1.506	0.896

データベース型：前方視点 (1ページ目 / 列数=2 / 行数=54)

統計データ2		1	2
群1		第1主成分	第2主成分
1	ロベス(D)	1.140	2.496
2	メヒア(西)	2.553	2.466
3	小林誠(巨)	-4.485	2.408
4	内川(ノ)	1.471	2.381
5	バレンティン	2.029	2.315
6	レアード(日)	1.901	2.246
7	デスパイネ(2.743	2.196
8	中田(日)	1.726	2.171
9	ゴメス(神)	0.472	1.923
10	村田(巨)	1.028	1.868
11	ギャレット(-1.260	1.690
12	ビシエド(中)	0.048	1.624
13	T-岡田(オ)	0.401	1.608
14	銀次(楽)	-2.266	1.500
15	堂上(中)	-2.240	1.391
16	貞谷(神)	-1.653	1.319

データベース型：前方視点 (1ページ目 / 列数=2 / 行数=54)

この結果から、z1に総合的な打撃力の強さが表され(値が小さいほど重症)が、z2に攻撃の際の俊足度の低さが要約されていると解釈できます。なおz1, z2のスコアは、その平均値は必ず0.0となり、その標準偏差は、それぞれ固有値の平方根($\sqrt{4.143} = 2.035$ 、 $\sqrt{2.963} = 1.721$)となります。

ここで、次節の因子分析との違いは、多変量情報を少数の因子に括り出しているだけ、分析に用いた13変数の共通性により、それらをグループ分けする機能を持たない点です。同じデータを使って、因子分析で解析した結果は、次節の例題で示していますので、結果を比較してみてください。



Rによる分析結果

```
> dat = read.csv("DB型 打撃成績.CSV", header=T)
> res = princomp(data=dat, ~打率+得点+安打+二塁打+三塁打+本塁打+打点+盗塁+犠打+犠飛+四球+三振+併殺打, cor=T)
> res$loadings
```

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13
打率	-0.276	0.298	-0.380			0.267	-0.158		0.495		-0.417	0.386	-0.118
得点	-0.318	0.343	0.216	-0.131	0.232					-0.110	0.613	0.331	-0.401
安打	-0.307	0.299	-0.287	0.157	0.249	0.245		-0.292		0.263	0.270	-0.502	0.321
二塁打	-0.375		-0.160	0.251			-0.541	0.170	-0.554	-0.350			
三塁打		0.399		0.435		-0.490	0.382	0.377	-0.116	0.291			
本塁打	-0.390	-0.184	0.302				0.214	0.200	0.277	-0.321		0.138	0.649
打点	-0.421	-0.194	0.144				0.227	0.223	0.170	-0.112	-0.239	-0.568	-0.488
盗塁		0.430	0.194	-0.204		0.497	0.395		-0.446	-0.213	-0.284		
犠打	0.296	0.104	0.171		0.695		-0.298	0.496	0.108			-0.113	0.111
犠飛	-0.229	-0.209	-0.309	-0.315	0.538	-0.341	0.256	-0.264	-0.216		-0.265	0.204	
四球	-0.193	0.249	0.193	-0.654	-0.252	-0.283	-0.294	0.123		0.362	-0.108	-0.123	0.179
三振	-0.212		0.594	0.354	0.173		-0.167	-0.335		0.396	-0.312	0.194	
併殺打	-0.186	-0.410	-0.167			0.401	0.109	0.447	-0.249	0.508	0.181	0.194	

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077	0.077
Cumulative Var	0.077	0.154	0.231	0.308	0.385	0.462	0.538	0.615	0.692	0.769	0.846	0.923	1.000

```
> summary(res)
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Standard deviation	2.03536	1.72144	1.21976	0.96938	0.90700	0.83330	0.71843	0.65140	0.59480	0.54236
Proportion of Variance	0.31867	0.22795	0.11445	0.07228	0.06328	0.05341	0.03970	0.03264	0.02721	0.02263
Cumulative Proportion	0.31867	0.54662	0.66107	0.73335	0.79663	0.85004	0.88975	0.92239	0.94960	0.97223

	Comp.11	Comp.12	Comp.13
Standard deviation	0.44693	0.31839	0.24472
Proportion of Variance	0.01537	0.00780	0.00461
Cumulative Proportion	0.98760	0.99539	1.00000