

DEA による回帰型データのクラスター分析

新村 秀一

成蹊大学一般研究報告 第45巻第3分冊

平成23年5月

BULLETIN OF SEIKEI UNIVERSITY, Vol.45 No.3

May, 2011

DEAによる回帰型データのクラスター分析

新村 秀一¹

1. はじめに

DEA (Data Envelopment Analysis, D-効率性分析, 包絡分析法) は, 複数の同質な組織や個人をDMU (Decision Making Unit, 意思決定主体) と呼び, 入力変数と出力変数に各DMUに最適な重みを付け, 入力と出力の比 (効率値) の最大値が1 (1になったDMUの集合を参照集合という) になるように制限して, DMUの効率性を相対的に評価する手法である。各DMUに最適な重みを付けても, 効率的な1になるDMUと, その重みを適用した他のDMUが1になり, 自分は非効率で1以下になるDMUもある。そして非効率なDMUは, 自分に最適な評価の重みで1になる効率的なDMUと比較して, 入力を減らすか出力を増やすことで, 何を改善すれば効率的になるかを検討し改善を促す手法である [1]。

分析したいDMUが k 個ある場合, k 個の異なった重みで全てのDMUの効率値を計算する。そして各DMUの重みで計算した効率値ベクトルは必ず一意である特定のDMUの組が効率的なパターンになるいずれかの組 (クラスター) に属することになる。これによって, 回帰型のデータは, 通常のクラスター分析と異なり一つの固定されたクラスターに属することになる。同一のクラスターに属するDMUは, 効率値1になるDMUを目標にして改善すれば, 自身も効率的になることができる。

評価法に関しては, 評価の悪いDMUは入力 (予算, 人員, 顧客層等) が少ないためと言い訳が先に立つ。しかし自分に最適な重みで効率的であるか否かを評価するので, これまでと違った新しい改善の参考指標として使えると考えられる。

改善活動は, 製造現場で成功をおさめたが, ブルーカラのものと考えられ, ホワイトカラーや知識人にとって受け入れ難かった。しかし, DEAによる評価法による改善は, 面従腹背になりがちな不透明な企業体における新しい意識改革に有効と考えられる。ただし, 多くの関係者が納得するように入出力の項目を見直し改善する必要があり, 性急な実施は慎むべきであろう。

本論文では, 学生データ [2] と2008年の卒業生の高橋君が統計レポート作成のために集めた野球選手のデータを用いる。学生データは, 筆者がこれまで最適線形判別関数

1 Shuichi Shinmura New Cluster Analysis of Data using Regression Analysis by DEA

[3]の研究に用い、SAS, SPSS, JMPの統計入門の解説書に用いている。しかし、データがDEAの要求する前提を満たしていないと考えられ、その影響の検討に用いる。一方、野球選手の分析はDEAを用いて広く分析されていて[4][5]、今回提唱するクラスター分析法の検証の比較に適した分野と考えられる。また将来、学生の成績評価法につながればと考えている。

2. 経営効率性分析あるいは包絡分析法とは

(1) DEA法

DEAは、企業の事業部や百貨店などの複数の店舗、自治体の複数の図書館などの各種事業体をDMUとみなし、その効率性を評価する手法である。あるいは利用局面を拡大して、野球選手の年俵が実績とどう関係しているか等の個人評価にも利用できる。1978年にテキサス大学のCharnes, Cooper and Rhodes [6]によって提案されたので、CCRモデルと呼ばれている。

評価手法としては、回帰分析や判別分析などの手法が思いつく。これらの統計手法は、分析対象のデータに共通の重みがつけられる。

しかし、DEA法はガソリンや電機で動くモータをイメージすればよい。入力であるガソリンに対し、モータの出力は損出が発生するので、出力／入力の比は1以下になる。1に近いほど、エネルギー効率が良いモータと考えられ、小馬力や大馬力のモータ毎に効率的なものが存在する。この場合は1入力1出力であるが、一般的には多入力多出力も扱える。ただし統計手法と異なり、あまり多くの入出力項目で評価できない難点がある。

(2) LPによる定式化 (CCRモデル)

例えばk個の大学の評価を考える。入力変数としては、予算、教員数、職員数、非常勤講師数等が考えられる。出力変数は、入学者数、卒業生数、就職率などが考えられる。m個の入力項目を x_1, \dots, x_m とし、n個の出力項目を y_1, \dots, y_n とする。h番目の大学(DMU_h)の入力項目を x_{1h}, \dots, x_{mh} とし、n個の出力項目を y_{1h}, \dots, y_{nh} とする。この場合、h番目のDMU_hの効率値は次の比になる。

$$DMU_h = (b_{1h} * y_{1h} + \dots + b_{nh} * y_{nh}) / (a_{1h} * x_{1h} + \dots + a_{mh} * x_{mh})$$

ここで、この比が1以下になるように重み (b_{1h}, \dots, b_{nh}) と (a_{1h}, \dots, a_{mh}) を決めてやればよい。重みを各DMUに無関係に一定(固定)にすれば、これまでの統計アプローチになる。DEA法の特徴は、DMUごとにその効率を最大になるように個別の重みを与える点である。しかし、その重みを他のDMUにも適用し効率値を1以下にするという制約を課すことにする。ここでh番目のDMU_hを考えると次の定式化になる。

$$MAX = (b_{1h} * y_{1h} + \dots + b_{nh} * y_{nh}) / (a_{1h} * x_{1h} + \dots + a_{mh} * x_{mh}); \quad (1)$$

$$(b_{1p} * y_{1p} + \dots + b_{np} * y_{np}) / (a_{1p} * x_{1p} + \dots + a_{mp} * x_{mp}) <= 1; \text{ for } p=1, \dots, k \quad (2)$$

ただし、測定値も重みも正の値と仮定しているのが難点である。この仮定によって、制約式(2)で絶対値が1以下という定式化が避けられる。このモデルは、数理計画法で

分数計画法と呼ばれている。これを次の同値なLPモデルに置き換えたものが一般的にCCRモデルとして知られている。すなわち、目的関数(1)の分母を1とすることで分数計画法はLPになる。

$$\text{MAX} = b_{1h} * y_{1h} + \dots + b_{nh} * y_{nh} ;$$

$$a_{1h} * x_{1h} + \dots + a_{mh} * x_{mh} = 1 ;$$

$$b_{1h} * y_{1p} + \dots + b_{nh} * y_{np} \leq a_{1h} * x_{1p} + \dots + a_{mh} * x_{mp} ; \text{for } p=1, \dots, k$$

ここで注意したいことは、各DMUに対して上のLPモデルをk個解く必要がある。すなわち、k個の大学(DMU)に対してk個の異なったLPモデルを繰り返し解く必要がある。

(3) 効率的フロンティア

DEAの特徴を、図2.1の1入力2出力を例にして説明する。例えば大学の場合、教職員数を入力とし、入学者数と就職者数を出力と考えればよい。各DMU_hに対してCCRモデルを1回解く。A、B、C、Dの大学の効率値が1である。これらを参照集合={A,B,C,D}と呼んでいる。効率値が1のA、B、C、Dを結んだ線分は、全体として凸体になり、効率的フロンティアと呼ばれている。どのような重み付けを行っても、考えている全てのDMUはこの凸体に内包される。これが包絡と呼ばれるゆえんである。大学Gは自分に最適な重みで評価して例えば0.7であったとする。DMU_gは、Gの良いところを取り入れて重み付けしても非効率的であり、入力か出力を改善する必要がある。すなわち仮想のP大学が努力目標になる。あるいは実態のあるBかCを目標にして改善をはかれば良い。この場合、AとDを努力目標にする必要がない。

しかし、多入力多出力のデータではこのような図を描くことができない。そこで本論文では、参照集合の組み合わせパターンの異なるものをクラスターと考え、DMUを一意に分割する新しいクラスター分析を提案したい。各クラスターには効率値が1のDMUが1個以上含まれ、非効率なDMUにとっての努力目標になる。

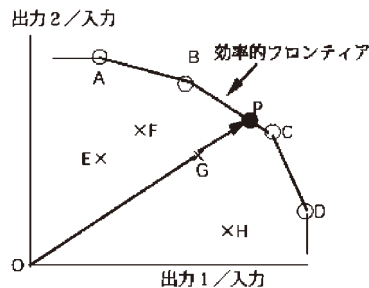


図2.1 1入力2出力

(4) 非効率的フロンティア

一方、次のようなモデルを考えることができる。

$$\text{MIN} = b_{1h} * y_{1h} + \dots + b_{nh} * y_{nh} ;$$

$$a_{1h} * x_{1h} + \dots + a_{mh} * x_{mh} = 1 ;$$

$$b_{1h} * y_{1p} + \dots + b_{nh} * y_{np} \geq a_{1h} * x_{1p} + \dots + a_{mh} * x_{mp} ; \text{for } p=1, \dots, k$$

このモデルは、図2.2に示すように、非効率的なフロンティアを見つけてくれる。出力をどこまで落とせばもっとも非効率であるかが分かる。大学A, B, Cの非効率率は1であり、大学Gは1.5だったとしよう。これが図の大学Pまで非効率率的になれば、1に下がる。CCRモデルで効率的なDMUに飴（評価をあげる）、逆CCRモデルで非効率的なDMUに鞭（評価を下げる）の様な使い分けをすればよい。DEAは、米国会計学会の会長で数理計画法の研究者でもあったCharnes & Cooperが開発した評価の手法である。効率性を評価したい組織や個人など（DMU）の入力と出力項目に、各DMUに最適な重みを求めて評価する手法である。繰り返しになるが回帰分析は、対象とするDMUに共通の重み（係数）を決める点異なる。

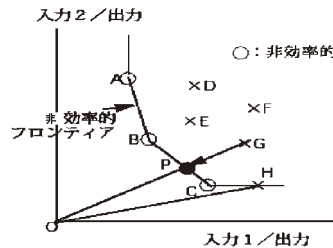


図2.2 逆CCR（非効率的なフロンティア）

3. LINGOのプログラムと分析結果

3.1 LINGOのプログラム

以下が4章で詳細に分析する40人の学生データ [7] のDEAのLINGO² [8] プログラムである。DEAの欠点として、入力と出力データは正の値で大きくなるほど効率が良くなるということを前提にしている点である。勉強時間と支出を入力とし成績を出力と考えると、勉強時間が増えれば成績も良くなるが、支出は少なくなれば多分成績が良くなると考えられるのでDEAの前提を逸脱していると考えられる。

SETSとENDSETSは集合節といい、1次元集合DMUとFACTORを定義している。DMUは40人の学生を表し、SCOREは各学生の効率値が入る1次元配列である。FACTORは2入力1出力の重みを表す集合である。DXFは1次元集合DMUとFACTORで定義される(40*3)の2次元集合であり、データが定義される2次元配列Fと各DMUの重みが出力される2次元配列Wを定義している。S40は(40*40)の2次元集合であり、配列Sに40個の重みで計算した40人の学生の効率値ベクトルが出力される。これを効率

2 経済学部で予算で大学の情報教育用のPCにインストール済み。

値配列Sと呼ぶことにする。

DATA：とENDDATAは、プログラムで使用するデータを与えたり、Excelから入力したり、分析結果を出力できる。「NINPUTS=2;」で入力変数がセル範囲Fの最初の2列であることを示す。「FACTOR=F1..F3;」で入力変数と出力変数名をオプションで定義している。「F=@OLE();」でセル名Fの(40*3)の2次元配列データをExcelから入力している。「WGTMIN=.00000;」で重みの下限値を0に、「BIGM=99999;」で上限値を99999に定義している。

2番目のDATA節の「@OLE('C:\LINGO12\SAMPLES\AOYAMA3VarS.XLS','W','SCORE','S'(注：指定しなくて空白でも良い))=W,SCORE,S;」で、Excelに計算した(40*3)の重みと、(40*1)の効率値と、(40*40)の40個の重みで計算した40人の学生の効率値ベクトル(クロス効率値という)をExcelのセル範囲名「W,SCORE,S」に出力する。配列Sの対角要素にはSCOREの値が入る。

MODEL:

SETS:

DMU:SCORE; FACTOR;
DXF(DMU, FACTOR): F, W;
S40(DMU,DMU):S;

ENDSETS

DATA:

NINPUTS = 2; DMU = 1..40; FACTOR= F1..F3;
F=@OLE() ;
WGTMIN = .00000;BIGM = 99999;

ENDDATA

MAX = @SUM(DMU: SCORE);
@FOR(DMU(I): SCORE(I) = @SUM(FACTOR(J) | J #GT# NINPUTS:
F(I, J) * W(I, J));
@SUM(FACTOR(J) | J #LE# NINPUTS: F(I, J) * W(I, J)) = 1;
@FOR(DMU(K):
[LE1] @SUM(FACTOR(J) | J #GT# NINPUTS: F(K, J) * W(I, J))
<= @SUM(FACTOR(J) | J #LE# NINPUTS: F(K, J) * W(I, J)));
@FOR(DXF(I, J): @BND(WGTMIN, W, BIGM));

CALC:

@SET('TERSEO', 1); !Solve main model;
@SOLVE();
!Compute S based on model results;
@FOR(DMU(I):

```

@FOR( DMU( K): S(k,i)=@SUM( FACTOR( J) | J #GT# NINPUTS: F(K, J)
* W( I, J)) /
@SUM( FACTOR( J) | J #LE# NINPUTS: F( K, J) * W( I, J)) );
!Print standard solution report;
@SOLU(); @SET( 'TERSEO', 0);
ENDCALC
DATA:
@OLE('C:\LINGO12\SAMPLES\base3V.XLS', 'SCORE', 'w', 'S')=SCORE,W,S;
ENDDATA
END
    
```

CALC:とENDCALCでは、前で定義されたCCRモデルを「@SOLVE();」で解いた後、40個の重みで効率値配列Sを計算している。もし、この部分をCCRモデルの後に定義すれば最適化モデルの一部と考えられ、非線形計画法になり、計算時間もかかる。

R	S	T	U	V	W	X	Z	AA	AH	AJ	BC
勉強	支出	成績	W1	W2	W3	SCORE	S2	S3	S10	S12	S31
12	4	100	0.038	0.136	0.006	0.614	0.500	0.614	0.405	0.263	0.139
9	2	100	0.000	0.500	0.010	1.000	1.000	1.000	0.577	0.357	0.185
7	3	95	0.056	0.202	0.009	0.866	0.633	0.866	0.626	0.422	0.226
10	2	90	0.000	0.500	0.010	0.900	0.900	0.848	0.474	0.290	0.150
7	3	90	0.056	0.202	0.009	0.820	0.600	0.820	0.593	0.400	0.214
7	3	90	0.056	0.202	0.009	0.820	0.600	0.820	0.593	0.400	0.214
8	3	85	0.053	0.191	0.009	0.733	0.567	0.733	0.505	0.333	0.177
5	4	85	0.052	0.185	0.008	0.711	0.425	0.711	0.654	0.500	0.283
6	5	85	0.042	0.150	0.007	0.575	0.340	0.575	0.537	0.415	0.236
3	3	85	0.196	0.137	0.012	1.000	0.567	1.000	1.000	0.810	0.472
6	3	85	0.060	0.214	0.010	0.821	0.567	0.821	0.630	0.436	0.236
2	4	80	0.375	0.063	0.013	1.000	0.400	0.792	1.000	1.000	0.667
10	4	80	0.041	0.147	0.007	0.532	0.400	0.532	0.375	0.250	0.133
6	3	80	0.060	0.214	0.010	0.772	0.533	0.772	0.593	0.410	0.222
9	5	75	0.037	0.133	0.006	0.451	0.300	0.451	0.360	0.254	0.139
5	2	75	0.082	0.294	0.013	0.997	0.750	0.997	0.703	0.469	0.250
3	5	75	0.154	0.108	0.009	0.692	0.300	0.580	0.692	0.652	0.417
5	3	75	0.064	0.227	0.010	0.770	0.500	0.770	0.634	0.455	0.250
5	4	75	0.052	0.185	0.008	0.627	0.375	0.627	0.577	0.441	0.250
7	4	75	0.047	0.168	0.008	0.569	0.375	0.569	0.459	0.326	0.179
4	5	70	0.133	0.093	0.008	0.560	0.280	0.517	0.560	0.483	0.292
7	3	70	0.056	0.202	0.009	0.638	0.467	0.638	0.462	0.311	0.167
6	3	70	0.060	0.214	0.010	0.676	0.467	0.676	0.519	0.359	0.194
4	4	70	0.147	0.103	0.009	0.618	0.350	0.618	0.618	0.500	0.292
3	3	70	0.196	0.137	0.012	0.824	0.467	0.824	0.824	0.667	0.399
4	6	65	0.122	0.085	0.007	0.476	0.217	0.412	0.476	0.433	0.271
3	3	65	0.196	0.137	0.012	0.765	0.433	0.765	0.765	0.619	0.361
5	5	65	0.118	0.082	0.007	0.459	0.260	0.459	0.459	0.371	0.217
3	5	65	0.154	0.108	0.009	0.600	0.260	0.503	0.600	0.565	0.361
3	2	60	0.098	0.352	0.016	0.955	0.600	0.955	0.818	0.600	0.333
1	6	60	1.000	0.000	0.017	1.000	0.200	0.432	0.692	1.000	1.000
5	2	60	0.082	0.294	0.013	0.797	0.600	0.797	0.563	0.375	0.200
3	6	60	0.250	0.042	0.008	0.500	0.200	0.396	0.500	0.500	0.333
1	8	60	1.000	0.000	0.017	1.000	0.150	0.327	0.545	0.857	1.000
2	5	60	0.353	0.059	0.012	0.706	0.240	0.487	0.655	0.706	0.500
2	6	55	0.333	0.056	0.011	0.611	0.183	0.379	0.532	0.611	0.456
3	7	55	0.240	0.040	0.008	0.440	0.157	0.317	0.418	0.440	0.306
3	7	50	0.240	0.040	0.008	0.400	0.143	0.288	0.380	0.400	0.278
3	10	40	0.214	0.036	0.007	0.286	0.080	0.167	0.240	0.296	0.222
2	5	40	0.353	0.059	0.012	0.471	0.160	0.325	0.436	0.471	0.333

図3.1 Excel上のデータと出力結果

3.2 分析結果

図3.1は、Excel上のデータと出力結果である。セル範囲「R2:T41」にはセル名F,「U2:W41」にはセル名W,「X2:X41」にはセル名SCORE,「Y2:BL41」にはセル名Sが定義してある。ただしセル名Sの40列のうち後で示すクラスターを表す5列のみを示す。この準備が終われば、LINGOのプログラムを実行するだけで、図3.1のように出力が得られる。

セル範囲Fの2行から41行には40人の学生が成績の降順で入っている。セル範囲Wには、40人の各学生の効率値を最大にする重みが入っている。 $W3 * \text{成績} / (W1 * \text{勉強} + W2 * \text{支出})$ で効率値が計算される。SCORE列の「X2:X41」には、各学生のDEAによる目的関数の値が入っている。学生2（勉強=9, 支出=2, 成績=100）、10（勉強=3, 支出=3, 成績=85）、12（勉強=2, 支出=4, 成績=80）、31（勉強=1, 支出=6, 成績=60）、34（勉強=1, 支出=8, 成績=60）番目の学生だけが効率値が1で効率的である。これら5人の学生が参照集合になる。これまでの分析手法では、成績が60点の学生31と34を評価することはなかった。しかし勉強時間がわずか1時間で60点を取るという点で効率的である。残りの35人の学生は、これら5人のいずれかの学生に対して非効率である。

これを知るためにセル範囲Sに、40個の重みで40人の学生の効率値を計算して、S1からS40の40列に出力する。S2は2番目の学生の重みで計算した40人の学生の効率値が「Z2:Z41」に出力されている。同じ100点を取った1番目の学生は、（勉強=12, 支出=4, 成績=100）であり2番目の学生に比べ勉強時間が3時間多く飲酒日数が2日多いので非効率と考えられ、効率値は0.5になる。1番目の学生は2番目の学生を手本とし、入力値を減らす努力をすべきという結果になる。

これは「努力をするな」という誤解を生み少し違和感をもつが、企業の部門評価では費用、従業員数、設備費用などを入力変数と考え、利益を出力と考えた場合の部門比較を考えれば納得いく。同じ利益である部門の比較をすれば、より多くの費用、従業員数、設備費用などを使用する部門は少なくするよう努力すべきという点である。具体的な実態のあるDMU同士の比較なので分かりやすい。

S1からS40の40個のクロス効率値の中で1になる学生のパターンを調べると表3.1の5個のパターンになる（この表は、図3.1でS1からS40の40列から同じ効率値パターンをもつ列を1個だけ残した。さらに40人の学生の行でSCORE=1にならないものを非表示にした）。これが今回提案する新しいクラスター分析である。クラスターS2は、2番目の学生の重みで計算した効率値でこの学生だけが1で残り39人は非効率である。このクラスターには注釈のように4番目(S4)の学生も同じクラスターなのでS4は非表示にしてある。すなわち4番目の学生の重みで計算しても2番目だけが1になる。参照集合に入っていない3番目の学生の重みで計算すると2番と10番の学生が効率値1であり、自分は非効率である。このクラスターには1,3,5-9,11,13-16,18-20,22,23,30,32番といった19人の学生の重み

で計算したのも同じになる。S10は9人、S12は8人が含まれる。31と34は参照集合に入っているが同じクラスター S31になる。

表3.1 新しいクラスター分析

SN	勉強	支出	成績	W1	W2	W3	SCORE	S2	S3	S10	S12	S31
2	9	2	100	0.000	0.500	0.010	1.000	1.000	1.000	0.577	0.357	0.185
10	3	3	85	0.196	0.137	0.012	1.000	0.567	1.000	1.000	0.810	0.472
12	2	4	80	0.375	0.063	0.013	1.000	0.400	0.792	1.000	1.000	0.667
31	1	6	60	1.000	0.000	0.017	1.000	0.200	0.432	0.692	1.000	1.000
34	1	8	60	1.000	0.000	0.017	1.000	0.150	0.327	0.545	0.857	1.000

S2 : 2,4 S3:1,3,5-9,11,13-16,18-20,22,23,30,32 S10 : 10,17,21,24-29 S12 : 12,33,35-40 S31 : 31,34

図3.2は、(成績/勉強)をX軸に、(成績/支出)をY軸にした散布図である。参照集合 = {2, 10, 12, 31, 34} を結んだものが効率的フロンティアである。支出が増えると成績が下がるためか、凸体になっていない。また表3.1で決めた5個のクラスターを丸で囲んである。

左上のクラスター S2は、成績が100点と90点の学生であり、成績は良いが勉強時間の多い努力型の学生である。次ぎのクラスター S3は、学生2 (勉強=9, 支出=2, 成績=100) と10 (勉強=3, 支出=3, 成績=85) を結ぶ効率的フロンティア直線と、原点から各学生の点を結ぶ直線との交点が仮想的な効率値1の学生である。実際には学生2と10と比較して改善を図れば良いだろう。

クラスター S10は、原点と学生10 (勉強=3, 支出=3, 成績=85) を結ぶ直線上に24 (勉強=4, 支出=4, 成績=70), 25 (勉強=3, 支出=3, 成績=70), 27 (勉強=3, 支出=3, 成績=65), 28 (勉強=3, 支出=5, 成績=65) があり、これらの4人の問題点は10番目の学生と比較すれば一目瞭然である。直線上にない17 (勉強=3, 支出=5, 成績=75), 21 (勉強=4, 支出=5, 成績=70), 26 (勉強=4, 支出=6, 成績=65), 29 (勉強=3, 支出=5, 成績=65) との問題点も明らかであるが、分析データによっては単純に10と比較しづらいこともある。

クラスター S12には、8人の学生が含まれている。

クラスター S31 (勉強=1, 支出=6, 成績=60) とS34 (勉強=1, 支出=8, 成績=60) は同じクラスターであり、勉強時間が少なく支出が多く成績の悪い学生である。しかし、勉強時間を多くすれば成績が上がる可能性があるため、大学に入って勉強に興味を失っている学生の指導に効果的であるかもしれない。

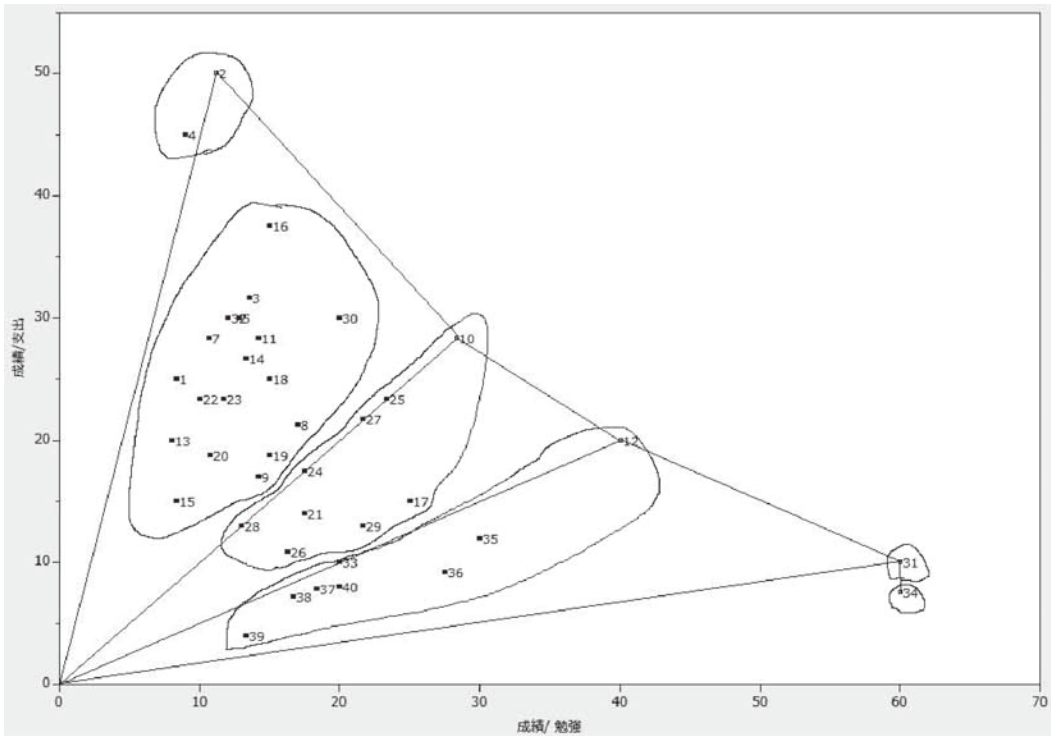


図3.2 (成績/勉強)をX軸に(成績/支出)をY軸にした散布図と効率的フロンティア

3.3 通常のクラスター分析と主成分分析

(1) 通常のクラスター分析

このデータを通常のクラスター分析で分析すると図3.3が得られる。DEAを参考にして学生を5個のクラスターに分けることにする。変数のクラスター分析の結果は解釈がしやすいが、ケースのクラスターは最終的なクラスター数の決定と解釈が難しい。これに対して、今回提案するクラスターは一意に決まるところがメリットである。

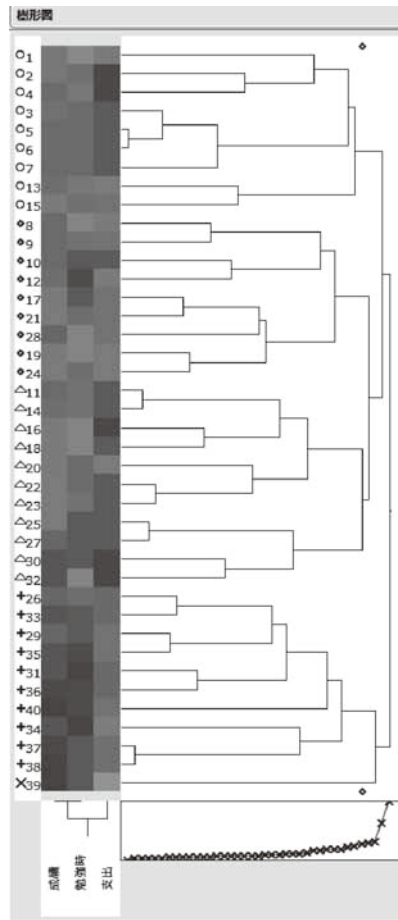


図3.3 クラスター分析

(2) クラスター分析を主成分分析で確認

クラスター分析で5個に分けた結果にマークを付け、主成分分析で因子負荷プロットとスコアプロットを描くと図3.4と図3.5が描かれる。因子負荷プロットから成績と勉強時間の値が大きい学生は第1象限で第1主成分軸より、支出の多い学生は第2象限に布置されることが分かる。

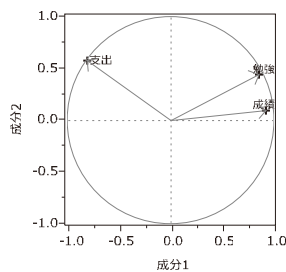


図3.4 因子負荷プロット

スコアプロットから、○の9人は成績が良く、△（11人）、菱形（9人）、+（10人）、×（1人）の順に悪くなる。以上の全体的な傾向は分かっても、個別指導の参考にはできない点
が統計手法の問題点である。

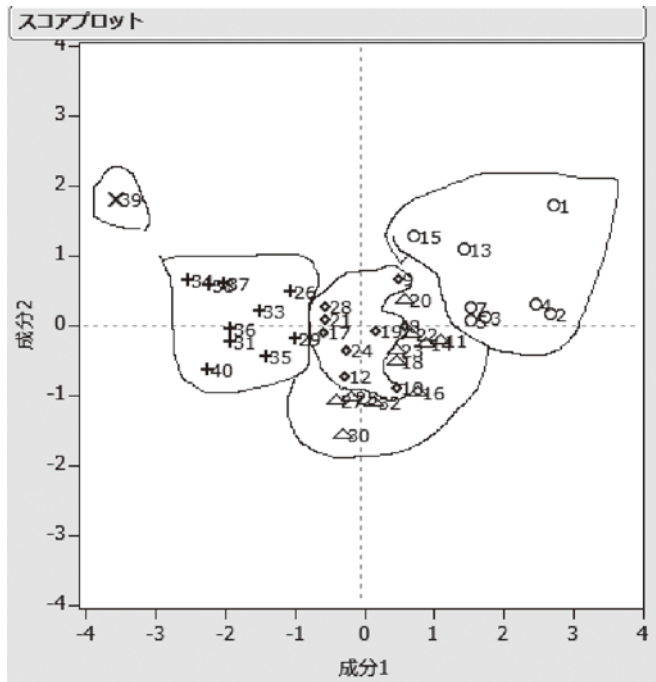


図3.5 スコアプロット

3.4 分析結果のクラスター分析と主成分分析

図3.6は40人のクロス効率値Sのクラスター分析である。変数のクラスター分析は縦線で分割したように、左からS3,S2,S10,S12,S31で表わされるクラスターに対応している。

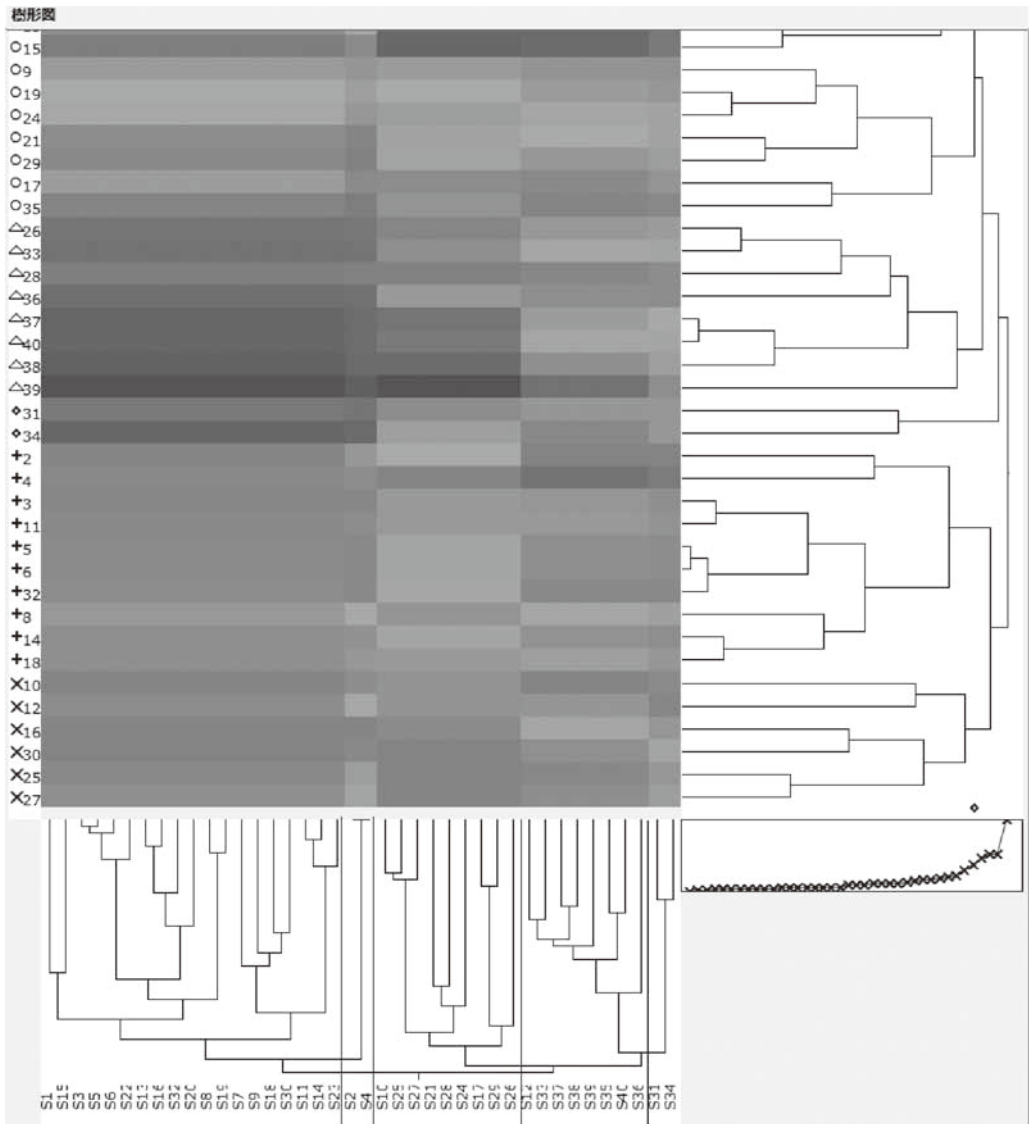


図3.6 効率値のクラスター分析

これは当然の結果であり、相関分析を行うと同じクラスターに属するS1からS40の相関係数は1になる。図3.7は固有値を表す。クラスター数の5個で累積寄与率は100になり、全ての情報が含まれている。

主成分分析: 相関係数行列から							
番号	固有値	寄与率	20	40	60	80	累積寄与率
1	25.6229	64.057					64.057
2	13.5207	33.802					97.859
3	0.7868	1.967					99.826
4	0.0646	0.162					99.988
5	0.0049	0.012					100.000
6	0.0000	0.000					100.000

図3.7 固有値

図3.8はスコアプロットと因子負荷プロットである。因子負荷プロットの1象限から4象限にかけて時計回りでS34,S12 (S38が表記されている), S10 (S24 が表記されている), S3 (S13が表記されている), S2と5個の因子負荷量が矢印で示されている。実際は同じクラスターに含まれるものが重なっている。

スコアプロットは、S2とS31が同じクラスターになっているが、他のものは変数のクラスターと厳密に対応していない。

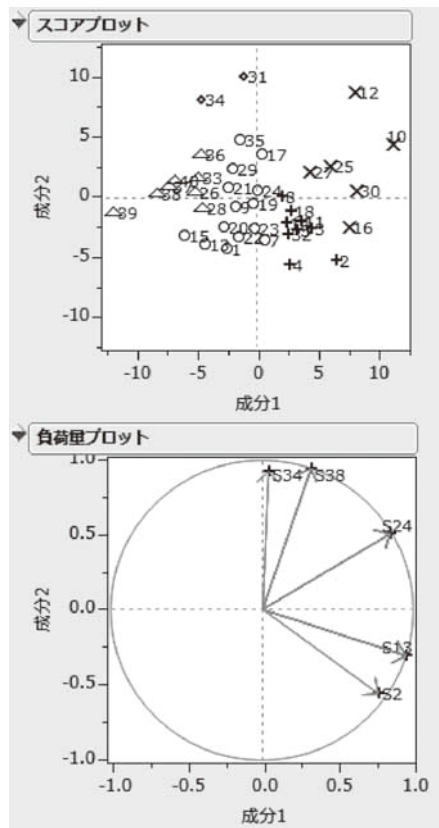


図3.8 スコアプロットと因子負荷プロット

4. 学生データ

ここでは、学生データを用いて統計分析する。図4.1に見る通り、勉強時間が増えれば成績が上がる。支出や飲酒日数が増えると成績が下がる。このようなデータは、勉強時間、支出、飲酒日数を説明変数として成績を目的変数とする重回帰分析で分析できる。

DEAは入力が出力と負の相関になると図2.1のように凸体にならず解釈がしにくい問題点がある。また飲酒日数のように0の値が表れると重みの上限をどうするかへの注意が必要になる。ここではDEAの前提に合うようにデータ値を変換しないで、本データを用いてDEAで注意すべき事項を含め検討する。

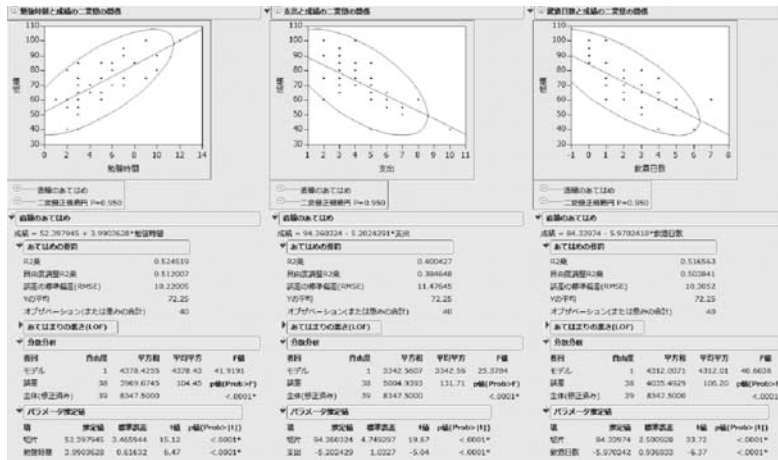


図4.1 学生データの散布図

4.1 1入力1出力

(1) 効率的とは

図4.2の散布図のような、勉強時間を入力とし成績を出力と考える1入力1出力の問題を考える。この時、原点と勉強時間=1, 成績=60点の学生を結んだ線が一番効率 (60点/時間) 的な境界である。すなわち12時間勉強し100点取った学生 (100/12=8.3点/時間) は、60*12=720点取らなければ非効率と考えている。すなわち、1時間当たり60点の結果が得られると考えている点に注意がいる。

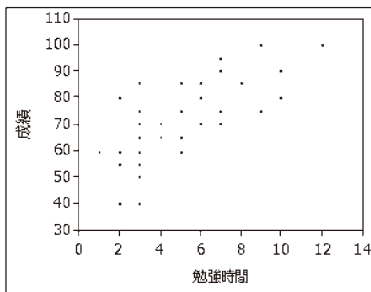


図4.2 散勉強時間を入力とし成績を出力と考える1入力1出力の散布図

このデータに対してLINGOのプログラムを実行すると、表4.1の出力が得られる。1列目が40人の学生の入力値である勉強時間で、2列目が出力の成績得点である。3列と4列は40人の学生に最適な重みである。これで40人の学生の目的関数の値を計算したものがSCOREである。S1列は、1番目の学生の重みを40人の学生に適用した値である。12時間勉強して100点を取った学生は、自分に最適な重みであるにかかわらず0.139であり、1時間勉強して60点の学生が最も効率的な1と評価されている。S1からS40まで重みを変えて計算しても同じ結果になるのでS3からS40を省略する。Ratio = $W2/W1=1/60$ で一定である。

表4.1 1入力1出力の結果

勉強	成績	W1	W2	Ratio	SCORE	S1	S2
12	100	0.0833333	0.0013889	0.017	0.139	0.139	0.139
9	100	0.1111111	0.0018519	0.017	0.185	0.185	0.185
7	95	0.1428571	0.0023810	0.017	0.226	0.226	0.226
10	90	0.1000000	0.0016667	0.017	0.150	0.150	0.150
7	90	0.1428571	0.0023810	0.017	0.214	0.214	0.214
7	90	0.1428571	0.0023810	0.017	0.214	0.214	0.214
8	85	0.1250000	0.0020833	0.017	0.177	0.177	0.177
5	85	0.2000000	0.0033333	0.017	0.283	0.283	0.283
6	85	0.1666667	0.0027778	0.017	0.236	0.236	0.236
3	85	0.3333333	0.0055556	0.017	0.472	0.472	0.472
6	85	0.1666667	0.0027778	0.017	0.236	0.236	0.236
2	80	0.5000000	0.0083333	0.017	0.667	0.667	0.667
10	80	0.1000000	0.0016667	0.017	0.133	0.133	0.133
6	80	0.1666667	0.0027778	0.017	0.222	0.222	0.222
9	75	0.1111111	0.0018519	0.017	0.139	0.139	0.139
5	75	0.2000000	0.0033333	0.017	0.250	0.250	0.250
3	75	0.3333333	0.0055556	0.017	0.417	0.417	0.417
5	75	0.2000000	0.0033333	0.017	0.250	0.250	0.250
5	75	0.2000000	0.0033333	0.017	0.250	0.250	0.250
7	75	0.1428571	0.0023810	0.017	0.179	0.179	0.179
4	70	0.2500000	0.0041667	0.017	0.292	0.292	0.292
7	70	0.1428571	0.0023810	0.017	0.167	0.167	0.167
6	70	0.1666667	0.0027778	0.017	0.194	0.194	0.194
4	70	0.2500000	0.0041667	0.017	0.292	0.292	0.292
3	70	0.3333333	0.0055556	0.017	0.389	0.389	0.389
4	65	0.2500000	0.0041667	0.017	0.271	0.271	0.271
3	65	0.3333333	0.0055556	0.017	0.361	0.361	0.361
5	65	0.2000000	0.0033333	0.017	0.217	0.217	0.217
3	65	0.3333333	0.0055556	0.017	0.361	0.361	0.361

3	60	0.3333333	0.0055556	0.017	0.333	0.333	0.333
1	60	1.0000000	0.0166667	0.017	1.000	1.000	1.000
5	60	0.2000000	0.0033333	0.017	0.200	0.200	0.200
3	60	0.3333333	0.0055556	0.017	0.333	0.333	0.333
1	60	1.0000000	0.0166667	0.017	1.000	1.000	1.000
2	60	0.5000000	0.0083333	0.017	0.500	0.500	0.500
2	55	0.5000000	0.0083333	0.017	0.458	0.458	0.458
3	55	0.3333333	0.0055556	0.017	0.306	0.306	0.306
3	50	0.3333333	0.0055556	0.017	0.278	0.278	0.278
3	40	0.3333333	0.0055556	0.017	0.222	0.222	0.222
2	40	0.5000000	0.0083333	0.017	0.333	0.333	0.333

(2) 効率的なDMUを省くと？

次に勉強時間1時間で60点取った効率的な学生2名を省いた38名を考えると、表4.2のように2時間勉強して80点 ($80/2=40$ 点/時間, $\text{Ratio}=2/80=0.025$) の学生が最も効率的である。すなわち、効率的なDMUの存在によって結果が大きく異なってくる点に注意がいる。

表4.2 効率的な学生2名を省いた38名

勉強	成績	W1	W2	Ratio	SCORE	S1	S2
12	100	0.083	0.002	0.025	0.208	0.208	0.208
9	100	0.111	0.003	0.025	0.278	0.278	0.278
7	95	0.143	0.004	0.025	0.339	0.339	0.339
10	90	0.100	0.003	0.025	0.225	0.225	0.225
7	90	0.143	0.004	0.025	0.321	0.321	0.321
7	90	0.143	0.004	0.025	0.321	0.321	0.321
8	85	0.125	0.003	0.025	0.266	0.266	0.266
5	85	0.200	0.005	0.025	0.425	0.425	0.425
6	85	0.167	0.004	0.025	0.354	0.354	0.354
3	85	0.333	0.008	0.025	0.708	0.708	0.708
6	85	0.167	0.004	0.025	0.354	0.354	0.354
2	80	0.500	0.013	0.025	1.000	1.000	1.000
10	80	0.100	0.003	0.025	0.200	0.200	0.200
6	80	0.167	0.004	0.025	0.333	0.333	0.333
9	75	0.111	0.003	0.025	0.208	0.208	0.208
5	75	0.200	0.005	0.025	0.375	0.375	0.375
3	75	0.333	0.008	0.025	0.625	0.625	0.625
5	75	0.200	0.005	0.025	0.375	0.375	0.375
5	75	0.200	0.005	0.025	0.375	0.375	0.375

7	75	0.143	0.004	0.025	0.268	0.268	0.268
4	70	0.250	0.006	0.025	0.438	0.438	0.438
7	70	0.143	0.004	0.025	0.250	0.250	0.250
6	70	0.167	0.004	0.025	0.292	0.292	0.292
4	70	0.250	0.006	0.025	0.438	0.438	0.438
3	70	0.333	0.008	0.025	0.583	0.583	0.583
4	65	0.250	0.006	0.025	0.406	0.406	0.406
3	65	0.333	0.008	0.025	0.542	0.542	0.542
5	65	0.200	0.005	0.025	0.325	0.325	0.325
3	65	0.333	0.008	0.025	0.542	0.542	0.542
3	60	0.333	0.008	0.025	0.500	0.500	0.500
5	60	0.200	0.005	0.025	0.300	0.300	0.300
3	60	0.333	0.008	0.025	0.500	0.500	0.500
2	60	0.500	0.013	0.025	0.750	0.750	0.750
2	55	0.500	0.013	0.025	0.688	0.688	0.688
3	55	0.333	0.008	0.025	0.458	0.458	0.458
3	50	0.333	0.008	0.025	0.417	0.417	0.417
3	40	0.333	0.008	0.025	0.333	0.333	0.333
2	40	0.500	0.013	0.025	0.500	0.500	0.500

4.2 2入力1出力の場合

勉強時間と飲酒日数を入力とし、成績を出力とした分析を行う。

4.2.1 飲酒日数に0がある場合

DEAの弱点は、入出力のデータが正の値であり、0が含まれていると重みの上限に問題が生じる。例えば入力変数の飲酒日数は、1週間に何日飲んでいるかという変数であるので、0が出てくる。この場合、この変数の重みを大きくしても、本人の重みは0であり効率的になる。しかし、他の学生は値が0でなければ分母は大きくなり、SCOREは小さくなり非効率になる。

SCOREをみると、10から12番までと、31番と34番の5人の学生が効率的である。他の学生はこの5人に対して非効率になる。

S1は1番目の学生に最適な重みを40人の学生に適用したものであり、10番目と11番目の85点を取った学生のSCOREが1になり最も効率的である。1番目の学生は100点を取っているが10番目の学生に比べて勉強時間が4倍多いので改善が望まれる。

表4.3 勉強時間と飲酒日数を入力とし、成績を出力とした分析

勉強	飲酒	成績	W1	W2	W3	SCORE	S1	S10	S11	S12	S31
12	1	100	0.07	0.20	0.0047	0.47	0.47	0.35	0.00	0.24	0.14
9	0	100	0.11	99999	0.0078	0.78	0.78	0.49	0.78	0.32	0.19
7	0	95	0.14	99999	0.0101	0.96	0.96	0.60	0.96	0.39	0.23
10	0	90	0.10	99999	0.0071	0.64	0.64	0.40	0.64	0.26	0.15
7	0	90	0.14	99999	0.0101	0.91	0.91	0.57	0.91	0.37	0.21
7	0	90	0.14	99999	0.0101	0.91	0.91	0.57	0.91	0.37	0.21
8	0	85	0.13	99999	0.0088	0.75	0.75	0.47	0.75	0.30	0.18
5	1	85	0.13	0.38	0.0088	0.75	0.75	0.65	0.00	0.47	0.28
6	1	85	0.11	0.33	0.0078	0.67	0.67	0.56	0.00	0.40	0.24
3	1	85	0.26	0.21	0.0118	1.00	1.00	1.00	0.00	0.77	0.47
6	0	85	0.17	99999	0.0118	1.00	1.00	0.63	1.00	0.40	0.24
2	2	80	0.44	0.06	0.0125	1.00	0.71	1.00	0.00	1.00	0.67
10	3	80	0.05	0.16	0.0037	0.30	0.30	0.29	0.00	0.22	0.13
6	2	80	0.13	0.10	0.0059	0.47	0.47	0.47	0.00	0.36	0.22
9	1	75	0.08	0.25	0.0059	0.44	0.44	0.34	0.00	0.23	0.14
5	1	75	0.13	0.38	0.0088	0.66	0.66	0.58	0.00	0.42	0.25
3	1	75	0.26	0.21	0.0118	0.88	0.88	0.88	0.00	0.68	0.42
5	2	75	0.15	0.12	0.0068	0.51	0.48	0.51	0.00	0.41	0.25
5	4	75	0.12	0.10	0.0055	0.41	0.31	0.41	0.00	0.38	0.25
7	1	75	0.10	0.30	0.0071	0.53	0.53	0.43	0.00	0.30	0.18
4	1	70	0.14	0.43	0.0101	0.71	0.71	0.65	0.00	0.48	0.29
7	1	70	0.10	0.30	0.0071	0.49	0.49	0.40	0.00	0.28	0.17
6	2	70	0.13	0.10	0.0059	0.41	0.41	0.41	0.00	0.32	0.19
4	1	70	0.14	0.43	0.0101	0.71	0.71	0.65	0.00	0.48	0.29
3	3	70	0.29	0.04	0.0083	0.58	0.41	0.58	0.00	0.58	0.39
4	2	65	0.18	0.14	0.0080	0.52	0.46	0.52	0.00	0.43	0.27
3	2	65	0.22	0.17	0.0098	0.63	0.51	0.63	0.00	0.57	0.36
5	3	65	0.14	0.11	0.0061	0.39	0.33	0.39	0.00	0.34	0.22
3	4	65	0.28	0.04	0.0080	0.52	0.31	0.47	0.00	0.52	0.36
3	1	60	0.26	0.21	0.0118	0.71	0.71	0.71	0.00	0.55	0.33
1	5	60	1.00	0.00	0.0167	1.00	0.26	0.55	0.00	1.00	1.00
5	1	60	0.13	0.38	0.0088	0.53	0.53	0.46	0.00	0.33	0.20
3	2	60	0.22	0.17	0.0098	0.59	0.47	0.59	0.00	0.52	0.33
1	7	60	1.00	0.00	0.0167	1.00	0.19	0.41	0.00	0.86	1.00
2	4	60	0.39	0.06	0.0111	0.67	0.30	0.52	0.00	0.67	0.50
2	3	55	0.41	0.06	0.0118	0.65	0.35	0.56	0.00	0.65	0.46
3	5	55	0.27	0.04	0.0077	0.42	0.22	0.35	0.00	0.42	0.31
3	3	50	0.29	0.04	0.0083	0.42	0.29	0.42	0.00	0.42	0.28
3	6	40	0.26	0.04	0.0074	0.30	0.13	0.23	0.00	0.30	0.22
2	4	40	0.39	0.06	0.0111	0.44	0.20	0.35	0.00	0.44	0.33

注：S1：1,8,9,13,15,16,20-22,24,32が含まれる。S11：2-7,11が含まれる。S10：10,14,17-19,23,26-28,30,33が含まれる。S12：12,25,29,35-40が含まれる。S31：31,34が含まれる。

S1のように10番 (3,1,85) と11番 (6,0,85) の学生を一番効率的にする重みは, 8 (5,1,85), 9 (6,1,85), 13 (10,3,80), 15 (9,1,75), 16 (5,1,75), 20 (9,1,75), 21 (4,1,70), 22 (3,1,75), 24 (5,4,75), 32 (5,1,60) 番の学生の重みである。これらは, 10番と11番の学生に対し非効率な学生群で, 10番と11番をお手本に頑張るべきである。

S11の11番目の学生のSCOREは1である。S11は, 11番だけが効率値1で, このクラスターには2,3,4,5,6,7,11が含まれる。これらの学生の飲酒日数は0日であり, 勉強時間が6時間の学生の成績は85点と効率的であり, 4番目の10時間の学生の90点より効果的である。勉強時間を10時間から $(6/85) * 5 + 6 = 6$ 時間21分に短縮すれば効率的になる。あるいは $6/85 = x/90$ を解くと $x = 6.35$ になる。

S10 (3,1,85) には14 (6,2,80), 17 (3,1,75), 18 (5,2,75), 19 (5,4,75), 23 (6,2,70), 26 (4,2,65), 27 (3,2,65), 28 (5,3,65), 30 (3,1,60), 33 (5,3,65) が含まれる。10と12番目の学生が効率的である。

S12 (2,2,80) には25 (3,3,70), 29 (3,4,65), 35 (2,4,60), 36 (2,3,55), 37 (3,5,55), 38 (3,3,50), 39 (3,6,40), 40 (2,4,40) が含まれる。12番と31番目の学生が効率的である。

S31 (1,5,60) と34 (1,7,60) は効率的であるが, 34の方が非効率に見える。飲酒日数の重みが両方とも0のため, この2名の学生では飲酒日数が評価されていないので, このような問題点が表れる。重みの下限値を0でなく恣意的に0.00005等に設定すると, 実行可能解がなくなることもあるので, 0にして出力結果の解釈に注意を払えば良いであろう。

このように, 学生全体を効率値でクラスター化すると図4.3のクラスターが得られる。すなわち, 左上から右下にS31,S12,S10,S1,S11のクラスターを表す。

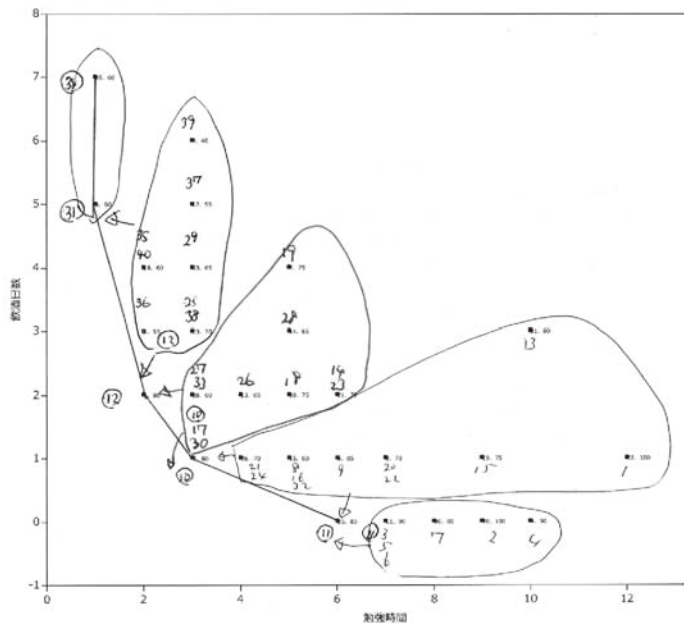


図4.3 2入力1出力

4.2.2 重みの最大値を1に変更

飲酒日数が0の重みは最大値999999になるので、他の重みを参考に1に変更する。変更によっても表4.4のように今回提案する表4.3のクラスター化には影響されないことが分かる。

表4.4 重みを1に変更

勉強	飲酒	成績	W1	W2	W3	SCORE	S1	S10	S11	S12	S31
12	1	100	0.07	0.20	0.00	0.47	0.47	0.35	0.39	0.24	0.14
9	0	100	0.11	1	0.01	0.78	0.78	0.49	0.78	0.32	0.19
7	0	95	0.14	1	0.01	0.96	0.96	0.60	0.96	0.39	0.23
10	0	90	0.10	1	0.01	0.64	0.64	0.40	0.64	0.26	0.15
7	0	90	0.14	1	0.01	0.91	0.91	0.57	0.91	0.37	0.21
7	0	90	0.14	1	0.01	0.91	0.91	0.57	0.91	0.37	0.21
8	0	85	0.13	1	0.01	0.75	0.75	0.47	0.75	0.30	0.18
5	1	85	0.13	0.38	0.01	0.75	0.75	0.65	0.55	0.47	0.28
6	1	85	0.11	0.33	0.01	0.67	0.67	0.56	0.50	0.40	0.24
3	1	85	0.26	0.21	0.01	1.00	1.00	1.00	0.67	0.77	0.47
6	0	85	0.17	1	0.01	1.00	1.00	0.63	1.00	0.40	0.24
2	2	80	0.44	0.06	0.01	1.00	0.71	1.00	0.40	1.00	0.67
10	3	80	0.05	0.16	0.00	0.30	0.30	0.29	0.20	0.22	0.13
6	2	80	0.13	0.10	0.01	0.47	0.47	0.47	0.31	0.36	0.22
9	1	75	0.08	0.25	0.01	0.44	0.44	0.34	0.35	0.23	0.14
5	1	75	0.13	0.38	0.01	0.66	0.66	0.58	0.48	0.42	0.25
3	1	75	0.26	0.21	0.01	0.88	0.88	0.88	0.59	0.68	0.42
5	2	75	0.15	0.12	0.01	0.51	0.48	0.51	0.31	0.41	0.25
5	4	75	0.12	0.10	0.01	0.41	0.31	0.41	0.18	0.38	0.25
7	1	75	0.10	0.30	0.01	0.53	0.53	0.43	0.41	0.30	0.18
4	1	70	0.14	0.43	0.01	0.71	0.71	0.65	0.49	0.48	0.29
7	1	70	0.10	0.30	0.01	0.49	0.49	0.40	0.38	0.28	0.17
6	2	70	0.13	0.10	0.01	0.41	0.41	0.41	0.27	0.32	0.19
4	1	70	0.14	0.43	0.01	0.71	0.71	0.65	0.49	0.48	0.29
3	3	70	0.29	0.04	0.01	0.58	0.41	0.58	0.24	0.58	0.39
4	2	65	0.18	0.14	0.01	0.52	0.46	0.52	0.29	0.43	0.27
3	2	65	0.22	0.17	0.01	0.63	0.51	0.63	0.31	0.57	0.36
5	3	65	0.14	0.11	0.01	0.39	0.33	0.39	0.20	0.34	0.22
3	4	65	0.28	0.04	0.01	0.52	0.31	0.47	0.17	0.52	0.36
3	1	60	0.26	0.21	0.01	0.71	0.71	0.71	0.47	0.55	0.33
1	5	60	1.00	0.00	0.02	1.00	0.26	0.55	0.14	1.00	1.00
5	1	60	0.13	0.38	0.01	0.53	0.53	0.46	0.39	0.33	0.20
3	2	60	0.22	0.17	0.01	0.59	0.47	0.59	0.28	0.52	0.33
1	7	60	1.00	0.00	0.02	1.00	0.19	0.41	0.10	0.86	1.00
2	4	60	0.39	0.06	0.01	0.67	0.30	0.52	0.16	0.67	0.50
2	3	55	0.41	0.06	0.01	0.65	0.35	0.56	0.19	0.65	0.46
3	5	55	0.27	0.04	0.01	0.42	0.22	0.35	0.12	0.42	0.31
3	3	50	0.29	0.04	0.01	0.42	0.29	0.42	0.17	0.42	0.28
3	6	40	0.26	0.04	0.01	0.30	0.13	0.23	0.07	0.30	0.22
2	4	40	0.39	0.06	0.01	0.44	0.20	0.35	0.11	0.44	0.33

4.2.3 DEAの一般的な表示法

図2.1のように成績／勉強と成績／飲酒に変更して散布図を描くと図4.4のグラフが描ける。ただし、飲酒に0があるのでこれを仮に0.5に置き換える。そして、31, 12, 10, 11を結ぶ効率的フロンティアを考えて、原点とこれらの4人の学生を結ぶ直線を描く。

学生10 (3,1,85) と原点を結ぶ直線の傾きは3で、17 (3,1,75), 30 (3,1,60), 14 (6,2,80), 23 (6,2,70) も傾きは3である。17と30は10と同じ入力値でありながら出力の成績が悪い。14と23は入力を2倍使っているが成績は悪い。

原点と26 (4,2,65) は傾きが2である。学生10 (3,1,85) と12 (2,2,80) を結ぶ直線は (成績／飲酒) = $-27/7 * (成績／勉強) + 314/7$ であり、交点は $(40 * 34 / 41, 80 * 34 / 41)$ になる。もし、65点の学生が勉強時間 = $65 * 41 / (40 * 34) = 1.96$ で、飲酒日数 = $65 * 41 / (80 * 34) = 0.98$ であればこの交点になり効率的になる。26番の学生は勉強時間と飲酒日数をほぼ半減して65点を維持すれば効率的になる。

11を含めて左上にいる6人は、飲酒日数が0であるのを0.5としたためその位置関係はいい加減である。0でなければ、効率的な11と原点を結ぶ直線状に残りの5人がプロットされる。

以上から無理に凸体にするために出力／入力という変換を行わず、図4.3のように入力変数そのもので散布図を描いても良いのではと考える。

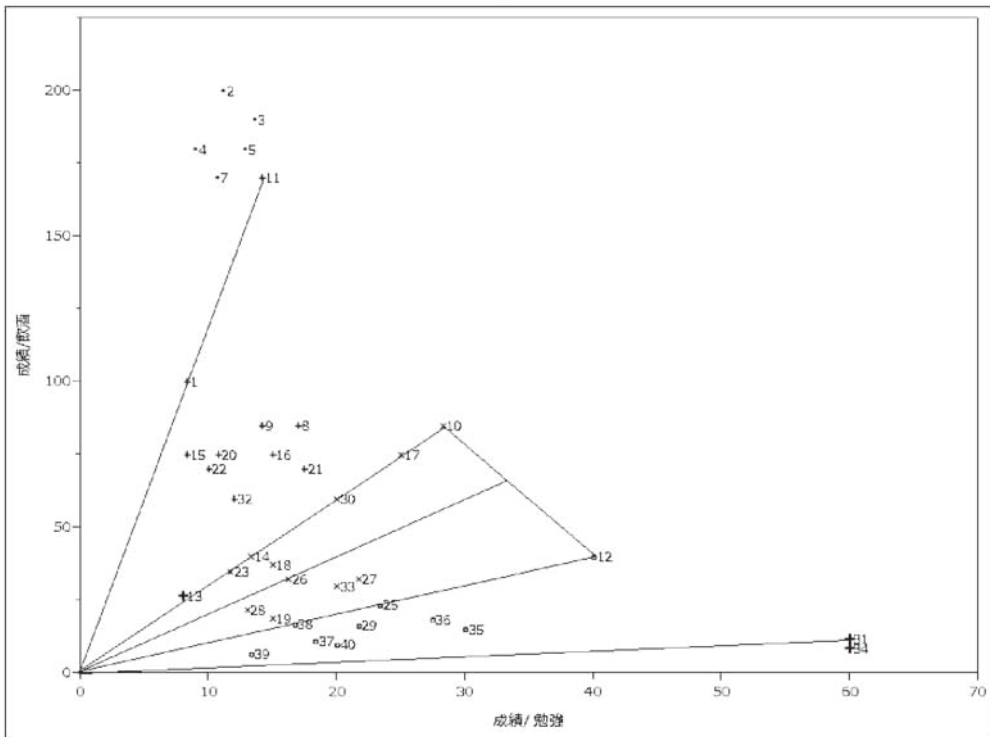


図4.4 図1.1の表記

4.2.4 クラスター分析と主成分分析による検討

図4.5は効率値配列Sのクラスター分析の結果である。変数のクラスターは、左からS1, S11, S10, S12, S31の5個のクラスターに対応している。



図4.5 クラスター分析

S1からS40の40個の変数の相関係数は、同じクラスターの変数は1になる。

主成分分析を行うと、第5主成分で累積寄与率は100%になった。図4.6は因子負荷量とスコアプロットである。因子負荷量は第1象限から第4象限に時計回りで、S31 (S34と表示), S32 (S39と表示), S10 (S28と表示), S1 (S16と表示), S11の5個のクラスターに対応している。スコアプロットは、図4.5のクラスター分析でケースを5個のクラスターに分けてマークの違いで表わしているが、S31以外は変数のクラスターと一致していない。

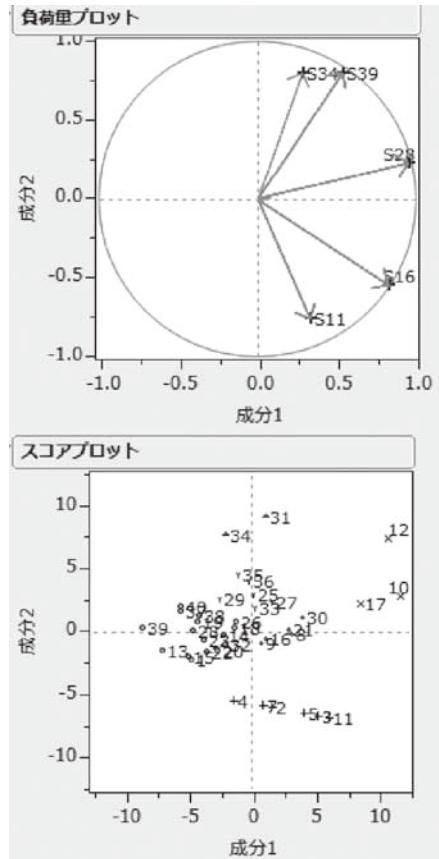


図4.6 因子負荷量とスコアプロット

4.3 4変数の場合

入力として勉強時間, 飲酒日数, 支出の3入力を用いる。

4.3.1 重みの上限99999

重みの上限を99999に設定して解くと表4.5の出力になる。参照集合は8人の学生 {2,3, 10,11,12, 16,31,34} である。クラスター数は10個になった。参照集合の31が12と同じクラスターになった。また, S7,S15, S25に代表されるクラスターは参照集合を含まないクラスターである。すなわち, 全て非効率なDMUから構成されているが, 参照集合の中の2個か3個が効率値が1になっている。

表4.5 重みの上限を99999に設定

Hour	Spence	Drink	Score	W1	W2	W3	W4	SCORE	S2	S3	S7	S10	S11	S12	S15	S16	S25	S34
12	4	1	100	0.038	0.136	0.000	0.006	0.61	0.00	0.00	0.00	0.47	0.00	0.24	0.58	0.61	0.41	0.14
9	2	0	100	0.000	0.500	999999	0.010	1.00	1.00	0.94	1.00	0.78	0.78	0.32	1.00	1.00	0.58	0.19
7	3	0	95	0.105	0.088	999999	0.011	1.00	0.63	1.00	1.00	0.96	0.96	0.39	1.00	0.87	0.63	0.23
10	2	0	90	0.000	0.500	999999	0.010	0.90	0.90	0.77	0.83	0.64	0.64	0.26	0.83	0.85	0.47	0.15
7	3	0	90	0.105	0.088	999999	0.011	0.95	0.60	0.95	0.95	0.91	0.91	0.37	0.95	0.82	0.59	0.21
7	3	0	90	0.105	0.088	999999	0.011	0.95	0.60	0.95	0.95	0.91	0.91	0.37	0.95	0.82	0.59	0.21
8	3	0	85	0.082	0.115	999999	0.010	0.82	0.57	0.81	0.82	0.75	0.75	0.30	0.82	0.73	0.50	0.18
5	4	1	85	0.125	0.000	0.375	0.009	0.75	0.00	0.00	0.00	0.75	0.00	0.47	0.75	0.71	0.65	0.28
6	5	1	85	0.111	0.000	0.333	0.008	0.67	0.00	0.00	0.00	0.67	0.00	0.40	0.63	0.57	0.54	0.24
3	3	1	85	0.167	0.000	0.500	0.012	1.00	0.00	0.00	0.00	1.00	0.00	0.77	1.00	1.00	1.00	0.47
6	3	0	85	0.167	0.000	999999	0.012	1.00	0.57	1.00	0.98	1.00	1.00	0.40	0.98	0.82	0.63	0.24
2	4	2	80	0.437	0.000	0.062	0.013	1.00	0.00	0.00	0.00	0.71	0.00	1.00	0.71	0.79	1.00	0.67
10	4	3	80	0.041	0.147	0.000	0.007	0.53	0.00	0.00	0.00	0.30	0.00	0.22	0.39	0.53	0.37	0.13
6	3	2	80	0.060	0.214	0.000	0.010	0.77	0.00	0.00	0.00	0.47	0.00	0.36	0.60	0.77	0.59	0.22
9	5	1	75	0.053	0.075	0.149	0.006	0.47	0.00	0.00	0.00	0.44	0.00	0.23	0.47	0.45	0.36	0.14
5	2	1	75	0.082	0.294	0.000	0.013	1.00	0.00	0.00	0.00	0.66	0.00	0.42	0.83	1.00	0.70	0.25
3	5	1	75	0.167	0.000	0.500	0.012	0.88	0.00	0.00	0.00	0.88	0.00	0.68	0.69	0.58	0.69	0.42
5	3	2	75	0.064	0.227	0.000	0.010	0.77	0.00	0.00	0.00	0.48	0.00	0.41	0.60	0.77	0.63	0.25
5	4	4	75	0.052	0.185	0.000	0.008	0.63	0.00	0.00	0.00	0.31	0.00	0.38	0.40	0.63	0.58	0.25
7	4	1	75	0.065	0.091	0.182	0.008	0.57	0.00	0.00	0.00	0.53	0.00	0.30	0.57	0.57	0.46	0.18
4	5	1	70	0.143	0.000	0.428	0.010	0.71	0.00	0.00	0.00	0.71	0.00	0.48	0.60	0.52	0.56	0.29
7	3	1	70	0.056	0.202	0.000	0.009	0.64	0.00	0.00	0.00	0.49	0.00	0.28	0.59	0.64	0.46	0.17
6	3	2	70	0.060	0.214	0.000	0.010	0.68	0.00	0.00	0.00	0.41	0.00	0.32	0.52	0.68	0.52	0.19
4	4	1	70	0.143	0.000	0.428	0.010	0.71	0.00	0.00	0.00	0.71	0.00	0.48	0.66	0.62	0.62	0.29
3	3	3	70	0.196	0.137	0.000	0.012	0.82	0.00	0.00	0.00	0.41	0.00	0.58	0.53	0.82	0.82	0.39
4	6	2	65	0.180	0.000	0.140	0.008	0.52	0.00	0.00	0.00	0.46	0.00	0.43	0.42	0.41	0.48	0.27
3	3	2	65	0.196	0.137	0.000	0.012	0.76	0.00	0.00	0.00	0.51	0.00	0.57	0.60	0.76	0.76	0.36
5	5	3	65	0.118	0.082	0.000	0.007	0.46	0.00	0.00	0.00	0.33	0.00	0.34	0.37	0.46	0.46	0.22
3	5	4	65	0.154	0.108	0.000	0.009	0.60	0.00	0.00	0.00	0.31	0.00	0.52	0.36	0.50	0.60	0.36
3	2	1	60	0.098	0.352	0.000	0.016	0.95	0.00	0.00	0.00	0.71	0.00	0.55	0.82	0.95	0.82	0.33
1	6	5	60	0.583	0.000	0.083	0.017	1.00	0.00	0.00	0.00	0.26	0.00	1.00	0.30	0.43	0.69	1.00
5	2	1	60	0.082	0.294	0.000	0.013	0.80	0.00	0.00	0.00	0.53	0.00	0.33	0.67	0.80	0.56	0.20
3	6	2	60	0.219	0.000	0.171	0.010	0.59	0.00	0.00	0.00	0.47	0.00	0.52	0.41	0.40	0.50	0.33
1	8	7	60	0.999	0.000	0.000	0.017	1.00	0.00	0.00	0.00	0.19	0.00	0.86	0.22	0.33	0.55	1.00
2	5	4	60	0.353	0.059	0.000	0.012	0.71	0.00	0.00	0.00	0.30	0.00	0.67	0.35	0.49	0.65	0.50
2	6	3	55	0.412	0.000	0.059	0.012	0.65	0.00	0.00	0.00	0.35	0.00	0.65	0.34	0.38	0.53	0.46
3	7	5	55	0.240	0.040	0.000	0.008	0.44	0.00	0.00	0.00	0.22	0.00	0.42	0.24	0.32	0.42	0.31
3	7	3	50	0.292	0.000	0.042	0.008	0.42	0.00	0.00	0.00	0.29	0.00	0.42	0.28	0.29	0.38	0.28
3	10	6	40	0.259	0.000	0.037	0.007	0.30	0.00	0.00	0.00	0.13	0.00	0.30	0.14	0.17	0.24	0.22
2	5	4	40	0.353	0.059	0.000	0.012	0.47	0.00	0.00	0.00	0.20	0.00	0.44	0.23	0.32	0.44	0.33

表4.6は、表4.5の10個のクラスター関係をまとめたものである。

表4.6 3入力1出力のクラスター関係

	Score	人数	クラスター
S2	1	2	2,4
S3	1	3	3,5,6
S7	0.82	1	7
S10	1	6	10,8,9,17,21,24
S11	1	1	11
S12	1	8	12,31,36,38,39
S15	0.47	2	15,20
S16	1	10	1,13,14,16,18,19,22,23,30,32
S25	0.82	6	25,26,27,28,29,33
S34	1	1	34

図4.7は、3個の入力のうち、勉強時間と飲酒日数の平面でクラスター関係を表示したものである。3次元を2次元に投影しているので、複雑なクラスター表現になっている。効率フロンティアも凸体でなくS10のところにくぼんでいるが、これはS10が支出の軸でフロンティアになっているためと考えられる。10個のクラスターは平面上では複雑に入り組んでいる。

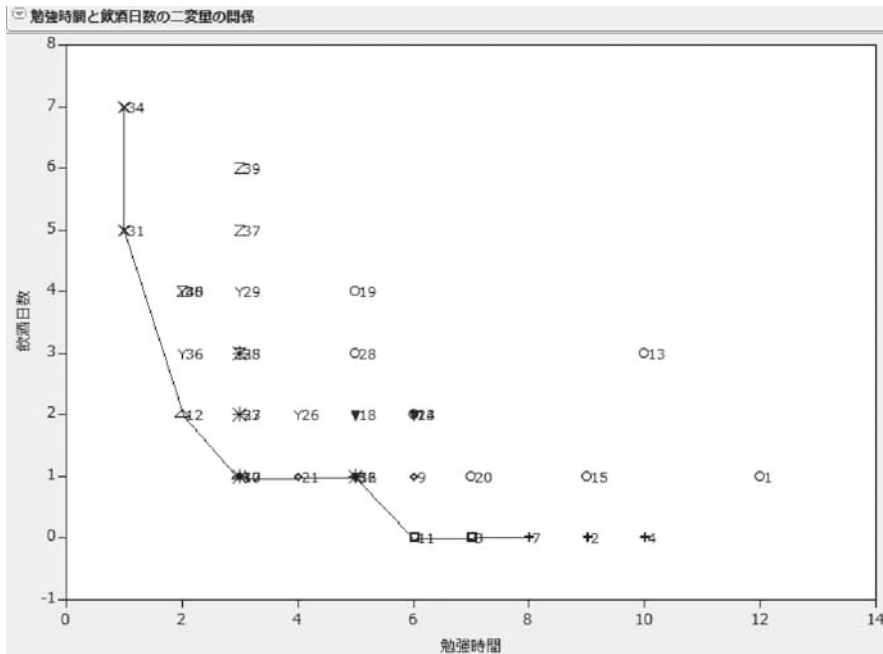


図4.7 3入力1出力のクラスター関係

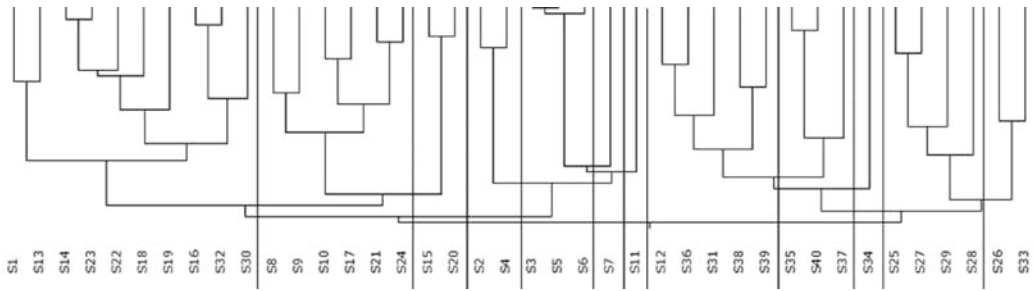


図4.8 効率値配列Sをクラスター分析した変数のクラスター

図4.8はクロス効率値Sをクラスター分析した変数のクラスターである。左から順に S16, S10, S15, S2, S3, S7, S11, S12, S35, S34, S25, S26ときれいにわかれている。

同じクラスターに属する相関係数は1である。図4.9は主成分分析の因子負荷量とスコアプロットである。累積寄与率は第10主成分で100%になった。因子負荷プロットは重なっているがほぼ10個の矢印が見える。第1と第2主成分ではほぼ半径が1のものとして1以下のものがある。スコアプロットは、きれいにクラスターに分かれていない。

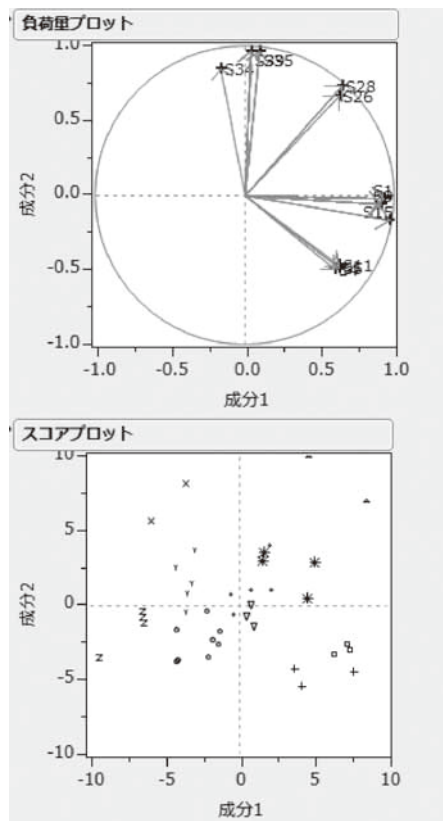


図4.9 主成分分析の因子負荷量とスコアプロット

4.3.2 重みの上限を1に設定

表4.7は重みの上限を1に設定したものである。クラスター関係は変わっていない。

表4.7 重みの上限を1に設定したもの

SN	Hour	Spence	Drink	Score	W1	W2	W3	W4	SCORE	S2	S3	S7	S10	S11	S12	S15	S16	S25	S34
2	9	2	0	100	0.000	0.500	999999	0.010	1.00	1.00	0.94	1.00	0.78	0.78	0.32	1.00	1.00	0.58	0.19
3	7	3	0	95	0.105	0.088	999999	0.011	1.00	0.63	1.00	1.00	0.96	0.96	0.39	1.00	0.87	0.63	0.23
7	8	3	0	85	0.082	0.115	999999	0.010	0.82	0.57	0.81	0.82	0.75	0.75	0.30	0.82	0.73	0.50	0.18
10	3	3	1	85	0.167	0.000	0.500	0.012	1.00	0.00	0.00	0.00	1.00	0.00	0.77	1.00	1.00	1.00	0.47
11	6	3	0	85	0.167	0.000	999999	0.012	1.00	0.57	1.00	0.98	1.00	1.00	0.40	0.98	0.82	0.63	0.24
12	2	4	2	80	0.437	0.000	0.062	0.013	1.00	0.00	0.00	0.00	0.71	0.00	1.00	0.71	0.79	1.00	0.67
15	9	5	1	75	0.053	0.075	0.149	0.006	0.47	0.00	0.00	0.00	0.44	0.00	0.23	0.47	0.45	0.36	0.14
16	5	2	1	75	0.082	0.294	0.000	0.013	1.00	0.00	0.00	0.00	0.66	0.00	0.42	0.83	1.00	0.70	0.25
25	3	3	3	70	0.196	0.137	0.000	0.012	0.82	0.00	0.00	0.00	0.41	0.00	0.58	0.53	0.82	0.82	0.39
31	1	6	5	60	0.583	0.000	0.083	0.017	1.00	0.00	0.00	0.00	0.26	0.00	1.00	0.30	0.43	0.69	1.00
34	1	8	7	60	0.999	0.000	0.000	0.017	1.00	0.00	0.00	0.00	0.19	0.00	0.86	0.22	0.33	0.55	1.00

飲酒日数の0を0.01に変更してもクラスター関係は変わらない。以上から、データ値に0があっても値の変換などをする必要がなく、最初は上限値99999を用いて分析し、他の重みを参考にして上限値を小さくすることで良いのではと考えている。

5. 野球選手の年俸評価

67人の野球選手の年俸を出力と考え、四球、本塁打、打率を入力変数と考え分析する。重回帰分析で変数選択するとこの順にモデルに取り込まれるので、最初は（四球、本塁打）の2入力の分析を行い、次に（四球、本塁打、打率）の3入力の分析を行う。

5.1 2入力1出力

（四球、本塁打）の2入力と年俸を出力と考え分析すると、表5.1のExcelシートが得られる。本塁打から推定年俸の列（E2：G68）にセル名Fを与えLINGOに読み込むと、SCORE（H2：H67）、重みを表すセル名W（I2：K68）に67人の選手に最適な重み出力される。この重みを選手ごとに計算した効率値がセル範囲S（L2：CA68）の67行67列のセル範囲に出力される。これらの67人の選手は、野村（41）、宮本（44）、城島（60）の3人が参照集合であり、4個のクラスターに分かれる。

表5.1 (四球, 本塁打) の2入力と年俵を出力とするExcelシート

SN	選手	チーム	リーグ	打率	本塁打	四球	推定年俵	SCORE	W1	W2	W3	S1	S41	S44	S60
1	福浦	口	0	0.35	18	58	5400	0.165	0.001	0.017	0.000	0.165	0.162	0.025	0.160
2	小笠原	日	0	0.34	32	63	11000	0.307	0.001	0.016	0.000	0.307	0.259	0.029	0.301
3	松中	ダ	0	0.33	36	57	15000	0.460	0.001	0.017	0.000	0.460	0.357	0.035	0.453
4	松井 (Matui)	巨	1	0.33	36	120	50000	0.738	0.000	0.008	0.000	0.738	0.732	0.118	0.718
5	ローズ (Rose)	近	0	0.33	55	83	17850	0.375	0.000	0.012	0.000	0.375	0.286	0.028	0.370
6	谷	オ	0	0.33	13	65	8000	0.237	0.015	0.012	0.000	0.219	0.237	0.052	0.212
7	古田	ヤ	1	0.32	15	43	20000	0.822	0.001	0.023	0.000	0.822	0.783	0.113	0.801
8	ベタジーニ	ヤ	1	0.32	39	120	21000	0.310	0.000	0.008	0.000	0.310	0.301	0.046	0.301
9	中村	近	0	0.32	46	104	30000	0.508	0.000	0.009	0.000	0.508	0.450	0.055	0.497
10	磯部	近	0	0.32	17	52	4000	0.136	0.001	0.019	0.000	0.136	0.132	0.020	0.132
11	鈴木	横	1	0.32	6	51	18000	0.739	0.021	0.017	0.000	0.630	0.739	0.254	0.608
12	金本	広	1	0.31	25	128	22200	0.336	0.008	0.006	0.000	0.308	0.336	0.075	0.299
13	真中	ヤ	1	0.31	7	38	8450	0.435	0.026	0.022	0.000	0.396	0.435	0.102	0.383
14	稲葉	ヤ	1	0.31	25	43	4700	0.191	0.001	0.023	0.000	0.191	0.154	0.016	0.188
15	バルデス	ダ	0	0.31	21	77	4000	0.093	0.012	0.010	0.000	0.092	0.093	0.016	0.089
16	ロベス	広	1	0.31	32	51	8000	0.274	0.001	0.019	0.000	0.274	0.214	0.021	0.270
17	松井	西	0	0.31	24	46	25000	0.954	0.001	0.021	0.000	0.954	0.798	0.088	0.936
18	ディアス	広	1	0.3	32	39	3630	0.161	0.001	0.025	0.000	0.161	0.112	0.010	0.160
19	高橋	巨	1	0.3	27	49	12000	0.429	0.001	0.020	0.000	0.429	0.352	0.038	0.422
20	佐伯	横	1	0.3	14	46	8300	0.319	0.001	0.021	0.000	0.319	0.316	0.050	0.311
21	柴原	ダ	0	0.3	7	46	11000	0.483	0.022	0.018	0.000	0.426	0.483	0.133	0.412
22	桧山	阪	1	0.3	12	29	4300	0.261	0.001	0.034	0.000	0.261	0.237	0.030	0.255
23	清原	巨	1	0.3	29	65	30000	0.812	0.001	0.015	0.000	0.812	0.718	0.088	0.795
24	石井	横	1	0.3	8	54	25000	0.939	0.019	0.016	0.000	0.825	0.939	0.265	0.797
25	元木	巨	1	0.29	9	37	8800	0.440	0.025	0.021	0.000	0.422	0.440	0.083	0.410
26	赤星	阪	1	0.29	1	50	1200	0.102	1.000	0.000	0.000	0.043	0.056	0.102	0.041
27	立浪	中	1	0.29	9	54	19000	0.701	0.019	0.015	0.000	0.626	0.701	0.179	0.606
28	小久保	ダ	0	0.29	44	62	18000	0.506	0.001	0.016	0.000	0.506	0.374	0.035	0.500
29	水口	近	0	0.29	3	60	5600	0.210	0.019	0.016	0.000	0.167	0.210	0.158	0.161
30	井出	日	0	0.29	11	46	5800	0.234	0.020	0.017	0.000	0.224	0.234	0.045	0.217
31	岩村	ヤ	1	0.29	18	32	5000	0.274	0.001	0.031	0.000	0.274	0.223	0.024	0.269
32	江藤	巨	1	0.29	30	73	24000	0.580	0.001	0.013	0.000	0.580	0.526	0.068	0.566
33	カブレラ	西	0	0.28	49	84	15000	0.313	0.000	0.012	0.000	0.313	0.251	0.026	0.308
34	メイ	口	0	0.28	31	46	3000	0.114	0.001	0.021	0.000	0.114	0.086	0.008	0.112
35	ラミレス	ヤ	1	0.28	29	27	5400	0.344	0.000	0.037	0.000	0.343	0.209	0.016	0.344
36	小関	西	0	0.28	3	46	5100	0.246	0.024	0.020	0.000	0.198	0.246	0.144	0.191
37	田口	オ	0	0.28	8	43	12000	0.545	0.023	0.019	0.000	0.496	0.545	0.127	0.481
38	ボーリック	口	0	0.28	31	107	16500	0.273	0.008	0.007	0.000	0.273	0.273	0.045	0.266
39	ビティエロ	オ	0	0.28	22	36	6900	0.335	0.001	0.027	0.000	0.335	0.264	0.027	0.330
40	仁志	巨	1	0.27	20	36	16000	0.779	0.001	0.027	0.000	0.779	0.637	0.068	0.765
41	野村	広	1	0.27	9	31	17500	1.000	0.029	0.024	0.000	1.000	1.000	0.165	0.972
42	金城	横	1	0.27	3	56	3900	0.156	0.020	0.017	0.000	0.125	0.156	0.110	0.120
43	大村	近	0	0.27	16	31	5200	0.294	0.001	0.032	0.000	0.294	0.247	0.028	0.289

44	宮本	ヤ	1	0.27	1	27	11800	1.000	1.000	0.000	0.000	0.782	1.000	1.000	0.753
45	葛城	オ	0	0.27	14	43	1200	0.049	0.001	0.023	0.000	0.049	0.048	0.007	0.048
46	今岡	阪	1	0.27	4	29	3700	0.262	0.036	0.030	0.000	0.227	0.262	0.078	0.220
47	吉岡	近	0	0.27	26	66	6700	0.179	0.001	0.015	0.000	0.179	0.165	0.022	0.175
48	中村	中	1	0.27	2	31	12500	0.894	0.036	0.030	0.000	0.721	0.894	0.530	0.694
49	小川	横	1	0.26	15	57	7000	0.223	0.016	0.013	0.000	0.218	0.223	0.040	0.212
50	大島	オ	0	0.26	1	75	8500	0.720	1.000	0.000	0.000	0.203	0.267	0.720	0.195
51	木村	広	1	0.26	7	61	4400	0.151	0.017	0.014	0.000	0.129	0.151	0.053	0.124
52	濱中	阪	1	0.26	13	53	900	0.031	0.018	0.015	0.000	0.030	0.031	0.006	0.029
53	アリアス	オ	0	0.26	38	49	6000	0.213	0.001	0.020	0.000	0.213	0.151	0.013	0.211
54	井端	中	1	0.26	1	49	3200	0.271	1.000	0.000	0.000	0.117	0.152	0.271	0.112
55	塩崎	オ	0	0.26	4	54	3800	0.154	0.020	0.017	0.000	0.126	0.154	0.081	0.121
56	小坂	口	0	0.26	1	77	8500	0.720	1.000	0.000	0.000	0.198	0.260	0.720	0.190
57	谷繁	横	1	0.26	20	65	14000	0.381	0.001	0.015	0.000	0.381	0.376	0.059	0.371
58	東出	広	1	0.26	5	35	2100	0.122	0.029	0.024	0.000	0.107	0.122	0.036	0.103
59	井口	ダ	0	0.26	30	61	4100	0.118	0.001	0.016	0.000	0.118	0.101	0.012	0.116
60	城島	ダ	0	0.26	31	31	18000	1.000	0.000	0.032	0.000	1.000	0.630	0.049	1.000
61	田中	日	0	0.26	20	57	14000	0.434	0.001	0.017	0.000	0.434	0.413	0.059	0.423
62	片岡	日	0	0.25	16	57	18000	0.564	0.016	0.013	0.000	0.560	0.564	0.095	0.544
63	金子	日	0	0.25	8	38	6200	0.311	0.025	0.021	0.000	0.290	0.311	0.066	0.281
64	初芝	口	0	0.25	16	55	10000	0.322	0.001	0.018	0.000	0.322	0.322	0.053	0.313
65	福留	中	1	0.25	15	56	4200	0.136	0.016	0.014	0.000	0.133	0.136	0.024	0.129
66	土橋	ヤ	1	0.25	2	39	7500	0.433	0.029	0.024	0.000	0.344	0.433	0.318	0.331
67	マクレーン	西	0	0.25	39	78	10000	0.225	0.001	0.013	0.000	0.225	0.191	0.022	0.221

注：出力の推定年俵の値が大きいののでW3は0表示になっている。単位を億円にすべきであった。

図5.1は、X軸を（年俵／四球）、Y軸を（年俵／本塁打）でプロットした散布図である。宮本と野村と城島の3選手が効率フロンティアになっている。

原点と宮本を結ぶ直線より上にいる赤星、大島、井端、小坂の5人をクラスター A（宮本を目標）、原点と宮本と野村を結ぶ三角形にある中村、鈴木、土橋ら27人をクラスター B（宮本と野村を目標）、野村と城島でできる三角形に含まれる両松井、吉田、清原、仁志ら33名をクラスター C（野村と城島を目標）になる。そしてラミラスと城島の2名がクラスター Dになる。以上をまとめると表5.2のクラスターができる。

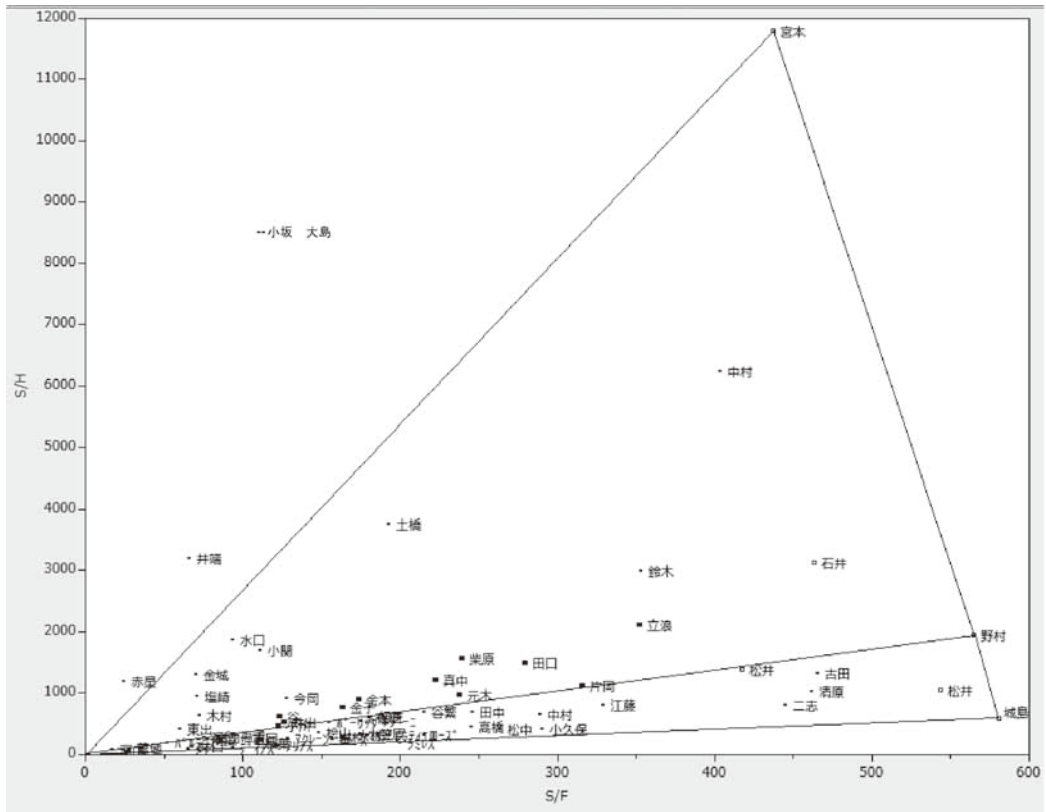


図5.1 4個のクラスター

クラスター Aは宮本を含む5名の選手からなり、宮本が効率値1の手本になる。クラスター Bは野村を含む27名の選手からなり、宮本と野村が効率値1の手本になる。クラスター Cは33名の選手からなり、城島と野村が効率値1の手本になる。クラスター Dは城島を含む2名の選手からなり、城島が効率値1の手本になる。

表5.2 4個のクラスター

クラスター	手本	人数	選手
A (S44)	宮本	5	宮本,赤星,大島,井端,小坂
B (S41)	宮本, 野村	27	野村,谷,鈴木,金本,真中,バルデス,柴原,石井,元木,立浪水口,井出,小関,田口,ポーリック,金城,今岡,中村,小川,木村,濱中,塩崎,東出,片岡,金子,福留,土橋
C (S1)	野村, 城島	33	福浦,小笠原,松中,松井,ローズ,古田,ベタジーニ,中村,磯部,稲葉,ロベス,松井 (西武),ディアス,高橋,佐伯,桧山,清原,小久保,岩村,江藤,カプレラ,メイ,ビティエロ,仁志,大村,葛城,吉岡,アリアス,谷繁,井口,田中,初芝,マクレーン
D (S60)	城島	2	城島,ラミレス

図5.1と表5.2から、非効率な選手が効率的な選手に比べ妥当であるか否かを検討し、もし正しければ自分の目標とする効率的な選手を明確にして目標設定することに意味があるろう。

この結果を見て気付くことは、外人選手の効率が悪いことである。スポーツ選手のように結果が全ての世界でも日本病が読み取れる。すなわち、「外国で過去の実績があるが、もう峠を越した選手を日本人選手以上に厚遇で迎える」、これはスポーツ界にとどまらず企業でも研究の世界でも認められる。筆者が企業人のとき、親会社から「ジーン・アムダール（IBM360の天才設計者と呼ばれ、その後富士通とCPUの共同開発をしていたが提携を解消した直後）との提携の可否の検討」を依頼された。その際私は、「あこがれの天才技術者と会ってみたいが、この技術革新の激しい世界で3度も才能を開花させることができるとは思えない」と回答して沙汰やみになった。これ以外にも、自分で判断できない意思決定者が過去の実績がある外人をことのほか有難がる風潮がある。

図5.2はクロス効率値SをJMPに入力し群平均法でクラスター分析を行った。変数のクラスターは、左からC,D,B,Aの4つに上手く分かれた。しかし必ずこのようにうまくいくかは今後の検討事項である。

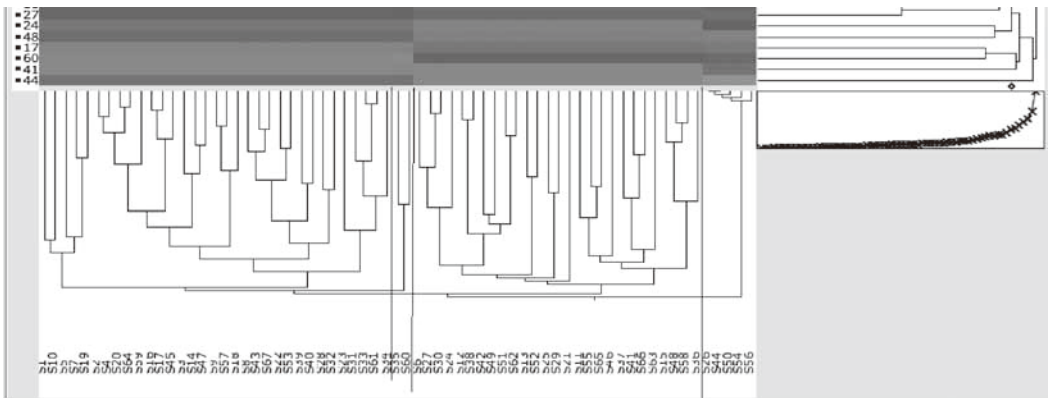


図5.2 セル範囲SをJMPに入力し群平均法でクラスター分析

図5.3は主成分分析の因子負荷量とスコアプロットである。固有値の累積寄与率は第3主成分までで100%になった。因子負荷量プロットも3個のグループになる。表5.4の4クラスターと食い違うが、クラスターDはほぼクラスターCと違いがないと考えられ、融合してもよさそうだ。

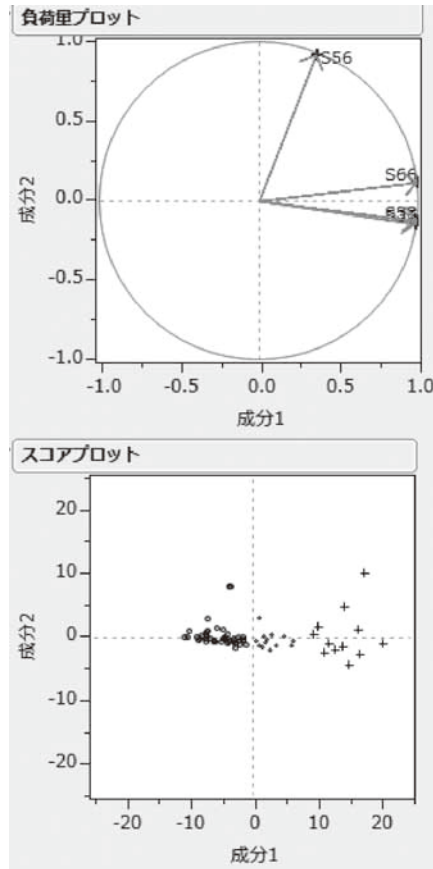


図5.3 効率値配列Sの主成分分析

5.2 3入力1出力

(四球, 本塁打, 打率) の3入力と年俵を出力と考え分析すると, 表5.3のExcelシートが得られる。これらの67人の選手の参照集合は, 2入力の3人に加えて松井 (4), 松井 (西鉄, 17), 石井 (24) の3人が加わり6人になった。そして11個のクラスターに分かれる。

表5.3 四球, 本塁打, 打率の3入力と年俵を出力とするExcelシート

SN	打率	本塁打	四球	推定年俵	SCORE	W1	W2	W3	W4	S1	S2	S4	S8	S13	S17	S24	S35	S41	S44	S60
1	0.35	18	58	5400	0.187	0.859	0.005	0.011	0.00003	0.187	0.179	0.149	0.104	0.168	0.177	0.122	0.160	0.187	0.025	0.165
2	0.34	32	63	11000	0.349	1.025	0.000	0.010	0.00003	0.343	0.349	0.233	0.216	0.266	0.332	0.160	0.301	0.343	0.029	0.307
3	0.33	36	57	15000	0.510	1.099	0.000	0.011	0.00003	0.490	0.510	0.300	0.299	0.362	0.491	0.199	0.453	0.490	0.035	0.460
4	0.33	36	120	50000	1.000	1.257	0.016	0.000	0.00002	1.000	1.000	1.000	1.000	0.794	0.862	0.662	0.718	0.999	0.118	0.738
5	0.33	55	83	17850	0.473	0.858	0.000	0.009	0.00003	0.435	0.473	0.275	0.364	0.295	0.421	0.164	0.370	0.435	0.028	0.375
6	0.33	13	65	8000	0.268	0.832	0.004	0.010	0.00003	0.268	0.252	0.259	0.164	0.253	0.241	0.229	0.212	0.268	0.052	0.219
7	0.32	15	43	20000	0.860	1.061	0.006	0.013	0.00004	0.859	0.815	0.616	0.411	0.793	0.847	0.522	0.801	0.860	0.113	0.822
8	0.32	39	120	21000	0.434	3.106	0.000	0.000	0.00002	0.419	0.423	0.406	0.434	0.326	0.362	0.261	0.301	0.419	0.046	0.310
9	0.32	46	104	30000	0.676	0.729	0.000	0.007	0.00002	0.651	0.676	0.524	0.624	0.478	0.586	0.324	0.497	0.650	0.055	0.508

10	0.32	17	52	4000	0.153	0.941	0.005	0.012	0.00004	0.153	0.146	0.118	0.083	0.136	0.145	0.096	0.132	0.153	0.020	0.136
11	0.32	6	51	18000	0.809	1.673	0.079	0.000	0.00004	0.741	0.670	0.730	0.381	0.781	0.674	0.809	0.608	0.741	0.254	0.630
12	0.31	25	128	22200	0.556	1.574	0.020	0.000	0.00003	0.439	0.427	0.556	0.471	0.372	0.364	0.400	0.299	0.439	0.075	0.308
13	0.31	7	38	8450	0.442	0.391	0.028	0.018	0.00005	0.415	0.375	0.335	0.180	0.442	0.402	0.354	0.383	0.415	0.102	0.396
14	0.31	25	43	4700	0.198	0.567	0.001	0.018	0.00004	0.194	0.195	0.118	0.101	0.154	0.198	0.085	0.188	0.194	0.016	0.191
15	0.31	21	77	4000	0.117	0.725	0.004	0.009	0.00003	0.117	0.114	0.110	0.086	0.100	0.104	0.083	0.089	0.117	0.016	0.092
16	0.31	32	51	8000	0.300	1.214	0.000	0.012	0.00004	0.290	0.300	0.177	0.173	0.216	0.292	0.118	0.270	0.290	0.021	0.274
17	0.31	24	46	25000	1.000	0.539	0.001	0.017	0.00004	1.000	1.000	0.645	0.541	0.807	1.000	0.467	0.936	1.000	0.088	0.954
18	0.3	32	39	3630	0.164	0.608	0.001	0.020	0.00005	0.152	0.161	0.081	0.080	0.110	0.164	0.054	0.160	0.152	0.010	0.161
19	0.3	27	49	12000	0.465	1.254	0.000	0.013	0.00004	0.458	0.465	0.294	0.265	0.357	0.456	0.205	0.422	0.458	0.038	0.429
20	0.3	14	46	8300	0.353	1.049	0.006	0.013	0.00004	0.353	0.335	0.274	0.183	0.325	0.337	0.232	0.311	0.353	0.050	0.319
21	0.3	7	46	11000	0.505	0.343	0.025	0.016	0.00005	0.487	0.443	0.447	0.243	0.505	0.451	0.468	0.412	0.487	0.133	0.426
22	0.3	12	29	4300	0.261	0.000	0.001	0.034	0.00006	0.239	0.224	0.151	0.095	0.231	0.252	0.134	0.255	0.240	0.030	0.261
23	0.3	29	65	30000	0.971	1.047	0.000	0.011	0.00003	0.959	0.971	0.712	0.670	0.748	0.900	0.484	0.795	0.959	0.088	0.812
24	0.3	8	54	25000	1.000	1.488	0.070	0.000	0.00004	1.000	0.919	1.000	0.564	1.000	0.899	1.000	0.797	1.000	0.265	0.825
25	0.29	9	37	8800	0.447	0.379	0.027	0.017	0.00005	0.444	0.409	0.343	0.201	0.447	0.432	0.330	0.410	0.444	0.083	0.422
26	0.29	1	50	1200	0.102	0.000	1.000	0.000	0.00008	0.053	0.047	0.063	0.027	0.060	0.047	0.095	0.041	0.053	0.102	0.043
27	0.29	9	54	19000	0.758	0.989	0.005	0.012	0.00004	0.758	0.701	0.742	0.433	0.745	0.683	0.713	0.606	0.758	0.179	0.626
28	0.29	44	62	18000	0.607	1.091	0.000	0.011	0.00003	0.560	0.607	0.335	0.413	0.382	0.556	0.205	0.500	0.560	0.035	0.506
29	0.29	3	60	5600	0.349	2.319	0.109	0.000	0.00006	0.215	0.193	0.271	0.129	0.230	0.185	0.349	0.161	0.215	0.158	0.167
30	0.29	11	46	5800	0.255	1.084	0.006	0.014	0.00004	0.255	0.238	0.215	0.134	0.243	0.239	0.193	0.217	0.255	0.045	0.224
31	0.29	18	32	5000	0.274	0.000	0.001	0.031	0.00005	0.258	0.253	0.154	0.116	0.219	0.272	0.118	0.269	0.258	0.024	0.274
32	0.29	30	73	24000	0.725	0.978	0.000	0.010	0.00003	0.716	0.725	0.570	0.561	0.555	0.654	0.380	0.566	0.715	0.068	0.580
33	0.28	49	84	15000	0.410	0.884	0.000	0.009	0.00003	0.381	0.410	0.262	0.354	0.262	0.357	0.156	0.308	0.380	0.026	0.313
34	0.28	31	46	3000	0.124	1.339	0.000	0.014	0.00004	0.119	0.124	0.070	0.071	0.087	0.121	0.046	0.112	0.119	0.008	0.114
35	0.28	29	27	5400	0.344	0.000	0.000	0.037	0.00006	0.282	0.302	0.132	0.128	0.200	0.328	0.088	0.344	0.282	0.016	0.343
36	0.28	3	46	5100	0.325	2.374	0.112	0.000	0.00006	0.237	0.212	0.255	0.121	0.262	0.213	0.325	0.191	0.237	0.144	0.198
37	0.28	8	43	12000	0.568	0.354	0.025	0.016	0.00005	0.565	0.519	0.499	0.285	0.568	0.526	0.491	0.481	0.565	0.127	0.496
38	0.28	31	107	16500	0.394	3.584	0.000	0.000	0.00002	0.375	0.375	0.388	0.394	0.298	0.321	0.255	0.266	0.375	0.045	0.273
39	0.28	22	36	6900	0.343	0.669	0.002	0.022	0.00005	0.331	0.334	0.197	0.167	0.263	0.343	0.141	0.330	0.331	0.027	0.335
40	0.27	20	36	16000	0.799	0.672	0.002	0.022	0.00005	0.780	0.777	0.480	0.390	0.637	0.799	0.354	0.765	0.780	0.068	0.779
41	0.27	9	31	17500	1.000	1.410	0.008	0.018	0.00006	0.999	0.923	0.716	0.427	1.000	1.000	0.675	0.972	1.000	0.165	1.000
42	0.27	3	56	3900	0.254	2.425	0.114	0.000	0.00007	0.160	0.144	0.200	0.096	0.171	0.138	0.254	0.120	0.160	0.110	0.125
43	0.27	16	31	5200	0.294	0.000	0.001	0.032	0.00006	0.283	0.275	0.174	0.128	0.245	0.294	0.136	0.289	0.283	0.028	0.294
44	0.27	1	27	11800	1.000	0.000	1.000	0.000	0.00008	0.780	0.672	0.664	0.291	1.000	0.763	1.000	0.753	0.780	1.000	0.782
45	0.27	14	43	1200	0.055	1.134	0.006	0.014	0.00005	0.055	0.053	0.043	0.030	0.049	0.053	0.035	0.048	0.055	0.007	0.049
46	0.27	4	29	3700	0.263	0.531	0.038	0.024	0.00007	0.230	0.204	0.184	0.092	0.263	0.226	0.218	0.220	0.230	0.078	0.227
47	0.27	26	66	6700	0.222	1.073	0.000	0.011	0.00003	0.221	0.222	0.178	0.168	0.174	0.202	0.121	0.175	0.221	0.022	0.179
48	0.27	2	31	12500	0.935	2.784	0.131	0.000	0.00007	0.763	0.668	0.684	0.314	0.914	0.728	0.935	0.694	0.764	0.530	0.721
49	0.26	15	57	7000	0.268	0.950	0.005	0.012	0.00004	0.268	0.258	0.244	0.177	0.237	0.242	0.194	0.212	0.268	0.040	0.218
50	0.26	1	75	8500	0.737	3.225	0.152	0.000	0.00009	0.286	0.257	0.490	0.215	0.302	0.234	0.737	0.195	0.286	0.720	0.203
51	0.26	7	61	4400	0.199	1.687	0.079	0.000	0.00005	0.168	0.155	0.198	0.111	0.165	0.145	0.199	0.124	0.168	0.053	0.129
52	0.26	13	53	900	0.037	1.009	0.005	0.013	0.00004	0.037	0.035	0.033	0.023	0.033	0.033	0.028	0.029	0.037	0.006	0.030
53	0.26	38	49	6000	0.245	1.321	0.000	0.013	0.00004	0.225	0.245	0.127	0.153	0.153	0.230	0.079	0.211	0.225	0.013	0.213
54	0.26	1	49	3200	0.278	3.235	0.152	0.000	0.00009	0.147	0.131	0.185	0.081	0.166	0.128	0.278	0.112	0.147	0.271	0.117
55	0.26	4	54	3800	0.227	2.220	0.105	0.000	0.00006	0.161	0.145	0.193	0.097	0.168	0.140	0.227	0.121	0.161	0.081	0.126

56	0.26	1	77	8500	0.739	3.235	0.152	0.000	0.00009	0.280	0.253	0.492	0.216	0.295	0.229	0.739	0.190	0.280	0.720	0.198
57	0.26	20	65	14000	0.480	0.849	0.005	0.011	0.00003	0.480	0.471	0.429	0.356	0.400	0.430	0.312	0.371	0.480	0.059	0.381
58	0.26	5	35	2100	0.126	0.449	0.032	0.021	0.00006	0.117	0.105	0.102	0.053	0.126	0.111	0.113	0.103	0.117	0.036	0.107
59	0.26	30	61	4100	0.144	1.140	0.000	0.012	0.00004	0.140	0.144	0.101	0.105	0.105	0.132	0.066	0.116	0.140	0.012	0.118
60	0.26	31	31	18000	1.000	0.000	0.001	0.031	0.00006	0.896	0.974	0.436	0.465	0.616	1.000	0.282	1.000	0.896	0.049	1.000
61	0.26	20	57	14000	0.528	0.934	0.005	0.012	0.00004	0.528	0.521	0.435	0.366	0.435	0.483	0.314	0.423	0.527	0.059	0.434
62	0.25	16	57	18000	0.693	0.954	0.005	0.012	0.00004	0.693	0.670	0.623	0.472	0.599	0.625	0.480	0.544	0.693	0.095	0.560
63	0.25	8	38	6200	0.326	1.296	0.007	0.016	0.00005	0.326	0.301	0.277	0.163	0.323	0.306	0.265	0.281	0.326	0.066	0.290
64	0.25	16	55	10000	0.395	0.978	0.005	0.012	0.00004	0.395	0.382	0.347	0.263	0.341	0.358	0.267	0.313	0.395	0.053	0.322
65	0.25	15	56	4200	0.165	0.973	0.005	0.012	0.00004	0.165	0.159	0.151	0.111	0.144	0.148	0.118	0.129	0.165	0.024	0.133
66	0.25	2	39	7500	0.587	2.914	0.137	0.000	0.00008	0.407	0.361	0.434	0.201	0.460	0.367	0.587	0.331	0.407	0.318	0.344
67	0.25	39	78	10000	0.299	0.966	0.000	0.010	0.00003	0.283	0.299	0.213	0.270	0.202	0.259	0.129	0.221	0.283	0.022	0.225

表5.4は11個のクラスターである。下線を引いた松井（巨人，4），松井（西武，17），石井（24），野村（41），宮本（44），城島（60）の6人が効率的になった。クラスター1から9までが2名以上の選手から構成されている。これらが実際にどのような意味をもっているかは，2011年度の応用発展演習の学生で野球に詳しいものがいれば課題として検討したい。

表5.4 3入力1出力の18個のクラスター

クラスター	手本	人数	選手
S44	44	2	26,44
S24	24,44	12	11,24,29,36,42,48,50,51,54-56,66
S41	24,41,17	7	7,10,20, 30,41,45,63
S2	4,17	15	2,3,5,9,16,19,23,28,32,33,34,47,53,59,67
S1	4,17,24	11	1,6,15,27,49,52,57,61,62,64,65
S13	24,41,44	6	13,21,25,37,46,58
S17	17,41,60	5	14,17,18,39,40
S60	41,60	4	22,31,43,60
S8	4	2	8,38
S4	24,4	2	4,12
S35	60	1	35

図5.4は，3入力を表示できないので，2入力の図5.1の散布図を利用した。6人の選手の効率的フロンティアは，3次元空間を2次元に投影しているので複雑な折れ線になっている。上の宮本と赤星はかなり離れているがクラスター S44である。赤星は宮本を手本に相当の努力が必要である。石井を含む12人のクラスター S24はかなり広い範囲に散在し，石井と宮本を手本とする選手群である。原点と野村を結ぶ直線の下には，2人の松井，野村，城島を手本とし，多くの選手がひしめいている。高収入を得るには，これらのクラスターに入る必要がある。

図5.6は主成分分析の因子負荷量とスコアプロットである。累積寄与率は第8主成分までで100%であった。

因子負荷プロットは、9から10程度の異なった因子があることが分かる。

以上から、表5.4の11個のクラスターはS35と同じくS4、S8あるいはS44を他と融合することも考えられる。

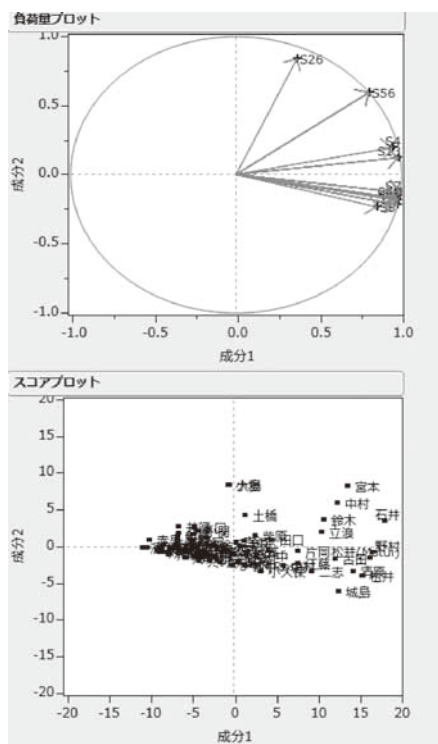


図5.6 主成分分析の因子負荷量とスコアプロット

以上から、DEAは3入力1出力程度のモデルになると2入力1出力に比べ表や図の作成に時間がかかる。しかし、これまでの評価法と異なり、異なった視点からの評価ができる点で重要である。

6. まとめ

DEAは、大学における利用面と教育にとっても重要と考える。

評価法としては、特に公的な大学の評価、科研費申請の基礎資料として導入されるべきであろう。外部資金の獲得に大学や研究者が注力しているが、結果についての議論が少ないようである。過去の入力と出力（評価が難しいので試行錯誤する必要がある）を分析し予算の使われ方が効率的かどうかの参考指標になる。また成蹊大学の情報図書館も外部評価が高く喜ばしいことであるが、折角データがあるのに外部評価者はなぜ客観

的な分析をしないのか不思議である。

一方、今後は年度予算の手当てが難しくなる傾向にある各大学にあって、適切な予算配分が問題になってくる。大学にあっては、各学部の各教員は自分の信念に基づいて教育と研究を行っているが、他者の位置づけはお互いに分かりづらい。一応委員会で議論した結果に任せるべきであるが、基本的な情報がないと議論しにくいであろう。ただし、不利な評価の場合は、本人や学部から評価項目に問題がないか意見を述べる機会をもち、数年かけて洗練させていく必要がある。

一方、学生には今後社会に出て役立つ「問題解決学」を教育に取り入れていく必要があろう。高校数学で分かる金利計算やローン計算は「21世紀の一般教養」として必要と考えている [9]。これに合わせて、評価法やプロジェクト管理 (PERT) はどのような組織でも必要な問題解決の技術である。

DEAによる評価法を、大学の各部門の評価、研究や教育への利用の普及のため、図3.1の形式でExcel上にデータ (配列F) さえ準備いただければ、分析結果のフィードバックを条件にLINGOによるDEAの分析をボランティアで引き受けてもかまいません。

文 献

- [1] 刀根薫 (1993). 経営効率性の測定と改善案— 包絡分析法 DEA による —. 日科技連.
- [2] 高森寛, 新村秀一 (1987). 統計処理エッセンシャル. 丸善.
- [3] 新村秀一 (2010). 最適線形判別関数. 日科技連.
- [4] 上田徹, 住舎俊宏 (2002). どの野球選手の攻撃力が優れているだろうか. オペレーションズ・リサーチ, 47,137-141.
- [5] 廣津信義, 上田徹 (2008). 経営効率分析法 (DEA) を利用した野球チームのラインナップ選定のための一手法—北京五輪野球日本代表候補選手を例として—. 順天堂大学スポーツ科学研究, 12, 1 - 10.
- [6] A.Charnes, W. W. Cooper and E. Rhodes (1978) . Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, 2, 429-444.
- [7] 新村秀一 (2004). JMP活用統計学とおき勉強法. 講談社.
- [8] 新村秀一 (2007). ExcelとLINGOで学ぶ数理計画法. 丸善.
- [9] 新村秀一 (2009). 数学のできる問題解決学. 成蹊大学一般研究報告. 42 (4), 1-52.
- [10] 新村秀一 (2011). 数理計画法による問題解決学. 日科技連.

PRINTED BY
SEIKO-SHA CO. LTD.
1-5-15, NISHITUTUJIGAOKA, CHOUFUSHI, TOKYO

Seikei University
3-3-1, Kichijoji-Kitamachi, Musashino-shi,
Tokyo, 180-8633 Japan