

Total-Scoreモデルにおける探索的方法について

小島 秀夫*

（1992年10月7日受理）

On an Exploratory Method for Total-Score Models

Hideo KOJIMA

（Received October 7, 1992）

はじめに

本稿の目的は、最近Goodmanによって開発されたTotal-Scoreモデルにおいて、最適モデルを発見する方法を示すことである¹⁾。Goodmanは、Total-Scoreモデルを展開した論文においてさまざまなモデルを提示し、それらのモデルの適合度を問題にしているが、それらのモデルの改良をどのような情報によって行えばよいのかということについては言及していない。たとえばGoodmanは、表1に示されるようなデータにおいて、複数ある変数のなかでAの変数のカテゴリーの順序を逆にした場合のモデルの適合度を問題にしている。しかしながら、問題になるのはなぜ変数Bあるいは変数Cではなく変数Aのカテゴリーの順序を逆転すればよいのかということである。さらに複雑なモデルを考える場合に、その根拠をどこに求めればよいのかは、困難な問題である²⁾。本研究では、その目的のために数量化2類を使用することを提示する。まず初めにTotal-Scoreモデルの論理構造を明らかにし、数量化2類を使用することの妥当性について検討しよう。

Total-Scoreモデルと数量化2類

Total-Scoreモデルを具体的に説明するために、表1のデータを使用することとする。表1のデータは、大学生を対象に過去1年間に「レストランに食事に出かけた」・「ゴルフ・テニスをした」・「二泊以上の国内旅行に出かけた」・「芝居見物・コンサートなどに出かけた」ことの有無を調べ、それを男女別に集計したものである³⁾。表1で0は「したことがない」を示し、1は「したことがある」を示している。たとえば、4つのうちにも「したことがない」人は、男女ともに4人であることが分る。4つの項目（A, B, C, D）への反応は（i, j, k, l）（i=0,1；j=0,1；k=0,1；l=0,1）と表わすことができる。したがって、反応パターン（i, j, k, l）の総合スコア（total-score）

*茨城大学教育学部社会情報研究室（〒310 水戸市文京2丁目1-1）。

表1 男女別にみた行動様式

反応パターン				性別(E)	
A	B	C	D	男性	女性
0	0	0	0	4	4
0	0	0	1	2	2
0	0	1	0	2	3
0	0	1	1	0	1
0	1	0	0	2	1
0	1	0	1	1	1
0	1	1	0	1	1
0	1	1	1	3	2
1	0	0	0	19	22
1	0	0	1	19	28
1	0	1	0	48	72
1	0	1	1	62	106
1	1	0	0	29	32
1	1	0	1	23	28
1	1	1	0	80	78
1	1	1	1	156	188

A: 「レストラン」 B: 「ゴルフ・テニス」 C: 「国内旅行」 D: 「コンサート」
 0 = 「なし」 1 = 「あり」

は、 $t=i+j+k+l$ と表わすことができる。表1は、男女別の4重クロス表とみなすこともできるし、 $\{A, B, C, D, E\}$ の5重クロス表とみなすこともできるが、 $\{ABCD, E\}$ の2重クロス表とみなすことも可能である。2重クロス表 $\{ABCD, E\}$ の場合、Total-Scoreモデルは、

$$F_{ijkls} = NP_{ijkl}^{ABCD} P_{s|t}^E \quad (t=i+j+k+l) \quad (1)$$

で表わされる。ここで、 F_{ijkls} はセル (i, j, k, l, s) の期待度数を示し、 N はサンプル数、 P_{ijkl}^{ABCD} はある個人が項目 (A, B, C, D) に対して (i, j, k, l) の反応をする確率を示す。 $P_{s|t}^E$ は、項目 (A, B, C, D) に対する総合スコア t ($t=i+j+k+l$)が与えられた場合に、ある個人が項目 E に関して S と反応する条件付き確率を示す。(1)式は、

$$F_{ijkls} = \eta \tau_{ijkl}^{ABCD} \tau_s^E \tau_{st}^{ET} \quad (2)$$

と書きかえられる。(2)式で τ_{st}^{ET} は変数 E と $ABCD$ 間の非独立を示し、その非独立は変数 $ABCD$ から得られる総合スコア t に依存していることを示す。

(1)式あるいは(2)式で示されるものがTotal-Scoreモデルのなかでもadditモデルとよばれるものであるが、このモデルの意味を理解するために(1)式の一部を次のように書きかえることとする。

$$P_{s|ijkl}^E = P_{s|t} \quad (t=i+j+k+l) \quad (3)$$

ここで $P_{s|ijkl}$ は、項目 (A, B, C, D) に対してある個人が (i, j, k, l) の反応をした場合に、項目 E に

対する個人の反応がSである条件付き確率を示す。さらに、

$$\gamma_s^E = \tau_s^E / \tau_{s+1}^E, \quad \gamma_{st}^{ET} = \tau_{st}^{ET} / \tau_{s+1,t}^{ET}$$

とすると、(2)式の一部は、

$$\Omega_{ijkl}^E = \Omega_{st}^E \quad (t=i+j+k+l) \quad (4)$$

のようにも書きかえられる。ここで Ω_{st}^E は、項目(A, B, C, D)に対する個人の総合スコアtが与えられた場合に、項目Eについて個人がS+1であるよりもSであるという条件付き確率を示す。したがって、換言すれば(i, j, k, l)への反応パターンによって個人が変数Eに関して、S+1であるかSであるかが決定されるということである。この考えは、数量化2類の考えと基本的に同じである。

ここで数量化2類についての数式による記述は行わないが、数量化2類とは説明変数の各カテゴリーに任意の数値を与えて、質的なデータとして与えられている外的基準のグループ分けを最大限にしようとする方法である⁴⁾。そこで、このTotal-Scoreモデルと数量化2類の類似性に注目し、数量化2類による分析結果を利用して、Total-Scoreモデルの改良を行うというのがここでの考えである。以下に、実際の適用例を示すこととする。

実 例 1

ここでは、表1のデータを分析してみよう。表1のデータの数量化2類による分析結果が表2に示されている。数量化2類では、分類の程度の良さを示すものとして相関比(η^2)の大きさが問題とされるが⁵⁾、ここでの目的のためには相関比の値の大きさは問題にしなくともよい。

表2に示された結果についてみてみよう。まず最初に偏相関係数の値に注目すると、項目(B)の「ゴルフ・テニス」が0.079と相対的に大きく、ついで項目(D)の「コンサート」が0.039となっている。項目(A)「レストラン」と項目(C)「国内旅行」の偏相関係数は、それぞれ0.021, 0.012と小さい。全体的に偏相関係数の値は小さいけれども、ここでの目的のためにはその点は問題にならず、注目すべき点は、偏相関係数の相対的な大きさである。

各カテゴリーに与えられるノーマライズド・スコアに注目すると、項目(B)「ゴルフ・テニス」において「なし」のスコアが-1.107, 「あり」のスコアが0.697となり、他の項目の場合とは異なり、プラスとマイナスが逆になっているのが注目される。この項目(B)以外の項目では「なし」にプラスの値が、「あり」にマイナスの値が与えられている。ここで、偏相関係数の相対的な大きさと、各カテゴリーに与えられたノーマライズド・スコアの符号のパターンに注目し、Total-Scoreモデルを設定してみよう。実際の分析ではGLIMが使用されているが、そのプログラムが本論文の末尾が示されている。

表3に、Total-Scoreモデルの分析結果が示されている。モデル H_0 は変数{ABCD}と変数{E}が独立であることを示している。モデルの適合度は L^2 (尤度比統計量)の大きさによって決定される

が、モデルH₀は自由度が15であるのに対し、L²は10.54でうまくデータに適合している。実際、表3に示されたそれぞれのモデルのAICを求めた結果でも、AICの値が-19.46で最小であり、このモデルが最適モデルであることが明らかにされる⁹⁾。しかしながら、ここでの目的はTotal-Scoreモデルを使用することであるから、分析をさらに進めることとする。モデルH₀は(1)式で示されるモデルである。モデルH₀が成立すれば、より多くのパラメータのあるモデルH₁は必ず成立する。モデルH₁の意味は、変数 {ABCD} と変数 {E} の非独立は総合スコア t に依存しているということである。換言すれば、総合スコア t のレベルが与えられた場合に、変数 {ABCD} と変数 {E} は独立であるということの意味している。このことは $[ABCD \otimes E | T_1]$ と表わされる⁷⁾。モデルH₁のL²は9.62となり、自由度は11でありモデルの適合度は良いが、モデルH₀と比較した場合に自由度の差は4であるのに対し、L²の差はわずか1にすぎず、モデルが改良されたとは認められない。

表2 数量化2類による分析結果

項 目	カテゴリー	ノーマライズド・ スコア	偏相関係数
(A) レストラン	なし	1.362	0.021
	あり	-0.041	
(B) ゴルフ・テニス	なし	-1.107	0.079
	あり	0.697	
(C) 国内旅行	なし	0.260	0.012
	あり	-0.070	
(D) コンサート	なし	0.548	0.039
	あり	-0.351	

$\eta^2=0.008$

表3 Total-Scoreモデルによる分析結果

モデル	独立条件	d.f.	L ²	P
H ₀	[ABCD \otimes E]	15	10.54	>.500
H ₁	[ABCD \otimes E T ₁]	11	9.62	>.500
H ₂	[ABD \otimes E T ₂]	20	415.1	<.001
H ₃	[ABCD \otimes E T ₃]	11	3.43	>.500
H ₄	[ABCD \otimes E T ₄]	7	8.68	>.500
H ₅	[ABCD \otimes E T ₅]	7	2.55	>.500
H ₆	[ABD \otimes E T ₆]	15	31.03	>.010
H ₇	[ABCD \otimes E T ₇]	9	2.92	>.500

モデルH₂は、数量化2類による分析で偏相関係数の値が一番小さかった変数Cを分析から除去したものであるが、L²の値はきわめて大きくモデルの適合度は悪い。変数Cを分析から除去することで、総合スコアの値も変わる。モデルH₃は、変数(B)「ゴルフ・テニス」のノーマライズド・スコアの符号に注目し、表1の0と1を逆転したものである。その点以外はモデルH₁と同じであるが、モデルH₁と同じ自由度であるが、L²は3.43ときわめて良い適合度を示している。モデルH₄は表2の偏相関係数の値に注目し、偏相関係数の値の似た変数どうしをまとめ、2次元の総合スコアを設定したものである。すなわち、

$$\Omega_{\text{aijkl}}^E = \gamma_{\text{a}}^E \gamma_{\text{b}}^E \gamma_{\text{c}}^E \gamma_{\text{d}}^E \quad (u = i+k, v = j+l) \quad (5)$$

と表わされるモデルである。ここでは偏相関係数の値がほぼ等しい項目Aと項目Cを1次元とし、項目Bと項目Dを他の次元とした。したがって、総合スコアは(0,0)、(0,1)、(1,0)などのように与えられる。このモデルの適合度もL²が8.68であり、非常に良い。モデルH₅はモデルH₄と基本的に同じ2次元の総合スコアを設定したものであるが、項目(B)のスコアを逆転した後、総合スコアを設定したものである。モデルH₅はモデルH₄よりもデータに適合している。

モデルH₆はモデルH₂と関連するモデルであるが、モデルH₆では項目(B)の数値が逆転されている。モデルH₆は統計的に有意でありモデルは成立しないが、モデルH₂と比較すると自由度の差が5であるのに対して、L²の差は415.1-31.3=383.8となり、項目(B)のカテゴリーの順序の影響の大きさが明らかにされる。モデルH₇はモデルH₅の特殊なものであり、v=j+l=2の場合にuの値はどのようなものでもよいというモデルである⁹⁾。このモデルもうまくデータに適合している。

ここでの分析の結果を要約しておこう。ここでの分析では各カテゴリーのノーマライズド・スコアの符号に注目することと、偏相関係数の大小に注目することによって、モデルの改良を行うことが可能であることが明らかにされた。特に、項目(B)のカテゴリーの順序を逆転することによって適合度が著しく改良されることが明らかにされた。

実 例 2

次の例についてみてみよう。表4は、階層帰属意識別にみた老後の生活についての不安感を調べたもので、1992年6・7月に調査が実施されたものである⁹⁾。変数Aは「経済的に安定して暮せるかどうかの不安」、変数Bは「大きな病気をしたり、事故に会う不安」、変数Cは「友人や家族から孤立してしまう不安」、変数Dは「生きがいがなくなってしまう不安」を示している。表4で0は「感じない」、1は「感じる」である。階層帰属意識は「中の上」と「上」を「上」にまとめ、「中の下」以下を「下」にまとめてある。

表5に数量化2類による分析結果が示されている。ここでも相関比は0.022と小さい。偏相関係数についてみると、項目(A)「経済的不安」が0.123と相対的に高い。項目(C)「孤立不安」と項目(D)「生きがい不安」の偏相関係数は、それぞれ0.062と0.053となっている。項目(B)「病気・事故不安」の偏相関係数は0.008となっている。ノーマライズド・スコアについてみると、項目(A)と項目(C)

が同じ方向に、項目(B)と項目(D)が同じ方向になっている。すなわち、項目(A)と項目(C)では「感じない」にマイナスの値が、「感じる」にプラスの値が与えられている。これに対して項目(B)と項目(D)については、「感じない」にはプラスの「感じる」にはマイナスの値が与えられている。

この数量化2類による分析結果を利用してTotal-Scoreモデルをあてはめた結果が、表6に示されている。モデル H_0 は、変数{ABCD}と変数{E}の独立を示すものであるが、 L^2 が33.37であり統計的に有意であり、モデルは成立しない。モデル H_1 は総合スコア t ($t=i+j+k+l$)を使用したモデルであるが、ここでも L^2 が23.34でありモデルは成立していない。

表4 階層帰属意識別にみた不安感

反応パターン				階層帰属意識	
A	B	C	D	上	下
0	0	0	0	52	132
0	0	0	1	20	59
0	0	1	0	30	51
0	0	1	1	82	191
0	1	0	0	3	5
0	1	0	1	4	7
0	1	1	0	3	14
0	1	1	1	15	41
1	0	0	0	7	11
1	0	0	1	1	6
1	0	1	0	6	4
1	0	1	1	42	53
1	1	0	0	1	0
1	1	0	1	0	4
1	1	1	0	6	3
1	1	1	1	36	44

A：「経済的不安」 B：「病気・事故の不安」
 C：「孤立してしまう不安」 D：「生きがい不安」
 0 = 「感じる」 1 = 「感じない」

表5 数量化2類による分析

項	目	カテゴリー	ノーマライズド・スコア	偏相関係数
(A)経済的不安		感じない	-1.13	0.123
		感じる	3.59	
(B)病気・事故不安		感じない	0.07	0.008
		感じる	-0.26	
(C)孤立不安		感じない	-1.56	0.062
		感じる	0.78	
(D)生きがい不安		感じない	1.25	0.053
		感じる	-0.68	

$\eta^2=0.022$

表6 Total-Scoreモデルによる分析結果

モデル	独立条件	d.f.	L ²	P
H ₀	[ABCD ⊗ E]	15	33.37	<.01
H ₁	[ABCD ⊗ E T ₁]	11	23.34	<.05
H ₂	[ACD ⊗ E T ₂]	15	36.45	<.01
H ₃	[ABCD ⊗ E T ₃]	11	22.58	<.05
H ₄	[ABCD ⊗ E T ₄]	8	8.40	>.05
H ₅	[ACD ⊗ E T ₅]	20	500	<.01
H ₆	[ABCD ⊗ E T ₆]	9	11.38	>.05

モデルH₂は、項目(B)の偏相関係数が小さいことを考慮し、項目(B)を除去して総合スコアt (t=i+k+l) をあてはめたモデルである。このモデルも統計的に有意であり成立していない。モデルH₃は、項目(C)「孤立不安」の数値0,1を逆転したものであるが、このモデルもL²が22.58で統計的に有意であり、成立していない。モデルH₄は偏相関係数の大きさに注目し、項目(B)(C)(D)をまとめたものである。さらに項目(C)の数値を逆転したものである。すなわち、

$$\Omega_{S|ijk}^S = \gamma_{S|ijk}^E \gamma_{SUV}^{SUUV} \quad (u=i, v=j+k+l) \tag{6}$$

と表わされるモデルである。このモデルの下でのL²は8.40となり、自由度は8であるからモデルは成立する。AICも-7.6で、表6に示されたモデルの中では最小値を示している。したがって、階層帰属意識の「上」「下」を分けるものは1次元の総合スコアではなく、2次元の総合スコアであることが明らかにされる。

モデルH₅はモデルH₂とほぼ同じものであるが、モデルH₅では項目(C)の数値が逆転されている点が異なる。このモデルH₅のL²は500でモデルは成立していない。モデルH₆はモデルH₄においてj+k+l=3の場合に、iのレベルの情報が必要ないというものであり¹⁰⁾、L²は11.38でありモデルは成立している。しかしながら、このモデルのAICは-6.12であり、モデルH₄の下でのAICよりは大きい。

要約と結論

本研究は、Total-Scoreモデルにおいて最適モデルを発見するために数量化2類によって得られる情報を利用することを試みたものである。分析の結果、以下のような点に注目することによって、モデルの改良が効率的に行われることが明らかにされた。

(1)数量化2類によって得られる偏相関係数の値に注目し、似た値の項目をまとめることによって、モデルの適合度は良くなる。

(2)同時に、各カテゴリーに与えられたノーマライズド・スコアの符号に注目することによって、各項目における各カテゴリーの符号が同じになるように、変数の数値を逆転することによってモデルの適合度は良くなる。

ここでの例では、各変数が2分類のものしか扱われていないが、3カテゴリー以上の場合も同様

に扱うことができる¹¹⁾。変数の数が増加し、カテゴリー数が増加するにつれて、数量化2類によって得られる情報はより役立つものと考えられる。

注

- 1) Total-Scoreモデルについては、次の文献を参照せよ。Goodman, L. 1990. "Total-score models and Rasch-type models for the analysis of a multidimensional contingency table, or a set of multidimensional contingency tables, with specified and/or unspecified order for response categories" In : Clogg, C. C. (ed), *Sociological Methodology 1990*, Basil Blackwell, Oxford. なお、Total-ScoreモデルのGLIMによるモデル構成については、次の文献を参照せよ。小島秀夫、1992. 「Total-ScoreモデルとGLIM」『茨城大学教育学部紀要（人文・社会科学，芸術）』41, PP.125-132.
- 2) 理論が明確な場合には最適モデルを理論によって求めればよいが、社会調査で得られるデータの場合には理論による方法は困難である。
- 3) この調査の一部は、以下で報告されている。小島秀夫・篠原清夫、1985. 「大学生の職業意識形成過程の研究—教育学部生の職業意識の分析—」『茨城大学教育学部紀要（教育科学）』34, PP.281-296.
- 4) 数量化2類については、以下の文献を参照せよ。岩坪秀一、1987. 『数量化法の基礎』（朝倉書店）。駒澤勉、1982. 『数量化理論とデータ処理』（朝倉書店）。駒澤勉・橋口捷久、1988. 『パソコン数量化分析』（朝倉書店）。
- 5) 数量化2類を実際に分析に使用する場合は、ここでの分析よりも多くの変数が使用されるのが普通であり、したがって相関比の値も大きくなる。
- 6) この結果からすると、ここでの例は適切なものではないといえる。したがって、ここではモデル探索のプロセスの例示としての意味しかないといえる。
- 7) これらの表記の方法については、次の文献を参照せよ。Andersen, E. B. 1990. *The Statistical Analysis of Categorical Data*. Springer-Verlag, New York.
- 8) すなわち、 $v=2$ の場合には項目Aと項目Cについての情報は不必要である。ということの意味している。
- 9) この調査は、茨城県からの委託で茨城大学教育学部情報教育研究室が実施したものである。調査目的は、茨城県男性の社会意識を明らかにすることである。この調査について報告書等ははまだ出されていない。
- 10) すなわち、 $v=3$ の場合には項目Aに関する情報は不必要であることを意味している。
- 11) ただし、組合せが複雑になりすぎ、コンピュータの容量の関係で実際に計算が可能かどうかは分からない。また、モデルが複雑になった場合の解釈は困難になる。

以下に、实例1で使用したGLIMの使用例を示しておく。

```

% C THIS IS GLIM RUNSTREAM FOR TOTAL-SCORE MODELS
% UNITS 32
% C READ DATA
% DATA DATA
% READ
4 2 2 0 2 1 1 3 19 19 48 62 29 23 80 156
4 2 3 1 1 1 1 2 22 28 72 106 32 28 78 188
% C READ ITEM A
% DATA A
% READ
1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
% C READ ITEM B
% DATA B
% READ
1 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2
1 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2
% C READ ITEM C
% DATA C
% READ
1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2
1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2
% C READ ITEM D
% DATA D
% READ
1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
% C READ ITEM E
% DATA E
% READ
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
% C READ INVERSED ITEM B
% DATA B1
% READ
2 2 2 2 1 1 1 1 2 2 2 2 1 1 1 1
2 2 2 2 1 1 1 1 2 2 2 2 1 1 1 1

```

¥C READ TOTAL-SCORE

¥DATA T

¥READ

1 2 2 3 2 3 3 4 2 3 3 4 3 4 4 5

1 2 2 3 2 3 3 4 2 3 3 4 3 4 4 5

¥C READ TOTAL-SCORE WHEN ITEM C IS DELETED

¥DATA T1

¥READ

1 2 1 2 2 3 2 3 2 3 2 3 3 4 3 4

1 2 1 2 2 3 2 3 2 3 2 3 3 4 3 4

¥C READ TOTAL-SCORE WHEN ITEM B IS INVERSED(ALL ITEMS INCLUDED)

¥DATA T2

¥READ

2 3 3 4 1 2 2 3 3 4 4 5 2 3 3 4

2 3 3 4 1 2 2 3 3 4 4 5 2 3 3 4

¥C READ TOTAL-SCORE WHEN (ITEM A+ITEM C, ITEM B+ITEM D)

¥DATA T3

¥READ

1 2 4 5 2 3 5 6 4 5 8 7 5 6 7 9

1 2 4 5 2 3 5 6 4 5 8 7 5 6 7 9

¥C READ TOTAL-SCORE WHEN (ITEM A+ITEM C, INVERSED ITEM B+ ITEM D)

¥DATA T4

¥READ

2 3 5 6 1 2 4 5 5 6 7 9 4 5 8 7

2 3 5 6 1 2 4 5 5 6 7 9 4 5 8 7

¥YVARIATE DATA

¥ERROR P

¥LINK L

¥FACTOR A 2 B 2 C 2 D 2 E 2 B1 2 T 5 T1 4 T2 5 T3 9 T4 9

¥C FIT THE MODEL A.B.C.D+E.T

¥FIT A.B.C.D+E.T

¥DISPLAY E M D R

¥C FIT THE MODEL A.B.D+E.T1

¥FIT A.B.D.+E.T1

¥DISPLAY E M D R

¥C FIT THE MODEL A.B1.C.D+E.T2 (ITEM B IS INVERSED)

¥FIT A.B1.C.D+E.T2

¥DISPLAY E M D R

¥C FIT THE MODEL A.B.C.D+E.T3(ITEM A+ITEM C,ITEM B+ITEM D)

¥FIT A.B.C.D+E.T3

¥DISPLAY E M D R

¥C FIT THE MODEL A.B.C.D+E.T4(ITEM A+ITEM C,INVERSED ITEM B+ITEM D)

¥FIT A.B1.C.D+E.T4

¥DISPLAY E M D R

¥C FIT THE MODEL A.B1.D+E.T2(ITEM C IS DELETED AND ITEM B IS INVERSED)

¥FIT A.B1.D+E.T2

¥DISPLAY E M D R

¥STOP