

## 第 2 部

## 第5章

# 地域格差の統計的分析

—台湾地域の事例を中心として—

### はじめに

輸出志向によって発展してきた台湾経済の1960年以降の特徴は、金融自由化や国際化に見られるような経済構造の変化によるサービス産業のめざましい発達と2回の石油ショックによる経済成長の鈍化にもかかわらず、アメリカ経済の景気回復にともない回復から拡大へと転じたことである。こうしたあまりにも急激なサービス産業の成長ぶりは生産額の比重の低下をまねき、その相対的生産性の低落は経済全体の成長率を低下させる要因となったといわれている。この間、経済成長を最優先させたために急激な経済成長による弊害は1975年以降、さまざまな形で表面化してきた。特に、地域発展の不均衡による人口の一部都市集中と、産業の都市集中はさまざまな地域格差の問題を生じさせている。

このような台北市および大都市集中の傾向に見られる台湾地域の地域格差を統計的に検討しようとする際にまず最初に問題になることは、地域ごとに観測されるいくつかの事象を通して、地域間に格差が存在するか否かということである。統計的な分析方法とは、こうした地域間の格差の有無を確かめることであり、その格差が存在する場合には、地域ごとに現われる事象の傾向や特徴的なパターンを定量的にとらえ、それらの相互関係を明らかにする

と同時に、その因果関係を分析する方法をいう。統計的分析による結果をもとにその国・地域が固有に持つ経済的あるいは社会的な要素を加味し具体的な解釈を与えるのは地域研究の分野である。

本章は得られたデータから地域間の特徴ある情報を引き出すための統計的手法の紹介を行い、それらの手法によって台湾地域の事例を検討する。地域間の格差を明らかにするための手法としてジニ係数や変動係数等がよく用いられるが、ここでは情報量をもとに台北市を中心とする北部地区集中の実態を明らかにする。さらに、この情報量を地域区分間による情報量と地域区分内の情報量に分解して、地域格差の状態をより詳細に検討する。また、地域格差の特徴を引き出すためには、地域間の変動を大きく同時に地域内の変動を小さくするように地域区分を行い、類似した地域の特徴と異なった地域の特徴をパターン化し、与えられたデータから地域性を見いだすことが必要である。こうした地域間の変動を検討する方法としては各種の統計的な手法が考えられているが、ここでは情報量基準である AIC (Akaike Information Criterion・赤池情報量基準) をクロス表に応用し、AIC にもとづく階層的なクラスターによって台湾地域の最適な地域分割を行い、そのクラスターごとの特徴を明らかにする手法を用いた。ここで AIC を用いるのは通常の尤度比検定による検定では同一の尤度関数の下でパラメータの全体集合とその部分集合の比較に限られており、階層構造を明らかにする場合には効果的であるが、AIC ではこうしたクラスターの階層構造に関わりなく判定できるので、モデルの選択の幅が広がるからである。

台湾地域のサービス業や北部地区への集中は、AIC によって得られた地域分割と行政区分による地域分割との比較を通して明らかにする。また、対数線形モデルを用いて同じ課題を別の視点から考えてみる。

この分析のためのデータとして就業者数<sup>(1)</sup>と基本所得<sup>(2)</sup>を使用した。前者のデータは地域区分×産業区分×年(1982～89年)の3重クロス表、後者のデータは産業区分がなく、地域区分×年(1976～89年)で構成されている。産業区分は(1)農林漁および狩猟業(牧畜業を含む)、(2)鉱業および土石採取業、(3)

製造業, (4)電気・ガス・水道業, (5)建設業, (6)商業, (7)運輸・倉庫・通信業, (8)金融・保険・不動産および商工サービス業, (9)社会および個人サービス業である。台湾地域の行政区分による地域分割は23市・県に分かれていて, その内訳は北部地区として台北市, 基隆市, 新竹市, 台北県, 宜蘭県, 桃園県, 新竹県の7市・県, 中部地区として台中市, 苗栗県, 台中県, 彰化県, 南投県, 雲林県の6市・県, 南部地区として高雄市, 台南市, 嘉義市, 嘉義県, 台南県, 高雄県, 屏東県, 澎湖県の8市・県, および東部地区として台東県, 花蓮県の2県の4地域区分である。<sup>(3)</sup>

## 第1節 情報量と K-L 情報量

標本空間を  $\Omega$  とするとき, ある仮説にもとづいて  $\Omega$  を分割した事象の集まりを  $\Omega = \{A_1 \cdots A_N\}$  とする。  $\Omega$  の分割であるから  $A_i \cap A_j = \phi$ , ( $i \neq j$ ) である。事象  $A_i$  の確率を  $\Pr(A_i) = p_i$  で表わし,  $p = \{p_1 \cdots p_N\}$  とする。この時確率  $p$  に関する情報量は,

$$H(p) = - \sum_{i=1}^N p_i \log p_i$$

で定義される。ただし,  $0 \log 0 = 0$  とする。この情報量  $H(p)$  はボルツマン型のエントロピーともいわれ, 上に凸の関数であり,  $0 \leq H(p) \leq \log N$  の範囲をとる。左辺の等号は確率的に見てある特定の要素  $i$  のみが存在し, それ以外の要素が存在しない場合,  $p_i = 1$ ,  $p_j = 0$  ( $i \neq j$ ) に成立する。また, 右辺の等号は全ての要素が等しい確率を持つ場合,  $p_i = N^{-1}$ , ( $i = 1 \cdots N$ ) に成立する。すなわち, 分布が完全に集中している状態は前者の例であり, 完全に平等な分布の状態は後者の例である。このことはエントロピーの値は分布の状態が完全に集中している時の値 0 から平等に近い状態に移動するにつれて増加し, 完全な平等になる分布の状態で上限  $\log N$  となることを示している。この性質を確率分布の集中度の度合を計る尺度として用いることがで

きる。

所得分配でしばしば使用される不平等度を表わすタイル尺度は、総所得を 1 に基準化した所得分配  $p = \{ p_1 \cdots p_N \}$  に対して、

$$T(p) = \log N - H(p) \quad \cdots \cdots (1)$$

で定義されている。タイル尺度の値は情報量とは逆に、分布の状態が完全に平等な時の値 0 から不平等な状態に移動するにつれて増加し、完全に集中した状態で上限  $\log N$  となる。

次に  $N$  個ある事象  $A$  を  $n$  個のグループに統合し、 $\Omega = \{ G_1 \cdots G_n \}$  になったとする。 $\Omega$  の要素は、

$$G_i = \bigcup_{h=1}^{N_i} A_{ih}, \quad A_{ih} \subset A_i, \quad (h=1 \cdots N_i) \quad \sum_{i=1}^n N_i = N$$

として構成される。 $A$  の要素は互いに共通部分を持たないので、 $G_i$  の確率を  $P_i$  とすれば、

$$\Pr \{ G_i \} = \sum_{h=1}^{N_i} \Pr \{ A_{ih} \} = P_i$$

として得られる。その時、情報量  $H(p)$  は次のように変形することができる。

$$H(p) = - \sum_{i=1}^n P_i \log P_i + \sum_{i=1}^n P_i \left\{ - \sum_{j=1}^{N_i} (p_{ij}/P_i) \log (p_{ij}/P_i) \right\} \cdots (2)$$

ここで  $w_j = \{ p_{i1}/P_i \cdots p_{iN_i}/P_i \}$  右辺の  $\{ \quad \}$  の中の式を  $H(w_i)$  と置けば、

$$\begin{aligned} H(p) &= H(P) + \sum_{i=1}^n P_i H(w_i) \\ &= HB + HW \end{aligned}$$

と書き直すことができる。右辺の第 1 項  $HB$  はグループ化された事象  $G$  の情報量であり、グループ間の情報量ともいわれる。第 2 項  $HW$  はグループ内の情報量  $H(w_i)$  に  $P_i$  がウエイトとしてついている形をしており、グループ内情報量ともいわれる。このことから全体の情報量はグループ間の情報量とグループ内の情報量に分解されることを示している。

標本空間  $\Omega$  に異なる確率分布  $P$  と  $Q$  が想定され、 $P$  を真の分布、 $Q$  を観測

値から得られた分布とする。P の分布と Q の分布の類似性を調べるための尺度として用いられる Kullback-Leibler (K-L) 情報量は以下のように定義される。

$$I(p; q) = \sum_{j=1}^N p_j \log(p_j/q_j)$$

K-L 情報量は、 $I(p; q) \geq 0$  であり、等号は  $p_j = q_j$ , ( $j=1 \cdots N$ ) の場合に成り立つ。このことは、 $I(p; q)$  の値が 0 の時、Q の分布 P の分布に完全に一致し、0 に近いほど Q の分布は P の分布に類似する。

前と同じように A をグループ化して G に統合すると、 $G_i$  に属する A の要素は  $G_i = \{ A_{ih} \mid A_{ih} \subseteq \Omega, h=1 \cdots N_i \}$  である。K-L 情報量は変形して、

$$\begin{aligned} I(p; q) &= \sum_{i=1}^n P_i \log(P_i/Q_i) + \sum_{i=1}^n P_i \left\{ \sum_{j=1}^{N_i} (p_{ij}/P_i) \log[(p_{ij}/P_i)/(q_{ij}/Q_i)] \right\} \\ &= I(P; Q) + \sum_{i=1}^n P_i I_i(u; v) \\ &= IB + IW \qquad \dots\dots(3) \end{aligned}$$

と表され、K-L 情報量に関しても全体の情報量  $I(p; q)$  がグループ間の情報量 IB とグループ内の情報量 IW に分離されることがわかる。 $u_i = \{ (p_{i1}/P_i) \cdots (p_{iN_i}/P_i) \}$   $v_i = \{ (q_{i1}/Q_i) \cdots (q_{iN_i}/Q_i) \}$  である。

## 第2節 情報量による所得と就業者数の地域格差の拡大

台湾地域における基本所得と就業者数の地域格差を(1)式で表されたタイルの尺度、ジニ係数および変動係数<sup>4)</sup>による集中度で示したのが第1表である。この3つの集中度はそれぞれ基準が異なり横並びの比較は意味をなさない。そこでこれらを平均0と分散1になるように標準化し、年毎の推移を示したのが第1図である。前述したように上昇すれば集中度は高まる。この図から3つの集中度はほぼ同じ傾向を示していることが確かめられる。同図の(a)より所得については、1982年から89年にかけて地域間の格差が広がってきてい

第1表 基本所得と就業者数の集中度の比較

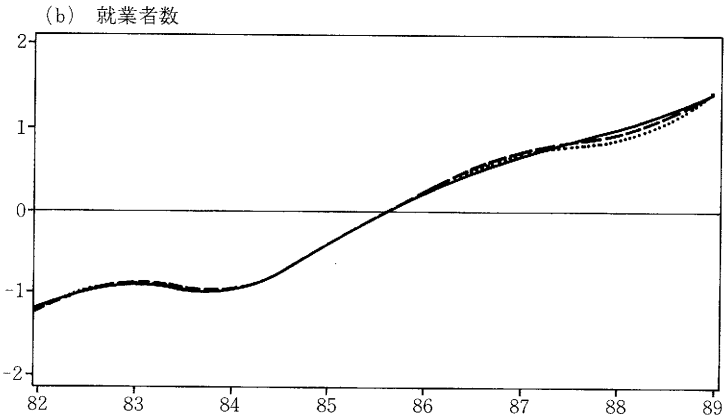
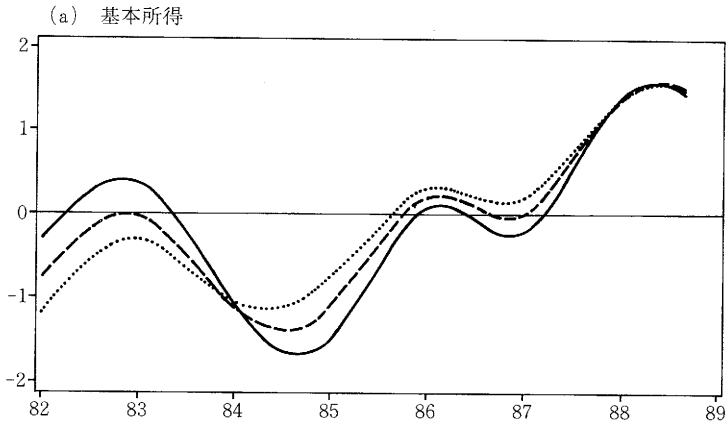
	変動係数	ジニ係数	タイル尺度	H(P)
(a) 基本所得				
1982	1.11108	0.46583	0.40774	2.72775
83	1.12761	0.47606	0.42066	2.71483
84	1.09230	0.46738	0.40143	2.73407
85	1.08198	0.47104	0.40125	2.73425
86	1.12091	0.48317	0.42515	2.71035
87	1.11414	0.48210	0.42200	2.71350
88	1.15366	0.49630	0.44718	2.68831
89	1.14906	0.49614	0.44733	2.68816
(b) 就業者数				
1982	0.68758	0.35029	0.20593	2.92956
83	0.69883	0.35519	0.21151	2.92399
84	0.69682	0.35422	0.21024	2.92526
85	0.71926	0.36155	0.22078	2.91472
86	0.74249	0.37059	0.23245	2.90305
87	0.76039	0.37639	0.24075	2.89475
88	0.77318	0.37879	0.24509	2.89041
89	0.78929	0.38621	0.25379	2.88171

(出所) 筆者作成。

る傾向がうかがえるが、84と85年、またわずかではあるが87、89年には縮小の時期があったことがわかる。同図の(b)からは就業者数についても1984年に縮小の時期はあったものの82年から89年にかけて格差が拡大していたことになる。1988年は変動係数が直線的に上昇するのに対して、ジニ係数とタイルの尺度は上昇の仕方が緩やかになっている。しかし、これらの集中度の間に大きな違いは見られないので、以下、タイルの尺度の逆の関係にある情報量  $H(p)$  ((1)式を参照) にもとづいて地域格差を考えることにして、他の係数による解釈は省略する。

就業者のデータは産業×地域×年の情報を持っている。そのため産業別に見た就業者数の地域間格差の全情報量と全情報量を(2)式に従って分解した地

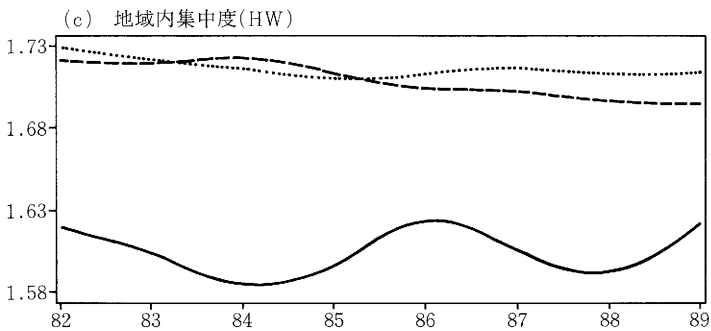
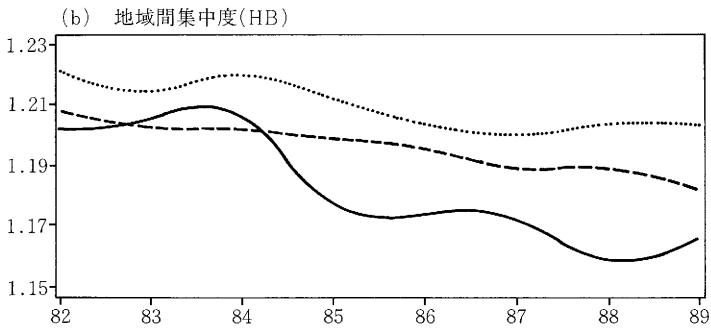
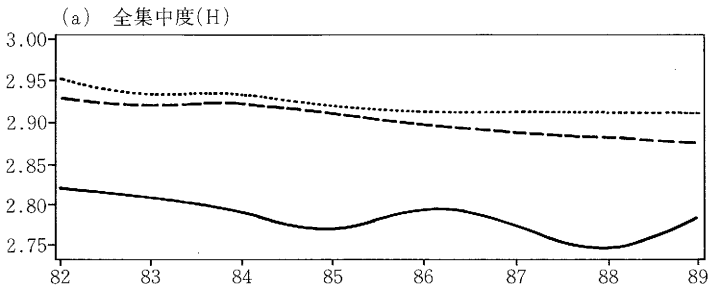
第1図 集中度の比較



(出所) 筆者作成。(——変動係数, .....ジニ係数, ----タイル尺度)



第2図 就業者の産業別集中度



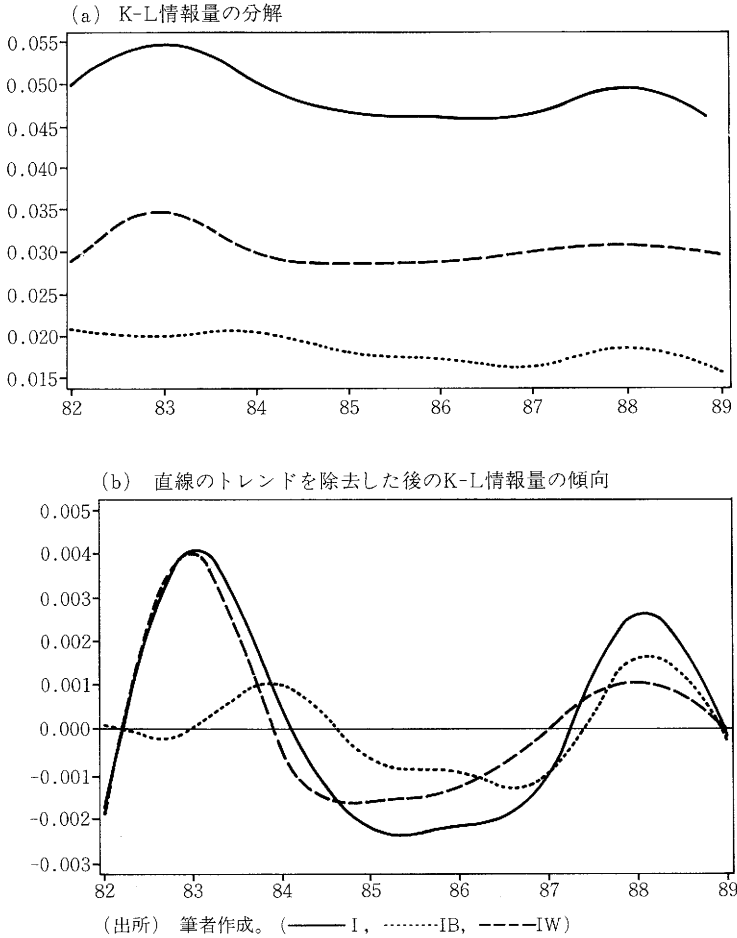
(出所) 筆者作成。(——第1次産業, .....第2次産業, ----第3次産業),  
地域区分は行政区分によって分割された4地域である。

地域区分間（グループ間）の情報量と地域区分内（グループ内）の情報量を年ごとに比べることができる。地域区分は行政区分によって分割された北部、中部、南部および東部の4地域である。情報量  $H(p)$  を用いて集中度の度合いを測るので情報量という言葉を集集中度に置き換えることにする。タイルの尺度とは逆に下降するほど集中の度合いが高くなる。第2図に3つの集中度が産業別に示されている。第1次産業の全集中度は1982年から85年までと87、88年は拡大しているが、86年と89年は縮小している。地域内集中度は全集中度とほぼ同じ傾向を示していることからこの間の拡大と縮小は主として地域内の影響によるものと判断される。特に地域間では1984年まで縮小していたのが急激に拡大に転じている。第2次産業は全集中度で見れば1985年までは格差は拡大し、86年以降は一定を保っている。この一定に保たれているように見える期間、地域間格差が拡大した分は地域内の縮小で調整されたようである。1984年にはわずかな地域間格差縮小の存在が確かめられる。第3次産業は1984年までは一定を保ち、その後拡大に変化している。地域内の格差縮小が1984年にあるが、地域内および地域間共に格差が拡大する傾向にある。

前述したように、全産業における就業者数の地域格差は1984年を除いて年々拡大しており、85年までは第1次と第2次産業の影響が強く86年以降では第3次産業の貢献度が大きくなる。1984年の格差の縮小は第2次と第3次産業の格差縮小によるものであり、その要因は2次産業の地域間と第3次産業の地域内の格差の縮小によるものと考えられる。

次に、所得と就業者の地域格差の集中度の類似性は(3)式の  $K-L$  情報量で測ることができる。この  $I(p; q)$  において  $p$  を基本所得、 $q$  を就業者数、地域の統合を行政区分による地域分割とする時、この結果は第3図に図示されているとおりになる。 $I(p; q)$  の値が0に近いほど類似の度合いが強い。同図の(a)の  $I$  より、1983年と88年にわずかな山が観測され、所得と就業者の間の地域間の片寄り具合の違いが拡大している状況を見ることができる。第1図によれば、所得の地域格差の拡大は83、86、88年に表れているこの結果を第3図と比べると、 $K-L$  情報量の拡大の時期は1986年を除いて所得の格差

第3図 K-L情報量のグラフ



の拡大の時期と一致しているように見える。しかし所得と直接に関わる家計消費について台湾地域では3年を周期とする短期循環および6～7年の中期循環を持っているといわれており、この経済変動の影響が所得にも反映しているとみなしなければならない。そのため、両者の拡大時期の一致を判断するこ

とは容易ではない。

1983年はIとIWに共通の上昇があり、IBは一定の状態にあることから、両者の違いの拡大は地域区分間による影響よりも、地域区分内の違いの拡大の影響によるものであることが想定される。このことはまた、第3図の(a)の各集中度から直線のトレンドを差し引いて変動のみを強調した同図の(b)から、1983年前後はIとIWが重なり合うことから明らかである。一方、1988年の両者の違いは逆に、地域区分間の影響が強い。以上のことから、台湾においては所得と就業者共に地域格差は拡大する方向にあり、第3次産業にその傾向が強く現われている。しかもその拡大要因は地域間格差より、地域内格差による影響と考えられる。

### 第3節 AICの値にもとづく台湾地域区分の分割

地域区分をA、産業区分をBとすると、 $A \times B$ によって表現される就業者数のクロス表について、産業区分を固定したまま地域区分の統合をAICの値を基準にしたクラスター分析によって試みる。分析は年ごとに行う。地域区分の数をR個、産業区分の数をC個とする。ここでCは産業分類数9に等しい。任意の2つの地域区分、 $A_h$ と $A_i$ を1つの区分 $A_h \cup A_i$ に統合することを考える。産業区分がjであり、 $A_h$ と $A_i$ に属する観測値をそれぞれ $n_{hj}$ と $n_{ij}$ とすれば、 $A_h \cup A_i$ の観測値は $n_{hj} + n_{ij}$ となり、地域区分の数が1つ減少する。補論1式より、このクロス表のAICの値(補論の「AICにもとづく分割のモデル」を参照のこと)、 $AIC(A_1, \dots, A_h \cup A_i, \dots, A_R)$ を、

$$AIC[R, (A_h, A_i)] = -2 \sum_{i=1}^{R-1} \sum_{j=1}^C n_{ij} \log \left\{ (N n_{ij}) / (n_i \cdot n_j) \right\} \\ + 2(R-2)(C-1)$$

で表すことにする。ここで、 $\{A_1 \dots A_h \cup A_i \dots A_R\}$ を改めて $\{A_1 \dots A_{R-1}\}$ と置き換える。元の地域区分はR個に区分されているので、R個の中から

2個を取り出す組合せの数、 $R(R-1)/2$ だけ統合の組合せが存在する。これらの組合せの中で、AICの値が最小になるものを最適な組合せとする。すなわち、組合せの中で産業区分による構成比が一番類似している地域をまとめるといふ方法を用いる。このようにして地域区分であるAの区分数をAICの値を最少にするようにしつつRから始まって、 $R-1$ 、 $R-2$ …… $2$ 、 $1$ と逐次統合していく階層的な方法を階層的なクラスター分析という。最適な区分数の決め方については、AICの値が最少となる区分数があればそれを最適なものとする。AICが単調に変化していく場合には、統合プロセスの各段階のAICがその段階までの値とそれ以降の値との間に大きなギャップの生じるところを一定の目安とする。しかしながらこれらの基準は絶対的なものではないので、得られた区分数によって分割されたグループが無理なく解釈ができるところで打ち切った方がよいと思われる。

台湾地域の23の地域区分の特徴を明らかにするため、地域間の格差を拡大させるような地域分けをする階層的なクラスターを構成してみよう。第2表には階層的なクラスターによる地域区分数とそれに対応するAICの値およびAICの差である $\Delta AIC$ が年ごとに示されている。また、 $\Delta AIC$ の変化は第4図に示されている。1982年から89年の全般にわたって共通していえる $\Delta AIC$ の変化は地域区分が6まではおだやかな減少を示しているが、地域区分が5より小さくなると年による変化が表れることである。1982年は地域区分が5から4にかけて変化するよりも4から3への変化の方が顕著に表れている。1983年と84年は区分が5から4にかけて82年に比べて急激な落ちを示し、4から3にかけては変化が少ない。第4表から階層的なクラスターの逐次的な統合プロセスを表わすAICの値の変化を見ることができる。またこのプロセスに対応するデンドログラム(樹形図)が第5図である。以上のことから行政区分の地域分割である4の地域区分は階層的なクラスターの打ち切りの個数としても一定の根拠があると思われる。

第2表 AICの値

(a) 絶対値の対数変換に符号をつけたAIC

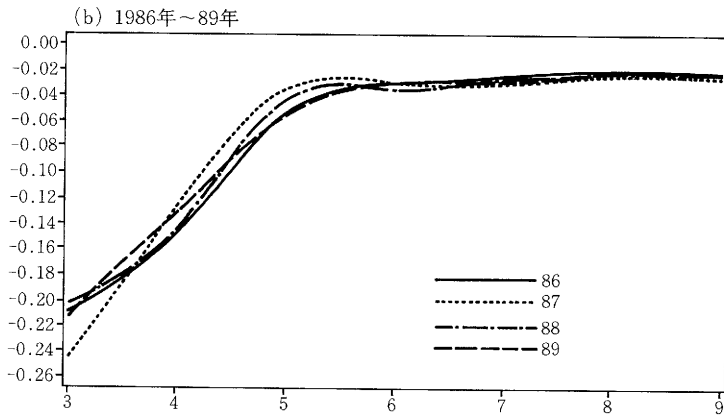
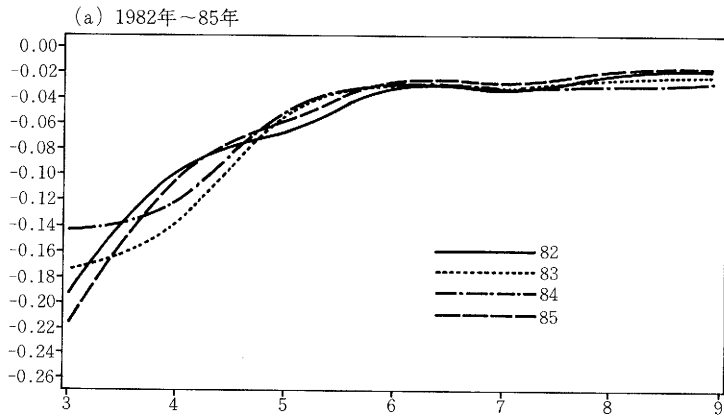
地域区分数	82	83	84	85	86	87	88	89
22	-14.1280	-14.1824	-14.2336	-14.2844	-14.2772	-14.2581	-14.3171	-14.3184
21	-14.1258	-14.1804	-14.2317	-14.2830	-14.2255	-14.2564	-14.3157	-14.3173
20	-14.1232	-14.1778	-14.2294	-14.2816	-14.2235	-14.2548	-14.3141	-14.3157
19	-14.1204	-14.1749	-14.2259	-14.2789	-14.2208	-14.2528	-14.3124	-14.3141
18	-14.1175	-14.1710	-14.2224	-14.2780	-14.2179	-14.2501	-14.3107	-14.3117
17	-14.1144	-14.1670	-14.2187	-14.2757	-14.2168	-14.2476	-14.3082	-14.3098
16	-14.1111	-14.1629	-14.2158	-14.2723	-14.2108	-14.2443	-14.3053	-14.3068
15	-14.1088	-14.1583	-14.2109	-14.2688	-14.2055	-14.2405	-14.3002	-14.3029
14	-14.1013	-14.1524	-14.2058	-14.2636	-14.2002	-14.2345	-14.2946	-14.2985
13	-14.0889	-14.1359	-14.1918	-14.2512	-14.1958	-14.2283	-14.2866	-14.2895
12	-14.0807	-14.1266	-14.1826	-14.2411	-14.1858	-14.2205	-14.2824	-14.2869
11	-14.0702	-14.1167	-14.1720	-14.2298	-14.1774	-14.2118	-14.2759	-14.2782
10	-14.0578	-14.1026	-14.1589	-14.2180	-14.1651	-14.2013	-14.2677	-14.2693
9	-14.0444	-14.0862	-14.1374	-14.2057	-14.1526	-14.1899	-14.2572	-14.2458
8	-14.0253	-14.1108	-14.1374	-14.2057	-14.1375	-14.1695	-14.2415	-14.2300
7	-13.9950	-14.0396	-14.1108	-14.1894	-14.1213	-14.1479	-14.2191	-14.2114
6	-13.9532	-14.0821	-14.0821	-14.1648	-14.1004	-14.1219	-14.1950	-14.1880
5	-13.8978	-14.0100	-14.0522	-14.1387	-14.0715	-14.0944	-14.1629	-14.1593
4	-13.8298	-13.9557	-14.0021	-14.0815	-14.0180	-14.0605	-14.1312	-14.1262
3	-13.7039	-13.8185	-13.8812	-13.9772	-13.8704	-13.9340	-13.9782	-13.9742
2	-13.6039	-13.4442	-13.7360	-13.7590	-13.6596	-13.6873	-13.7772	-13.7648
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

(b)  $\Delta$ AIC

地域区分数	82	83	84	85	86	87	88	89
22	-0.00229	-0.00200	-0.00191	-0.00131	-0.00165	-0.00168	-0.00142	-0.00105
21	-0.00260	-0.00266	-0.00227	-0.00143	-0.00205	-0.00168	-0.00160	-0.00166
20	-0.00271	-0.00285	-0.00347	-0.00174	-0.00268	-0.00238	-0.00165	-0.00166
19	-0.00295	-0.00395	-0.00352	-0.00190	-0.00291	-0.00238	-0.00178	-0.00232
18	-0.00310	-0.00398	-0.00367	-0.00233	-0.00311	-0.00244	-0.00245	-0.00272
17	-0.00332	-0.00411	-0.00389	-0.00312	-0.00442	-0.00329	-0.00292	-0.00421
16	-0.00427	-0.00455	-0.00397	-0.00378	-0.00439	-0.00385	-0.00510	-0.00429
15	-0.00538	-0.00596	-0.00504	-0.00520	-0.00528	-0.00601	-0.00585	-0.00498
14	-0.00724	-0.00899	-0.00746	-0.00543	-0.00717	-0.00622	-0.00607	-0.00603
13	-0.00815	-0.00937	-0.00856	-0.00698	-0.00725	-0.00773	-0.00614	-0.00663
12	-0.00815	-0.00937	-0.00923	-0.01008	-0.00834	-0.00872	-0.00822	-0.00871
11	-0.01052	-0.00988	-0.01130	-0.01130	-0.01231	-0.01051	-0.00822	-0.01392
10	-0.01236	-0.01409	-0.01392	-0.01180	-0.01513	-0.01138	-0.01049	-0.01450
9	-0.01340	-0.01640	-0.02137	-0.01237	-0.01513	-0.02039	-0.01572	-0.01576
8	-0.01907	-0.02256	-0.02137	-0.01623	-0.01618	-0.02167	-0.02243	-0.02344
7	-0.03031	-0.02401	-0.02873	-0.02464	-0.02091	-0.02593	-0.02115	-0.02344
6	-0.03182	-0.02966	-0.02991	-0.02612	-0.02887	-0.02749	-0.03212	-0.02841
5	-0.06540	-0.05425	-0.05009	-0.05723	-0.05352	-0.03399	-0.04165	-0.0428
4	-0.09963	-0.13724	-0.12085	-0.10423	-0.14759	-0.03399	-0.04165	-0.13110
3	-0.19432	-0.17430	-0.14519	-0.21825	-0.21081	-0.24675	-0.20102	-0.20940
2	-13.6039	-13.6442	-13.7360	-13.7590	-13.6596	-13.6873	-13.7772	-13.7648

(注) 第4図を見やすくするために、 $\log(|AIC|)$ に符号を付けた値を改めてAICと置いた。

(出所) 筆者作成。

第4図  $\Delta$ AICのグラフ

(注) 各 $\Delta$ AICを区別し易いように、 $\Delta$ AICの絶対値に対数変換した値を用いた。x軸は地域区分の個数を表わす。

(出所) 筆者作成。





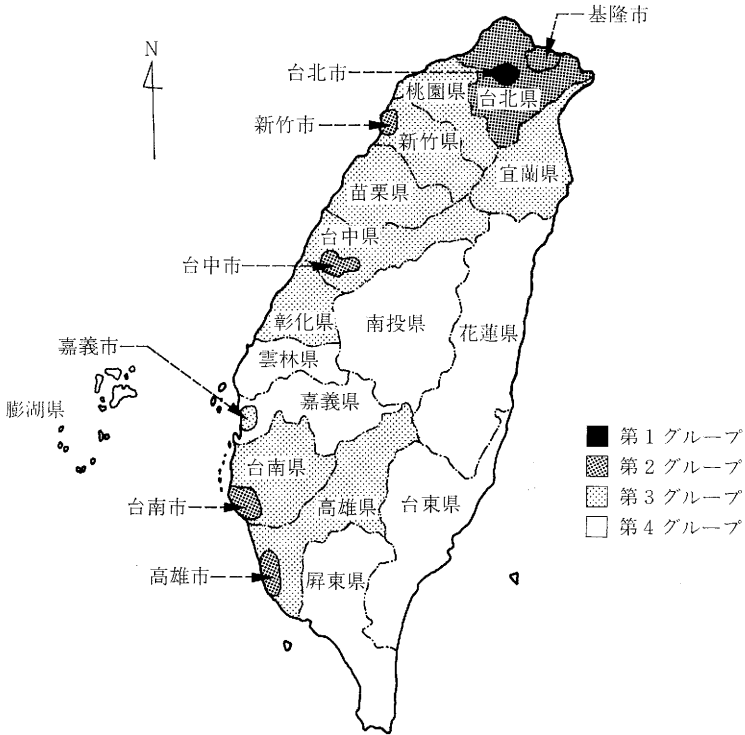
第6図 AICの最小化による地域区分の推移

グループ	市県	82	83	84	85	86	87	88	89	
1	台北市	22	22	22	22	22	22	22	22	
	基隆市	・	・	・	⑰	→	⑰	・	・	
	台中市	・	⑲	・	・	・	・	・	・	
	高雄市	・	⑳	・	・	・	・	・	・	
2	基隆市	⑰	→	⑰	・	・	・	⑰	→	⑰
	台中市	⑲	→	⑲	→	→	→	→	→	⑲
	高雄市	⑳	→	⑳	→	→	→	→	→	⑳
	台北県	1	1	1	1	1	1	1	1	1
	新竹市	18	18	18	18	18	18	18	18	18
	台南市	21	21	21	21	21	21	21	21	21
3	嘉義市	⑳	→	⑳	→	→	→	⑳	→	・
	桃園県	・	③	→	③	・	・	③	→	③
	嘉義市	・	・	・	・	・	・	・	・	⑳
	桃園県	③	→	③	→	③	→	③	→	③
	宜蘭県	2	2	2	2	2	2	2	2	2
	新竹県	4	4	4	4	4	4	4	4	4
	苗栗県	5	5	5	5	5	5	5	5	5
	台中県	6	6	6	6	6	6	6	6	6
	彰化県	7	7	7	7	7	7	7	7	7
	台南県	11	11	11	11	11	11	11	11	11
4	高雄県	12	12	12	12	12	12	12	12	12
	南投県	・	・	・	⑧	→	⑧	・	・	・
	台東県	・	・	・	⑭	→	⑭	・	・	・
	南投県	⑧	→	⑧	・	・	⑧	→	→	⑧
	台東県	⑭	→	⑭	・	・	⑭	→	→	⑭
	雲林県	9	9	9	9	9	9	9	9	9
4	嘉義県	10	10	10	10	10	10	10	10	10
	屏東県	13	13	13	13	13	13	13	13	13
	花蓮県	15	15	15	15	15	15	15	15	15
	澎湖県	16	16	16	16	16	16	16	16	16

(注) 表に示されている数字は、市県の動きを表わすために使用したもので、特に意味を持たない。

(出所) 筆者作成。

第7図 地理的な分布 (1989年)



(出所) 筆者作成。

断するため、1989年を例として地図に表わしたのが第7図である。第6図によれば、第1のグループを構成する地域の中で年に対して変化を示さないのは台北市のみである。1983年に台中市と高雄市が第2のグループからこのグループに移動してくるが、翌84年にはまた元のグループへ戻る。1985年から87年にかけて基隆市がこのグループに属している。地理的には南北の両極をなす主要都市によって構成されるグループである。第2のグループで年に対して変化を示さないのは台北県、新竹市、台南市である。台中市と高雄市は1983年のみ第1のグループに属すが、それ以外の年はこのグループに属して

第3表 産業界の就業者数構成比

(a) AIC最小化による地域分割

年	W1	W1	W3	W4	W5	W6	W7	W8	W9
1. 第1グループ									
82	1.84275	0.491400	23.4644	0.859951	7.86241	28.9926	8.35381	5.89681	22.2359
83	1.64512	0.470035	24.2068	0.940071	7.11804	28.5371	8.46063	6.11046	22.5617
84	1.83697	0.229621	23.0769	0.918485	7.11825	29.7359	8.216636	6.19977	22.6177
85	1.54867	0.221239	22.2345	0.884956	6.26555	30.6416	8.51770	6.09407	23.3407
86	1.35258	0.104275	22.7260	0.834202	6.35630	30.6589	8.44630	6.26552	23.18491
87	0.89242	0.094135	21.7824	0.185083	5.98486	31.0108	8.29438	7.88446	24.0452
88	0.82523	0.094833	19.8643	0.145143	7.52800	31.4272	7.52800	7.52800	22.7078
89	0.85238	0.093197	19.8909	0.145593	5.76459	35.3644	7.92111	8.38770	23.0196
2. 第2グループ									
82	7.47051	1.70380	36.0419	0.524246	8.65007	19.1350	7.68895	2.79598	15.9895
83	7.13689	2.04508	36.5192	0.417362	7.92989	20.0751	7.01169	2.79533	16.0684
84	5.89760	0.95962	38.7443	0.439824	7.23711	20.7517	6.87733	2.91883	16.2795
85	6.11929	0.81332	38.3811	0.503486	7.35864	20.9227	6.70023	2.90473	16.2645
86	6.75229	0.73394	39.1193	0.477064	7.15956	20.1101	6.56881	3.06587	16.0367
87	5.69247	0.55236	40.4026	0.451232	7.39327	19.6450	6.52551	3.01978	16.1338
88	4.18095	0.37697	39.4106	0.445511	7.57368	20.8362	6.51131	3.82263	17.0322
89	3.72671	0.35959	38.8562	0.424975	7.68223	21.4122	6.70134	4.18437	16.8721
3. 第3グループ									
82	23.4270	0.892459	36.3231	0.356983	7.71977	13.4315	4.28380	1.56180	12.0036
83	24.1510	0.818666	37.5131	0.301724	7.06897	12.8488	3.96552	1.46552	11.8103
84	23.5587	0.456242	40.2323	0.290336	6.59477	12.2356	3.56698	1.45168	11.6134
85	22.3887	0.374332	39.7836	0.332917	6.89956	13.0670	3.62047	1.62297	12.1099
86	22.2848	0.402253	40.2253	0.321802	6.35559	13.3146	3.78117	1.64294	12.6307
87	20.4305	0.391389	40.9045	0.391389	6.49706	13.1898	3.87476	1.76125	12.7593
88	18.1676	0.509005	40.7945	0.352388	6.96946	13.9389	3.91543	1.87940	13.2733
89	17.7357	0.463679	40.0696	0.270479	7.26430	14.4127	4.01855	2.35703	13.44030
4. 第4グループ									
82	39.0476	0.136054	22.0408	0.340136	8.29932	12.3129	3.46939	1.36054	12.8932
83	38.2883	0.76778	21.8390	0.401896	7.76282	13.7246	3.3742	1.33869	12.7891
84	39.7193	0.65662	21.5899	0.35309	7.15240	13.3317	3.4803	1.49424	14.1335
85	36.2996	0.128041	22.2793	0.256082	6.37803	13.4443	3.38309	1.79257	14.9203
86	33.9526	0.128123	24.2793	0.320307	7.36707	13.3733	3.58744	1.60154	14.9203
87	33.0769	0.128205	23.7179	0.320513	8.07692	14.8812	3.52564	1.85897	14.8077
88	31.8717	0.196335	23.8874	0.261780	8.24607	14.8560	3.59248	2.09424	14.9859

(出所) 筆者作成。

いる。基隆市は1982年から84年, 88, 89年はこのグループに属す。1983, 84年は桃園県が第3のグループから移動してきて, 85年に元に戻る。1982年から88年にかけてこのグループに属していた嘉義市が1989年に第3のグループへ移動する。このグループは台北県や基隆市を含む北部の主要都市と台北市

(b) 行政区分による地域分割

年	W1	W2	W3	W4	W5	W6	W7	W8	W9
1. 北部地区									
82	7.20439	1.91856	32.6155	0.537314	8.84886	20.3211	7.75255	3.52388	17.2979
83	7.50371	2.11738	33.6553	0.520059	7.91233	20.2823	7.23083	3.56612	17.1620
84	6.43777	1.00143	35.5508	0.536481	7.47496	21.1373	6.93848	3.71950	17.2031
85	6.03418	0.87199	34.7401	0.592954	7.32473	22.1137	7.03057	3.62149	17.6142
86	6.33454	0.75883	34.7410	0.560871	7.05033	21.9789	7.09337	3.76933	17.6839
87	5.27950	0.62112	35.5280	0.528750	7.32919	21.6770	7.08075	4.06832	17.8882
88	3.95257	0.45607	35.3907	0.516874	7.35786	23.1681	6.81058	4.52147	17.7258
89	3.72891	0.44432	34.8328	0.503107	7.16188	23.7052	7.04350	5.20864	17.4015
2. 中部地区									
82	26.6915	0.383142	33.3333	0.328607	7.17022	14.0120	3.77668	1.86037	12.6437
83	26.4737	0.421053	34.2632	0.315789	6.31673	14.2105	3.35842	1.78947	12.8421
84	24.8457	0.308642	36.8827	0.305642	6.21572	13.7953	3.18930	1.90041	12.8426
85	24.2132	0.253807	36.4467	0.304569	6.29244	14.1946	3.14721	1.97970	12.8426
86	23.5756	0.294695	37.3772	0.245580	5.92414	14.1946	3.14721	2.06257	13.0157
87	22.5434	0.192768	38.0058	0.395356	6.01119	14.1118	3.27002	2.33477	13.3430
88	21.2165	0.340633	37.0803	0.340633	6.42336	14.5985	3.74896	2.33477	13.2713
89	19.7312	0.336054	34.6338	0.288046	6.96111	15.2184	3.60058	2.88046	14.4503
3. 南部地区									
82	24.0367	0.366972	30.5881	0.458716	8.02752	15.7798	5.13761	1.76897	13.7156
83	23.5692	0.450653	30.8092	0.450653	7.16983	16.6742	4.91212	1.75755	14.1956
84	22.6543	0.217014	32.0313	0.477431	7.16983	16.6742	4.91212	1.73611	14.4531
85	24.0892	0.085727	31.0759	0.424633	7.07246	16.2886	4.75090	1.73611	14.4531
86	23.7521	0.124792	31.4050	0.373376	6.90516	16.0560	4.57130	2.39359	14.4021
87	20.9350	0.162602	33.6585	0.365884	7.03252	16.2502	4.59150	1.82157	14.7225
88	18.9385	0.160836	32.9715	0.402091	7.59952	17.2597	4.65825	2.37234	15.0439
89	18.0490	0.197472	32.2670	0.315956	7.89889	17.7330	4.89731	2.72512	15.9143
4. 東部地区									
82	36.7589	0.395257	16.9960	0.395257	9.8814	14.6245	4.74308	1.63103	14.6245
83	36.0000	0.400000	15.2000	0.400000	9.6000	15.0000	5.20000	1.60000	15.0000
84	38.3721	0.387597	17.4419	0.397597	9.2031	13.9535	4.26957	1.55039	14.3411
85	33.2031	0.390625	18.7500	0.390625	8.2031	14.9531	4.68750	1.79688	17.9688
86	28.7356	0.383152	17.6245	0.376284	8.4291	16.0920	5.74713	2.99895	19.9234
87	27.6923	0.384615	17.6923	0.759231	8.4615	16.5935	5.76923	2.00000	20.0000
88	26.7658	0.371747	17.8439	0.371747	9.9937	18.2156	4.83271	2.97393	19.9309
89	24.7402	0.333701	18.8976	0.367701	10.2362	18.1102	4.72641	2.86220	15.1517

(出所) 筆者作成。産業区分は、W1：農林業および狩猟業(畜産業を含む)、W2：鉱業および土石採取業、W3：製造業、W4：電気・ガス・水道業、W5：建設業、W6：商業、W7：運輸・倉庫・通信業、W8：金融・保険・不動産および加工サービス業、W9：社会および個人サービス業。

以外の市によって構成される。

第3のグループで年に対して変化を示さないのは宜蘭県、新竹県、苗栗県、台中県、彰化県、台南県、高雄県であり、第4のグループでは雲林県、嘉義県、屏東県、花蓮県、澎湖県である。1985年には第3のグループに属してい

た南投県と台東県が第4のグループへ移動し、翌86年に元へ戻る。第3グループは南北の主要都市の衛星圏を構成し、第4グループは山岳地帯や東部の地区を含む。

次に、地域区分ごとに産業間の就業者数の構成比を表わしたのが第3表の(a)である。これらの結果から各地域区分の特徴を見ると、農業に比較的高い比率を持つのは第3と第4のグループであり、その中でも第4のグループについては1982年には約40%を占めている。この2つのグループに比べて第1と第2のグループの農業の比率は極端に低い。農業の特徴はすべての地域区分において1982年から89年にかけて毎年減少の傾向にあることである。

製造業では第2と第3のグループの比率が高い。第3次産業の発達により第2次産業の比重は下降傾向にあるといわれているものの、製造業については第2、第3および第4のグループはわずかではあるが毎年増加する傾向にあるが、第1のグループのみは減少している。

第3次産業の中で比率の高い商業は農業とは対象的なパターンを示し、第1グループの比率が高く約30%を超え、次いで第2グループの約20%と続く。第3と第4グループはともに比率が低く12%から14%である。商業も毎年増加する傾向にあり、特に第1グループの増加は著しい。社会および個人サービスは相対的に第1グループの比率が高く、傾向としてはすべての地域区分でわずかではあるが増加している。

第3表の(b)には行政区分による地域分割の産業間の構成比も示されている。第2節からわかるように、台湾の地域格差は地域間よりも地域内の格差によって生じており、第3図の(b)に表わされた地区別の構成比も、実際には相当のばらつきがあると思われる。

## 第5節 AICの分布による行政区分の有意性

行政区分による地域は4つに分割されている。また第3節で述べたように

AICにもとづく階層的なクラスターでも地域区分を4にすることを根拠を与えている。そこで台湾地域の23市・県を任意に4つに統合したクロス表のAICから経験分布を求め、その分布から行政区分による地域分割がどのような位置にあるかを調べてみる。説明を簡単にするため、対象年を1982年と89年の2年間とする<sup>(5)</sup>。23の地域区分を任意に4つに統合した地域区分を $\{A_i(4) \mid i=1 \cdots 4\}$ として表わせば、統合した地域区分から得られるAICの値は、(1)式から、

$$\text{AIC} \{ A_1(4), A_2(4), A_3(4), A_4(4) \}$$

となる。

4つの地域区分は疑似一様乱数の1から4までを発生させ、この数字を市・県に付けられた一連番号に順次に対応させて統合を行うという方法を用いる。すなわち、 $U_i$ を $A_i$ に対応する一様乱数とすれば、統合された4つの地域区分は、

$$A_1(4) = \{ A_i \mid U_i = 1 \}, \dots, A_4(4) = \{ A_i \mid U_i = 4 \}$$

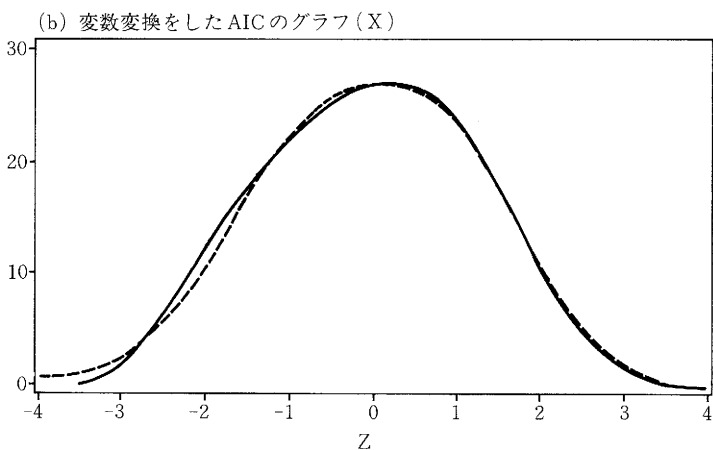
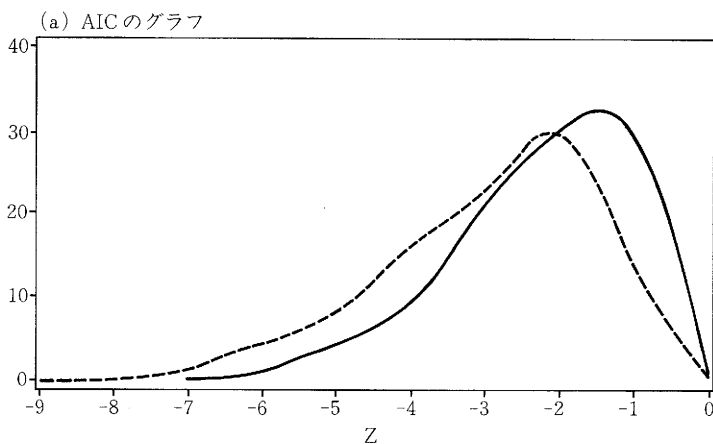
である。AICの値は各統合ごとに計算する。この計算を繰り返し行い、年ごとに900個の標本を取る。標本ごとに地域区分が23あるので、2年間の合計で41,400個の乱数を発生させたことになる。この内訳は1の表れた頻度が10,188 (24.6%)回、2の頻度は10,365 (25.0%)回、3は10,484 (25.3%)回、4は10,363 (25.0%)回であり、このことから均一的な地域の割り振りが行われたと見てよいであろう。

1982年と89年について、4つに統合したクロス表から得られたAICの分布が第8図の(a)に示されている。これらの分布は極端に左に偏りを持った形をしている。分布のピークを中央に移動させ、正規分布に近似させるため次のような変換を行う。

$$x = (| \text{AIC} |)^p$$

ここで $| \cdot |$ は絶対値を表し、 $p$ の値には正規分布の近似の程度を確かめるためのコロモゴロフ=スミノフのD統計量が小さくなる値を与える。その結果は第4表に示されているとおり、1982年では $p=0.328$ 、89年では $p=$

第 8 図



(出所) 筆者作成。(——82年, .....89年)

0.261の時にそれぞれ最小のDの値を持ち、正規性の仮説検定は棄却域の確率<sup>(6)</sup>が15%以上となり棄却されない。したがって、変換された変数 $x$ の分布は1982年では平均55.234と標準偏差10.3612を持つ正規分布、89年では平均26.1854と標準偏差3.5033を持つ正規分布に従ってほぼ分布していると仮定

第4表 D統計量

p	1982年		1989年	
	D	棄却域の確率	D	棄却域の確率
0.258	0.0227408	*	0.0260732	0.141
0.259	0.0226623	⋮	0.0259915	0.144
0.260	0.0225877	⋮	0.0259113	0.147
0.261	★0.0225143	⋮	0.0258307	0.15
0.262	0.0225560	⋮	0.0257508	*
0.263	0.0226568	⋮	0.0256696	⋮
0.264	0.0227572	⋮	0.0255897	⋮
0.265	0.0228591	⋮	0.0255091	⋮
0.266	0.0229597	⋮	0.0254287	⋮
0.267	0.0230597	*	0.0253477	⋮
0.324	0.0287354	0.071	0.0206688	⋮
0.325	0.0288339	0.069	0.0205854	⋮
0.326	0.0289325	0.067	0.0205016	⋮
0.327	0.0290305	0.065	0.0204180	⋮
0.328	0.0291287	0.064	★0.0203343	⋮
0.329	0.0292267	0.062	0.0204314	⋮
0.330	0.0293245	0.060	0.0205437	⋮
0.331	0.0294224	0.058	0.0206561	⋮
0.332	0.0295207	0.056	0.0207683	⋮
0.333	0.0296186	0.054	0.0208805	*

(注) ★はpの値を表のように指定した時、Dの最小値を示す点を表わす。

\*は棄却域の確率が15%以上を表わす。

(出所) 著者作成。

することができる。xの分布は第8図の(b)に示されている。これは行政区分による地域分割から得られたAICの有意性を正規分布によって検討することの根拠を与える。

行政区分による地域分割から得られたAICをxに変換しそれを標準化した結果が第5表に示されている。有意水準を10%とすれば、棄却域の確率は1982年および89年ともに0.10よりも小さいので、AICが平均値と等しくなるという帰無仮説は棄却される。すなわち、AICを基準にとると、行政区分による地域分割は有意となる。



第5表 行政区分による地域分割から得られたAICの標準化の結果

年	AIC	p	X	Z	棄却域の確率
82	-456,120	0.328	71.8088	1.5997	0.0548
89	-565,369	0.261	31.7129	1.5804	0.0570

(出所) 著者作成。

## 第6節 対数線形モデルによる交互効果

次に対数線形モデル(補論の「対数線形モデル」を参照のこと)を就業者数のデータに適用して行政区分による地域分割を中心に台湾の地域経済の特徴を検討する。大まかな産業のくくりのみを問題にするので、産業区分としては、第1次産業を農業部門、第2次産業を鉱業部門、製造業、電気・ガス・水道業部門、建設業部門、第3次産業を残りの業種とする。地域区分と産業区分に年を付け加え、3変数を使って対数線形モデルを試みる。年は1982年から89年までとする。

地域区分、産業区分、年をそれぞれ変数としてA, B, Cで表わす。標本数が多いので飽和モデルが最適なモデルとして選択される。主効果や交互効果のすべての変数の組み合わせが有意であり、その結果は第6表に示されている。この結果によると、地域区分の主効果は北部地区、中部地区、南部地

第6表 飽和モデルの結果

## (a) 地域区分の主効果

変数	北部	中部	南部	東部
A	0.5171	0.4466	0.6166	-1.5803

## (b) 産業区分の主効果

変数	第1次産業	第2次産業	第3次産業
B	-0.5361	0.2427	0.2934

(c) 年の主効果

変数	82	83	84	85	86	87	88	89
C	-0.0548	-0.0298	-0.0108	-0.0032	0.0297	0.0341	0.0200	0.0148

(d) 地域区分と産業区分の交互効果

	第1次産業	第2次産業	第3次産業
北部	-0.8759	0.3925	0.4834
中部	0.2198	0.0507	-0.2705
南部	0.1514	-0.0283	-0.1231
東部	0.5047	-0.4149	-0.0898

(e) 地域区分と年の交互効果

	82	83	84	85	86	87	88	89
北部	-0.0224	0.0170	-0.0077	-0.0100	0.0265	0.0298	-0.0231	-0.0101
中部	-0.0122	0.0018	-0.0050	0.0001	-0.0038	0.0050	0.0028	0.0114
南部	-0.0117	-0.0209	-0.0078	0.0053	0.0006	0.0009	0.0096	0.0240
東部	0.0463	0.0021	0.0204	0.0046	-0.0233	-0.0356	0.0107	-0.0252

(f) 産業区分と年の交互効果

	82	83	84	85	86	87	88	89
第1次産業	0.1229	0.1223	0.0876	0.0473	0.0147	-0.0624	-0.1554	-0.1770
第2次産業	-0.0398	-0.0620	-0.0079	-0.0189	-0.0201	0.0253	0.0541	0.0693
第3次産業	-0.0831	-0.0603	-0.0797	-0.0284	0.0054	0.0371	0.1013	0.1077

(g) 地域区分、産業区分および年の交互効果

		82	83	84	85	86	87	88	89
北部	第1次産業	0.0484	0.0780	0.0030	-0.0017	0.0657	0.0135	-0.0947	-0.1122
	第2次産業	-0.0273	-0.0115	-0.0140	-0.0067	-0.0287	-0.0010	0.0520	0.0371
	第3次産業	-0.0211	-0.0665	0.0110	0.0083	-0.0370	-0.0125	0.0427	0.0751
中部	第1次産業	-0.0199	-0.0197	-0.0379	-0.0227	-0.0129	0.0258	0.0642	0.0231
	第2次産業	-0.0249	-0.0000	0.0144	0.0151	0.0342	0.0107	-0.0214	-0.0280
	第3次産業	0.0448	0.0197	0.0235	0.0076	-0.0213	-0.0365	-0.0428	0.0050
南部	第1次産業	-0.0426	-0.0596	-0.0590	0.0344	0.0547	0.0240	0.0330	0.0151
	第2次産業	0.0188	0.0263	0.0019	-0.0253	-0.0186	0.0146	-0.0000	-0.0177
	第3次産業	0.0238	0.0333	0.0571	-0.0091	-0.0361	-0.0386	-0.0330	0.0026
東部	第1次産業	0.0141	0.0013	0.0939	-0.0100	-0.1075	-0.0633	-0.0025	0.0740
	第2次産業	0.0334	-0.0148	-0.0023	0.0169	0.0131	-0.0243	-0.0306	0.0086
	第3次産業	-0.0475	0.0134	-0.0916	-0.0068	0.0944	0.0876	0.0331	-0.0827

区および東部地区の係数はそれぞれ0.5171, 0.4466, 0.6166および-1.5803となっている<sup>(7)</sup>。対数線形モデルの係数は各変数間の変動が0となるように推計されるので、その解釈は相対的な変数間の変動になる。

$$\sum_{i=1}^4 u_i^A = \sum_{j=1}^3 u_j^B = 0$$

したがって係数の相対的な動きから、北部地区、中部地区、南部地区の3地区は就業者が多く集中しているのに対し、東部地区だけは逆に過疎になっていることがわかる。これらの中でも南部地区の係数が一番大きい。

産業区分の主効果から判断すれば、農業が-0.5361と減少傾向にあるのに対して、第2および第3次産業はともに係数がプラスの符号を持つため、集中していることがわかる。さらに第3次産業の係数0.2934に対して第2次産業の係数が0.2427であることから、第2次産業に比べて第3次産業の集中度合いが強いことが見て取れる。年の効果は産業区分や地域区分に比べて小さく、1982年以降では年による影響はそれほど顕著ではなかったと判断される。

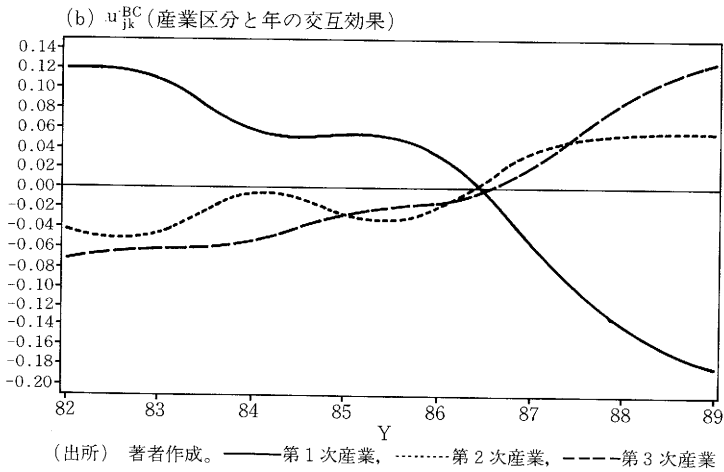
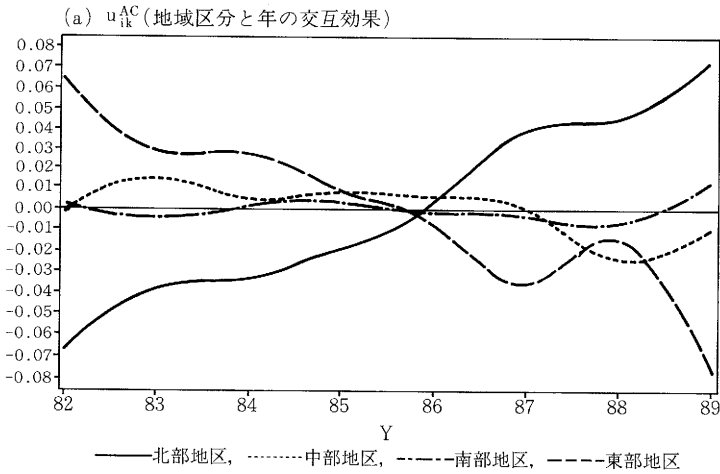
地域区分と産業区分の交互の関係を調べることにする。対数線形モデルでは交互作用の変動が0になるように定義されているので、係数の関係は、

$$\sum_{i=1}^4 u_{ij}^{AB} = \sum_{j=1}^3 u_{ij}^{AB} = \sum_{i=1}^4 \sum_{j=1}^3 u_{ij}^{AB} = 0$$

となる。相対的な関係として北部地区の農業の就業者数が極端に少なく、この部分を北部地区の第2、第3次産業あるいは中部や東部地区の農業が吸収していることが想定される。台湾地域全体では増加の方向にある第3次産業にも地域差が現われており、北部地区のみ集中が激しく他の3地区の就業者数の集中度合いは相対的に小さい。

地域区分と年、産業区分と年の交互効果は第9図に示されている<sup>(8)</sup>。第9図の(a)によれば、北部地区の就業者数の増加に対して、他の地区の増加の程度は極端に低い。同時に、同図の(b)からは第2次と第3次産業の増加に対して、第1次産業の減少の状態が明らかにされる。特に1987年以降、第2次産業の

第9図 対数線形モデルの係数の比較



上昇に歯止めがかかったのに対して第3次産業の増加は著しい。

### おわりに

台湾地域では都市計画および区域計画にもかかわらず、人口と産業活動の計画的な統合ができなかったことが、人口と産業活動の南北両極の集中と地域間格差の拡大をもたらした原因といわれている。これらの地域格差を改善するため、経済建設委員会では全国国土総合開発計画を策定し、実施してきた。しかしながら、地域格差は台北市を中心とする北部地区に集中する傾向があり、それがさまざまな構造的な問題を引き起こす要因となっている。

情報量による集中度によれば台湾経済の発展にともなって所得と就業者数共に地域格差は拡大する方向にあり、第3次産業にその傾向が強く現われている。しかもその拡大要因は第3次産業では地域間格差および地域内格差が同時に拡大しているが、第1次と第2次産業ではどちらかといえば地域間格差より地域内格差による影響と考えられる。就業者数は台北市を中心とする北部地区に集中しており、北部地区で就業者数の約40%を占める。対数線形モデルからも同じような結果が確認され第2および第3次産業の就業者数が第1次産業の分を吸収し、しかも地域間でその傾向に違いがあることが読みとれる。北部地区では第2および第3次産業の増加、第1次産業の減少がみられる。特に1986年以降では第2次産業の増加に歯止めがかかり、第3次産業の顕著な増加が特徴的である。南部地区と東部地区はともに第1次産業の比率が高く、東部の方が圧倒的にその比重が大きい。

AICにもとづくクラスターからわかることは、1980年代前半から87年にかけて台北市を中心に基隆市、台中市、高雄市で構成される主要都市がひとつのグループを作っていたことである。しかしながら1988年以降になるといまままで維持されてきた台北市と基隆市との間の類似点が薄れ基隆市は台北県や新竹市などのグループに属すようになる。これに対して台北市は他の地域と

は異なった産業構造のパターンを示すようになり、台北市のみでひとつのグループを構成する。このことから1980年代の前半は南北にまたがる両極化の傾向から、80年代の後半に向かって台北市の一極化へと変化してきたことが想定される。北部地区については集中している傾向が見られるものの台北市のみの就業者数の増加はそれほど著しい伸びを示さず、台北市の一極化がそのまま集中化であるという結論には至っていない。

最後に行政区分による地域の分割については対数線形モデルによりその違いは明らかであるが、23地域をいろいろ組み合わせるということを想定した時、行政区分による地域分割は統計的には有意であるという結果に達した。

[注] \_\_\_\_\_

- (1) 使用したデータは、行政院主計處発行の以下の統計書に記載されている。1982年のデータは『中華民國勞工統計月報 特載：中華民國勞動力調査專案補充研究分析 1982』から得た。1983年から86年までは『中華民國勞工統計月報 (特載), 中華民國七十三年, 勞動力調査專案補充研究分析』1984~86年。1987年から89年までは『中華民國七十六年臺地區人力資源統計年報』台北 1988年である。データの連続性は『中華民國七十八年臺地區人力資源統計年報』台北 1990年に記載されている年次別の人口総額が各統計書にある総額と一致することで確かめることができる。
- (2) 所得分布のデータは、Directorate-General of Budget, Accounting & Statistics, Executive Yuan, Republic of China, *Report on the Survey of Personal Income Distribution in Taiwan Area, Republic of Taiwan* に記載されている。1976年から86年まではこの資料の中の“Distribution of Personal Income by Areas,” 1987, 88, 89年は“Personal Distributed Factor Income by Areas”の各表中の“Primary Income”を使用した。金額の単位はNT\$ 1,000である。
- (3) 台湾地域の行政区分による地域分割は1981年以前までは、台北市と高雄市の2つの行政院直轄市のほか台湾地区として、基隆市、台中市、台南市の3市、台北県、宜蘭県、桃園県、新竹県、苗栗県、台中県、彰化県、南投県、雲林県、嘉義県、台南県、高雄県、屏東県、台東県、花蓮県、澎湖県の16県の合計21地区に分かれていたが、その後、新竹市と嘉義市の2市が行政区分で分割されたため、2直轄市、5市、16県の23地区となっている。地域区分についてはここにあげたもの以外に金門地区があるが、この地区は特殊な地区で

あるため今回の分析の対象からは除外した。

- (4)  $n$  個の観測値を  $x = |x_1 \cdots x_n|$  とする。観測値の平均を  $\mu$ , 分散を

$$V(x) = n^{-1} \sum_{i=1}^n (x_i - \mu)^2$$

とした時、分散の平方根を平均で割った値  $V(x)^{1/2} / \mu$  を変動係数という。また、ジニ係数は、

$$G(x) = (2n^2 \mu)^{-1} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|$$

で定義される。

- (5) 産業区分と年の 2 変数を考慮した対数線形モデル

$$\log m_{ij} = u + u_i^A + u_j^C + u_{ij}^{AB} \\ (i = 1 \cdots 3, j = 1 \cdots 8)$$

の結果からわかるように1982年から89年までの間で82年と89年の産業区分が極端に区別されることからこの2年を選んだ(結果は省略した)。

- (6) 得られたデータが正規分布からのサンプルであるかどうかの仮説検定は Kolomogorov-Sminov の  $D$  統計量が用いられるが、ここでは SAS のプロシジャー UNIVARIATE に従って、標本サイズが50より大きい場合には、データの分布は平均と分散が標本平均と標本分散に等しい正規分布に等しいという仮説が検定される。検定統計量の有意確率は次のようにして得られる。まず、

$$\left\{ n - 0.01 + 0.85 / (n^{1/2}) \right\}^{1/2} D$$

が計算される。ここで  $n$  は有効データ数である。この値を用いて、ステイブンスがシミュレーションで求めた棄却限界値の範囲で線形補間することによって有意性が検定される (Stephens, M. A., "EDF Statistics for Goodness of Fit and Comparison," *Journal of the American Statistical Association*, 1974年)。

- (7) 対数線形モデルが,  $\log m_{ijk} = u + u_i^A + u_j^B$

と  $A$  と  $B$  の主効果のみの場合、北部、中部、南部および東部地区の係数は0.7709, 0.3658, 0.5402および-1.6069となる。

- (8) 地域区分×年の交互効果は対数線形モデルの、

$$\log m_{ijk} = u + u_i^A + u_k^C + u_{ik}^{AC}$$

を用いた。産業区分×年の交互効果も同じような方法を用いた。

## 〔補論〕

## AIC にもとづく分割表のモデル化

母集団から得られた標本の各要素が、第1の特性についてR個の級、第2の特性についてC個の級に分類されるとき、この2つの特性に関して同時に観測度数を求めたものをR×C分割表とよぶ。分割表の2つ特性を変数AとBでそれぞれ表わし、各特性はA×Bに分かれているとする。変数AとBの特性を $A = \{A_1 \cdots A_R\}$ 、 $B = \{B_1 \cdots B_C\}$ とそれぞれ分割する。分割なので、 $A_i \cap A_j = \phi$ 、 $B_i \cap B_j = \phi$  ( $i \neq j$ )となる。すなわち共通集合は空集合である。変数がAがi番目の級、変数Bがj番目の級をとる時、その確率を $\Pr \{A_i \times B_j\} = p_{ij}$ 、 $i = 1 \cdots R$ 、 $j = 1 \cdots C$ とする。また、それに対応する観測度数を $n_{ij}$ で表わす。Nを母集団から得られた標本数とすると、

$$\sum_{i=1}^R \sum_{j=1}^C p_{ij} = 1, \quad \sum_{i=1}^R \sum_{j=1}^C n_{ij} = N$$

となる。標本数Nが多ければ、確率の集合 $\{p_{ij}, i = 1 \cdots R, j = 1 \cdots C\}$ のもとで観測度数 $\{n_{ij}\}$ が得られる確率は多項分布に従って分布すると仮定することができる。観測値が得られた時、パラメータ $p_{ij}$ に関するこのモデルの対数尤度は、

$$\begin{aligned} \log L(\{p_{ij}\}) &= \sum_{i=1}^R \sum_{j=1}^C n_{ij} \log p_{ij} \\ &+ \sum_{i=1}^R \sum_{j=1}^C \log n_{ij}! + \log N! \end{aligned}$$

と表わすことができる。この時、AICは次のように定義される。

$$AIC = -2 (\text{分布の最大対数尤度}) + 2 (\text{自由パラメータの数})$$

最適分布の選択はK-L情報量と同じようにAICの小さな値を持つ分布が良い分布になるという考えをもとに行う。このモデルの最大対数尤度は、多項分布の(R-1)(C-1)個の自由パラメータの最尤推定値を対数尤度へ代入した



ものであるから、このモデルの AIC は以下ようになる。

$$AIC(A_1 \cdots A_R) = -2 \sum_{i=1}^R \sum_{j=1}^C n_{ij} \log(N n_{ij}) / (n_{i.} n_{.j}) + 2(R-1)(C-1) \quad \cdots(1)$$

ここで、 $n_{i.} = \sum_{j=1}^C n_{ij}$ 、 $n_{.j} = \sum_{i=1}^R n_{ij}$  である。このモデルに対応する AIC の値を比較することによって、最適なモデルを選ぶことができ、一番多くの情報量を有する  $R \times C$  の分割表を決定することができる（坂元慶行『カテゴリカルデータのモデル分析』共立出版 1985年）。

### 対数線形モデル

地域区分、産業区分および年をそれぞれ変数として A, B, C で表わすことにする。A のカテゴリーの数を  $N_A$ 、B の数を  $N_B$ 、C の数を  $N_C$  とした時、3 変数の飽和对数線形モデルは次のように表わすことができる。

$$\log m_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC} \quad \cdots(2)$$

ここで  $i = 1 \cdots N_A$ 、 $j = 1 \cdots N_B$ 、 $k = 1 \cdots N_C$  であり、3 重クロス表のセル  $(i, j, k)$  の期待頻度を  $m_{ijk}$  とする。この時、すべてのセルの平均を

$$u = (N_A N_B N_C)^{-1} \sum_i \sum_j \sum_k \log m_{ijk}$$

とする。特性  $A_i$  についての主効果は、 $A_i$  に属する期待頻度の対数  $\log m_{ijk}$  の平均から全平均の差として定義される。

$$u_i^A = (N_B N_C)^{-1} \sum_j \sum_k \log m_{ijk} - u$$

特性 B と C の主効果も同じように定義される。主効果だけでは説明できない部分を交互効果という。交互効果 AB は

$$u_{ij}^{AB} = N_C^{-1} \sum_k \log m_{ijk} - (u_i^A + u_j^B) - u$$

となる。交互効果 BC と AC についても同様である。2 変数の場合は 2 変数の交互効果までですべての効果を説明できるが、3 変数になると主効果と 2 変数の交互効果でも説明がつかない部分が残る。この残差を 3 次の交互効果として表し、

$$u_{ijk}^{ABC} = \log m_{ijk} - (u_{ij}^{AB} + u_{jk}^{BC} + u_{ik}^{AC}) \\ - (u_i^A + u_j^B + u_k^C) - u$$

で定義する。飽和モデルは以上のすべての組み合わせを用いた補論の(2)の式で表わされる。データを適切に説明するモデルは主効果や交互効果の有意性を判断することにより階層的な方法により決めることができる(松田紀之『質的情報の多変量解析』朝倉書店 1988年)。