農業地域区分における主成分分析の利用について

川口, 雅正
九州大学農学部農業計算学講座

李, 鍾相
九州大学農学部農業計算学講座

https://doi.org/10.15017/23440

出版情報：九州大學農學部學藝雜誌. 47 (3/4), pp.163-179, 1993-03. 九州大學農學部

九大農学芸誌 (Sci. Bull. Fac. Agr., Kyushu Univ.)
第47巻 第3・4号 163-179 (1993)

農業地域区分における主成分分析の利用について

川口雅正・李 鍾相
九州大学農学部農業計算学講座

(1992年10月30日受理)

On the Application of Principal Component Analysis in Agricultural Regional Land Classification

Tsunemasa KAWAGUCHI and Jong-Sang LEE
Seminar of Econometric Analysis in Agriculture, Faculty of Agriculture
Kyushu University 46-07, Fukuoka 812

I 緒 言

国際分業を基礎とする日本経済の昭和30年代以降の高度成長は，日本国民の平均的所得水準を大幅に上昇させたが，その反面都市と農山村のアンバランスな発展，農林業と非農林業とのアンバランスな発展をもたらし，また種々の環境問題をひきおこした．このような状況の下で，おもに昭和50年代以降，農業構造の改善と農業生産力の維持発展，農林業の公益的機能の維持発展に向けて，生産および居住の場としての農村の総合的な整備の必要性が強く指摘されるようになり，地域計画ないし農村計画に関する関心が高まり多くの研究が行われるようになった．たとえば地域計画ないし農村計画について総論的な考察を行ったものとして，天間(1966)，和田(1975 a, b, 1980)，窪谷(1978, 1988)，武藤 (1979)，相川 (1985) などの研究が知られている．

地域計画に関する考察は学際的に種々の側面を考慮して行う必要があるが，中でも空間的要素を無視することは許されない．空間的要素は一般経済理論においては無視されがちであるが，空間的要素がなければ地域に関する問題もおこりえない．このような空間的要素の役割に焦点を当てた社会経済問題の研究は地域科学 (Regional Science) と呼ばれ，従来の立地論を含む新たな研究分野を形成しつつある (Isard, 1975)．しかし農業経済の分野においては，空間的要素の重要性は早くから認識されていた．というのは，農業生産は空間的に大きな広がりを持ち，その主要かつ再生産不可能な労働手段である土地（気候を含む）は位置により大きく異なり，また農産物の物的特性のゆえにその市場条件も位置により大きく異なるからであろう．

空間は連続した広がりをもち，空間的要素は無数の側面をもっている．従って空間的要素そのものをあるがままに認識することは不可能である．そこで分析対象とする側面の理論的概念を形成し，その理論的概念（たとえば立地論的概念など）による単純化を行い，その単純化に基づいて空間的要素の役割を認識するのに好都合な空間的な単位を明らかにし，連続した空間的広がりをその単位を基に区分して（つまり地域区分を行い）空間的要素に関する認識を行わざるをえないであろう．かくて空間的要素の認識の手段として適切な地域区分は実際上不可欠であろう．

地域区分を行う場合，区分の対象となる連続した空間的広がりの範囲を明確にし，上述のように適切な単位（空間的単位）を選ぶ必要がある．そして分析対象とする空間的要素の役割を認識するのに好都合な標識を選び，それぞれの単位の標識を調査等によって明らかにする必要がある．ただし最初の段階では範囲も単位も標識も理論的かつ抽象的にのみ規定され，単位間に見られる社会経済的空間的諸関係も考慮されるが，実際に調査等によって標識を明らかにする際には，単位間に見られる社会経済的空間的諸関係のほとんどは捨象せざるをえず，また調査等に伴う社会的制約や技術的制約のため操作可能な実際的な範囲や単位や標識を利用せざるをえない．このようにして各単位の標識を明らかにし，その標識に基づいて分析対象とする空

間的要素の役割を認識するのに好都合な単位の分類区分（つまり地域区分）を行い，調査等の過程でほとんど捨象されていた単位間の社会経済的空間的諸関係も加象して，空間的要素の認識のためにその地域区分を利用することとなる．

以上のような地域区分の手順は，社会経済統計学でいう統計の作成および加工利用の手順と酷似したものである（川口，1981，26頁および40―41頁）．ただし調査等の過程でほとんど捨象されていた単位間の空間的諸関係が，単位の分類区分の段階で加象されざるをえないという点は，地域区分の手順に独特な側面であり，鈴木（1973）はこれを「分級図の作成問題」と呼んでいる．また実際に調査等によって明らかにされる各単位の標識としては，単位間に見られる社会経済的空間的諸関係はほとんど含まれていないのであるから，結節地域（分極地域や核地域ともいう）や計画地域という観点から単位の分類区分を行うことは困難であり，実際上は均質地域（等質地域ともいう）という観点から地域区分を行わざるをえないであろう．そしてその地域区分を利用する過程で，捨象されていた単位間の諸関係を加象することにより（児島，1961も参照），結節地域や計画地域に関する認識を深めざるをえないであろう．かくて本稿では均質地域という観点からの地域区分に焦点を当て，そこで利用される主成分分析の妥当性について考察を行いたい．

なお地域区分を行う際の一つの基本的な問題点は，理論的な範囲や単位や標識と，操作可能で実際的なものとして利用した範囲や単位や標識とのギャップに関する問題であり，社会経済統計学でいう信頼性の問題（川口，1981，39頁；喜多，1982，13―19頁）に相当するものである．つまり実際に利用した範囲や単位や標識は分析の対象とした空間的要素の客観的作用を正しく反映しうるかどうかという問題である．言うまでもなく，望ましい地域区分はその空間的要素の客観的作用を正しく反映し，その作用が安定し固定していると見なしうる将来に向って，その作用の認識に役立つものでなければならない．同時に種々の社会的制約や技術的制約の下で簡便かつ実際的なものでなければならず，両者を両立させることは容易ではない．このような問題点については，本稿ではふれないが，たとえば空間的要素の作用を各単位の期待農業所得という側面でとらえた金沢（1973），和田（1973，1975 c, d，1976 a, b，1980）の研究を参照していただきたい．

II 科学的認識過程における統計的操作の役割

一般に科学的認識は次のような過程を経て深まってゆく．最初に過去の何らかの理論的仮説と経験に基づいて一定の理論的仮説が形成され，その理論的仮説は新たな経験に基づいて検証される．検証の結果，理論的仮説と経験との間に矛盾が見いだされるならば，その理論的仮説は経験に基づいて修正され，その修正された理論的仮説は更に新たな経験に基づいて検証される．このような理論的仮説の形成（修正）と検証を通して科学的認識は深まってゆく．統計資料という形で経験を獲得しかつ理論的仮説の形成と検証のために統計資料を利用するための手法である統計的操作はその過程でどのような役割をはたすのであろうか．

統計的操作は理論的仮説の形成のために利用されることもあれば，また理論的仮説の検証のために利用されることもある．そして仮説の形成のために利用される統計的操作と仮説の検証のために利用される統計的操作はその性格を異にする．この性格の違いに注目することは，主成分分析の性格を考える上で極めて重要である．そこでこの性格の違いについて少し詳しく分析しておこう．

林（1974，29―32頁）や林ら（1970，2―5頁）は，寺田寅彦の言葉を借りれば，仮説形成のための統計的操作は「最初のことば」に相当し，仮説検証のための統計的操作は「最後のことば」に相当するという意味のことを述べており，また両者の機能は大きく異なり，両者を混同すれば大きな誤りをおかすと述べている．つまり「最初のことば」は混とんとした事象に斧を入れ，その中に何らかの手がかりを得ようとする場合に有効なものであり，「ああであろうか，こうであろうか」と試行錯誤的に考える際に有効なものである．それは知識のひらめきを誘うものであり，そのひらめきによって得られるものは一つの解釈であり仮説であって，確かに「こうである」と止めを刺すようなものではない．両者を混同することは大きな誤りであり，止めを刺すには「最後のことば」によらなければならない．

理論的仮説の形成過程は大変複雑で試行錯誤的なものであり，仮説は帰納によっても形成され，また発想によっても形成される（川口，1981，31―32頁）．従って仮説形成のための統計的操作は，統計資料の示唆する情報をおおらかに探索し発見するためのものでなくてはならない．記述統計学や従来の社会経済統計学は，

統計資料を加工整理して統計的法則（原因機構が明らかにされていない安定的結果ないし規則性）を導出し，統計資料の示唆する情報を探索し発見するための手法の研究に大きな関心を払った．その後統計資料に一定の確率分布モデルを仮定し，そのモデルの妥当性を確率論的に検証することを主な目的とする推測統計学の発展に伴い，「この統計資料から何が言えるか」といったタイプの，統計資料の示唆する情報をおおらかに探索し発見するための，記述統計学的手法は時代おくれの古くさいものとして不当にも軽視されるようになった．けれども近年のコンピューターの普及とそれを基礎とする多変量解析等の進歩により事情は大きく変わりつつある．というのは，多変量解析等の手法が記述統計学的手法に変わって，統計資料の示唆する情報をおおらかに探索し発見するために好んで用いられるようになってきたからである．

理論的仮説の検証のための統計的操作を「確証的(confirmatory)」統計的操作と呼び，上述のような仮説形成のための統計的操作を「探索的(exploratory)」統計的操作と呼ぶことにしよう．近年確証的統計的操作の幅（視野）のせまさが問題とされるようになり，また探索的統計的操作の重要性が再認識されるようになってきている（竹内ら編，1989，12—13頁，316—317頁，950—951頁；柳井•髙木訳，1981）．このような流れを背景として，多変量解析等の探索的統計的操作の利用がコンピューターの普及につれて今後ますます増加するものと考えられる．しかしそれとともに，林(1974)や林ら(1970)の危惧が現実のものとなる可能性が大きいように思われる．というのは，多変量解析等の探索的統計的操作と種々の確証的統計的操作は数学的にますます複雑になってきており，応用を主とする大多数の人々にとって，両者は紛らわしくて区別しにくく，またそれらが「最初のことば」なのか「最後のことば」なのか理解しにくくなってきていると考えられるからである．かくて「最初のことば」を「最後のことば」として述べるような大きな誤ちが今後おこる可能性が大きいように思われる．次節ではこのような観点より主成分分析の利用について考察を行うこととする．

III　主成分分析の探索的利用と確証的利用

今 N 個の個体からなるある集団 G があり，各個体は P 種類の標識によって識別され，それらの標識は量的な標識，つまり変数（変量）であるものとしよう．従って P 個のそれらの変数を X_1, X_2, ……, X_p, 第 i 個体におけるそれらの変数の値を，A_{i1}, A_{i2}, ……, A_{ip}で表すことにすれば，第 i 個体は簡潔に$(A_{i1}, A_{i2}, \cdots\cdots, A_{ip})$と表される．なお地域区分の場合を考えると，空間的単位が個体に当たり，空間的要素の役割を認識するのに好都合なものとして選ばれた空間的単位の量的標識が変数 X_1, X_2, ……, X_pに当たる．集団 G はこの場合，区分の対象となる連続した空間的広がりの範囲に含まれる，すべての空間的単位の集団である．そして通常変数 X_1, X_2, ……, X_pの間にはかなりの相関があり，個体数 N は変数の個数 P より大きい．集団 G とその個体は表1によって簡潔に要約される．表1には基準化した変数 x_1,, x_2, ……, x_pと第 i 個体におけるそれらの値 a_{i1}, a_{i2}, ……, a_{ip}も以後の説明の都合上つけ加えられている．また個体識別のためではな

表1　集団 G の構成

個体	変数	基準化した変数	Y y
	$X_1, X_2, \cdots\cdots, X_P$	$x_1, x_2, \cdots\cdots, x_P$	
1	$A_{11}, A_{12}, \cdots\cdots, A_{1P}$	$a_{11}, a_{12}, \cdots\cdots, a_{1P}$	Y_1 y_1
2	$A_{21}, A_{22}, \cdots\cdots, A_{2P}$	$a_{21}, a_{22}, \cdots\cdots, a_{2P}$	Y_2 y_2
⋮	⋮ ⋮ ⋮ ⋮ ⋮	⋮ ⋮ ⋮ ⋮ ⋮	⋮ ⋮
N	$A_{N1}, A_{N2}, \cdots\cdots, A_{NP}$	$a_{N1}, a_{N2}, \cdots\cdots, a_{NP}$	Y_N y_N
平均	$\bar{X}_1, \bar{X}_2, \cdots\cdots, \bar{X}_P$	0　0 ……　0	$\bar{Y}$　0
標準偏差	$S_1, S_2, \cdots\cdots, S_P$	1　1 ……　1	S_Y　1

（備考）平均、標準偏差および変数の基準化の算式は次のとおりである。

$$\bar{X}_j=(1/N)\sum_{i=1}^{N}A_{ij}$$

$$S_j=\sqrt{(1/N)\sum_{i=1}^{N}(A_{ij}-\bar{X}_j)^2}$$

$$a_{ij}=(1/S_j)(A_{ij}-\bar{X}_j)$$

$$i=1,2,\cdots\cdots,N$$
$$j=1,2,\cdots\cdots,P$$

いが，別の変数Yとそれを基準化したyの第i個体における値Y_iとy_iもつけ加えられている．

主成分分析は互いに相関のある多数の変数のもつ情報を，互いに無相関な小数の合成変数に要約し，分析の見通しをよくする手法であり，多変量解析の中でも最も基本的な手法と考えられている．そしてその手法の変形まで考慮に入れると，多様な利用のしかたが可能であり，特定の問題への実際の利用に当たっては，利用のしかたを明確にするためにいくつかの重要な選択を行わねばならない．その変形や多様な利用のしかたについては，Rao (1964)，奥野ら (1971, 159-257頁, 384-391頁)，Kendall *et al.* (1983, pp.320-339)，Jolliffe (1986) を参照していただきたい．ここでは農業地域区分への利用に当たっての選択について考えてみよう．

最も基本的かつ重要な選択は，主成分分析が適切な統計的操作として必要とされるのかどうか，という点についての判断を行うことであろう．つまり，分析対象とする空間的要素の役割がある1つの変数Yによって測られ，各個体におけるYの値が何らかの方法で直接測定される場合には，Yの値に基づいて直接地域区分を行うことができ，主成分分析は不要である．たとえば，空間的要素の作用を各集落の期待農業所得でとらえる場合，各集落の期待農業所得（または実際上期待農業所得と見なしうるもの）を直接測定しうるのであれば，その測定された期待農業所得に基づいて集落を単位とした地域区分を行うことができ，主成分分析は不要である．なお全く異なる観点からではあるが，山中 (1980, 120-138頁) は，生産農業所得統計に示される市町村単位の生産農業所得と，農業センサス農業集落カード等より，農業集落単位の生産農業所得の推計を行っている．ただし推計の精度については検討を要するであろう．また山中 (1980, 70-98頁) は同じ地域を対象とし，経営規模に関する12変数を用いて農業集落を単位とする主成分分析（考え方としては確証的な分析）を行っているが，その第1主成分と上述の生産農業所得の推計値との間には高い正の相関があることを指摘している．

空間的要素の役割がある1つの変数Yによって測られ，変数Yと変数X_1, X_2, ……, X_pとの関数関係が何らかの確証的統計的操作によって計測される場合には，各個体におけるX_1, X_2, ……X_pの値より各個体のYの値を計算で求めることができ，従って各個体のYの値を直接測定しうる場合と同様に，Yの値に基づいて直接地域区分を行うことができるので，主成分分析は不要である．なお能美 (1988) は，生産農業所得統計に示される市町村単位の生産農業所得を従属変数とし，農業センサス市町村統計書に示される10個の変数（同集落カードにも示されている）を独立変数とする重回帰分析を行い，その重回帰式と農業集落カードを利用して各農業集落の生産農業所得の推計を行い，その推計された各集落の生産農業所得に基づいて地域区分を行っている．ただし市町村単位の重回帰式を集落単位にもそのまま利用しうるのかどうか全く検討されておらず，推計の精度については検討を要する．また能美はその10個の変数（上記の山中の場合と比較するとかなり性格が異なる）を利用した主成分分析（探索的な分析）を試み，その主成分と生産農業所得との相関がほとんどないことを指摘し，主成分に関する解釈ないし仮説を安易に「最後のことば」とみなしてはいけないことを指摘している．

以上の二つの場合と同様に，分析の対象とする空間的要素の役割が，何らかの確証的統計的操作によって数量的に変数X_1, X_2, ……X_pから直接推計される場合には，原則として主成分分析は不要であると考えられる．ただし精度の高い推計は通常容易ではない．従って主成分分析が必要とされるのは，そのような直接的な推計が困難な場合だけであり，その空間的要素の役割と変数X_1, X_2, ……, X_pとの漠然とした相関関係しか分からない場合だけであろう．以下このような場合を前提として，農業地域区分への主成分分析の利用に当たっての主な選択について考えてみよう．

第一に，研究の対象となる母集団とその個体を研究の目的に照らして明確にし，かつ主成分分析の直接の対象となる集団Gと母集団との関係を明確にする必要がある．ここで考えるべき点は，集団Gは母集団に外ならず（または母集団とみなされ），最終的な研究の対象となるのか，それとも集団Gはそれと異なる母集団から無作為に抽出された一つの標本にすぎず（または標本とみなされ），集団Gの主成分分析はその母集団に関する統計的推論をするための手段にすぎないのか，という点である．地域区分の場合を考えると，集団Gは研究の対象となる母集団と同じであり，標本とは考えにくい．従って確率論を基礎とする統計的推論は問題とならず，母集団たる集団Gの集団的性格を記述統計学的に要約し明らかにすることだけが問題となる．

なお主成分分析が統計的推論のために利用されることは種々の制約のためにまれである (Jolliffe, 1986, pp. 39-46)．統計的推論の中でも，実際の分析でとり上

げ考慮すべき主成分の個数を決めるための仮説検定（Jolliffe, 1986, pp. 43-45 ; Anderson, 1963）は注目されるが，実際の経験によるとこれも有用なものとは考えにくく，多くの場合経験的な規則が利用されている（Jeffers, 1967 ; 奥野ら，1971，194頁および226-227頁：Jolliffe, 1986, pp. 92-114）．また変数の分散や共分散を求めるための，表1に示すような算式の分母は，集団 G を母集団とみなすか（この時 N），標本とみなすか（この時 $N-1$）によって若干異なるが，この違いは相関係数の計算や本稿の以後の分析に影響を与えることはない．そしてたとえ集団 G が本来標本であったとしても，多くの場合そうであるように，その集団的性格を記述することだけが問題とされるのであれば，それは集団 G を母集団とみなすことに等しい．

第二に，どのような変数を利用するか，明確にする必要がある．この点は上述のように，研究の目的によって規定されるものであり，母集団や個体の選択と共に最も重要な点である．地域区分の場合には，空間的要素の役割を認識するのに最も適した変数を利用する必要がある．なお統計資料等として直接利用しうる変数だけでなく，それらを対数変換したものも，場合によっては非線形の変数間の関係を分析するために利用する方がよいかもしれない（Jeffers, 1967 ; Jolliffe, 1986, pp. 209-211 および pp. 226-227）．しかしここで統計的推論の確率論的前提条件を満たすために変数の変換をする必要はない．また地域区分の場合に，そのような変数の変換が必要かどうか，今後の経験によって判断せざるを得ない．

第三に，各変数の値をどのような測定尺度（原点と測定単位によって規定される）によって測定するか，明確にする必要がある．つまり表1の第 i 個体における変数 X_1，X_2，……，X_pの値 A_{i1}，A_{i2}，……，A_{ip}はある測定尺度によって測定されたものであり，異なった測定尺度を利用すれば異なった値が得られ，特に測定単位が異なれば各変数の分散や共分散も異なる．従って測定尺度の変更，特に測定単位の変更は，各変数の値を直接利用する主成分分析の結果に複雑な影響を与える．このことは表1の記号を利用して $N\times P$ 行列 A と $P\times P$ 対角行列 S

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1P} \\ a_{21} & a_{22} & \cdots\cdots & a_{2P} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots\cdots & a_{NP} \end{bmatrix},\ S=\begin{bmatrix} S_1 & 0 & \cdots\cdots & 0 \\ 0 & S_2 & \cdots\cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots\cdots & S_P \end{bmatrix}$$

を導入すると，行列につけたプライムはその転置を示すものとして，変数 X_1，X_2，……，X_pの分散共分散行列 W が

$$W=\frac{1}{N}(AS)'AS=S\ (\frac{1}{N}A'A)S=SRS$$

ここで，$R=\frac{1}{N}A'A$ は変数 X_1，X_2，……，X_pの相関行列

と表され，A と R は測定尺度を変更しても不変であるが，S と W は測定単位を変更すると変化し，従って W の固有値と固有ベクトルも複雑に変化することより明らかである．

各変数の測定尺度が異質で性格の異なるものである場合に，測定尺度を選択するための適切な基準は何もなく，測定尺度が異なれば各変数の値を直接利用する主成分分析の結果は複雑に変化し，同種の分析の相互比較も困難となり，かつ S_1，S_2，……，S_pの値が大きく異なるような場合には実際上有益な分析結果が得られることは少ないと言われている．また統計的推論をする上では各変数の値を直接利用する方が望ましいが，ここでは統計的推論は問題とならない．このような問題点は，測定尺度の選択に影響されることのない基準化した変数 x_1，x_2，……，x_pを利用して主成分分析を行うことによって軽減される（Jolliffe, 1986, pp. 16-21 および pp. 32-35）．基準化した変数の分散共分散行列は相関行列 R に外ならず，測定尺度によって影響されることはない．また基準化した変数を利用する主成分分析は計算上も便利であり，同種の分析の比較にも適している．地域区分の場合を考えると，種々の測定尺度の利用が考えられるので，また同種の分析を比較検討する上からも，基準化した変数を利用する方がよいように思われる．

以上のような選択に基づいて主成分分析を行い，求められた主成分（合成変数）のうち分散の大きい方からある少数個を何らかの形で利用して，均質地域という観点からの農業地域区分を行うものとしよう．この時問題となるのは，主成分とその解釈を探索的な観点から利用するのか，それとも確証的な観点から利用するのかという点である．

Rao（1964, pp. 329-330 および pp. 344-356）は，どのような場合にはっきりとした目的をもって確証的な観点から主成分分析を利用しうるのか，またどのような場合に研究の初期に探索的な観点から主成分分析を利用しうるのかという点について検討を行った．しかし Jeffers（1967, p. 236），奥野ら（1971，4頁）お

よびJolliffe (1986, p. 20, p. 41およびp. 46) が述べているように，実際上主成分分析は多くの場合純粋に記述的かつ探索的な観点から利用されており，何らかのはっきりとした目的に従って確証的な観点から利用されることは極めて少ないと考えられる．

農業地域区分の場合においても，主成分分析は多くの場合純粋に記述的かつ探索的観点から利用されており，確証的な観点から利用されることは極めて少ないと考えられる．もしそうだとすれば，その主成分の解釈（ただ一とおりとは限らない）は知識のひらめきによって得られる一つの解釈であり仮説であって，その解釈の正しさが経験によって確証された訳ではないのである(児島，1971も参照)．その解釈は次に続くべき確証的な研究を容易にし，その研究にヒントを与えるにすぎず，「最後のことば」にはなりえないと考えられる．そこで次に，主成分分析を利用したわが国の農業地域区分に関する過去の研究を参照しつつ，以上の点を具体的に検討してみよう．

山中(1980, 13-69頁)は農業集落を単位とした地域区分に焦点を当て，従来から最もよく利用されている「指標組合せ法」と主成分分析とを比較検討する中で，主成分分析には性格の異なる二通りの利用のしかたがあることを指摘している．第一の利用のしかたは，とり上げたすべての変数を一度に利用して主成分分析を行い，分散の大きい方から少数個の主成分を求め，その少数個の主成分を何らかの形で利用して地域区分を行うという利用のしかたである．このような例としては，たとえば伊香(1979)，児島(1971, 1973, 1974)，窪谷（1977，ただし一部の変数を対数変換した場合も同じ変数とみなすことにする)，松原（1974)，能美(1988)，鈴木(1973)，和田(1975 d, 1976 a, b, 1980, 120-163頁)，山田（1975）などの研究がある．

第二の利用のしかたは，とり上げた変数を同種ないし同質の変数からなるいくつかの変数群に分類し，それぞれの変数群ごとに別々に主成分分析を行い，それぞれの変数群ごとに分散の大きい方から少数個（多くの場合1個）の主成分を求め，それらのすべての変数群の主成分を何らかの形で利用して地域区分を行うという利用のしかたである．このような例としては，たとえば窪谷（1971, 1975)，和田（1980, 165-185頁）の研究がある．ただし窪谷の研究では変数群が真に同種ないし同質の変数からなっているかどうか疑わしく，和田の研究はもともと第一の利用のしかたで作成した二種類の区分を重ね合わせたものである．従って窪谷と和田の研究は本来の第二の利用のしかたの適切な事例とは言えない．

さらに山中(1980, 13-69頁)は，第二の利用のしかたは経済指数作成のための利用（Feeney and Hester, 1967；Jolliffe, 1986, pp. 61-63）と同じ性格のものであり，同種ないし同質の変数を一種の指数として要約するための利用であり，求められる主成分の意味は明らかでその解釈に関する問題はほとんど生じないと述べている．また，第一の利用のしかたは複雑な地域農業構造を把握するためのものであり，最も一般的なものであるが，求められる主成分の意味を事前に予想することはできず，分析の結果を待って解釈をつけ加えざるを得ず，その解釈に関して大きな問題が生じると述べている．そしてこの二つの利用のしかたを明確に区別し使い分ける必要があり，そうすることによって主成分の解釈の問題は解決されると述べている（この最後の点の説明は極めて簡単であり，真意はよく分からない)．

以上のような山中の指摘は大筋として適切なものであり，主成分分析の利用のしかたとして探索的な利用のしかた（第一の利用のしかた）と確証的な利用のしかた（第二の利用のしかた）があること，そして探索的な利用のしかたが最も一般的であるが，それは主成分の解釈上大きな問題を伴うことを示唆している．けれども探索的な利用はもともと知識のひらめきを誘うためのものであり，いろんな解釈ないし仮説を引きだすためのものではないだろうか．そのような解釈ないし仮説を無理やり最終的な結論とみなそうと努力することこそ見当違いと言うべきであろう．それは続くべき確証的な研究を容易にし，その研究にヒントを与えればよいのであり，それ以上のことを期待すべきではない．この点については次節でより詳しく述べる．続くべき確証的な研究においては確証的統計的操作，たとえば重回帰分析の利用や主成分分析の確証的な利用等が互いに補完的になされるべきであろう．和田(1975 d, 1454頁，1980, 130頁)は「つまり，分級図を作るためにはどうしても一定の主観的判断は必要なのであり，その意味では主成分分析による処理は一つの参考資料と考え，従来の経験的方法による分級図をチェックするものとして，それと平行してなされる(行なう)のが良いと思われる．」と述べているが，おそらく主成分分析の探索的利用を念頭において上述のような観点から述べたものと推察される．

IV　農業地域区分における主成分分析の探索的利用の問題点

能美（1988）が事例分析によって示したように，主成分に関する解釈ないし仮説を安易に確証されたものと結論づけることは大変危険である．この点に関してもう少し一般的な形で考察しておこう．

表1の記号を利用するものとし，変数 Y が空間的要素の役割を測るための変数であり，変数 X_1，X_2，……X_pは Y と相関関係があるものとしよう．変数 Y を従属変数とし X_1，X_2，……，X_pを独立変数とする重回帰分析をすると，高い決定係数（重相関係数）が得られ，変数 X_1，X_2，……，X_pは空間的要素の役割を認識するのに好都合な変数であるものとしよう．このような場合にも，変数 X_1，X_2，……，X_pの主成分分析によって得られる（分散の大きい方から少数個の）主成分と Y との相関がほとんどなく，それらの主成分は空間的要素の役割を認識するのに好都合な変数とはなりえないことがありうる．このことは基準化した変数を用いると次のように示される．

今 $q=(y_1,\ y_2,\ \cdots\cdots,\ y_N)'$なる記号を導入し，$A$ の階数は P に等しいものとして，重回帰分析により q を変数 x_1，x_2，……，x_pで説明される部分 Ab と残差 ε に分割し

$$q=Ab+\varepsilon,\ \text{ただし}\ b=(b_1,\ b_2,\ \cdots\cdots,\ b_p)'=(A'A)^{-1}A'q\ \text{は偏回帰係数の列ベクトル},$$
$$\varepsilon=(\varepsilon_1,\ \varepsilon_2,\ \cdots\cdots,\ \varepsilon_N)'\ \text{は残差の列ベクトル}$$

と表わそう．また相関行列 R の固有値 λ_jとその固有列ベクトル h_jは

$$Rh_j=\lambda_j h_j\ ;\ j=1,\ 2,\ \cdots\cdots,\ P$$
$$\text{ただし}\ \lambda_1 \geqq \lambda_2 \geqq \cdots\cdots \geqq \lambda_p > 0$$

なる関係をみたし，h_1，h_2，……，h_pは正規直交系をなすから $P\times P$ 行列 H と D

$$H=[h_1, h_2, \cdots\cdots, h_P],\ D=\begin{bmatrix} \lambda_1 & 0, & \cdots\cdots, & 0 \\ 0 & \lambda_2, & \cdots\cdots, & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0, & \cdots\cdots, & \lambda_P \end{bmatrix}\ \text{ただし}\ D\ \text{は対角行列で正則}$$

を導入すると，$H'H=HH'=I_p$（P 次単位行列）が成立し H は直交行列となる．そして

$$RH=HD,\ \text{よって}\ R=HDH'\ \text{および}\ D=H'RH$$

なる関係が成立する．ここで P 個の主成分 C_1，C_2，……，C_pは N 次列ベクトルで

$$C_1=Ah_1,\ C_2=Ah_2,\ \cdots\cdots,\ C_p=Ah_p$$

と表わされるから

$$[C_1C_2\cdots\cdots C_P]=AH$$
$$\frac{1}{N}(AH)'AH=H'\left(\frac{1}{N}A'A\right)H=H'RH=D$$

なる関係が成立し，P 個の主成分は互いに直交し C_jの分散 $\frac{1}{N}C_j'C_j$は λ_jに等しいことが分る．以上のような記号を利用して上述の重回帰式を変形すると

$$q=Ab+\varepsilon=A(HH')b+\varepsilon=AH(H'b)+\varepsilon$$
$$=[C_1C_2\cdots\cdots C_P]\,\beta+\varepsilon$$
$$\text{ただし}\ \beta=(\beta_1、\beta_2、\cdots\cdots,\ \beta_P)'=H'b$$

と表わされ，q は基準化されており C_1，C_2，……，C_p，ε は互いに直交するから

$$q'q/N=(\beta_1^2C_1'C_1+\beta_2^2C_2'C_2+\cdots\cdots+\beta_p^2C_p'C_p+\varepsilon'\varepsilon)/N$$
$$=\lambda_1\beta_1^2+\lambda_2\beta_2^2+\cdots\cdots+\lambda_p\beta_p^2+\sigma^2$$
$$=1\qquad\qquad \text{ただし}\ \sigma^2=\varepsilon'\varepsilon/N$$

が成立し，重回帰分析の決定係数は

$$\text{決定係数}=b'(A'q)/q'q\qquad \text{ここで}\ b=H\beta,\ \ q'q=N$$
$$=\beta'H'(A'A)H\beta/N\qquad (A'A)H\beta=A'q$$
$$=\beta'H'RH\beta\qquad R=(1/N)A'A$$
$$=\beta'D\beta$$
$$=\lambda_1\beta_1{}^2+\lambda_2\beta_2{}^2+\cdots\cdots+\lambda_P\beta_P{}^2$$

と表わされる．さらに主成分 C_jと変数 Y（y でも同じ）との相関係数は

$$q'C_j/N\sqrt{\lambda_j}=\beta_j\lambda_j/\sqrt{\lambda_j}=\sqrt{\lambda_j}\,\beta_j=\alpha_j,\ j=1,2,\cdots\cdots,p$$

と表わされることが分かる．

かくて高い決定係数がえられることは，σ^2が小さいことを意味する．そのような場合にも，たとえば第三主成分まで考慮するものとして，Y とそれらの主成分との相関係数は，

$$\alpha_1=\sqrt{\lambda_1}\,\beta_1,\ \alpha_2=\sqrt{\lambda_2}\,\beta_2,\ \alpha_3=\sqrt{\lambda_3}\,\beta_3$$

に等しいから，たとえ λ_1，λ_2，λ_3がある程度大きくても β_1，β_2，β_3が小さければ，それらの相関係数は小さなものとなってしまう．従って問題となるのは β_1，β_2，β_3が小さな場合であり，これは

$$q=\beta_1C_1+\beta_2C_2+\beta_3C_3+\cdots\cdots+\beta_PC_P+\varepsilon$$

なる関係から明らかなように，変数 Y (y でも同じ) が C_1, C_2, C_3の成分をほとんど含まず，それ以外の分散の小さな主成分と小さな残差 ε によって構成されている場合である．このような場合には最初の三つの主成分は空間的要素の役割を認識するのに好都合な変数（合成変数）とは言えない．

以上述べた点を能美（1988）の事例を用いて具体的に分析してみよう．能美の事例では，$N=28$, $P=10$ で対称な相関行列 R は表2に示すとおりである．この相関行列の固有値と固有列ベクトルを，倍精度の Power Method (ベキ乗法) で，長さが1になるよう基準化したベクトルの，各要素の前回との差の絶対値がすべて 10^{-13}より小さくなるまで計算をくり返して求めると，表3のような結果が得られる．なお能美は単精度でかつ収束精度 10^{-5}で計算しているので(私信による)，若干誤差が出ており，以後の計算結果にも影響すると思われたので，再計算をしその結果を利用することとした．再計算にあたって，固有値問題とベキ乗法に関して古屋(1959)，Jolliffe(1986, pp. 235-246)，森ら(1982, 81-118頁)，Morrison(1967, pp. 221-258)，日本数学会編（1968, 417-423頁），清水（1965），Wilkinson (1965, pp. 570-647)，Wilkinson and Reinsch(1971, pp. 191-201)を参照し，独自のプログラムで計算した．固有値問題はマルコフ過程を利用した農業経済分析（神谷, 1962；久保, 1974；上路, 1979；Fiedler, 1986, pp. 87-111）とも密接に関連することが知られている．

表2　変数間の相関係数

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y
X_1	1.00000	-0.37899	0.02568	0.25472	0.15768	-0.23794	-0.15519	-0.28275	0.62137	-0.43952	0.71730
X_2	-0.37899	1.00000	0.49085	0.20197	0.32557	0.37475	0.41084	0.41986	-0.73565	0.52157	0.01402
X_3	0.02568	0.49085	1.00000	0.12443	0.00768	0.11020	0.15755	0.04166	-0.21105	0.15134	0.17956
X_4	0.25472	0.20197	0.12443	1.00000	0.37717	0.35714	0.44558	0.25202	0.00071	-0.14937	0.56351
X_5	0.15768	0.32557	0.00768	0.37717	1.00000	0.56185	-0.02601	0.51228	-0.24730	0.00460	0.29009
X_6	-0.23794	0.37475	0.11020	0.35714	0.56185	1.00000	0.22650	0.56481	-0.36582	-0.06669	0.09513
X_7	-0.15519	0.41084	0.15755	0.44558	-0.02601	0.22650	1.00000	0.14950	-0.24420	0.20082	0.22554
X_8	-0.28275	0.41986	0.04166	0.25202	0.51228	0.56481	0.14950	1.00000	-0.47572	0.11524	-0.10045
X_9	0.62137	-0.73565	-0.21105	0.00071	-0.24730	-0.36582	-0.24420	-0.47572	1.00000	-0.81674	0.09351
X_{10}	-0.43952	0.52157	0.15134	-0.14937	0.00460	-0.06669	0.20082	0.11524	-0.81674	1.00000	0.03187

(注) 能美(1988)の第1表より引用．最後の Y の列を除いたものが本稿の相関行列 R であり，Y の列が列ベクトル r_Yである．

表3　相関行列 R の固有値と固有列ベクトル（再計算）

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
3.6017369	2.0797800	1.3417605	0.9851752	0.7522828	0.4197657	0.3444757	0.2588871	0.1693485	0.0467876
h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}
-.2800727	0.3768746	0.2504856	0.3296476	0.4146324	0.0957566	0.2906556	0.3670518	-.4420440	-.1114195
0.4527133	-.0658700	0.2094253	0.1907456	-.0109300	-.0263192	0.2256374	-.6080021	-.5026341	0.1908844
0.1804287	-.0176743	0.5172550	0.5974813	-.4277731	0.0812305	-.1029481	0.1919516	0.3238688	-.0259533
0.1538943	0.4844785	0.3400250	-.2361957	0.2301842	0.0692767	-.7017240	-.1366431	-.0195759	0.0306035
0.2448219	0.4314602	-.2816664	0.2966851	0.3605029	-.2374173	0.2385252	-.2330396	0.5357283	-.0448382
0.3270617	0.3504116	-.2251020	-.0761328	-.3834623	-.5220176	-.0124765	0.4206029	-.2756658	0.2029172
0.2393068	0.0740146	0.5046545	-.5846064	0.0072658	-.0059213	0.5111153	0.1285170	0.2331759	-.0927304
0.3527825	0.2266553	-.3344292	-.328877	-.1274890	0.8029655	0.0934056	0.1668294	-.0190970	0.1154078
-.4634582	0.2578743	0.1030948	-.0410506	-.1775697	0.0546562	0.1505634	-.1727379	0.1472884	0.7730936
0.3113649	-.4316234	0.0554429	0.0618597	0.5168588	-.0457698	-.1091302	0.3669991	0.0851808	0.5355009

(注) 能美（1988）は最初の三つの固有値と固有ベクトルの計算結果を示しているが，第一の固有ベクトルとして上表の符号を変えたものを示している．この表の4行目以下が本稿の直交行列 H である．

上述の説明から分るように，β の値は能美(1988)の示したデータを用いて次のように求められる．つまり最も直接的な求めかたは

$$b=(A'A)^{-1}A'q=(NR)^{-1}A'q=R^{-1}\left(\frac{1}{N}A'q\right)=R^{-1}r_Y$$

ただし$\frac{1}{N}A'q=r_Y=(r_{1Y},r_{2Y},\cdots\cdots,r_{10Y})'$であり，$r_{jY}$ $(j=1,2,\cdots\cdots,10)$はX_jとYとの相関係数

なる式よりbを求め，表3の行列Hを利用して$\beta=$H'bと求めることであろう．表2の相関係数r_{jY}と倍精度で計算した表4の逆行列R^{-1}（計算法については表4の注を参照のこと）を利用してbおよびβを求めると表5のようになる．従って主成分C_jとYとの相関係数 $\alpha_j=\sqrt{\lambda_j}\ \beta_j$は，表3の$\lambda_j$を利用して求めると，表5のような値となる．その結果から明らかなように，Yと第1主成分C_1との相関係数は0.00623でほとんど0に等しく，第2主成分C_2との相関係数は0.4868（決定係数0.2370)，第3主成分C_3との相関係数は0.4513（決定係数0.2037）となりかなり小さな値である．これに対して，Yとの相関係数が最も大きいのは第5主成分C_5であり，その値は0.4987(決定係数0.2487)となり，第8主成分C_8との相関係数も0.3782(決定係数0.1430）と比較的に大きい．このように第1主成分は空間的要素の役割を認識するのに好都合な変数（合成変数）とはいえず，第2主成分と第3主成分も，第1主成分よりもよいがそれほど好都合とはいえそうにない．

なお以上の計算結果の中で能美の示した値と異なるものがある．たとえばYと第2および第3主成分との相関係数は，能美によると0.296および−0.135となり，本稿の値とかなり異なる（能美，1988，158頁)．ただし能美によるこれらの値は能美（1988）の第3表および第1表とも矛盾する．というのはYと第j主成分との相関係数はより直接的に

$$\frac{1}{N}\left(\frac{1}{\sqrt{\lambda_j}}Ah_j\right)'q=\frac{1}{\sqrt{\lambda_j}}h_j'\left(\frac{1}{N}A'q\right)=\frac{1}{\sqrt{\lambda_j}}h_j'r_Y$$

と表されるので，能美の第1表よりr_Yを第3表よりh_jとλ_j ($j=1$，2，3でh_1はその符号を変えて利用)を読み取ってこの式で計算をすると

Yと第1主成分との相関係数＝0.0062338
Yと第2主成分との相関係数＝0.4867892
Yと第3主成分との相関係数＝0.4513374

表4　相関行列Rの逆行列R^{-1}

2.6823476	0.2295457	-0.5362822	-0.6046604	-1.1703857	0.4622754	0.1217035	0.1434677	-2.3404917	-0.8662699
0.2295457	3.9767332	-1.3646006	0.0478784	-1.0292255	-0.6685229	-1.0438554	0.1270065	3.1620918	1.0673314
-0.5362822	-1.3646006	1.6367789	-0.0720713	0.6343566	-0.3266814	0.2859934	0.0899983	-0.2192433	-0.0540212
-0.6046604	0.0478784	-0.0720713	1.8679636	-0.4145667	-0.1786643	-0.8998631	-0.1161172	0.3040368	0.4315955
-1.1703857	-1.0292255	0.6343566	-0.4145667	2.6743271	-1.2230273	0.8211962	-0.6404548	-0.1432115	-0.4374608
0.4622754	0.6685229	-0.3266814	-0.1786643	-1.2230273	2.9895931	-0.5930762	-0.0054408	2.8368306	2.5189342
0.1217035	-1.0438554	0.2859934	-0.8998631	0.8211962	-0.5930762	1.8824143	-0.1209386	-1.1527472	-0.9286740
0.1434677	0.1270065	0.0899983	-0.1161172	-0.6404548	-0.0054408	-0.1209386	2.1209544	1.9128672	1.3106159
-2.3404917	3.1620918	-0.2192433	0.3040368	-0.1432115	2.8368306	-1.1527472	1.9128672	13.2336338	8.4099901
-0.8662699	1.0673314	-0.0540212	0.4315955	-0.4374608	2.5189342	-0.9286740	1.3106159	8.4099901	7.2094483

(注) 修正Gram-Schmidt法（Björck, 1967；中川・小柳，1982，65-72頁）による直交化を利用した独自のプログラムで倍精度で計算した．

表5　b，βおよびαの計算結果

b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
0.9612685	-.1510981	0.0356630	0.3202597	-.0919931	0.3235714	0.0980534	-.1505750	-.2208528	0.4148989
β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
0.0032831	0.3375474	0.3896323	0.1641563	0.5750164	-.1785090	-.0515587	0.7432190	-.4536245	-.0323190
α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}
0.0062308	0.4867920	0.4513282	0.1629349	0.4987361	-.1156548	-.0302608	0.3781569	-.1866754	-.0069908

(注) ここで$\sum_{j=1}^{10}\alpha_j^2=\sum_{j=1}^{10}\lambda_j\beta_j^2=0.908179$なる関係が成立し，これは重回帰分析の決定係数に等しい．

という本稿の結果とほとんど同じ結果が得られるからである．私信によると，能美の計算にはかなり計算誤差が生じており，精度が低下しているのではないかと推察される．そこで検算の意味も含めて次のような算式による計算も試みた．つまり

$$\beta=H'b=H'R^{-1}r_Y=H'HD^{-1}H'r_Y=D^{-1}H'r_Y$$

ここで$R^{-1}=HD^{-1}H'$

なる関係式を利用して，相関行列Rの逆行列R^{-1}を求めることなくβを求めてみた．その結果少なくとも有効数字6桁までは上述の結果と全く一致した．このことから本稿の計算結果にはあまり計算誤差は生じていないものと推察される．言うまでもなく能美の指摘の内容は正しいものであり，計算誤差のためにその意義が低下することはないであろう．

V　農業地域区分における主成分分析の確証的利用

主成分分析の確証的利用は，探索的利用と互いに補完的なものであり，探索的な研究過程（この効率化は研究全体の効率化にとって極めて重要であるが通常これは表面に出てこない）に続く確証的な研究過程で重要な役割をはたすものと考えられる．しかし農業地域区分に関する上述のような諸研究を参照する限り，主成分分析の確証的利用に関する本格的な研究はほとんど見られない．そこで本節では主成分分析の確証的利用について若干考察を行ってみよう．

確証的利用はすでに述べたように理論的仮説の検証のための主成分分析の利用である．従って検証されるべき理論的仮説が既に形成されており，そこで利用される諸概念は密接に関連づけられ，それぞれの概念の意味は明確になっていなければならない．そしてこの段階では，諸概念間の量的関係を明らかにし，理論的仮説の概念構成が適切であったかどうか，またその量的関係は理論的仮説と矛盾しないものであるかどうか，というような点について検討を行うことが主な研究課題となろう．このような研究内容から考えると，また互いに相関のある多数の変数のもつ情報を互いに無相関な少数の合成変数に要約し，分析の見通しをよくするという主成分分析の性格から考えると，同種ないし同質の変数を一種の指数として要約するための主成分分析の利用が，その確証的利用の一つの代表的な場合であると考えられる．

同種ないし同質の変数を一種の指数として要約するための方法は従来の指数論（米沢，1972；伊大知，1971）においてもよく研究されており，主成分分析のこのような利用に当たっては，その成果も十分に考慮すべきであろう．指数論の成果も取り入れ，かつそれをより発展させるような形で主成分分析を利用するためには，Rao（1964，pp. 340-343）によって既に示唆されているように，数理計画法を補完的に利用して，主成分分析を必要に応じて変形する必要があるものと考えられる．以下この点について考察を行ってみよう．

今同種ないし同質の変数としてP個の変数$X_1, X_2, \cdots\cdots, X_p$があり，それらの変数の値およびそれらを基準化した変数$x_1, x_2, \cdots\cdots, x_p$の値が表1のように表わされるものとしよう．また変数$X_1, X_2, \cdots\cdots, X_p$を要約する一種の指数を作成し，その指数を表1の変数Yで表すものとしよう．ただしここでは前節までと異なり，変数Y（指数Y）とそれを基準化したyの第i個体における値Y_iとy_iは未知数であり，それらの値は指数として満たすべきいくつかの条件を満たすようこれから求められるものとする．なおこれまでに導入した行列などの記号は形式的にはそのまま利用するものとしよう．従って$q=(y_1, y_2, \cdots\cdots, y_N)'$はここでは$y_1+y_2+\cdots\cdots+y_N=0$および$y_1^2+y_2^2+\cdots\cdots+y_N^2=N$なる条件をみたす未知数の列ベクトルとなる点に注意していただきたい．

最初に指数Yと変数$X_1, X_2, \cdots\cdots, X_p$との相関係数の平方和（決定係数の合計）がなるべく大きくなるよう指数Yを定めるものとすれば，どのような指数Yが得られるか考えてみよう．ただし指数Yは変数$X_1, X_2, \cdots\cdots, X_p$の合成変数である必要はなく，その値は自由に定めてよいものとする．指数Yの値はその基準化した値と適当な測定尺度（原点と測定単位）によって定まるので，まず基準化した値を求め，その後で適当な測定尺度を選ぶことにしよう．言うまでもなく相関係数の値は測定尺度とは無関係である．

上述の相関係数の平方和は容易に分るように次のような式で表わされる．この

$$(\text{平方和})=\left(\frac{1}{N}q'A\right)\left(\frac{1}{N}q'A\right)'=\left(\frac{q}{\sqrt{N}}\right)'\left(\frac{1}{N}AA'\right)\left(\frac{q}{\sqrt{N}}\right)$$
$$=\left(\frac{q}{\sqrt{N}}\right)'\Omega\left(\frac{q}{\sqrt{N}}\right)\quad \text{ただし}\ \Omega=\frac{1}{N}AA'$$

$$\text{ここで}\quad \left(\frac{q}{\sqrt{N}}\right)'\left(\frac{q}{\sqrt{N}}\right)=\frac{1}{N}q'q=1$$

$$\left(\frac{q}{\sqrt{N}}\right)'\theta=0$$

ただし$\theta=(1\ 1\ \cdots\cdots\cdots\ 1)'$はその要素がすべて1の$N$次列ベクトル

式から推察されるように，また以下述べるように，平方和を最大にするqと実対称行列Ωの固有列ベクト

ルとの間には密接な関係がある．なお Ω は $N\times N$ 実対称行列であり $P\times P$ 相関行列 $R=(1/N)A'A$ と異なる点に注意していただきたい．

よく知られているように Ω の固有値と固有列ベクトルは次のように与えられる（Rao, 1964, p. 332 ; Gower, 1966, pp. 331-332)．つまり，簡単のため A の階数は P に等しいものとして，これまでと同様に相関行列 R の固有値と固有列ベクトルを λ_j と h_j $(j=1, 2, \cdots\cdots, P)$ で表わせば

$$\Omega\left(\frac{Ah_j}{\sqrt{N\lambda_j}}\right)=\frac{1}{N}AA'\left(\frac{Ah_j}{\sqrt{N\lambda_j}}\right)=\frac{1}{\sqrt{N\lambda_j}}A\left(\frac{1}{N}A'A\right)h_j$$

$$=\frac{1}{\sqrt{N\lambda_j}}ARh_j=\lambda_j\left(\frac{Ah_j}{\sqrt{N\lambda_j}}\right)$$

$$\left(\frac{Ah_j}{\sqrt{N\lambda_j}}\right)'\left(\frac{Ah_j}{\sqrt{N\lambda_j}}\right)=\frac{1}{N\lambda_j}h_j'A'Ah_j=\frac{1}{\lambda_j}h_j'Rh_j$$

$$=1$$

$$\left(\frac{Ah_j}{\sqrt{N\lambda_j}}\right)'\theta=\frac{1}{\sqrt{N\lambda_j}}h_j'A'\theta=0 \quad j=1, 2, \cdots\cdots, P$$

なる関係が成立し，λ_j と $(1/\sqrt{N\lambda_j})Ah_j=(1/\sqrt{N\lambda_j})C_j$ は Ω の固有値と固有列ベクトルであることが分る．また A の階数は P に等しいので

$$A'L_j=0, \quad L_j'L_j=1, \quad L'_iL_j=0\,(i\neq j)$$
$$i, j=1, 2, \cdots\cdots, (N-P)$$

なる条件をみたす $(N-P)$ 個の N 次列ベクトル L_1, L_2, ……, $L_{(N-p)}$ が存在する．そして

$$\Omega L_j=\frac{1}{N}AA'L_j=\frac{1}{N}A(A'L_j)=0=0\,L_j$$
$$j=1, 2, \cdots\cdots, (N-P)$$

なる関係が成立するので，Ω は固有値 0 を持ち，ベクトル L_1, L_2, ……, $L_{(N-p)}$ は固有値 0 に対応する固有列ベクトルであることが分る．かくて Ω の固有値とそれに対応する固有列ベクトルは

固有値 λ_j と固有列ベクトル $\dfrac{1}{\sqrt{N\lambda_j}}C_j$
$(j=1, 2, \cdots\cdots, P)$

固有値 0 と固有列ベクトル L_j
$(j=1, 2, \cdots\cdots, N-P)$

となり，$P+(N-P)=N$ 個の固有列ベクトルは正規直交系をなす（正規行列 *normal matrix* の相異なる固有値に対応する固有列ベクトルは直交する，古屋, 1959, 107-108 頁)．従って容易に分るように，相関係数の平方和を最大にする q は

$$\frac{1}{\sqrt{N}}q=\frac{1}{\sqrt{N\lambda_1}}Ah_1, \quad \text{つまり}\ q=\frac{1}{\sqrt{\lambda_1}}C_1$$

であり，平方和の最大値は λ_1 である．

以上の分析から明らかなように，相関係数の平方和を最大にすることだけを要求するのであれば，最善の指数 Y は結局のところ第 1 主成分 C_1 を基準化した $(1/\sqrt{\lambda_1})C_1$ と適当な測定尺度によって定まることとなる．ここで注意すべき点は，変数 X_1, X_2, ……, X_p の合成変数（一次結合）として指数 Y を作成する，というような固苦しい条件は何も付けることなく上述の結論が得られることである．このことは第 1 主成分 C_1 の解釈の自由度をいっそう拡大するものと考えられる．なぜならば，第 1 主成分 C_1 を上述のような指数 Y と関連づけて自由に解釈することができるからである．

なお Ω の固有列ベクトルからなる $N\times N$ 直交行列 U

$$U=\left[\frac{C_1}{\sqrt{N\lambda_1}}\ \frac{C_2}{\sqrt{N\lambda_2}}\cdots\cdots\frac{C_P}{\sqrt{N\lambda_P}}\ L_1\ L_2\cdots\cdots L_{(N-P)}\right]$$

および次のような条件をみたす任意の N 次列ベクトル $\delta=(\delta_1, \delta_2, \cdots\cdots, \delta_N)'$

$$\theta'(U\delta)=\delta_{P+1}(\theta'L_1)+\delta_{P+2}(\theta'L_2)+\cdots\cdots$$
$$+\delta_N(\theta'L_{(N-P)})=0$$
$$(U\delta)'(U\delta)=\delta'\delta=\delta_1{}^2+\delta_2{}^2+\cdots\cdots+\delta_N{}^2=N$$

を導入すると，任意の指数 Y を基準化して得られる q は $q=U\delta$ と表わされる．そして A の第 j 列を $A_j(j=1, 2, \cdots\cdots, P)$ で表わせば，任意の指数 Y と変数 X_j との相関係数は，

$$\frac{1}{N}q'A_j=\frac{1}{N}\delta'U'A_j \qquad (j=1, 2, \cdots\cdots, P)$$
$$=\frac{1}{N}\left(\delta_1\frac{C_1'A_j}{\sqrt{N\lambda_1}}+\delta_2\frac{C_2'A_j}{\sqrt{N\lambda_2}}+\cdots\cdots+\delta_P\frac{C_P'A_j}{\sqrt{N\lambda_P}}\right)$$

と表わされる．従ってもし δ_{p+1}, δ_{p+2}, ……, δ_N の中に 0 でないものがあるならば，それらをすべて 0 とし，かつ δ_j $(j=1, 2, \cdots\cdots, P)$ の絶対値を $\delta_1{}^2+\delta_2{}^2+\cdots\cdots+\delta_p{}^2=N$ となるようすべて比例的に大きくすることによって，指数 Y と X_j $(j=1, 2, \cdots\cdots, P)$ との相関係数の絶対値をすべて修正前より大きいかまたは少なくとも等しくすることができる．従って指数 Y と変数 X_1, X_2, ……, X_p との相関係数の絶対値が大きいほどよいのであれば，

$$q=\left[\frac{C_1}{\sqrt{N\lambda_1}}\ \frac{C_2}{\sqrt{N\lambda_2}}\cdots\cdots\frac{C_P}{\sqrt{N\lambda_P}}\right][\delta_1\ \delta_2\cdots\cdots\delta_P]'$$
$$=A\left[\frac{h_1}{\sqrt{N\lambda_1}}\ \frac{h_2}{\sqrt{N\lambda_2}}\cdots\cdots\frac{h_P}{\sqrt{N\lambda_P}}\right][\delta_1\ \delta_2\cdots\cdots\delta_P]'$$

ただし，$\delta_1{}^2+\delta_2{}^2+\cdots\cdots+\delta_P{}^2=N$

というふうに，q が A の P 個の列の合成変数(一次結合）として表わされる場合だけを考えればよいことになる．そこで以下そのような場合だけを考えることにする．

上述のように指数 Y と変数 X_1, X_2, ……, X_pとの相関係数の平方和を最大にするには $q=(1/\sqrt{\lambda_1})C_1=(1/\sqrt{\lambda_1})Ah_1$とすればよい．しかし固有列ベクトル h_1の要素の中には多くの場合負の値も含まれており，そのような場合 q は A の列の非負の一次結合（係数が非負)とはならない．また q と $X_j(j=1, 2,\cdots\cdots, P)$ との相関係数は，因子負荷量行ベクトル

$$\frac{1}{N}\left(\frac{1}{\sqrt{\lambda_1}}Ah_1\right)'A=\frac{1}{\sqrt{\lambda_1}}h_1'R=\sqrt{\lambda_1}\,h_1'$$

の第 j 要素に等しいから，そのような場合 q と負の相関をもつ変数 X_jも存在することになる．従って変数 X_1, X_2, ……, X_pの非負の一次結合として指数 Y を構成し，その指数 Y と変数 $X_j(j=1, 2,\cdots\cdots, P)$ との相関係数がなるべく一様に正の大きな値となるようにしたい時には第１主成分を基準化した q は必ずしも最善とは言えない．そして実際の指数作成においてはこのような問題によく直面するものと考えられる．

このような場合に望ましい指数 Y を求めるための一つの方法は次のような方法であろう．つまり q を A の列の非負の一次結合として

$$q=Au,\ \ u=(u_1,\ u_2,\ \cdots\cdots,\ u_p)'\geqq 0$$

と表せば，q が基準化されていることより P 次列ベクトル u は

$$q'q=(Au)'Au=u'A'Au=N$$
よって $u'Ru=1$

なる条件をみたさねばならない．そして指数 Y と X_j $(j=1, 2,\cdots\cdots, P)$ との相関係数を第 j 要素とする P 次列ベクトルが

$$\frac{1}{N}A'q=\frac{1}{N}A'Au=Ru$$

と表わされることより，これらの相関係数がなるべく一様に正の大きな値となるようにするには，Ru の最小の要素をなるべく大きくすればよいであろう．つまり $\theta_p=(1, 1,\cdots\cdots, 1)'$をその要素がすべて１の P 次列ベクトルとして，また η^2をある小さな正数として

$u\geqq 0$, $u'Ru=1$, $Ru\geqq l\theta_p$, $l\geqq\eta^2$
なる条件の下で l を最大にする u を求めよ

なる問題を解けばよいことになる．

この問題は一種の非線形計画問題であり，次のように変形して解くと便利であろう．つまり

$$\frac{1}{l}u=v=(v_1,\ v_2,\ \cdots\cdots,\ v_p)'$$
よって $u=lv$

とおけば問題は

$v\geqq 0$, $v'Rv=(1/l^2)$, $Rv\geqq\theta_p$
$l\geqq\eta^2$なる条件の下で l を最大にする v を求めよ

と変形される．そこで

$v\geqq 0$, $Rv\geqq\theta_p$なる条件の下で $v'Rv$ を最小にする v を求めよ

なる二次計画問題を解くことができ，$v'Rv$ の最小値が正となるならば最初の問題は解けたことになる．容易に分るように A の階数が P で R が正則（正値定符号)ならば，$v'Rv=0$ となるのは $v=0$ の時だけであるが，$v=0$ は $Rv\geqq\theta_p$なる条件をみたさない．従って少なくとも一つの実現可能解が存在し R が正則ならばこの問題は必ず解けることが分る．

逆にこの二次計画問題の最適解を $v=v^*$とし，$v^{*\prime}Rv^*>0$ が成立するものとすれば，ある正の値 l^*に対して

$$v^*\geqq 0,\ \ Rv^*\geqq\theta_p,\ \ v^{*\prime}Rv^*=\{1/(l^*)^2\}$$

が成立する．よって $u^*=l^*v^*$とすれば

$$u^*\geqq 0,\ \ Ru^*\geqq l^*\theta_p,\ \ u^{*\prime}Ru^*=1$$

が成立する．よってこれまでの説明から明らかなように

$q=Au^*$とすれば
$$q'q=u^{*\prime}A'Au^*=N(u^{*\prime}Ru^*)=N$$
$$(1/N)A'q=(1/N)A'Au^*=Ru^*\leqq\theta_p$$

が成立し，また $\theta_p\geqq Ru^*\geqq l^*\theta_p$が成立する．このことから明らかなように $0<l^*\leqq 1$ となることが分る．以上のように実際の指数作成においては，数理計画法を補完的に利用して，主成分分析を変形する必要がある場合が多いものと考えられる．

二次計画法を補完的に利用して主成分分析を変形する上述の方法の適用例を，能美（1988）によって取り上げられた事例を利用して明らかにしておこう．この方法は，本節の初めに述べたように，同種ないし同質の変数を一種の指数として要約するためのものである．もちろん異種ないし異質の変数を要約するための方法としても形式的には適用可能であるが，そのような形式的な利用をしても有意義な結果が得られることは少ないであろう．たとえば能美によって取り上げられた10個の変数 X_1～X_{10}は同種ないし同質とは考えられないが，表2に示されるこの10個の変数の相関行列 R を用いて形式的にこの方法を適用すると表6のような結果が得られる．つまり表6のような u を用いて $q=Au$ なる指数を作ると，q は基準化された値となり，q と X_jとの相関係数は Ru の第 j 要素（表6の Ru 列の上から j 番目の値）となり，すべての相関係数の値は0.135314以上でそれほど異ならない．しかし異種ないし異質の変数が共存するため，それらの相関係数の平方和は0.362186と小さく，得られた指数 q は形式的なものにすぎず，このような指数 q を求める実質的な意義は少ないように思われる．ただし，前節の主成分分析の結果とこの結果を比較することにより，この方法の性格がより明確になるものと考えられる．本稿で利用した二次計画問題の解法については川口・李（1992）を参照していただきたい．

ところで5つの変数，X_4：乳用牛飼養頭数/戸，X_5：肉用牛飼養頭数/戸，X_6：豚飼養頭数/戸，X_7：採卵鶏飼養羽数/戸，X_8：ブロイラー出荷羽数/戸，は畜産部門の規模を示すという点でかなり同質的である．もちろん真に同種ないし同質と考えてよいかどうか検討を要するが，上述のように10個の変数 X_1～X_{10}を考える場合よりもかなり同質的と考えてよいであろう．そこでこの5つの変数を要約し一種の畜産部門の規模指数を作成する場合を考えてみよう．この5つの変数の相関行列を表2より作成すると次の表7のようになる．この表7の相関係数（相関行列 R_5）を用いて上述の方

表6　二次計画法によって変形された主成分分析の結果
（X_1～X_{10}の10変数）

v		u		Ru
v_1	0	u_1	0	0.288430
v_2	4.545457	u_2	0.615064	0.135314
v_3	0.096998	u_3	0.013125	0.135314
v_4	0	u_4	0	0.233793
v_5	0	u_5	0	0.229888
v_6	6.192541	u_6	0.837937	0.135314
v_7	0	u_7	0	0.248248
v_8	4.351228	u_8	0.588782	0.135314
v_9	22.51443	u_9	3.046518	0.135314
v_{10}	16.91453	u_{10}	2.288773	0.135314

$l=0.135314$　　$u_i=lv_i$

$(Ru)'Ru=(0.288430)^2+(0.135314)^2+\cdots\cdots\cdots+(0.135314)^2=0.362186$

表7　変数 X_4～X_8の相関係数

	X_4	X_5	X_6	X_7	X_8
X_4	1.00000	0.37717	0.35714	0.44558	0.25202
X_5	0.37717	1.00000	0.56185	-0.02601	0.51228
X_6	0.35714	0.56185	1.00000	0.22650	0.56481
X_7	0.44558	-0.02601	0.22650	1.00000	0.14950
X_8	0.25202	0.51228	0.56481	0.14950	1.00000

法を適用すると表8のような結果が得られる.

表8から明らかなように，基準化した変数 x_jの値にウェイト u_j ($j=4$, 5, 6, 7, 8)を乗じてそれらを合計して指数 q を作ると，指数 q は基準化された値となり，指数 q と X_j ($j=4$, 5, 6, 7, 8)との相関係数はすべて0.666920に等しくなる.そしてこれらの相関係数の平方和は2.223912となり，相関行列 R_5 の最大の固有値2.43317とほとんど同じ大きさであることが分る.この基準化された指数 q と適当な測定尺度（原点と測定単位）を組合せて利用すれば，一種の畜産部門の規模指数を作成しうるのではなかろうか.もちろんここでは適用例を示すことを主な目的としており，それが適当な規模指数かどうかは検討を要しよう.

二次計画法を補完的に利用するこの方法の性格を明らかにするために，相関行列 R_5の固有値と固有列ベクトルおよび各主成分の因子負荷量を表9に示している.特に第1主成分の因子負荷量と上述の相関係数0.666920との大小関係に注目していただきたい.

上述の方法は変数 X_1, X_2, ……, X_pの非負の一次結合として指数 Y を構成し，指数 Y と変数 X_j ($j=1$, 2, ……, P)との相関係数をなるべく一様に正の大きな値とするためのものであった.しかしそれらの相関係数の重要性に差があると考えられる場合にも，次のような若干の修正をするだけで上述の方法は同様に有効である.つまり $\theta_p=(1, 1, \cdots\cdots 1)'$なる P 次列ベクトルの第 j 要素 ($j=1$, 2, ……, P) を，指数 Y と X_jとの相関係数の重要性に応じて調整し，重要性が低いものほど1よりもより小さな正数とすればよい.なおより詳細な説明と事例分析は別稿に譲ることとする.

表8 二次計画法によって変形された主成分分析の結果（X_4～X_8の5変数）

v		u		R_5u
v_4	0.241213	u_4	0.160870	0.666920
v_5	0.644025	u_5	0.429513	0.666920
v_6	0.134757	u_6	0.089872	0.666920
v_7	0.817306	u_7	0.545078	0.666920
v_8	0.410989	u_8	0.274097	0.666920
$l=0.666920$　$u_i=lv_i$ $(R_5u)'R_5u=5\times(0.666920)^2=2.223912$				

（注）ここでの v と u は5次元の列ベクトルであるが，混乱することはないと思われるので，簡単のため上のような記号を用いることとした.

表9 相関行列 R_5の固有値と固有列ベクトルおよび因子負荷量

固有値	2.43317	1.17964	0.642093	0.422009	0.323094	備考
固有列ベクトル	0.425499	0.444918	-0.612407	0.239758	-0.434133	X_4
	0.487874	-0.372192	-0.418607	-0.033663	0.668646	X_5
	0.528962	-0.144588	0.233495	-0.719511	-0.356480	X_6
	0.257578	0.757945	0.372796	-0.071725	0.463736	X_7
	0.484544	-0.261026	0.506190	0.646949	-0.149369	X_8
因子負荷量	0.663719	0.483231	-0.490726	0.155752	-0.246767	X_4
	0.761015	-0.404241	-0.335432	-0.021868	0.380067	X_5
	0.825107	-0.157038	0.187101	-0.467410	-0.202628	X_6
	0.401787	0.823212	0.298724	-0.046594	0.263594	X_7
	0.755822	-0.283503	0.405614	0.420272	-0.084903	X_8

VI 要　約

近年その重要性が指摘されている地域計画や農村計画の基礎として，一般に何らかの地域区分が不可欠である．しかしそのような地域区分は，実際には均質地域（同質地域）の観点から出発せざるをえず，従ってここでは均質地域としての地域区分が基本的に重要である（1節）．そこで均質地域としての農業地域区分に焦点を当て，その一つの分析手法として主成分分析を利用する場合の問題点について考察を行い，次のような点を明らかにした．

統計資料を作成しそれらを理論的仮説の形成と検証の際に利用するための手法である統計的操作は，仮説検証のための（確証的）統計的操作と仮説形成のための（探索的）統計的操作に区別され，両者の性格は大きく異なる．そして多様な形で利用される主成分分析の性格を明らかにする上で，両者の性格の違いに注目することは極めて重要である（II節）．

主成分分析の利用のしかたは多様であり，農業地域区分への利用に当たっては，利用のしかたを明確にするために，いくつかの重要な選択を行う必要があることを具体的に示した．そして確証的な観点から主成分分析を利用する場合と探索的な観点から利用する場合とは明確に区別されねばならないこと，探索的な利用の主な目的は最終的な何らかの結論を得ることではなくそのためのヒントを得る点にあること，また農業地域区分への利用は純粋に記述的かつ探索的なものがほとんどであり，確証的観点からの利用は例外的で極めて少ないこと，を明らかにした（III節）．

農業地域区分の手法として主成分分析を探索的な観点から利用する場合の基本的な問題点は，最初に選択した変数がたとえ全体として区分の目的にあった適切なものであったとしても，それらの変数から得られる分散の大きい少数個の主成分は必ずしも区分の目的にあった適切なものではないこと，従ってまたそれらの主成分に関する解釈は一つのヒントにすぎず経験によって確証されたものとは考え難いことである．このような点を理論的に，また事例分析によって明らかにした（IV節）．

主成分分析の確証的利用の一つの代表的な場合は，農業地域区分における同種ないし同質の変数を一種の指数として要約するための利用である．この場合従来の経済指数論などの成果も取り入れ，それをより発展させるような形で主成分分析を利用するためには，数理計画法を補完的に利用して主成分分析を変形せざるをえないことも多い．その具体例として，同種ないし同質の変数の非負の一次結合としてある望ましい性質をもつ指数を作成するための方法が，二次計画法の補完的利用により主成分分析からどのように形成されるかを理論的に，また事例分析によって明らかにした（V節）．

文　献

相川哲夫　1985　農村空間整備論．農林統計協会，東京

Anderson, T. W.　1963 Asymptotic Theory for Principal Component Analysis. *The Annals of Mathematical Statistics*, **34** : 122-148

Björck, Å.　1967 Solving Linear Least Squares Problems by Gram-Schmidt Orthogonalization. *BIT*, **7** : 1-21

Feeney, G. J. and D. D. Hester 1967 Stock Market Indices : A Principal Components Analysis. *In* "Risk Aversion and Portfolio Choice." ed. by D. D. Hester and J. Tobin, John Wiley & Sons, Inc., New York, pp. 110-138

Fiedler, M. 1986 *Special Matrices and Their Applications in Numerical Mathematics*. transl. by K. Segeth and P. Přikryl, Martinus Nijhoff Publishers, Dordrecht (The Netherlands)

古屋　茂　1959　行列と行列式(増補版)．培風館，東京

Gower, J. C. 1966 Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika*, **53** (3·4) : 325-338

林　知己夫　1974　数量化の方法．東洋経済新報社，東京

林　知己夫・樋口伊佐夫・駒沢　勉　1970　情報処理と統計数理．産業図書，東京

伊香厚雄　1979　主成分分析を利用した農業地域区分の方法．武藤和夫・森島　賢編著：地域農業計画の方法と実際．明文書房，東京，114-127 頁

伊大知良太郎編　1971　経済統計講義．青林書院新社，東京

Isard, W. 1975 *Introduction to Regional Science*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 青木外志夫・西岡久雄監訳　1980　地域科学入門（I）—1980　地域科学入門（II）—1985　地域科学入門（III）．大明堂，東京

Jeffers, J. N. R. 1967 Two Case Studies in the Application of Principal Component Analysis. *Applied Statistics*, **16** : 225-236

Jolliffe, I. T. 1986 *Principal Component Analysis*. Springer-Verlag, New York

神谷慶治　1962　職業移動マトリックス—マルコフモデルについて—．神谷慶治・沢村東平監修：新しい農業分析．東京大学出版会，東京，176-197 頁

金沢夏樹　1973　経済的土地分級の意義と課題．金沢夏樹編著：経済的土地分級の研究一農業への適用一．東京大学出版会，東京，3-31頁

川口雅正　1981　統計学の学問的性質について．九大農学芸誌，**36**（1）：25-46

川口雅正・李　鍾相　1992　空間均衡分析における均衡解の一意性について一Ⅰ．理論的考察．九大農学芸誌，**46**（3・4）：177-198

Kendall, M., A. Stuart and J. K. Ord 1983 *The Advanced Theory of Statistics, Vol. 3, Fourth Edition*. Charles Griffin & Company Limited, High Wycombe (England)

喜多克己訳　1982　アメリカ農業経済学会：アメリカ合衆国における農業経済データ体系（*A Survey of Agricultural Economics Literature*, Vol. 2, Part IV, 1977 の抄訳で統計研究参考資料 No. 11）．法政大学日本統計研究所，東京

児島俊弘　1961　農業の経済的地帯分画の考え方と指標についての覚え書一農業における経済的地帯形成の諸問題(一)一．農業総合研究，**15**(1)：45-70

児島俊弘　1971　主成分分析における結果解釈過程の構造一農業部門結合の地域型検出に関連して一．農業総合研究，**25**（2）：189-215

児島俊弘　1973　農業集落の主観的証価と客観的評価（I）一論理のわく組み一．農業総合研究，**27**（4）：1-29

児島俊弘　1974　農業集落の主観的評価と客観的証価（II）一事例の解析一．農業総合研究，**28**（1）：1-42

久保嘉治　1974　マルコフ過程による動態分析．工藤元・西村正一・高山　崇・久保嘉治編著：改稿近代農業経済学．明文書房，東京，433-454頁

窪谷順次　1971　地域計画のための農業集落の判別．農業総合研究，**25**（1）：205-226

窪谷順次　1975　農業開発計画のための地域分類．農業総合研究，**29**（1）：33-75

窪谷順次　1977　首都圏における緑地空間分布の規則性について．農業総合研究，**31** (2)：133-161

窪谷順次　1978　地域計画のあり方・考え方．地域計画研究会編：地域計画一その理論と実験一．農林統計協会，東京，59-103頁

窪谷順次　1988　地域計画論一都市・農村の調整，環境問題と関連させて一．農業総合研究所，東京

松原茂昌　1974　主成分分析．工藤　元・西村正一・高山　崇・久保嘉治編著：改稿近代農業経済学．明文書房，東京，413-426頁

森　正武・名取　亮・鳥居達生　1982　数値計算（岩波講座情報科学一18）．岩波書店，東京

Morrison, D. F. 1967 *Multivariate Statistical Methods*. McGraw-Hill, Inc., New York

武藤和夫　1979　地域農業計画の課題と手順．武藤和夫・森島　賢編著：地域農業計画の方法と実際．明文書房，東京，1-12頁

中川　徹・小柳義夫　1982　最小二乗法による実験データ解析．東京大学出版会，東京

日本数学会編　1968　岩波数学辞典　第2版，岩波書店，東京

能美　誠　1988　期待農業所得分級法に関する考察一重回帰分析による接近一．農業経済研究，**60**（3）：150-159

奥野忠一・久米　均・芳賀敏郎・吉澤　正　1971　多変量解析法．日科技連，東京

Rao, C. R. 1964 The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhya*, Series A, **26** (Part 1) : 329-358

清水留三郎　1965　行列の固有値問題．山内二郎・森口繁一・一松　信編：電子計算機のための数値計算法Ⅰ．培風館，東京，173-195頁

鈴木福松　1973　分級図の作成問題．金沢夏樹編著：経済的土地分級の研究一農業への適用一．東京大学出版会，東京，209-215頁

竹内　啓ら編　1989　統計学辞典．東洋経済新報社，東京

天間　征　1966　地域農業計画の諸問題．農業技術研究所報告 H，(34)：43-57

上路利雄　1979　マルコフ過程．武藤和夫・森島　賢編著：地域農業計画の方法と実際．明文書房，東京，72-82頁

和田照男　1973　経済的土地分級方法の基本問題．金沢夏樹編著：経済的土地分級の研究一農業への適用一．東京大学出版会，東京，107-126頁

和田照男　1975 a　農村土地利用計画の課題と実態一農村土地利用計画の課題と方向（1）一．農業および園芸，**50**（9）：1083-1090

和田照男　1975 b　農村土地利用計画の基本問題一農村土地利用計画の課題と方向（2）一．農業および園芸，**50** (10)：1203-1210

和田照男　1975 c　優良農地の保全と経済的土地分級［1］一農村土地利用計画の課題と方向（3）一．農業および園芸，**50** (11)：1335-1338

和田照男　1975 d　優良農地の保全と経済的土地分級［2］一農村土地利用計画の課題と方向（4）一．農業および園芸，**50** (12)：1453-1456

和田照男　1976 a　農地の経済的土地分級の実際［1］一農村土地利用計画の課題と方向（5）一．農業および園芸，**51**（3）：389-393

和田照男　1976 b　農地の経済的土地分級の実際［2］一農村土地利用計画の課題と方向（6）一．農業および園芸，**51**（4）：503-507

和田照男　1980　現代農業と土地利用計画一土地利用転換と計画手法一．東京大学出版会，東京

Wilkinson, J. H. 1965 *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford

Wilkinson, J. H. and C. Reinsch 1971 *Handbook for Automatic Computation, Volume II, Linear Algebra*. Springer-Verlag, Berlin・Heidelberg

山田三郎　1975　アジア諸国の農業特性と農業地域類型一主成分分析による農業特性の類型化一．農業経済研究，**47**（1）：1-13

山中　守　1980　農業集落区分の計量経済学的研究．九州大学大学院農学研究科学位請求論文
柳井晴夫・高木廣文訳　1981　F.ハートウイグ・B.E.デアリング：探索的データ解析の方法．朝倉書店，東京
米沢　治文　1972　経済統計計量分析．日本評論社，東京

Summary

Some kind of regional land classification is indispensable in general as the basis of regional or rural planning, the importance of which has been stressed recently in Japan. And in practice, the most fundamental kind of regional land classification is the classification of land area into homogeneous regions (Section I). Principal Component Analysis (P. C. A.) is well known as an important analytical tool for such a classification. But it is also well known that researchers must solve the difficult problem of how to make sense of Principal Components (P. C.) when they try to apply P. C. A. to such a classification. In this paper, We study the problem of how to make sense of P. C. in such a regional land classification, and the conclusion of this paper is as follows.

Statistical Operations (S. O.), namely procedures for producing statistical data and for using the statistical data to form and test a theoretical hypothesis, should be classified into the confirmatory S. O. for testing a theoretical hypothesis and the exploratory S. O. for forming a theoretical hypothesis, and it should be noted that the character of the both are fundamentally different. Giving full attention to this difference in character is essential for understanding the character of P. C. A. which is used in various ways (Section II).

We can make use of P. C. A. in various ways, and when we try to apply P. C. A. to agricultural regional land classification, we must make some important choices, which we show concretely in section III, to fix the way to application. And the confirmatory way of application should be distinguished precisely from the exploratory way of application, the main objective of which is not to give some conclusion of the analysis but to give some hint for the following analysis. Most of the applications of P. C. A. to agricultural regional land classification are of purely descriptive and exploratory nature, and that of confirmatory nature is rare (Section III).

The several P. C. which correspond to large eigenvalues and usually used for the classification may not be suitable for the objective of classification if the variables selected for the P. C. A. are suitable for the objective of classification as a whole. This is the fundamental problem of the exploratory way of application of P. C. A. to agricultural regional land classification. Because of this fact, the interpretation of P. C. given by researchers should be thought as a hint or a hypothesis which is not confirmed by experience. We account for the above points theoretically and by a case study (Section IV).

Application of P. C. A. to making some kind of index number which represents several selected variables of the same nature is a typical and main example of the confirmatory way of application of P. C. A. to agricultural regional land classification. To take into consideration the results of the theory of index number and other related useful results in the application, we often need to extend P. C. A. in many directions in combination with mathematical programming. We show theoretically and by a case study an example of how to extend P. C. A. in combination with quadratic programming for making some kind of index number which is a nonnegative linear combination of selected variables of the same nature and has some desirable properties (Section V).