

経営効率性による企業と大学の評価法 (1)

－DEA普及のための方策1－

新 村 秀 一

普通、評価対象のデータがある場合、重回帰分析による予測や、判別分析などの統計手法の利用が考えられる。この場合、目的変数の評価項目を計測しやすいあるいは時間的に先行する説明変数の加重和で予測することが中心である。

しかし、入力（投入、説明変数）と出力（産出、目的変数）の比（出力／入力あるいは産出／投入）という単純な考えで評価対象が効率的か非効率的かという視点で、各評価対象の良いところを最大限に見つけ、非効率的であればどの評価対象を手本にし、どこを改善すれば良いかが分かるDEA法（Data Envelopment Analysis, 包絡分析法あるいは経営効率分析法）を紹介する。

本研究では、DEA法が本来対象とする経営組織体のほか、学生の成績データの評価、44車種の車の価格性能比を例としてケータイなどの各種製品のマーケティングへの利用、そしてこれまで重回帰分析や判別分析で取り上げられてきた分野への適用を紹介する。

また統計手法と比較することで、この新しい手法の理解を容易にすることを試みる。

1. 基本的な考え方

1.1 代表的なDEAで扱うデータ

DEA法は、1978年にテキサス大学のCharnes, Cooper and Rhodesによって提案されたので、一般的にCCRモデルと呼ばれている。特に、CharnesとCooperは経営科学で大きな足跡を残し、Cooper教授は会計学の大家でもある。そのため、企業（銀行、メーカー、デパートやスーパーといった流通業、商社など）や政府や自治体などの部門（事業部、流通業の店舗、メーカーの工場、公的な病院や図書館）を念頭にこの手法が開発された。しかし効率性で考えることのできない大学においても、自己規制と合意形成に利用できると思われる。

本研究では、DEA法が本来対象とするこれらの経営組織体のほか、44車種の車の価格性能比を例としてケータイなどの商品のマーケティングへの応用を取り上げる。その後で、重回帰分析や判別分析で取り上げる、学生の成績データ [5] [6] の評価、野球選手の年俵評価、日経新聞の優良企業100社と非有優良企業30社の判別や平均給与の分析を行い適用への注意点を述べる。

表1.1は、日本における、この分野の指導的立場にある刀根 [4] [14] が説明に用いた東京

都23区の区立図書館に関する有名なデータである。本データは、DEA法が本来対象とする経営組織体の評価のための代表事例と考えてほしい。データの取得年が昭和61年と古いが、学生に最近のデータで現状を再評価するテーマを課すことも考えている。また、実名で取り上げているので、最新のデータでない方が良いと考えた。

表1.1 東京都23区の区立図書館 (昭和61年)

SN	区	床面積 (千㎡)	蔵書数 (千冊)	職員数 (人)	貸出数 (千冊)	登録者数 (千人)
1	千代田	2249	164	26	105	6
2	中央	4617	339	30	315	18
3	台東	3873	282	51	542	16
4	荒川	5541	401	78	848	31
5	港	11381	363	69	759	57
6	文京	10086	542	114	1439	66
7	墨田	5434	508	61	840	35
8	渋谷	7524	339	74	541	33
9	目黒	5077	511	84	1562	65
10	豊島	7029	394	68	978	41
11	新宿	11121	510	96	930	47
12	中野	7072	527	92	1345	56
13	品川	9348	602	127	1165	70
14	北	7781	529	96	1349	37
15	江東	6235	394	77	1101	58
16	葛飾	10593	516	101	1070	46
17	板橋	10866	567	118	1708	103
18	江戸川	6500	468	74	1223	47
19	杉並	11469	768	103	2300	85
20	練馬	10868	670	107	1901	70
21	足立	10717	845	120	1910	89
22	大田	10716	1259	242	3055	98
23	世田谷	10888	1149	202	4096	191

区立図書館を評価する場合、床面積、蔵書数、職員数といった図書館の運営に投入される経営資源がある。刀根は、分館数や図書館予算、区の人口なども考慮すべきと指摘している。

その結果、貸出数や登録者数といった評価項目が産出される。重回帰分析の立場では、図1.1のように投入量は入力変数であり、産出量は目的変数である。DEAでは3入力1出力モデルという。

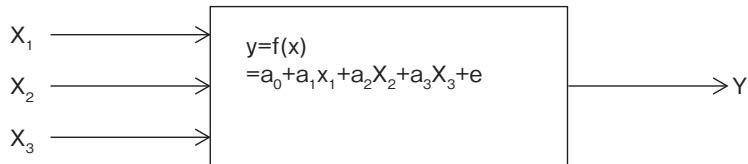


図1.1 重回帰分析

そして産出量の y を $y=a_0+a_1*x_1+a_2*x_2+a_3*x_3+e$ というように予測する。しかし、表1.1のように産出量が2個ある場合（3入力2出力モデル）、2個の重回帰式を別々に考える必要がある。しかしDEAでは、 $(b_1*y_1+b_2*y_2)/(a_1*x_1+a_2*x_2+a_3*x_3)$ という比でもって効率性を考える。

1.2 回帰による通常のアプローチとDEA法のアプローチ

表1.1のようなデータがあると、回帰分析の利用が頭に浮かぶ。職員数と蔵書数を説明変数とし、貸出数を単回帰分析すると図1.2の結果が得られる。単回帰式は次のとおりである。

$$\text{貸出数} = -226.8 + 16.4 * \text{職員数}$$

$$\text{貸出数} = -441.5 + 3.3 * \text{蔵書数}$$

貸出数／職員数の比は定数項の-226.8を無視すれば16.4と一定である。23区立図書館は、結局職員数が1単位（人）増えれば、平均的に16.4（千冊）だけ貸出数が増えることが期待できる。図に書き込んだ回帰直線の下にくる大田区などの図書館は、この平均を下回っていて1人当たりの貸出数を16.4（千冊）まで増やす努力が必要である。世田谷区は職員数と蔵書数の両方の回帰分析でも回帰直線の上にくる図書館である。見えにくいだが、足立区は貸出数／職員数では直線の上にあるが、貸出数／蔵書数では下にきている。

本研究では、統計分析はJMPを用いる。[6]に2024年まで使える評価版が添付してある。

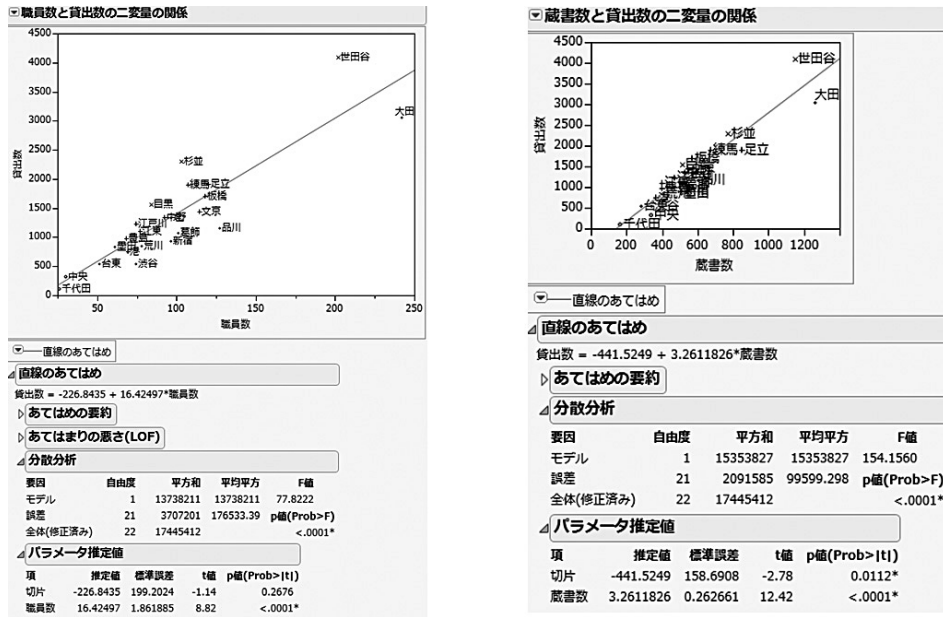


図1.2 回帰分析のアプローチ

図1.3は、DEAの考え方である。これらはいずれも1入力1出力モデルである。統計は分析対象をデータとしてとらえ、そこから役に立つ情報を引き出す学問である。もう一つの問題解決学の有力な学問である数理計画法では、分析対象を数学的に記述しこれをモデルと呼んでいる。

左の貸出数/職員数では、原点から貸出数/職員数の比が2300/103 ≃ 23と一番大きい杉並区を通る直線を引いている。杉並区が投入資源の職員数に対して、産出量の貸出数の比が一番大きく、他の図書館のお手本になるわけである。ほぼその下にある練馬区は、職員数が107人で貸出数が1901であるので、貸出数を2300/103*107 ≃ 2389まで増やす必要がある。2389-1901=488(千冊)足りないが、手本の杉並を1とすると、1901/2389 ≃ 0.79と非効率である。この原点を通る直線は、効率的フロンティアという。評価対象はこの直線の下にきて全て非効率的である。すなわち、効率的な杉並を手本(専門書では、少し硬い表現で優位集合あるいは参照集合といっている)にして改善を図る必要がある。

右の貸出数/蔵書数は、世田谷が手本になる。

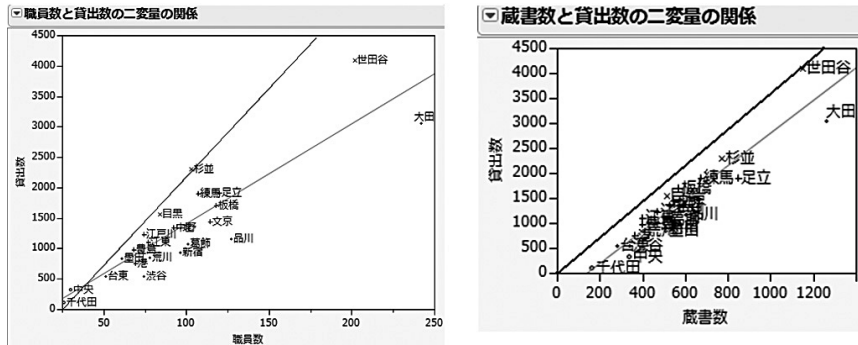


図 1.3 DEAのアプローチ

1.3 回帰分析によるアプローチの問題点

回帰分析によるアプローチの問題点は以下の通りである。

- ・ 複数の入力変数の線形和で1個の目的変数の値を予測する方法である。この目的変数が、評価を表す指標である場合に、重回帰分析は評価手法と考えられる。
- ・ 2個以上の複数の出力、すなわち多入力多出力が扱えない。2組の複数の変数を線形結合して2組の合成変数を作り、相関を最大にする正準相関分析があるが、統計ではそれほど普及していない。
- ・ 評価対象の誰を手本にして、何を改善すれば良いかが分からない。

一方、DEA法と異なり、統計手法の適用はデータに制約が少ない。DEA法は、一般的に次のようなデータに関する制約がある [13]。

- ・ 入力データは正の値である。負や0の値があれば、変換する必要がある。
- ・ 入出力の項目の選択は、専門家の意見などを聞く必要がある。
- ・ 一般的に入力データは値が小さいものほど好ましく、出力は大きなものほど好ましい。
- ・ 単位は任意にとっても良い。

これに対して、筆者は次のような対応策を考えた。

- ・ 入力データに負や0の値があれば、出力にはそのような制約がないので、入力と出力を交換することを検討する。また0がある場合、2.2で紹介する汎用DEAプログラムの重みの上限を変更することで対応できる。
- ・ DEA法で分析するデータが、重回帰で分析できる重回帰型DEAデータ、あるいは判別分析で分析できる判別型DEAデータの場合、重回帰分析や判別分析の分析結果を利用できる。特に変数選択法の利用は有効である。
- ・ データの単位は任意にとっても良いが、最大値が1から10未満に変換すれば、重みの比較が容易になる。

1.4 包絡分析法とは

DEA法は、企業の事業部、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の評価対象すなわち意思決定主体 (Decision Making Unit, DMU) の効率性を評価する手法である。あるいは、野球選手の年俵が成績とどう関係しているかも評価できる。また各教員が教育というお互いの信念で行っている場合、研究や教育に関する分析をベースに、自分自身がその結果を解釈し、軌道修正し、合意形成に利用することが考えられる。

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。これらは、分析対象のデータに共通の重みをつけて行われる。

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、モータの出力はロスが発生するので、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る。この場合、1入力1出力であるが、多くの場合は多入力多出力であることが多い。例えば表1.1の図書館データでは、評価対象 (DMU) は23個ある。これらの評価対象には、3個の入力と2個の出力がある。

一般化すれば、 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。そして、 k 個の評価対象 (DMU) がある。1番目の図書館 (DMU₁) の測定値を $(x_{11}, \dots, x_{m1}), (y_{11}, \dots, y_{n1})$ とする。この場合、DMU_h の効率性は次の比になる。

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

ここで、DMU_h の重み (b_{1h}, \dots, b_{nh}) と (a_{1h}, \dots, a_{mh}) を用いて、すべての評価対象で効率値を計算し、この比が1以下になるような制約条件のもとで、DMU_h を最大にする重みを数理計画法の入門である線形計画法 (Linear Programming, LP) で決めてやればよい。高校数学で習う「領域の最大/最小問題」は、実はLPの理論そのものを絵で扱っている。しかし、重みをDMU_i ($i=1, \dots, 23$) 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法の一番の特徴は、DMU_i ごとにその効率を最大になるように個別の重みを求める点である。

次が、LPでDMU_h を最大化する重みを求めるLPモデルである。 k 個の評価対象でDMU_i の重みを用いて効率値が1以下になる制約を課していることに注意してほしい。

$$\begin{aligned} \text{MAX} &= (b_{1h} * y_{1h} + \dots + b_{nh} * y_{nh}) / (a_{1h} * x_{1h} + \dots + a_{mh} * x_{mh}) ; \\ (b_{1h} * y_{1p} + \dots + b_{nh} * y_{np}) / (a_{1h} * x_{1p} + \dots + a_{mh} * x_{mp}) &\leq 1; \text{ for } p=1, \dots, k \end{aligned}$$

このモデルは、数理計画法で分数計画法と呼ばれ、非線形最適化という解法が必要になる。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。

$$\begin{aligned} \text{MAX} &= b_{1h} * y_{1h} + \dots + b_{nh} * y_{nh}; \\ a_{1h} * x_{1h} + \dots + a_{mh} * x_{mh} &= 1; \\ (b_{1h} * y_{1p} + \dots + b_{nh} * y_{np}) &\leq (a_{1h} * x_{1p} + \dots + a_{mh} * x_{mp}) ; \text{ for } p=1, \dots, k \end{aligned}$$

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回解く必要がある。すなわち、23区の図書館（DMU）を最大化する目的関数を23個の異なったLPモデルを繰り返し解く必要がある。

これを表1.1のデータを例にして説明すると次のようになる。例えば千代田区（DMU₁）を最大化するモデルは次のようになる。ただし、h=1であるが、他の22区でも計算する必要があるので変数hにしてある。

$$\text{MAX} = b_{1h} * 105 + b_{2h} * 6;$$

$$a_{1h} * 2249 + a_{2h} * 164 + a_{3h} * 26 = 1;$$

$$b_{1h} * 105 + b_{2h} * 6 \leq a_{1h} * 2249 + a_{2h} * 164 + a_{3h} * 26; \text{ (千代田区の制約)}$$

・

・

$$b_{1h} * 4096 + b_{2h} * 191 \leq a_{1h} * 10888 + a_{2h} * 1149 + a_{3h} * 202; \text{ (世田谷区の制約)}$$

目的関数の千代田区の値（効率値）は、最大値が1になるように最適な重みがLPで選ばれる。1になれば、千代田区は効率的に運用されていると考える。もし、0から1未満の値であれば、制約式に含まれる他の図書館の効率値が1になるものがあるためである。自分に一番有利な重みであるにもかかわらず、他の図書館が高く評価されれば、千代田区は謙虚に反省し、どこを改善すべきかを手本であるその図書館を目標に改善する必要がある。これがこれまでの評価法にない、新しい視点である。

DEA法の効率性の考え方を、図1.4の1入力2出力モデルで説明する。例えば、野球選手の場合、打席数が多ければ得点と年俵が上がるのが期待できる。そこで打席数を入力とし、得点と年俵を出力とした1入力2出力モデルを考えればよい。各DMU_iに対してCCRモデルを1回解く。A, B, C, Dの効率値が1で、Gは例えば0.7であったとする。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。これが包絡分析と呼ばれるゆえんである。

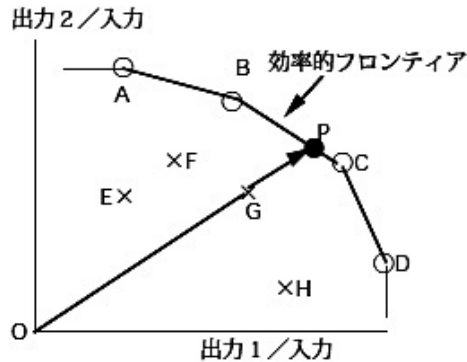


図1.4 1入力2出力

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、出力1／入力と出力2／入力を改善する必要がある。すなわちP点が努力目標になる。あるいは、BとCの線分のどこでも良い。しかし実態のあるBとかCを手本にするのが实际的である。

一方、次のようなモデルを考えることができる。

$$\text{MIN} = b_1 * y_{1h} + \dots + b_n * y_{nh};$$

$$a_1 * x_{1h} + \dots + a_m * x_{mh} = 1;$$

$$(b_1 * y_{1p} + \dots + b_n * y_{np}) \geq (a_1 * x_{1p} + \dots + a_m * x_{mp}); \text{ for } p=1, \dots, k$$

このモデルは、図1.5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。A、B、Cの非効率値は1であり、Gは1.5だったとしよう。これが図のP点まで非効率的になれば、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。

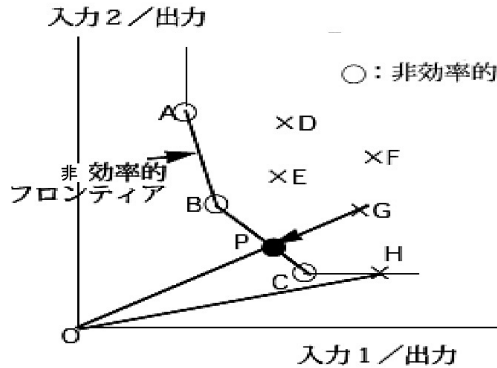


図1.5 逆CCRモデル

2. DEAの普及のために

2.1 DEA普及のための提言

DEA普及のために、次の点を提案する。

- ・誰もが使える汎用プログラムを開発した [8]。
- ・2入力1出力あるいは1入力2出力でしか、手本（参照集合という重い言葉も普及を妨げている）と非効率な評価対象（DMU）の関係が視覚的に説明できない。入力と出力の合計が4以上の多入力多出力の関係を説明するために、「DEAクラスター」を提案したい。2.2で紹介するクロス効率値で効率値が1になる同じパターンをもつDMUを、同じDEAクラスターに属する構成員と考えると、すべてのDMUは一つのDEAクラスターに所属する。DEAクラスターには、効率値が1となる手本がある。非効率なDMUは、効率的な手本と比較しどこを改善すべきかの方法を提案する。
- ・CCRモデルの双対モデルを考えて、各評価対象の手本を求めることができる。しかし、これを理解するためには、相当の数値計画法の知識が要求され、企業への普及の障害になる。また、改善方法も計算できるが知識がないと理解しにくい。これに代わって、非効率な評価対象の入力の1個の値を固定し、手本と同じ構成比をもつ理想状態を計算し、実際の値との差で改善点を見つける1入力固定改善法を提案する。
- ・DEA法は、入力と出力に選ぶ変数が重要である。少なくとも重回帰分析や判別分析で分析できるデータ（回帰型DEAデータと呼ぶ）は、説明変数が入力、目的変数が出力と考えられる。これらのデータでは、重回帰や判別分析の変数選択法で入力変数を決めるなど、統計手法の利用が考えられる。
- ・入力データが0を含む場合、これを正に変換して対応することが行われてきたが、納得できる解決策はなかった。汎用DEAプログラムの重みの上限値の決め方を変えることで対応する

方法を提案する。

2.2 LINGOによるDEAの汎用プログラム

図2.1はLINGO（リンゴ）というシカゴ大学ビジネススクールのL.Schrage教授が創設者であるLINDO Systems Inc.の数理計画法ソフト [1] [7] [8] で記述された汎用プログラムである。図書館データの2入力1出力モデルで考えている。本稿では、CCRモデルだけで議論する。内容の詳細を理解する必要はないが、汎用DEAプログラムを簡単に説明する。LINGOは、普通の数式表現の分かり易い自然表記で数理計画法モデルを記述する方法がある。この場合は、一般的な数式表現で理解しやすい。しかし、大規模で複雑なモデルを記述する場合は、以下の集合（SET）表記が便利である。大きくは、SETS:とENDSETSの集合節と、DATA:とENDDATAのDATA節と、その後のCCRモデルを記述したMODEL部分と、その後で最適計算で得られた重みからクロス効率値を計算するCALC節からできている。

集合節では、1次元集合の評価対象を表すDMU（23個の要素をもつ）と入出力の変数を表すFACTOR（3個の2入力1出力の変数）を定義する。SCOREはDMUと同じ属性をもつ1次元配列である。DXFは、行がDMU、列がFACTORで作られる（23行*3列）の2次元集合で、データを2次元配列Fで、各DMUの重みを配列Wで定義している。S40は（23行*23列）の2次元集合で、クロス効率値が2次元配列Sに出力される。

最初のDATA節では、「NINPUTS = 2;」で入力データFの最初の2列が入力であることを指定している。「WGTMIN = .00000001;」は、重みの下限を0に設定している（ $|u| \leq 10^{-6}$ を0と判定している）。「BIGM = 99999;」は重みの上限を99999に制限している。入力データが0を含む場合、0に対応する重みが99999になるが、他のDMUの重みの最大値を調べてその2倍程度に再設定して解くことも考えられる。

「F=@OLE();」は、Excel上の23行3列のセル範囲Fに入力したデータを、@OLE関数でLINGOに入力し、LINGOの配列Fの値にしてLP計算に用いている。

この後、このFの値を用いて23個のDMUに対して各区の図書館の効率値を最大化する23組の重みが計算される。

CALC節では、LPで計算された重みを用いて、SCOREとクロス効率値を計算し、「@OLE()=SCORE, W, S;」で効率値SCORE、重みW、クロス効率値Sを、Excel上の同名のセル範囲に出力している。

重要なことは、Excel上に、F、SCORE、W、Sのセル範囲を指定し、Fに入力データを与え、この汎用プログラムの下線を引いた3カ所の数値を問題に応じて変更して実行すれば良い。これらをExcel上に与えてDATA節で、「DMU, FACTOR, NINPUTS=@OLE();」で入力すれば、プログラムを変更することなく、任意の問題に対応でき統計ソフトと同じ完全な汎用プログ

ラムになるが、その都度変更するのと手間が変わらないのでこちらを採用している。

```

MODEL:
SETS:
    DMU/L.23/:SCORE;
    FACTOR/L.3;
    DXF(DMU, FACTOR) : F, W;
    S40(DMU, DMU) :S;
ENDSETS
DATA:
    NINPUTS =2;WGTMIN = .0000001;BIGM = 9999999;
    F=@OLE();
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();
    @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)) ) ;
        @SOLU();
    @SET('TERSEO', 0) ;
    @OLE() =SCORE, W, S;
ENDCALC
END

```

図2.1 LINGOの汎用DEAプログラム

2.3 DEAクラスタの提案

図2.1の汎用プログラムで、蔵書数と職員数を入力とし、貸出数を出力とした2入力1出力のDEA分析を行うことにする。図2.2のExcel上の「C2:E24」にセル範囲名Fを定義し、入力データを定義する。計算前に、「G2:G24」にセル範囲名SCORE, 「H2:J24」にセル範囲名W, 「K2:AG24」にセル範囲名Sを定義すると計算結果が図のように出力される。

すなわち、汎用DEAプログラムでFを入力し実行すると、図2.2のSCORE, W, Sに効率値, 重み, クロス効率値Sが出力される。実は、クロス効率値SをこれまでのDEA研究では利用してこなかった。今後クロス効率値を統計ソフトで分析すれば、さらに新しい利用法も出てくると考える。

		F 163.523												
	A	B	C	D	E	F	G	H	I	J	K	AC	AD	AG
1	SN	区	蔵書数	職員数	貸出数	登録者数	SCORE	W1	W2	W3	S1	S19	S20	S23
2	1	千代田	163.523	26	105.321	5.561	0.193	0.002	0.025	0.002	0.193	0.181	0.193	0.181
3	2	中央	338.671	30	314.682	18.106	0.470	0.000	0.033	0.001	0.392	0.470	0.392	0.261
4	3	台東	281.655	51	542.349	16.498	0.540	0.004	0.000	0.001	0.529	0.476	0.529	0.540
5	4	荒川	400.993	78	847.872	30.81	0.593	0.002	0.000	0.001	0.553	0.487	0.553	0.593
6	5	港	363.116	69	758.704	57.279	0.586	0.003	0.000	0.001	0.556	0.492	0.556	0.586
7	6	文京	541.658	114	1438.75	66.137	0.745	0.002	0.000	0.001	0.657	0.565	0.657	0.745
8	7	墨田	508.141	61	839.597	35.295	0.616	0.000	0.016	0.001	0.590	0.616	0.590	0.463
9	8	渋谷	338.804	74	540.821	33.188	0.448	0.003	0.000	0.001	0.385	0.327	0.385	0.448
10	9	目黒	511.467	84	1562.27	65.391	0.897	0.001	0.008	0.001	0.897	0.833	0.897	0.857
11	10	豊島	393.815	68	978.117	41.197	0.705	0.001	0.010	0.001	0.705	0.644	0.705	0.697
12	11	新宿	509.682	96	930.437	47.032	0.512	0.002	0.000	0.001	0.488	0.434	0.488	0.512
13	12	中野	527.457	92	1345.19	56.064	0.719	0.001	0.007	0.001	0.719	0.655	0.719	0.715
14	13	品川	601.594	127	1164.8	69.554	0.543	0.002	0.000	0.000	0.478	0.411	0.478	0.543
15	14	北	528.799	96	1348.59	37.467	0.715	0.002	0.000	0.001	0.700	0.629	0.700	0.715
16	15	江東	394.158	77	1100.78	57.727	0.783	0.003	0.000	0.001	0.729	0.640	0.729	0.783
17	16	葛飾	515.624	101	1070.49	46.16	0.582	0.002	0.000	0.001	0.541	0.475	0.541	0.582
18	17	板橋	566.708	118	1707.65	102.967	0.845	0.002	0.000	0.000	0.752	0.648	0.752	0.845
19	18	江戸川	467.617	74	1223.03	47.236	0.787	0.001	0.009	0.001	0.787	0.740	0.787	0.734
20	19	杉並	768.484	103	2299.69	84.51	1.000	0.000	0.010	0.000	1.000	1.000	1.000	0.839
21	20	練馬	669.996	107	1901.47	69.576	0.849	0.001	0.006	0.000	0.849	0.796	0.849	0.796
22	21	足立	844.949	120	1909.7	89.401	0.729	0.000	0.005	0.000	0.729	0.713	0.729	0.634
23	22	大田	1258.98	242	3055.19	97.941	0.681	0.001	0.000	0.000	0.640	0.565	0.640	0.681
24	23	世田谷	1148.86	202	4096.3	191.166	1.000	0.001	0.000	0.000	1.000	0.908	1.000	1.000

図2.2 2入力1出力の分析結果

G列のSCORE(効率値)をみると、杉並区(19)と世田谷区(23)の効率値が1で、参照集合といわれる。W1列は蔵書数、W2は職員数、W3は貸出数の重みである。この重みを用いてK列からAG列の23列にクロス効率値が出力される。

K列は、千代田区の重み(H2:J2)=(0.002, 0.025, 0.002)を用いて計算した23区の効率値である。K2の千代田区の効率値は $0.002 \times 105.321 / (0.002 \times 163.523 + 0.025 \times 26) = 0.193$ で計算できる。千代田区は、自分に最適な重みで計算したにもかかわらず効率値は0.193にしかならない。それは、同じ重みで杉並と世田谷を計算すると1と効率的になるからである。ただし23区全

での重みで計算した効率値 (K2:AG2) の最大値になる。千代田区の重みで計算すると、杉並と世田谷の効率値が1になるので、千代田区はこの2図書館を手本にどこを改善すべきかを考え効率的になることが期待できる。

SCOREの効率値は、S1からS23のクロス効率値の対角要素の値でもある。クロス効率値では、SCOREで効率的であることが分かった2図書館だけの効率値が1になる可能性がある。残りの21図書館の効率値は1にならない。

この2図書館の効率値が1になるパターンを、表2.1のDEAクラスター S1と呼ぶことにする。これに属する図書館は、構成員欄の6図書館である。

S19列は、杉並の重みで計算した効率値であり、杉並だけが効率的であり、表2.1のDEAクラスター S19になる。この構成員は、2, 7, 19の3図書館である。このDEAクラスターには手本の杉並も含まれるので、杉並を手本に残りの2図書館は改善すれば良い。

S23列は世田谷の重みで計算した23図書館の効率値で、世田谷だけが1になる。このパターンを持つ図書館は14あり、世田谷も含まれる。他の13図書館は、世田谷を手本に改善すれば良い。

問題はDEAクラスター S1の手本の杉並と世田谷は、DEAクラスター S1の構成員でないが、6図書館はこれらの2図書館を参考に改善すれば良い。

表2.1 入力1出力のDEAクラスター

DEA	手本	n	構成員
S19	<u>19</u>	3	2, 7, 19
S1	19, 23	6	1, 9, 10, 12, 18, 20
S23	<u>23</u>	14	3-6, 8, 11, 13-17, 21-23

図2.3は、貸出数/蔵書数をX軸に、貸出数/職員数をY軸にプロットした散布図である。杉並と世田谷が参照集合で、原点と杉並を結ぶ直線の左上の三角形がDEAクラスター S19に対応し、杉並を含む3図書館が構成員である。杉並と世田谷と原点を結ぶ三角形がDEAクラスター S1で、杉並と世田谷は6図書館の手本になるが、自らは構成員ではない。原点と世田谷と世田谷からおろした垂線でできる直角三角形がDEAクラスター S23である。世田谷を含む14図書館が構成員で、他の13図書館は世田谷を手本にすればよい。

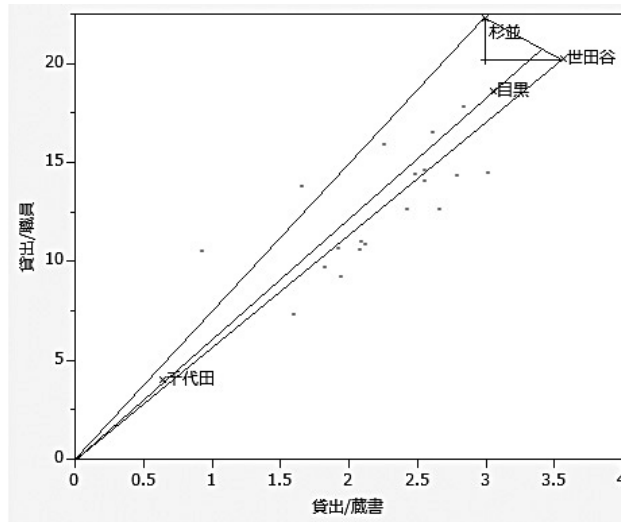


図2.3 貸出数／蔵書数をX軸に、貸出数／職員数をY軸にプロットした散布図

DEAのこれまでの説明では、クラスター S1に含まれる目黒は非効率である。そこで原点から目黒を通る直線と、杉並と世田谷を結ぶ直線の交点が仮想的な手本になる。しかし、現実に応用する場合、そのような仮想的なDMUを考え比較することは普及の妨げになる。実際問題として、杉並と世田谷を手本にして改善した方が分かり易い。

この場合に問題が生じる。目黒は、世田谷と比較してXとYの値を大きな方に改善すれば良い。杉並と比較すると、目黒のXの値が大きい。このような問題点のない1入力固定改善法を提案する。

2.4 改善方法の提案

DEAの一番重要な点は、手本を参考に非効率な評価対象が改善できる点である。しかし、従来は1.4で紹介したCCRモデルの双対モデルを解かないと、手本や改善点が分からなかった。しかし、双対モデルの理解は相当な数理計画法の知識が必要になる。

そこで、入力データFから、LINGOの汎用DEAプログラムを実行して、重みWと各評価対象の効率値SCOREと、各重みで評価対象の効率値を計算したクロス効率値Sというデータだけで、数理計画法の知識なくこれを行う方法を本研究で説明する。

手本はクロス効率値SからDEAクラスターを見つけることで簡単に分かる。そして非効率な評価対象は、入力項目の一つを固定して手本と考える入出力データの構成比になるようにすれば、効率的になる。そして実際の値との差を取ることで、どの項目を増やしたり減らしたりすればよいかの改善点が簡単に分かる（1入力固定改善法）。

2.4.1 2入力2出力

蔵書数と職員数を入力とし、貸出数を出力とする2入力1出力の分析は図2.3で理解できる。以下では、図で理解できない2入力2出力と3入力2出力を取り上げる。

表2.2は、2入力2出力の場合である。板橋、杉並、世田谷が参照集合で、表2.3の4個のDEAクラスターができた。S23の手本は板橋（17）と世田谷（23）であり、クラスターの構成員でもある。残りの4図書館は、板橋と世田谷を手本にして改善すれば良い。

S19は、杉並（19）と世田谷が手本であるが、世田谷が構成員でないので、残りの6図書館は杉並を手本とすればよいであろう。

S1は世田谷がS17は板橋が手本であるので、世田谷と板橋を手本に改善する。

表2.2 2入力2出力の場合

SN 区	蔵書数	職員数	貸出数	登録者数	SCORE	S1	S17	S19	S23
17 板橋	567	118	1708	103	1.000	0.922	1.000	0.752	1.000
19 杉並	768	103	2300	85	1.000	0.867	0.605	1.000	0.750
23 世田谷	1149	202	4096	191	1.000	1.000	0.916	1.000	1.000

表2.3 4個のDEAクラスター

DEA	手本	n	構成員
S1	世田谷	9	1-4, 6, 14, 16, 21, 22
S23	板橋, 世田谷	6	5, 11, 13, 15, 17, 23
S19	杉並, 世田谷	7	7, 9, 10, 12, 18, 19, 20
S17	板橋	1	8

表2.4（上）は、DEAクラスター S1の9個の図書館と最後の行は手本の世田谷を表す。図書館名の後の最初の4列は、現在のデータである。次の4列は、手本の世田谷の2入力と2出力の構成比（1149:202:4096:191）と、各図書館の蔵書数を固定して手本の世田谷の構成比と同じになるように他の3変数の値を比例計算したものである。この値が実現できれば、非効率な8図書館で効率値が全て1になる。最後の4列は、元のデータからこの目標の値を引いたものである。

千代田区、中央区、足立区は職員数が不足しているので、これらを増員して出力の貸出数と登録者数を増やすことができれば増員を検討すればよい。それ以外の6図書館の職員数は多い。例えば大田区では、21人配置転換し、その上で貸し出し数を1434（千冊）、登録者数

を112人増やす努力が必要になる。

表2.4 (下) は、各図書館の職員数を固定して手本の世田谷の構成比と同じになるよう他の3変数の値を計算したものである。千代田区と中央区と足立区は非効率であるにもかかわらず、世田谷より蔵書数が多い。良い点を悪く改変する必要はないであろう。その上で、現状の職員数で貸出数と登録者数を増やす努力が必要である。残り6図書館は蔵書数を増やすことで、貸出数と登録者数を増やすことができるか否かを検討する必要がある。蔵書数が少なく職員数の多い大田区は、職員数を固定して改善目標を考えると多分実行可能領域をはみだし、貸出数を1852冊増やすという達成困難な目標になる。

他のDEAクラスターも同様な方法で改善点分かる。

表2.4 手本との比較 (上:蔵書数を固定, 下:職員数を固定)

SN	区	蔵書数 職員数				蔵書数 職員数				蔵書数 職員数			
		蔵書数	職員数	貸出数	登録者数	蔵書数	職員数	貸出数	登録者数	蔵書数	職員数	貸出数	登録者数
1	千代田	164	26	105	6	164	29	583	27	0	-3	-478	-22
2	中央	339	30	315	18	339	60	1208	56	0	-30	-893	-38
3	台東	282	51	542	16	282	50	1004	47	0	1	-462	-30
4	荒川	401	78	848	31	401	71	1430	67	0	7	-582	-36
6	文京	542	114	1439	66	542	95	1931	90	0	19	-493	-24
14	北	529	96	1349	37	529	93	1885	88	0	3	-537	-51
16	葛飾	516	101	1070	46	516	91	1838	86	0	10	-768	-40
21	足立	845	120	1910	89	845	149	3013	141	0	-29	-1103	-51
22	大田	1259	242	3055	98	1259	221	4489	209	0	21	-1434	-112
23	世田谷	1149	202	4096	191	1149	202	4096	191	0	0	0	0

SN	区	蔵書数 職員数				蔵書数 職員数				蔵書数 職員数			
		蔵書数	職員数	貸出数	登録者数	蔵書数	職員数	貸出数	登録者数	蔵書数	職員数	貸出数	登録者数
1	千代田	164	26	105	6	148	26	527	25	16	0	-422	-19
2	中央	339	30	315	18	171	30	608	28	168	0	-294	-10
3	台東	282	51	542	16	290	51	1034	48	-8	0	-492	-32
4	荒川	401	78	848	31	444	78	1582	74	-43	0	-734	-43
6	文京	542	114	1439	66	648	114	2312	108	-107	0	-873	-42
14	北	529	96	1349	37	546	96	1947	91	-17	0	-598	-53
16	葛飾	516	101	1070	46	574	101	2048	96	-59	0	-978	-49
21	足立	845	120	1910	89	682	120	2433	114	162	0	-524	-24

22 大田	1259	242	3055	98	1376	242	4907	229	-117	0	-1852	-131
23 世田谷	1149	202	4096	191	1149	202	4096	191	0	0	0	0

2.4.2 3入力2出力の場合

表2.5は、3入力2出力の場合である。板橋、杉並、世田谷の3図書館が手本(参照集合)になり、表下のS7からS23の5個のDEAクラスターができた。

表2.5 3入力2出力の場合

SN 区	床面積	蔵書数	職員数	貸出数	登録者数	W1	W2	W3	W4	W5
17 板橋	10866	567	118	1708	103	0.00E+00	1.76E-03	0.00E+00	1.84E-04	6.65E-03
19 杉並	11469	768	103	2300	85	0.00E+00	0.00E+00	9.71E-03	4.35E-04	0.00E+00
23 世田谷	10888	1149	202	4096	191	0.00E+00	0.00E+00	4.95E-03	0.00E+00	5.23E-03

SN 区	S7	S8	S17	S19	S23
17 板橋	0.804	1.000	1.000	0.648	0.922
19 杉並	1.000	0.605	0.727	1.000	0.867
23 世田谷	1.000	0.916	1.000	0.908	1.000

表2.6は、S7からS23の5個のDEAクラスターの詳細である。DEAの問題点の一つは、変数が増えるとDEAクラスターが増える点である。DEAクラスターS23、S17とS19には、それぞれ参照集合の図書館が含まれている。これらのクラスターは、この参照集合を手本に改善をはかれば良いだろう。S7とS8は、参照集合はクラスターに含まれていない。S7の参照集合は杉並(19)と世田谷(23)であるが、構成員数の多いS23と融合しU23とする。S8の参照集合は板橋(17)でS17と融合し、U17とする。

表2.6 3入力2出力の場合の5個のDEAクラスターである

融合	DEA	手本	n	構成員
U23	S23	<u>23</u>	10	1-4, 6, 14, 16, 21-23
	S7	19, 23	6	7, 9, 10, 12, 18, 20
U17	S17	<u>17, 23</u>	5	5, 11, 13, 15, 17
	S8	17	1	8
S19	S19	<u>19</u>	1	19

図2.4はクロス効率値を階層型の群平均法で分析した結果である。左からの4変数は表2.6のDEAクラスター S23の一部に対応している。次のS5からS15の6変数は、DEAクラスター U17の6変数に対応している。3番目のS3からS16はDEAクラスター S23の一部に対応している。4番目のS7からS20はDEAクラスター S7の6変数とDEAクラスター S19の1変数に対応している。最後のS22は、DEAクラスター S23の構成員である。以上から、S19はU23に融合し、U23とU17の2つの融合DEAクラスターで考えることもありうる。

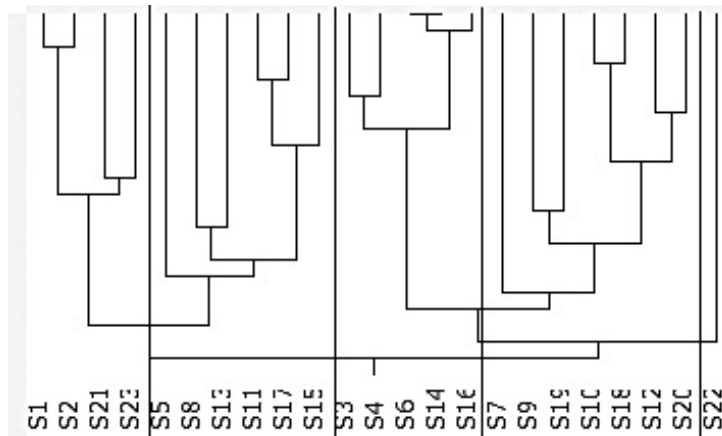


図2.4 クロス効率値を階層型の群平均法で分析した結果

表2.7は、3個の参照集合で、改善点の見つけ方の考え方を再び確認する。ここでは世田谷を手本とすることを考える。残りの板橋と杉並の構成比を世田谷に比例させる。そのために入力変数の一つ決めてその値を固定し、世田谷と同じ構成比になるようにする。3列から7列は実際のデータである。次の5列は、床面積を固定して、世田谷の構成比に比例させた改善目標値である。非効率なDMUは、残りの入力と出力の値をこの構成比と比較して改善すれば良い。次の5列は、実際の値からこの改善目標値を引いたものである。床面積を固定しているため0である。板橋と杉並の蔵書数、職員数の入力と貸出数と登録者数の値が負である。板橋と杉並は効率的であるが、世田谷に比べ蔵書数と職員数が少ないので、この絶対値だけ改善することで貸出数と登録者数の絶対値だけ改善できれば、世田谷並みにさらに効率的になる。すなわち、効率的な参照集合にも階層があることが分かる。

杉並区がこの改善目標を達成できない場合、床面積が広すぎて実行可能領域をはみだしていることが考えられる。表2.8から4185 (千 m^2) から5917 (千 m^2) のフロアを貸会議室に転用することが考えられる。

表2.7 構成比

SN 区	固定時					変動時					変動時				
	床面積	蔵書数	職員数	貸出数	登録者数	床面積	蔵書数	職員数	貸出数	登録者数	床面積	蔵書数	職員数	貸出数	登録者数
17 板橋	10866	567	118	1708	103	10866	1146.5	201.6	4088.0	190.8	0	-579.8	-83.6	-2380.4	-87.8
19 杉並	11469	768	103	2300	85	11469	1210.2	212.8	4314.9	201.4	0	-441.7	-109.8	-2015.2	-116.9
23 世田谷	10888	1149	202	4096	191	10888	1148.9	202.0	4096.3	191.2	0	0	0	0	0

表2.8は、蔵書数と職員数で固定した結果ある。蔵書数で固定した場合（左）、板橋区は床面積も職員数も多く、登録者数は多いが貸出数が少ない問題点分かる。杉並区は、床面積が広い割に職員が少なく、貸出数も登録者数も少ない。職員数を約32人増やすことで、貸出数を440（千冊）、登録者数を43（千人）増やすことができるか検討すべきである。

職員数を固定すると（右）、板橋区は床面積が多い割に蔵書数が少なく、出力も少ない。蔵書数を104（千冊）増やすことで出力が改善されるか検討すべきである。杉並区は、床面積が広すぎるが、その分蔵書数も多いので、貸出数も多い。ただし、登録者数を改善すべきである。

表2.8 構成比（蔵書数と職員数で固定）

SN 区	蔵書数で固定					職員数で固定				
	床面積	蔵書数	職員数	貸出数	登録者数	床面積	蔵書数	職員数	貸出数	登録者数
17 板橋	5495.197	0	18.35799	-312.967	8.669158	4505.7	-104.4	0	-685.2	-8.7
19 杉並	4185.926	0	-32.1195	-440.355	-43.3625	5917.2	182.7	0	211.0	-13.0
23 世田谷	0	0	0	0	0	0	0	0	0	0

表2.9で、U23の16図書館にS19の杉並を加えた改善点を考える。最初の4列は、床面積を固定したものであり、全て0なので省いてある。大田区の蔵書数と職員数だけが正であり、残りの値はすべて負である。大田区は、蔵書数を減らす必要はなく職員数を減らしたうえで、貸出数と登録者数を増やす必要がある。他の図書館は、蔵書数と職員数を増やすことで貸出数と登録者数が増やせるかどうか検討すべきである。

次の4列は、蔵書数を固定した結果である。大田区は床面積が少ないが職員数が多く、出力不足である。千代田区、中央区、墨田区、目黒区、豊島区、中野区、江戸川区、杉並区、練馬区、足立区は、床面積の箱モノが過大で、職員数が少なく、出力が伴っていない。その他は、床面積も職員数も多く、出力が伴っていないことが分かる。

最後の4列は、職員数を固定した結果である。大田区は入力と出力の両方が不足である。千代田区、中央区、墨田区、目黒区、豊島区、中野区、江戸川区、杉並区、練馬区、足立区は、

床面積の箱モノが過大で、蔵書数が多く、出力が伴っていない。床面積も蔵書数も削減しても意味がないので、現状で蔵書数の多さをアピールし出力の改善が必要である。その他の図書館は、床面積が多い割に蔵書集が少なく、出力が伴っていないことが分かる。魅力ある蔵書を増やして出力が改善できるかが課題である。

表2.9 U23の16図書館と杉並の改善点 (床面積と蔵書数と職員数で固定)

SN	区	蔵書数	職員数	貸出数	登録者数	床面積	職員数	貸出数	登録者数	床面積	蔵書数	貸出数	登録者数
1	千代田	-73.8	-15.7	-740.8	-33.9	699.3	-2.8	-477.7	-21.6	847.6	15.6	-421.9	-19.0
2	中央	-148.5	-55.7	-1422.3	-63.0	1407.3	-29.5	-892.9	-38.2	3000.0	168.0	-293.7	-10.3
3	台東	-127.0	-20.9	-914.8	-51.5	1203.7	1.5	-461.9	-30.4	1124.0	-8.4	-491.9	-31.8
4	荒川	-183.7	-24.8	-1236.8	-66.5	1740.7	7.5	-581.9	-35.9	1336.7	-42.6	-733.9	-43.0
6	文京	-522.6	-73.1	-2355.8	-110.9	4952.6	18.8	-492.5	-24.0	3941.3	-106.7	-873.0	-41.7
7	墨田	-65.2	-39.8	-1204.8	-60.1	618.2	-28.3	-972.2	-49.3	2146.0	161.2	-397.4	-22.4
9	目黒	-24.2	-10.2	-347.8	-23.7	229.7	-5.9	-261.4	-19.7	549.3	33.7	-141.1	-14.1
10	豊島	-347.9	-62.4	-1666.3	-82.2	3296.7	-1.2	-426.0	-24.3	3363.7	7.1	-400.8	-23.2
12	中野	-218.8	-39.2	-1315.5	-68.1	2073.2	-0.7	-535.5	-31.7	2113.1	4.2	-520.5	-31.0
14	北	-292.2	-48.4	-1578.8	-99.1	2769.5	3.0	-536.9	-50.5	2606.5	-17.2	-598.2	-53.4
16	葛飾	-602.1	-95.5	-2914.8	-139.8	5706.3	10.3	-768.0	-39.6	5149.0	-58.8	-977.7	-49.4
18	江戸川	-218.2	-46.6	-1222.4	-66.9	2068.3	-8.2	-444.3	-30.6	2511.3	46.7	-277.6	-22.8
19	杉並	-441.7	-109.8	-2015.2	-116.9	4185.9	-32.1	-440.4	-43.4	5917.2	182.7	211.0	-13.0
20	練馬	-476.8	-94.6	-2187.3	-121.2	4518.3	-10.8	-487.4	-41.9	5100.6	61.4	-268.4	-31.7
21	足立	-285.9	-78.8	-2122.3	-98.8	2709.3	-28.6	-1103.0	-51.2	4248.9	162.5	-523.7	-24.2
22	大田	128.3	43.2	-976.4	-90.2	-1215.6	20.6	-1433.7	-111.5	-2328.0	-117.4	-1852.3	-131.1
23	世田谷	0	0	0	0	0	0	0	0	0	0	0	0

表2.10で、U17の6図書館を板橋区を手本に改善点を考える。最初の4列は、床面積を固定したものである。港区、渋谷区、新宿区は、すべてが負である。床面積に対し、蔵書数と職員数が少ない。これらを増やすことで、出力が改善できるか問題である。品川区は蔵書数と職員数が多いのに、出力が伴ってなく問題である。江東区は、登録者数だけが負である。現状維持で、登録者数を増やす必要がある。

次の4列は、蔵書数を固定した結果である。港区と新宿区は床面積が過大で職員数が不足している。渋谷区は床面積が広く職員数も多いのに出力不足である。品川区は、床面積が狭いが職員数が多く、出力不足である。江東区は床面積を分館などで増やし職員数を増やすことで、出力を改善できるかが課題である。

最後の4列は、職員数を固定した結果である。港区と新宿区は床面積と蔵書数が多いが、

出力が不足している。渋谷区は床面積が広く蔵書数が少なく、出力不足である。品川区は、床面積が狭く蔵書数も少なく、出力不足である。江東区は床面積が狭いが蔵書数が多いのに、出力を改善することが課題である。分館などを増やして出力が改善できるかが課題である。

表2.10 U17の6図書館の改善点 (床面積と蔵書数と職員数で固定)

SN	区	蔵書数	職員数	貸出数	登録者数	床面積	職員数	貸出数	登録者数	床面積	蔵書数	貸出数	登録者数
5	港	-230.5	-54.6	-1029.9	-50.6	4418.7	-6.60805	-335.463	-8.69671	5027.2	31.7	-239.8	-2.9
8	渋谷	-53.6	-7.7	-641.6	-38.1	1027.8	3.454195	-480.087	-28.3704	709.7	-16.6	-530.1	-31.4
11	新宿	-70.3	-24.8	-817.3	-58.4	1348.4	-10.126	-605.373	-45.5738	2280.9	48.6	-458.8	-36.7
13	品川	114.1	25.5	-304.3	-19.0	-2186.9	1.736033	-647.965	-39.7515	-2346.8	-8.3	-673.1	-41.3
15	江東	69.0	9.3	120.9	-1.4	-1322.5	-5.07162	-86.926	-13.8888	-855.5	24.4	-13.5	-9.5
17	板橋	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0

3 学生の成績データ

学生の成績データは、成績を目的変数として、勉強時間、飲酒日数、支出、喫煙の有無、性別を説明変数とする重回帰分析で扱える回帰型DEAデータである。このようなデータでは、重回帰分析による分析を事前に行い、その情報を参考にすべきである。

3.1 統計分析

表3.1は、逐次変数増加法の結果である。勉強時間、飲酒日数、支出、喫煙の有無、性別の順に回帰モデルに取り込まれた。Cp統計量、AIC、BICは2変数モデルを選んだ。

表3.1 逐次変数選択法の結果

Var.	Cp	AIC	BIC
1 勉強時間	13.57	304.08	308.48
2 飲酒日数	<u>2.54</u>	<u>294.36</u>	<u>299.97</u>
3 支出	2.91	295.15	301.83
4 喫煙の有無	4.42	297.36	304.95
5 性別	6.00	299.83	308.15

表3.2は、5個の説明変数による重回帰分析の結果である。勉強時間の回帰係数は正であるが、それ以外は負であることが示すように。目的変数とこれら4変数は負の相関がある。

表3.2 5個の説明変数による重回帰分析の結果

	B	SE	t	p
切片	75.45	7.36	10.25	<.0001
性別	-2.02	3.12	-0.65	0.52
勉強時間	2.26	0.72	3.15	0.00
支出	-1.67	1.23	-1.36	0.18
喫煙の有無	-2.80	3.41	-0.82	0.42
飲酒日数	-2.50	1.43	-1.75	0.09

図3.1は、変数増加法で選ばれた勉強時間と飲酒日数の散布図である。合格群（70点以上）は、勉強時間が0時間から12時間までX軸方向に広く布置している。不合格群は、勉強時間は6時間以下で、飲酒日数が1日から7日までY軸方向にばらついている。

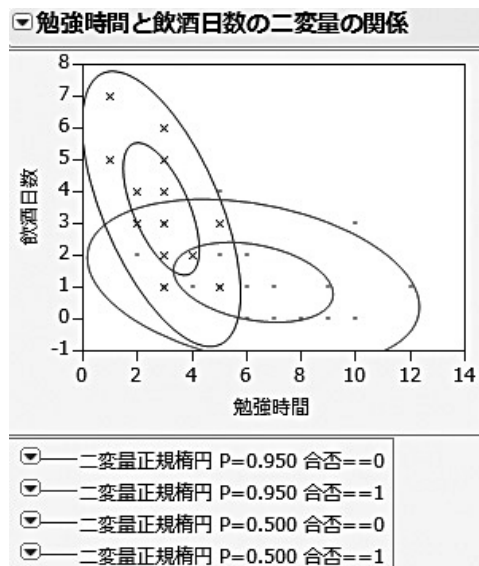


図3.1 変数増加法で選ばれた勉強時間と飲酒日数の散布図

3.2 飲酒の扱い

飲酒は、0から7の数値である。DEAの入力は正であることが期待されているので、1を足した新変数の飲酒1で分析し、表3.3が得られた。参照集合は6名である。そして、表3.4の6個のDEAクラスターが得られた。

表3.3 重みの下限が0

SN	飲酒1	勉強	成績	W1	W2	W3	S10	S2	S3	S12	S29	S31
2	1	9	100	1	0.000	0.010	0.876	1.000	1.000	0.541	0.328	0.185
3	1	7	95	1	0.026	0.011	1.000	0.950	1.000	0.646	0.400	0.226
10	2	3	85	0.332	0.112	0.012	1.000	0.425	0.523	1.000	0.781	0.472
12	3	2	80	0.175	0.238	0.013	0.772	0.267	0.337	1.000	1.000	0.667
31	6	1	60	0.000	0.999	0.017	0.336	0.100	0.128	0.583	1.000	1.000
34	8	1	60	0.000	0.999	0.017	0.255	0.075	0.096	0.458	0.871	1.000

表3.4 6個のDEAクラスター

融合	DEA	手本	n	構成員
S10	S10	3, 10	18	1, 8-11, 13-18, 20-24, 30, 32
S2	S2	2	2	2, 4
S3	S3	2, 3	4	3, 5-7
U12	S12	10, 12	8	12, 19, 25-28, 33, 38
	S29	12, 31	6	35-37, 39, 40, 29
S31	S31	31, 34	2	31, 34

成績／勉強時間をX軸に、成績／飲酒1をY軸にプロットしたのが図3.2である。図の参照集合を表す番号は学生の成績に対応している。表3.4のDEAクラスター S2は、原点と100点の左の直角三角形に対応している。DEAクラスター S3は、原点と100点と95点の三角形に対応している。DEAクラスター S10は、原点と95点と85点の三角形に対応していて構成員が8人と一番多い。成績が100点だが、勉強時間が12時間と一番多い1番目の学生は、効率値が0.585と非効率で図の×で示す。飲酒日数が少ない猛勉強型であれば、上のS2かS3の構成員になる。X軸の値が8前後に位置している。経営の指標と違って、教育的には勉強時間の多いことを非難できないが、もう少し効率的に行うべきであろう。

DEAクラスター S31は、原点と60点とそこからX軸に垂線を下ろした直角三角形に対応し、構成員の31と34番目の学生は手本でもある。しかし、自分自身以外の追従者がいない。

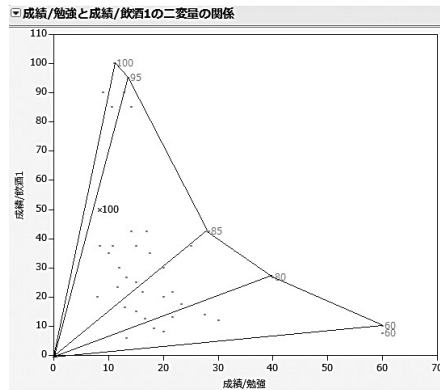


図3.2 重みの下限が0.001

経営データの場合は勉強時間を作業時間、成績を利益に読み替えれば、DEAクラスターS31には、多くの追従者もできて効率性の意味も良く理解できる。すなわち、重回帰型データでもDEA法を利用すべきと考えているが、解釈には注意する必要がある。

3.3 飲酒の扱い2

入力値が0を含む場合、0を小さな正の実数に置き換えた分析も行われている。0のまま分析すると、表3.5のように飲酒日数が0に対応した重みのW1の値(5列目)が、重みの上限に設定した99999になる。そして例えば2番目の学生の重みで計算したS2列とS3列の効率値(10列目と11列目)は、飲酒日数が0以外の正になる学生の効率値が0になってしまう。そこで5列のW1の99999以外の最大値を調べると0.5であるので、汎用プログラムの重みの最大値を2倍の1に設定して(BIGM=1)解くと、SCORE2で始まる最後の4列の計算結果になる。重みの上限を99999に設定したものと1に設定したものの効率値SCOREとSCORE2は同じである。ただし、クロス効率値で0になったものだけが非零に変更される。

表3.5 入力が0の問題

SN	飲酒 日数	勉強	成績	W1	W2	W3	SCORE	S1	S2	S3	SCORE2	S1	S2	S3
1	1	12	100	0.20	0.07	0.00	0.471	0.47	0.00	0.00	0.471	0.47	0.34	0.37
2	0	9	100	99999	0.11	0.01	0.784	0.78	0.78	0.78	0.784	0.78	0.78	0.78
3	0	7	95	99999	0.14	0.01	0.958	0.96	0.96	0.96	0.958	0.96	0.96	0.96
4	0	10	90	99999	0.10	0.01	0.635	0.64	0.64	0.64	0.635	0.64	0.64	0.64
5	0	7	90	99999	0.14	0.01	0.908	0.91	0.91	0.91	0.908	0.91	0.91	0.91

6	0	7	90	99999	0.14	0.01	0.908	0.91	0.91	0.91	0.908	0.91	0.91	0.91	
7	0	8	85	99999	0.13	0.01	0.750	0.75	0.75	0.75	0.750	0.75	0.75	0.75	
8	1	5	85		0.38	0.13	0.01	0.750	0.75	0.00	0.00	0.750	0.75	0.43	0.50
9	1	6	85		0.33	0.11	0.01	0.667	0.67	0.00	0.00	0.667	0.67	0.40	0.46
10	1	3	85		0.50	0.17	0.01	1.000	1.00	0.00	0.00	1.000	1.00	0.50	0.60
11	0	6	85	99999	0.17	0.01	1.000	1.00	1.00	1.00	1.000	1.00	1.00	1.00	
12	2	2	80		0.22	0.28	0.01	1.000	0.71	0.00	0.00	1.000	0.71	0.28	0.35
13	3	10	80		0.16	0.05	0.00	0.297	0.30	0.00	0.00	0.297	0.30	0.15	0.18
14	2	6	80		0.25	0.08	0.01	0.471	0.47	0.00	0.00	0.471	0.47	0.24	0.28
15	1	9	75		0.25	0.08	0.01	0.441	0.44	0.00	0.00	0.441	0.44	0.29	0.33
16	1	5	75		0.38	0.13	0.01	0.662	0.66	0.00	0.00	0.662	0.66	0.38	0.44
17	1	3	75		0.50	0.17	0.01	0.882	0.88	0.00	0.00	0.882	0.88	0.44	0.53
18	2	5	75		0.12	0.15	0.01	0.508	0.48	0.00	0.00	0.508	0.48	0.23	0.28
19	4	5	75		0.10	0.12	0.01	0.411	0.31	0.00	0.00	0.411	0.31	0.13	0.16
20	1	7	75		0.30	0.10	0.01	0.529	0.53	0.00	0.00	0.529	0.53	0.33	0.38
21	1	4	70		0.43	0.14	0.01	0.706	0.71	0.00	0.00	0.706	0.71	0.38	0.45
22	1	7	70		0.30	0.10	0.01	0.494	0.49	0.00	0.00	0.494	0.49	0.31	0.35
23	2	6	70		0.25	0.08	0.01	0.412	0.41	0.00	0.00	0.412	0.41	0.21	0.25
24	1	4	70		0.43	0.14	0.01	0.706	0.71	0.00	0.00	0.706	0.71	0.38	0.45
25	3	3	70		0.15	0.19	0.01	0.583	0.41	0.00	0.00	0.583	0.41	0.16	0.21
26	2	4	65		0.14	0.18	0.01	0.520	0.46	0.00	0.00	0.520	0.46	0.21	0.25
27	2	3	65		0.17	0.22	0.01	0.634	0.51	0.00	0.00	0.634	0.51	0.22	0.27
28	3	5	65		0.11	0.14	0.01	0.394	0.33	0.00	0.00	0.394	0.33	0.14	0.18
29	4	3	65		0.04	0.28	0.01	0.520	0.31	0.00	0.00	0.520	0.31	0.12	0.15
30	1	3	60		0.50	0.17	0.01	0.706	0.71	0.00	0.00	0.706	0.71	0.35	0.42
31	5	1	60		0.00	1.00	0.02	1.000	0.26	0.00	0.00	1.000	0.26	0.09	0.12
32	1	5	60		0.38	0.13	0.01	0.529	0.53	0.00	0.00	0.529	0.53	0.30	0.35
33	2	3	60		0.17	0.22	0.01	0.585	0.47	0.00	0.00	0.585	0.47	0.20	0.25
34	7	1	60		0.00	1.00	0.02	1.000	0.19	0.00	0.00	1.000	0.19	0.07	0.08
35	4	2	60		0.06	0.39	0.01	0.667	0.30	0.00	0.00	0.667	0.30	0.11	0.14
36	3	2	55		0.06	0.41	0.01	0.647	0.35	0.00	0.00	0.647	0.35	0.13	0.17
37	5	3	55		0.04	0.27	0.01	0.423	0.22	0.00	0.00	0.423	0.22	0.08	0.10
38	3	3	50		0.15	0.19	0.01	0.417	0.29	0.00	0.00	0.417	0.29	0.12	0.15
39	6	3	40		0.04	0.26	0.01	0.296	0.13	0.00	0.00	0.296	0.13	0.05	0.06
40	4	2	40		0.06	0.39	0.01	0.444	0.20	0.00	0.00	0.444	0.20	0.07	0.09

表3.6は、重みの上限を1にして分析した結果で、5人が効率的になった。そして5個のDEAクラスターができた。表3.3の2番と3番が非効率になり。11番が効率的になった。

表3.6 6個のDEAクラスター

SN	飲酒 日数	勉強	成績	W1	W2	W3	SCORE	S10	S11	S12	S29	S31
10	1	3	85	0.50	0.17	0.01	1.000	1.00	0.67	1.00	0.77	0.47
11	0	6	85	1.00	0.17	0.01	1.000	1.00	1.00	0.63	0.40	0.24
12	2	2	80	0.22	0.28	0.01	1.000	0.71	0.40	1.00	1.00	0.67
31	5	1	60	0.00	1.00	0.02	1.000	0.26	0.14	0.55	1.00	1.00
34	7	1	60	0.00	1.00	0.02	1.000	0.19	0.10	0.41	0.86	1.00

表3.7は、5個のDEAクラスターの詳細である。S29はS12に融合し、15人のU12にする。

表3.7 5個のDEAクラスター

DEA	手本	n	構成員
S10	<u>10</u> , 11	16	1, 8-10, 13-17, 20-24, 30, 32
S11	<u>11</u>	7	2-7, 11
S12	10, <u>12</u>	9	12, 18, 19, 25-28, 33, 38
S29	12, 31	6	29, 35-37, 39, 40
S31	<u>31</u> , <u>34</u>	2	31, 34

図3.3はクロス効率値を階層型クラスター分析の群平均法の分析結果である。左から1番目のS1からS24の16変数は、DEAクラスター S10に対応している。2番目のS2からS11の7変数はDEAクラスター S11に対応している。3番目のS12からS33の9変数はDEAクラスター S12に対応している。4番目のS29からS39の6変数はDEAクラスター S29に対応している。最後のS31とS34の2変数はDEAクラスター S31に対応している。

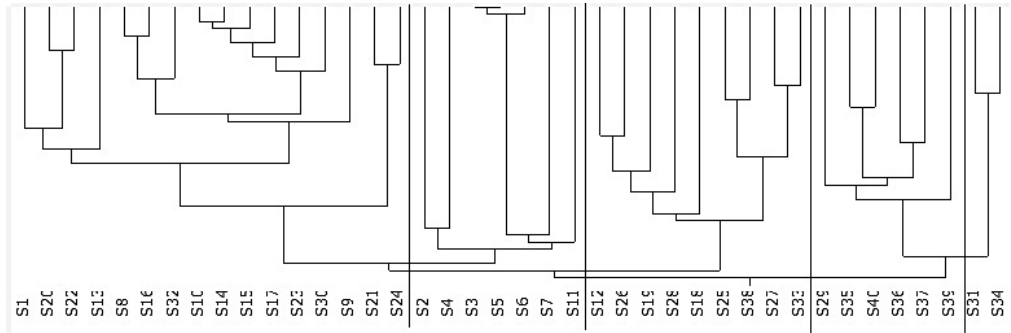


図3.3 クロス効率値を階層型クラスター分析の群平均法の分析結果

以上から入力で0を含む場合に、1足して0をなくすか、0を小さな値に変換するか、重みの上限值の変更で対応する方法があり、結果は微妙に異なってくる。どの方法が良いかは今後の検討課題である。

表3.8はクラスター S10を10番目の学生を手本にして、構成比で改善点を考えたものである。成績のように下側と上側に0と100点の打ち切りがあるデータに適用すると、100点以上の改善が要求されるという実行可能領域をはみだす問題が明らかになる。問題点の明らかこれらの学生の指導に使えるかもしれない。

表3.8 クラスター S10の改善

SN	SCORE	飲酒 日数	勉強 成績	飲酒 日数	勉強 成績	飲酒 日数	勉強 成績	飲酒 日数	勉強 成績	飲酒 日数	勉強 成績		
1	0.471	1.0	3.0	85.0	0	9.0	15.0	4.0	12.0	340.0	-3.0	0.0	-240.0
8	0.750	1.0	3.0	85.0	0	2.0	0.0	1.7	5.0	141.7	-0.7	0.0	-56.7
9	0.667	1.0	3.0	85.0	0	3.0	0.0	2.0	6.0	170.0	-1.0	0.0	-85.0
13	0.297	3.0	9.0	255.0	0	1.0	-175.0	3.3	10.0	283.3	-0.3	0.0	-203.3
14	0.471	2.0	6.0	170.0	0	0.0	-90.0	2.0	6.0	170.0	0.0	0.0	-90.0
15	0.441	1.0	3.0	85.0	0	6.0	-10.0	3.0	9.0	255.0	-2.0	0.0	-180.0
16	0.662	1.0	3.0	85.0	0	2.0	-10.0	1.7	5.0	141.7	-0.7	0.0	-66.7
17	0.882	1.0	3.0	85.0	0	0.0	-10.0	1.0	3.0	85.0	0.0	0.0	-10.0
20	0.529	1.0	3.0	85.0	0	4.0	-10.0	2.3	7.0	198.3	-1.3	0.0	-123.3
21	0.706	1.0	3.0	85.0	0	1.0	-15.0	1.3	4.0	113.3	-0.3	0.0	-43.3
22	0.494	1.0	3.0	85.0	0	4.0	-15.0	2.3	7.0	198.3	-1.3	0.0	-128.3
23	0.412	2.0	6.0	170.0	0	0.0	-100.0	2.0	6.0	170.0	0.0	0.0	-100.0

24	0.706	1.0	3.0	85.0	0	1.0	-15.0	1.3	4.0	113.3	-0.3	0.0	-43.3
30	0.706	1.0	3.0	85.0	0	0.0	-25.0	1.0	3.0	85.0	0.0	0.0	-25.0
32	0.529	1.0	3.0	85.0	0	2.0	-25.0	1.7	5.0	141.7	-0.7	0.0	-81.7
10	1.000	1.0	3.0	85.0	0	0	0	1.0	3.0	85.0	0.0	0.0	0.0

飲酒日数を固定した場合、1番の学生は勉強時間が多いことで成績も15点アップしているがもっと上げないと非効率ということだ。8番と9番は勉強時間が多いことで、やっと目標を達成している。13番から32番の12人は、勉強時間が足りているのに、成績が悪く指導の対象になる。33番と38番はもっと勉強し成績を上げるように指導すべきである。

3.4 飲酒日数の扱い3

DEAの大きな問題点の一つは、入力変数が正という条件があることである。そこで、成績を入力とし、飲酒日数と勉強時間を出力とする1入力2出力モデルを考えることで解決できる。これは、今後取りあげる日本車44車種の機能でもって価格を予測することを考えた場合、各機能の価格比で性能／価格を論じることに対応している。表3.9に示す通り、手本は2人で3個のDEAクラスターが得られた。

表3.9 5個のDEAクラスター

DEA	手本	n	構成員
S13	13, 39	21	10, 12-14, 17-19, 23, 25-30, 33, 35-40
S1	13	17	1-9, 11, 15, 16, 20-22, 24, 32
S31	39	2	31, 34

図3.4は、クロス効率値Sを階層型クラスター分析の群平均法の変数のクラスター分析の結果である。左のS1からS24はDEAクラスターS1、次のS10からS38の21人はDEAクラスターS13、最後はS31に対応している。

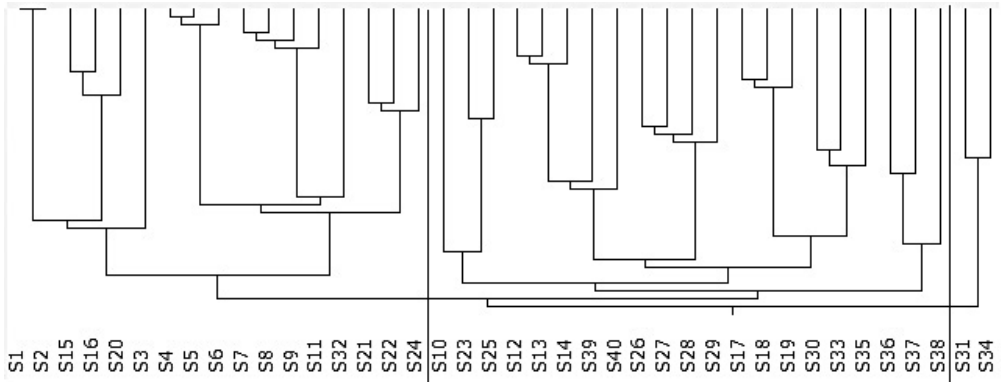


図3.4 クロス効率値Sを階層型クラスター分析の群平均法の変数のクラスター分析

表3.10の2列から4列は入力データFで、次の3列は望ましくない13番目の学生の構成比で他の39人の学生を1入力固定改善法による結果である。望ましくない学生を手本にしているので、1番から18番、20番から24番、26、27、30、32、33番の28名の学生は、飲酒日数と勉強時間が負であるが、34番の学生よりこの絶対値だけ飲酒日数と勉強時間が少ない望ましい学生であることを示す。表4.2で手本になった10番の学生は34番の学生に比べて飲酒日数は2.2日／週少なく、勉強時間が7.6時間少なく効果的に勉強していると解釈することになる。

19、25、29、31、34から40番の11名の学生は、飲酒日数が正で、成績は負である。特に34番以降の成績の悪い学生の飲酒日数が正の値が大きいので、この日数だけ減らして勉強時間に振り向けるよう指導すべきであろう。

表3.10 改善点の検討

SN	成績	飲酒日数	勉強	成績	飲酒日数	勉強	成績	飲酒日数	勉強
1	100	1	12	0	-2.8	-0.5	0	-14.0	4.5
2	100	0	9	0	-3.8	-3.5	0	-15.0	1.5
3	95	0	7	0	-3.6	-4.9	0	-14.3	-0.1
4	90	0	10	0	-3.4	-1.3	0	-13.5	3.3
5	90	0	7	0	-3.4	-4.3	0	-13.5	0.3
6	90	0	7	0	-3.4	-4.3	0	-13.5	0.3
7	85	0	8	0	-3.2	-2.6	0	-12.8	1.6
8	85	1	5	0	-2.2	-5.6	0	-11.8	-1.4
9	85	1	6	0	-2.2	-4.6	0	-11.8	-0.4

10	85	1	3	0	-2.2	-7.6	0	-11.8	-3.4
11	85	0	6	0	-3.2	-4.6	0	-12.8	-0.4
12	80	2	2	0	-1.0	-8.0	0	-10.0	-4.0
14	80	2	6	0	-1.0	-4.0	0	-10.0	0.0
15	75	1	9	0	-1.8	-0.4	0	-10.3	3.4
16	75	1	5	0	-1.8	-4.4	0	-10.3	-0.6
17	75	1	3	0	-1.8	-6.4	0	-10.3	-2.6
18	75	2	5	0	-0.8	-4.4	0	-9.3	-0.6
19	75	4	5	0	1.2	-4.4	0	-7.3	-0.6
20	75	1	7	0	-1.8	-2.4	0	-10.3	1.4
21	70	1	4	0	-1.6	-4.8	0	-9.5	-1.3
22	70	1	7	0	-1.6	-1.8	0	-9.5	1.8
23	70	2	6	0	-0.6	-2.8	0	-8.5	0.8
24	70	1	4	0	-1.6	-4.8	0	-9.5	-1.3
25	70	3	3	0	0.4	-5.8	0	-7.5	-2.3
26	65	2	4	0	-0.4	-4.1	0	-7.8	-0.9
27	65	2	3	0	-0.4	-5.1	0	-7.8	-1.9
28	65	3	5	0	0.6	-3.1	0	-6.8	0.1
29	65	4	3	0	1.6	-5.1	0	-5.8	-1.9
30	60	1	3	0	-1.3	-4.5	0	-8.0	-1.5
31	60	5	1	0	2.8	-6.5	0	-4.0	-3.5
32	60	1	5	0	-1.3	-2.5	0	-8.0	0.5
33	60	2	3	0	-0.3	-4.5	0	-7.0	-1.5
34	60	7	1	0	4.8	-6.5	0	-2.0	-3.5
35	60	4	2	0	1.8	-5.5	0	-5.0	-2.5
36	55	3	2	0	0.9	-4.9	0	-5.3	-2.1
37	55	5	3	0	2.9	-3.9	0	-3.3	-1.1
38	50	3	3	0	1.1	-3.3	0	-4.5	-0.8
40	40	4	2	0	2.5	-3.0	0	-2.0	-1.0
39	40	6	3	0	4.5	-2	0	0	0
13	80	3	10	0	0	0	0	-9	4

最後の3列は、39番目の成績が40点と最低で、成績が最低点で飲酒日数が6日／週で勉強時間が3時間と問題の多い学生を手本とした分析結果である。飲酒日数と勉強時間の両方が

負の25名の学生は、問題がない学生である。残りの15名の学生は、40点の学生に比べて勉強時間／得点の比で勉強時間が多い学生である。特に1番目の学生は、39番目の学生が40点取るのに3時間勉強しているので、もし比例配分が正しければ100点取るのに7.5時間の目標が決まり4.5時間多く勉強し、15番目の学生は3.4時間と多い学生である。

4. まとめ

本研究では、DEA法が広く社会で使われるための以下の提言を行った。

- 1) 使いやすくして問題のサイズに依存しない汎用DEAプログラムの提供が重要である。
- 2) これまでのDEA研究では、参照集合はCCRモデルの双対問題に変換して求めていた。このことが、企業などの線形計画法に堪能でない潜在ユーザーの利用を妨げる障害であった。一方、汎用DEAプログラムが与えられた評価対象のデータFを入力して実行すれば、最適な重みWとクロス効率値Sが簡単に得られる。すなわち、 $S=F*W$ で得られるクロス効率値のデータから、効率値が1になるものを探せば簡単に各評価対象の参照集合が分かる。
- 3) また、2入力1出力（あるいは1入力2出力）の場合に、出力／入力1と出力／入力2とで散布図を描いて参照集合と非効率な評価対象の関係が図示できるが、入力と出力の変数の和が4以上では図示できない。そこで、同じ参照集合をもつ評価対象を同じDEAクラスターの構成員と考えれば、この図表示で得られる原点と参照集合で分割される領域に対応する。また、クロス効率値を階層化クラスター分析で変数のクラスター分析をすれば、ほぼ得られたクラスターの結果に対応していることを示した。そして、各DEAクラスターで非効率なDMUは手本を参考にして改善する方法を提案した。
- 4) 学生のデータは、DEA法が対象とする経営の効率性を考える評価対象ではない。しかし、重回帰分析や判別分析で取り上げられたデータを、効率性という視点で分析を行えば、DEA法の新たな適用領域が拡大できる。4.3の表4.5で31番や34番のような60点の学生を統計的には注目しなかった。しかし、筆者が行った2010年の1年生を対象とした130人前後の必修科目の「統計入門」[10]で、中間試験が30点台で下から10位以内、期末が80点台でトップテンの受講生がいた。高校は、数Ⅲまでが必修の受験校出身者である。結局、能力があるのに勉強しなかったが、余りの成績の悪さに少しは勉強したのであろう。また2011年の中間試験で70点の学生は期末試験で99点取りトップになった[11]。期末試験の範囲になってから、予習することがこの結果になった。すなわち、1年生の時にしっかりと勉強に興味を持たせることが大学教育で必要であり、本来は能力があるが勉強時間の少ない学生を見つけるのにDEA法は役に立つことが分かる。

(成蹊大学経済学部教授)

文献

- [1] L.Schrage(2003). Optimizer Modeling with LINGO.(LINDO Systems Inc.
- [2] S.Shinmura(2012). Relationship between the DEA cluster and the lower limit of weights. Proceedings of DEA Symposium 2012, pp.63-68.
- [3] T.Ueda(2007). Application of Multivariate analysis for DEA. Proceedings of DEA Symposium 2007, pp.96-101.
- [4] K.Tone(2007). Introduction to Efficiency Analysis of a company-DEA(1). Operations Research, 32/12(1988).
- [5] 高森寛, 新村秀一 (1987). 統計処理エッセンシャル。丸善。
- [6] 新村秀一 (2004). JMP活用 統計学とっておき勉強法。講談社。
- [7] _____ (2007). ExcelとLINGOで学ぶ数理計画法。丸善。
- [8] _____ (2011). 数理計画法による問題解決法。日科技連出版社。
- [9] _____ (2011). DEAによる回帰型データのクラスター分析。成蹊大学一般研究報告, 45/3, 1-37.
- [10] _____ (2011). 問題解決学としての統計入門。第7回統計教育の方法論ワークショップ-問題解決力育成を目指した統計教育の方法論-, 1-10.
- [11] _____ (2012). 2011年と2012年度の統計入門の分析-2012年度の半筋授業の影響分析からセンター試験の分析へ-。第8回統計教育の方法論ワークショップ-問題解決力育成を目指した統計教育の方法論-, (2012)
- [12] _____ (2012). 統計教育16年の総括。第8回統計教育方法論のワークショップ-問題解決能力を目指した統計教育の方法論-, 46-55.
- [13] 末吉俊幸 (2001). DEA-経営効率分析法-。朝倉書店。
- [14] 刀根薫 (1993). 経営効率性の測定と改善-抱絡分析法DEAによる-。日科技連出版社。

Efficiency method in a company (1)

- How to extend the usage of DEA -