AIC および双対尺度法によるクロス表の カテゴリ統合分析法

小 木 元*

A METHOD OF CROSS TABULATED DATA ANALYSIS BY
CATEGORY POOLING USING AIC AND DUAL SCALING

Hajime OGI

In this paper, first, similarity between categories is formulated using AIC. Then the relation between dual scaling solution and category pooling process conducted by AIC is discussed.

Second, a method analyzing cross tabulated data is proposed, which is useful for handling multicategorized variables. In this method, categories are pooled into some groups so as to minimize AIC of cross table composed by the pooled category groups. Then, some tables that are not significantly different from the MAICE table are selected, and several inductive results are inferred and confirmed by comparing dual scaling solutions on the selected tables and by analyzing the category pooling process. This confirmation procedure also provides deeper understanding of dual scaling solutions.

Applying the method to two cross tables of vehicle possession in Tukuba Science City, it is shown that proposed method is practically efficient for drawing conclusions, even in the cases that several categories have rather small samples and the table apparently has no significance.

1. は じ め に

クロス表を一方の変数によってもう一方の変数を説明する目的で分析することを考える。説明変数が多いカテゴリにわかれている場合、各カテゴリにおける目的変数に対する度数分布をそのカテゴリの反応と呼ぶことにすれば、反応構成比の類似したカテゴリは目的変数に照らして区別が少ないと考えられる。このような反応構成比の類似したカテゴリを統合してクロス表を再構成すれば、見易い簡潔な表となる。

反応構成比の類似関係を変数間の統計的独立性を用いて定義すれば、いくつかのカテゴリにおける反応をとり出してクロス表の部分表を作成し、部分表の独立性を検定し、その結果が有意でないとき、すなわち、反応が部分表の説明変数側のカテゴリ区分に依らないとき、これらのカテゴリは類似性を有しているといえる。

このような変数間の相互依存性の指標として赤池 (1976) による AIC (Akaike's information criterion) を用いることを考えると、類似性を有するカテゴリグループでは、

$$\text{AIC(部分表)} > 0 \quad (1.1)$$

となる。

ところで、このようなカテゴリ間の類似関係は同値関係ではない。すなわち、類似関係を $\sim$ で定義すると、

$$(C_i \sim C_j) \text{ かつ } (C_j \sim C_k) \rightarrow (C_i \sim C_k)$$

等の場合があり得る。従って、カテゴリの統合を考えるためには、類似関係にあるカテゴリグループのうちどれを統合するかについて別の基準を考える必要がある。

カテゴリの統合によってクロス表を再構成したとき、新しいクロス表 (以下、統合表) では元のクロス表 (以下、原表) より変数間の相互依存度が高まること期待される。従って、統合するカテゴリグループ

* 工業技術院製品科学研究所 (Industrial Products Research Institute)

の選択を、統合表における変数間の相互依存度を基準として行えば、統合表の見易さ、簡潔さについて統計的な意味を与えることができる。

通例クロス表における独立性の検定に用いられる適合度検定 (Pearson の χ^2 , X^2 統計量) や、尤度比 χ^2 (Y^2 統計量) は、統合の仕方によって自由度が異なる場合の相互依存度を比較するには適切ではないが、AIC は異なる自由度を持つクロス表の変数間の相互依存度を直接比較することができる。

さらに、AIC には、

$$\text{AIC (統合表)} = \text{AIC (原表)} - \sum \text{AIC (部分表)} \quad (1.2)$$

という関係があり、上述のカテゴリ間の類似関係によるカテゴリ統合が統合表の変数間の相互依存度を高めるという期待に直接応える基準である。また、この関係から、統合表の AIC は、類似関係にあるカテゴリグループ内の相違点に関する情報を除去したあとのカテゴリグループ間の相違点に関する情報量を示しているとみなすことができる。

しかし、AIC によるモデル選択では、AIC の最も小さいモデル (MAICE: 最小 AIC の推定値, Minimum AIC Estimation の略) が複数個存在することが考えられ、さらに、K-L 情報量の推定量としての推定誤差の範囲内で MAICE に近いモデルが複数個存在する場合がある。これらの場合には、AIC の面からはどのモデルがより良いとも言い難い。すなわち、カテゴリ間の類似性及び相違点について、MAICE 表のカテゴリ統合が示すものとは別様のものが同じ程度に良いということになる。この場合には、これらの統合表に共通に含まれている特徴が最も重要な特徴であるといえるであろう。そこで、本論文では、複数の統合表を同時に選択し、これらを並べて比較検討を行う一連の方法を提案する。

クロス表のカテゴリ間の相互依存関係を分析する方法としては、この他に数量化による方法がある。本論文ではまず数量化法とカテゴリ統合の関係の検討を行い、複数の統合表を比較する一方法としてこれを利用する。

AIC を指標とするカテゴリ統合の評価は、坂元・石黒・北川 (1983)、坂元 (1985) によって研究され、MAICE モデルが予測を行うために一番有効なモデルであることが示された。しかし、単一のモデルから予測を行うのではなく、AIC によって棄却されない複数

のモデルを比較分析することによって効果的な予測が行われること、さらに、AIC に上記のような類似関係の指標としての側面があり、カテゴリ統合の効果を分析することがそれ自体で要因分析に寄与する情報を提供することについてはこれまで十分に検討されているとはいえない。また、坂元はカテゴリ統合を数量化法等を行うさいにあらかじめ施しておく処理としているが、前処理としてのカテゴリ統合が数量化法の処理結果にどのような影響を及ぼすものか、そのような前処理が必要とされるのは数量化法にとってどのような要件によるものかについては明らかにしていない。

本論文では、これらの点に焦点をあて、分析例を通してカテゴリ統合によるクロス表分析法の検討を行う。なお、本論文で数量化法を同値な方法である Nishisato (1980)、西里 (1982) の双対尺度法の形式で扱ったのは、双対尺度法では、カテゴリ統合の特殊問題として、順位性カテゴリを取り扱うために SDM 法がすでに提起されていたことによる。

2. 双対尺度法とカテゴリ統合

本節では双対尺度法の記法として西里 (1982) に従うことにする。

双対尺度法と変数間の相互依存性との関係として、

$$X^2 = f_i \text{trace}(C_1) = f_i \sum \eta_k^2$$

という関係がある。 $SS_i = f_i$ であるから、 X^2 統計量は、双対尺度法において特徴数量が説明する個体間平方和の全ての解についての総和に等しい。

データ行列 F が独立に近く、 $X^2 \ll \chi_{DF}^2$ であるとき、このクロス表に双対尺度法をあてはめても、個体間の特徴を十分に表現する特徴数量は得られない。双対尺度法によって個体—カテゴリ間の相互依存の様態を分析するさい、個体—カテゴリ間の相互依存の度が低ければ、もとより有意な解は得られない。個体間平方和が 0 である解 $\mathbf{x}_0 = \mathbf{1}$ が無意味であるのと同様に、解全体の個体間平方和の総和を示す X^2 統計量が分析自体の有意味度を示す。

しかし、 $X^2 \ll \chi_{DF}^2$ であることは、データ行列が有効な情報を全く含んでいないこととは異なる。データ行列が含む情報の大きさは、自由度によらず X^2 それ自体の大きさに対応するからである。それでは、データ行列が含む有効な情報のみを抽出して、それを処理するような分析はできないであろうか。この問題は前節に述べたカテゴリ統合によって応えることができるで

あろう。

カテゴリ統合は前節では AIC によって定義され、AIC とカテゴリ統合の間には (1.2) 式の関係があった。 X^2 統計量については、この式のような原表と統合表、部分表の関係を定式化することはできないが、部分表において個体—カテゴリが完全に相互独立であれば、統合による X^2 の変化は 0 となり、 X^2 は減少せず自由度だけが減少することになる。また、 X^2 と Y^2 は共に漸近的に χ^2 分布に従う統計量で、原表が相互独立に近い場合には大きな差がないこと、 Y^2 については (1.2) 式と同様な関係が成立することを考慮すると、原表のなかで特に相互独立に近い部分についてカテゴリ統合を行うことはやはり個体—カテゴリ（グループ）間の相互依存度を高めることになる。

注意すべきことは、カテゴリの統合は X^2 を増やす操作ではないことである。カテゴリ統合は X^2 や Y^2 を減少させるが、それ以上に自由度を減少させる可能性があるということである。このことから、AIC を小さくするカテゴリ統合は、データ行列から弁別効果の低いカテゴリ区分についての情報を捨棄する操作であるとみなすことができる。

カテゴリ統合による双対尺度解についてもう少し検討するために、カテゴリ統合操作行列 H を以下のよう定義する。

$$H = (h_{ij}) \quad i=1, 2, \dots, J, \quad j=1, 2, \dots, J'$$

ただし、 J は原表のカテゴリ数、 J' は統合後のカテゴリ（グループ）数とし、 $J' \leq J$ であり、カテゴリ i が統合によってグループ j に合併されるとき $h_{ij}=1$ とし、それ以外のときは $h_{ij}=0$ とする。

以下、統合表における双対尺度法についての行列、ベクトルその他に $\hat{}$ をつけて示すことにする。

カテゴリ統合によって、データ行列は $\hat{F} = FH$ 、列の周辺度数の対角行列は $\hat{D} = H'DH$ となる。さらに無意味解 $\hat{x}_0 = \mathbf{1}$ について、 $\mathbf{x}_0 = H\hat{x}_0$ となるから、固有方程式の行列は $\hat{C}_0 = BC_0B'$ 、 $\hat{C}_1 = BC_1B'$ となる。ただし $B = (H'DH)^{-1/2}H'D^{1/2}$ とする。

B については $BB' = I$ で、 $B'B$ の固有値は J' 個の 1 以外は 0 である。 C_1 および $B'B$ が共に対称であることを考慮すると、行列の絶対値最大の固有値を表す関数を $\lambda_{\max}(\cdot)$ として

$$\begin{aligned} \hat{\eta}_1^2 &= \lambda_{\max}(\hat{C}_1) \leq \lambda_{\max}(B'BC_1) \\ &\leq \lambda_{\max}(B'B)\lambda_{\max}(C_1) = \eta_1^2 \end{aligned}$$

従って、 $\hat{\eta}_1^2 \leq \eta_1^2$ となって、カテゴリ統合によって、第 1 固有値が大きくなることはない。

双対尺度法で得られた第 1 解の有意性は

$$\chi_{I+J-3}^2 \sim -[f_i - 1 - (I+J-1)/2] \log(1 - \eta_i^2)$$

によって検定される。カテゴリ統合を行うと、一般には固有値の減少効果が右边を小さくする。しかし、自由度の減少は右边には $2f_i$ に対する相対的な減少としてきてくるのに比較して、左边では著しい効果がある。従って、 f_i が充分大きく、かつ、カテゴリ統合による固有値の減少が僅かであれば、個体間、カテゴリ間の弁別性はより高くなるものと考えられる。

最大固有値問題 $(\hat{C}_1 - \hat{\eta}_1^2 I)\hat{w}_1 = 0$ は、

$$B(C_1 - \eta_1^2 I)B'\hat{w}_1 = 0$$

であり、原表の第 1 解との関係は次のように考えることができる。

特殊な場合として、 H による統合操作が原表の第 1 解と全く矛盾しない場合、 $B'Bw_1 = w_1$ となり、

$$B'\hat{w}_1 = w_1, \quad \hat{w}_1 = Bw_1, \quad \hat{\eta}_1^2 = \eta_1^2$$

が導ける。これは $w_1 = D^{1/2}H\alpha$ 、すなわち $x_1 = H\alpha$ と書き表せる場合である。統合されるカテゴリグループのそれぞれのカテゴリが第 1 解において同じ値をもつことに相当する。このとき第 1 解に対応する固有値は変化せず、解 x_1 は、 $\hat{x}_1$ の重み付き平均となる。

とくに、統合操作 H が部分表において完全に独立なカテゴリをグループにまとめる操作であるときには H はどの解とも矛盾しない。この場合には、第 2 番目以降の解についても同様になり、 X^2 統計量は変化しない。

もう一つの特種な場合として、 H が第 1 解と全く矛盾する場合がある。このとき、 $Bw_1 = 0$ 、すなわち $H'D^{1/2}w_1 = 0$ で、 $\hat{C}_1$ は C_1 の w_1 に対応する成分をもたない。統合表の第 1 解は原表の第 1 解とは直交するベクトルとなる。

一般の場合には、 $Bw_1 \neq 0$ 、しかし、 $B'Bw_1 \neq w_1$ となる。 $C_1 = W\Lambda W'$ と書き表したとき、

$$B'\hat{w}_1 = W\beta, \quad \hat{w}_1 = BW\beta, \quad \beta = W'B'\hat{w}_1$$

と表せる。このとき β が W の中の w_1 に対応する成分のみをもてば第 1 の場合になり、 w_1 に対応する成分が 0 であれば第 2 の場合になる。一般的には、統合操作 H が部分表において完全には独立でないカテゴリをグループにまとめる操作であるとき、統合される

カテゴリグループが完全独立に近ければ第1の場合に近い解が得られ、そうでなければ第1の場合から大きく外れた解が得られるものと予想される。

以上、変数間の相互依存性に基づくカテゴリ統合と双対尺度法の関係について、 X^2 、固有値および解の変動の面から検討を行った。 X^2 や固有値の変動からみて、カテゴリ統合は、双対尺度法の意味でも原表の有効な情報のみを抽出する効果的な方法となり得る。次節以降では双対尺度法を利用して、AICによって選択された複数の統合表と原表の比較、統合表相互の比較を行う。

3. カテゴリ統合による分析法

第1節で反応構成比の類似したカテゴリを統合するという観点から、部分表として個体—カテゴリ間が相互独立であるもののみを統合することを示した。しかし、この基準だけでは組み合わせ論的に多数の統合表が残り、分析法として効率がよくない。そこで、本論文ではカテゴリ統合による統合表の選択基準として、第1節、第2節をふまえてさらに3つのカテゴリ統合表選択基準を用いる。

- (1) X^2 , Y^2 , AIC に照らして、統合表における個体—カテゴリグループ間が相互依存関係にある。
- (2) 双対尺度分析によって有意な解を与える。
- (3) AIC が最も小さい統合表 (MAICE 表) であるか、または MAICE 表と有意な差がない。

第1の基準は、個体とカテゴリ (グループ) の統計的な相互依存関係が確実な統合表のみをその相互関係

分析のための基礎とすることを示し、カテゴリ統合による分析の結果について一つの統計的な保証を約束するための基準であり、第2の基準は、個体—カテゴリ間の相互依存関係の分析手段として用いる双対尺度法の解の有効性を保証するものである。また、第3の基準は、AIC がクロス表の持つ情報量を反映していることをふまえて、統合表のなかで最も情報量の多い表のみを分析の基礎とすることを示す。

これらの基準によって選択された統合表を AIC の小さい順、すなわち変数間の相互依存度の大きい順に並べて比較分析を行う。

以下、個体—カテゴリ関係の具体的な分析例のなかで、カテゴリ統合を行うことによる分析結果、複数の統合表を比較することによる分析結果を順次示すことによって、カテゴリ統合による分析法を具体的に提示し、その効果を説明する。

分析例として、筑波研究学園都市周辺居住世帯における乗用車および軽乗用車の保有台数 (昭和52年アンケート調査による) を居住地域から説明する問題を考えてみる。以下に示すように分析例のクロス表は乗用車では変数間の相互依存度はきわめて高く、軽乗用車では変数間の相互依存度はきわめて低い。

[分析例1]

表1aに同地域における乗用車の保有台数分布を示す。同表では、 $AIC = -26.94$, $X^2 = 127.96$, $Y^2 = 62.94$ ($\chi^2_{18}(0.5\%) = 37.16$) で、居住地域と乗用車保有台数の相互依存度はきわめて高いといえる。

各居住地域における反応構成比 (表1b) をみると、

表1. 居住地域に対する世帯の乗用車保有台数分布 (原表)

a. 原表

居住地域	A.	B.	C.	D.	E.	F.	G.	H.	I.	J.	合計
乗 用 車	0台	28	26	41	34	2	20	25	1	0	178
	1台	106	70	102	42	4	45	24	6	2	405
	2台	1	2	5	1	0	0	0	3	3	16
	合計	135	98	148	77	6	65	49	10	5	599

b. 原表の反応構成比 (単位%)

居住地域	A.	B.	C.	D.	E.	F.	G.	H.	I.	J.	合計
乗 用 車	0台	20.7	26.5	27.7	44.2	33.3	30.8	51.0	10.0	0.0	16.7
	1台	78.5	71.4	68.9	54.5	66.7	69.2	49.0	60.0	40.0	66.7
	2台	0.7	2.0	3.4	1.3	0.0	0.0	0.0	30.0	60.0	16.7
	合計	20.7	26.5	27.7	44.2	33.3	30.8	51.0	10.0	0.0	16.7

—(分析結果 1-1)—

全世帯の 2/3 が乗用車を 1 台保有し、一方、2 台保有の世帯は 3% 以下である。居住地域別にみると、1 台保有世帯は A, B, C の地域で多く、I, G, D の地域で少ない。2 台保有は I, H, J の地域で顕著に多く、E, F, G の地域では全くない。乗用車を保有しない世帯は G, D の地域で多く、I, H, J, A で少ない。

しかし、これらの観察は各居住地域のサンプル数の相違を無視したもので、I, E, J, H などのサンプル数の少ない地域について、A, B, C などの地域と同等の精度で結論を出すことには無理がある。

同様のことが双対尺度法の解 (図 1: 有意) についてもいえる。第 1 軸は 2 台保有世帯を 0~1 台保有世帯と峻別し、第 2 軸は 1 台保有世帯をさらに区分している。これを居住地域についてみると、

—(分析結果 1-2)—

I は H や J とともに遠く離れ、全体の中で特異な地域として示されている。H や J もその他の地域とは距離がある。第 1 軸の主要な意味はこれら 3 地域相互の区別およびその他の地域との区別にあり、これに比べると A~G における差はわずかで第 2 軸によって微妙に区別されている。

しかし、H, I, J の地域はサンプル数も少なく、クロス表の主要部がむしろそれ以外の地域にあることは困惑を招く。図示された顕著な特徴だけでなく、図では僅かにしか表現されていない特徴にもっと注目しなければならない。

3 つのカテゴリ統合表選択基準により選択されたカテゴリ統合の結果を AIC および統合表と双対尺度解の統計情報とともに表 2a に示す。ここで No. 0 は原表を示す。原表において相互依存度が充分高い場合、第 1 の基準、第 2 の基準は容易に満たされ、選択は主に第 3 の基準 (AIC の有意差を 1.0 としている) によって行われている。表に示された統計情報はみな 0.5% 水準で有意である。

最初に MAICE 表 (統合表 No. 1: 表 3a) を分析しよう。MAICE 表では、原表から得た双対尺度解での距離が大きい H, I, J は同じグループに分類され、距離的に近い A, B, ……、G がいくつかのグループに分類された。このことは AIC による統合表の選択法は、双対尺度解を解釈する上で解空間における距離をカテ

図 1. 乗用車保有についての双対尺度分析による解 (原表)

- : 居住地域 (A~J)
- : 乗用車保有台数 (0~2 台)

ゴリ間の dissimilarity とみなす分析法とは全く異なることを示している。

MAICE 表における反応構成比は表 3b のようになる。表 3b によると、

—(分析結果 1-3)—

1 台保有世帯は AEF で多く、DG, HIJ で少ない。
2 台保有世帯は HIJ で多く、AEF, DG で少ない。
0 台保有世帯は DG で多く、HIJ で少ない。BC は同調査における平均的な保有台数分布をしている。

反応構成比の観察結果は双対尺度分析 (図 2) によっても確認される。同図と図 1 を比較すると、カテゴリ統合によって双対尺度の全体的傾向は変化していない。保有台数については第 1 軸方向に若干の圧縮がみられ、これはカテゴリ統合による第一固有値の減少に対応している。居住地域については、第 1 軸方向への僅かな変動を除いて統合されたカテゴリグループでのサンプル数による重み付き平均にはほぼ一致する。

表2. AICにより居住地域カテゴリの統合(乗用車保有について)

a. MAICE との AIC の差が1.0 以下, 有意水準 0.5%

No	カテゴリグループ	AIC	X^2	Y^2	DF	χ_1^2	DF	χ_2^2	DF
0	A : B : C : D : E : F : F : G : H : I : J	-26.94	127.96	62.94	18	113.63	10	23.54	8
1	AEF : BC : DG : HIJ	-42.53	103.78	54.53	6	89.36	4	20.46	2
2	AF : BCE : DG : HIJ	-42.42	103.76	54.41	6	89.13	4	20.64	2
3	AF : BC : DEG : HIJ	-42.29	103.48	54.29	6	89.39	4	20.13	2
4	A : BCEF : DG : HIJ	-42.05	104.20	54.05	6	87.58	4	22.49	2
5	A : BC : EF : DG : HIJ	-41.88	106.14	57.88	8	89.69	5	22.51	3
6	A : BCF : DEG : HIJ	-41.84	103.84	53.84	6	87.74	4	21.97	2
7	AE : BCF : DG : HIJ	-41.81	103.94	53.82	6	87.68	4	22.12	2
8	AEF : BCJ : DG : HI	-41.79	108.35	53.79	6	94.49	4	20.61	2
9	ABCEF : DG : HIJ	-41.65	100.64	49.64	4	87.32	3	19.14	1
10	AF : BCEF : DG : HI	-41.63	108.29	53.63	6	94.21	4	20.80	2
11	AF : BCJ : DEG : HI	-41.55	108.06	53.55	6	94.52	4	20.29	2

b. MAICE との AIC の差が2.0 以下 (a. の続き)

No	カテゴリグループ	AIC	X^2	Y^2	DF	χ_1^2	DF	χ_2^2	DF
12	A : BCE : F : DG : HIJ	-41.48	105.89	57.47	8	89.43	5	22.49	3
13	ABCF : DEG : HIJ	-41.46	100.31	49.46	4	87.45	3	18.69	1
14	A : BC : DEG : F : HIJ	-41.35	105.62	57.35	8	89.69	5	21.98	3
15	AE : BC : F : DG : HIJ	-41.33	105.72	57.33	8	89.62	5	22.14	3
16	ABC : EF : DG : HIJ	-41.32	102.50	53.32	6	88.84	4	19.61	2
17	ABEF : C : DG : HIJ	-41.24	103.40	53.24	6	89.21	4	20.20	2
18	AEF : BC : DG : HJ : I	-41.20	121.73	57.20	8	109.92	5	20.82	3
19	A : BCJ : EF : DG : HI	-41.14	110.71	57.14	8	94.78	5	22.70	3
20	ABF : CE : DG : HIJ	-41.10	103.30	53.10	6	88.93	4	20.35	2
21	AF : BCE : DG : HJ : I	-41.09	121.72	57.09	8	109.69	5	21.01	3
22	ABF : C : DEG : HIJ	-41.02	103.07	53.02	6	89.27	4	19.82	2
23	AF : BC : DEG : HJ : I	-40.97	121.44	56.97	8	109.95	5	20.50	3
24	ABCE : F : DG : HIJ	-40.90	102.25	52.90	6	88.66	4	19.51	2
25	A : BCEFJ : DG : HI	-40.85	108.32	52.85	6	92.14	4	22.66	2
26	ABEF : CJ : DG : HI	-40.80	108.34	52.80	6	94.75	4	20.36	2
27	ABC : DEG : F : HIJ	-40.79	101.98	52.79	6	88.84	4	19.08	2
28	A : BCEF : DG : HJ : I	-40.73	122.16	56.73	8	108.04	5	22.89	3
29	A : BCEJ : F : DG : HI	-40.69	110.42	56.69	8	94.47	5	22.67	3
30	A : BCFJ : DEG : HI	-40.68	107.99	52.68	6	92.35	4	22.14	2
31	AE : BCFJ : DG : HI	-40.66	108.09	52.66	6	92.29	4	22.29	2
32	A : BCJ : DEG : F : HI	-40.61	110.19	56.61	8	94.78	5	22.16	3
33	AE : BCJ : F : DG : HI	-40.60	110.29	56.60	8	94.72	5	22.31	3
34	ABF : CEJ : DG : HI	-40.58	108.15	52.58	6	94.37	4	20.51	2
35	ABF : CJ : DEG : HI	-40.57	108.01	52.57	6	94.81	4	19.98	2
36	A : BC : EF : DG : HJ : I	-40.56	124.09	60.56	10	110.17	6	22.94	4

表3. 居住地域に対する世帯の乗用車保有台数分布（統合表 No.1）

a. 統合表 No.1

居住地域	AEF	BC	DG	HIJ	合計
乗用車					
0台	50	67	59	2	178
1台	155	172	66	12	405
2台	1	7	1	7	16
合計	206	246	126	21	599

b. 統合表 No.1 の反応構成比(単位%)

居住地域	AEF	BC	DG	HIJ	合計
乗用車					
0台	24.3	27.2	46.8	9.5	29.7
1台	75.2	69.9	52.4	57.1	67.1
2台	0.5	2.8	0.8	33.3	2.7

図2. 乗用車保有についての双対尺度分析による解（統合表 No.1）

- ：居住地域（AEF, BC, DG, HIJ）
- ：乗用車保有台数（0～2台）

しかし、このような MAICE 表によるグループ分けは唯一のものではない。表2aにみるように MAICE 表と統合表 No.2 以下との AIC の差はわずかであり、この差は有意とはいえないからである。また、双対尺度解の全体的傾向の一致は表2aのすべての統合表にみられる。ここでこれらの統合表を全体として観察し

表4. カテゴリグループの形成状況（乗用車保有について）

カテゴリ	A	BC	DG	HI	E	F	J
A	—	1			4	7	
BC	1	—			3	4	3
DG			—		3		
HI				—			8
E	4	3	3		—	5	1
F	7	4			5	—	
J		3		8	1		—

よう。

表2a では、A, BC, DG, HI の4個の核となるグループに E, F, J の地域が結合したり離れたりすることでカテゴリグループのバリエーションが構成されている。これらの結合、離散の状況をみると、表4のようになる。E の地域は A, BC, DG のどの核ともほぼ同じように結合し、これらの核の中間的性格を持っていることがわかる。E が小規模サンプルであることを考慮すると、これは E については2台保有世帯が少ない（ない）ことのほか特別な特徴を結論付けられないことを示している。F は A, BC と結合し、特に A との結合が多い。F が中規模サンプルであることを考慮すると、F については2台保有世帯が少ない（ない）ことのほか、1台保有世帯がやや多いとみなしてよい。J は BC, HI と結合し、特に HI との結合が多い。J が小規模サンプルであることを考慮すると、J については2台保有世帯がやや多いことのほか特別な特徴は結論付けられない。

以上をまとめると、

（分析結果 1-4）

筑波研究学園都市周辺地域における乗用車保有動向は、全体として1台保有世帯が多いなかで、0台保有世帯の比較的多い地域である D, G, 1台保有世帯の一層多い地域である A, F, 同地域の平均的保有傾向を示す B, C の地域と、これらと比べて2台保有世帯が非常に多い H, I, J の地域によって特徴付けられている。

分析結果の（1-3）と（1-4）の大きな違いは小規模サンプルの E の評価にある。MAICE 表では E は A, F と同じグループに統合されたが、表4の分析から、この統合は B, C, D, G などとの統合であっても AIC や

表 5. 居住地域に対する軽乗用保有台数分布 (原表)

a. 原表

居住地域	A.	B.	C.	D.	E.	F.	G.	H.	I.	J.	合計
軽乗用車	0 台	126	94	144	71	6	61	47	9	5	568
	1 台	9	3	4	6	0	4	2	1	0	30
	2 台	0	1	0	0	0	0	0	0	0	1
	合計	135	98	148	77	6	65	49	10	5	599

b. 原表の反応構成比 (単位)

居住地域	A.	B.	C.	D.	E.	F.	G.	H.	I.	J.	合計
軽乗用車	0 台	93.3	95.9	97.3	92.2	100	93.8	95.9	90.0	100	94.8
	1 台	6.7	3.1	2.7	7.8	0.0	6.2	4.2	10.0	0.0	5.0
	2 台	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2

双対尺度解に大きな影響を与えないこと、すなわち E の地域の MAICE 表における位置付けを強調することは誤りであることが示された。

[分析例 2]

表 5a に同地域における軽乗用車の保有台数分布を、表 5b にその反応構成比を示す。同地域では、軽乗用車を保有しているのは全体の 5% であり、95% の世帯は軽乗用車を保有していない。

表 5a では $AIC=24.85$, $X^2=12.63$, $Y^2=11.15$ ($\chi^2_{18}(25\%)=21.60$) で、居住地域と軽乗用車保有台数の相互依存関係があるとはみなしがたい。同表を直接双対尺度分析にかけても得られる解は 25% 水準でさえ有意でない。それでは軽乗用車の保有状況は居住地域とは無関係であると断言できるであろうか。また、表 5b にみられる居住地域による保有台数分布の差……特に、比較的サンプル数の大きい A, D と B, C との差などは無視できるものであろうか。

前例と同様にカテゴリ統合を行うと表 6a のようになる。原表において相互依存度がきわめて低い場合、統合表の選択は主に第 1 の基準、第 2 の基準 (有意水準を 5% としている) によって行われた。表に示された統計情報は原表 (No. 0) を除いて 5% 水準で有意である。従って、これらの統合表からみると居住地域と軽乗用車保有台数とは独立であるとはいえない。

MAICE 表 (統合表 No. 1: 表 7a) における反応構成比は表 7b のようになる。このグループ分けは表 5b における全体の 0 台保有世帯あるいは 1 台保有世帯の割合を基準とする分類と一致する。従って、

図 3. 軽乗用車保有についての双対尺度分析による解
(a) 原表 (図左) (b) 統合表 No.1 (図右)

(分析結果 2-1)

1 台保有世帯は ADFHJ で多く、BCEGI で少ない。また、2 台保有世帯の有無は有意な特徴とはいえない。

MAICE 表に対する双対尺度分析の結果を図 3 に示

表 6. AIC による居住地域カテゴリの統合（軽乗用車保有について）

a. MAICE との AIC の差が 1.0 以下，有意水準 5%

No	カテゴリグループ	AIC	X^2	Y^2	DF	χ_1^2	DF	χ_2^2	DF
0	A: B: C: D: E: F: G: H: I: J	24.85	12.63	11.15	18	8.73	10	3.83	8
1	ADFHJ: BCEGI	-3.06	6.53	7.06	2	6.54	2		
2	ADFGHJ: BCEI	-2.93	6.22	6.92	2	6.23	2		
3	ADFHJ: BCEG	-2.65	6.13	6.65	2	6.13	2		
4	ADEFHJ: BCGI	-2.58	6.05	6.57	2	6.05	2		

b. MAICE との AIC の差が 2.0 以下（a. の続き）

No	カテゴリグループ	AIC	X^2	Y^2	DF	χ_1^2	DF	χ_2^2	DF
5	ADFGHIJ: BCE	-2.55	5.88	6.55	2	5.88	2		
6	ADEFGHJ: BCI	-2.48	5.81	6.48	2	5.82	2		
7	ADEFHIJ: BCG	-2.20	5.67	6.19	2	5.68	2		
8	ADEFGHIJ: BC	-2.13	5.50	6.13	2	5.49	2		
9	ADFJ: BCEGHI	-2.07	5.64	6.07	2	5.64	2		
10	ADFGJ: BCEHI	-1.88	5.30	5.88	2	5.30	2		
11	ADFH: BCEGIJ	-1.78	5.33	5.78	2	5.33	2		
12	ADFIJ: BCEGH	-1.70	5.26	5.70	2	5.26	2		
13	ADEFJ: BCGHI	-1.63	5.19	5.63	2	5.18	2		
14	ADFGH: BCEIJ	-1.61	5.05	5.61	2	5.05	2		
15	ADFGIJ: BCEH	-1.55	4.99	5.55	2	4.98	2		
16	ADEFGJ: BCHI	-1.48	4.93	5.48	2	4.92	2		
17	ADFHI: BCEGJ	-1.42	4.97	5.42	2	4.97	2		
18	ADHJ: BCEFGI	-1.40	5.22	5.40	2	5.22	2		
19	ADEGH: BCGIJ	-1.35	4.90	5.35	2	4.90	2		
20	ADFHJ: B: CEGI	-1.35	10.70	9.35	4	7.51	3	3.19	1
21	ADFGHI: BCEJ	-1.29	4.75	5.29	2	4.75	2		
22	ADEFIJ: BCGH	-1.29	4.84	5.29	2	4.83	2		
23	ADFHJ: BI: CEG	-1.24	10.40	9.24	4	7.54	3	2.85	1
24	ADEFGH: BCIJ	-1.23	4.69	5.23	2	4.69	2		
25	ADFHJ: BE: CGI	-1.22	10.34	9.22	4	7.55	3	2.79	1
26	ADEFGIJ: BCH	-1.17	4.64	5.17	2	4.63	2		
27	ADFHJ: BEI: CG	-1.14	10.09	9.14	4	7.60	3	2.49	1

す。左に示した原表に対する解と比較すると，2 台保有世帯に対応する尺度がやや小さくなったほかは，カテゴリ統合によって双対尺度の全体的傾向は変化していない。前例と同様に，保有台数について若干の圧縮がみられ，居住地域について統合されたカテゴリグループでのサンプル数による重み付き平均にほぼ一致する。従って，カテゴリ統合によって得た解は原表の内部情報のなかで重要な情報のみを抽出したもので，そ

の抽出情報のみに限定すればそれは有意な情報であるといえる。原表が持っていた情報は，10 個の地域全ての相違を表現するには有意でないが，MAICE 表のように 2 個の地域群の相違を表現する程度には有意であるといえることもできる。

表 6a のカテゴリグループを観察すると，核となる ADFHJ, BC の 2 個のグループに E, G, I の地域が結合離散することでバリエーションが構成されている。

表7. 居住地域に対する軽乗用車保有台数分布 (統合表 No1)

a. 統合表 No1

居住地域	ADFHJ	BCEGI	合 計
軽乗用車	0 台	272	296
	1 台	21	9
	2 台	0	1
	合計	293	306
			599

b. 統合表 No1 の反応構成比 (単位%)

居住地域	ADFHJ	BCEGI	合 計
軽乗用車	0 台	92.9	96.7
	1 台	7.1	2.9
	2 台	0	0.3
			0.2

表8. カテゴリグループの形成状況 (軽乗用車保有について)

カテゴリ	ADFHJ	BC	E	G	I
ADFHJ	—		1	1	1
BC		—	3	3	3
E	1	3	—	2	2
G	1	3	2	—	2
I	1	3	2	2	—

カテゴリグループの形成状況 (表8) をみると, E, G, I の3地域は結合, 離散について同等で, BC との結合が多い. E, I については軽乗用車保有世帯がまったくないが, これらが小サンプルであることを考慮すると, 特別な結論は付けられない. G についてはこれが中規模サンプルであることから, 軽乗用車保有についてはほぼ中間的でどちらかといえば非保有世帯がやや多い地域とみなしてよい. 以上をまとめると,

(分析結果 2-2)

筑波研究学園都市周辺地域における軽乗用車保有動向は, 全体として保有世帯が非常に少ないなかで, 比較的保有世帯の多い A, D, F, H, J の地域, および一層保有世帯の少ない B, C, G の地域によって特徴付けられている.

この例の分析では, 原表では相互依存度が低いとみられる変数間の関係から, AIC を用いたカテゴリ統合によって重要な情報—それだけを取り出せば有意な情

報を抽出し, さらに本方法によってそのなかで最も重要な情報のみを抽出することができたといえよう.

以上, 分析例1ではカテゴリ統合によるクロス表分析法を具体的に提示しながらその特徴を説明し, 分析例2では原表の変数間相互依存性が低い場合にも本方法で最も有効な情報のみを抽出できることを示した. 本節で行ったカテゴリ統合によるクロス表分析法の概念の手順を図4に示す.

4. 考 察

AIC は, 真の確率分布を近似する複数のモデルがあるとき, 真の分布とモデルの分布との近さをはかることによってモデルの良さを定義するものといえる. 本論文ではその近似モデルを解釈する立場から AIC によるモデル選択を捉え直し, カテゴリ統合を行うことによって, 原表から見易い簡潔な統合表が作成されることと, そのとき失われる情報がカテゴリ間の類似関係に対応することを冒頭で指摘した. 本論文で注目していることは, クロス表のもつ主要な情報は何かという問題に対してこのような設定が実用的な理解の場を提供することである. 作成された統合表が解釈し易いのは, 原表から類似関係にある情報を捨象して, 特徴的な相違点を強調しているためであるといえることができる.

さらに分析例に見るように, MAICE と同程度に簡潔な複数のカテゴリ統合表を全体的にながめるとかなり一貫した統合のパターンが存在する. このことは, 統合表で強調された情報のうちで, さらにそのパターンが示すカテゴリグループの相違点は AIC の意味でクロス表のもつ最も重要な特徴であることを示している.

同時にそのパターンは, その最も重要な特徴を強調するために失われた情報を表しており, カテゴリ間の類似関係として明瞭な特徴となるカテゴリグループを示していると解釈することができる. その類似関係は双対尺度法の解空間における距離関係とは一致しないが, 分析結果を解釈することははるかに容易である.

前節で提案した分析方法は, MAICE 表との AIC の差が有意でない統合表をすべて抽出し, それらを通観することによってそのパターンを発見するものであった. 坂元ら (1983) によれば AIC の差の有意水準は 1~2 とされている. この有意水準を 2.0 とすることの是非を調べてみよう.

分析例1では MAICE との AIC の有意差を 2.0 ま

図4. カテゴリ統合によるクロス表分析法

図5. AICの増加に伴う双対尺度解の推移（乗用車保有）

(a) 第1の解 (b) 第2の解

図6. AICの増加にともなう双対尺度解の推移(軽乗用車保有)

で緩めても追加された統合表の統計情報はみな0.5%水準で有意である(表2b). また, 統合表の内容も, 双対尺度分析の解のうえではさほど大きく変化していない(図5a, b). 従って, この例ではAICの差が1.0ないし2.0程度以下であれば大きな情報損失は起こっていないといえる.

分析例2で同様にAICの有意差を2.0まで緩めると, 追加された統合表の統計情報はごく一部を除いて5%水準で有意ではない(表6b). 双対尺度分析の解を見ると2つの解が得られる統合表(No. 20, 23, 25, 27)が含まれ, これらの統合表の解では地域Bにおける2台保有世帯の存在が特徴として強調されているが, それ以外の統合表ではMAICE表の解と比べてさほど大きく変化していない(図6). 2台保有世帯の存在強調をMAICE表との有意な相違とみなすとすれば, この例ではAICの差の有意水準は, 1.7程度とすべきことになる.

このように双対尺度解からみると, カテゴリ統合によるAICの差の有意水準を一律とすることはよくないことがわかる.

さらにこの有意水準の変更の効果を調べてみよう. 表9~10に分析例1および2で選択基準を緩めた場合のカテゴリグループの形成状況を示す.

前節では, カテゴリグループの形成状況を分析するために「核」という考え方を導入し, 基準によって抽出された統合表のなかで組み合わせが変わらないカテゴリの統合を核と呼んだ. 核という扱いは多数の統合表を全体的に分析するうえで便利な概念であった.

表9によると, DとGは有意水準を緩めても常に同一のカテゴリグループに属している. BとC, HとIはいずれも30回以上同一のカテゴリグループに属していることから, そのほかのカテゴリ統合とは質的な相違があるように思われる.

BCという結合が成立しない場合は6回あり, この

表9. カテゴリグループの形成状況(乗用車保有: 有意水準を緩めた場合)

カテゴリ	A	B	C	D	E	F	G	H	I	J
A	—	11	5		11	17				
B	11	—	30		12	15				10
C	5	30	—		12	8				13
D				—	11		36			
E	11	12	12	11	—	13	11			4
F	17	15	8		13	—				3
G				36	11		—			
H								—	31	23
I								31	—	18
J		10	13		4	3		23	18	—

表10. カテゴリグループの形成状況(軽乗用車保有: 有意水準を緩めた場合)

カテゴリ	AD	F	J	H	G	E	I	BC
AD	—	18	15	13	8	7	7	
F	18	—	14	12	9	8	7	1
J	15	14	—	9	10	8	9	4
H	13	12	9	—	8	9	9	6
G	8	9	10	8	—	10	11	11
E	7	8	8	9	10	—	10	12
I	7	7	9	9	11	10	—	12
BC		1	4	6	11	12	12	—

ときにはBはAと結合している. HIという結合が成立しない場合は5回あり, このときにはHはJと結合している. AIC差が1.0以内にある統合表を上位, AIC差が2.0以内にある統合表を下位と呼ぶことにすれば, これらの場合はみな下位に起こっている. 一方, Aと結合しないBCという結合は26回あり, Jと結合しないHIという結合は13回ある. 従って, BC, HIとい

表 11. 核に関連した部分表の統計情報（乗用車保有）

部分表	AIC	X^2	Y^2	DF
AB	2.07	1.94	1.93	2
BC	3.53	0.45	0.47	2
ABC	2.97	4.78	5.03	4
HI	2.22	1.50	1.78	2
HJ	3.56	0.43	0.44	2
HIJ	4.88	2.77	3.12	4

う結合は回数においても出現の仕方についても AB, HJ という結合より重要であろう。

しかし、ABC という結合は上位で 2 回、下位で 2 回しかなくカテゴリグループの形成状況を分析するのにあまり寄与しないが、HIJ という結合は上位で 8 回、下位で 10 回あり、HI という結合との重要度の違いは判断し難い。

さらに、これらの部分表の統計情報(表 11)では、BC は AB, ABC より AIC も大きく、 X^2 , Y^2 についての上側確率からも独立性の高さを示しているが、HI は HJ や HIJ より AIC が小さく、 X^2 , Y^2 による上側確率も小さくなっている。

H, I, J が小サンプルであることも含めて、カテゴリグループの統合状況分析のための補助的手段としての核という概念についてはなお検討する必要がある。

表 10 によると、BC という核は変化しないが、ADFHJ という核は AD, F, J, H に分解する。AD と BC は全く結合せず、それぞれ軽乗用車保有世帯が多い地域と少ない地域を代表している。F はこれらに対する結合の割合からみて AD とほぼ一体である。J と H は同様にみると AD に近いが BC との結合も僅かに増える。表 8 をみるとこれらの地域はもともと軽乗用車保有世帯の割合が AD よりも多い地域で、BC との結合が増えているのはサンプル数が少ないことによる。E, G, I の結合状況は表 8 にみられた傾向を一層緩やかにしたもので、AD と BC との中間的な性格を強めている。

このように AIC の差の有意水準を緩めた場合、核がさらに分割され、各カテゴリの性格を一層詳しくみることができるとは、しかし、表 9 の分析例のように核を判

定する基準があいまいになったり、表 10 の分析例のように X^2 , Y^2 にたいする有意水準 (5%) が緩むことで分析結果の有意性が変化したりする場合がある。AIC の差の有意水準の効果的な設定には、なお検討を重ねる必要があろう。

5. ま と め

本論文では AIC を指標とするカテゴリ統合表に表れる特徴の意義の検討を行い、それがクロス表のもつ情報のうちでカテゴリ間の類似点に基づくものと相違点に基づくものによって構成されていることを示した。また、統合表の良さについて AIC を指標としたときに複数の統合表が同程度に良い場合、それらを比較検討することによって最も重要な特徴が抽出されることを示した。さらに、カテゴリ統合の効果を双対尺度法との関係で検討し、AIC を小さくするカテゴリ統合はデータ行列から弁別効果の低いカテゴリ区分についての情報を捨象する操作であることを示した。

これらを踏まえて本論文では、カテゴリ統合によるクロス表の特徴抽出法について提案を行うとともに、分析例を用いてその有効性を示し、双対尺度法を利用しながら検討を行った。

考察に述べたように、AIC の有意な差の範囲、補助的分析手段としての核の概念、もともと変数間の関連性の低いクロス表における詳細な分析の有意さの程度など、分析手続きとしての本方法にはなおいくつかの疑問点が残っている。これらを含めてクロス表データの分析方法を今後なお説明する必要があると考えている。

参 考 文 献

- 赤池弘次 (1976) 情報量基準 AIC とは何か、数理科学, No. 153, 5-11.
- Nishisato, S. (1980) *Analysis of Categorical Data —Dual Scaling and Its Applications*. University of Toronto Press.
- 西里静彦 (1982) 質的データの数量化、朝倉書店。
- 坂元慶行・石黒真木夫・北川源二郎 (1983) 情報量統計学、共立出版。
- 坂元慶行 (1985) カテゴリカルデータのモデル分析、共立出版。

—1986 年 4 月受付, 8 月改稿受付—