

地域別経済指標に基づく静岡SDモデルの開発：
地域統計データによる地域経済圏の統計的分析

メタデータ	言語: jpn 出版者: 公開日: 2012-05-18 キーワード (Ja): キーワード (En): 作成者: 上藤, 一郎, 浅利, 一郎, 山下, 隆之, 高瀬, 浩二 メールアドレス: 所属:
URL	https://doi.org/10.14945/00006675

地域別経済指標に基づく静岡SDモデルの開発 —地域統計データによる地域経済圏の統計的分析—

上藤一郎・浅利一郎・山下隆之・高瀬浩二

はじめに

筆者たちは、県内地域別の政策課題にアプローチするため、過去数年にわたりシステムダイナミクス (System Dynamics, 以下SD) による独自のモデル開発に取り組み、開発された「静岡システムダイナミクス・モデル」(静岡SDモデル) を用いて様々な地域経済分析を行ってきた。¹⁾その研究経緯を簡単に振り返ると、先ず2007年に最初の全県版の静岡SDモデル(I)を完成させ、それを用いて県内各産業の経済成長やその成長が静岡県の環境問題に与える影響などを分析した。その後この静岡SDモデルは、2010年および2011年に更なる改訂を加え、データ更新や経済セクターの設計変更を行うなどしてより精度を高めてきた。²⁾

こうした静岡SDモデルの開発とそれによる分析に並行して、筆者たちは分析の基礎となる地域統計データの整備にも力を注いできた。その成果の一つが、総務省統計局による「オーダーメイド集計」のサービスを利用して、既存の公的統計では得られない地域表章の集計データ(国勢調査再集計)を活用したことである。より具体的には、労働移動から見た静岡県内における経済圏を数量化IV類によって分類し、更にその分析結果の精度を検討したことであり、これらについては次のような成果を得ることができた。

先ず県内経済圏の統計的分類については、労働力移動の点から見た各市町間の経済的相互依存関係を分析するため、親近性の「近さ」を分析する統計手法(数量化IV類およびクラスター分析)を用いて、一つの新しい知見を見出すことができた。これまで静岡県では、現行の総合計画(2006年)に基づく伊豆半島、東部、中部、志太榛原・中東遠、西部の5地域区分を基準として、各地域の特徴や課題を明らかにしてきた。この5地域区分は、主に歴史的な地域の結びつきを反映した行政上の区分であるが、それに対し筆者達は前述の手法を用いて再検討を行い、静岡県内をより経済的な結びつきの強い5地域区分(伊豆、東部、富士・富士宮、中部、西部地域)もしくは6地域区分(伊豆、東部、富士・富士宮、中部、志太榛原・中東遠、西部地域)に再編することができることを明らかにした。更にそこで得られた知見に基づき、新しい6地域区分の人口動態および経済成長の将来予測を行うため、静岡SDモデル(II)を開発しその研究成果を公表した。

一方、オーダーメイド集計データの精度については、上藤・浅利・山下・高瀬[14]で明らかにしている。ここでいう精度とは、オーダーメイド集計データを利用することに起因する「誤差」を意味しており、それは統計表章上静岡県内における市町村間労働移動者数を10人単位で集計・表示することから生じる乖離を意味する。再集計されたデータにこのような誤差が含まれることは、データ保護を目的とする公的統計の法的制約により不可避である。またこのような誤差は、ランダムに発生しているとは言い難く、それ故誤差の正規性を仮定した推論が困難である。誤差のランダム性、正規性を仮定できない以上、「分析結果の精度」を検討するにはオーソドックスな統計的推論の方法に依拠することができず、このため上藤・浅利・山下・高瀬[7]では、この精度問題を解決する一つの試みとして、感度分析を行い数量化IV類による分析結果の変化を検討した。

¹⁾ SDモデルの特徴や詳細については、例えば、Yamashita [15], 山下 [16], 山下・高瀬 [18] 等を参照。

²⁾ この研究経緯については山下・上藤・高瀬 [19] を参照。

そこで本稿では、このような一連の研究経緯の総括として、上藤・浅利・山下・高瀬 [19] で行った数量化Ⅳ類による統計的分類とその感度分析の概要を示す。³⁾その際、これまで取り上げなかった地図情報を付加して、視覚的な方法による分析結果の要約を試みる。次いで、これらの分析で用いられたデータの比率補正法を取り上げる。前述の統計的分類に用いられるデータは、基本的には各市町間の労働移動率が用いられるが、小地域区分で「比率」を用いる場合、異なる人口規模による比較可能性の問題が不可避免的に発生する。すなわち、人口規模や労働力人口の規模が異なる場合の比率による比較である。例えば、人口規模や労働力人口の規模が大きい地域と小さい地域の労働移動率を比較する場合、見かけ上の数値上は同じであっても必ずしも同列に扱ってよいというものではない。そこで本稿では、人口規模に見合う比率に補正する試みとして経験的ベイズ推定法を取り上げる。経験的ベイズ推定法はもともと人口動態分析でしばしば用いられる補正法であるが、本稿では筆者たちが問題とする地域経済分析への応用拡張とその問題点を検討する。

1. 労働力移動から見た静岡県内における経済圏の統計的分類

一般に静岡県内における経済圏は、伊豆半島、東部、中部、志太榛原・中東遠、西部の5地域に分類される。しかしながら、県内地域経済の特性をより的確に分析するには、経済圏や文化圏の異なる志太榛原・中東遠をさらに分離し6地域に区分する方が、実態を反映するという意味で有効である場合も多い。そこで筆者たちは、これら5地域あるは6地域が労働力移動の点でおのおの自立した経済圏を形成しているかどうか統計データを用いて検証した。利用したのは2005年における国勢調査のオーダーメイド集計データで、そのデータを数量化Ⅳ類とクラスター分析に適用することによって、静岡県における経済圏の統計的分類を試みた。具体的には、各市区町村間の労働力（総数）の流出・流入において各市区町村間の親近性（依存度）を計測し、更に親近性のある市町村をグループ別けることによって、相互依存関係の強い市町村からなる経済圏を分類した。先に指摘した静岡6地域がこの分析結果に一致すれば、その地域分類の妥当性が労働力移動の点で検証されたことになる。この検証のため、先ず数量化Ⅳ類を適用し、親近性の「近さ」を2次元の座標軸においてスケージングした。更に、得られた2次元の数量化スコアをクラスター分析によって分類し、相互依存関係の強い市町村同士をグループ化した。

1.1 地図情報による分析結果の概要

分析結果を述べると、先に述べた静岡県による5地域分類、あるいはより細分化された6地域分類と完全には一致していないが、その不一致にはある共通したパターンが示された。具体的に述べると、第一に、どちらの分類においても、志太榛原地域と中東遠地域が静岡市を中心とする中部地域と浜松市を中心とする西部地域に分断され、少なくとも労働力の流出入という点では自立した経済圏を確立していないことが推察できた。一部異なる市町村も見られるが、概ね志太榛原地域の市町村が中部地域、中東遠地域の市町村が西部地域と相互依存関係を結んでいることが確認された。

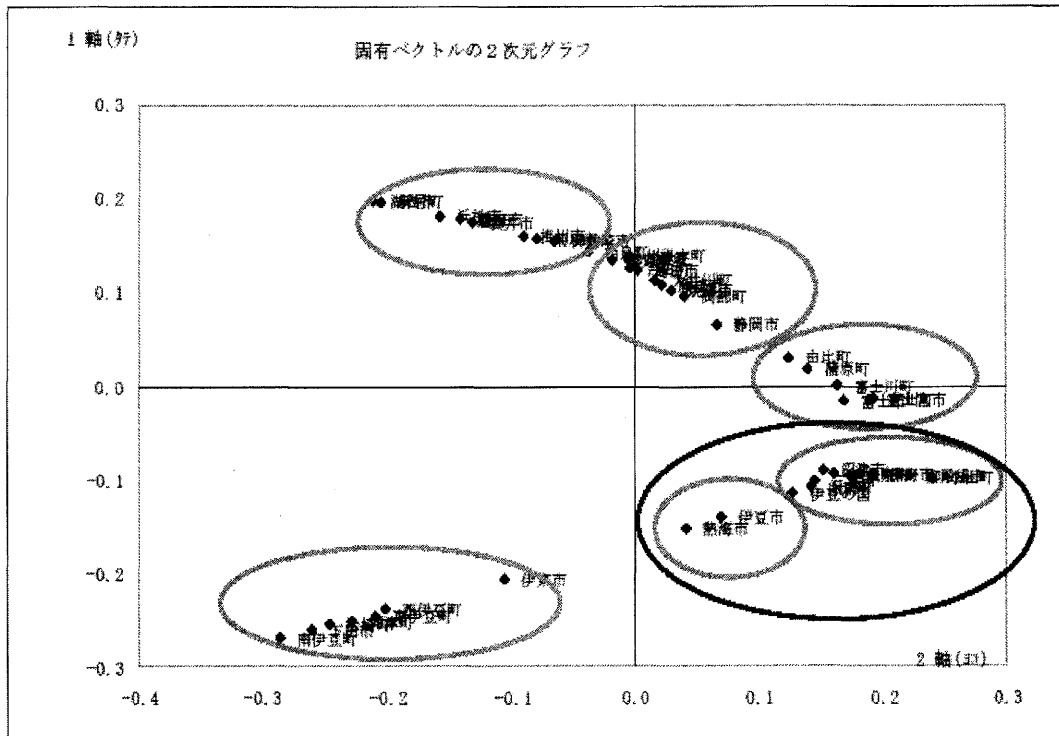
第二の特徴として、東部地域が伊豆地域における一部の市町村と依存関係にあることが示された。しかしその反面、富士市や富士宮市を中心とする地域が東部地域と中部地域（由比町）から分離して、独自の経済圏（富士・富士宮地域）を形成していることも確認できた。この結果、伊豆地域は前述のように東部地域との依存関係を有する市町村、伊豆地域独自の経済圏を形成している市町村に分離していることが判明した。

³⁾ 感度分析の詳細な結果については上藤・浅利・山下・高瀬 [19] を参照。

なお付言すれば、ここで用いた数量化Ⅳ類と呼ばれる分析手法は、親近性をどのような指標で計測するかによって分析結果が異なる可能性がある。筆者たちもこの点を考慮し、いくつかの指標によって試算を行っている。⁴⁾しかしながらいずれの指標においても、志太榛原地域・中東遠地域の分散化と富士市・富士宮市を中心とする「富士・富士宮地域」の自立性については、同様の結果が得られ従来分類とは異なる独自の知見を見出すことができた。

そこでこれらの結果を地図情報で視覚的に確認する。それに先立ち、数量化Ⅳ類とクラスター分析による地域分類の結果を図1「固有ベクトルの2次元グラフ」で見てください。

図1 固有ベクトルの2次元グラフ



数量化Ⅳ類を適用した数量化Ⅳ類とは、林知己夫によって開発された外的基準のない多変量データ解析の方法であり、数学的には計量的多次元尺度構成法の一つであると見なし得る。今、 n 個の対象において任意の対象 i と j の間の親近性(依存度)を表す指標を e_{ij} とする。但し $i, j=1, 2, \dots, n, i \neq j$ であり、 e_{ii} ($i=j$)は考慮されない。また e_{ij} は正負の値をとり得るどのような統計量であってもよいが、 e_{ij} が大きいほど親近性が高く、逆に小さいほど親近性が低くなるよう定義されなければならない。ここで対象 i についてある未知量 x_i を付与し

⁴⁾ 新規性の尺度の定義とそれらの計算結果については、山下・上藤[18]および山下・上藤・高瀬 [19] を参照。

$$Q = -\sum_i \sum_j e_{ij} (x_i - x_j)^2 \quad (i \neq j) \quad (1)$$

を、平均 $\frac{\sum_i x_i}{n} = 0$ 、分散 $\frac{\sum_i x_i^2}{n} = 1$ の下で最大化する $x_i, (i=1,2,\dots,n)$ を求める。その結果得られた固有ベクトル x を対象 i の座標軸とし、この尺度に基づく座標値 $x_i, (i=1,2,\dots,n)$ ⁵⁾ によって対象間の相互依存関係を把握することが数量化IV類の目的である。これは即ち、親近性の尺度で測られた数値を n 次元ユークリッド空間における点と見なし、これら各点の距離から見た分布特性、つまり類似した対象間の距離は近く、類似していない対象間の距離は遠くに布置するような特性を、固有ベクトルで尺度化された1次元もしくは2次元の座標軸から把握することを意味する。⁶⁾

図1はこの分析結果を示している。これは、数量化IV類によって求められた第1固有ベクトルと第2固有ベクトルに対応する各市町の数量スコアを示した散布図であるが、更にクラスター分析によって5分類と6分類した各市町のグループを楕円で類別している。この図を見ると、市町名が重なって少々判別しにくいものの、大きく二つの特徴が明らかになっていることがわかる。

一つは、前述の通り5分類および6分類で見ても、志太榛原地域と中東遠地域が静岡市を中心とする中部地域と浜松市を中心とする西部地域に分断され、少なくとも労働力の流出入という点では自立した経済圏を確立していないことである。また東部地域が伊豆地域における一部の市町村と依存関係にあること、その一方で富士市や富士宮市を中心とする地域が、東部地域と中部地域（由比町）から分離して独自の経済圏（富士・富士宮地域）を形成していることも示されている。更に6分類で見ると、5分類では東部地区に含まれる一部の伊豆地域（熱海市、伊豆市）が分離することもわかる。

もう一つの特徴は、グループ分けされた市町の位置情報が第1軸（第1固有ベクトル）に対応している点である。第1軸の上に位置するグループから、西部地域、中部地域、富士・富士宮地域、東部地域、伊豆地域の順に布置されており、第1軸の上・下が静岡県内市町の西・東に対応していることがわかる。つまりそれは、第1固有ベクトルが地理情報を表す軸として解釈可能であることを示している。

以上の分析結果を、図2～5の地図情報で確認しておこう。図2は、静岡県が総合計画（2006年）で採用している5分類を、更に志太榛原地域と中東遠地域に細分して6分類にしたものである。図3は、図1で見た数量化IV類による2次元グラフを地図上で示したものである。この図3の地図情報に、労働移動の手段となる交通網の情報を加えたものが図4と図5で、図4は鉄道網の情報を加えたもの、図5は図4に加えて幹線道路の情報を加えたものである。これらを見れば、数量化IV類による地域分類が交通網と対応していることが理解できよう。

⁵⁾ この座標値を数量化理論では「数量化スコア」と呼ぶ。

⁶⁾ 数量化IV類を始めとする数量化理論の基本的な考え方や論理については、開発者自身による林[2]・[19]を参照。また数量化IV類は、理論的には多次元尺度構成法の一つのヴァリエーションで見なすこともできる。このような観点から数量化IV類と多次元尺度構成法の関係を論じた文献については、齊藤・宿久[6]を参照。

図2 静岡県6地域分類

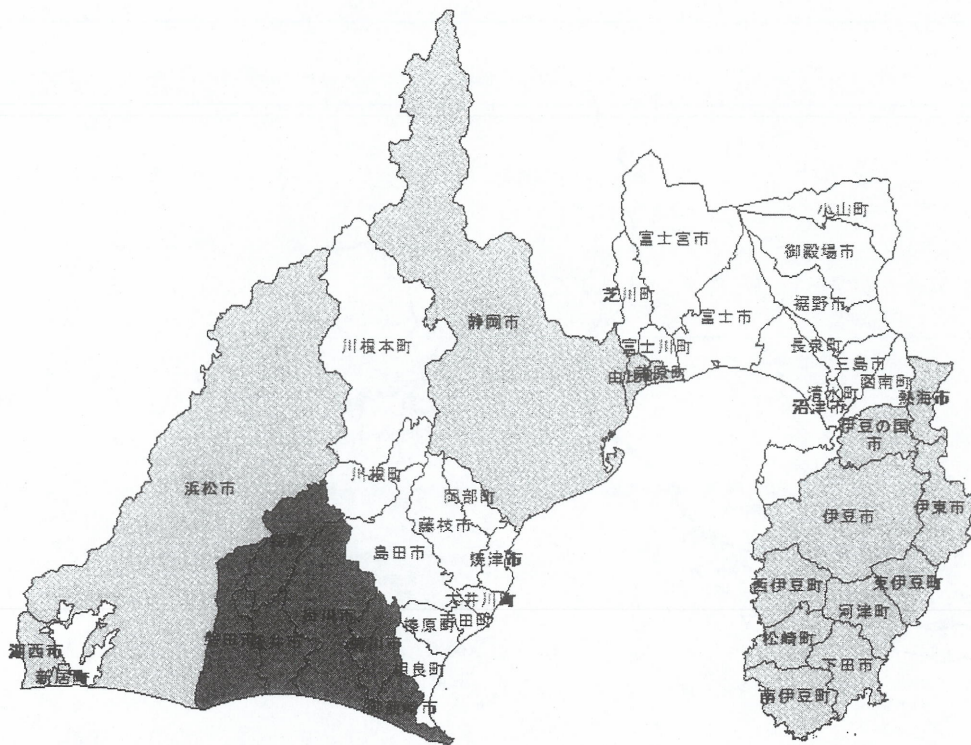


図3 数量化IV類による5分類

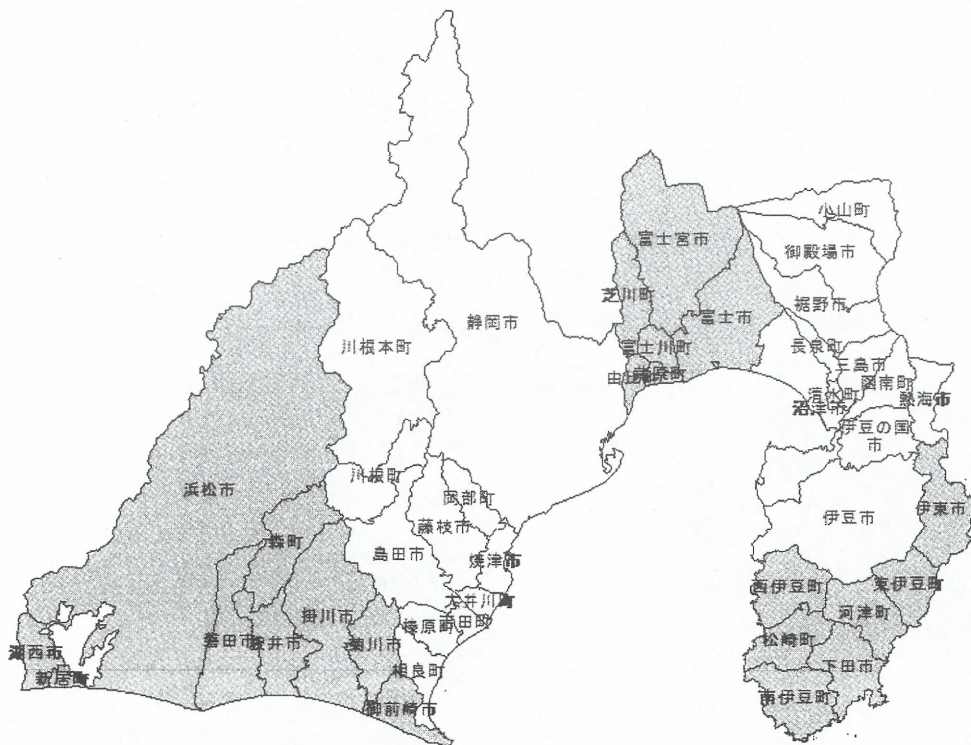


図4 数量化IV類による5分類（鉄道）

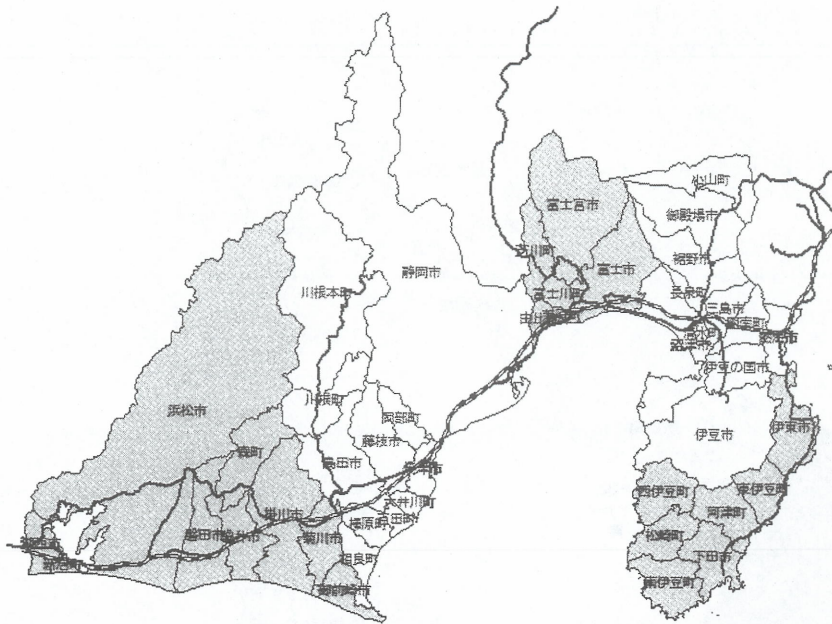
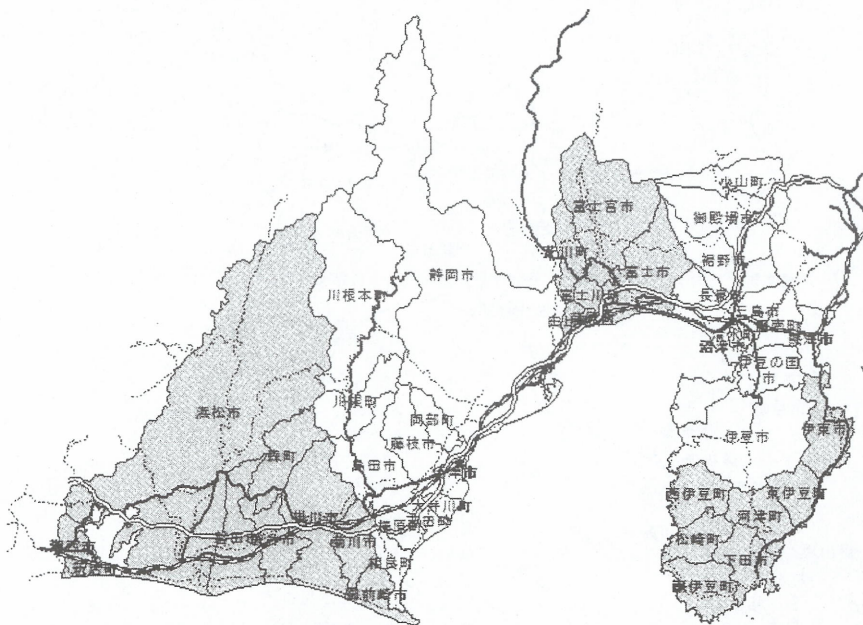


図5 数量化IV類による5分類（鉄道・道路）



1.2 オーダーメイド集計による統計データとその精度

本研究における統計的分析は、本来、県内地域別の政策課題にアプローチするため、「静岡システムダイナミックス・モデル」（静岡SDモデル）を開発し、それに付随してモデル分析に必要な統計データを整備することに端を発している。この研究の一環として、不足していた地域統計データの作成作業を行い静岡県の経済成長をめぐる政策シミュレーションを試みてきた。その過程で、総務省統計局による「オーダーメイド集計」のサービスを利用し、既存の公的統計では得られない地域表章の集計データ（国勢調査再集計）を活用した。本研究の分析結果はその一つの結果である。

分析で必要とされるデータは、「静岡県における市町村間の産業別就業者の流出・流入」であるが、公的統計でこのような統計情報を含むデータは、個票単位で産業別就業者の移動先（勤務地所在地）情報が含まれている『国勢調査』以外存在しない。しかしながら、細分化された産業別市町村間の県内移動については、人口 20 万人以上の都市の場合を除いてデータが公表されていない。このため、この分析においては、統計法の改正に伴い 2009 年度より実施されるようになった「委託による統計の作成等（オーダーメイド）」の制度を利用し、それにより得られた 2005 年度国勢調査の再集計データを用いることとした。このような独自のデータを用いて分析を試みたことが、本研究の一つの特徴であると見なし得る。

主要な結論については前述のとおりであるが、その際、看過できないのは、オーダーメイド集計データを利用することに起因する「誤差」の問題である。この誤差とは、統計表章上、静岡県内における市町村間労働移動者数を 10 人単位で集計・表示することから生じる乖離を意味する。再集計されたデータにこのような誤差が含まれることは、本研究で扱った国勢調査のオーダーメイド集計データに限らず、細分化された地域表章のオーダーメイド集計データ一般についても言えることである。

匿名データとは異なりオーダーメイド集計データは、細分化された地域表章データの利用が可能であるという点で、地域分析にとって今後より一層の拡充が期待される。しかしその一方で、前述のような誤差が生じることは、データ保護を目的とする公的統計の法的制約上不可避の問題である。また、このような誤差は、ランダムに発生しているとは言い難く、それ故誤差の正規性を仮定した統計的推測が困難であることは、オーダーメイド集計データの利用者にとって受容せざるを得ない問題となる。従って「分析結果の精度」を検討する場合、オーソドックスな統計的推測の方法に依拠して特定のパラメーター（例えば特定の市町村間の労働移動率）の有意性を調べることは必ずしも妥当ではない。

問題は、そのような誤差を含んだことによって分析結果が、誤差を含まない場合に比べてどの程度変化するのか、という点である。幸い筆者たちが利用した再集計データについては、全産業の市町村間労働移動の場合、本来の公表されている国勢調査データでも利用できるためその比較が可能である。そこで上藤・浅利・山下・高瀬[7]では、このオーダーメイド集計データの精度問題を解決する一つの試みとして、感度分析による分析結果の変化を検討した。⁷⁾

感度分析とは、データにある揺らぎ（摂動）を与え、それによって選択された統計的方法による分析結果がどのように変化するかを分析する方法である。本来は回帰分析を対象に発展した方法であるが、様々な多変量解析の方法にも拡充されている。特に本節で用いた数量化の方法に対する感度分析については、田中 [8]・[9]、Tanaka [10] による一連の研究によって、いくつかの影響関数の性質が明らかにされている。しかしながらこれらの研究は、主に数量化Ⅲ類を対象とした感度分析であり、数量化Ⅳ類については取り上げられていない。また上藤 [11] でも、田中豊の先行研究を基礎として、同じく数量化Ⅲ類における摂動の与え方と影響関数の問題を取り上げたが、数量化Ⅳ類については言

⁷⁾ 詳細な分析結果については、上藤・浅利・山下・高瀬 [19] を参照。

及していない。このため上藤・浅利・山下・高瀬[14]では、上藤 [11] で試みた分析を拡張して数量化IV類の影響関数を求めた。

先ず既存の国勢調査データに基づく「総数として就業者の市町村別県内移動」と、オーダーメード集計に基づく「総数として就業者の市町村別県内移動」の乖離率を見ておこう。ここで乖離率とは以下のように定義される。

$$\text{乖離率} = \frac{\text{再集計値} - \text{国勢調査値}}{\text{国勢調査値}} \quad (2)$$

実際に本研究で使用したデータで見ると、表1のようになる。

表1 全産業における乖離率

項目	乖離率
全産業非流出入	0.2%
全産業流出入	1.2%
全産業合計	0.1%

なお、これは全産業における就業者の県内市町村移動を表章した統計データであるが、産業別についても各々「10人単位の表章」原則が存在している。このため再集計データの産業別合計と全産業の再集計データが一致しないことも付言しておかなければならない。

そこで次に与えた摂動とその影響関数について見る。Saltelli-Chan-Scott [1] は摂動の多様性を指摘しているが、これは言い換えれば、摂動の与え方について厳密なルールが確立されていないことを意味する。このため上藤・浅利・山下・高瀬[14]では、上藤 [11] に基づいて以下の二つの摂動を与えて分析を行った。

摂動(1)

全ての変数、レコードに対して、(国勢調査値－再集計値)×0.02の摂動を与える。これは、国勢調査値に対する乖離率が0.1%であることから、再集計値に対して毎回0.0002%の摂動を累積的に与え、最終的には国勢調査値に至ることを意味する。再集計値による親近性尺度の計算から始まり、このような摂動を逐次与えながら国勢調査値よる親近性尺度の計算結果まで合計50回の計算を繰り返し、分析結果の感度を調べた。

摂動(2)

全ての変数において、静岡市から新居町までの44市町村についてレコードを逐次国勢調査値に入れ換え、合計44回の計算を繰り返し、分析結果の感度を調べた。

これら二つの摂動を与えた場合の影響関数については、上藤 [11] で用いた二つのタイプの影響関数を適用した。その一つは、任意の固有値 $\theta(i)$ に着目した影響関数で

$$\left. \frac{\partial \theta_i(\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = \theta(i), \quad i=1, \dots, (n-1) \quad (3)$$

と定義される。本稿で用いた数量化IV類の分析では二つの固有ベクトルを採用しているため、影響関数は、固有値 $\theta(1)$ と $\theta(2)$ の場合について調べる必要が出てくる。もう一つは、標本スコア $Y(i)$ の変化に着目した影響関数で以下のように定義される。

$$\left\| \frac{\partial Y(\varepsilon)}{\partial \varepsilon} \Big|_{\varepsilon=0} \right\| = \|Y(i)\| \quad (4)$$

(3) 式で定義される影響関数は、特定の固有値に着目した局所的な影響関数であるのに対して、(4) 式で定義される影響関数は、数量化IV類で分析上必要とされる固有値、固有ベクトルに基づく標本スコアに着目した総合的な影響関数であると言えよう。

再集計値による親近性指標の計算から始まり、逐次国勢調査値による新規性指標の計算結果まで摂動(1)を与えながら、合計50回の数量化IV類の計算を行い、クラスター分析によってそれぞれ6分類した所、28回目の感度から29回目の感度を与えたとき、静岡市を中心とするクラスターと富士市・富士宮市を中心とするクラスターに大きな変化が認められた。

この感度分析は、固有に着目した影響関数については、分析結果の変化をよく表している結果が得られた。しかしながら、標本スコアの変化に着目した影響関数では、11回目の感度と16回目の感度に大きな変化が認められたものの、それ以外の感度は全て一定で変化が認められなかった。一方、実際の数量化IV類とクラスター分析の結果においては、これらの感度を与えた場合でも、当初(出発点)の分析結果と比べて全く変化はなかった。つまりこれは、標本スコアによる影響関数が数量化IV類の場合における感度を十分に反映していないことを意味している。

摂動(2)は、すべての変数において、静岡市から新居町までのレコードを逐次国勢調査値に入れ換え、それによってデータに揺らぎを与えるものである。従ってこの場合は、一つの市町村が国勢調査値であることを除き、それ以外の市町村は再集計値であることを原則とする。この摂動を用いた数量化IV類による分析結果は、再集計値による分析と比較して大きな変化が認められず、静岡市と熱海市のデータを入れ替えた場合に多少の変化が認められるに止まった。人口規模の大きい静岡市の場合、その変化も大きく、全ての影響関数においてもの傾向が反映されたものであると理解できる。しかしながら熱海市の場合は、固有値 $\theta(1)$ の影響関数にその感度が示されているものの、固有値 $\theta(2)$ の影響関数においてはそれが示されなかった。またいずれの場合においても、急激な変化のパターンは認められず、数量化IV類とクラスター分析による分析結果に大きな変化は認められなかったとする前述の指摘を裏付ける結果となった。

これに対して、標本スコアによる影響関数では大きな変化を繰り返す感度のパターンを示しており、実際の分析結果の安定性を十分反映していない結果となった。これは、摂動(1)の場合と同様、数量化IV類の場合、この影響関数による感度評価に問題があることを示唆しているものと思われる。この点についての理論的解明については今後の更なる検討が必要となろう。

以上の分析結果をまとめると次のようになる。オーダーメードの再集計データを利用することによる分析結果の精度については、約0.06%の感度(28回から29回目の感度)をかけた場合に分析結果の変化が認められた。その変化は、再集計データを用いた数量化IV類による分析結果を否定するほどのものではなかったが、留意すべき問題点も含まれているように考えられる。例えば、このような変化が、再集計に伴う誤差に基づくものなのか、そうではなく、摂動の与え方の問題や感度による変化を評価する影響関数の問題なのを明らかにすることが上げられる。また仮にこの変化が誤差に伴うも

のであれば、それを補正するために適切な方法を検討する必要もあろう。

2. 経験的ベイズ推定法による市町間労働移動率の補正

小地域区分に再集計されたデータの精度については、比率の補正も重要な問題として検討する必要がある。前述の統計的分類に用いられるデータは、基本的には各市町間の労働移動率に加工して利用されるが、小地域区分で「比率」を用いる場合、異なる人口規模による比較可能性の問題が不可避免的に発生する。すなわち、人口規模や労働力人口の規模が異なる場合の比率による比較である。人口規模あるいは労働力人口の規模が大きい地域と小さい地域の労働移動率を比較する場合、見かけ上の数値上は同じであっても必ずしも同列に扱ってよいというものではない。例えば、表2は本研究が分析の対象とした2005年の労働流入率のうち、焼津市と岡部町から静岡市に流入する流入率を農業部門と全部門について示したものである。この比率を見ると、全産業に比べて農業部門では両市町の流入率はきわめて高いが、実際のところ静岡市への農業部門の総流入者数は35名（推定値）であり、このうち焼津市が10名（推定値）、岡部町が5名（推定値）となっているに過ぎない。⁸⁾一方、表3を見れば明らかなように、両市町の全産業全地域の流出者数に占める農業部門の比率は、それぞれ0.19%と0.13%ときわめて低い数値を示している。

表2 静岡市の労働流入率

従業地 常住地	静岡市の流入率	
	全産業	農業
焼津市	2.79%	28.57%
岡部町	0.40%	14.29%

表3 全地域への流出数

従業地 常住地	全地域への流出数		全流出数に占める農業の割合
	全産業	農業	
焼津市	27000	50	0.19%
岡部町	3950	5	0.13%

問題は、このような比率の数値を絶対的な情報と見なすとき、例えば静岡市への流入者数について焼津市では全産業では2.79%を占めているのに対して農業部門では28.57%と10.24倍の高さであるのに対して、岡部町では35.73倍となっており、相対的に大きな依存性があると判断されやすい。しかしながら、焼津市と岡部町とでは人口規模や労働力人口規模が大きく異なっていること、また全産業に占める農業部門の就業者が相対的に少ない状況の下で、このような経済的に強い「相互依存関係」を認めることには問題が残る。これらの問題は、小地域で細分類されたデータを比率という情報に加工するという事情に起因している。このような問題を回避し、人口規模や労働力人口規模に見合った比率に補正する方法は統計学的には様々な方法が提案されているが、本研究では経験ベイズ推定法を取り上げ検討することとした。この研究はあくまでもまだ試験的な段階であり、実際の応用には多くの検討課題を残している。そこで本稿では、丹後[14]・[15]に従って経験的ベイズ推定法の基本的な考え方を示し、その応用上の問題点を労働流出率について簡単に述べることとする。⁹⁾

今、任意の流出元（常住地）からの流出先を i 市町村、任意の産業を j とすると、 O_{ij} を j 産業における i 市町村への労働流出量、 n_j を j 産業における流出元の就業者総数、 N_j を j 産業の県内総労

⁸⁾ 前述のように、オーダーメード集計値は10人単位の表章によることから、総数と年齢別区分の合計は必ずしも一致しない。そこで本稿では、便宜的に推定値を総数と年齢別区分合計の平均値として求めた。

⁹⁾ これに関連して、ベイズ推測論の基本的な思想や論理については松原 [19] を参照。

働流出量¹⁰⁾、 N を県内全域の就業者総数とする。但し $i=1,2,\dots,a, j=1,2,\dots,k$ とする。ここで流出元の j 産業における i 市町村への労働流出率（ i 市町村への労働流入量） θ_{ij} は、以下のように定義される。

$$\theta_{ij} = \frac{O_{ij}}{n_{.j}} \quad (5)$$

丹後[11]・[12]では、死亡率の補正を目的として、事前分布にガンマ分布を仮定してその分布に含まれる未知パラメーターを最尤法によって求め、最終的には比率の補正值としての経験的ベイズ推定量を求めている。同様の仮定を置いて、ここで問題とする労働流出率の経験的ベイズ推定量を求めれば次のようになる。

よく知られているようにベイズの定理は、原因に相当するパラメーターを θ 、事前分布を $g(\theta)$ 、事後分布を $g(\theta \setminus x)$ 、結果に相当する観測値 x の確率密度関数を $f(x \setminus \theta)$ としたとき、以下のような関係が成り立つことを示している。

$$g(\theta \setminus x) = \frac{g(\theta)f(x \setminus \theta)}{\int g(\theta)f(x \setminus \theta)d\theta} \quad (6)$$

ベイズ推測でしばしば問題となるは、どのような確率分布を事前分布として想定すべきなのかという点であるが、丹後[11]・[12]では以下のようなガンマ分布を想定している。

$$g(\theta) = \frac{1}{\Gamma(\beta)} \alpha(\alpha\theta)^{\beta-1} \exp(-\alpha\theta) \quad (7)$$

但し $\theta \geq 0$ である。

そこで筆者たちも最初の取り組みとして、丹後[11]・[12]と同様この分布を事前分布として仮定することとした。(7)式からも明らかなようにガンマ分布は、パラメーター θ 以外にも未知パラメーター α 、 β が含まれている。一方、この分布の期待値は $\frac{\beta}{\alpha}$ 、分散が $\frac{\beta}{\alpha^2}$ 、であるため、筆者たちの問題として

いる O_{ij} については、次のように定式化される負の2項分布に従うことになる。

$$P(O_{ij} \setminus \alpha, \beta) = \frac{\Gamma(O_{ij} + \beta)}{\Gamma(\beta)O_{ij}!} \left(\frac{\alpha}{\alpha + n_{.j}} \right)^{\beta} \left(\frac{n_{.j}}{\alpha + n_{.j}} \right)^{O_{ij}} \quad (8)$$

この(8)式から、未知パラメーター α 、 β は最尤法によって求めることができる。¹¹⁾そこで求められたパラメーター α 、 β の最尤推定量 $\hat{\alpha}$ 、 $\hat{\beta}$ を利用して、最終的には以下のように θ_{ij} の経験的ベイズ推定量 $\hat{\theta}_{ij}$ が導出される。

$$\hat{\theta}_{ij} = \frac{O_{ij} + \hat{\beta}_{.j}}{n_{.j} + \hat{\alpha}_{.j}} \cdot \frac{N_{.j}}{N} \quad (9)$$

以上が本研究で問題とする経験的ベイズ推定法の概略である。なお、筆者たちが問題とする静岡県内の産業別・市町村別労働移動率への適用とその具体的な問題点の析出は今後の課題として取り組む予定ではあるが、それに先立ち本稿では理論的な問題をいくつか指摘しておきたい。第一は、やはりベイズ推定において常に問題となる事前分布の仮定である。筆者たちが検討の対象としたポワソン・

¹⁰⁾ 自市町村への流出=非流出を含んでいるので、これは j 産業における就業者総数を意味する。

¹¹⁾ 丹後[11]によれば、反復収束法により最尤推定値を求めることができるとしているが、初期値の推定にはモーメント法を利用することを勧めている。

ガンマモデルによる丹後[11]・[12]や平, 佐伯, 中田[4]の先行研究は, もともと市町村間の死因別・年齢別死亡率を補正する目的で開発されたものであり, そこには死亡率の地域特性はないということが前提でモデルが構築されている。しかし筆者たちが問題とする産業別の労働移動率は, 地域特性の影響を大きく受ける可能性があり, ここで示したポワソン・ガンマモデルの仮定に代えて新たな確率分布の仮定を模索する必要があると出た。第二に, 例えこのようなベイズ推定法を用いたとしても, 前節で見たようなオーダーメード集計に関わる誤差は補正することが難しいということである。もともとランダムな誤差を想定できない性質の誤差であるため, これをどのように評価するかは理論的にも経験的にも難しい問題である。しかしながら, オーソドックスな統計的推測とは異なり, ベイズ推測の場合事前分布に工夫を加えれば, このような問題をある程度克服できる可能性も否定できない。それらについては今後応用研究を進めていく過程で更なる検討を加えていく。

今後の課題—むすびに代えて—

SD モデルを地域分析に応用するために今後検討すべき課題の一つは, 小地域表章の公的統計を利用する際に生じるデータの精度の問題である。繰り返しになるが, オーダーメード集計データを利用する場合, 統計法上の制約より誤差が含まれることは不可避である。本稿で扱った事例で言うと, この誤差とは, 統計表章上, 静岡県内における市町村間労働移動者数を 10 人単位で集計・表示することから生じる乖離を指す。この誤差に加えて, 小地域表章故に生ずるデータの精度の問題もある。

例えば, 本稿で問題とした市町村間就業者移動は, 全産業だけではなく産業別に検討する必要があるが, 農林業のような県全体でも就業者数が相対的に少ない産業部門を取り上げる場合, 一つの問題が生じることは既に指摘した。市町村別にこの産業部門の労働移動率(流出率・流入率)を見たとき, もともと分子に相当する市町村別労働移動量に対して, 分母となる就業者総数は人口規模の大きさによって大きく変化するため, 人口規模の小さい市町村(従って就業者総数の少ない市町村)では, 労働移動量のわずかな変化でも大きくその比率の計算値に影響するからである。

このような小地域であるという事情から生じる比率の問題を補正する試みとして, 本稿では経験的ベイズ推定法を取り上げた。この推定法は, これまでも「人口動態統計」における市町村別死因別死亡率の計算などで応用されている研究例がいくつか存在しており, 本稿では事前分布としてガンマ分布を仮定したポワソン・ガンマモデルによる場合の基本的な問題点を検討した。残された問題としては, 経験ベイズ推定法を実際に応用し, そこから生ずる経験的な問題点を探ることであろう。仮定すべき確率モデルの妥当性も含めて, 経済データの精度を担保する方法を今後も検討していきたい。

謝 辞

本研究を進めるに当たり, 研究協力者として静岡県から塚本高士, 平野潤, 勝山敏司の各氏, 並びに財団法人静岡総合研究機構から片岡達也氏にご参加いただき, 様々な有益なご助言を賜りました。記して感謝の意を表します。

本研究は, 平成 23 年度人文学部学部長裁量経費の支援を受けて進められたものである。併せて, 平成 23 年度科学研究費補助金・基盤研究(C)「人口動態と地域ダイナミクス」(代表: 山下隆之, 課題番号 21530257) の助成を受けた。

参考文献

- [1] Saltelli, A., Chan, K. and Scott, E. M., *Sensitivity Analysis*, Wiley, 2000.
- [2] 林知己夫『数量化—理論と方法』朝倉書店, 1993年.
- [3] 林知己夫『行動計量学序説』朝倉書店, 1993年.
- [4] 平子哲夫・佐伯則英・中田正「人口動態市区町村別統計へのベイズ統計の応用について—(1)標準化死亡比への応用—」『厚生指標』第46巻第10号, 1999年, pp.3-11.
- [5] 松原望『ベイズ統計学概説—フィッシャーからベイズへ—』培風館, 2010年.
- [6] 齋藤堯幸・宿久洋『関連性データの解析法—多次元尺度構成法とクラスター分析法』共立出版, 2006年.
- [7] 高瀬浩二・山下隆之・上藤一郎「静岡県東部への企業進出に関するヒアリング調査報告—静岡県の魅力乗数—」『経済研究(静岡大学)』, 第15巻第3号, 2011年, pp.63-70.
- [8] 田中豊「数量化法における感度分析」『数理科学』, No.245, 1983年, pp.32-37.
- [9] 田中豊「数量化理論における感度分析とその応用」『品質』, 第14号, 1984年, pp.21-28.
- [10] Tanaka, Y., “Sensitivity analysis in Hayashi’s third method of quantification”, *Behaviormetrika*, No.16, 1984, pp.31-44.
- [11] 丹後俊郎「死亡指標の経験的ベイズ推定量について—疾病地図への応用—」『応用統計学』第17巻第2号, 1988年, pp.81-96.
- [12] 丹後俊郎『統計モデル入門』朝倉書店, 2000年.
- [13] 上藤一郎(1999)「数量化Ⅲ類の理論と農林業関連データへの応用」, 金子治平編『農村社会情報の収集およびデータ解析に関する理論的研究』平成9年度～平成10年度科学研究費補助金研究成果報告書, 1999年, pp.26-37.
- [14] 上藤一郎・浅利一郎・山下隆之・高瀬浩二「地域別経済指標に基づく静岡SDモデルの開発—モデル分析に利用する地域統計データの整備とその精度—」『地域研究(静岡大学)』, 第2号, 2011年, pp.1-12.
- [15] Yamashita, T., “A System Dynamics Approach to the Regional Macro-economic Model,” *Preliminary Conference Proceeding*, International System Dynamics Conferences, 2011.
<<http://www.systemdynamics.org/conferences/2011/prelimproceed.html>>
- [16] 山下隆之「地域マクロ経済のSDシミュレーション」『システムダイナミクス』, Vol.9, 2011年, pp.1-13.
- [17] 山下隆之・高瀬浩二「経済と環境のシステムダイナミクス・モデル：静岡県経済の将来予測」『経済研究(静岡大学)』, 第11巻4号, 2007年, pp.221-242.
- [18] 山下隆之・上藤一郎「地域経済内の相互依存性に関する研究—静岡県を事例として—」日本経済政策学会中部部会 OnLine ワーキングペーパー, No.2, 2011年.
<<http://www.soec.nagoya-u.ac.jp/jepa/>>
- [19] 山下隆之・上藤一郎・高瀬浩二「静岡県内市町村の相互依存性に関する研究」『経済研究(静岡大学)』第15巻第4号, 2011年, pp.195-211.
- [20] 山下隆之・上藤一郎・高瀬浩二「静岡SDモデルの開発—システムダイナミクスによる地域経済分析—」『経済研究(静岡大学)』第16巻第4号, 2012pp.157-172.