教育調査のための学区の層別*

水　野　欽　司

I　はじめに

教育調査や実験のために，多くの小，中学校から少数の対象校を選ぶとき，それらが研究の目的・意図に照らして偏りなく選ばれるように配慮すべきことはいうまでもない。

現実の多くの学校には，さまざまな意味の‘学校差’が存在する。当面それが主たる課題でない場合でも，調査や実験の結果にそれが影響してくることは否定できない。したがって実験に際しては，‘学校差’をブロック要因とするような計画を立てる必要がある。またこのような‘学校差’に積極的に注目し，学校を適当に層別して調査を行なうならば，調査測定の精度を高める上に有効であることは標本抽出理論の教えるところである。学校を選んで行なう教育心理学的な研究においては，調査地域の学校をなんらかの観点で事前に分類―層別―しておくことが，研究を進める上できわめて有効であることに異論はないといえる。

ところで，学校分類のこのような重要性，有効性にもかかわらず実際にそれを行なおうとすると適切な資料が簡単には入手できないのが現状であろう。

一般に調査対象の層別は調査内容と相関が高いと予想される既知の特性を用いて行なわれる。したがってすべての場合に通用する層別資料というものは原理的には存在しない。しかし，学校またはその一部の学級を対象とする研究においては児童・生徒の種々の能力，パーソナリティ，態度，行動などを調査・実験の内容としている。それゆえ，ある程度一般性のある層別の情報を見出し，それを整備しておくことは無意味ではなく，むしろ積極的に心がけるべきことと考えられる。このような層別のための学校資料は教育研究者あるいは教育行政担当者の手で整備されるべきものである。しかし，それが十分でなかったり，利用が困難な事情にあるときでも入手できる資料は極力活して行く努力を払うべきである。

一方それと同時に，このような資料が得られたとき能率よく層別を行なう技術的な方法を用意しておくことも必要であろう。

以下，本稿では，資料が特性項目の各区分に対する度数分布の形をとる場合（連関表）のクラスター化の手法と国勢調査で得られるデモグラフィックな資料を用いて，

1）「学校」に代るものとしての「学区」（名古屋市174学区）の層別を試みること
2）層別の手段としての手法の有効性について検討を加えること

を目的とする。

層別に用いた資料および方法の有効性，妥当性は，もちろん，層別に基づく教育的な調査・実験の諸結果を待って確かめられるべきことであるが，それは将来の機会に待つことにして，ここでは試算結果の報告にとどめる。

II　学区の層別とその方法

1.　層別の資料

一般に‘学校差’といってもその意味内容は多様である。その内容を強いてわければ次の三つの側面を挙げられよう。

1）児童・生徒側の特性……知能・学力の水準，パーソナリティ・態度・行動の特徴など
2）教育側の特性……教師のいろいろな質的・量的特性，教育方針，学校の設備とその規模など
3）家庭・地域環境の特性……親の教育態度，家族構成，職業・収入などの条件，居住地域の産業的構造など

これらの変数は学校層別の補助情報という前に，教育調査や実験が目的とする分析変数でもある。まさしく，これらについて数多くの研究が従来から行なわれている。しかし学校の層別という目標に限ってみれば，全学

*　本研究のデータ処理には名大大型電子計算機センターの FACOM 230-60 を利用した。

校についてこれらの特徴を整理した資料はまず存在しないといってよい。

資料入手の容易性という方向から，教育調査・実験の内容と関連の深いものを求めると国および地方自治体が行政的立場で行なう統計の資料を挙げることができる。その一つは5年ごとに行なわれる国勢調査の「学区」別の「各才別」，「最終卒業学校別」，「労働力状態別」，「産業分類別」，「職業分類別」などの人口データである。

これらのデータは‘学校’でなく「学区」別のデータであるが，1学区が一つの公立小学校に対応すること*を考慮すれば個々の学校に代るものであり，上で挙げた3）の家庭・地域環境の特性を表現するものとみなすことができる。特に「最終卒業学校別」や「職業分類別」は多くの意識調査（世論調査，市場調査など）の重要なフェイスシート項目として使われるように，態度・行動のちがいを識別するよい項目であり，しかも他の，たとえば「各才別人口」などと比べて地域間での差異が大きい。またこれらの項目は地域の社会・経済的な水準や質に密接に相関している。したがって，こどもの親も含めた地域住民のこどもに対する教育的な影響の質的な差を，これら項目データが反映すると期待できる。実際にこれらが児童・生徒の学力・態度・行動と有効に関連するかどうかについては，これもまた調査・実験によって検証しなければならないが，ここではまず「学区」別のデモグラフィックなデータによる公立小学校の層別を試みる。

ところで，このようなデモグラフィックなデータによる地域分類は，従来しばしば行なわれている。たとえば，元のデータから「二次産業就業人口比率」といった指標を作成し，適当な比率段階をもうけて分類するような仕方である。このような単独指標化による層別では，たとえば，二次産業以外の一次産業，三次産業の構成についての情報は無視されるのが欠点である。

ここでは，一歩を進めて，多数個のカテゴリー区分が存在するとき，特定区分にのみ着目するのではなく，全区分における比率構成のパターンを一括して扱い，比率構成の似ている「学区」を集めて分類することを計る。このように比率構成パターンによる分類では，単独指標による分類と異なり，手作業によるのは著しく困難であるから，適当な数値基準を用い，自動的に行なうクラスター分析の考え方を採用するのが有効である。これは単に作業の容易性のみならず，主観を排除した合理的な方法であるといえる。ここでは，層別に用いる項目の質的なカテゴリーにおける度数分布をそのまま活かすクラスター化の方法を用いる。

本稿はこの方法の有効性をテストする目的を含めている。地域を表わす指標を多種類用いて，通常の変量型のクラスター分析を行なっても，地域分類の目的は果せるわけであるが，ここではそれは行なわない。

* 名古屋市の公立中学校の場合は1～4学区が1中学校に対応する。

2. 層別の方法

学区の分類は，伝達情報量（McGill, 1954）を基準測度とする連関表のカテゴリー合併の方法を適用して行なう。

表1は，項目 x と y の2元の連関度数表である。両項目は，いわゆる名義尺度に属するものとする。項目 x の相互排反的な l 個のカテゴリーを添字 i $(i=1, 2, \cdots\cdots, l)$ で，項目 y の m 個のカテゴリーを添字 j $(j=1, \cdots\cdots, m)$ で示すとして，表1の n_{ij} は項目 x のカテゴリー i と項目 y のカテゴリー j に同時に該当する度数を表わす。また $n_{i\bullet}$, $n_{\bullet j}$ はそれぞれ項目 x, y の各カテゴリーの周辺度数，n は総度数である。

同じことを，項目 x, y のみでなく一般に三つ以上の項目について扱えば，多元連関表になる。この連関表の各細胞 $n_{ij\cdots k}$ は各項目の特定カテゴリーに同時に該当するものの度数を示す。

表1 連関度数表（2元）

		項目 y 1	2	……	j	……	m	計
項目 x	1	n_{11}	n_{12}	……	n_{1j}	……	n_{1m}	$n_{1\bullet}$
	2	n_{21}	n_{22}	……	n_{2j}	……	n_{2m}	$n_{2\bullet}$
	⋮	⋮	⋮		⋮		⋮	⋮
	i	n_{i1}	n_{i2}	……	n_{ij}	……	n_{im}	$n_{i\bullet}$
	⋮	⋮	⋮		⋮		⋮	⋮
	l	n_{l1}	n_{l2}	……	n_{lj}	……	n_{lm}	$n_{l\bullet}$
計		$n_{\bullet 1}$	$n_{\bullet 2}$	……	$n_{\bullet j}$	……	$n_{\bullet m}$	n

問題は，項目 x の l 個のカテゴリー群を，他の項目への反応模様が似ているものどうしを合併して $(k<l)$ なる k 個の群に縮小することである。

1） 最適基準

最適合併の数値基準として，ここでは伝達情報量を用い，これを最大にとどめるようなカテゴリーの合併を行なう。この基準が妥当であることは，伝達情報量がそもそも連関強度の測度で χ^2 と同義に扱えることから，明らかである。実際，他の項目への反応パターンが全く比

例的に等しい項目 x の二つのカテゴリーを合併しても，伝達情報量の損失は0である。

表1を総数 n を分母とする比率 p_{ij} の行列とし，多元連関表ならば比率 $p_{ij\cdots k}$ と考える。また各項目の周辺比率を $p_i, p_j, \cdots, p_k$ とし，さらにこれらを確率とみなすと，伝達情報量 $T(x{:}y{:}\cdots{:}z)$ は次式で与えられる。

$$T(x{:}y{:}\cdots{:}z)=U_{\max}(x,y,\cdots,z)-U(x,y,\cdots,z)$$

ここで，$U(x,y,\cdots,z)$ はいわゆる結合不確定度（joint uncertainty）で，$U_{\max}(x,y,\cdots,z)$ はその可能な最大値である。これらはまた次のように表現される。

$$U_{\max}(x,y,\cdots,z)=U(x)+U(y)+\cdots+U(z)$$

$$U(x,y,\cdots,z)=U(x)+U_x(y,\cdots,z)$$

ただし

$$U(x)=-\sum_i p_i\log p_i$$

$$U(y)=-\sum_j p_j\log p_j$$

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

$$U(z)=-\sum_k p_k\log p_k$$

$$U_x(y,\cdots,z)=-\sum_i\sum_j\cdots\sum_k p_{ij\cdots k}\log(p_{ij\cdots k}/p_i)$$

$U(x),\cdots,U(z)$ は各項目の周辺不確定度（marginal uncertainty），$U_x(y,\cdots,z)$ は項目 x に関する条件つき不確定度（conditional uncertainty）である。

以上により伝達情報量は，

$$T(x{:}y{:}\cdots{:}z)=U(y)+\cdots+U(z)-U_x(y,\cdots,z)$$

となる。項目 x のカテゴリーを合併して少数のグループにする際，$U_x(y,\cdots,z)$ に着目し，これを極力小さくするように計れば伝達情報量を大きくすることができる。

2）　カテゴリー合併の手順

項目 x のカテゴリー合併の方式には，いわゆる系統的（hierarchical）方法と非系統的な方法の二通りを区別できるであろう。

①　系統合併法

項目 x の全カテゴリー l 個が存在する状態から出発し，二つのカテゴリーのすべての組合せから最適な組を求めて合併し単独のカテゴリーにする。次に（$l-1$）個のカテゴリーについて，同様にして最適な対を合併する，という処理を反復してカテゴリー数を一つずつ減らしてゆく。

項目 x と y の2元の連関表の場合でいえば，項目 x のカテゴリー群のうち，カテゴリー a と b を合併した後の伝達情報量を $T'(x:y)$，合併前を $T(x:y)$ とすると，合併による基準の変化 $\varDelta T(x:y)$ は，

$$\varDelta T(x:y)=T(x:y)-T'(x:y)$$

$$=\sum_j p_{aj}\log\left(\frac{p_{aj}}{p_a}\right)+\sum_j p_{bj}\log\left(\frac{p_{bj}}{p_b}\right)$$

$$-\sum_j(p_{aj}+p_{bj})\log\left(\frac{p_{aj}+p_{bj}}{p_a+p_b}\right)$$

この $\varDelta T(x:y)$ は，必ず $\varDelta T(x:y)\geqq 0$ である。二つのカテゴリーを合併し，項目 x のカテゴリー数を一つ減らすごとに，伝達情報量は低減する。この低減 $\varDelta T(x:y)$ が最も0に近いカテゴリーの対 a，b を探して，合併を続ければよい。この方式では，合併の基準は，伝達情報量の損失の最小化として用いられる。

カテゴリーの逐次合併の終了基準として，はじめの伝達情報量に対するこの損失の累和を使用できる。たとえば，損失比率10％を許容限度と指定しておき，この値を越える直前で合併を終らせる。しかし，あらかじめ最終カテゴリー数を定め無条件にその数になるまでを行ないたい場合も実際には多いであろう。

この計算手続は3元以上の連関表でも全く同じで，上田（1967）の2元の例は3元以上の場合に容易に拡張できる。伝達情報量の損失 $\varDelta T(x{:}y{:}\cdots{:}z)$ は項目 y から項目 z までの各カテゴリーの交絡を単一のカテゴリーとし，それをあらためて h とすれば，

$$p_{ih}=p_{ij\cdots k}$$

$$\varDelta T(x{:}y{:}\cdots{:}z)=T(x{:}y{:}\cdots{:}z)-T'(x{:}y{:}\cdots{:}z)$$

$$=\sum_h p_{ah}\log\frac{p_{ah}}{p_a}+\sum_h p_{bh}\log\frac{p_{bh}}{p_b}$$

$$-\sum_h(p_{ah}+p_{bh})\log\left(\frac{p_{ah}+p_{bh}}{p_a+p_b}\right)$$

で，全く2元の場合と同型であり，伝達情報量の変化量に関する限り，同じ計算でよい。

系統合併法は最終的なカテゴリー合併のみでなく，その下位の構成も合せて知りたいとき，また最終的なカテゴリー数についてあらかじめ決定できないような場合に適している。結果から判断して適当なものを採択すればよい。

②　順次修正法*

初めに最終的なカテゴリー数を定めて，その条件下で，基準 $T(x:y)$ を極力大きくするカテゴリー合併を実現する手法である。

初めに仮りの合併を指定しておき，個々のカテゴリーのグループの所属を他に移すと仮定して，もしも基準 $T(x:y)$ が改善される（大きくなる）ならば所属を変更し，そうでなければ元のままにする，という操作を反復し，所属の変更による $T(x:y)$ の増加が完全に認め

*　適当な名称がない。クラスター分析では山登り法（hill-climbing method）などの呼称がある。(Friedman & Rubin, 1967)

られなくなるまで行なう。

この方式は‘真の’基準最大化を保証しないが，それに近い結果が得られると期待できる。

あるカテゴリー i が所属する合併カテゴリー群を α とし，所属変更の相手先カテゴリー群を β で表わすと i の α から β への変更による基準の変化 $\Delta T(x:y)$は，移動前を $T(x:y)$，移動後を $T'(x:y)$ として，

$$\Delta T(x:y)=T'(x:y)-T(x:y)$$
$$=-\sum_j p_{\alpha j}\log\left(\frac{p_{\alpha j}}{p_\alpha}\right)-\sum_j p_{\beta j}\log\left(\frac{p_{\beta j}}{p_\beta}\right)$$
$$+\sum_j (p_{\alpha j}-p_{ij})\log\left(\frac{p_{\alpha j}-p_{ij}}{p_\alpha-p_i}\right)$$
$$+\sum_j (p_{\beta j}+p_{ij})\log\left(\frac{p_{\beta j}+p_{ij}}{p_\beta+p_i}\right)$$

$\Delta T(x:y)$ が正値であれば，$T(x:y)$ は増大，負値であれば減小する。したがって計算では $\Delta T(x:y)$ が最大の正値となる相手先カテゴリー群を探して所属変更を行なえばよい。3元以上の伝達情報量を扱う場合も，系統合併法によるときと同様に，2項目以上の交絡カテゴリーを単一カテゴリーとすれば，上式の形で $\Delta T(x:y:\cdots:z)$ を算出することができる。

順次修正法は最終的カテゴリー数をなんらかの根拠，仮説により事前に決めなければならない。また初めに与える仮りの合併で最終結果が影響を受けるのが欠点と考えられる。この初期分類をたとえば乱数などによりランダムに行なうことは，かえって不適当であると考えられる。完全ランダム化は，初めに均質なグループを作ってしまうことに当り，収束を遅くするだけでなく，いわゆる局所的最大点（local maximum）に至る危険があると思われる。

Ⅲ 名古屋市における174学区の層別*

伝達情報量 $T(x:y)$ を基準として連関度数表をデータとする「学区」の層別を名古屋市における174学区について試算してみる。「学区」が先の項目 x に当る。

「学区」の名称と区域は昭和47年5月1日現在のものである。（図1）

1. 学区層別の資料

計算に用いた資料は昭和45年度国勢調査（45年10月1日現在）の一部である。**

層別に用いる項目（項目 y）は，次の2項目とする。

a）「最終卒業学校の種類別人口」

b）「職業別就業者数」

以下，簡単にa）を「学歴」，b）を「職業」と略称することにする。

「学歴」の国勢調査におけるカテゴリー区分は，

（1 小学・高小・新中），（2 旧青学），（3 旧中・新高），（4 短大・高専），（5 大学），（6 在学者），（7 未就学者）

の7区分であるが，実際の計算ではこのうちの「在学者」，「未就学者」を除き，「旧青学」を「旧小・高小・新中」と合せて卒業者のみの以下の4区分とした。

(1) 小学・高小・新中・旧青学

(2) 旧中・新高

(3) 短大・高専

(4) 大学

「職業」の国勢調査の区分は次の10区分である。

（1 専門的技術的職業従事者），（2 管理的職業従事者），（3 事務従事者），（4 販売従事者），（5 農林漁業作業者），（6 採鉱採石作業者），（7 運輸通信従事者），（8 技能工，生産工程作業者および単純労働者），（9 保安職業従事者），（10 サービス職業従事者）

今回の計算では，内容的に似ている区分，また該当人数の少ない区分を配慮して，1と2，4と10をそれぞれ合併すると同時に，6から9までを一括して扱うことにした。ただし，（5農林漁）は該当人数が少ないが，その特殊性を考えて独立区分とした。結局，計算に使用したカテゴリーは以下の5カテゴリーである。

(1) 専門・技術・管理的

(2) 事務

(3) 販売・サービス

(4) 農林漁

(5) 技能・労務的（6～9）

以上により学区の層別計算に使用したデータは174学区×4学歴，および174学区×5職業の2種類の連関度数表である。連関度数表の各欄はその区分の該当人数（男女計）である。ただし「職業」については20%抽出からの推計であって本来の人数ではない。

層別の計算は，「学区×学歴×職業」の3元連関表が得られないため，それぞれ2元連関表の場合として別々

* この計算は，名古屋市青少年問題協議会，家庭教育問題調査専門委員会における「親子の価値観のずれに関する研究調査」の標本抽出の検討を兼ねて行なわれた。

** 表の詳細は，資料：名古屋市編「名古屋の人口」（昭.48.2月）の262～265ページ，および298～301ページを参照。

なお，これら資料の入手および内容理解について名古屋市教育委員会社会教育課の協力を戴いた。

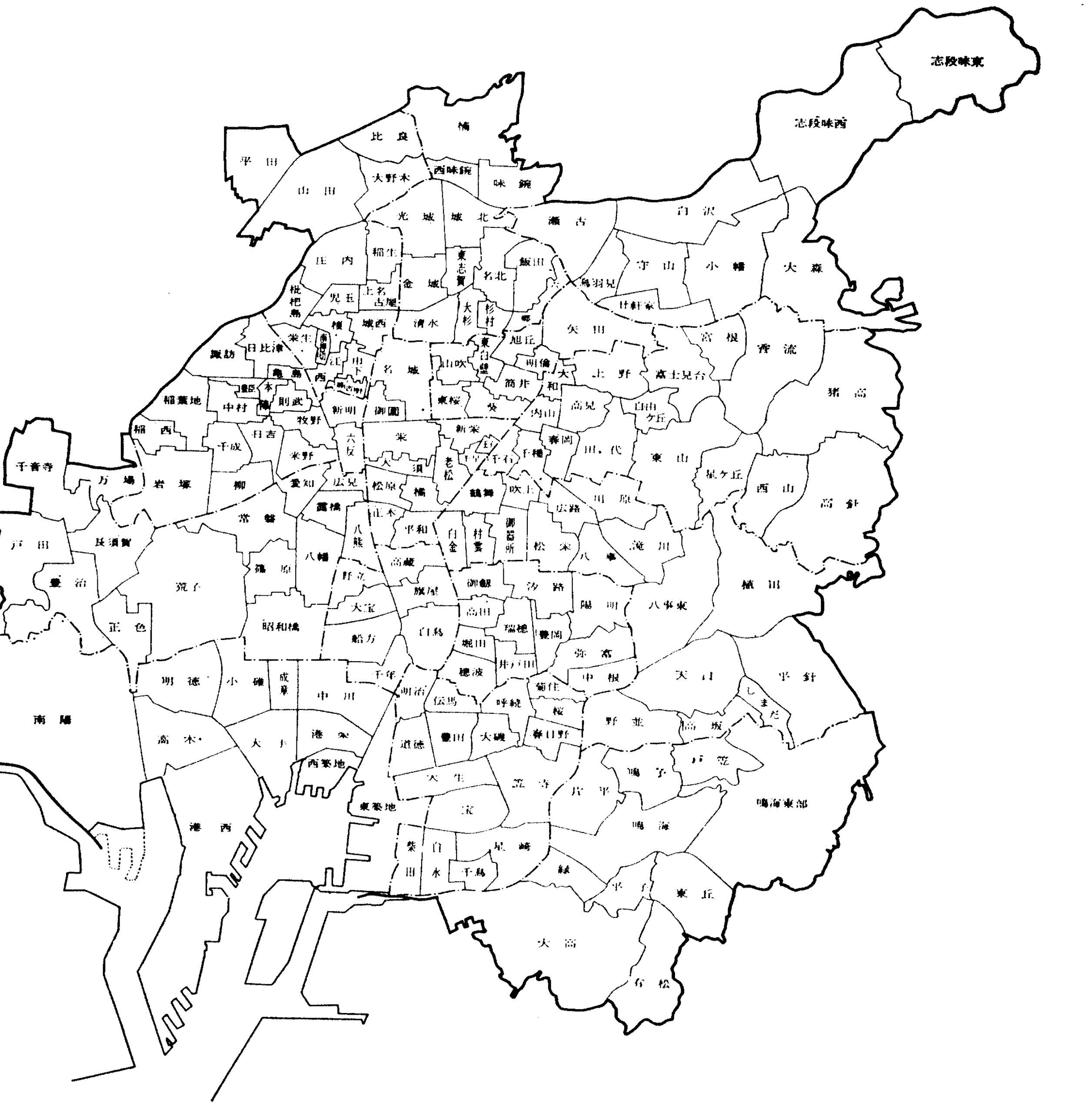

図１　名古屋市174学区の区域（昭和47年５月１日現在）
（名古屋市編「名古屋の人口」より転載）

に行なう。

2. 層別計算の結果

計算は２種類の層別データおよび２種類の手法の組合せで４通り行なわれたが，そのうち系統合併法の場合を中心に述べ，非系統的な順次修正法の結果についてはそれとの対比として触れる。

1）　「学歴別人口」による層別

174学区×４学歴のはじめの連関表（比率）における伝達情報量 $T(x:y)$ およびそれに関係する諸量は次の通りである。

$T(x:y)$	0.0508 (bit)
$U_{\max}(x, y)$	8.8140
$U(x, y)$	8.7632
$U(x)$	7.3192
$U(y)$	1.4948

表 2　学歴構成による名古屋市 174 学区の分類—系統合併法による—*

実人数　　(単位：人)

グループ ＼ 学歴		小学，高小，新中，旧青学	旧中，新高	短大，高専	大　学	グループ計 (　%)
Aグループ	A_1	8,273	18,436	4,184	7,659	38,552 (　2.72)
	A_2	15,226	23,894	6,170	9,206	54,496 (　3.84)
		23,499	42,330	10,354	16,865	93,048 (　6.56)
Bグループ	B_1	19,760	25,824	5,117	8,092	58,793 (　4.15)
	B_2	32,470	37,557	6,685	9,495	86,207 (　6.08)
		52,230	63,381	11,802	17,587	145,000 (10.22)
Cグループ	C_1	44,995	41,636	6,534	10,332	103,497 (　7.30)
	C_2	128,904	105,027	13,919	18,732	266,582 (18.80)
		173,899	146,663	20,453	29,064	370,079 (26.09)
Dグループ	D_1	117,305	81,074	8,826	11,067	218,272 (15.39)
	D_2	203,959	116,716	11,156	13,512	345,343 (24.35)
		321,264	197,790	19,982	24,579	563,615 (39.74)
Eグループ	E_1	140,308	66,733	5,109	6,201	218,351 (15.40)
	E_2	20,779	6,276	501	573	28,129 (　1.98)
		161,087	73,009	5,610	6,774	246,480 (17.38)
市　全　体		731,979	523,173	68,201	94,869	1,418,222 (100.0)

(　) は市全体に対するグループ人口のパーセント，A_1，A_2，…，E_2は途中の10グループ段階のもの

構成比　　(単位：%)

グループ ＼ 学歴		小学，高小，新中，旧青学	旧中，新高	短大，高専	大　学
Aグループ	A_1	21.46	47.82	10.85	19.87
	A_2	27.94	43.85	11.32	16.89
		25.25	45.49	11.13	18.13
Bグループ	B_1	33.61	43.92	8.70	13.76
	B_2	37.67	43.57	7.75	11.01
		36.02	43.71	8.14	12.13
Cグループ	C_1	43.47	40.23	6.31	9.98
	C_2	48.35	39.40	5.22	7.03
		46.99	39.63	5.53	7.85
Dグループ	D_1	53.74	37.14	4.04	5.07
	D_2	59.06	33.80	3.23	3.91
		57.00	35.09	3.55	4.36
Eグループ	E_1	64.26	30.56	2.34	2.84
	E_2	73.87	22.31	1.78	2.04
		65.35	29.62	2.28	2.75
市　全　体		51.61	36.89	4.81	6.69

5グループにおける情報量

$T'(X:Y)$	0.0459	(bit)
$U(X,Y)$	3.517	
$U(X)$	2.068	
$U(Y)$	1.495	
$T(X:Y)$ の損失率	9.57%	

10グループにおける情報量

$T'(X:Y)$	0.049	(bit)
$U(X,Y)$	4.373	
$U(X)$	2.926	
$U(Y)$	1.495	
$T(X:Y)$ の損失率	4.43%	

* 計算は名古屋大学大型計算機センターの FACŌM 230-60 および教育心理学教室の NEAC 1240 による。この計算と結果の整理は，名古屋大学教育心理学教室の佐々木雅子，幸村京子の両嬢に負うところが大きい。

表3　名古屋市174学区の学歴構成による分類　—系統合併法による—

Aグループ (10学区)	A_1 (3)	(千) 西山，東山，(緑) 鳴子・戸笠
	A_2 (7)	(千) 田代，自由ケ丘，富士見台，星ケ丘，(昭) 八事，滝川，(瑞) 陽明
Bグループ (16学区)	B_1 (6)	(千) 高見，猪高，(昭) 川原，八事東，(瑞) 汐路，(緑) 東丘
	B_2 (10)	(千) 上野，宮根，(東) 山吹，旭丘，(中) 名城，御園，(昭) 松栄，御器所，(瑞) 豊岡，(守) 小幡
Cグループ (45学区)	C_1 (13)	(千) 春岡，大和，高針，(東) 東白壁，東桜，(中) 老松，(昭) 鶴舞，広路，平針，高坂・しまだ，(瑞) 弥富・中根，(緑) 片平，鳴海
	C_2 (32)	(千)，千種，千石，内山，香流，(東) 明倫，矢田，筒井，葵，(北) 飯田，清水，東志賀，城北，(西) 稲生，(中村) 稲西，(中) 新栄，栄，千早，大須，松原，橘，正木 (昭) 村雲，吹上，植田，野並，(瑞) 瑞穂，(南) 菊住，桜・春日野，(守) 白沢，廿軒家，大森，守山
Dグループ (70学区)	D_1 (27)	(北) 金城，光城，味鋺・西味鋺，楠，(西) 榎，児玉，城西，山下，那古野，大野木 平田，比良，(中村) 稲葉地，中村，新明，牧野，千成，岩塚，(中) 平和，(昭) 天白，(瑞) 御劔，高田，堀田，(熱) 高蔵，旗屋，白鳥，(緑) 平子
	D_2 (43)	(北) 六郷，名北，杉村，大杉，(西) 江西，栄生，枇杷島，庄内，山田，上名古屋，(中村) 諏訪，豊臣，本陣，則武，亀島，六反，米野，日吉，柳，(昭) 白金，(瑞) 穂波，井戸田，(熱) 千年，船方，大宝，(中川) 常盤，八熊，荒子，戸田，(港) 東築地，西築地，(南) 明治，呼続，大磯，豊田，笠寺，柴田，(守) 鳥羽見，志段味西 志段味東，(緑) 鳴海東部，有松，大高
Eグループ (33学区)	E_1 (28)	(中村) 日比津，(熱) 野立，(中川) 愛知，広見，露橋，八幡，昭和橋，篠原，長須賀，千音寺，万場，(港) 港楽，大手，港西，高木，明徳，小碓，成章，中川，(南) 伝馬，道徳，大生，宝，白水，千鳥，星崎，(守) 瀬古，(緑) 緑
	E_2 (5)	(西) 南押切，(中) 王子，(中川) 正色，豊治，(港) 南陽

伝達情報量 $T(x:y)$ は小さく，もともと「学区」間で学歴構成に極端な開きがないことがいえる。

① 系統合併法による場合

全174学区から出発し，合併をくりかえし，グループを一つずつ減じて最終的な5グループに縮小した連関表は表2に掲げた通りである。（表2では10グループまでの状態と合せて表示した。）

表2から，五つのグループは高い学歴構成から低い学歴構成まで順序づけられたものであることがわかる。

名古屋市全体の平均的な学歴構成に近いのは，C，Dのグループで174学区のうち115学区がこれに属する。それに対して，A，B両グループは，所属学区数は少ないが，Aの大学卒率（18.1%）が市全体の6.7%に比べ3倍近いことが示すように極めて高学歴構成で特徴づけられる学区である。逆にEグループは，大学卒率が2.8%と市全体の半分以下で学歴構成の低い方に偏っている。

学歴の区分は，本来序列をもったものであるが，この層別方法は分類グループまで序列をつける方法ではない。あくまで名義的な区分に対する度数分布の型に従いグループをわける。したがってたとえば両極の「大学卒」と「旧小・新中卒」のみが多く，中間の少ないグループが得られてもよい。しかし，結果的にはグループに関しても学歴の序列に対応して単純な順序が成立する分類となった。

名古屋市174学区の最終的なグループ所属は表3の通りである。表中の(千)などは区名の頭文字である。

Aグループ……大学卒が20%弱を占める高学歴地域で，千種区の中心部および昭和区，瑞穂区の一部の学区が含まれる。これらは以前からの住宅地区，および市東部の新興住宅団地を抱える学区である。郊外の新興住宅地区ということができる。

Bグループ……Aグループに準ずる高学歴学区である。地理的にAグループに接しているが市の中心部側に位置するものが多く，住宅地域と商業地域が混在している学区が多い。また，いずれはAグループに移るようなごく新しい郊外住宅地域が含まれている。

Cグループ……市の中心部の商業地域の学区が多くを占める。名古屋市として古くに発展した市街地が多いのが特徴である。

Dグループ……70学区が含まれる平均的な地域で，商店，工場，住宅が多様に混りあっている。市の北，西，南部一帯に拡がっている。

Eグループ……市の南西部に位置する工場地帯に属する学区である。学歴構成の点で他の4グループに比べて低い。

全体として名古屋市の場合，これらのグループが地理的にも一つのまとまりをなす傾向がある。概して市の新興郊外部を含めた東部地区の学歴構成が高く，次いで中心の市街部，続いて北，西，南部地区となっている。

② 順次修正法による場合

学歴データにより，初めに5グループと指定して分類計算を行なった結果は，表4の通りである。表4は最終5グループの学歴構成を示す。

この計算における初期分類は174区学を次のように5グループに割り当てる方法で行なった。

(千種区，昭和区，緑区の42学区)
(瑞穂区，東区，中区の　30学区)
(守山区，北区，西区の　37学区)
(中村区，熱田区，南区の39学区)
(中川区，港区の　　　　26学区)

これは，各区の全体における比率構成をみて，似たものを合せたものである。結果的な伝達情報量 $T(x:y)$ は 0.0463 bit で，系統的な方法 (0.0459) と比べて僅かに高い。伝達情報量の損失率でいえば8.73%で，系統的な方法の9.57%に優っている。しかし，グルーピングの効率に関しては両方法に実質的な差はないといってよいだろう。

学区のグループ所属に関しては若干の異動がみられる。具体的な学区名は省略して，表5では順次修正法の5グループ（A′～E′）と系統合併法の5グループ（A～E）とがどの程度合致しているかをグループの対ごとに学区数で示した。系統合併法ではDグループが70学区

表4　学歴構成による名古屋市174学区の分類　―順次修正法―

実人数　　(単位：人)

学歴／グループ	小学，高小 新中，旧青学	旧中，新高	短大，高専	大学	グループ計 (%)
A′グループ	25,084	44,361	10,722	17,706	97,873 (6.90)
B′グループ	80,439	90,481	16,050	23,997	210,967 (14.88)
C′グループ	198,514	157,373	20,246	27,406	403,539 (28.45)
D′グループ	273,427	161,377	15,843	19,372	470,019 (33.14)
E′グループ	154,515	69,581	5,340	6,388	235,824 (16.63)
市全体	731,979	523,173	68,201	94,869	1,418,222 (100.0)

(　) は市全体に対するグループ人口のパーセント

構成比　　(単位：%)

学歴／グループ	小学，高小 新中，旧青学	旧中，新高	短大，高専	大学
A′グループ	25.63	45.33	10.96	18.09
B′グループ	38.13	42.89	7.61	11.37
C′グループ	49.19	39.00	5.02	6.79
D′グループ	58.17	34.33	3.37	4.12
E′グループ	65.52	29.51	2.26	2.71
市全体	51.61	36.89	4.81	6.69

5グループにおける情報量

T′(X：Y)…0.0463 (bit)
U(X, Y) … 3.598
U(X) ………2.150
U(Y) ………1.495
T(X：Y) の損失率 8.73%

表5　系統合併法と順次修正法による「学歴」グループの比較

系統合併 ＼ 順次修正	A′グループ (11学区)	B′グループ (24学区)	C′グループ (49学区)	D′グループ (59学区)	E′グループ (31学区)
A グループ (10 学 区)	10				
B グループ (16 学 区)	1	15			
C グループ (45 学 区)		9	36		
D グループ (70 学 区)			13	57	
E グループ (33 学 区)				2	31

（数字はグループに属する学区数）

を含む大グループであったが，順次修正法では最大のグループが59学区と少なく，その分だけ他グループに分散している。しかし，両分類方式の結果は，一部を除けばおおむね対応しているということができる。

2）「職業別人口」による層別

層別計算に入る前の連関表（174学区×5職業）の伝達情報量 $T(x:y)$ とこれに関係する諸情報量は次の通りである。

$T(x:y)$	0.0956 (bit)
$U_{\max}(x, y)$	9.2429
$U(x, y)$	9.1474
$U(x)$	7.3206
$U(y)$	1.9223

「職業」では，制約係数 $D(D=T(x:y)/U(y))$ が0.050であり，この値は「学歴」の0.034より高く，学区が学歴構成よりも職業構成の差異をよく示すことが認められる。同類の産業活動がある地域にかたまって営まれ，就業者の居住分布もそれに応じて定まるのは都市形態の大きな特徴である。「学歴」よりも産業活動と直接に結びつく「職業」が学区の差を反映するのは当然であろう。

① 系統合併法による場合

計算手順に従い，174グループ（学区）を順次縮小して5グループにした結果を表6に示す。途中段階の10グループの結果も合せて掲げてある。

最終5グループにおける伝達情報量 $T(x:y)$ の損失率は30.75%で，これは「学歴」の9.57%に比べてかなり大きい。グループ化による情報要約が粗いというべきであろう。（「学歴」での損失率にほぼ一致する「職業」のグループ数は17である。）最終5グループは，それぞれ特定の職業区分の人口が高く，よくその特徴を示しているといえる。しかし，「学歴」のように単純な一次元的順序を示すものではなく，構成比率パターン全体を同時的に扱う今回の手法の長所を結果的によく活かした分類になっている。

各グループに含まれる学区名は表7に示される。表6と表7を合せて最終5グループの特徴を挙げると

a グループ……いわゆるホワイトカラーの多い学区で「専門・技術・管理職」，「事務職」を合せると65%になる。千種区，昭和区，瑞穂区の20学区が含まれる。住宅地域，住宅団地地区に加えて，一部の商住地域がはいる。

b グループ……a グループに準ずるグループで23学区が属する。ホワイトカラーの多い住宅地域であるが，概して a グループよりもさらに郊外の地域や新興の住宅地域が多い。そのことは「農林漁」に従事する人口11.18%によっても示される。

c グループ……「販売・サービス」の41.6%が示すように名古屋市の最も中心的な商店街，または古くからの市街部を含む商業地域である。中区の大部分をはじめとする21学区がこれに属する。

d グループ……名古屋市全体の職業構成に最も近い平均的なグループで174学区の中の61学区がこれに当る。

e グループ……「技能・労務的」の56.27%や「農林漁」3.17%により特徴づけられる49学区である。e グループは，その下部グループ e_2 が示すように市の外周部の農業色をなお濃く残している学区と，市の南，西部工場地帯の学区とを含んでいる。

② 順次修正法による場合

表 6　職業構成による名古屋市 174 学区の分類　—系統合併法による—

実人数　　　　　　　　　　　　　　　　　　　　　　　　　　（単位：人）

職業 グループ		専門，技術管理	事務	販売	農林漁	技能，労務他	グループ計（%）
a グループ	a_1	14,450	11,830	13,145	100	8,120	47,645（4.70）
	a_2	18,960	19,810	22,300	175	22,705	83,950（8.27）
		33,410	31,640	35,445	275	30,825	131,595（12.97）
b グループ		21.800	27,150	28,975	1,495	47,750	127,170（12.53）
c グループ	c_1	3,525	3,085	11.465	5	3,190	21,270（2.10）
	c_2	12,060	15,335	32,460	25	24,440	84,320（8.31）
		15,585	18,420	43,925	30	27,630	105,590（10.41）
d グループ	d_1	25,130	41,980	50.770	350	117.005	235,235（23.18）
	d_2	15,095	25,730	49.365	365	66,260	156,815（15.46）
		40,225	67,710	100,135	715	183,265	392,050（38.64）
e グループ	e_1	11,460	21,625	22,215	4,600	64,860	124,760（12.30）
	e_2	970	2,105	1,825	2,800	7,050	14,750（1.45）
	e_3	7,590	17,325	19,640	785	73,400	118,740（11.70）
		20,020	41,055	43,680	8,185	145,310	258,250（25.45）
市全体		131,040	185,975	252,160	10,700	434,780	1,014,655（100.0）

（　）は市全体に対するグループ人口のパーセント，a_1，a_2，b，…，e_3 は途中の10グループ段階のもの

構成比　　　　　　　　　　　　　　　　　　　（単位：%）

職業 グループ		専門，技術管理	事務	販売サービス	農林漁	技能，労務他
a グループ	a_1	30.33	24.83	27.59	0.21	17.04
	a_2	22.58	23.60	26.56	0.21	27.05
		25.39	24.04	26.93	0.21	23.42
b グループ		17.14	21.35	22.78	1.18	37.55
c グループ	c_1	16.57	14.50	53.90	0.02	15.00
	c_2	14.30	18.19	38.50	0.03	28.98
		14.76	17.44	41.60	0.03	26.17
d グループ	d_1	10.68	17.85	21.58	0.15	49.74
	d_2	9.63	16.41	31.48	0.23	42.25
		10.26	17.27	25.54	0.18	46.75
e グループ	e_1	9.19	17.33	17.81	3.69	51.99
	e_2	6.58	14.27	12.37	18.98	47.80
	e_3	6.39	14.59	16.54	0.66	61.82
		7.75	15.90	16.91	3.17	56.27
市全体		12.91	18.33	24.85	1.05	42.85

5グループにおける情報量

$T'(X:Y)$　0.066（bit）

$U(X,Y)$　3.986

$U(X)$　2.130

$U(Y)$　1.922

$T(X:Y)$ の損失率　30.75%

10グループにおける情報量

$T'(X:Y)$　0.081（bit）

$U(X,Y)$　4.865

$U(X)$　3.023

$U(Y)$　1.922

$T(X:Y)$ の損失率　15.13%

表7　名古屋市174学区の職業構成による分類　—系統合併法による—

aグループ（20学区）	a_1（8）	（千）田代，自由ケ丘，西山，星ケ丘，東山，（昭）八事，滝川，（瑞）陽明
	a_2（12）	（千）高見，上野，富士見台，（東）山吹，旭丘，（昭）松栄，御器所，鶴舞，川原，八事東，（瑞）豊岡，汐路
bグループ（23学区）		（千）大和，宮根，香流，猪高，高針，（東）明倫，（北）清水，東志賀，（中）正木，（昭）村雲，吹上，広路，植田，平針，高坂・しまだ，野並，（瑞）弥富・中根（守山）小幡，大森，（緑）片平，鳴子・戸笠，鳴海，東丘
cグループ（21学区）	c_1（4）	（中）名城，御園，栄，大須
	c_2（17）	（千）春岡，千種，内山，（東）東白壁，葵，東桜，（西）巾下，那古野，（中村）中村，新明，六反，牧野，（中）新栄，千早，老松，松原，橘
dグループ（61学区）	d_1（33）	（東）矢田，（北）飯田，名北，大杉，金城，城北，光城，（西）児玉，枇杷島，庄内稲生，（中村）日比津，稲西，（瑞）御劔，高田，穂波，井戸田，瑞穂，（熱）旗屋船方，大宝，（中川）愛知，露橋，八幡，（港）港楽，（南）明治，呼続，菊住，桜・春日野，大磯，笠寺，（守山）鳥羽見，廿軒家
	d_2（28）	（千）千石，（東）筒井，（北）六郷，杉村，（西）榎，城西，江西，栄生，上名古屋，（中村）稲葉地，豊臣，本陣，則武，亀島，米野，日吉，千成，（中）王子，平和，（昭）白金，（瑞）堀田，（熱）高蔵，白鳥，野立，（中川）広見，八熊，正色（港）西築地
eグループ（49学区）	e_1（24）	（北）味鋺・西味鋺，楠，（西）大野木，平田，比良，（中村）諏訪，柳，岩塚，（昭）天白，（中川）常盤，荒子，戸田，千音寺，万場，（港）高木，明徳，小碓，（守山）白沢，守山，志段味東，（緑）緑，平子，有松，大高
	e_2（5）	（中川）豊治，長須賀，（港）南陽，（守山）志段味西，（緑）鳴海東部
	e_3（20）	（西）南押切，山田，（熱）千年，（中川）昭和橋，篠原，（港）東築地，大手，港西，成章，中川，（南）伝馬，豊田，道徳，大生，宝，柴田，白水，千鳥，星崎，（守）瀬古

初めに5グループを指定し，仮分類としては「学歴別」と同様に比率構成の似ているものを合せて14区を次のように配分した。

（千種区，昭和区の33学区）

（緑区，瑞穂区，守山区の28学区）

（中区，中村区，東区の37学区）

（北区，熱田区，西区の35学区）

（中川区，南区，港区の41学区）

これより出発し，分類の順次修正計算を行なった結果の最終的なグループ別職業構成を表8に示す。

結果的な伝達情報量 $T(x:y)$ は0.070 bit で系統合併法に比べてやや高い。伝達情報量の損失率においても26.7%で，系統合併法の30.75%よりやや低い。しかし，各グループに含まれる学区の内容では，系統合併法の結果と異なる。2種類の分類の対応関係を学区数で表わしたのが表9である。

「専門・技術・管理」や「事務」が多い a′グループと「販売・サービス」の多い b′グループは，系統合併法の a グループ，b グループとほぼ対応している。しかし，c′d′e′ グループと c，d，e グループは相互に入り組んだ関係になっている。

「学歴」と比べて両分類法の間の差異が大きいのは，もともとこれらのグループの学区の職業構成は市全体の平均的構成に近く相互の分離が小さいこと，損失率でみるとおり，グループ化が粗すぎるため，両方式の差が強調され易いためと思われる。

3. 層別結果の検討

1）「学歴」と「職業」による層別の比較

2種類のデータによる学区の層別を比較すると，伝達情報量 $T(x:y)$ の損失率が示すように「職業」の方が情報要約が粗くなる傾向が認められたが，いずれの場合も特徴的な5グループを抽出している。

2種類の層別は異なる資料から得られたものである

表 8　職業構成による名古屋市 174 学区の分類　—順次修正法—

実人数　　(単位：人)

職業／グループ	専門，技術管理	事務	販売	農林漁	技能，労務他	グループ計 (%)
a′ グループ	37,635	36,430	38,970	645	36,345	150,025 (14.79)
b′ グループ	42,885	63,270	88,085	1,050	146,050	341,340 (33.64)
c′ グループ	17,925	22,060	51,505	40	35,090	126,620 (12.48)
d′ グループ	21,390	43,455	53,180	1,475	154,650	274,150 (27.02)
e′ グループ	11,205	20,760	20,420	7,490	62,645	122,520 (12.08)
市全体	131,040	185,975	252,160	10,700	434,780	1,014,655 (100.0)

(　) は市全体に対するグループ人口のパーセント

構成比　　(単位：%)

職業／グループ	専門，技術管理	事務	販売	農林漁	技能，労務他
a′ グループ	25.09	24.28	25.98	0.43	24.23
b′ グループ	12.56	18.54	25.81	0.31	42.79
c′ グループ	14.16	17.42	40.68	0.03	27.71
d′ グループ	7.80	15.85	19.40	0.54	56.41
e′ グループ	9.15	16.94	16.67	6.11	51.13
市全体	12.91	18.33	24.85	1.05	42.85

5 グループにおける情報量

T′(X : Y)… 0.070 (bit)

U(X, Y)…… 4.042

U(X)……… 2.190

U(Y)……… 1.922

T(X : Y) の損失率　26.7%

表 9　系統合併法と順次修正法による「職業」グループの比較

	a′ グループ (24学区)	b′ グループ (54学区)	c′ グループ (25学区)	d′ グループ (43学区)	e′ グループ (28学区)
a グループ (20学区)	20				
b グループ (23学区)	4	16	1		2
c グループ (21学区)			21		
d グループ (61学区)		38	3	20	
e グループ (49学区)				23	26

数字はグループに属する学区数

が，各グループの所属学区を比べるとき，比較的よく似た構造を示している。表10は「学歴」と「職業」の各 5 グループに関して個々の学区がどう対応しているかを系統合併法の場合についてみたものである。表10の数字はグループの各組合せに属する学区の数を示す。これから「学歴」による層別と「職業」による層別の両者の関連が高いことが認められる。これはもともと「職業」と「学歴」との相関が高いという理由から当然といえる。

結果の利用上の見地からいえば，層別は調査・実験を有効にする補助手段であるから，どちらか一方を用いればよいであろう。あるいは両結果を合せて適当な層別修正を視察で行なうのもよいと思われる。

ところで以上のような地域分類に関して考慮しなければならない要因として「年令」があげられる。一般に，年令の若い層に高学歴が多く，高年令層では学歴が概して低い。したがって，たとえば既述の結果は，年令的に若い層が新興郊外部に，年令的に高い層が旧市街部に多く居住するための随伴現象であるということも考えられないわけではない。

表11，表12は系統合併法による「学歴」，「職業」の学区グループの年令階級分布を比較したものである。「学歴」ではAグループで30〜40代が，C，Dグループで

表10　「学歴」と「職業」による分類の比較（系統合併法による場合）

学歴別＼職業別		a グループ		b グループ	c グループ		d グループ		e グループ		
		a₁	a₂		c₁	c₂	d₁	d₂	e₁	e₂	e₃
Aグループ	A₁	2		1							
	A₂	6	1								
Bグループ	B₁		4	2							
	B₂		6	2	2						
Cグループ	C₁		1	8		4					
	C₂			10	2	7	9	2	2		
Dグループ	D₁					5	6	8	8		
	D₂					1	13	14	8	2	5
Eグループ	E₁						5	2	6	1	14
	E₂							2		2	1

（数字は学区の数）

表11　「学歴」の５グループの年令別構成(系統合併法)
（単位：%）

年令区分＼グループ	A	B	C	D	E
0 ～ 14才	23.9	21.9	22.4	22.6	25.1
15 ～ 19才	7.6	8.9	8.0	8.8	8.6
20 ～ 29才	22.2	24.8	24.5	23.2	22.3
30 ～ 39才	18.5	16.6	17.3	16.7	17.5
40 ～ 49才	13.2	12.3	11.7	11.9	11.8
50 ～ 64才	9.7	10.5	10.9	11.2	10.2
65 才 以 上	4.8	5.1	5.3	5.4	4.4

表12　「職業」の５グループの年令別構成(系統合併法)
（単位：%）

年令区分＼グループ	a	b	c	d	e
0 ～ 14才	21.4	25.6	17.3	21.7	26.0
15 ～ 19才	8.7	7.2	9.2	8.7	8.4
20 ～ 29才	24.1	23.2	25.5	23.4	22.6
30 ～ 39才	16.4	19.3	14.7	16.4	18.4
40 ～ 49才	13.0	11.2	12.2	12.1	11.3
50 ～ 64才	10.7	9.2	13.8	11.8	9.2
65 才 以 上	5.6	4.3	7.1	5.7	4.0

は50代以上が他と比べてやや多い傾向が認められる。「職業」でみると，c，dグループでは50代以上が，a，b，eグループでは30～40代がやや多い傾向がある。しかし，年令構成の差異が学歴や職業の構成比を強く規定しているとは認めがたい。

その他では，一般に「学歴」，「職業」の両分類結果に‘住宅団地’のあり・なしがよく効く傾向が認められた。学区のように比較的狭い地域範囲では，一つの住宅団地の存在が全体の構成に大きく影響するのは当然であろう。このことは，層別データの取得時期と調査・実験の時期の間隔が大きいとき，結果の利用において注意すべきことを示している。もし，その期間に住宅団地出現の事実があれば，層別の修正が必要となる。ただし，修正の方向は一概にいえず，住宅団地の種類によりそれぞれ入居資格条件（収入など）が異なるから，それらを配慮して行なうことになる。

2）層別の方法について

系統合併法と非系統的な順次修正法の両分類における最適化基準（伝達情報量の損失最小）の達成値には大きな差異が認められなかった。

順次修正法では分類の下位構成が得られない代りに基準の最適化において優ると期待されたが，結果的には「学歴」ではほとんど変らず，「職業」でもその差異は小さかった。これについては，そもそも最高値がそこに存

在するのか，あるいは順次修正方式（山登り方式）自体にあるのか，あるいは最初の仮分類に帰因する局所的極大点であるのか，その理由の断定は困難である。

ここでは仮分類の点について検討を加えておく。今回の計算では仮分類として名古屋市14区を似ている区どうしを5グループに分割して与えた。比較のために，学区を配列順に交代的に分配した場合（ランダムとみなされる）と，系統合併法の最終5グループをあらためて初期分類とした場合について試算した結果は，それぞれ収束までの反復演算時間は異なったが，分類の基準（伝達情報量）に関しては「学区」，「職業」ともに全く同じであった。分類の学区内容についても「職業」では全く同じであったが，「学歴」では系統合併法の最終5グループを初期分類とした場合が他の二つとわずかな相違（グループ一致率で約94%）を示した。結局，仮分類の影響は認められなかった。

系統的と非系統的な両方式で分類された学区の内容をみると，「学歴」データではほぼ同様といってよく，両方式のいずれを採るかはさして重要事ではないと思われる。しかし，「職業」のように伝達情報量の損失が比較的大きい場合には両方式の差異が大きくなるのが認められた。このような場合に最終的なグループ化としてどちらの方式の結果を採択すべきかが問題になる。順次修正法では初めにグループ数を仮定し仮分類が必要であることを考慮すると，系統合併法の完全に客観的な結果を採用するのが適当だと考えられる。

順次修正法については，今回は仮分類の影響がほとんど認められなかったが，なお多くの検討が必要のように思われる。

両方法の電算機計算時間の比較では，順次修正法は系統合併法に比べて大幅に短かかった。もちろん，これはプログラミングの巧拙が関係し，仮分類の与え方が関与してのことであるが，演算時間の点では順次修正法が系統合併法に対して有利であることが一般に認められそうである。

Ⅳ　おわりに

1）　本稿では学校を層別することを目標に，国勢調査における「学歴別」および「職業別」人口の資料を用いて“学区”の層別を試みた。調査，実験の対象地域の全学校に渡って共通に与えられる資料はなかなか存在しない現状にあっては，デモグラフィックなデータはなんといっても利用し易い資料である。この点はこのような目標の層別に対して長所である。

しかし，果して学歴別や職業別人口構成による「学区」の層別が，単に市や区を更に細分した地域特徴の記述にとどまらず，学校における教育的な調査，実験に関して高い層別効果をあげるかどうかについては，今後の検討を待たなければならない。「学歴別」や「職業別」が，こどもへの教育態度などに関して果して異なるのか，といった問題は，たとえば教育態度調査，教育世論調査の結果から確かめることができるが，逆に，数多くのこどもについての調査，実験の成果を親のデモグラフィックな属性の観点で整理することも有効であろう。この逆の方向の努力が現状では不足しているように思われる。このような情報のフィード・バックが盛んになされて真に有効な層別資料も見出されるし，適切な層別計画も可能となって研究の質をあげることができる。

2）　学区の層別にあたり，伝達情報量を基準とする連関表の区分合併の方法を用いたが，ここでは「学区×学歴」，「学区×職業」のそそれぞれ2元の場合として計算した。一般的には3元以上の連関として行なえば，より内容豊かな層別が得られると期待できる。しかし，3元以上の資料は実際には入手し難いこと，概して計算量が増大することが3元以上に拡張するときの問題点である。

多元連関表が得られなくてもその周辺連関のデータは利用可能な場合が多い。たとえば，3元データで項目 x と y，項目 x と z について連関表が存在するという場合である。このとき，項目 y と項目 z の偏連関が小さいとあらかじめ仮定できれば，次のようにして，完全3元連関表使用の計算と大差ない結果が得られる。

完全な3元連関表での伝達情報量 $T(x:y:z)$ は，

$$T(x:y:z)=U(y)+U(z)-U_x(y,\ z)$$

であるが，もしも項目 y と項目 z が，項目 x の各区分の中では独立とすると，

$$p_{ijk}=p_{ij}\cdot p_{ik}/p_i$$

ただし，p_{ijk} は3元連関表での比率。p_{ij}，p_{ik} は項目 x と y，x と z における周辺比率。p_i は項目 x に関する周辺比率。

$$\begin{aligned} U_x(y,\ z) &= -\sum_i\sum_j\sum_k p_{ijk}\log(p_{ijk}/p_i) \\ &= -\sum_i\sum_j p_{ij}\log(p_{ij}/p_i)-\sum_i\sum_k p_{ik}\log(p_{ik}/p_i) \\ &= U_x(y)+U_x(z) \end{aligned}$$

となる。したがって，このとき項目 y と項目 z を区別せず，あたかも単独項目のように横に並べた2元連関表を考えれば（ただし $\sum_j p_{ij}=p_i$，$\sum_k p_{ik}=p_i$ としておく），Ⅱで述べた2元の計算手順に帰着する。

本稿では連関表の区分合併の方法を「学区」の層別に関連して述べてきた。しかし，単に学区の分類と限らず，一般の連関表に適用できる方法である。計算自体は簡単でプログラミングも容易である。また更に手を加えれば利用内容を豊かにすることもできる。たとえばアルゴリズムの一部を若干変更すると次のような条件つきカテゴリー合併に使用できる。

(1) 無差別に相手カテゴリーと合併できるのではなく，あらかじめ指定した許される範囲内で合併を進める方式。たとえば地域の分類を無差別に行なうと，遠く離れた地域が同じグループにまとめられるケースが現れる。それで差し支えない場合はよいが，交通事情の調査で，地理的な位置も同時に考慮したいような場合には困ることになる。このときは地域の合併の際，隣接の（境界を接している）地域としか合併しないという条件を付して行なえばよい。

(2) 項目 x のカテゴリーが一次元的な序列をなす場合にその方向に沿ってのみ合併を行なう方式。これは(1)と同類であるが，たとえば項目 x が細かい年令段階区分からなり，これをもっと粗い区分でまとめようとする場合に当る。この場合，合併は必らずその年令段階の直前，直後の段階に対して行なうような条件を付けたアルゴリズムにすればよい。

文　　献

Friedman, H. P. and Rubin, J. 1967 On some invariant criteria for grouping data. *Journal of the American Statistical Association*, **62**, 1159—1178.

McGill, W. J. 1954 Multivariate information transmission. *Psychometrika*, **19**, 97—116.

上田尚一　1967 統計調査における情報理論の応用(2) ——データの要約またはグループわけ—— 総理府統計局研究彙報，17号，55—71.

ON THE STRATIFICATION OF "SCHOOL ZONES" FOR EDUCATIONAL SURVEY

Kinji MIZUNO

The purposes of this paper were to examine on the effectiveness of the following clustering procedure for contingency table, and to apply this procedure to the stratification of 174 school zones in Nagoya City based on the Census data.

The procedure is summarized as follows :

Concerning the frequency distribution of item Y (table-head item), the mutually similar categories of item X (table-side item) are combined to the same group under the condition of minimizing the loss of transmitted information $T(x:y)$. Consequently, initial contingency table is reduced to a smaller one, which has several distinct groups combining similar categories of item X. This procedure can be easily extended to the case of multiway contingency table.

Two algorithms, hierarchical and non-hierarchical (so-called hill-climbing), were examined against two kinds of Census data (1970), the educational distribution and the occupational one of Nagoya habitants, by school zone. Each school zone corresponds to each category of item X.

The results of the computation with educational data showed five characteristic groups ordered one-dimensionally from 'high-educated' to 'low-educated'. Similarly, five groups were distinguished in the occupational case. These were 'residential zones in the surburbs', 'commercial zones', 'industrial zones' and so on. In both educational and occupational data, the achieved criteria $T(x:y)$ by two algorithms were not so much different from each other. Larger difference between the two algorithms was recognized in the contents of school zones in each group at the occupational case, in which the loss of transmitted information was greater than the other.

On the whole, it was found that both these clustering methods and the data of this kind might be useful for the stratification of school zones. However, whether such stratification is actually valid for various educational surveys or not is in need of further studies.