Title	正準相関分析の経営学への適用可能性 : Redundancyの解釈をめぐって
Sub Title	Applicability of Canonical Correlation Analysis to Business Administration : Issues of Redundancy
Author	岡本, 大輔(Okamoto, Daisuke)
Publisher	
Publication year	1985
Jtitle	三田商学研究 (Mita business review). Vol.28, No.4 (1985. 10) ,p.30- 46
JaLC DOI	
Abstract	
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234698-19851025-04053848

三田商学研究
28 巻 4 号
1985 年 10 月

正準相関分析の経営学への適用可能性

——Redundancy の解釈をめぐって——

岡 本 大 輔

1. はじめに

今日，経営学の実証分析において最も多く用いられている手法は重回帰分析である。これはある分析をする際に考慮すべき要因は，単一では十分でない，という理由による。企業を分析するには，トップ・マネジメント，製品，組織，財務，経営関係，といった様々な要因が考えられるが，これらは独立の要因ではなく，それぞれがお互いに複雑に絡みあっている[1)]。それ故，これら複数の要因を同時に考慮しなければならない。ここに，多変量を同時に扱える重回帰分析を用いる意義がある。

しかし重回帰分析においても，目的を設定する基準変数は単変量である，という制約がある。一般に組織の目的を考えると，構成員，利害関係者の目的と考えることができるが，それは固定的で単一のものではない。Cameron＝Whetten はその理由として以下の4つを挙げている[2)]。

1. 組織の構成員である個人は自分の目的を正しく認識できない。
2. しかもそれは，時とともに変化し得る。
3. 複数の目的が同時に追求される。
4. それら複数の目的は，互いに相反することもある。

組織を企業に限定すれば，究極的目的は長期の維持発展[3)]，とまとめられるが，これを測定するためには様々な指標が考えられる。少なくとも維持のためにはある程度以上の収益を上げ，発展のためには成長しなければならない。それ故，収益性と成長性という2つの指標が必要となり，単一指標というわけにはいかない。そこで重回帰分析においては，複数の指標を何らかの方法で1つの指標に合成するという操作を行なうわけであるが，もしも，合成せずに同時に複数の基準変数を扱える

1） 清水龍瑩〔1984〕p. 56.
2） Cameron＝Whetten〔1983〕p. 12.
3） 清水龍瑩〔1984〕p. 4.

手法があれば，その方が望ましい。そこで期待されるのが，正準相関分析 (Canonical Correlation Analysis) という手法なのである。

2. 正準相関分析

正準相関分析とは，2つの変数群において複数の変数同士の相関関係を見るための手法である。実際には，それぞれの変数群において線型結合を考え，お互いの相関が最大となるような線型結合の係数を求める。

今，2つの変数群を $\mathbf{Y}$, $\mathbf{X}$ とする。

$$\mathbf{Y}=[Y_1\ Y_2\ Y_3 \cdots\cdots Y_p]$$

$$\mathbf{X}=[X_1\ X_2\ X_3 \cdots\cdots X_q]$$

ここで2つの変数群の相関を見るには単相関分析が考えられるが，この場合，pq 個の相関係数が算出される。仮に，$p=8$, $q=10$ とすれば，80個の相関係数となってしまう。そこで，各変数群それぞれを線型結合によりまとめて，その相関を見ようというわけである。つまり，

$$t=\sum_{j=1}^{p} a_j Y_j = a_1 Y_1 + a_2 Y_2 + \cdots\cdots + a_p Y_p$$

$$u=\sum_{j=1}^{q} b_j X_j = b_1 X_1 + b_2 X_2 + \cdots\cdots + b_q X_q$$

という線型結合をつくり，その相関係数 R を求める。ここで，R が最大となる線型結合 t_1, u_1 を第1正準変量 (canonical variate)，その係数 a_{1j}, b_{1j} を第1正準係数 (cononical coefficient)，そして R_1 を第1正準相関係数 (canonical correlation coefficient) と呼ぶ。

次に，第1正準変量 t_1, u_1 とは互いに無相関，という条件の下で新たな正準変量の組が，その相関が最大となるように求まる。それが第2正準変量 (t_2, u_2)，第2正準係数 (a_{2j}, b_{2j})，第2正準相関係数 (R_2) となる。さらに，第1，第2正準変量とは無相関，という条件で，第3正準変量が求まり，このプロセスは第 s 正準変量 ($s=\min(p, q)$) が求まるまで続けられる。まとめれば，それぞれの正準変量は，それ以前に算出された正準変量とは無相関という条件の下で最大の正準相関係数を持つものとして定義される。

このように正準相関分析とは，2つの変数群において，それぞれの線型結合同士の最大相関を求めるものであり，相関関係を調べるための総合特性値を求める手法であるといえる[4)]。従って，考え方としては単相関分析の拡張であり，重回帰分析のように，どちらが説明変数で，どちらが被説明変数（基準変数）という区別はない。しかし，どちらかを基準変数として考えることに何ら支障はな

4） 詳しくは，Green〔1978〕pp. 260-285.

い。また，それが実用上，有用であることは第1章で述べた通りである。そこで本論文では一応，重回帰分析の拡張として，正準相関分析を考えていくことにする。

3. 意味のある正準変量

正準相関分析は，Hotelling によって1936年に考案された手法であるが[5)]，実用上，解釈上多くの問題があり，実際にはあまり用いられていない[6)]。そこで本章では，その点について考察を行なう。

3—1 正準相関係数のチェック

前述の如く，正準相関係数は複数個，算出される。すなわち，1つの変数群が p 個，もう1つの変数群が q 個の変数から成っていれば，$\min(p, q)$ 個の正準相関係数が得られることになる。ただしこれらは，あくまでも計算のプロセスとして機械的に求められるのであって，現実問題として，それらに意味があるかどうかは別問題である。そこで算出される複数個の正準相関係数のそれぞれに意味があるか否か，というチェックを行なう必要がある。しかし現在，確立されたチェック方法がないというのが，正準相関分析がその有用性にもかかわらず，あまり用いられていない大きな理由の1つである。

本論文では，2つのチェックを考える。1つは正準相関係数自体の大きさのチェック，もう1つは構造相関という考え方によるチェックである。では，まず正準相関係数自体のチェックを考えてみよう。

このチェックは統計的有意性検定により，算出された正準相関係数が，統計的に有意なほど大きいかを見る。具体的には，算出された2つの正準変量 t と u は無相関であるという帰無仮説を立て，通常の有意性検定を行なう。この場合，Wilks の Λ を用いた，Bartlett の χ^2 検定が使える[7)]。

第 d 正準相関係数の検定

$$\chi^2 = -\{m-1-(p+q+1)/2\}\log \Lambda_d$$

自由度は $(p+1-d)(q+1-d)$

ここで，$\Lambda_d = \prod_{j=d}^{s}(1-R^2{}_d)$，$m$：サンプル数

この検定により，一応どの正準相関係数に着目したら良いかがわかる。しかし，筆者はこれだけでは十分でないと考える。というのはまず，この統計的有意性検定を行なってもまだ，複数の正準

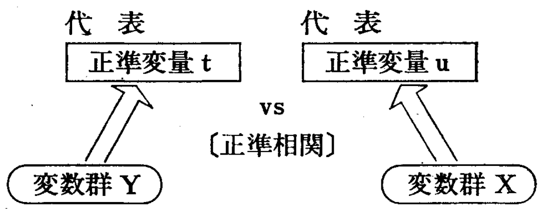
5) Hotelling〔1936〕
6) わずかに例を挙げれば，Waugh〔1942〕，Tintner〔1946〕，清水龍瑩，藤森三男〔1969〕，そして最近では，Reimann〔1975〕
7) 詳しくは，クーリー＝ローンズ〔1973〕pp. 180–181. また，各記号については APPENDIX 参照。

相関係数が残る可能性がある。その場合もしも互いに矛盾しない結果（モデル）が残るならば問題はないが，矛盾することも考えられる。すると，どちらを重視すべきかは，この統計的有意性検定ではわからないのである。そのような場合，正準相関モデルの解釈は非常に難しくなる。また，1回の統計的有意性検定の結果は，その検定の性質上絶対視するわけにはいかないという問題もある。[8]
そこで，第2のチェックを考えるわけである。

3—2　構造相関

構造相関（structure correlation）とは，算出された正準変量とオリジナル変数との相関係数として定義される。例えば，変数群 Y に関する第1正準変量 $t_1=\sum_{j=1}^{p} a_{1j}Y_j$ の構造相関としては，t_1 と Y_1, t_1 と Y_2, ……t_1 と Y_p というように p 個の相関係数が得られる。この構造相関という考え方には，以下のような重要な意義がある。正準相関分析はそもそも2つの変数群の相関関係を調べることを目的としている。この目的のためには，通常，単相関分析が用いられるが，各変数群の変数の数が多いと，単相関係数の数もそれに比例して増えてしまう。そこで正準変量 t，u という，いわば各変数群の総合特性値をつくって，その相関関係を調べようというのが正準相関分析のアプローチである（図1参照）。ところが，この総合特性値であるべき正準変量 t と u は，その相関関係を最大化するという目的でのみ，作られる。つまり t と u は，その正準係数 a_j, b_j によって決まるわけだが，その算出過程は，正準相関係数の最大化が唯一の条件なのである。[9] 故に，算出された正準変量が総合特性値として適切である，という保証はどこにもないのである。このように考えてくると，前述のような正準相関係数の大きさだけを問題にした統計的有意性検定を行なうことは片手落ち，ということがわかる。たとえ2つの正準変量が無相関である，という帰無仮説が棄却され，正準相関係数が統計的に有意，という結果を得たとしても，そのもととなる正準変量がオリジナル変数の総合特性値として無意味であったとしたら，その正準相関係数も無意味である。そこで，正準変量の総合特性値としての資格をチェックする必要が生ずるわけである。そして，そのチェックを行なうのが，構造相関の考え方なのである。図1で考えれば，構造相関の大きさは，各変数群から各正準変量に引かれた矢印の太さで示されており，その太さをチェックしようというのである。

図1　正準相関分析と構造相関



構造相関とは，矢印の太さを表わす指標である。

以上，構造相関の考え方について述べてきたが，如何に重要なチェックであるかがわかる。ただ，実用上には問題がある。最初に述べたが，実に多くの構造相関係数が算出されてしまうのであ

8）　岡本大輔〔1984，a〕
9）　もちろん，第2正準変量以降においては，それ以前に算出された正準変量と無相関，という条件がつくが，ここでの議論には関係ない。

る。[10] もしも，$p=8$, $q=10$であれば，$(8+10)\times 8=144$個！もの構造関係が得られてしまう。これらを，さらに要約するのが次章で述べる Redundancy である，と筆者は考えるのである。

4. Redundancy

4—1 Redundancy の定義

Stewart=Love は，1968年に “A General Canonical Correlation Index” という論文を発表し，その中で Redundancy という考え方を示した。[11] まず，Redundancy の定義を以下に示す。[12]

Y に関する第 d Redundancy $R_{Yd}{}^2=R_d{}^2\left[\sum_{c=1}^{p} g_{dc}{}^2/p\right]$

X に関する第 d Redundancy $R_{Xd}{}^2=R_d{}^2\left[\sum_{c=1}^{q} h_{dc}{}^2/q\right]$

Y に関する Redundancy $\overline{R_Y{}^2}=\sum_{d=1}^{s} R_{Yd}{}^2$

X に関する Redundancy $\overline{R_X{}^2}=\sum_{d=1}^{s} R_{Xd}{}^2$

ここで，g_{dc}はベクトル $\mathbf{g}_d=\mathbf{R}_{yy}\mathbf{a}_d$ の第 c 要素，h_{dc}はベクトル $\mathbf{h}_d=\mathbf{R}_{xx}\mathbf{b}_d$ の 第 c 要素。[13]
$\mathbf{R}_{yy}$, $\mathbf{R}_{xx}$ は各変数群の相関行列，$\mathbf{a}_d$, $\mathbf{b}_d$ は各変数群の第 d 正準係数ベクトル。

Stewart=Love は，この Redundancy の方が正準相関係数よりも，2つの変数群 Y，X の情報の重複度を正しく示していると主張した。すなわち，正準変量を評価する際には，正準相関係数ではなく，この Redundancy を見るべきであるとした。ところが Stewart=Love は上記の論文で，Redundancy の定義について詳しく述べなかったため，様々な議論をまきおこしてしまった。[14] その議論は主に，Psychological Bulletin という学術雑誌上において行なわれたが，うまく議論がかみ合わず，明確な結論は得られていない。そこで，以下では Redundancy を再考し，筆者なりの解釈を述べ，その有用性を考えていくことにする。

4—2 Redundancy の考え方

Redundancy 批判の主なものは，その式の意味が不明確なため，いくつかの誤解を生じ，その

10) 図1を正確に書くならば，各変数群からの矢印は1本ではなく，各変数群の変数の数だけ引ける。
11) Stewart=Love〔1968〕
12) 説明の都合上，表記法を若干，修正した。補足をするならば，Redundancy はまず，各正準変量ごとに計算され（$R_{Yd}{}^2$, $R_{Xd}{}^2$），夫々をY について加えたものが変数群 Y の Redundancy（$\overline{R_Y{}^2}$），Xについて加えたものが変数群 X の Redundancy（$\overline{R_X{}^2}$）となる。
13) 記号については，APPENDIX 参照。
14) 例えば，Nicewander=Wood〔1974, 1975〕，Miller〔1975〕など。

指標の意味もおかしいというものであった。そこで本節では Redundancy の式の意味を考えてみる。

$$\overline{R_{Y^2}} = \sum_{d=1}^{s} R_d^2 \left[\sum_{c=1}^{p} g_{dc}^2 / p \right] \quad \cdots\cdots (1)$$ [15]

まず，(1)式の右辺の［ ］内に着目する。この［ ］内は，$\frac{1}{p}\sum_{c=1}^{p} g_{dc}^2$ と変形でき，さらに定義により，

$$g_{dc} = [\mathbf{g}_d]_c = [\mathrm{R}_{yy}\mathbf{a}_d]_c = [\mathrm{R}_{yy}\mathbb{A}]_{cd} \quad \cdots\cdots (2)$$ [16]

ここで(2)の等式の右から2番目の式から見ていくと，R_{yy} は変数群 $\mathbf{Y}$ の p 行 p 列の相関行列であり，

$$\mathrm{R}_{yy}^{17)} = \frac{1}{m-1} \begin{pmatrix} \sum y_{i1}^2 & \sum y_{i1}\ y_{i2} \cdots\cdots\cdots\cdots & \sum y_{i1}\ y_{ip} \\ \sum y_{i2}\ y_{i1} & \sum y_{i2}^2 \cdots\cdots\cdots\cdots & \sum y_{i2}\ y_{ip} \\ \vdots & & \vdots \\ \sum y_{ic}\ y_{i1} & \cdots\cdots\cdots\cdots\cdots\cdots & \sum y_{ic}\ y_{ip} \\ \vdots & & \vdots \\ \sum y_{ip}\ y_{i1} & \cdots\cdots\cdots\cdots\cdots\cdots & \sum y_{ip}^2 \end{pmatrix}$$

ここで，y_{ij} は Y_{ij}〔第 i サンプル，第 j 変数〕を $N(0,1)$ に基準化したもの，各$\sum$は$\sum_{i=1}^{m}$を示し，m はサンプル数である。

$\mathbf{a}_d$ は第 d 正準係数ベクトルであり，

$$\mathbf{a}_d = \begin{pmatrix} a_{d1} \\ a_{d2} \\ \vdots \\ a_{dp} \end{pmatrix}$$

$\mathbb{A}$ ((2)の等式の一番右の式) は $\mathbf{a}_d$ を横に並べた p 行 s 列の正準係数行列である。

$$\mathbb{A} = [\mathbf{a}_1\ \mathbf{a}_2 \cdots\cdots\cdots \mathbf{a}_d \cdots\cdots \mathbf{a}_s]$$

$$= \begin{pmatrix} a_{11} & a_{21} \cdots\cdots & a_{d1} \cdots\cdots & a_{s1} \\ a_{12} & a_{22} \cdots\cdots & a_{d2} \cdots\cdots & a_{s2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{2p} \cdots\cdots & a_{dp} \cdots\cdots & a_{sp} \end{pmatrix}$$

これらの定義により(2)式を説明すれば，g_{dc} はベクトル $\mathbf{g}_d$，ベクトル $\mathrm{R}_{yy}\mathbf{a}_d$ の第 c 要素であり，行列 $\mathrm{R}_{yy}\mathbb{A}$ の c 行 d 列要素であるということになる。すなわち，g_{dc} を解釈するためには，行列

15) ここでは，変数群$\mathbf{Y}$を例にとって説明するが，変数群$\mathbf{X}$についても，全く，同様である。
16) 〔 〕の右下の添字は，各ベクトル，または行列の要素を表わす。
17) この式は y_{ij} の分散共分散行列だが，$N(0, 1)$ に基準化されているので，相関行列となる。以下，変数はすべて，$N(0, 1)$ に基準化されているものとして考える。

$\mathbb{R}_{yy}\mathbb{A}$ の c 行 d 列要素を見れば良い。行列 $\mathbb{R}_{yy}\mathbb{A}$（$p$ 行 s 列）を書き下してみると，

$$\mathbb{R}_{yy}\mathbb{A}=\frac{1}{m-1}\begin{pmatrix} \sum_{j=1}^{p}\left[a_{1j}\sum_{i=1}^{m}y_{i1}y_{ij}\right] & \sum_{j=1}^{p}\left[a_{2j}\sum_{i=1}^{m}y_{i1}y_{ij}\right]\cdots\cdots & \sum_{j=1}^{p}\left[a_{dj}\sum_{i=1}^{m}y_{i1}y_{ij}\right]\cdots\cdots & \sum_{j=1}^{p}\left[a_{sj}\sum_{i=1}^{m}y_{i1}y_{ij}\right] \\ \sum_{j=1}^{p}\left[a_{1j}\sum_{i=1}^{m}y_{i2}y_{ij}\right] & \sum_{j=1}^{p}\left[a_{2j}\sum_{i=1}^{m}y_{i2}y_{ij}\right]\cdots\cdots & \sum_{j=1}^{p}\left[a_{dj}\sum_{i=1}^{m}y_{i2}y_{ij}\right]\cdots\cdots & \sum_{j=1}^{p}\left[a_{sj}\sum_{i=1}^{m}y_{i2}y_{ij}\right] \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^{p}\left[a_{1j}\sum_{i=1}^{m}y_{ic}y_{ij}\right] & \sum_{j=1}^{p}\left[a_{2j}\sum_{i=1}^{m}y_{ic}y_{ij}\right]\cdots\cdots & \sum_{j=1}^{p}\left[a_{dj}\sum_{i=1}^{m}y_{ic}y_{ij}\right] & \vdots \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{p}\left[a_{1j}\sum_{i=1}^{m}y_{ip}y_{ij}\right] & \sum_{j=1}^{p}\left[a_{2j}\sum_{i=1}^{m}y_{ip}y_{ij}\right]\cdots\cdots & \cdots\cdots\cdots\cdots & \sum_{j=1}^{p}\left[a_{sj}\sum_{i=1}^{m}y_{ip}y_{ij}\right] \end{pmatrix}$$

故に行列 $\mathbf{R}_{yy}\mathbf{A}$ の c 行 d 列要素は

$$\frac{1}{m-1}\sum_{j=1}^{p}\left[a_{dj}\sum_{i=1}^{m}y_{ic}y_{ij}\right]=\frac{1}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{p}y_{ic}y_{ij}a_{dj}$$

と書ける。(2)式より，

$$g_{dc}=\frac{1}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{p}y_{ic}y_{ij}a_{dj} \qquad \cdots\cdots(3)$$

ところで，第 i サンプル，第 d 正準変量 t_{id} を考えてみると，正準変量の定義により，

$$t_{id}=\sum_{j=1}^{p}y_{ij}a_{dj}$$

この第 d 正準変量 t_{id} と変数群 $\mathbf{Y}$ の第 c 変数 y_{ic} との相関係数 $r(y_c, t_d)$ を考えてみると，

$$\begin{aligned} r(y_c, t_d) &= \frac{1}{m-1}\sum_{i=1}^{m}y_{ic}t_{id} \\ &= \frac{1}{m-1}\sum_{i=1}^{m}y_{ic}\left[\sum_{j=1}^{p}y_{ij}a_{dj}\right] \\ &= \frac{1}{m-1}\sum_{i=1}^{m}\sum_{j=1}^{p}y_{ic}y_{ij}a_{dj} \qquad \cdots\cdots(4) \end{aligned}$$

ここで(3)式と(4)式とを比べてみると，全く同じ式である。

$$g_{dc}=r(y_c, t_d) \qquad \cdots\cdots(5)$$

つまり，g_{dc} は変数群 $\mathbf{Y}$ についての第 d 正準変量と，変数群 $\mathbf{Y}$ の第 c 変数との相関係数であることが証明された。これは正に，前章で述べた構造相関である。[18] すると，g_{dc}^2 はいわば，構造決定

18)　この証明は，$\mathbf{R}_{yy}$ と $\mathbf{A}$ との積の直接的証明であるが，行列を分解して以下の様に証明することもできる。今，変数群 $\mathbf{Y}$ を基準化した行列を $\mathbf{Y}_s$ とすれば，$\mathbf{R}_{yy}=\mathbf{Y}_s'\mathbf{Y}_s$　∴　$\mathbf{R}_{yy}\mathbf{A}=(\mathbf{Y}_s'\mathbf{Y}_s)\mathbf{A}=\mathbf{Y}_s'(\mathbf{Y}_s\mathbf{A})$
ここで，$(\mathbf{Y}_s\mathbf{A})$ はいわば正準変量行列であるから，$\mathbf{Y}_s'(\mathbf{Y}_s\mathbf{A})$ は構造相関行列である。

係数と言え，さらに(1)式の右辺 [ ] 内の$\frac{1}{p}\sum_{c=1}^{p} g_{dc}^2$ は，構造決定係数の平均値，ということになる。構造相関は前章で説明した通り，算出された正準変量がもとの変数をどれだけ説明しているかを示しているので，これを各変数に関する平均値にすれば，もとの変数群をどれだけ説明しているかを示すことになる。すなわち，(1)式の [ ] 内，$\frac{1}{p}\sum_{c=1}^{p} g_{dc}^2$ は，オリジナル変数群（この場合は変数群Y）の総合特性値としての資格を示しており，これは，いわば代表権の大きさを示している，と考えられる。

次に，(1)式の右辺の$\sum_{d=1}^{s}$の中，すなわち，変数群Yに関する第 d Redundancy $R_d^2\left[\sum_{c=1}^{p} g_{dc}^2/p\right]=R_{Yd}^2$ について考えてみよう。正準相関分析においては，正準相関係数が大きく，しかもそれぞれの正準変量 t_d, u_d がオリジナルの変数群 Y，X を良く表わしている（代表権が大きい）ことが望ましい，ということを前章で述べた。そこで，正準相関係数に，代表権の大きさを示す $\left[\sum_{c=1}^{p} g_{dc}^2/p\right]$ をかければ，この 2 つの基準を同時に満たしているか，という指標ができる。ただここで，代表権の方の $\left[\sum_{c=1}^{p} g_{dc}^2/p\right]$ は前述のように相関係数ではなく決定係数のレベルになっているので，正準相関係数の方も同じレベルにするため，2 乗しなければならない。それが，第 d Redundancy $R_d^2\left[\sum_{c=1}^{p} g_{dc}/p\right]$ $=R_{Yd}^2$ となるわけである。まとめれば，変数群 Y に関する第 d Redundancy とは，前章で述べた 2 つのチェックの基準を同時にチェックするための総合指標であると言える。さらにこれを，(1)式のように，すべての正準変量について加えれば，全体の Redundancy が求まる，というわけである。

以上の考え方により，Redundancy は，算出された正準変量，そして正準相関係数の有用性を示しており，単に正準相関係数を見るよりも，Redundancy を見る方が良いという Stewart＝Love の主張は，正しいと言える。

4—3 Redundancy というネーミング

Redundancy の考え方は今のところ，ほとんど認められていない。例えば，多変量解析の標準的テキストで Redundancy を紹介しているテキストは数少ない。また，コンピュータのパッケージ・プログラムで Redundancy という統計量を算出してくれるソフトは，筆者の知る限り，存在しない。Redundancy の意味が不明確なため，多くの議論を呼んでしまい明確な結論が得られていなかったことが，その大きな理由である。

日本では，Cooley＝Lohnes の *Multivariate Data Analysis* の翻訳において Redundancy が紹介されているが，冗長性と訳されている[19)]。Redundancy を冗長性と訳すと，この値が小さい方が望ましいという誤解が生じてしまう。原語の Redundancy のままでも同様の誤解は十分に考えられる。もともと Redundancy（冗長性）とは，情報理論の用語であり，情報が誤りなく伝達されたか否

19) クーリー＝ローンズ〔1973〕pp. 175-179.

かをチェックするための“余分”な情報という意味で使われていた言葉である。[20] この場合のRedundancy は良い意味で使われているので，Stewart=Love は，情報の重複度を示す指標として Redundancy と名づけたのであろう。しかし，やはり上記の如き誤解は避けられない。つまり，Redundancy が一般に認められないもう1つの理由は，そのネーミングの悪さにあったのである。少なくとも日本語では，総合特性度，代表度などの意訳が必要と思われる。

4—4 Redundancy の利用法

これまでは説明の都合，変数群Yの Redundancy を例にとって説明してきた。しかし実際には変数群Xについても全く同じ統計量が算出される。そこで利用に際しては，両方の変数群の Redundancy を見なければならない。

Redundancy の基本的要素には前章で述べた構造相関が入っており，各変数群の総合特性値としての資格を見ているわけであるから，その資格に差があってはならない。図1にもどって考えてみるならば，2本の矢印の太さに差があってはならないのである。これは例え，一方の正準変量の総合特性値としての資格が高くても(矢印が太くても)，もう一方の正準変量の総合特性値としての資格が低ければ(矢印が細ければ)，その正準相関係数に意味がなくなってしまうことを考えれば，明らかである。それ故，実際には，2つの変数群の Redundancy を見て，両者が同程度に大きい場合，初めて，その正準相関係数，正準相関モデルが統計的に信頼できるということになるのである。

5. 数 値 例[21]

5—1 サンプル・データ

中部経済同友会では，中部地区と関東地区の中堅企業に関する調査を行なった。以下にその調査概要を示す。

調査機関：中部経済同友会中堅企業委員会 (名古屋相互銀行社長　加藤千麿委員長)

主　　査：慶應義塾大学商学部　清水龍瑩教授 (調査員は，名古屋経済大学経済学部　篠原光伸講師と筆者)

調査対象：資本金3,000万—10億円の中堅製造企業

調査時期：1984年12月

20) 中山伊知郎〔1972〕p. 1006.

21) 本章の計算には，慶應義塾大学計算センター三田計算室の FACOM M-360を使用した。パッケージ・プログラムは，富士通 ANALYST (CANCORR) を利用し，TSS 端末，Redundancy の計算には，NEC PC-9801F2 を利用した。

調査方法：アンケート調査（有効回答：中部地区142社，関東地区140社），社長面接調査及び工場見学（中部地区4社，関東地区4社）

分析方法：QAQF（慶應義塾経営力評価グループ開発，定性要因の定量分析手法）[22]

分析結果：中堅企業の一般的成長条件，中部・関東地区中堅企業の共通成長要因，中部地区と関東地区の違い。[23]

本章では，上記のサンプル・データをそのまま用いることにする。

5―2　分析目的とステップ

上記の調査で得られた結果のうち，中部・関東地区の違いについてまとめると，中部では，クローズド思考，安定思考，製造思考の企業の業績が良く，関東では逆に，オープン思考，成長思考，製品思考の企業の業績が良いことがわかった。この結果は単に統計的処理から生まれたものではなく，社長面接調査によっても，確認されている。また，他にも同様の結論を出している研究が多い。例えば，名古屋大学の瀧澤菊太郎教授は中部地区の優良企業には安全性を重視して堅実経営を行なう企業が多いことを指摘し，その理由を以下の様にまとめている。第1に徳川時代以来，質実勤倹の気風が強かったこと，第2に政治の中心東京と，商売の中心大阪の中間であり，積極的に経営を展開するに必要な情報が欠けていたこと，第3に東京，大阪と異なり，大手財閥に属さない地方財閥が多く，自分の城は自分で守るほかなかったこと。[24]

そこで本章では，正準相関分析により中堅企業の分析を行ない，果して，従来の諸説，従来の研究結果の検証ができるのか，を見ることにする。以下に，その分析のステップを示す。

1. 変数群Yには業績指標，変数群Xには経営要因を設定する。
2. 中部データと関東データを分け，それぞれ別々に正準相関分析を行なう。
3. 前章までの考え方を基に，統計的に信頼できる正準相関モデルを選び出す。
4. 中部モデルと関東モデルを比較し，従来の諸説，すなわち，中部企業の安定思考と関東企業の成長思考などの違い，が検証できるかを見る。

5―3　分析結果

(1)　正準相関分析　Take 1

中部経済同友会調査のデータは，88の変数から成っているが，これらすべてを正準相関分析モデルに入れても，良い結果は期待できない。というのは，正準相関係数は大きくなるが，構造相関の

22)　詳しくは，清水龍瑩〔1981〕pp. 259-267.
23)　詳しくは，中部経済同友会〔1985〕，清水龍瑩〔1985〕
24)　瀧沢菊太郎他〔1985〕pp. 45-51.

平均は小さくなり，結局は Redundancy が小さくなってしまう，と予想されるからである。そこで，以下の７変数を選んだ。

変数群Ｙとして

Y_1：売上高伸率（４年間移動平均伸率）

Y_2：利益率（使用総資本経常利益率）

変数群Ｘとして

X_1：社長の年齢

X_2：社長の持株比率（社長の親族も含む）

X_3：最高意思決定機関の運営方法（役員の参加度）

X_4：新鋭設備比率（主要設備のうち，取得後３年以内のものの割合，取得原価ベース）

X_5：製品技術開発力（製品技術の独自開発度）

表１　正準相関分析　Take 1

中　部

正準変量	正準相関係数 R	R の２乗	χ^2 値	自由度	有意確率
1	0.40534	0.16430	26.668	10	0.003
2	0.15915	0.02533	3.335	4	0.503

変数群	正準変量	構造決定係数の平均値	Redundancy	Redundancyの割合
Y	1	0.44747	0.07352	0.90231
	2	0.31425	0.00796	0.09769
			計　0.08148	
X	1	0.04997	0.00821	0.89046
	2	0.03987	0.00101	0.10954
			計　0.00922	

関　東

正準変量	正準相関係数 R	R の２乗	χ^2 値	自由度	有意確率
1	0.39466	0.15576	21.637	10	0.017
2	0.08949	0.00801	0.981	4	0.913

変数群	正準変量	構造決定係数の平均値	Redundancy	Redundancyの割合
Y	1	0.44498	0.06931	0.97236
	2	0.24594	0.00197	0.02764
			計　0.07128	
X	1	0.06067	0.00945	0.95166
	2	0.05935	0.00048	0.04834
			計　0.00993	

結果は表1にまとめられている。$p=2$, $q=5$ なので $s=\min(p, q)=2$ となり，2つの正準相関係数が算出された。しかし，中部モデルにおいても関東モデルにおいても，χ^2 検定で統計的に有意[25]となったのは，第1正準変量だけであった。Redundancy を見ると，中部の Y に関する第1 Redundancy が 0.07352, X に関する第1 Redundancy が 0.00821であり，1ケタ違う。すなわち，図1で言えば矢印の太さが全く異なっている状態である。これは，2つの正準変量の総合特性値としての資格，代表度が異なっており，Y の正準変量は Y を良く表わしているが，X の正準変量は X を良く表わしていないということを意味する。この場合の正準相関係数は無意味である。第2 Redundancy を見ると，Y，X ともに小さい。この場合の正準相関係数も無意味である。関東モデルにおいても，Y に関する第1 Redundancy が0.06931, X に関する第1 Redundancy が 0.00945 と，これまたケタが違っている。そして第2 Redundancy も小さい。まとめると，この Take 1 では，中部においても関東においても，統計的検定で有意となったのは第1正準相関係数だけで，しかもその Redundancy は良い値が得られていない。すなわち，統計的に信頼できるモデルは得られなかったというのが結論である。

Redundancy は正準相関係数の2乗と，構造決定係数の平均値との積によって求まることを前章で述べたが，この Take 1 の Redundancy の低さの原因は，後者の構造決定係数の平均値の低さにあった。これは平均値の性質上，1つでも構造相関の低い変数があると，その影響が全体に出てしまうためと考えられる。実際，Take 1 では，X_3 の構造相関が極端に低かったのである。モデルに入れる変数が多すぎると，この様な状態が起きる確率が高まるので，予め，7変数を選んだが，それでもまだ多いことがわかった。[26]

(2)　正準相関分析　Take 2

Take 1 の7変数でも，変数の数が多すぎることがわかったので，X の変数を減らしながら，試行錯誤をかさね，最終的に，5変数モデルを得た。5変数とは，Y_1, Y_2, X_2, X_4, X_5 である。その結果が表2にまとめられている。Take 1 同様，2つの正準相関係数が求まり，5％で統計的に有意となったのは第1正準相関分析だけであった。これは中部モデル，関東モデル，共通である。中部モデルにおいて第1 Redundancy を見ると，Y について0.08789, X について0.04387である。同じではないが，Take 1 とは異なり，同じケタなので，一応，モデルの経営学的解釈を行なう価値がある程度の水準であると判断した。関東モデルにおいても，第1 Redundancy が Y について 0.07946, X について0.04328，とほぼ同様な結果が得られている。第2正準相関係数については，中部，関東ともに，χ^2 検定で有意とならなかったし，Redundancy もかなり低いので，信頼できない。そこで，Take 2 で，中部，関東ともに第1正準相関係数だけを，統計的に信頼できると

25)　有意水準は5％，有意水準の議論については岡本大輔〔1984, a〕

26)　もちろん，7変数で多い，というのは本章で用いているサンプル，変数においてのことである。しかしこのように，Redundancy を変数選択の基準として用いることができる，という点は注目に値する。

し，モデルの解釈を行なうことにする。

表2　正準相関分析　Take 2

中　部

正準変量	正準相関係数 R	R の2乗	χ^2 値	自由度	有意確率
1	0.36511	0.13331	20.733	6	0.002
2	0.12402	0.01538	2.031	2	0.362

変数群	正準変量	構造決定係数の平均値	Redundancy	Redundancyの割合
Y	1	0.65930	0.08789	0.94373
	2	0.34072	0.00524	0.05625
			計　0.09313	
X	1	0.32908	0.04387	0.89257
	2	0.34327	0.00528	0.10743
			計　0.04915	

関　東

正準変量	正準相関係数 R	R の2乗	χ^2 値	自由度	有意確率
1	0.36029	0.12981	17.683	6	0.007
2	0.04916	0.00242	0.302	2	0.860

変数群	正準変量	構造決定係数の平均値	Redundancy	Redundancyの割合
Y	1	0.61212	0.07946	0.98831
	2	0.38788	0.00094	0.01169
			計　0.08040	
X	1	0.33339	0.04328	0.97742
	2	0.41528	0.00100	0.02258
			計　0.04428	

表3　正準相関分析（Take 2）のモデル

中部の正準相関モデル

正準相関係数　$R=0.36511$　Redundancy　$R_{Y1}{}^2=0.08789$

Redundancy　$R_{X1}{}^2=0.04387$

0.850369〔売上高伸率〕+0.288802〔利益率〕

0.302288〔社長持株比〕+0.959352〔新鋭設備比〕−0.047983〔製品技術力〕

関東の正準相関モデル

正準相関係数　$R=0.36029$　Redundancy　$R_{Y1}{}^2=0.07946$

Redundancy　$R_{X1}{}^2=0.04328$

0.479280〔売上高伸率〕+0.770279〔利益率〕

−0.183318〔社長持株比〕+0.920630〔新鋭設備比〕+0.361964〔製品技術力〕

5—4　正準相関モデルの解釈

表3が，最終的に得られた正準相関モデルである。ここでは，各正準係数の符号についてのみ，解釈を行なう。というのは，各正準係数の大きさは，重回帰分析と同様，各変数の貢献度の目安にはなるが，お互いの変数群の中での相関があるため，正しい貢献度とは言えない[27]。また，重回帰分析の t 値に匹敵するような統計量も存在せず，係数の大きさに関して明確なことは言えない，と考えるからである[28]。

ではまず，変数群 **Y** の正準変量 t_1 について見てみよう。「売上高伸率」，「利益率」ともに正の符号を持っており，これは中部モデル，関東モデル共通である。企業にとって売上高伸率も利益率も高い方が望ましい指標であるから，正準変量 t_1，は業績の総合特性値として考えて良い。もしも，Y_1 と Y_2 の係数の符号が正と負に分かれてしまえば，正準変量 t_1 の解釈は難しくなるが，幸い，両方とも正になったので，このような解釈が可能となったわけである。では業績の総合特性値としての t_1 に，各経営要因がどのように関連しているかを見るため，変数群 **X** の正準変量 u_1 について見てみよう。

まず，「社長の持株比率」であるが，中部モデルでは正，関東モデルでは負，となっている。すなわち，

『中部の企業では，社長の持株比率の高い企業の業績が良く，関東の企業では逆に，低い企業の業績が良い』

これは中部企業のクローズド思考，関東企業のオープン思考を表わしている。中部では“企業というものを同族を中核とした共同体であるという企業一家的考え方”[29]があり，クローズドな経営を行なう方が高業績につながるのである。

次に「新鋭設備比率」を見ると，中部，関東ともに正の符号となっている。すなわち，

『一般に，新鋭設備比率の高い企業の業績は良い』

中部においても関東においても，積極的な設備投資，新鋭設備の導入は技術水準を高め，効率的な生産体制を確立し，コストダウンを可能とし，業績を高めるのであろう。

では，どのような技術を重視するのか，というのが3つ目の「製品技術開発力」である。ここでは技術を2つに分けて考えている。“製品の品質・機能に関する技術（製品技術）と，製造に関する

27）岡本大輔〔1984，b〕pp. 31-32.

28）この問題については，今後，考えるべき大きな課題である，と筆者は考えている。

29）中部経済同友会〔1985〕p. 17.

技術(製造技術)[30]”である。そして「製品技術開発力」は前者の技術を外部から導入せずに，独自に開発する度合を示した変数である。この変数の係数の符号が，中部では負，関東では正となっている。すなわち，

『中部では製品技術を外部から導入する企業の業績が良く，関東では独自に開発する企業の業績が良い』

ここで，「新鋭設備比率」と「製品技術開発力」とを考慮してまとめれば，

『中部でも関東でも，技術は製品戦略の中核であり，重要な経営要因であるが，その力の入れ所が異なっている。中部では製造思考であり，製品技術よりも製造技術に力を入れる企業の業績が良く，関東では製品思考であり，製品技術に力を入れる企業の業績が良い』

全体としては，中部の安定思考，関東の成長思考としてまとめられる。中部では，クローズドで，独立的な経営を行ない，製造技術を重視してコストダウンを図る企業が高業績を上げている。関東では逆に，オープンな経営を行ない，製品技術を重視して，新機能を持った製品の開発を行なう企業が，高業績を上げている。

以上，中部企業，関東企業を正準相関分析によって比較してみた結果，統計的に信頼のおけるモデルにおいて，従来の諸説，従来の研究結果と矛盾しない結論が得られた。今回の分析では，従来の方法に劣らないというのが一応の結論である。しかし，現在，最も信頼され，最も多く使用されている重回帰分析に比べて，複数の基準変数を同時に扱えるという利点のある正準相関分析は，今後，大いに研究され，改良され，用いられていく価値のある分析手法である。

6. 要　　約

社会科学においてはほとんどの場合，複数の要因が複雑にからみあっており，一面的な解釈では何も言えず，多面的な解釈が必要となる。重回帰分析を用いる意義はそこにあるが，基準変数については単変量である。そこで正準相関分析を，複数の説明変数，複数の基準変数という視点で考えてみた。

正準相関係数は，そのプロセスの性質上，複数個，算出されてしまい，その解釈が難しい。そこで筆者は Stewart=Love の Redundancy という考え方を利用し，正準相関分析の利用法をまと

30) 清水龍瑩〔1984〕p. 111.

めてみた。Redundancy は，その式の解釈が困難なため，あまり用いられていないが，正準相関係数のチェックには極めて有用である。すなわち，正準相関係数の大きさと，正準変量の総合特性値としての資格，という2つの基準を同時に満たしているか否かを示しているのが Redundancy なのである。

Redundancy を利用することにより，正準相関分析の有用性は相当，高まったと言える。最後の数値例では，中部企業と関東企業を比較し，その違いをかなり明確に示すことに成功した。ただ，Redundancy 自体のチェック基準や，各正準係数の個別のチェックなど，残された問題は多い。正準相関分析の複数の基準変数を同時に扱えるという利点を考えると，これらの問題点を解決していく意義は大きい。正準相関分析は，今後，さらに研究され，改良されていく価値のある分析手法である。

REFERENCES

奥野忠一他『多変量解析法』日科技連，1971年。

岩田暁一，黒田昌裕「正準相関分析のための計算プログラム」『産業研究所シリーズ』No.191（1966年）pp. 1–39.

クーリー=ローンズ『行動科学のための多変量解析』井口晴弘他（訳）鹿島出版会，1973年。

清水龍瑩『企業成長論』中央経済社，1984年。

　　　　『現代企業評価論』中央経済社，1981年。

　　　　「中堅企業の企業経営と成長要因」『三田商学研究』第28巻3号（1985年）pp. 1–21.

清水龍瑩，藤森三男「経営力評価のための多変量解析モデル」『三田商学研究』第12巻5号（1969年）pp. 102–145.

瀧沢菊太郎他『無借金企業』有斐閣，1985年。

中部経済同友会『中堅企業の成長の条件』1985年。

中山伊知郎（編）『統計学辞典』東洋経済新報社，1972年。

ボルチ=ファング『応用多変量解析』中村慶一（訳）森北出版，1976年。

岡本大輔「経営学研究における統計的有意性検定」『三田商学研究』第27巻5号（1984年）pp. 1–12. a.

　　　　「財務データによる企業評価」『三田商学研究』第27巻1号（1984年）pp. 16–32. b.

Cameron, K. S. and Whetten, D. A., *Organization Effectiveness* Academic Press N. Y. 1983.

Gleason, T. C., "On Redundancy in Canonical Analysis", *Psychological Bulletin* (Vol. 83, No. 6) 1976, pp. 1004–1006.

Green, P.E., *Analyzing Maltivariate Data* The Dryden Press, Illinois, 1978.

Holland, T. R., Levi, M. and Watson, C. G., "Canonical Correlation in the Analysis of a Contingency Table", *Psychological Bulletin* (Vol. 87, No. 2) 1980, pp. 334–336.

Hotelling, H., "Relations between Two Sets of Variables", *Biometrica* (Vol. 28) 1936, pp. 321–377.

Isaac, P. D. and Milligan, G. W., "A Comment on the Use of Canonical Correlation in the Analysis of Contingency Tables", *Psychological Bulletin* (Vol. 93, No. 2) 1983, pp. 378–381.

Knapp, T. R., "Canonical Correlation Analysis : A General Parametric Significance-Testing System", *Psychological Bulletin* (Vol. 85, No. 2) 1978, pp. 410–416.

Miller, J. K., "In Defense of the General Canonical Correlation Index : Reply to Nicewander and Wood", *Psychological Bulletin* (Vol. 82, No. 2) 1975, pp. 207–209.

Nicewander, W. A. and Wood, D. A., "Comments on "A General Canonical Correlation Index"", *Psychological Bulletin* (Vol. 81, No. 1) 1974, pp. 92–94.

- "On the Mathematical Bases of the General Canonical Correlation Index : Rejoinder to Miller", *Psychological Bulletin* (Vol. 82, No. 2) 1975, pp. 210–212.

Reimann,. B. C., "Organizational Effectiveness and Management's Public Values : A Canonical Analysis", *Academy of Management Journal* (Vol. 18. No. 2) 1975, pp. 224–241.

Stewart, D. and Love, W., "A General Canonical Correlation Index", *Psychological Bulletin* (Vol. 70, No. 3) 1968, pp. 160–163.

Tintner, G., "Some Applications of Multivariate Analysis to Economic Data", *Journal of the American Statistical Association* (Vol. 41) 1946, pp. 472–500.

Waugh, F. W., "Regressions between Sets of Variables", *Econometrica* (Vol. 10) 1942, pp. 290–310.

APPENDIX　記号一覧

m：サンプル数

p, q：変数数，$s=\min(p, q)$

$\mathbf{Y}, \mathbf{X}$：２つの変数群

$\mathbf{a}_d$：第 d 正準係数ベクトル($\mathbf{Y}$)，$\mathbf{a}_d'=(a_{d1}\ a_{d2}\cdots\cdots a_{dp})$

$\mathbf{b}_d$：第 d 正準係数ベクトル($\mathbf{X}$)，$\mathbf{b}_d'=(b_{d1}\ b_{d2}\cdots\cdots b_{dq})$

t_d：第 d 正準変量($\mathbf{Y}$)，$t_d=\sum_{j=1}^{p} a_{dj}\ Y_j$

u_d：第 d 正準変量($\mathbf{X}$)，$u_d=\sum_{j=1}^{q} b_{dj}\ X_j$

R_d：第 d 正準相関係数，t_d と u_d の単相関係数

$\mathbf{R}_{yy}$：$\mathbf{Y}$の相関行列

$\mathbf{R}_{xx}$：$\mathbf{X}$の相関行列

最後に，本論文執筆にあたり，常に貴重な助言をして下さった，慶應義塾大学商学部教授，清水龍瑩先生に深く感謝いたします。そして論文執筆中に行われた慶應義塾大学商学会報告会では，岩田暁一教授をはじめ，多くの先生方に貴重なコメントをいただいたことを明記し感謝いたします。また，本論文執筆の動機となった正準相関分析の議論の多くは，清水龍瑩研究会で生まれたものであり，その出席者である黒川行治助教授，篠原光伸講師（名古屋経済大学）の両先生，そして大学院生の梼浩二君，松永久君，海保英孝君，古川靖洋君の諸君に感謝いたします。

(1985年７月脱稿)