第46巻第11号「厚生の指標」1999年10月

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 (2)合計特殊出生率への応用

佐伯　則英*1　平子　哲夫*2　中田　正*3

Ⅰ はじめに

人口動態統計の地域別統計については，毎年の都道府県別統計の他，5年に1回，国勢調査の年を中心とした5年間のデータをまとめて保健所・市区町村別統計を作成しており，今回の「平成5年～平成9年人口動態保健所・市区町村別統計」で3回目の刊行になる。

これは，5年間のデータをプールすることにより，指標の安定性を確保しようとしているものであるが，5年間データをまとめても人口規模の小さな町村においては，数値の安定性を確保していない場合がある。合計特殊出生率(*TFR*，*Total Fertility Rate*)は，「1人の女子が仮にその年次の年齢別出生率で一生の間に生むとした時の子供数に相当する」指標として有名であるが，過去3回のデータをみると（表1），A村及びB村は，昭和60年(昭和58年～昭和62年のデータ)から，平成2年(昭和63年～平成4年のデータ）へ一度下がってから，平成7年（平成5年～平成9年のデータ）になると上がっている。また，C村は3回分下がり続けているが，D村は一度上がって，また，下がっているというように，人口規模の小さな市区町村においては合計特殊出生率の数値が安定していないのが見てとれる。

表1　人口規模の小さな町村における合計特殊出生率の変化の例

	昭和60年		平成2年		平成7年	人口規模
A 村	2.2	↘	1.5	↗	1.7	1 300
B 村	2.5	↘	1.7	↗	2.2	1 500
C 村	1.9	↘	1.3	↘	1.0	1 500
D 村	1.7	↗	2.1	↘	2.0	1 500
E 村	2.0	↘	1.9	↘	1.2	1 600
F 村	1.9	↗	2.3	↘	2.0	1 800
G 村	2.0	↘	1.9	↘	1.6	1 900
H 村	1.9	↘	1.5	↘	1.4	1 900
I 村	1.6	↘	1.4	↘	1.4	2 100
J 町	1.5	↘	1.2	↗	1.6	2 100

作成期間を10年あるいは15年と拡張する方法も考えられるが，この方法では必ずしも安定性を確保する迄には到らないばかりでなく，分子(例えば出生数，死亡数）と分母（例えば人口）との乖離が非常に大きくなる問題も生ずる。

他の方法として，当該市町村と周囲の市町村の情報を組み合わせて，より安定した指標を得る方法が考えられており，これは，ベイズ統計による方法として定式化されている。ベイズ・モデルを用いることにより数値が安定化し，従前比較が困難であった人口規模の小さな市区町村の経年的変化あるいは地域間の比較が可能になる。

本稿では，平成5年～平成9年人口動態保健所・市区町村別統計の中に掲載されている「合計特殊出生率（ベイズ推定値)」について，ベイズ推定の算出の考え方及び実際の計算方法について解説し，結果の精度評価及び考察を行う。

Ⅱ 合計特殊出生率

合計特殊出生率(*TFR*，*Total Fertility Rate*)は，15歳から49歳までの女子の年齢別出生率を合計したもので，1人の女子が仮にその年次の

*1厚生省大臣官房統計情報部管理企画課(前人口動態統計課)課長補佐　*2同人口動態統計課課長補佐　*3同前人口動態統計課長

年齢別出生率で一生の間に生むとした時の子供数に相当する指標である。

出生力の指標として用いられることが多く，わが国では，昭和41年「ひのえうま」の合計特殊出生率1.58を，平成元年に下回ったため「1.57ショック」という言葉も生まれた。現在，平成10年の概数ベースで1.38となっている。また，都道府県別の合計特殊出生率については，毎年公表しており，平成10年概数ベースでは，沖縄県，島根県，福島県が高く，東京都，千葉県，北海道，京都府など大都市を含む地域で低いという傾向があった（図1）。

この合計特殊出生率は，15歳から49歳までの女子の年齢別出生率を合計したものであることから以下のように表すことができる。

$$\text{合計特殊出生率}：TFR=\sum B_i/N_i^f$$
$$=\sum \theta_i \quad \cdots\cdots\cdots\cdots\cdots\cdots(1.1)$$

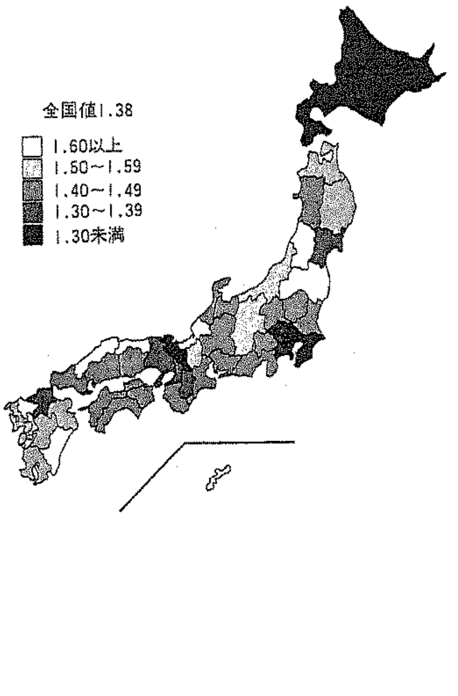
ただし，B_i＝母の年齢階級iの出生数（15≦年齢≦49）

N_i^f＝年齢階級iの女子人口（15≦年齢≦49）

θ_i＝年齢階級iの女子の出生率（15≦年齢≦49）

市区町村別の合計特殊出生率は5歳階級別のデータに基づき算出を行っているため，年齢階級iは（15歳～19歳），（20歳～24歳），……，（45歳～49歳）の7階級に分けて算出し，それらの年齢階級別出生率を合算することにより市区町村別の合計特殊出生率が得られることになる。

図1　都道府県別合計特殊出生率（平成10年概数）



合計特殊出生率（ベイズ推定値）の算出に際しては，まず，このような5歳階級別の女子の出生率の推定を行い，最後にそれらを合算することにより，合計特殊出生率（ベイズ推定値）の算出を行っている。

Ⅲ　ベイズ推定値の算出の考え方（5歳階級別出生率）

以下，5歳階級別出生率のベイズ推定値算出の考え方を解説する。特別な断りがない限り，人口，出生数及び出生率について，ある年齢階級iに関するものとする。

まず，出生率の事前分布としてベータ分布を選択する。ベータ分布は，2つの正の実数α，βで特徴づけられる区間［0，1］上の連続分布で，比率の分布としてしばしば使われるものである。また，この分布は$\alpha=\beta=1$のとき，区間［0，1］上の一様分布となる。

母数空間を$\Theta=[0,\ 1]$，母数（出生率）をθとすると，事前分布の密度関数は以下のようになる。

$$p(\theta)=\frac{1}{B\ (\alpha,\ \beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\cdots\cdots(2.1)$$

ただしB（α，β）はベータ関数で，

$$B\ (\alpha,\ \beta)=\int_0^1 x^{\alpha-1}\ (1-x)^{\beta-1}dx\ \cdots(2.2)$$

である。

ベータ分布の基本的性質として，確率変数$\tilde{x}$がベータ分布に従うとき，

$$\text{平均}：E(\tilde{x})=\frac{\alpha}{\alpha+\beta}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(2.3)$$

$$\text{分散}：V(\tilde{x})=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\cdots(2.4)$$

となる。

観測によってN（人口）とB（出生数）を得るわけであるが，偶然変動を大きく受けるのはBであり，Nはそれほど変動しないと考える。ここでは人口は既知とし，出生数は確率変数$\tilde{B}$の実現値と考える。更に出生率θが真の時，$\tilde{B}$が2

項分布Bin (N, θ) に従うと仮定する。このとき，$\tilde{B}$の確率密度関数f ($B|N$, θ) は次のようになる。

$$f\ (B|N,\ \theta)={}_{N}C_{B}\,\theta^{B}\ (1-\theta)^{N-B}\ \cdots\cdots(2.5)$$

これを出生率θの関数とみれば尤度関数となる。

以上より，観測によりN, Bを得たという条件のもとで母数θの事後分布の密度関数をp ($\theta|B$, N)とすると，事前分布$p(\theta)$は(2.1)として得られているので，ベイズの定理より

$$p(\theta|B,N)=\frac{f\ (B|N,\ \theta)\ p\ (\theta)}{\int_0^1 f\ (B|N,\ \theta)\ p\ (\theta)\ d\theta}$$

$$=\frac{{}_NC_B\,\theta^B\ (1-\theta)^{N-B}p\ (\theta)}{\int_0^1 {}_NC_B\,\theta^B\ (1-\theta)^{N-B}p\ (\theta)\ d\theta}$$

$$=\frac{1}{B(\alpha',\beta')}\theta^{\alpha'-1}(1-\theta)^{\beta'-1}\quad(2.6)$$

ただし$\alpha'=\alpha+B$, $\beta'=\beta+N-B$

このように，事後分布もα'及びβ'をパラメータとするベータ分布となる。事後分布の平均値及び分散は，ベータ分布の基本的性質((2.3)及び(2.4)) より

$$E\ (\tilde{\theta}|B,\ N)=\frac{\alpha+B}{\alpha+\beta+N}\ \cdots\cdots\cdots\cdots(2.7)$$

$$V(\tilde{\theta}|B,N)=\frac{(\alpha+B)\ (\beta+N-B)}{(\alpha+\beta+N)^2(\alpha+\beta+N+1)}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(2.8)$$

となる。

Ⅳ 実際の計算方法

以上のようなベイズモデルにより合計特殊出生率（ベイズ推定値）を算出する場合，当該市区町村の出生率の事前分布（ベータ分布）の平均，分散を設定する必要がある。今回の算出にあたっては，当該市区町村を含む，より大きな地域（二次医療圏）の市区町村の出生率の平均，分散に等しいとして決定した。

二次医療圏とは，都道府県において設定され，概ね30万人程度の人口規模を持つ日常生活圏として全国を覆うように定められている区域であり，主として一般の入院医療を提供する病院の病床整備を図るべき区域のことである。日常生活圏ということであるため，出生や死亡の発生に関してある程度同一性が考えられることから，全国的な状況観察に適するものとして採用した。算出にあたっては，全国的な状況を観察するものとして，算出結果の統一性を保つため，当該市区町村を含む二次医療圏の数値に基づき，ベイズ・モデルを適用し，市区町村の数値を推定している。

今，地域をK二次医療圏，K二次医療圏内の市区町村をjとし，j市区町村の人口をN_j，出生率を$q_j(=B_j/N_j)$として対応するK二次医療圏内の市区町村の出生率の平均Q_K及び分散V_Kを以下のとおりに求めた。このとき，ウェイトw_jの取り方により，Q_KはK二次医療圏における出生率と一致する。

$$Q_K=\sum_K w_j q_j\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(3.1)$$

$$V_K=\sum_K w_j\ (q_j-Q_K)^2\cdots\cdots\cdots\cdots\cdots\cdots\cdots(3.2)$$

ただし$w_j=\dfrac{N_j}{\sum_K N_j}$, $q_j=\dfrac{B_j}{N_j}$

このとき，事前分布にベータ分布を規定し，平均について(2.3)=(3.1)，分散について(2.4)=(3.2)となるため

$$\frac{\alpha_K}{\alpha_K+\beta_K}=Q_K\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(3.3)$$

$$\frac{\alpha_K\beta_K}{(\alpha_K+\beta_K)^2\ (\alpha_K+\beta_K+1)}=V_K\cdots\cdots\cdots(3.4)$$

これらをα_K, β_Kについて解いて，

$$\alpha_K=Q_K\left\{\frac{Q_K\ (1-Q_K)}{V_K}-1\right\}\ \cdots\cdots\cdots\cdots(3.5)$$

$$\beta_K=(1-Q_K)\left\{\frac{Q_K\ (1-Q_K)}{V_K}-1\right\}\ \cdots(3.6)$$

となる。このα_K, β_KがK二次医療圏の市区町村に対する出生率の事前分布のパラメータとなる。このとき，事後分布の平均及び分散は(2.7)及び(2.8)より

$$E\ (\tilde{\theta}|B_j,N_j)=\frac{\alpha_K+B_j}{\alpha_K+\beta_K+N_j}\ \cdots\cdots\cdots\cdots(3.7)$$

$$V(\tilde{\theta}|B_j,N_j)$$

$$=\frac{(\alpha_K+B_j)\ (\beta_K+N_j-B_j)}{(\alpha_K+\beta_K+N_j)^2(\alpha_K+\beta_K+N_j+1)}\cdots(3.8)$$

となる。

つまり，①K二次医療圏内の各市区町村の人口Nと出生数Bから，式(3.1)及び(3.2)により，事前分布の平均Q_K及び分散V_Kが計算できる。②このQ_K及びV_Kに基づき，K二次医療圏のパラメータα_Kとβ_Kが式(3.5)及び(3.6)により算出できる。

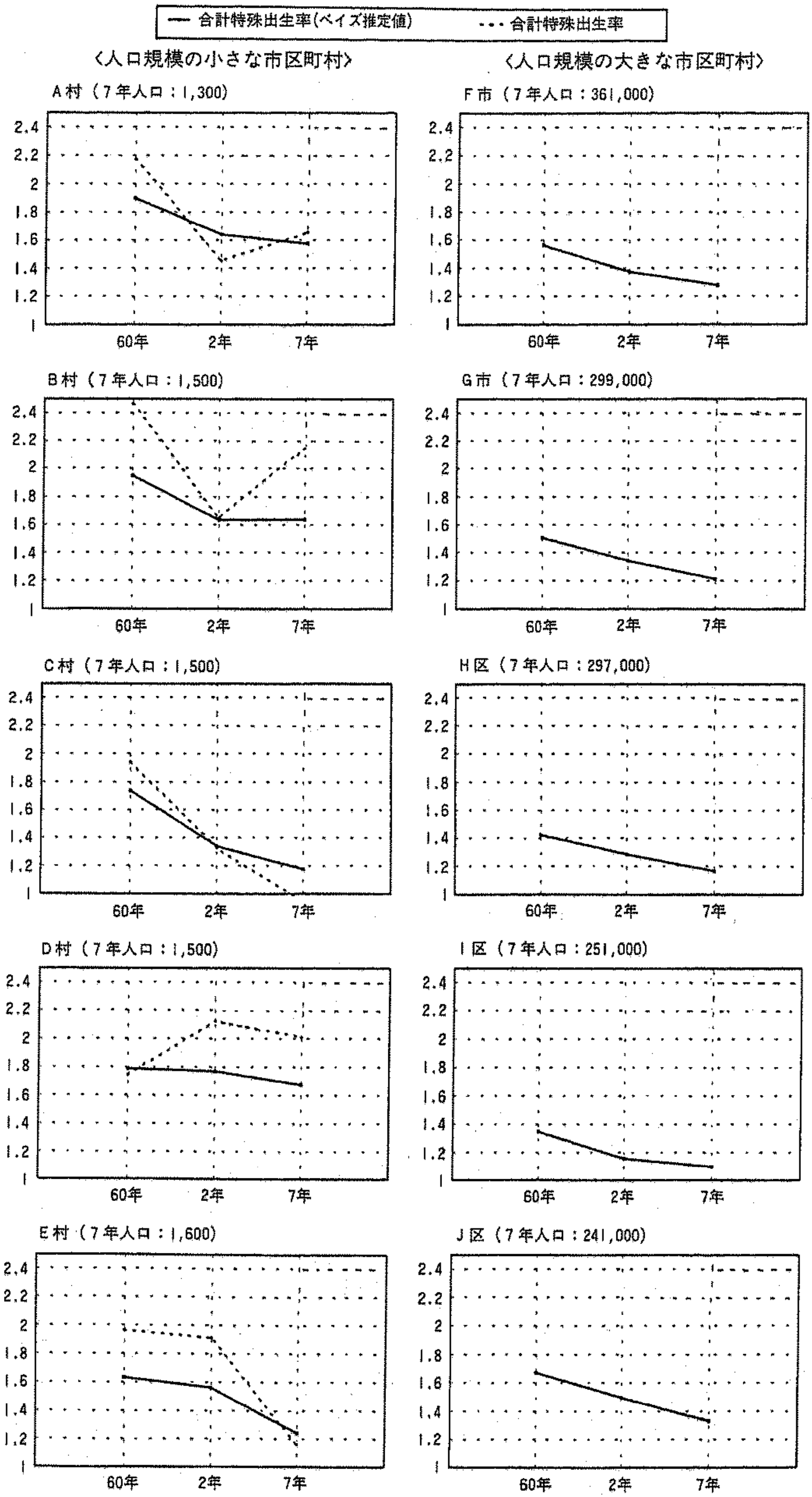
最後に，③式(3.7)により，当該j市区町村のベイズ推定値を得ることができる。また，(3.8)により，このときのベイズ推定値の分散も計算できることとなる。

合計特殊出生率（ベイズ推定値）は，このように5歳階級別に得られたベイズ推定値を合計することにより得られることになる。

V 結果及び精度

図2は，得られた合計特殊出生率（ベイズ推定値）と従前の合計特殊出生率との比較として，人口規模の小さな市区町村と大きな市区町村をそれぞれ5ずつ例としてあげたもので

図2 合計特殊出生率と合計特殊出生率（ベイズ推定値）の比較
－人口規模の小さな市区町村と人口規模の大きな市区町村－



である。前出の人口規模の小さな市区町村であるA村からE村において，合計特殊出生率(ベイズ推定値)の方は，偶然によるものと考えられる極端な値がなくなり，傾向がはっきりと見て取れるようになっている。これに対して，人口規模の大きな市区町村では従来の値と変化は無い。

また，図3は，ある県における合計特殊出生率（ベイズ推定値）の合計特殊出生率に対する割合を示したものであるが，人口規模が2万〜3万人以上の市区町村では，ほとんど1，つまり数値の変化がないのに対し，それ以下の人口規模の市区町村では上下ともに補正されていることが見てとれる。

図4は，合計特殊出生率と合計特殊出生率（ベイス推定値）の比較を，北海道を例に挙げて地図で比較したものである。従前の合計特殊出生率による地図では，高い地域と低い地域が混沌としていて判別がつきづらいが，合計特殊出生率（ベイズ推定値）による地図では，札幌，旭川，函館といった都市の周辺部分で出生率が低く，その他の地域では出生率が高いということがよくわかる。

合計特殊出生率（ベイズ推定値）について，日本全国の市区町村の地図を書いてみると，東京，大阪，北九州，札幌を中心とした，いわゆる都市部において低く，沖縄，九州南部，山陰地方，東北地方

図3　市区町村の人口規模とTFR（ベイズ推定値）のTFRに対する割合の関係
－A県－

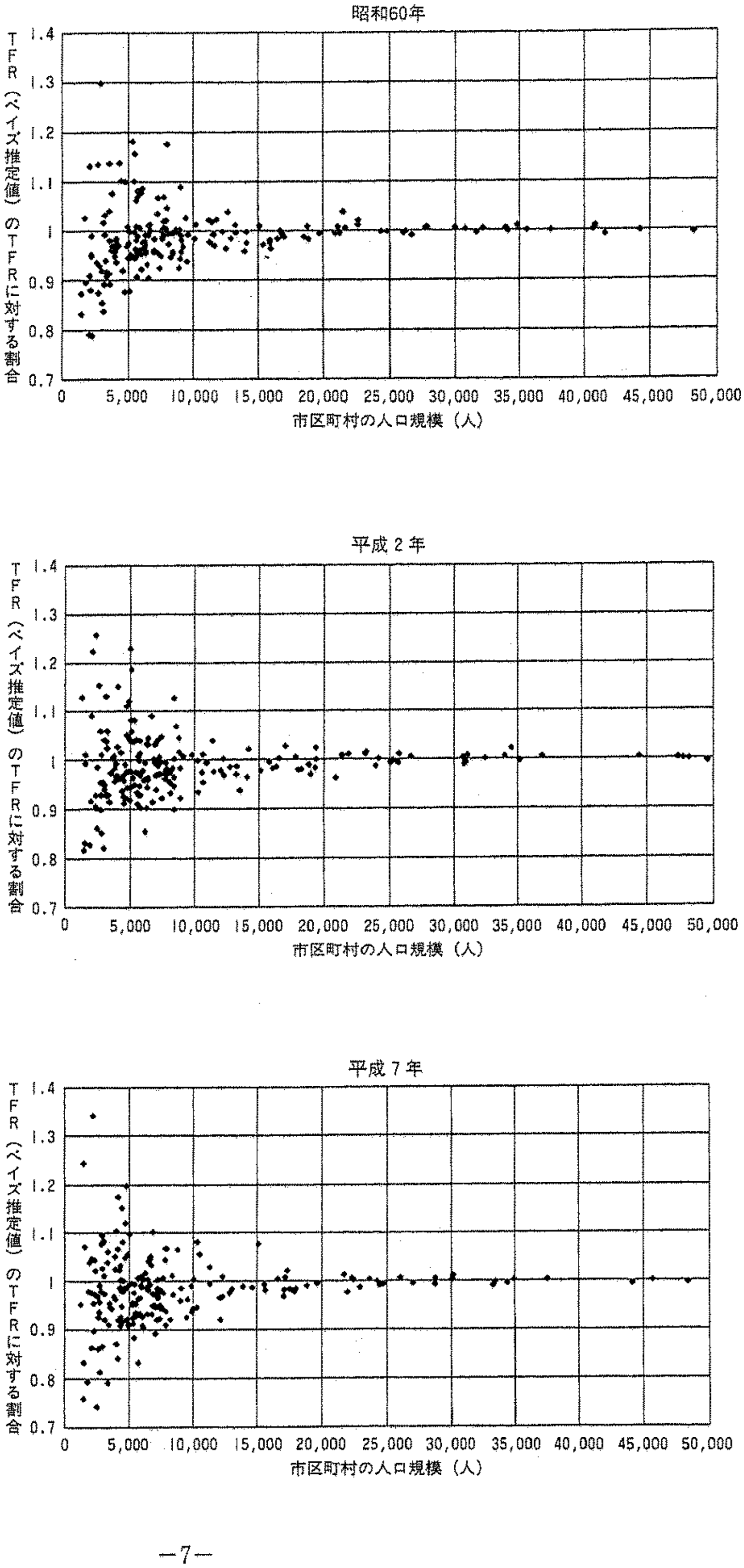
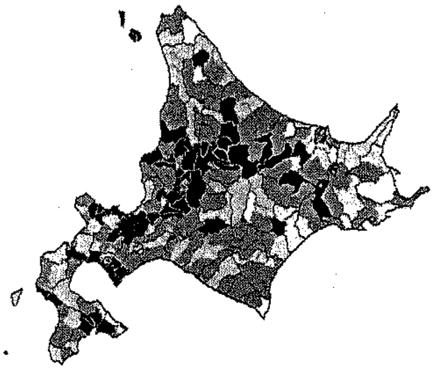


図4　市区町村別合計特殊出生率と合計特殊出生率（ベイズ推定値）の比較
－北海道－

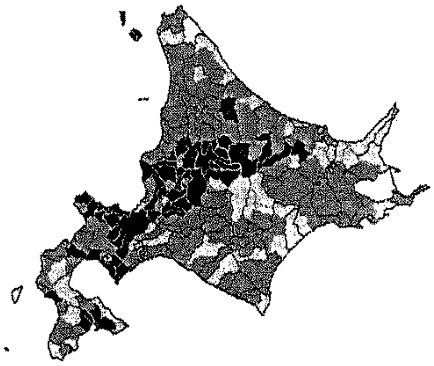
凡	例
	出生率が高い地域
	出生率がやや高い地域
	出生率が平均的な地域
	出生率がやや低い地域
	出生率が低い地域

昭和60年（昭和58年～昭和62年）

合計特殊出生率

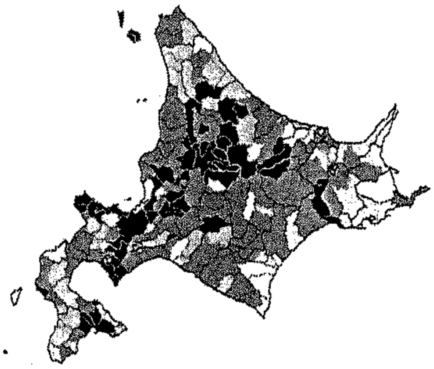


合計特殊出生率（ベイズ推定値）

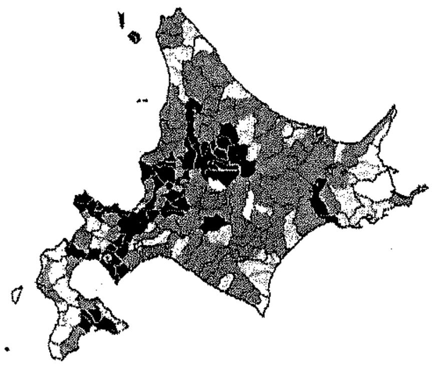


平成２年（昭和63年～平成４年）

合計特殊出生率

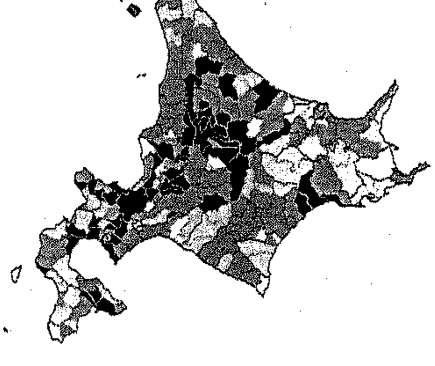


合計特殊出生率（ベイズ推定値）

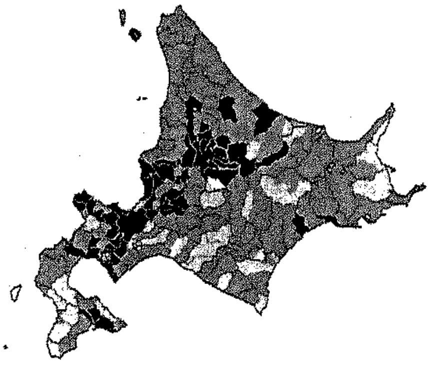


平成７年（平成５年～平成９年）

合計特殊出生率



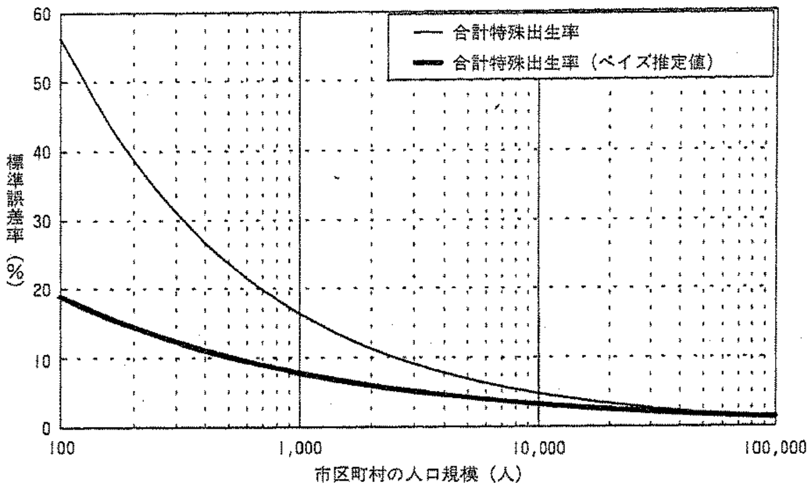
合計特殊出生率（ベイズ推定値）



などの地域で出生率が高いということがわかった。また，都道府県別にみた場合には，出生率が低い北海道をみると，県内すべての市区町村で出生率が低いということではなく地域差があるということが，今回の市区町村別のデータをみることによって理解できる。

指標の精度を示しているのが図5である。これは，市区町村の人口規模と合計特殊出生率及び合計特殊出生率（ベイズ推定値）の標準誤差率（推計値の大きさに対する標準誤差の百分率）の関係を示している。合計特殊出生率は2項分布，合計特殊出生率（ベイズ推定値）はベータ分布の分散を考えているため，両者の比較は正確には問題があるが，傾向は読みとることができ，合計特殊出生率（ベイズ推定値）では，人口規模のより小さな市区町村において，合計特殊出生率に比べ大幅に数値が安定するようになっていることがわかる。

図5　市区町村の人口規模と標準誤差率との関係



Ⅵ 考　　察

観測された人口N，出生数Bに対して，事後分布における推定値をみると，分母は$\alpha+\beta+N$，分子は$\alpha+B$となっている。事前分布の平均即ち二次医療圏全体の出生率が$\alpha/(\alpha+\beta)$で与えられていることを考慮すれば，事後分布における推定値は二次医療圏の出生率を，観測された（市区町村の）人口，出生数で修正したものであると考えられる。従って，たとえ人口，出生数ともに0であっても出生率が推定できないということはなく，その場合にはその市区町村の出生率はその市区町村が含まれる二次医療圏の出生率となる。

このように，今回の事前分布の設定は，ある地域の出生率は基本的にはその地域を含むより大きな地域の出生率に従う，という素朴な仮定に基づくという点で，極めて自然なものであるといえる。

合計特殊出生率（ベイズ推定値）についての市区町村地図については，いわゆる都市部において低く，都市から離れた地域で出生率が高いということがはっきりとわかった。都道府県別の分析によると，都道府県別の合計特殊出生率は都市化と強い関係を持っていることがわかっている（佐伯（1998））が，市区町村別にみてもこのことが明らかになったといえよう。

市区町村の人口規模についてみると，人口規模が2万～3万人以下の市区町村において値が安定化され，人口2万～3万人程度以上の市区では従前の合計特殊出生率の数値と変わりはない。このため，合計特殊出生率(ベイズ推定値)は，人口規模の小さな市区町村において有効な指標であるといえる。

今回のベイズ・モデルは，極めてシンプルなモデルを採用しており，事前分布の設定に時系列要素を入れるなども考えられるが，業務として5年周期で提供していくデータであるという制約上，あまり複雑なモデルの採用は不可能であり，また，今回のモデルでも実務上の使用には耐えられるものと考えている。

Ⅶ おわりに

市区町村において，地域間の比較や経年的な

動向を合計特殊出生率で見る場合，特に出生数が少ない場合には，数値が大幅に上下し，その地域の出生の動向を把握することが困難であった。これは，標本数（出生数）が少ないため，偶然変動の影響を受け，数値が不安定な動きを示すためである。

今回，算出された合計特殊出生率（ベイズ推定値）により，出現数の少なさに起因する偶然性の影響を減少させ，地域間比較，経年比較に耐えうるより安定性の高い指標を得ることができた。

本指標の性格を十分に理解された上で，各地域における保健・医療・福祉活動の基礎資料として活用されることを願うものである。

文　献

1）厚生省大臣官房統計情報部，「平成５～９年　人口動態保健所・市区町村別統計」，人口動態統計特殊報告，1999
2）鈴木雪夫，「統計学」，朝倉書店，1987
3）鈴木雪夫，「人口動態統計の小地域分析手法に関する研究」，平成９年度厚生行政科学研究報告書，1998
4）鈴木雪夫，「小地域における死亡指標－ベイズ統計学からのアプローチ」，平成元年度厚生科学研究報告書，1990
5）佐伯則英，「人口動態からみた少子現象」，子供家庭福祉情報VOL.14 p62-69，1998