

慶應義塾大学学術情報リポジトリ

Keio Associated Repository of Academic resouces

Title	社会学的類型構成の計量的方法：クラスター分析とその応用
Sub Title	A quantitative approach to "type-construction" in sociology : techniques of "cluster analysis" and their application
Author	井関, 利明(Izeki, Toshiaki) 堀内, 四郎(Horiuchi, Shiro)
Publisher	慶應義塾大学大学院社会学研究科
Publication year	1972
Jtitle	慶應義塾大学大学院社会学研究科紀要：社会学心理学教育学 (Studies in sociology, psychology and education). No.12 (1972.),p.31- 43
Abstract	
Notes	論文
Genre	Departmental Bulletin Paper
URL	http://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN0006957X-00000012-0031

社会学的類型構成の計量的方法

— クラスタ分析とその応用 —

A Quantitative Approach to "Type-Construction" in Sociology

— Techniques of "Cluster Analysis" and their Application —

井 関 利 明

Toshiaki Izeki

堀 内 四 郎

Shiro Horiuchi

1. 類型の方法論的性格と有用性

「類型」は、さまざまな学問領域で使用されてきた伝統的な分析用具の一つである。とくに社会学の分野においては、「類型」は、有用な概念的装置として、データの選択、記述、比較、説明などのために、積極的に活用されてきた。リースマンの「伝統志向型」「内部志向型」「他人志向型」、ウェーバー (M. Weber) の「伝統的支配」「カリスマ的支配」「法制的支配」、テンニース (F. Tönnies) の「ゲマインシャフト」と「ゲゼルシャフト」などは、その好例であろう。

しかし、従来、「類型構成」とその利用は、必ずしも方法論的基礎づけの明確でないままに、多数の社会学者によって、行われてきたように思われる。これに対して、近年においては、科学方法論の研究者達によって、「類型」に関する旧来の誤解・誤用が指摘され、「類型」の方法論的性格を明確化する作業が、積み重ねられてきている。このような一連の研究において成果をおさめた学者として、ベッカー (H. Becker)¹⁾、ラザスフェルト (P. F. Lazarsfeld)²⁾、ヘンペル (C. G. Hempel)³⁾、マキニー (J. C. McKinney)⁴⁾ などの名前をあげることができよう。

「類型」が、その方法論的基礎づけの曖昧なままに、用いられることの多かった理由の一つは、従来、「類型構成」の標準化された手続きを明示することが、きわめて稀だったことに求められるであろう。本論文の狙いは、社会学における「類型構成」の、標準化された手続きの一例を示すことにある。そのための基礎的な作業と

して、本節では、まず「類型」の方法論的性格を考察し、さらに、「類型」の分析用具としての有用性について論ずることにしたい。

ここでは、マキニーの著名な定義を出発点として、「類型」の方法論的性格に考察を加えてみよう。彼によれば、「類型」とは、「一組の規準を、意図的かつ計画的に選択し、抽象し、組み合わせ、かつ(時には)強調したものである。そしてその一組の規準は、経験的対応物を持ち、経験的事例を比較する基礎として役に立つ。」⁵⁾ ここでいう「規準」(criteria)は、「類型」を組立てている要素 (elements) ないし特性 (characteristics) を意味している。マキニーの定義に関連して、さしあたり、つぎの4点を確認しておくことが、必要であろう。

1) 「類型」は、けっして具体的な経験的対象またはその集合を指示するものではない。それは、分析用具として設定された概念的構成体 (conceptual construct) に他ならない。すなわち、「類型」とは、経験的対象を分析するための、より有効な概念的装置の設定を目的として、さまざまな対象の複雑多様な諸特性のなかから、一組の特性を選択し、抽象し、強調し、組み合わせることによって、構成されるものである。したがって、分析的な概念のレベルと、操作的で観察可能なレベルとを分けて考える場合に、それは、基本的には、前者のレベルに位置づけられることになる。

2) しかし、「類型」を有効な分析用具にするためには、なんらかの操作化 (operationalization) の手続きが必要である。すなわち、「類型」を組立てている諸特性の測定手続きを明示することによって、操作的で観察可

能なレベルにおいて、使用できるようにしなければならない。これによって、多数の経験的対象を「類型」と比較し、それぞれが、いかなる側面で、どの程度まで「類型」と異なっているのかを測定することが可能になる。マキニーのいう「経験的対応物 (empirical referents) をもち、経験的事例を比較する基礎として役に立つ一組の規準」とは、以上の意味での「測定手続きの明示された諸特性」を指示するものであって、彼はけっして、「類型」と1対1で対応するような具体的対象を見出すように要求しているわけではない。

3) 「類型」は、一般に複数の諸特性の複合体である。しかし、稀には、「類型」が単一の特性によって構成されることもある。「内向的性格」と「外向的性格」のように、一次元的連続体上の両極を表わす場合が、それである。このような「類型」は、'polar types' と呼ばれる。だが、一次元的連続体が、相互に関連の強い、いくつかの変数の合成されたものとして解釈されるならば、polar types も、やはり複数の諸特性の複合体として、表現されることになる。社会学の領域では、テンニースの「ゲマインシャフト」と「ゲゼルシャフト」や、スペンサー (H. Spencer) の「軍事型社会」と「産業型社会」の例にみられるように、複合的性格の polar types が多い。

4) 「類型」を「分類概念」や「平均」と混同してはならない⁹⁾。まず、「類型」は、「分類概念」とは異なった性格を有している。「分類概念」は、特定の対象のあるクラスに所属させる概念である。任意の対象とあるクラスとの関係は、つぎの2通りしかない。すなわち、対象がそのクラスに属しているか、または属していないかのどちらからである。特定の対象が、なんらかの「類型」に正確に該当すること (すなわち、典型的事例) は、稀であろう。経験的対象は、ある「類型」について、きわめて近似しているとか、あまり近似していないとか、または、ある側面では該当しているが、他の側面では該当していない、などと判定されるわけである。

また、「類型」は、必ずしも平均的な対象を指示するものでもない。むしろ、多くの場合、分析用具として有効な「類型」を構成するためには、ある側面が著しく強調されなければならない。このようにして構成された「類型」は、かえって平均からの偏差のきわめて大きなものとして、表現されることになる。

「類型」の方法論的性格に関しては、以上の4点を確認しておくにとどめたい。つぎに、分析用具としての「類型」の有用性に、焦点を当てることにしよう。この

点に関しては、さまざまな角度から検討することができ、ここでは、以下の4点を指摘しておきたい。

1) 「類型」は、データ選択の方向を示唆するという意味で、索出的価値 (heuristic value) を持つ。研究者は、経験的対象を「類型」と比較することによって、特定化された分析視点に立って、対象を識別することができる。したがって、「類型」は、分析の目的に沿って事実を選択し、その相対的重要度を決定する作業において、きわめて有効な用具になりうるものである。

2) 説明および予測との関連においても、「類型」は役に立つ。「任意の対象 x に関して、類型 t にたいする x の近似性が高いほど、 x が特性 c を示す確率も高い。(ただし特性 c は、類型 t を組立てている諸特性のなかには、含まれていないものとする)」という命題 P が成立している場合には、 t にきわめて近似している特定の対象 a に関して、説明または予測を行うことが可能である。すなわち、 a が特性 c を示すことが判明しているならば、命題 P に基づいて、それを説明することができる。また、 a が特性 c を示すか否かが明らかでない場合には、命題 P に基づけば、それを予測することができるわけである。

上述したような命題 P が成立する場合としては、つぎの2通りが考えられる。i) 「もし x が t ならば、 x は特性 c を示す」という命題 Q が、演繹的に推論される場合には、命題 Q に依拠して、命題 P が説明または予測のために利用される。経済学における「ホモ・エコノミクス」の人間像や、化学における「理想気体」など、理念化 (idealization) によって構成された類型は、概してこのような性格を持っている。ii) 類型 t を組立てている諸特性と、特性 c の間の関連が強いことが、相関係数の計算などによって、経験的に検証されている場合にも、命題 P を利用することができる。パスナリティの類型論に基づいて、特定個人の行動を説明または予測するという事例は、これに該当する。

3) 「類型」の使用は、思考を効率化する。「類型」は諸特性の複合体であるが、複数の特性の組合せの数は、分析のために採用される変数の増加によって、幾何級数的に大きくなるものである。一例として、ダミー変数形式 (0-1 形式) の変数が、6個採用される場合を想定してみよう。組合せの数は、 $2^6=64$ 個にもなる。しかし、64個のなかから、一定の分析視点に立って、とくに重要度の高い数個の組合せを「類型」として選定し、残りの組合せを「類型」からの「偏差」(deviation) と解釈することによって、思考の効率化を図ることができ

るわけである。

4) 「類型」は、分類体系を構築する場合に、役立つことがある。分類は、特定の側面に関して、複数の対象を観察または測定し、共通な特性の発見された対象を、同一のクラスに所属させる作業である。しかし、一つのクラスと他のクラスとの境界を明確に規定することが容易な場合もあれば、困難な場合もある。例をあげれば、性別、年齢、学歴、所得水準などの人口学的要因に関しては、クラス間の境界は明瞭である。特定の人間が「30代」のクラスに所属するかどうかは、彼の誕生日を知っていれば、簡単に判定できる。ところが、パーソナリティ、ライフ・スタイル、あるいはなんらかの行動パターンなどをとりあげて分類を試みる場合には、その作業はより困難なものとなり、各クラス間の境界は、曖昧になりがちである。

このような場合には、以下のような手続きが有効であろう。すなわち、まず数個の「類型」を構成する。つぎに、それぞれの「類型」に対応させて、同数のクラスを設定する。各クラスは、対応する「類型」に近似している対象の集合を、指示するものである。各クラスの境界の明確化は、つぎのようにして可能になる。すなわち、対応する「類型」との近似性が、どの程度以上の対象を所属させるかという基準を、それぞれのクラスごとに、操作的に規定するわけである。

以上に理解されるように、つまり「類型」は有用性の高い分析用具であり、とくに社会学においては、頻繁に活用される重要な概念的装置といえよう。しかし、方法的基礎づけの曖昧なままで、類型が使用される場合が多かったことも、否定できない。前述したように、理由の一つは、従来、「類型構成」の標準化された手続きを明示することが、きわめて稀であったことに求められるであろう。もっとも一般的な「類型構成法」は、「直観的設定」であったように思われる。

これに対して、本論文では、「類型構成」の、一つの標準化された手続きを提示してみたい。この手続きにおいては、分析用具としてより有効な「類型」を構成するために、「クラスター分析」と呼ばれる計量分析の手法が採用される。「クラスター分析」は、重回帰分析、判別分析、主成分分析などの多変量解析法と比較した場合、まだ利用の歴史の浅い手法である。とくに社会学の領域では、これまで注目されることが少なかった。だが、近年、生物学、医学、薬学などの分野で広く活用されてきており、計量地理学やマーケティングにおいても、普及しつつある手法である。社会学者は、とくに

「類型構成」との関連において、この手法に注目すべきであろうというのが、われわれの主張である。

以下においては、第2節で「クラスター分析」と「類型構成」の関連を論じ、第3節で「クラスター分析」の手法の一部を説明し、第4節で「クラスター分析」による「類型構成」の事例を紹介することにしよう。

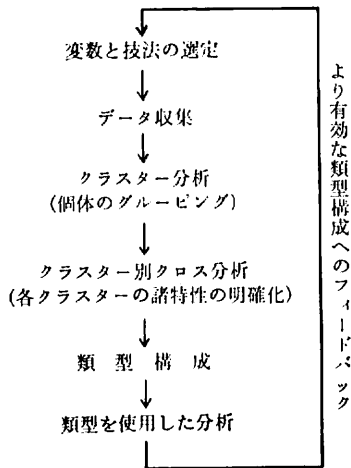
2. 類型構成とクラスター分析

「クラスター分析」の基本的アイディアは、つぎのように要約できる。すなわち、相互に近似性の高い「もの」同志は、できるだけ同一グループに、相互に近似性の低い「もの」同志は、できるだけ異なるグループに所属するように、 n 個の「もの」を、 m 個 ($m < n$) のグループに分割する、ということである。このグループを「クラスター」と呼ぶのである。

以上のようなアイディアに基づいて、多数の手法が考案されている。したがって、「クラスター分析」という名称は、単一の手法ではなく、一群の手法を指示するものである。これらの手法は、大別すると、「O分析」と「V分析」の2種類になる。クラスター化される「もの」が、対象 (objects) または個体 (individuals) の場合には、「O分析」と呼び、変数 (variables) の場合は、「V分析」と呼ぶ。「O分析」と「V分析」の区別は、計量分析のきわめて基本的な性格に基づくものである。なぜならば、計量分析のためのデータは、一般に、 n 個の個体を行にとり、 k 個の変数を列にとったデータ行列 ($n \times k$) として、与えられるからである。すなわち、「O分析」と「V分析」の区別は、データ行列の行と列の区別に対応し、さらには主語と述語の区別に対応するものといえよう。「類型構成」との関連で、直接的に問題となるのは「O分析」であり、以下で「クラスター分析」という語を使用する場合は、もっぱら「O分析」を指すことにする⁷⁾。

「クラスター分析」は、すでに繰り返したように、類型構成のための有効な手法として、利用可能である。図1は、「クラスター分析」による類型構成法の、一連の手順を示している。本節では、まず図1の各ステップを簡単に解説し、つぎに、このような方法の利点について論じてみたい。

第1ステップ……初めに、データの収集に先立って、一定の分析視点から、戦略的に重要な変数が選定され、その測定手続きが明示されなければならない。また、クラスター分析の手法は幾通りも考案されているので、分析の目的やデータの性格 (変数の個数、変数の尺度の種



〈図1〉 クラスタ分析による類型構成法

類、変数間の相関関係、個体数、個体間の同質性の程度など)に応じて、もっとも適合的な技法が選定されるように、配慮しなければならない。

第2ステップ……選定された変数について、データを収集する。

第3ステップ……収集されたデータを、「クラスタ分析」の選定された特定手法を用いて解析する。解析の結果、相互に近似性の高い個体からなる、いくつかのクラスタが形成される。

第4ステップ……各クラスタの諸特性とクラスタ間の差異をみるために、クラスタ別クロス分析が行われる。すなわち、クラスタ区分を1個の変数とみて、他の変数とのクロス集計を実施する。

第5ステップ……クロス分析によって明確化された、各クラスタの平均値的もしくは最頻値的な諸特性とクラスタ間の差異に基づいて、それぞれのクラスタごとに、類型が構成される。

なお、クラスタの諸特性と類型とは、明確に区別されなければならない。両者の差異に関しては、以下の2点が重要である。(1)クラスタの諸特性は、操作的で観察可能なレベルにおいて、統計解析の結果として記述されるが、類型は、分析的なレベルで、概念的構成体として記述される。(2)類型を構成する際には、一定の分析視点に立って、クラスタの諸特性につきのような処理が加えられる。すなわち、i) 重要度の高い特性を強調し、ii) 重要度のきわめて低い特性を捨象し、iii) 相互に関連が強いと解釈される複数の特性を概括する、というような処理である。

第6ステップ……このようにして構成された類型を、

データの選択や、経験的対象の識別、記述、比較、説明、予測などに利用する。実際に用いることによって、構成された類型の、分析用具としての有効性がテストされて、修正すべき点が発見される。テストの結果は、より有効な類型を構成するための、再度のクラスタ分析へ(具体的には、新たな変数と技法の選定へ)とフィード・バックしていくわけである。

つぎに、以上のようなステップを踏んで行われる類型構成法の利点を、論ずることにしたい。ここでは、とくに直観的な類型構成法との比較において、以下の2点を指摘しておこう。

1) クラスタ分析による類型構成法は、標準化された手続きであるという意味において、「科学的」である。すなわち、同一の分析視点に立ち、かつ同一の理論的公準と概念枠組から出発しているかぎり、以上の各ステップの手順を踏むことによって、いかなる研究者でも同一の類型を構成することができるはずである。

2) クラスタ分析による類型構成法は、直観的な方法と比較した場合、他の条件が同じならば、分析用具としての有効性の低い類型を構成してしまう危険性が、相対的に小さい。これは、分析のために採用される変数の個数の増加によって、諸特性の組合せの数が、幾何級数的に大きくなる事情に関連している。この場合、第1節で述べたように、とくに重要度の高い数個の組合せのパターンを、類型として選択し、他の組合せパターンをとる対象を識別、記述、または比較するための規準として使用することが、要請される。

クラスタ分析による類型構成法においては、経験的対象が比較的とりやすい諸特性の組合せ(すなわち、各クラスタの、平均値的ないし最頻値的な特性の組合せ)のパターンのいくつか、重要度の高いものとして選び出される。この組合せパターンに基づいて、類型が構成されるわけである。一方、直観的な方法による類型構成は、どのような組合せパターンがとられやすく、どのような組合せパターンがとられにくいという経験的データに基づくことなしに行われる。したがって、他の条件が同一ならば、直観的な類型構成法においては、つぎのような危険性が、より大きくなる。すなわち、重要度の低い組合せパターンを選択することによって、分析用具としての有効性の低い類型を構成してしまう危険性である。

3. クラスタ分析の手法

クラスタ分析としては、これまで数多くの手法が考

<表1> 距離および類似度の指標

ユークリッド距離	$d_{\alpha\beta}^2 = \sum_{i=1}^p (x_{\alpha i} - x_{\beta i})^2 \quad (1)$ <p>または</p> $d_{\alpha\beta}^2 = \frac{1}{p} \sum_{i=1}^p (x_{\alpha i} - x_{\beta i})^2 \quad (2)$
荷重ユークリッド距離	$d_{\alpha\beta}^2 = \sum_{i=1}^p k_i (x_{\alpha i} - x_{\beta i})^2 \quad (3)$ <p>k_i: 定数</p>
標準ユークリッド距離	$d_{\alpha\beta}^2 = \sum_{i=1}^p \frac{(x_{\alpha i} - x_{\beta i})^2}{S_i^2} \quad (4)$ $S_i^2 = \frac{1}{n-1} \sum_{\alpha=1}^n (x_{\alpha i} - \bar{x}_i)^2$ $\bar{x}_i = \frac{1}{n} \sum_{\alpha=1}^n x_{\alpha i}$
マハラノビスの汎距離	$d_{\alpha\beta}^2 = (x_{\alpha} - x_{\beta})' \Sigma^{-1} (x_{\alpha} - x_{\beta}) \quad (5)$ $x_{\alpha} = (x_{\alpha 1}, x_{\alpha 2}, \dots, x_{\alpha p})'$ $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)'$ $\Sigma = \frac{1}{n} \sum_{\alpha=1}^n (x_{\alpha} - \bar{x})(x_{\alpha} - \bar{x})'$
2進法ユークリッド距離	$d_{\alpha\beta} = \frac{b+c}{p} \quad (0-1 \text{ 型データ}) \quad (6)$ $b = \sum_{i=1}^p x_{\alpha i}(1-x_{\beta i})$ $c = \sum_{i=1}^p (1-x_{\alpha i})x_{\beta i}$
個体間積和	$q_{\alpha\beta} = \sum_{i=1}^p \frac{(x_{\alpha i} - \bar{x}_i)(x_{\beta i} - \bar{x}_i)}{S_i^2} = \sum_{i=1}^p z_{\alpha i} z_{\beta i} \quad (7)$ $z_{\alpha i} = \frac{x_{\alpha i} - \bar{x}_i}{S_i}$
個体間相関係数	$r_{\alpha\beta} = \frac{\sum_{i=1}^p (z_{\alpha i} - \bar{z}_{\alpha})(z_{\beta i} - \bar{z}_{\beta})}{\sqrt{\sum_{i=1}^p (z_{\alpha i} - \bar{z}_{\alpha})^2 \sum_{i=1}^p (z_{\beta i} - \bar{z}_{\beta})^2}} \quad (8)$
一致係数	$S_{\alpha\beta} = \frac{a+d}{p} \quad (0-1 \text{ 型データ}) \quad (9)$ $a = \sum_{i=1}^p x_{\alpha i} x_{\beta i}$ $d = \sum_{i=1}^p (1-x_{\alpha i})(1-x_{\beta i})$
類似比	$P_{\alpha\beta} = \frac{a}{a+b+c} \quad (0-1 \text{ 型データ}) \quad (10)$

案されてきている⁸⁾。本節では、そのなかから、類型構成に利用する場合に、比較的有効と思われる幾通りかの手法を選んで、紹介したい。ここで解説される諸手法は、3個のグループ(すなわち、①閾値を設定する手法、②階層的な手法、③分割の妥当性の指標を使用する手法)に分けることができる。以下、順を追って紹介していこう。

なお、いかなる手法においても、データとしては、 p 個の変数に関する、 n 個の個体の測定値が与えられているものとし、各測定値を $x_{\alpha i} (\alpha=1, 2, \dots, n; i=1, 2, \dots, p)$ と表わす。この $x_{\alpha i}$ は、必ずしも原データまたはその標準化されたものに限定されない。むしろ成分得点、因子得点(R技法)、因子負荷量(Q技法)、林式数量化理論第Ⅲ類または第Ⅳ類の手法によって付与される数値などが使用される場合が多い。

また、大部分の手法においては、個体間の距離または類似度の計算が行われる。表1⁹⁾は、比較的多く利用される距離または類似度の指標を、整理したものである。

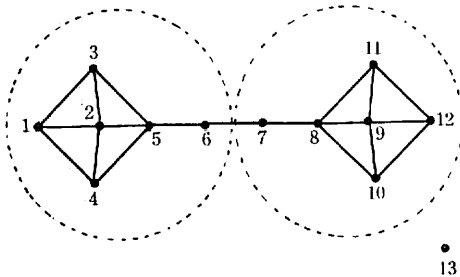
初めに、閾値(threshold value)を設定する手法について、紹介しよう。ここでは、グリーン(P. E. Green)の手法¹⁰⁾、スニース(P. H. Sneath)の「単一連結法」(single linkage method)¹¹⁾、ウィシャート(D. Wishart)の「モード分析」(mode analysis)の、3つの手法をとりあげることにしたい。

1) グリーンの手法は、以下の手順で行われる。①閾値 D が、研究者によって任意に決定される。②すべての個体相互間の距離 $d_{\alpha\beta}$ を計算する。③距離が最小の2個体によって、クラスター C_1 を形成する。④ C_1 の重心、および、残りの個体それぞれの、 C_1 の重心からの距離を計算する。⑤ C_1 に所属していない個体のなかで、 C_1 の重心からの距離が最小のものを、 C_1 に所属させる。⑥以上の④⑤の手順を反復し、 C_1 の個体

数をふやしてゆく。⑦もし、 C_1 の重心と、 C_1 に所属していない個体との距離の最小値が D を越えたならば、反復を停止する。⑧ C_1 に所属していない個体のなかで、距離が最小の2個体によって、クラスター C_2 を形成し、④⑤⑥⑦の手順を反復する。⑨ すべての個体が、いずれかのクラスターに所属するか、あるいは、どのクラスターにも所属していない個体間の距離の最小値が D を越えるまで、同様にして、 C_3, C_4, C_5, \dots を形成していく。

2) スニースの「単一連結法」は、以下の手順で行われる。①閾値 D が、研究者によって任意に決定される。②すべての個体相互間の距離 $d_{\alpha\beta}$ を計算する。③ $d_{\alpha\beta} \leq D$ であるような個体 α と β を、すべて連結する。④直接的もしくは間接的(すなわち、他の個体を介して)に連結されている個体のあつまりのそれぞれを、クラスターとする。

3) ウィシャートの「モード分析」は、単一連結法にたいする批判から生まれたものである。彼の論点を、図を用いて述べてみよう。図2は、13個の個体があり、



<図2> 鎖効果

そのうち12個が、単一連結法によって1個のクラスターを形成している状態を、示したものである。13個の点は、それぞれ個体を表すものであり、個体13番のみは、クラスターに所属せずに孤立している。よくみると、個体2番を中心とする「かたまり」と、個体9番を中心とする「かたまり」があることが、理解されるであろう。したがって、1番から12番までの個体によって、1個のクラスターが形成されている状態よりは、1番から6番までの個体によって、クラスター C_1 が、また7番から12番までの個体によって、クラスター C_2 が形成されているという状態の方が、望ましいわけである。しかし、単一連結法を利用すると、6番と7番の連結によって、 C_1 と C_2 が融合されてしまう。このような効果を、「鎖効果」(chain effect)と呼ぶ。「モード分析」は、鎖効果を抑止するために、考案された手法である。

「モード分析」は、以下の手順で行われる。①閾値 D および「密度閾値」(density threshold) k が、研究者

によって、任意に決定される。ただし、 k は正の整数であり、3, 4, 5 などの数値が選定されることが多い。

②すべての個体相互間の距離 $d_{\alpha\beta}$ を計算する。③各個体 α ごとに、 α を中心として、半径 D の超球面内に入る、 α 以外の個体の数を調べる。個体数が k 以上ならば、 α を「稠密点」(dense point)、 k 未満ならば、 α を「非稠密点」(non-dense point)と呼ぶ。④稠密点のみを単一連結法と同様の手順で連結し、クラスターを形成する。(①から④までの手順を、 $k=1$ として行う場合には、単一連結法と同一の結果になる。)⑤非稠密点のなかで、もっとも近い稠密点との距離が D 以下のものを、その稠密点が所属しているクラスターに入れる。この段階におけるクラスターを、「核(nuclei)レベル」のクラスターと呼ぶ。⑥残りの非稠密点を、それぞれ、もっとも近い稠密点が所属しているクラスターに入れる。この段階におけるクラスターを、「完全(complete)レベル」のクラスターと呼ぶ。

図2のデータを、 $k=3$ として、以上の手順を踏んで処理してみよう。まず③で、6, 7, 13番を除く各個体が、稠密点になる。つぎの④で、1番から5番までの個体によって、クラスター C_1 が、また8番から12番までの個体によって、クラスター C_2 が形成される。そして、⑤では、6番が C_1 、7番が C_2 に、さらに⑥では、13番が C_2 に入れられる。

つぎに、階層的的手法(hierarchical method)について、紹介しよう。階層的的手法にもさまざまな種類があるが、大別すると、2つの系列に整理される。一方は、すべての個体が1個のクラスターに所属している状態から出発し、クラスターの分割(partition)を反復していく手法である。もう一方は、すべてのクラスターが、それぞれ1個ずつの個体によって形成されている状態から出発し、クラスターの融合(fusion)を繰り返していく手法である。

後者の系列に属する手法は、きわめて数多く考案されており、しかも、広く普及しているものが多い。ここでは、そのなかから、ランス(G. N. Lance)とウィリアムズ(W. T. Williams)の「組合せ手法」(combinational method)¹³⁾と、ウィシャートの「階層的モード分析」(hierarchical mode analysis)¹⁴⁾をとりあげることにした。

4) ランスとウィリアムズの「組合せ手法」は、以下の手順で行われる。①すべての個体相互間の距離 $d_{\alpha\beta}$ を計算する。②初めは、すべてのクラスターが、それぞれ1個ずつの個体によって形成されているので、クラ

スター数は n 個である。 n 個のクラスターのなかで、距離 D_{fg} が最小のクラスター C_f と C_g を融合する。クラスター数は $(n-1)$ 個になる。③ 同様に、クラスター数が 1 個になるまで、融合を反復する。

組合せ手法においては、クラスター C_h と C_l が融合して C_g が形成される場合に、他のクラスター C_f と C_g との距離 D_{fg} を、融合前の D_{fh} , D_{fl} および D_{hl} だけから算出する。個体間距離としてユークリッド距離を使用する場合には、 D_{fg} の計算式として、以下のような 2 通りが考えられている。

$$D_{fg} = a_h D_{fh} + a_l D_{fl} + b D_{hl} + c |D_{fh} - D_{fl}| \quad (11)$$

$$D_{fg}^2 = a_h D_{fh}^2 + a_l D_{fl}^2 + b D_{hl}^2 + c |D_{fh}^2 - D_{fl}^2| \quad (12)$$

ここで、 a_h , a_l , b , c はすべてパラメーターである。

ランスとウィリアムズは、パラメーターを変えることによって、つぎの 6 種類の組合せ手法を導出した。すなわち、i) 最近隣法 (nearest neighbor method), ii) 最遠隣法 (furthest neighbor method), iii) メディアン法 (median method), iv) 重心法 (centroid method), v) グループ平均法 (group-average method), vi) 可変法 (flexible method) という、6 通りの手法である。ここでは、それぞれについて、簡単な説明を加えるにとどめておこう。なお、i) ii) iii) においては (11) 式が、また iv) v) vi) においては、(12) 式が使用されている。

i) 最近隣法においては、(11) の係数として、 $a_h = a_l = \frac{1}{2}$, $b = 0$, $c = -\frac{1}{2}$ が選定され、

$$D_{fg} = \frac{1}{2} D_{fh} + \frac{1}{2} D_{fl} - \frac{1}{2} |D_{fh} - D_{fl}|$$

$$= \begin{cases} D_{fl} & (D_{fh} \geq D_{fl} \text{ のとき}) \\ D_{fh} & (D_{fh} \leq D_{fl} \text{ のとき}) \end{cases} \quad (13)$$

となる。したがって、クラスター間の距離 D_{fg} は、それぞれのクラスターに所属する個体間の距離の最小値、すなわち $\text{Min. } \{d_{\alpha\beta} | \alpha \in C_f, \beta \in C_g\}$ である。

ii) 最遠隣法においては、(11) の係数として、 $a_h = a_l = \frac{1}{2}$, $b = 0$, $c = \frac{1}{2}$ が選定され、

$$D_{fg} = \frac{1}{2} D_{fh} + \frac{1}{2} D_{fl} + \frac{1}{2} |D_{fh} - D_{fl}|$$

$$= \begin{cases} D_{fh} & (D_{fh} \geq D_{fl} \text{ のとき}) \\ D_{fl} & (D_{fh} \leq D_{fl} \text{ のとき}) \end{cases} \quad (14)$$

となる。したがって、クラスター間の距離 D_{fg} は、それぞれのクラスターに所属する個体間の距離の最大値、すなわち $\text{Max. } \{d_{\alpha\beta} | \alpha \in C_f, \beta \in C_g\}$ である。

iii) メディアン法においては、(12) の係数として、 $a_h = a_l = \frac{1}{2}$, $b = -\frac{1}{4}$, $c = 0$ が選定され、

$$D_{fg}^2 = \frac{1}{2} D_{fh}^2 + \frac{1}{2} D_{fl}^2 - \frac{1}{4} D_{hl}^2 \quad (15)$$

となる。したがって、この手法では、 C_g の代表点として、 C_h の代表点と C_l の代表点との中点がとられるわけである。

iv) 重心法においては、クラスター C_h , C_l , C_g に所属する個体数を、それぞれ n_h , n_l , n_g 表わすとき、(12) の係数として、 $a_h = \frac{n_h}{n_g}$, $a_l = \frac{n_l}{n_g}$, $b = -\frac{n_h n_l}{n_g^2}$, $c = 0$ が選定され、

$$D_{fg}^2 = \frac{n_h}{n_g} D_{fh}^2 + \frac{n_l}{n_g} D_{fl}^2 - \frac{n_h n_l}{n_g^2} D_{hl}^2 \quad (16)$$

となる。したがって、クラスター間の距離は、それぞれのクラスターの重心間の距離である。

v) グループ平均法においては、(12) の係数として、 $a_h = \frac{n_h}{n_g}$, $a_l = \frac{n_l}{n_g}$, $b = c = 0$ が選定され、

$$D_{fg}^2 = \frac{n_h}{n_g} D_{fh}^2 + \frac{n_l}{n_g} D_{fl}^2 \quad (17)$$

となる。この手法では、クラスター間の距離は、それぞれのクラスターに所属する個体間の距離の平均である。

vi) 可変法は、(12) で、 $a_h = a_l$, $b < 1$, $c = 0$, $a_h + a_l + b = 1$ とすることによって、融合後の空間密度を制御しようとする手法である。すなわち、 b の値が 1 に近づくほど空間が濃縮され、また b の値が負になれば、空間の拡散が起こることになる。

このように、一口に組合せ手法といっても、さまざまな種類があるわけである。

5) ウィンシャートの「階層的モード分析」は、前述の「モード分析」を、階層的なアルゴリズムに直したものである。「モード分析」においては、閾値 D と密度閾値 k がともに固定されていたが、「階層的モード分析」では、 k の値のみを固定しておいて、 D の値を逐次的に大きくし、稠密点の個数を増してゆく、という工夫がなされている。クラスターは稠密点のみによって形成されるものであり、またクラスター間の距離は、それぞれのクラスターに所属する稠密点間の距離の最小値と定められている。

「階層的モード分析」は、以下の手順で行われる。① 密度閾値 k が、研究者によって、任意に決定される。 $k=1$ として行えば、最近隣法と同一の結果になる。② すべての個体相互間の距離 $d_{\alpha\beta}$ を計算する。③ 各個体 α ごとに、 α から k 番目に近い個体との距離を $PD(\alpha)$

とする。④ PD の値が小さい順に個体を並べ、 $KP(1)$, $KP(2)$, ..., $KP(n)$ と表わす。⑤ PD の添字 α を、対応する KP の添字と同じものに入れかえる。たとえば、 $KP(2)$ ならば $PD(2)$ とする。⑥ D の初期値を $PD(1)$ として、以後 ⑦ ⑧ ⑨ の要領で、 D の値を逐次的に大きくしてゆく。 $KP(\alpha)$ として表わされる個体は、 $PD(\alpha) \leq D$ のとき、稠密点と呼ばれる。したがって、 $KP(1)$, $KP(2)$, ..., $KP(n)$ の順番で、稠密点になっていく。

⑦ さて、 $KP(\alpha-1)$ までが、稠密点になっているものとしよう。すなわち、 $PD(\alpha-1) \leq D$ である。ここで、 $PMIN$ と $DMIN$ の大小を比較する。ただし、 $PMIN$ は、次に稠密点になる順番の個体 $KP(\alpha)$ から、 k 番目に近い個体までの距離（すなわち、 $PMIN = PD(n)$ ）である。また $DMIN$ は、異なったクラスターに所属する2個の稠密点の最小距離である。⑧ $PMIN > DMIN$ ならば、 D の値を $DMIN$ まで大きくして、距離 $DMIN$ の2個のクラスターを融合し、⑦に戻る。⑨ $PMIN \leq DMIN$ ならば、 D の値を $PMIN$ まで大きくして、個体 $KP(\alpha)$ を稠密点にする。この場合、i) $KP(\alpha)$ を中心とする半径 $PMIN$ の超球面内に稠密点がなければ、 $KP(\alpha)$ は単独で1個のクラスターを形成する。ii) 同超球面内に稠密点が1個だけあるならば、 $KP(\alpha)$ は、その稠密点が所属しているクラスターに入れられる。また稠密点が2個以上あっても、すべて同じクラスターに所属していれば、まったく同様である。iii) 同超球面内に、所属するクラスターの異なる2個以上の稠密点があるならば、それらのクラスターと $KP(\alpha)$ が、1個のクラスターに融合される。これら i) ii) iii) のいずれの場合にも、⑦に戻る。⑩ 以上の ⑦ ⑧ ⑨ の手順を、 $D = PD(n)$ になるまで、すなわち、すべての個体が稠密点になるまで反復する。⑪ すべての個体が稠密点になったときに、クラスターが2個以上あれば、組合せ手法の

なかの最近隣法を用いて、クラスターが1個になるまで、融合を繰り返す。

なお、階層的手法においては、個体の融合過程が、図3のような樹形図 (dendrogram) として表現される。

6) さいごに、分割の妥当性の指標を使用する手法を紹介しよう。この手法は、フリードマン (H. P. Friedman) とルービン (J. Rubin) によって展開されているものであり¹⁶⁾、つぎの4つのステップから成り立っている。① クラスターの個数 m が、研究者によって任意に決定される。② n 個の個体を m 個のクラスターに分割する場合の、分割の妥当性の指標を設定する。③ なんらかの方法で、 n 個の個体を m 個のクラスターに分割して、分割の初期状態とする。④ 初期状態から出発して、分割を変化させ、試行錯誤的に改良していく。すなわち、変化させる度に指標を計算し直して、指標の値ができるだけ大きく（または小さく）なるような分割の状態を追求する。

以下においては、まず3種類の指標（すなわち、i) $\text{tr } W$, ii) $|T|/|W|$, iii) $\text{tr } W^{-1}B$ ）について解説し、つぎに、試行錯誤的に分割を改良していくための3通りの手法（すなわち、i) 丘登り法、ii) 強制移動法、iii) 再配置法）を説明しよう。

ここで、全変動（偏差平方和・積和行列）を T 、クラスター間の変動を B 、クラスター内の変動の和を W と表わすと、

$$T = B + W \quad (18)$$

の関係がある。クラスター C_k 内の変動を $W^{(k)}$ とすれば、 $T, B, W, W^{(k)}$ は、すべて $p \times p$ の行列である。ここで、

$$T = (t_{ij}) \quad B = (b_{ij})$$

$$W = (w_{ij}) = \sum_{k=1}^m W^{(k)} \quad W^{(k)} = (w_{ij}^{(k)})$$

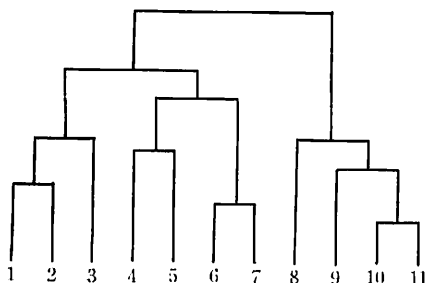
と表わし、クラスター C_k の個体数を n_k 、変数 i についての総平均値を \bar{x}_i 、変数 i についての、 C_k 内の平均値を $\bar{x}_i^{(k)}$ とすると、各行列の要素は、

$$t_{ij} = \sum_{\alpha=1}^n (x_{\alpha i} - \bar{x}_i)(x_{\alpha j} - \bar{x}_j) \quad (19)$$

$$b_{ij} = \sum_{k=1}^m n_k (\bar{x}_i^{(k)} - \bar{x}_i)(\bar{x}_j^{(k)} - \bar{x}_j) \quad (20)$$

$$w_{ij} = \sum_{k=1}^m w_{ij}^{(k)} \quad (21)$$

$$w_{ij}^{(k)} = \sum_{\alpha \in C_k} (x_{\alpha i} - \bar{x}_i^{(k)})(x_{\alpha j} - \bar{x}_j^{(k)}) \quad (22)$$



<図3> 樹形図

(数字は、各個体を表わしている)

と表わされる。

T が一定であるから、 W を小さくすれば、 B は大きくなる。ところで、それぞれのクラスターが、相互に近似性の高い個体によって形成されるようにするためには、 W を小さくしなければならない。このような見解に基づいて、フリードマンとルービンは、つぎのような指標を提案している。

$$\text{tr } W \rightarrow \text{Min.} \quad (23)$$

$$\frac{|T|}{|W|} \rightarrow \text{Max.} \quad (24)$$

$$\text{tr } W^{-1} B \rightarrow \text{Max.} \quad (25)$$

このなかで、 $|T|/|W|$ は、ウィルクス (S. Wilks) の A 統計量の逆数である。計算の便宜のために、代って $\log(|T|/|W|)$ が使用される場合が多い。

以上のような指標を設定したならば、指標をできるだけ大きく ($\text{tr } W$ の場合のみ、小さく) するような分割の状態を、追求しなければならない。理想的な手法は、 n 個の個体を m 個のクラスターに分割するとき、可能な分割状態をすべて試みて、それぞれの場合について指標を計算し、最良の分割状態を選び出すことである。しかし、この手法では、個体数 n がやや大きければ、天文学的計算時間を費す結果になる。したがって、最良の分割に近い状態に到達するために、試行錯誤的に分割を変化させて、指標ができるだけ大きく (または小さく) なるような状態を求める手法が開発された。それが、つぎに述べる丘登り法、強制移動法および再配置法である。

i) 丘登り法 (hill-climbing passes) の手順は、以下の通りである。① 適当な分割を初期状態として与え、指標の計算を行う。② 各個体を、それが所属しているクラスター以外の各クラスターに移動させ、そのつど指標を計算する。③ もしも、どのクラスターに移動させても、指標の値が改良されなければ、もとのクラスターに入れたままにしておく。改良されたならば、最良になったときのクラスターに所属させる。④ どのように個体を移動させても、指標が改良されることはない、という状態に到達するまで、すべての個体について、② ③ の手順を反復する。

ii) 強制移動法 (forcing passes) の手順は、以下の通りである。① 適当な分割を初期状態として与え、指標の計算を行う。② クラスター C_1 に所属している個体のなかで、 C_1 の重心からもっとも距離の遠い個体 α を、 C_1 以外のクラスターで、 α からもっとも近い重心をもつものへ移動させる。③ 移動の度に、各クラスターの重心、

クラスター内の分散・共分散行列、個体と重心とのマハラノビスの汎距離、および分割の妥当性の指標を、計算し直す。④ 以上の②③の手順を、 C_1 に所属する個体が1個になるまで反復する。⑤ 指標の値がもっとも改良されたときの、分割の状態に戻す。⑥ 同様の手順を、 C_2, C_3, \dots, C_m について反復する。

iii) 再配置法 (reassignment passes) の手順は、以下の通りである。① 適当な分割を初期状態として与え、指標の計算を行う。② 各クラスターの重心、分散・共分散行列、および個体と重心とのマハラノビスの汎距離を計算する。③ 各個体を、もっとも近い重心をもつクラスターに所属させる。このとき、強制移動法と異なって、②のそれぞれの値を固定し、個体が移動しても計算し直さない。④ すべての個体について、③の手順を反復する。⑤ 指標の計算を行う。⑥ 以上の②から⑤までの手順を、指標の値が改良されなくなるまで繰り返す。

なお、フリードマンとルービンは、丘登り法、強制移動法、および再配置法を、組合せて反復試行するように提示している。

この手法においては、クラスター数 m が、研究者によって任意に決定されなければならないことが、ひとつの限界になっている。しかし、多数のクラスター分析の手法のなかでも、もっとも緻密に展開されているものの一つであり、類型構成の目的でクラスター分析を利用する場合にも、きわめて有効な手法であろうと考えられている。

4. クラスター分析による類型構成の事例

最後に、クラスター分析による類型構成法を、フィールド調査のデータに適用した事例を紹介しよう。本節でとりあげる調査研究は、「消費者生活体系と広告情報」というタイトルのもとに、筆者らを含む6名の共同研究として行われたものである¹⁰⁾。調査の基本的狙いは、消費者の生活体系が異なれば、情報探索行動と広告にたいする態度に関して、いかなる差異が生ずるであろうか、という点を明らかにすることにあった。われわれは、一連の調査結果の集計・分析作業の最終ステップとして、数個の生活体系類型を構成し、各類型に特有な情報探索行動と広告にたいする態度のパターンを確定することを試みた。以上のような事情により、この調査研究は、クラスター分析による類型構成法の好個の事例となったのである。

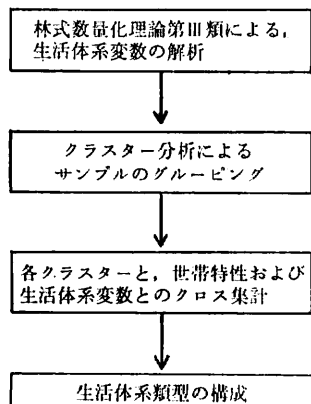
なお、ここでいう「生活体系」概念は、主体の生活平

面と行動領域への、「システム」概念の意識的適用から生まれたものである。複合概念としての「生活体系」については、さしあたり、簡単につきのように定義しておこう。

「生活体系」とは、4つの機能的諸領域（すなわち、① 適応、② 目標達成、③ 統合連帯、④ パターン維持と緊張処理）と、3つの構成要素（すなわち、① 生活構造、② 生活意識、③ 生活行動）との相互依存的なサブシステムから構成される複合体であって、周期的に反復される安定したリズムをもち、日々、週、月、年、あるいは一生を通じてのサイクルとなって現われる、均衡持続の傾性を示すシステムである。生活体系の形成・維持・発展のためには、① 環境への適応（生活資源の獲得と欲求の触発・解消をも含む）、② 生活目標の設定と達成、③ 統合の形成と維持、④ 価値・規範の内面化の保持と緊張の処理、の4つの機能的緊張事態（functional exigencies）が、日々、週、月、年ごとに解決されなければならない¹⁷⁾。

われわれの調査研究においては、「生活体系」概念を、分析のためのキー概念として設定し、生活体系変数 61（設問数 16）——生活構造変数 26＋生活意識変数 35——が、フィールド調査の質問紙に採用された。なお、世帯特性 19（設問数 5）も、質問紙に組込まれた。世帯特性は、一般に人口学的要因と呼ばれるものだが、「生活体系」コンテキストにおいては、生活体系規定要因として解釈されている。

フィールド調査は、昭和 45 年 10 月に、東京都目黒区および台東区の 428 世帯を対象として実施された。調査方法としては、質問紙による面接調査法を採用し、主婦もしくは主婦の役割担当者にたいする面接を通して、対象世帯に関する情報を得ることにした。



<図 4> 生活体系類型の構成手順

ここで、本調査における、生活体系の類型構成の分析ステップを、簡単に説明しておこう。図 4 にみられるように、類型構成にいたる集計・分析の手続きは、大別して、つぎの 4 つのステップから成り立っている。

第 1 ステップ……生活体系諸変数を、林式数量化理論第 Ⅲ 類の手法を用いて解析し、生活体系の基本的軸を構成する¹⁸⁾。

第 2 ステップ……数量化 Ⅲ 類の手法によって、それぞれの軸ごとに、各サンプル（調査対象）に付与された数値をデータとして、クラスター分析を行う。解析の結果、数個のクラスターが形成され、各クラスターは、相互に近似した「生活体系プロフィール」を持ったサンプルのあつまりとなる。

第 3 ステップ……サンプルのクラスター区分と、生活体系変数および世帯特性とのクロス集計を行う。集計結果を分析して、各クラスターの生活体系の諸特性と、クラスター間の差異を明確化する。

第 4 ステップ……各クラスターの統計的諸特性の検出を通じて、クラスターごとに「生活体系類型」を構成する。

なお、当然のことながら、第 1 ステップから第 3 ステップまでは、コンピューターによる処理をとまらう。計測機械は、IBM 360 および GE 635 を使用した。

本調査では、第 1 ステップから第 4 ステップまでの分析が、並行して 2 通り行われた。これは、第 1 ステップにおいて、数量化 Ⅲ 類の手法を用いて解析されるためのデータが、2 組作成されたことによる。以下の論述では、まず数量化 Ⅲ 類の解析結果を簡単に紹介し、つぎにクラスター分析を用いて構成された各類型について、説明を加えることにしたい。

数量化 Ⅲ 類の解析のためのデータは、つぎのような要領で、2 組作成された。まず、質問紙に採用された生活構造変数と生活意識変数のなかから、とくに代表的なものと思われる 23 変数が選択され、それぞれの変数についての、各サンプルの測定値が、解析用データとしてまとめられた。つぎに、とくに生活意識に焦点を合わせることによって、19 個の生活意識変数が選択され、各サンプルの測定値がまとめられた。以下では、前者を X 案、後者を Y 案と呼ぶことにしたい。解析用データが 2 組も作成されたのは、われわれの使用した数量化 Ⅲ 類のコンピューター・プログラムの制約上、質問紙に採用された生活体系変数のすべてを、同時にインプットできなかったためである。

X 案・Y 案のそれぞれについて、数量化 Ⅲ 類の手法

による解析が行われた。両案ともに第5軸まで求めたところ、固有値が高く、かつ解釈可能な軸の数とその意味づけは、つぎのような結果になった。

X—I 軸 (生活水準).....どの程度豊かな生活を送っているか、また、どれだけ多くの財を保有しているか、という側面に関連している。すなわち、消費生活の水準の高低を表わす軸と解釈される。

X—II 軸 (流動性—固定性).....生活が流動的で変化に富んでいるか、あるいは生活が固定的で変化に乏しいか、という側面を表わす軸である。

Y—I 軸 (流動性—固定性).....X—I 軸と同様な性格の軸である。

Y—II 軸 (生活充実志向).....充実した生活にたいす

る願望が強く、生活向上の努力をしているか、あるいは、充実した生活にたいする志向が薄弱なのか、という側面を表わす軸である。

Y—III 軸 (合理性).....わりきった態度で、生活を合理的に営んでいるかどうか、という側面を示す軸と解釈される。

なお、X—I 軸と X—II 軸は、X 案の解析によって、また Y—I 軸、Y—II 軸および Y—III 軸は、Y 案の解析によって構成されたものである。

X 案・Y 案の数量化 III 類による解析結果のそれぞれについて、2通りのクラスター分析が行われた。X 案による方を A グループ、Y 案による方を B グループと呼んでおこう。計算方法としては、第3節で紹介

<表2> 各クラスターの特性と構成された類型 (A グループ)

クラスター No.	世帯特性	生活体系の諸特性	構成された類型	典型的イメージ
A—I クラスター	商工サービス業 (自営)、管理職、自由業が多く、所得水準が高い。40代が多く、東京出身者・山の手居住者が多い。	食生活・住生活洋風化度、外出志向、革新性、社交性、高級品志向のスコアが高く、堅実性のスコアが低い。耐久財保有水準が高く、レジャー費の割合が大きい。	「富裕洗練型」 ・生活水準が高い ・壮年 ・洗練された生活	住宅街の開業医夫婦。裕福で、スマートな生活を送っている。
A—II クラスター	専門技術職、事務職が多く、所得水準が低い。20代・30代が多く、学歴が低い。	耐久財保有水準が低く、また、食生活・住生活洋風化度、対人関心度、権威主義 (上下関係意識)、社会的感性、計画性、銘柄固定性、高級品志向などのスコアが低い。世代間移動アスピレーションのスコアのみ高い。	「安直萎縮型」 ・生活水準が低い ・青年 ・社会的参加度が低い ・消極的、非活動的	区役所の無気力職員世帯。元気がなく、イージー・ゴーイングな感じ。
A—III クラスター	専門技術職、事務職が多く、所得水準が低い。10代・20代が多い。	食生活洋風化度、外出志向、対人関心度のスコアが高く、生活目標では「その日その日を楽しむ」が多い。一方、堅実性、計画性、銘柄固定性、高級品志向、権威主義 (上下関係意識)、個別主義 (義理・恩意識) のスコアは低い。	「軽薄不均等型」 ・生活水準が低い ・青年 ・社会的参加度が低い ・享乐的 ・生活体系の均衡が不安定	繁華街の商店員。多少、おっちょこちょいの「今どきの若い連中」。
A—IV クラスター	所得水準は中程度で、30代・40代・50代が多い。下町居住世帯が多く、居住歴が長い。	堅実性、計画性、家事愛好度、個別主義 (義理・恩意識) のスコアが高い。一方、外出志向、消費志向、社会的感性、食生活・住生活洋風化度のスコアは低い。	「保守慎重型」 ・生活水準は中程度 ・中年 ・堅実 ・保守的	町工場の主人夫婦。勤勉実直な頑固者。
A—V クラスター A—VI クラスター	他のクラスターの中間的な位置を占め、明確な特徴を示していないので省略する。			

〈表3〉 各クラスターの特性と構成された類型 (B グループ)

クラスター No.	世帯特性	生活体系の諸特性	構成された類型	典型的イメージ
B-I クラスター	所得水準が低く、地方出身者が多い。学歴も低く、年齢では50代が多い。	革新性、達成要求、対人関心度、社会的感性、外出志向、社交性、融和性、高級品志向、銘柄固定性などのスコアがすべて低い。家計支出額が小さくて、耐久財保有水準が低く、耐久財要求水準は、それに輪をかけて低い。	「閉鎖停滞型」 ・生活水準が低い ・老年 ・社会的参加度が低い ・消極的、非活動的 ・ライフ・アップ動機づけが弱い	停年退職後のビルの守衛夫婦。ひっそりと暮らしている感じ。
B-II クラスター	商工サービス業(自営)が多い。所得水準は、Bグループシステムの中では、最高である。	生活意識については、中間的なスコアが多く、はっきりした特色がみられない。銘柄固定性、消費志向、家事愛好度のスコアが高く、革新性、社会的感性のスコアが低い、という程度である。しかし、耐久財保有水準は高く、家計支出額も大きい。	「余裕安定型」 ・生活水準が高い ・中年 ・生活意識は中庸穏健的	老舗の中年主人夫婦。ゆとりがあって、落ち着いている。
B-III クラスター	事務職、専門技術職、管理職が多く、所得水準はやや高い方である。30代・40代が多い。	堅実性、達成要求、計画性、家事愛好度、権威主義(上下関係意識)、個別主義(義理・恩意識)、世代間移動アスピレーションなどのスコアが、いずれも高い。食費の割合が大きく、貯蓄志向も大きい。	「堅実前向き型」 ・生活水準は中程度 ・壮年 ・堅実 ・積極的	大企業勤務の中堅サラリーマン世帯。まじめなマイ・ホーム主義者
B-IV クラスター	商工サービス業(自営)、専門技術職、事務職が多く、所得水準は中程度である。学歴は高く、20代が多い。	外出志向、消費志向、社交性、社会的感性、食生活・住生活洋風化度のスコアが高い。一方、堅実性、計画性、家事愛好度、銘柄固定性、世代間移動アスピレーションなどのスコアは低く、貯蓄性向も小さい。	「享楽感覚型」 ・生活水準は中程度 ・青年 ・享乐的 ・衝動的	団地・社宅住まいのレジャー志向若夫婦。遊び好きで、ややうわついた感じ。
B-V クラスター	20代の単身世帯が多く、所得水準は低い。地方出身者・下町居住者が多く、居住歴は短い。	世代間移動アスピレーション、衣生活洋風化度(主婦の結婚式参列の服装)、堅実性のスコアが高い。一方、権威主義(上下関係意識)、個別主義(義理・恩意識)、社交性、家事愛好度、銘柄固定性、高級品志向、住生活洋風化度のスコアは低い。	「浅慮消極型」 ・生活水準が低い ・青年 ・社会的参加度が低い ・消極的 ・合理的	地方出身の無口なブルー・カラー。パツとしない毎日を、淡々とすごしている感じ。
B-VI クラスター	商工サービス業(自営)が多く、所得水準は高い。学歴も高く、年齢別では30代が多い。	革新性、社会的感性、外出志向、融和性、社交性、権威主義(上下関係意識)、高級品志向、達成要求、のスコアがいずれも高く、堅実性のスコアは低い。食費の割合が小さくて、レジャー費の割合が大きい。	「開放革新型」 ・生活水準が高い ・壮年 ・社会的参加度が高い ・積極的、活動的 ・ライフ・アップ動機づけが強い	青年エグゼクティブ世帯。新しいものをどしどしとり入れ、交際も活発である。

された数種類の技法のなかから、分割の妥当性の指標として $\log(|T|/|N|)$ を用いる手法が選択された。データには、数量化 III 類の手法によって、それぞれの軸ごとに、各サンプルに付与された数値が使用された。クラスターの個数に関しては、分析の目的、従来の経験、サンプル数などを考慮して、6個のクラスターに分割するのが適当であろう、ということに落ち着いた。

つぎに、A グルーピングと B グルーピングの 2 通りのクラスター区分を、それぞれ一つの変数として、生活体系諸変数および世帯特性とのクロス集計を行った。集計結果の分析によって、生活体系に関する各クラスターの諸特性と、クラスター間の差異が明確化された。

こうして明らかにされた各クラスターの統計的諸特性のなかから、一定の分析視点に立って、ある特性を強調し、ある特性を捨象し、相互に関連の強い複数の特性を概括することによって、各クラスターごとに「生活体系類型」が構成された。表 2 および表 3 は、各クラスターの諸特性、構成された生活体系類型、および各類型についての具体的なイメージを整理したものである。

本節で紹介した研究例は、あくまでもささやかな試験的な応用例にすぎない。しかし、表 2・表 3 は、クラスター分析による類型構成法の有用性を、ある程度示唆するものと思われる。クラスター分析の積極的利用によって、社会学における類型論的アプローチは、主観性と恣意性の強いものから、より有効なより精密な方法へと移行してゆくであろう、というのが、本論におけるわれわれの結論である。

(注)

- 1) Howard Becker, "Constructive Typology in the Social Sciences," *American Sociological Review*, 5 (February, 1940).
- 2) Paul F. Lazarsfeld and Allen H. Barton, "Quantitative Measurement in the Social Sciences: Classification, Typologies, and Indices" in D. Lerner and H. D. Lasswell, ed., *The Policy Sciences*, (Stanford University Press, 1951), pp. 155-192.
- 3) Carl G. Hempel, "Typological Methods in the Natural and Social Sciences" in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, (The Free Press, 1955)

pp. 155-171.

- 4) John C. McKinney, *Constructive Typology and Social Theory*, (Appleton-Century-Crofts, 1966).
- 5) *Ibid.*, p. 3. なおマキニーは、とくに「構成された類型」(constructed type) という語を使用している。
- 6) 類型とその近似物との判別に関しては、*ibid.*, pp. 14-16 参照。
- 7) V 分析に関しては、安田三郎「社会統計学」(丸善, 1969) pp. 84-100 参照。
- 8) クラスター分析の諸手法は、次書において、手際よく整理されている。奥野忠一、芳賀敏郎、久米均、吉沢正共著「多変量解析法」(日科技連, 1971) pp. 391-411.
- 9) この表の作成に際しては、以下の文献を参考にした。①奥野ほか、前出, p. 395. ② D. Whishart, *A FORTRAN II Program for Numerical Classification*, (Computing Laboratory, Univ. of St. Andrews, 1968), p. 20.
- 10) Ronald B. Frank and Paul E. Green, "Numerical Taxonomy in Marketing Analysis" in J. Ardnt, ed., *Insights into Consumer Behavior*, (Allyn and Bacon, 1968), pp. 101-126.
- 11) Robert R. Sokal and Peter H. A. Sneath, *Principles of Numerical Taxonomy*, (Freeman & Company, 1963), pp. 180-181.
- 12) D. Whishart, *op. cit.*
- 13) G. N. Lance and W. T. Williams, "A General Theory of Classificatory Sorting Strategies 1—Hierarchical Systems," *Computer Journal*, Vol. 9 (1967), pp. 373-380.
- 14) D. Whishart, *op. cit.*
- 15) H. P. Friedman and J. Rubin, "On Some Invariant Criteria for Grouping Data," *Journal of American Statistical Association*, Vol. 62 (1967), pp. 1159-1178.
- 16) この調査研究は、吉田秀雄記念事業財団の助成金により、慶応義塾大学商学部教授 村田昭治を主管として、慶大商学部助教授 清水猛、地域開発研究所主任研究員 青池慎一、横浜国立大学経営学部助手 川島行彦に筆者らを加えた、計 6 名の共同研究として行われた。
- 17) 「生活体系」概念に関しては、吉田正昭、村田昭治、井関利明共編「消費者行動の理論」(丸善, 1969) 第 4 章参照。
- 18) 林式数量化理論に関しては、林知己夫、樋口伊佐夫、駒沢勉共著「情報処理と統計数理」(産業図書, 1970) pp. 223-350 参照。