Title	主成分分析(Principal Components)の経済分析への応用
Sub Title	An application of the method of principal components to economic analysis
Author	佐藤, 保
Publisher	慶應義塾経済学会
Publication year	1954
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.47, No.2 (1954. 2) ,p.160(52)- 173(65)
JaLC DOI	10.14991/001.19540201-0052
Abstract	
Notes	資料
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-19540201-0052

て、彼等は、それが、政治權力・指導的經濟勢力との聯繫を缺き、且、彼等に內在する極度に保守的な性格に基因するものであることを見究めやうとはせず、徒らに、彼等の經濟的活動分野を收縮することに專念した。それ故、少くとも、明治三〇年代に於ける彼等の經濟的勢力には一つの限界が存在してゐたと見られるのである。それにも、拘らずこの過程を通じて、彼等の林業經營に對する關心は强化され、山林への意識に斬新さが加えられた點を見逃してはならない。この意味に於て、彼等は確かに、當時の我が國林業を擔ふ一群の人々といふことが出來るであろう。

（筆を措くに當り、貴重な資料を貸與された土井家に對して深々な謝意を表すると共に、本稿が筆者の獨力に依らざることを明記し、協力を惜まれなかつた幾多の人々に對し御禮申上ぐる次第である。）

（註一）本文四〇〜四二頁參照。

（註二）「三重縣の林業」昭和一五年・三重縣經濟部林務課刊・一三〇頁。

（註三）石渡貞雄・前掲書・二五三頁。

（註四）「貨幣制度調査會報告附錄」三〇八頁。埼玉縣外七縣人民生計及冠婚葬祭ニ關スル調査・(乙)三重縣調査參照。

（註五）藤田武夫・日本資本主義と財政・上・二〇六〜七頁。

（註六）土井家による資本貸付が廣範圍に亘り、且、自からも數種の事業―海運業・水産業等―を主催したことは、現地調査によつて確認することが出來た。

（註七）取得山林面積の狹隘化は、明治三一年以降增々顯著となつた。その概要は次の通りである。一筆當り面積は單位以下切捨。

年次	所有面積	前年度ニ比シ增加面積	筆數	一筆當リ平均面積	備考
	町反畝歩	町反畝歩		町反畝歩	
明治31年	185 9 3 8		159	1 2 3	相續
〃 35年	188 8 8 8	2 9 5 0	164	1 1 5	買得
〃 40年	190 7 3 4	1 8 4 26	170	1 1 2	同上

（註八）土井家に關する限り、「明治の初期には相當の森林小所有者が存在してゐた…それが明治三〇年前後景氣の變動に伴ひ、森林を賣却するものが增加し、漸次小所有林が大所有者に集中されて行く傾向…」（林業經濟實態調査・前掲書・一三九頁）は、さして顯著でなかつたといはざるを得ない。

主成分分析 (Principal Components) の經濟分析への應用

佐藤　保

一、序

普通使われる多元回歸と共に、その一般化を示すともいうべき多元解析の方法が發展してきた。これらの方法は經濟分析に應用されて有效な結果を與え得るかどうかは今後の問題ともいえるのであるが、Tintner はその著 Econometrics の中で判別分析 (Discriminant Analysis)、正準相關 (Canonical Correlation) 加重回歸 (Weighted Regression)、主成分分析 (Principal Component) の四つを說明し經濟分析に應用している。(註一) この內前者が非常に密接に關連を持つたものであることが Kullback によつて示されている。(註二) 判別分析とは若干の資料からある特性を選んで資料を二分しこれらのグループの間に或る意味に於て最もうまく區別を與えるような線型函數を決定することであり、主成分分析は變數の組をできる限り少いより基本的な成分の組に分解することと、總分散が各成分によつてどれくらい說明されうるかを示すものであり、正準相關は二組の變數があるとき各組に含まれる變數の一次結合によつて表し、これを正準變數と呼べばこれら二つの正準相關を極大にしようとするものである。加重回歸は變數のすべての組が誤差を伴う場合を取扱うものである。それぞれ特徵のあるものであるがここでは主成分分析について少し詳しく述べて見たいと思う。

註一　G. Tintner Econometrics 1952 p. 93.

註二　B. Kullback An Application of Information Theory of Multivariate Analysis. Annals of Mathmatical Statistics Vol. 23. 1952 p. 88.

この原理が經濟的變數に應用される時どのような意味をもちどのような利點をもつているであろうか、Tintner は次に Stone の使用した例を引いて十七の變數が三箇の要素によつてその分散の殆んど全部が說明され、相關を調べることによつて第一要素として所得、第二要素として所得の變化率、第三要素として傾向値を選び、結局その三つで變數の總分散の九七%が說明できる。それ故に十七の變數のこれら三つの說明變數に對する重回歸を計算することは有效であろう。と述べている。又主成分分析の他の應用は一般指數の面である。現在使われている指數、例えば生產指數を例にとれば附加價値をウェイトとした加重平均のものが多いが、これらの經濟的指數に對して統計的指數ともいうべきものを考えることができよう。卽ち主成分分析の應用によつて種々の生產量の分散の何%が說明されているかがわかるからである。(註三) 以下この原理と日本經濟への適用を示すことにする。

註三　G. Tintner Econometrics p. 109—110.

二、主成分分析の原理

先にも述べた如く原理は變數の組をより基本的な要素と呼ばれる獨立な成分の組に分解することであり、これは Hotelling によつて工夫された。又 Girshik は systematic component と誤差の和からなるところの變數の組がある時最小の誤差を伴

う變數の線型函數を發見すること、又要求される線型函數と各變數の相關係數の平方和が極大になるように與えられた變數の線型函數を見出すことを示した。以下順次やや煩雜なきらいもあるが途中計算もすべて省略せずに説明してゆくことにする。

今標準化された變數 $z_1 \cdots\cdots z_p$ の組をより基本的な變數の組、$u_1 \cdots\cdots u_p$ によつて置換えようとする。即ち元の變數をXとすれば、

$$z=\frac{X-\bar{X}}{s_x} \quad \begin{array}{l}\bar{X}\text{ は X の平均}\\ s_x\text{ は X の標準偏差}\end{array}$$

そこで

(1)
$$z_1=k_{11}u_1+k_{12}u_2+\cdots\cdots+k_{1p}u_p$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$z_p=k_{p1}u_1+k_{p2}u_2+\cdots\cdots+k_{pp}u_p$$

とした時變數 $u_1 \cdots\cdots u_p$ は主成分と云われる。

それらは orthogonal 即ち $\sum_{t=1}^{N} u_{it}u_{jt}=0 \quad (i \neq j)$ と假定される。これは次の事を意味する。即ちもとの系列 $X_1X_2 \cdots\cdots X_p$ に各々 $u_1 \cdots\cdots u_p$ 迄の主成分が働き各系列がN個あるとすれば各變數について各々違つた主成分を掛け合せて合計したものは零になるということである。Z_i 間の相關係數は

(2) $r_{ij}=k_{i1}k_{j1}+k_{i2}k_{j2}+\cdots\cdots+k_{ip}k_{jp}$

$$(i, j=12\cdots\cdots p)$$

の如く示される。これは次のように證明できる。一般に

$$Z=\sum_{i=1}^{p}\alpha_i u_i$$

に於いて u_i の平均 $E(u_i)=a_i$ 分散 $\sigma_{u_i}^2$ としの平均 $E(Z)=a$ 分散 σ_z^2 とすると u_iu_j の共分散は $r_{ij}\sigma_i\sigma_j$ で表わせる。平均値の性質から

$$a=E(Z)=\sum_{i=1}^{p}\alpha_iE(u_i)=\sum_{i=1}^{p}\alpha_i a_i$$

を得る。Zの分散は

$$\sigma_z^2=E[(z-a)^2]=E[(\sum_{i=1}^{p}\alpha_iu_i-\sum_{i=1}^{p}\alpha_ia_i)^2]$$
$$=E[(\sum_{i=1}^{p}\alpha_i\ (u_i-a_i)^2]$$
$$=E[\sum_{i,j=1}^{p}\alpha_i\alpha(u_i-a_i)(u_j-a_j)]$$
$$=\sum_{i,j=1}^{p}\alpha_i\alpha_jE[(u_i-a_i)(u_j-a_j)]$$
$$\sigma_z^2=\sum_{i,j=1}^{p}\alpha_i\alpha_j\sigma_i\sigma_jr_{ij}$$

又

$$z_1=\sum_{i=1}^{p}k_{1i}u_i \qquad z_2=\sum_{i=1}^{p}k_{2i}u_i$$

とし z_1 z_2 間の相關係數を求めれば

$$E(z_1)=\sum_{i=1}^{p}k_{1i}a_i \quad E(z_2)=\sum_{i=1}^{p}k_{2i}a_i$$

z_1 と z_2 の共分散は

$$E\{[z_1-E(z_1)][z_2-E(z_2)]\}$$
$$=E\{[\sum_{i=1}^{p}k_{1i}(u_i-a_i)][\sum_{j=1}^{p}k_{2j}(u_j-a_j)]\}$$
$$=\sum_{i,j=1}^{p}k_{1i}k_{2j}\sigma_i\sigma_jr_{ij}$$

相關係數の公式から

$$r_{z_1z_2}=\frac{\sum_{i,j=1}^{p}k_{1i}k_{2j}\sigma_i\sigma_jr_{ij}}{\sqrt{\sum_{i,j=1}^{p}k_{1i}k_{1j}\sigma_i\sigma_jr_{ij}}\sqrt{\sum_{i,j=1}^{p}k_{2i}k_{2j}\sigma_i\sigma_jr_{ij}}}$$

現在の場合は $r_{ij}(i \neq j)=0$ $\sigma_i=\sigma_j=1$ であるから式は

$$\frac{k_{11}k_{21}+k_{12}k_{22}+k_{13}k_{23}+\cdots\cdots+k_{1p}k_{2p}}{\sqrt{(k_{11}^2+k_{12}^2+\cdots\cdots+k_{1p}^2)(k_{21}^2+k_{22}^2+\cdots\cdots+k_{2p}^2)}}$$

しかるに

$$\sum_{i=1}^{p}k_{1i}^2=1 \qquad \sum_{i=1}^{p}k_{2i}^2=1$$

であるから結局

$$r_{z_1z_2}=k_{11}k_{21}+k_{12}k_{22}+k_{13}k_{23}+\cdots\cdots+k_{1p}k_{2p}$$

となるので

$$r_{z_iz_j}=k_{i1}k_{j1}+k_{i2}k_{j2}+\cdots\cdots+k_{ip}k_{jp}$$
$$(i, j=12\cdots\cdots p)$$

となる。

(3) $S_i=k_{1i}^2+k_{2i}^2+\cdots\cdots+k_{pi}^2$

$$(i=12\cdots\cdots p)$$

と置けば、これは明らかに i 番目の主成分 u_i のすべての標準化された變數 z_i の分散に對して寄與する部分を表わしている。そこで最初の主成分の寄與を極大ならしめるようとする。即ち條件(2)の下に

(4) $S_1=\sum_{i=1}^{p}k_{i1}^2$

を極大ならしめようとするわけである。Lagrange の乘數 μ_{ij}、を導入して

(5) $F=\sum_{i=1}^{p}k_{i1}^2-\sum_{j=1}^{p}\sum_{i=1}^{p}\sum_{s=1}^{p}\mu_{ij}k_{is}k_{js}$

をつくり k_{i1} について偏微分し簡單化すると

(6) $k_{i1}-\sum_{j=1}^{p}\mu_{ij}k_{j1}=0 \qquad (s=1)$

$$-\sum_{j=1}^{p}\mu_{ij}k_{js}=0 \qquad (s=2, 3\cdots\cdots p)$$

となる。次にこの組織に於ける各方程式に k_{i1} を乘じて i に關して合計すると

(7) $\sum_{i=1}^{p}k_{i1}^2-\sum_{i=1}^{p}\sum_{j=1}^{p}\mu_{ij}k_{i1}k_{j1}=0 \qquad (s=1)$

$$-\sum_{i=1}^{p}\sum_{j=1}^{p}\mu_{ij}k_{i1}k_{js}=0 \quad (s=2, 3, \cdots\cdots p)$$

(6)の最初の方程式から

$$\sum_{i=1}^{p}\mu_{ij}k_{i1}=k_{j1}$$

を得る。

$$\sum_{i=1}^{p}k_{i1}^2=\lambda_1$$

と置くと

$$(8)\quad \lambda_1-\sum_{j=1}^{p}k_{j1}^2=0 \qquad (s=1)$$

$$-\sum_{j=1}^{p}k_{j1}k_{js}=0 \qquad (s=2,3,\cdots\cdots p)$$

この組織に於ける各方程式に k_{is} を乘じて s について合計する。

$$(\lambda_1-\sum_{j=1}^{p}k_{j1}^2)k_{i1}-\sum_{j=1}^{p}\sum_{s=2}^{p}k_{j1}k_{js}k_{is}=0$$

$$=\lambda_1k_{i1}-\sum_{j=1}^{p}k_{j1}k_{j1}k_{i1}-\sum_{j=1}^{p}\sum_{s=2}^{p}k_{j1}k_{js}k_{is}=0$$

$$(9)\quad =\lambda_1k_{i1}-\sum_{j=1}^{p}\sum_{s=1}^{p}k_{j1}k_{js}k_{is}=0$$

$$(i=1,2\cdots\cdots p)$$

(2)を用いることによつて

$$r_{11}k_{11}+r_{12}k_{12}+\cdots\cdots+r_{1p}k_{p1}=\lambda_1k_{11}$$

$$(10)\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$r_{1p}k_{11}+r_{2p}k_{21}+\cdots\cdots+r_{pp}k_{p1}=\lambda_1k_{p1}$$

一次の同次方程式はその行列式が零である時にのみ零以外の解をもつ。

$$(11)\quad \begin{vmatrix} (1-\lambda_1) & r_{12}\cdots\cdots\cdots\cdots & r_{1p} \\ \cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots \\ r_{1p} & r_{2p}\ \cdots\cdots\cdots & (1-\lambda_1) \end{vmatrix}=0$$

(11)の最大根は最初の主成分に結びついており標準化された變數の分散の大部分を説明する。(8)を制限式として最初の主成分の係數を計算できる。同樣にして第二の主成分を考え λ_2 を計算すれば最初の主成分が取除けられた後、標準化された變數の分散を説明する。

次に第二の方法、これは經濟分析により有效であるといわれるが、先ず z_i は二つの部分から成ると考える。

$$z_i=m_i'+y_i'$$

ここで m_i' は眞の値、y_i' は確率誤差（random error）である。確率誤差は同じ分散 σ^2 をもち互に獨立である。問題とされているのは一次函數

$$u=k_1z_1+k_2z_2+\cdots\cdots\cdots\cdots+k_pz_p$$

を見出すことであるがこれは u の分散が 1 という條件の下に $k_1^2+k_2^2+\cdots\cdots\cdots+k_p^2$ に比例するような誤差の分散を極小にするような方法によつて行はれる。即ち

$$(12)\quad \sigma^2\sum_{i=1}^{p}k_i^2$$

を極小にすることになる。(12)は u の分散で、（ここでは u_1 のみが考えられているから時に記號はついていない）z_i の分散を説明した殘りを示しており $\sum k_i^2$ は z_i の分散の内 u で説明されない部分を表す。u の分散が1であるという條件の下に(12)を極小にする。

$$(13)\quad \sum_{i=1}^{p}\sum_{j=1}^{p}k_ik_jr_{ij}=1$$

u の分散は先に述べた如く求められ $\sigma_i=\sigma_j=1$ であるから(13)式を得る。Lagrange の乘數を導入して新しい函數をつくる。

$$(14)\quad \mathrm{F}=\sigma^2\sum_{i=1}^{p}k_i^2-\mu\sum_{i=1}^{p}\sum_{j=1}^{p}k_ik_jr_{ij}$$

k_i について偏微分し $\mu=\sigma^2/\lambda$ と置くことによつて(10)と同樣な組織を得る。即ち一般に

$$\mathrm{F}=\sum_{i=1}^{p}\sum_{j=1}^{p}k_ik_jr_{ij}=k_1k_1r_{11}+k_1k_2r_{12}+\cdots\cdots+k_1k_pr_{1p}$$

$$+k_2k_1r_{21}+k_2k_2r_{22}+\cdots\cdots\cdots+k_2k_pr_{2p}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+k_pk_1r_{p1}+k_pk_2r_{p2}+\cdots\cdots+k_pk_pr_{pp}$$

$$\frac{\partial \mathrm{F}}{\partial k_1}=(k_1r_{11}+k_2r_{12}+\cdots\cdots\cdots+k_pr_{1p})$$

$$+(k_1r_{11}+k_2r_{21}+\cdots\cdots\cdots+k_pr_{p1})$$

$$r_{ij}=r_{ji}$$

$$\frac{\partial \mathrm{F}}{\partial k_1}=2\sum_{j=1}^{p}k_jr_{1j}$$

これを現在の式に應用すれば

$$\mathrm{F}=\sigma^2\sum_{i=1}^{p}k_i^2-\mu\sum_{i=1}^{p}\sum_{j=1}^{p}k_ik_jr_{ij}$$

$$\frac{\partial \mathrm{F}}{\partial k_1}=\sigma^2 2k_1-\mu 2\sum_{j=1}^{p}k_jr_{1j}=0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial \mathrm{F}}{\partial k_p}=\sigma^2\,2\,k_p-\mu 2\sum_{j=1}^{p}k_jr_{pj}=0$$

これより

$$k_1r_{11}+k_2r_{12}+\cdots\cdots\cdots\cdots+k_pr_{1p}=\frac{\sigma^2}{\mu}k_1$$

$$k_1r_{21}+k_2r_{22}+\cdots\cdots\cdots\cdots+k_pr_{2p}=\frac{\sigma^2}{\mu}k_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$k_1r_{p1}+k_2r_{p2}+\cdots\cdots\cdots+k_pr_{pp}=\frac{\sigma^2}{\mu}k_p$$

となつて(10)と同樣になる。今度は(13)を制限式として k_i について解くことができる。これにより

$$u=k_1z_1+k_2z_2+\cdots\cdots\cdots\cdots+k_pz_p$$

を決定できる。次に

$$u=k_1z_1+k_2z_2+\cdots\cdots\cdots\cdots+k_pz_p$$

に於てもとの標準化された變數 $z_1z_2\cdots\cdots\cdots z_p$ と u との相關係數のすべての平方和が極大になるように係數 k_i を決定しよう。即ち

(15)　$T=\sum_{i=1}^{p}(\sum_{j=1}^{p}k_j r_{ij})^2$

を極大にしようとする。uとz_i間の相關係數は先に述べた相關係數の公式に於て、uの分散もz_iの分散も共に1であるから分母は1となり分子は$\sum_{j=1}^{p}k_j r_{ij}$となる。そこでTはすべての標準化された變數$z_1 \cdots\cdots z_p$とuとの相關係數の平方和である。條件式は前と同樣uの分散が1ということであるから Lagrange の乘數を導入して

(16)　$G=\sum_{i=1}^{p}(\sum_{j=1}^{p}k_j r_{ij})^2-\lambda\sum_{i=1}^{p}\sum_{j=1}^{p}k_i k_j r_{ij}$

(16)をk_sについて偏微分し0と置くことによつて(17)式を得る。

(17)　$\sum_{i=1}^{p}\sum_{j=1}^{p}k_j r_{ij} r_{sj}-\lambda k_j r_{js}=0 \quad (s=12\cdots\cdots p)$

今$[r^{ij}]$を相關行列$[r_{ij}]$の逆行列とする。この行列の要素を乘じ加えることによつて(10)と同樣の方程式組織を得る。即ち逆行列の要素の性質

$$r^{11}=\frac{R_{11}}{R} \qquad r^{12}=\frac{R_{21}}{R} \cdots\cdots r^{1p}=\frac{R_{p1}}{R}$$

（Rは相關行列の行列式，R_{ij}は餘因數を示す）

$$r^{11}r_{11}+r^{12}r_{21}+r^{13}r_{31}+\cdots\cdots+r^{1p}r_{p1}=1$$
$$r^{11}r_{12}+r^{12}r_{22}+r^{13}r_{32}+\cdots\cdots+r^{1p}r_{p2}=0$$
$$\cdots\cdots\cdots\cdots$$
$$r^{11}r_{1p}+r^{12}r_{2p}+r^{13}r r_{3p}+\cdots\cdots+r^{1p}r_{pp}=0$$

を利用し(17)をsについて解いて置き

$$\sum_{i=1}^{p}\sum_{j=1}^{p}k_j r_{ji}k_{1j}-\lambda\sum_{j=1}^{p}k_j r_{j1}=0 \quad r^{11}\ r^{21}\cdots\cdots r^{p1}$$
$$\sum_{i=1}^{p}\sum_{j=1}^{p}k_j k_{ji}k_{2j}-\lambda\sum_{j=1}^{p}k_j r_{j2}=0 \quad r^{12}\ r^{22}\cdots\cdots r^{p2}$$
$$\cdots\cdots\cdots\cdots$$
$$\sum_{i=1}^{p}\sum_{j=1}^{p}k_j r_{ji}k_{pj}-\lambda\sum_{j=1}^{p}k_j r_{jp}=0 \quad r^{1p}\ r^{2p}\cdots\cdots r^{pp}$$

とし例えば$(r^{11}r^{12}\cdots\cdots r^{1p})$迄を掛けて加えれば

$$k_1r_{11}+k_2r_{12}+\cdots\cdots k_pr_{1p}=\lambda k_1$$

が得られ以下順次同樣に掛け加えることによつて(10)と同樣の組織が得られる。即ち

$$u=k_1z_1+k_2z_2+\cdots\cdots+k_pz_p$$

に於けるk_iは誤差分散を極小にし、相關の平方和を極大にするという二面性を同時にもつように定められたことになるわけである。

註、G. Tintner Econometrics p. 102—106.
ウィルクス小河原正巳譯數理統計學 P49—51 參照。

三、生産指數への應用

以上の主成分分析の原理を應用して生産指數をつくつて見よう。その前に計算に使われる記號を説明しておく。一次の同時方程式の組織

$$(A-\lambda I)x=0$$

を解くとき、すべての潜在根が實根でそして二根は等しくないとすれば行列Aを高累迄高める。

$$(A^p-I\lambda^p)x=0$$

ことによつて容易に得られる。この組織を解くために

$$x_1^{(1)}=1 \quad x_2^{(1)}=1\cdots\cdots \quad x_n^{(1)}=1$$

をつくり　$A^p x^{(1)}=x^{(2)'}$

をつくる。$x_i^{(2)'}$を$x_i^{(2)'}$の最大のものである$x_j^{(2)'}$で割ることによつて$x^{(2)}$を得る。即ち

$$x_i^{(2)}=\frac{x_i^{(2)'}}{x_j^{(2)'}}$$

この手續は$x_i^{(k-1)}$と$x_i^{(k)}$の値が充分一致する迄續けられる。次いで

$$A\cdot x^{(k)}=x_1'$$

を計算すればx_1'の最大の要素が最大根λ_1である。（註二）

註一　G. Tintner Econometrics p. 351—354.

資料として一橋大學經濟研究所偏、解説經濟統計を使う。x_1は製造工業の内耐久財工業の生産指數、x_2は製造工業の内非耐久財工業、x_3は鑛業、x_4は農業の各生産指數を示す。x_1 x_2 x_3はGHQ生産指數を使うことにする。この指數は基準年度は一九三二—三六年、ウェイトは附加價値を用いラスペイレス式である。品目として鑛業九品目、製造工業、耐久財工業として、金属工業、機械器具工業、窯業、製材業で二三品目、非耐久財工業として纖維工業、化學工業、食品工業、印刷業で二五品目である。ウェイトを示すと

鑛業	10.43
製造工業耐久財	44.15
製造工業非耐久財	45.42

農業生産指數は農林省、農林統計月報と第二八次農林省統計表とを連絡してつくられたもの、品目八〇、（耕種六六、養蠶一、畜産一三）基準は一九三四—三六年、ラスペイレス式。農業生産指數のみ基準時が違つているのでこれを修正して前三者と合わせることにする。期間は昭和五年以前は資料がないのでやむをえず、昭和五年より一四年迄の一〇年間を取ることにする。昭和一四年は恐らく戰前として取り得る最後の年と思われる。指數を示すと、

昭和	5	6	7	8	9	10	11	12	13	14
x_1	66.0	57.8	65.5	80.1	99.4	118.8	136.1	173.4	194.7	214.4
x_2	80.4	81.3	85.0	93.2	103.8	106.9	111.0	131.0	129.4	128.6
x_3	88.2	81.3	81.8	92.5	100.8	106.5	118.3	127.1	136.3	143.3
x_4	103.5	92.9	97.3	109.7	91.5	97.3	104.2	109.9	106.7	115.4

各々の間の相關係數を求めると

$r_{12}=0.958 \quad r_{13}=0.994 \quad r_{14}=0.672$

$r_{23}=0.995 \quad r_{24}=0.581 \quad r_{34}=0.690$

從って相關行列と⑽⑾に相當するものは

$$\begin{array}{c|cccc} & x_1 & x_2 & x_3 & x_4 \\ \hline x_1 & 1.000 & 0.958 & 0.994 & 0.672 \\ x_2 & 0.958 & 1.000 & 0.995 & 0.581 \\ x_3 & 0.994 & 0.995 & 1.000 & 0.690 \\ x_4 & 0.672 & 0.581 & 0.690 & 1.000 \end{array}$$

$$1.000k_{11}+0.958k_{21}+0.994k_{31}+0.672k_{41}=\lambda k_{11}$$
$$0.958k_{11}+1.000k_{21}+0.995k_{31}+0.581k_{41}=\lambda k_{21}$$
$$0.994k_{11}+0.995k_{21}+1.000k_{31}+0.690k_{41}=\lambda k_{31}$$
$$0.672k_{11}+0.581k_{21}+0.690k_{31}+1.000k_{41}=\lambda k_{41}$$

$$\begin{vmatrix} 1.000-\lambda & 0.958 & 0.994 & 0.672 \\ 0.958 & 1.000-\lambda & 0.995 & 0.581 \\ 0.994 & 0.995 & 1.000-\lambda & 0.690 \\ 0.672 & 0.581 & 0.690 & 1.000-\lambda \end{vmatrix}=0$$

となる。以下先に述べた方法で計算を進める。

$$A=\begin{pmatrix} 1.000 & 0.958 & 0.994 & 0.672 \\ 0.958 & 1.000 & 0.995 & 0.581 \\ 0.994 & 0.995 & 1.000 & 0.690 \\ 0.672 & 0.581 & 0.690 & 1.000 \\ \hline 3.624 & 3.534 & 3.679 & 2.943 \end{pmatrix}$$

$$A^2=\begin{pmatrix} 3.357384 & 3.295462 & 3.404890 & 2.586458 \\ 3.295402 & 3.245350 & 3.343142 & 2.492326 \\ 3.404890 & 3.343142 & 3.454161 & 2.626063 \\ 2.586458 & 2.492326 & 2.626063 & 2.265245 \\ \hline 12.644134 & 12.376280 & 12.828256 & 9.970092 \end{pmatrix}$$

$$A^4=\begin{pmatrix} 40.3898 & 39.5887 & 41.0021 & 31.6978 \\ 39.5887 & 38.7807 & 40.1629 & 31.0339 \\ 41.0021 & 40.1629 & 42.1837 & 32.1582 \\ 31.6978 & 31.0339 & 32.1582 & 24.9201 \\ \hline 152.6784 & 149.5663 & 155.5069 & 119.8100 \end{pmatrix}$$

$$A^8=\begin{pmatrix} 5884.5238 & 5764.7308 & 5995.0281 & 4617.3256 \\ 5764.7308 & 5647.3770 & 5872.9834 & 4523.3287 \\ 5995.0281 & 5872.9834 & 6107.8450 & 4704.0253 \\ 4617.3256 & 4523.3287 & 4704.0253 & 3623.0146 \\ \hline 22261.6083 & 21808.4199 & 22679.8818 & 17467.6942 \end{pmatrix}$$

$x^{(1)}=\{1 \quad 1 \quad 1 \quad 1\}$

$x^{(2)\prime}=\{22261.6083 \quad 21808.4199 \quad 22679.8818 \quad 17467.6942\}$

$x^{(2)}=\{0.981558 \quad 0.961355 \quad 1.000000 \quad 0.770185\}$

$x^{(3)\prime}=\{20869.1771 \quad 20444.3350 \quad 21251.6909 \quad 16357.6703\}$

$x^{(3)}=\{0.982001 \quad 0.962001 \quad 1.000000 \quad 0.770558\}$

$x^{(4)\prime}=\{20.877.2303 \quad 20452.2245 \quad 21269.5088 \quad 16381.4333\}$

$x^{(4)}=\{0.981557 \quad 0.961575 \quad 1.000000 \quad 0.770184\}$

$x^{(5)\prime}=\{20870.4348 \quad 20445.5672 \quad 21262.5858 \quad 16376.1018\}$

$x^{(5)}=\{0.981557 \quad 0.961575 \quad 1.000000 \quad 0.770184\}$

$x_1{}'=\{3.414310 \quad 3.344384 \quad 3.463862 \quad 2.678465\}$

$x_1=\{0.985695 \quad 0.965507 \quad 1.000000 \quad 0.773260\}$

$\lambda_1=3.464$ とする。

$$1.000k_{11}+0.958k_{21}+0.994k_{31}+0.672k_{41}=3.464k_{11}$$
$$0.958k_{11}+1.000k_{21}+0.995k_{31}+0.581k_{41}=3.464k_{21}$$
$$0.994k_{11}+0.995k_{21}+1.000k_{31}+0.690k_{41}=3.464k_{31}$$
$$0.672k_{11}+0.581k_{21}+0.690k_{31}+1.000k_{41}=3.464k_{41}$$

$$-2.464k_{11}+0.958k_{21}+0.994k_{31}=-0.672k_{41}$$
$$0.958k_{11}-2.464k_{21}+0.995k_{31}=-0.581k_{41}$$
$$0.994k_{11}+0.995k_{21}-2.464k_{31}=-0.690k_{41}$$
$$0.672k_{11}+0.581k_{21}+0.690k_{31}=2.464k_{41}$$

$$k_{11}=1.3000k_{41}$$
$$k_{21}=1.2738k_{41}$$
$$k_{31}=1.3189k_{41}$$

$k_{11}{}^2+k_{21}{}^2+k_{31}{}^2+k_{41}{}^2=3.464$ より

$k_{41}{}^2=0.572373 \quad k_{41}=0.756$

結局

$k_{11}=0.983 \quad k_{21}=0.963 \quad k_{31}=0.997 \quad k_{41}=0.756$

$k_{11}{}^2=0.966 \quad k_{21}{}^2=0.927 \quad k_{31}{}^2=0.994 \quad k_{41}{}^2=0.572$

なる。以上の結果により次の事が云える。

$\lambda_1=3.464$

より四つの標準化された變數の總分散は四であるから最初の主成分は總分散の約八六%を説明する。

$k_{11}=0.983 \quad k_{21}=0.963 \quad k_{31}=0.997 \quad k_{41}=0.756$

$k_{11}{}^2=0.966 \quad k_{21}{}^2=0.927 \quad k_{31}{}^2=0.994 \quad k_{41}{}^2=0.572$

より最初の主成分はz_1の分散の約九七%、z_2の約九三%、z_3の約99%、z_4の約五七%を説明する。次に指數をつくって見よう。λ_1の値と各k_{i1}の比例關係は同じであるから

$$\sum_{i=1}^{4}\sum_{j=1}^{4} k_i k_j r_{ij}=1$$

の中に導入してk_iを決定できる。その結果は

$k_1=0.283 \quad k_2=0.277 \quad k_3=0.287 \quad k_4=0.218$

そこですべての變數に對して誤差の分散を極小にし、相關係數の平方の極大にされた和をもつところのuは

$$u=0.283z_1+0.277z_2+0.287z_3+0.218z_4$$

によつて示される。以上が統計的意味であるが次にその經濟的意味を考える。先ずk_{31}が最も大きい。從つてz_3の係數が最も大きいということはその系列が他の系列に對して最も高い相關度をもつていたと云うことである。即ち鑛業が他のものに最も大きな相關をもつていたと云うことであり生產指數全體の動きを見る時最も代表性をもつていることがわかる。しかしこれはそ

のまゝ經濟的重要性をもつていると云うことではない。その意味では從來の指數の方がすぐれているといえるであろう。その點で主成分分析の原理が異質的な集りの間に適用されてどれほどの經濟的意味を持ちうるかは疑問である。同種類のものの間でならばその代表性と重要性が一致する可能性が強いからその動きの傾向を知る上に有效であると考えられる。次にTintnerの計算したアメリカのものと比べて見よう。Tintner の分析は一九一九年—一九三九年迄のもので日本の一九三〇年—一九三九年迄のものとは一概に比べることはできないが、彼のつくつた指數は

$$u=0.269462z_1+0.292772z_2+0.313815z_3+0.269918z_4$$

となつている。$z_1\ z_2\ z_3\ z_4$は日本の場合と同樣の意味をもつている。即ちこの指數で鑛業が最も大きなウエイトを持つている。これは單に偶然の一致であるか、或は鑛業というものが生産發展の中で中間的性格を持つているために最も大きな代表性をもつているのか判然とはわからないが興味ある一致といえるであろう。しかし以下の順序は違う。アメリカは$z_3\ z_2\ z_4\ z_1$の順、日本は$z_3\ z_1\ z_2\ z_4$の順である。ここで注目されるのはアメリカの場合は四つの係數があまり違わない。即ち部門分割ができにくいと云うことである。これに對して日本の場合、はじめの相關係數の計算から豫想される如く鑛工業と農業では相當に係數の値が違う。これは最初の主成分は鑛工業の分散の大部分を説明し、第二の主成分は農業部門の分散を説明するといつてよいであろう。これによつてこの期間に於ては鑛工業部門と農業部門とはあまり關係がない經濟を想像できる。このことは單に指數の動きだけを見ても云えることではあるが主成分分析も又よくこのことを示しているということである。次に農業生産指數を除いて他の三つのものでもう一度指數をつくつて見る。相關行列は次の如くなる。

$$\begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \end{array}\begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{pmatrix} 1.000 & 0.958 & 0.994 \\ 0.958 & 1.000 & 0.995 \\ 0.994 & 0.995 & 1.000 \end{pmatrix} \end{array}$$

これより同樣な方法で指數をつくれば

$$u=0.323z_1+0.323z_2+0.327z_3$$

となる。相關行列からわかる如くz_2の係數はz_1の係數より大きくなるはずであるがその差は殆んどないので小數第三位ぐらいでは差はないと考える。又相關行列から豫想される如く各ウエイトは殆んど等しい。今これと前に述べたGHQ鑛工業綜合指數との相關を調べてみよう。この相關が高いであろうことははじめから豫想のつく所である。何故なら相關は對應關係が問題となるのであるからすべてのものが同じ方向に動いている時、例えウエイトがどう變ろうとも相關が高くなることは當然だからである。しかし係數を調べることは意義のあることであるから、もとの資料を標準形に直すと

昭和	5	6	7	8	9	10	11	12	13	14
z_1	−0.906	−1.228	−1.216	−0.705	−0.317	−0.051	+0.500	+0.911	+1.341	+1.668
z_2	−1.293	−1.246	−1.052	−0.623	−0.068	+0.094	+0.308	+1.356	+1.272	+1.230
z_3	−1.007	−1.189	−1.016	−0.747	−0.391	−0.033	+0.285	+0.974	+1.367	+1.730

これよりuは

昭和	5	6	7	8	9	10	11	12	13	14
u	−1.038	−1.187	−1.064	−0.672	−0.250	+0.002	+0.354	+1.048	+1.289	+1.500

鑛工業生産指數（vで表わしておく）は

v	76.1	73.4	78.2	88.7	102.1	110.9	120.0	145.8	154.9	163.5

兩者の間の相關係數を計算すると

$$r_{uv}=0.993$$

となつて非常に高い相關が示される。Tintner は卸賣物價指數で、農産物價格と食料價格とその他のものの價格と三つに分けて指數をつくり普通の指數との相關を調べ相關係數 0.991 と計算されたことからその指數は價格の變動の大部分を充分よく説明するものであり殘りの變動は無視されると述べている。しかし必ずしも常にそうなるとはいえないであろう。例えば日本の場合數値の小さい方から順位をつけてみると、

昭和	5	6	7	8	9	10	11	12	13	14
製造工業非耐久財	1	2	3	4	5	6	7	10	9	8
製造工業耐久財	3	1	2	4	5	6	7	8	9	10
製造業綜合	2	1	3	4	5	6	7	8	9	10
鑛業	3	1	2	4	5	6	7	8	9	10
鑛工業綜合	2	1	3	4	5	6	7	8	9	10

となる。前に述べた如くウエイトは耐久財、非耐久財共殆んど等しく約四五、鑛業一〇であるから綜合指數には製造業の綜合と同じ順位がついている。これに對して現在つくられた指數の順位は鑛業に最もウエイトが置れるから順位は

3　1　2　4　5　6　7　8　9　10

となつてもとの鑛業と一致している。そこで鑛工業がいつも平行的に進むものであるならば相關は高くなるが違つた動きをする場合に相關は低くなるであろう。即ち相關が高いということは經濟發展の正常な形を示すものであり相關の低いということは不均衡的動きを表すといつてよいであろう。次にこの指數の性質を二三述べれば、最も代表的、或は中間的なものにウエイトを置く以上その動きの變動性は當然小さくなるであろう。又

普通の指數では一次の同次函數、卽ち係數の總和が1になることが要求されるがこの指數ではそのようになる場合もあろうが一般的にはそうならない。しかしこの指數は普通の指數とは別の原理から作られており標準形を使つていることからuは期待値0、分散1で標準正規分布をなし、常に0を中心として動きを示すので普通の意味での倍數的關係は成立しない。卽ちこの指數はそのままでは數値的意味をあまりもたないのである。示すものは傾向的な動きでありその意味で一次の同次函數でなければならないという制限はこの場合あまり大きな意味を持たないのではないかと考えられる。主成分分析の應用はなを多くの制限や問題を含んでいる。例えば從來の指數と如何なる數値的對應關係をつくるか、或は生產に於ける代替性の問題等が擧げられるのであろう。しかし同種類の物の傾向的動きの豫測に使うことは有效であろうし今後の改善發達が期待されるであろう。最後に相關圖と兩指數の動きを圖示して終りとする。

附記　本稿は多くの方々の御指導を受けたが特に田島一郎先生及び小尾惠一郎先生に多大の御指導をいただいたことに對して感謝の意を表したい。