

影響関数を用いた構造方程式モデルによるセグメンテーション方法の検討*

Segmentation Method Using Influence Function in Structural Equation Model*

寺部慎太郎**・岩井俊英***

By Shitnaro TERABE**・Toshihide IWAI***

1. はじめに

交通計画や都市計画などの公共事業において、本来は事業主体と周辺住民の意識は一致しているべきであるが、実際には、計画の受益者兼負担者である市民と、市民のために計画を立案・執行する行政の間で計画への意識に差が生じていることが多い。また、近年は市民の価値観は多様化し、計画者が市民の意識を簡単には把握しづらい状況になっている。そこで意識調査などを行い市民の意識を調べることで良く行われるが、それに基づいた計画や施策を行政が打ち出しても、それに対する市民の反応は一樣ではなく、効果が予想よりも小さくなってしまいうことも少なくない。

そこで、ある部分で共通な考え方を持った人でグループ分け（セグメンテーション）をすれば、人々の意識を把握しやすくなるのではないかと、そしてそれぞれのグループにターゲットを絞って働きかければ、全体に同じ方法で働きかけるよりも成果を得ることができるのではないかと考えられる。

本研究では、意識調査の分析に良く用いられる構造方程式モデルから得られる意識構造を基準に、何らかの意味で同質な人ごとにセグメンテーションをする方法について検討することを目的とする。その際には、セグメンテーションの指標として、影響関数から得られる値を用いることを試みた。

2. 影響関数を用いたセグメンテーション方法

(1) 影響関数

影響関数は、主に回帰分析などのロバスト推定のために導入されたもので、各観測値がモデルの推定結果にどのような影響を与えているかを数値化して、その結果から外れ値を探すために用いられる¹⁾。構造方程式モデルではその適合度関数が、 $G(\underline{s}, \hat{\underline{\theta}})$ と表されるとき ($\underline{s} = \text{vech}(S)$: 観測変数の分散共分散行列 S の対角成分を含む下三角行列の要素をベクトル化したもの)、個人 i の観測変数ベクトル $\underline{z}_i$ の、推定パラメータベクトル $\hat{\underline{\theta}}$ に対する影響関数は、

$$\text{EIF}(\underline{z}_i; \hat{\underline{\theta}}) = - \left[\frac{\partial^2 G(\underline{s}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T} \right]^{-1} \left[\frac{\partial^2 G(\underline{s}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{s}^T} \right] \text{EIF}(\underline{z}_i; \underline{s})$$

となる。ただし、

$$\text{EIF}(\underline{z}_i; \underline{s}) = \text{vech}[(\underline{z}_i - \bar{\underline{z}})(\underline{z}_i - \bar{\underline{z}})^T - S]$$

である²⁾。これは、あるサンプルを推定母集団から外すときにパラメータがどう変動するかを定義したものであるが、寺部・屋井³⁾は推定する対象に別のサンプルを新たに加える際のパラメータの変動も同様に定義できると考え、 $\underline{z}$ に含まれない個人 j の観測変数ベクトル $\underline{y}_j$ について

$$\text{EIF}_{\Phi}(\underline{y}_j; \hat{\underline{\theta}}) = - \left[\frac{\partial^2 G(\underline{s}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T} \right]^{-1} \left[\frac{\partial^2 G(\underline{s}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{s}^T} \right] \text{EIF}_{\Phi}(\underline{y}_j; \underline{s})$$

ただし、

*キーワード：意識調査分析，市民参加

**正会員，博（工），東京大学大学院工学系研究科社会基盤工学専攻（〒113-8656文京区本郷7-3-1，TEL:03-5841-6135，FAX:03-5841-8506，E-MAIL:shin@trip.t.u-tokyo.ac.jp）

***学生会員，東京大学大学院工学系研究科社会基盤工学専攻

$$\text{EIF}_{\text{ap}}(\underline{y}_j; \underline{s}) = \text{vech}[(\underline{y}_j - \bar{\underline{z}})(\underline{y}_j - \bar{\underline{z}})^T - S]$$

とした．上の影響関数から各個人の各パラメータに対する影響係数が得られる．

(2) セグメンテーション指標

ここで，分析者が着目した部分的な構造（ある一つのパスや，複数のパス，変数間の相関など）の影響係数をもとにセグメンテーションをすることになるが，本研究ではモデル全体への適合度に着目してセグメンテーションを行うため， χ^2 （カイ二乗）値とGFI（適合度指標）を取り上げた．

a) χ^2 値

χ^2 値は，サンプル数Nと構造方程式モデルの適合度関数 $G(\cdot)$ とで計算されるが，あるサンプルを取り除いたときに χ^2 値がどれだけ変化するかを調べ（変化量を $\Delta \chi^2$ とする），セグメンテーションを行う．ここで $\Delta \chi^2$ 値を次のように定める．

$$\Delta \chi^2 = (N-2)G(S_{(i)}, \hat{\Sigma}(\hat{\theta}_{-(i)})) - (N-1)G(S, \hat{\Sigma}(\hat{\theta}))$$

(i) : i 番目のサンプルを落としたときの値

χ^2 値は，定義より小さいほどモデルを支持することを意味するので， $\Delta \chi^2$ の値が大きくなれば，そのサンプルは χ^2 値を小さくするものである，即ち，そのサンプルはモデル支持ということになる．

では， $\Delta \chi^2$ の値がいくつ以上になったときにサンプルを支持とすればよいだろうか．高田・屋井・寺部⁴⁾は，0を基準とし，0以上ならばそのサンプルをモデル支持，0未満ならばモデル不支持としていた．しかし，ほとんどのサンプルは $\Delta \chi^2$ の値が0付近に集中しているため，これらのサンプルの中から支持サンプルのみを取り出すのは決して容易なことではない．そこで今回は， $\Delta \chi^2$ の基準を0ではなく χ^2/N （N：サンプル数）とする．これを基準としたのは以下の理由からである．

この式を変形すると， $\Delta \chi^2 / \chi^2 > 1/N$ となる．この不等式の左辺は，各サンプルが抜けることにより χ^2 値を何％増加（減少）させたかを表すものである．高田らの方法だと， χ^2 の値が100の時の $\Delta \chi^2 = 5$ と， χ^2 の値が20の時の $\Delta \chi^2 = 5$ は同じ価値であったが，今回提案した方法ならば， $\Delta \chi^2$ の

値が同じであっても χ^2 の値が異なれば価値が異なる（すなわち χ^2 の値が大きくなるほど価値が小さくなる）という意味を含ませることができる．

また，右辺を $1/N$ とすれば，サンプルの数が増えるほど，一つのサンプルの全体に与える影響が小さくなることを表現することができる．

b) GFI

χ^2 値は，サンプル数に影響されやすく，サンプル数が多くなると χ^2 値が増加し，悪いモデルだと判定されやすくなってしまいう性質がある．この欠点を補うためにつくられたのがGFI(Goodness of Fit Index)である．GFI は次の式で表される．

$$\text{GFI} = 1 - \frac{\text{tr}((\hat{\Sigma}(\hat{\theta}))^{-1}(S - \hat{\Sigma}(\hat{\theta})))^2}{\text{tr}((\hat{\theta})^{-1}S)^2}$$

ただし，

S：観測変数の分散共分散行列

$\Sigma(*)$ ：モデルパラメータの分散共分散行列

θ ：モデルパラメータ

である．

GFI は最大値が1になるようにできており，1に近いほどモデルがよいことを表す．また，GFI は負の値もとりうる．ここで，先ほどの χ^2 値と同じようにサンプルを一つ取り除いたときにGFI がどれだけ変化するかを調べ，（ ΔGFI とする）セグメンテーションを行う．

$$\Delta \text{GFI} = \text{GFI}_{(i)} - \text{GFI}$$

ただし， $\text{GFI}_{(i)}$ は i 番目のサンプルを取り除いたときのGFIを表す．GFI は， χ^2 値と反対で大きな値ほどよいわけだから， ΔGFI の値が負になればそのサンプルはモデルを支持することになる．

よって，各サンプルの ΔGFI を求めて，それが負の値（0以下）であればモデル支持で，正の値（0以上）であれば不支持ということになるが，これも χ^2 値による方法のところで述べたのと同じ理由から，0を基準とするのは適当でないものと思われる．また，今回の研究で求めた各サンプルの ΔGFI は，大きさが1よりも非常に小さかったため， χ^2 値の場合と同じような方法でセグメンテーショ

ンをすることができない．そのため，セグメンテーションの基準を $-GFI / (10 \times N)$ とした．

(3) セグメンテーションの手順

- セグメンテーションは以下の手順で行われる．
- ①意識調査の実施と基礎分析
 - ②構造方程式モデルによる意識構造モデルの作成（複数）と，セグメンテーションの基準となるモデルの選定
 - ③セグメンテーションの指標の算出
 - ④閾値の設定

①で得られた意識調査のデータ及びその基礎分析の結果から，②で意識構造モデルを作成する．普通の分析の場合は，ここで複数のモデルを構築し，それらの適合度などを比較検討し，最も良いモデルを1つ採用するが，本研究の手法では，その複数のモデルをセグメンテーションの基準とするため，いくつかに絞ることが行われる．（後述のケーススタディでは3つのモデルに絞られた．）次に③では2．（2）で述べたセグメンテーション指標を算出し，④でその指標の閾値を設定する事によって，セグメンテーションが終了する．

3. ケーススタディ

(1) 使用したデータ

本研究では，長期交通計画策定に対する市民の参加意識を調べた調査⁵⁾を題材にした．この調査は横浜市内在住の市民を対象にして家庭訪問留置回収方式で1997年11月に行われたもので，配布数が2257票，回収数が1873票，回収率は83%である．本研究では必要な設問への無回答などを除いた1044サンプルを使用した．

(2) 作成した構造方程式モデル

モデルは7つの共通因子（「地域活動への参加」「社会への関心の高さ」「環境意識」「他人への思いやり」「行政依存」「定住・永住」「計画関連事項に関する認知状況」）の組み合わせから作成され，各モデルの適合度や類似性から以下の3つがセグメンテーションの基準モデルとして選択された．

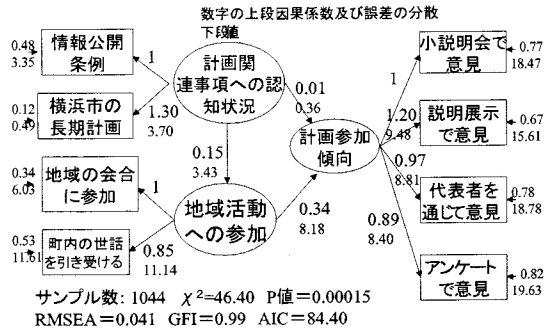


図1 地域活動モデル

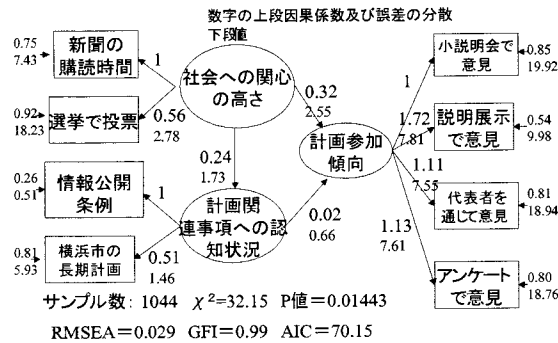


図2 社会関心モデル

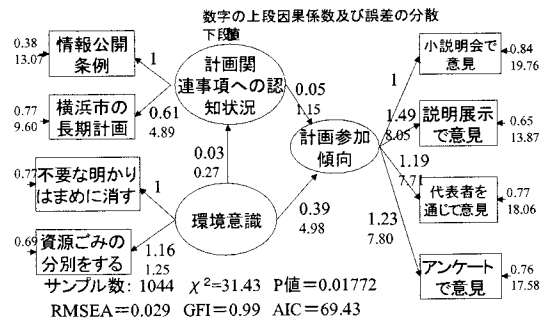


図3 環境意識モデル

(3) セグメンテーションの指標の算出

それぞれのモデルについて，セグメンテーション指標である $\Delta \chi^2$ と ΔGFI が算出された．各サンプルがこの2つの指標を持つので，その散布図の一部を図4，図5に示す．このように，両指標の性質が現れていることがわかる．また，この散布図は対象となるモデルによって異なるため，そのモデル間の関係を見るためにも有用であろう．

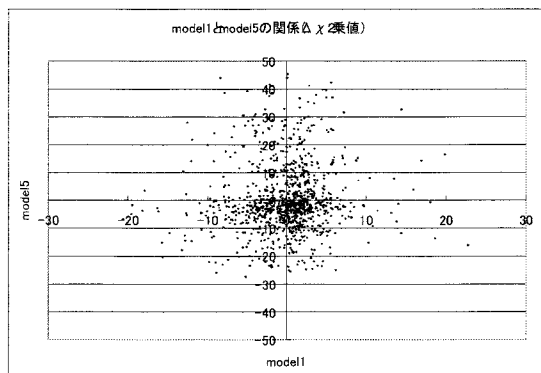


図4 地域活動モデルと社会関心モデルの関係
($\Delta \chi^2$ の散布)

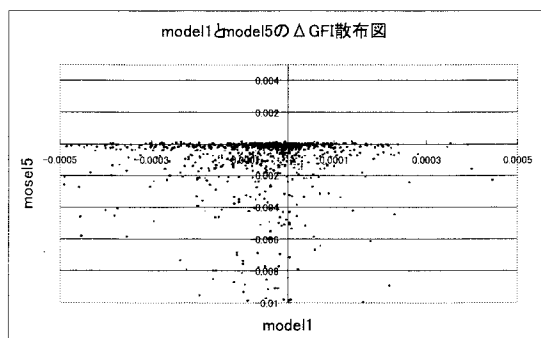


図5 地域活動モデルと社会関心モデルの関係
(ΔGFI の散布)

(4) 形成されたセグメンテーション

閾値を前述のように設定し、その値により各サンプルがそれぞれのセグメンテーション（モデル）に帰属するか否かが決定され、状況によっては複数のモデルに重複して帰属することもあり得る。得られたセグメンテーションをベン図にして表したものが図6、図7である。

4. おわりに

本研究の目的は、意識構造に基づいたセグメンテーション方法の問題点を解決することを意識しながら、その方法を検討することであった。今回は2通りのセグメンテーション方法を検討し、それぞれの方法について意味のあるセグメンテーションができるかどうか調べてみた。それぞれのセグメンテーション方法の適用可能性については次の通りである。

χ^2 値による方法 対象全サンプル サンプル数: 1044

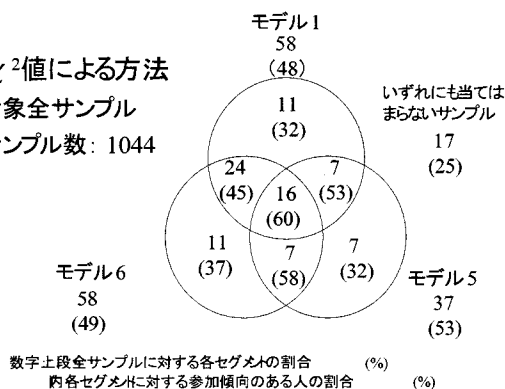


図6 得られたセグメンテーション
($\Delta \chi^2$ による)

GFIによる方法 対象全サンプル サンプル数: 1044

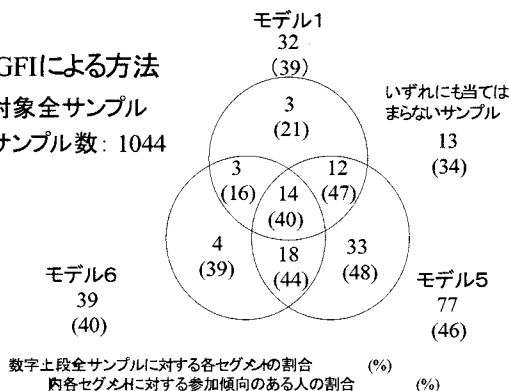


図7 得られたセグメンテーション
(ΔGFI による)

χ^2 値による方法では、所属セグメント数が多いほど計画への参加傾向が高くなる傾向にあり、これもセグメンテーション方法として利用できる可能性があることが分かった。GFIによる方法では、各サンプルの ΔGFI の値が非常に小さかったために、サンプル間の違いがはっきりせず、結局望ましい結果を得ることはできなかった。

参考文献

- 1) 田中豊：「多変量解析における感度分析」，行動計量学，19巻1号，pp.3-17，1992
- 2) Tanaka et al.: "Sensitivity Analysis in Structural Equation Models", COMPSTAT, pp.493-498, 1992.
- 3) 寺部慎太郎・屋井鉄雄：「共分散構造モデルを用いた内生的なセグメンテーション手法」，日本行動計量学会第24回大会発表論文抄録集，pp.102-103，1996
- 4) 高田伸二・屋井鉄雄・寺部慎太郎：「交通計画に対する市民意識の構造分析」，土木学会第53回年次学術講演会講演概要集，pp.212-213，1998
- 5) 寺部慎太郎・屋井鉄雄・関健太郎：「長期交通計画策定に対する市民参加意識の分析」，土木計画学研究・論文集，No.16，pp.161-166，1999