

2020年度 統計データ分析コンペティション
特別賞（統計活用） [大学生・一般の部]

社会増減が合計特殊出生率に与える影響

堤敬司（京都府政策企画部企画統計課）

論文の概要

都道府県及び市区町村別の年齢別女性人口と出生数を用いて合計特殊出生率（TFR）の推定を行い、女性の年齢階級別転入超過率を特徴量として重回帰モデル及びランダムフォレストによる分析を行って、女性の年齢別転入超過率が出生率に有意に影響していることを示している。また、結婚して出産を迎える女性の出産コア年齢である30-34歳の女性の転入超過が自治体別でみた場合のTFRの向上に資することを指摘している。

論文審査会コメント

ランダムフォレストの利用例として、考察も面白かった。

社会増減が合計特殊出生率に与える影響

堤 敬司

京都府政策企画部企画統計課

第1章 研究のテーマと目的

我が国では、合計特殊出生率（以下「TFR」という。）「1.57ショック⁽¹⁾」を契機に、様々な少子化対策が講じられてきた。しかしながら、低下トレンドから抜け出すことはできず、TFRの全国値は2005年に1.26まで低下した後、2015年には1.45と一旦は回復していたが、厚生労働省「人口動態統計月報年計（概数）」によると、2019年は1.36と再び低下の様相を呈している。人口置換水準2.07を下回る状況が1974年以来46年連続で続いている状況である（図1）。

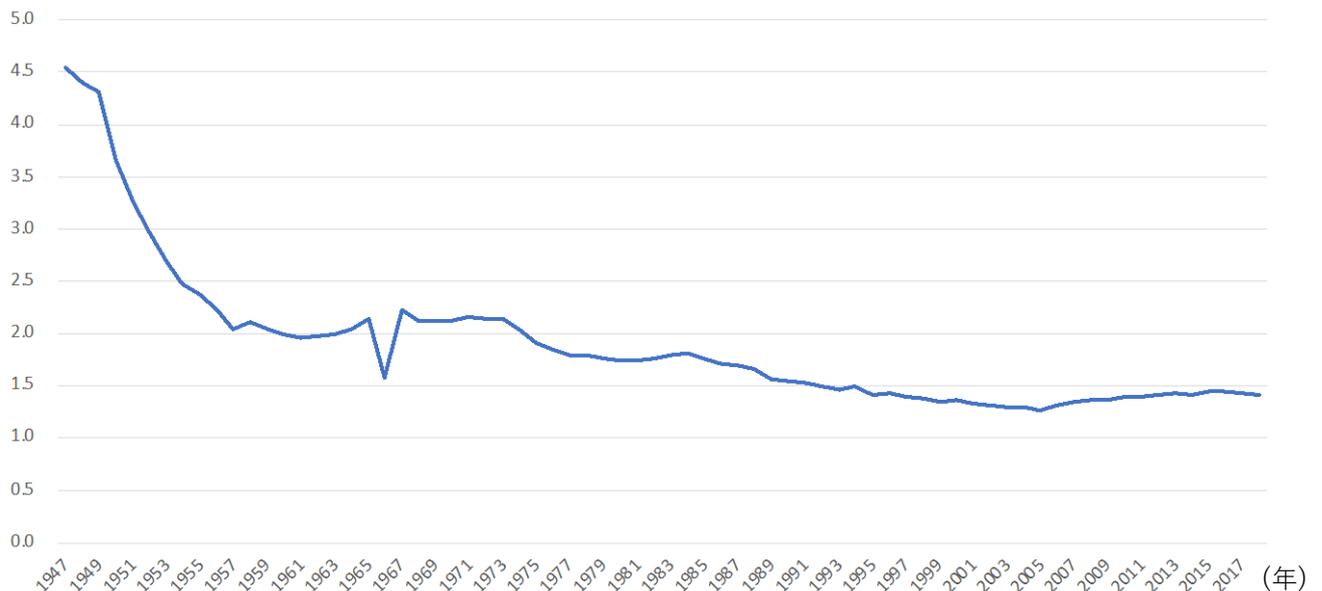


図1 TFR全国値(確定値)の推移(1947 - 2018年)

出典：厚生労働省「人口動態調査」

少子化に係る諸問題が取り沙汰されて久しいが、今生まれた新生児は「すでに起こった未来⁽²⁾」として今後約100年間にわたり人口ピラミッドの1つの年齢階級を構成し、当該年齢階級の人口を事後的に増加させることは移民などの一部の例外を除き不可能である。このような少子化に伴う人口減少が漫然と続く場合、「日本の子ども人口時計2020年版⁽³⁾」によると西暦3280年には、我が国の子どもの数は1人になると試算されている。人口減少による影響は、集落・コミュニティの消滅、経済規模の縮小、国防力の低下など枚挙にいとまがなく、確実に我が国の国力を奪っていくことについては論を俟たないであろう。人口減少を一刻も早く食い止め、社会の安定を図るためにも、要因を的確に解析しTFRを上昇させていくことが強く求められる。

TFRに影響する諸要因の分析については、都道府県別や市町村別でのTFRを用いた先行研究⁽⁴⁾が既に行われている。沖縄県(2017年TFR1.94)と東京都(2017年TFR1.21)では状況が異なり、全国規模の処方箋だけではなく、地域の実情に応じた施策も求められてくるため、都道府県別及び市町村別の分析は有用であると考えられる。

$$\text{期間合計特殊出生率} = \sum_{x=15}^{49} \frac{\text{Children}(x)}{\text{Mothers}(x)}$$

表1 TFRの変数の定義

変数名	出典
Children (x): x歳の女性が出産した子どもの数	厚生労働省「人口動態統計」
Mothers (x): x歳の日本人女性人口(10/1時点)	総務省「推計人口」または総務省「国勢調査」

※通常、都道府県別及び市区町村別のTFRの算出に際しては、利用できる統計の制限から5歳階級別人口を用い、5を乗じることで便宜的に算出されている。

TFRには、「期間合計特殊出生率」と「コーホート合計特殊出生率」があるが、厚生労働省が公表するデータなど一般的には前者が用いられている。上記定義式と共に考えると、分母の日本人女性人口に攪乱要因があるのではないかと考えた。京都府は地域性として学生が多く、「出産を想定しない女性」が相対的に多く存在するため、低めに出る傾向（TFRが薄まる傾向）があるとの仮説を設定する。

第2章 研究の方法と手順

2.1 年齢階級別データの取得及び加工

仮説の検証にあたって、年齢階級5歳階級別の女性の転入超過率に着目し分析を行っていく。都道府県レベルでの分析、そして市町村レベルで京都府のローカル分析を行う。TFRについては、都道府県別は教育用標準データセット(SSDSE)から取得し、市区町村別は別途自力で近似するという方法をとる。厚生労働省「人口動態保健所・市区町村別統計」において、市区町村別TFRが公表されているが、5年平均となっている点や分母が5年間固定されている点には一定の留意が必要である。すなわち、市区町村別で時系列変化による差異の分析が難しい等のデメリットがある。今回は、分母のx歳の日本人女性人口を、「推計人口」「国勢調査」ではなく、総務省「住民基本台帳に基づく人口、人口動態及び世帯数調査」を用いてTFRを近似させる方法で分析を行っていく。

2.2 年齢階級別データを用いた分析

上記で取得・加工したデータは以下の重回帰モデル及びランダムフォレストにより、TFRの高低に影響を与える特徴量（日本人女性の年齢階級別転入超過率）を抽出し、結果の解釈及び考察につなげていく。

$$\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_7 x_7$$

\hat{y} : TFR推定値 β_0 : 切片 β_i : 変回帰係数 x_i : 5歳階級別日本人女性の転入超過率（15歳～49歳）

第3章 データセットと加工

本稿の分析においては、以下のデータを用いた。必要に応じ、e-Statからデータを取得し加工を行った。また、総務省「住民基本台帳に基づく人口、人口動態及び世帯数調査」を用いたTFRの近似を行う。

3.1 使用データセットと加工

表2 使用データセット

使用変数	出典
A. 都道府県別 TFR	厚生労働省「人口動態統計」(2014-2017)*
B. 総人口 (人)	総務省「人口推計」(2017)*
C. 短期大学学生数 (人)	文部科学省「学校基本調査」(2017)*
D. 大学学生数 (人)	文部科学省「学校基本調査」(2017)*
E 1519. 女性転入超過数(15-19 歳) (人) ～E 4549. 女性転入超過数(45-49 歳) (人)	総務省「住民基本台帳人口移動報告」(2014-2017)
F 1519. 女性人口(15-19 歳) (人) ～F 4549. 女性人口(45-49 歳) (人)	
G 1519. 出生数 (15-19 歳) (人) ～G 4549. 出生数(45-49 歳) (人)	総務省「住民基本台帳に基づく人口、人口動態及び世帯数調査」(2013-2017)
H. 市区町村別 TFR-A	厚生労働省「人口動態統計」(2013-2017)

「*」は、SSDSE から取得したもの。それ以外は e-Stat から取得した。

表3 データセットの加工

変数名	加工方法
I. 人口 1000 人あたり学生数 (人)	$(C+D) / B * 1000$
J1519. 女性転入超過率(15-19 歳) (%) ～4549. 女性転入超過率(45-49 歳) (%)	$E1519 / F1519 * 100$ ～ $E4549 / F4549 * 100$
K. 市区町村別 TFR-B	$(G1519 / F1519 * 5) + (G2024 / F2024 * 5) + \dots + (G4549 / F4549 * 5)$

※ 5 歳階級別市区町村別の転入超過数については、2013 年以前は e-Stat に掲載がないため、2014 年～2017 年の 4 年分の転入超過率を算出 (都道府県も市区町村に合わせるため同一の期間分を算出)

3.2 市区町村別 TFR-B の妥当性

市区町村別 TFR の分析に用いられるものとして、先行研究において子ども女性比⁽⁴⁾⁽⁵⁾を用いた分析があるが、今回は上記 TFR-B を別途計算し、散布図に描き、回帰による決定係数 R²で妥当性を検証した (図 4 及び図 5)。

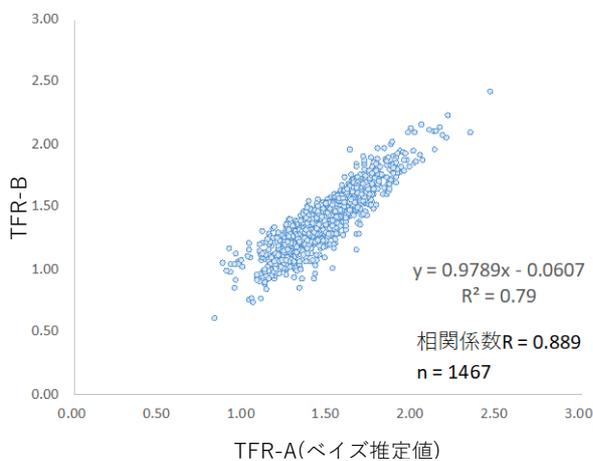


図4 TFR-A(ベイズ推定値)と TFR-B

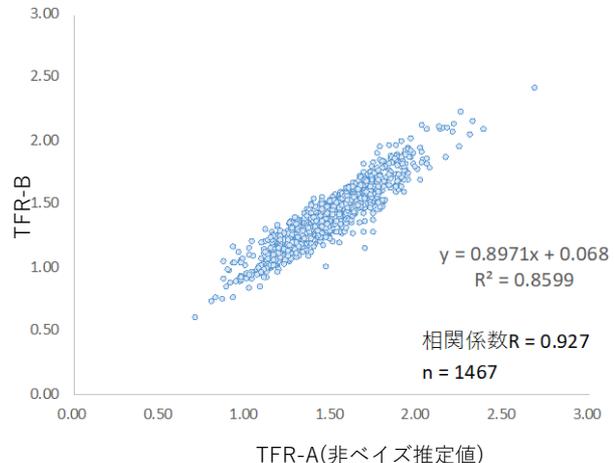


図5 TFR-A(非ベイズ推定値)と TFR-B

人口が少ない市区町村は、ばらつきが大きくなるため「人口動態保健所・市区町村別統計」の例にならない、TFRの標準誤差が0.1以上の市区町村については、外れ値として除去した。その結果、市区町村数が1884から1467となった。TFR-Aは2013年～2017年（平成25年～29年）の平均値になっていることから、同一期間のTFR-Bの平均値をとった。また、TFR-Aにはベイズ推定値と非ベイズ推定値の両方の掲載があるため、両方との関係性を検証した。

決定係数 R^2 をみると、TFR-A（ベイズ推定値）とTFR-Bの関係では0.7900、TFR-A（非ベイズ推定値）とTFR-Bの関係では0.8599と、一定の妥当性が確認できる。これによりTFR-Aでは5年平均であったが、TFR-Bは各年毎の市区町村別TFRを用いることができる。なお、本稿ではTFR-Aとの時系列の整合性を保つため、2013年から2017年までの各年のTFR-Bを求めたが、最新のデータ（2020年9月1日時点）では2018年まで計算できるため、最新のデータを用いた様々なTFRの分析も可能である。

第4章 データ分析の結果

4.1 都道府県別TFRによる分析

5歳階級別市区町村別の転入超過数については、2013年以前はe-Statに掲載がないためSSDSEとの時系列の整合性を保つために、都道府県別TFR及び市区町村別TFRはともに2014年から2017年までの4年分を分析対象とする。

まず、都道府県別TFRについて、上記2.2で示した重回帰モデル式を適用した結果、以下表4のとおり15-19歳及び30-34歳の女性の転入超過率で有意となった。15-19歳の女性の転入超過率が1%上昇すると、TFRに対して0.04程度の下押し圧力となることが分かる。また逆に30-34歳の転入超過率が1%上昇すると、TFRに対して0.12程度の上昇圧力となることが分かる。

また、社会増減の影響を除くため特徴量（説明変数）を全て0とした場合（以下「特徴量ゼロモデル」という）、上記の重回帰モデル式では切片のみが残ることになる。すなわち15-49歳の女性の社会増減の影響を除くと、TFRは1.46程度となる。全国の2014-2017年のTFR公表値の平均は1.44であるが、当該重回帰モデルの95%C.I.内に収まることにも注目しておきたい。

表4 都道府県別TFRに係る重回帰分析結果

変数	偏回帰係数	標準誤差	t 値	95% C.I. 下限	95% C.I. 上限	VIF
切片	1.457 **	0.011	133.482	1.435	1.479	
15-19歳女性転入超過率	-0.039 **	0.012	-3.206	-0.064	-0.015	4.421
20-24歳女性転入超過率	-0.012 †	0.007	-1.656	-0.027	0.002	6.503
25-29歳女性転入超過率	-0.003	0.021	-0.120	-0.044	0.039	3.160
30-34歳女性転入超過率	0.116 *	0.049	2.373	0.020	0.213	2.805
35-39歳女性転入超過率	0.150 †	0.078	1.911	-0.005	0.304	2.227
40-44歳女性転入超過率	-0.065	0.118	-0.557	-0.297	0.167	2.266
45-49歳女性転入超過率	0.088	0.132	0.670	-0.172	0.349	2.735
サンプルサイズ	188 (47 都道府県 * 4 年分)					
決定係数 R^2	0.408					
補正決定係数 R^2	0.385					
標準誤差	0.106					

† p<0.1 * p<0.05 ** p<0.01

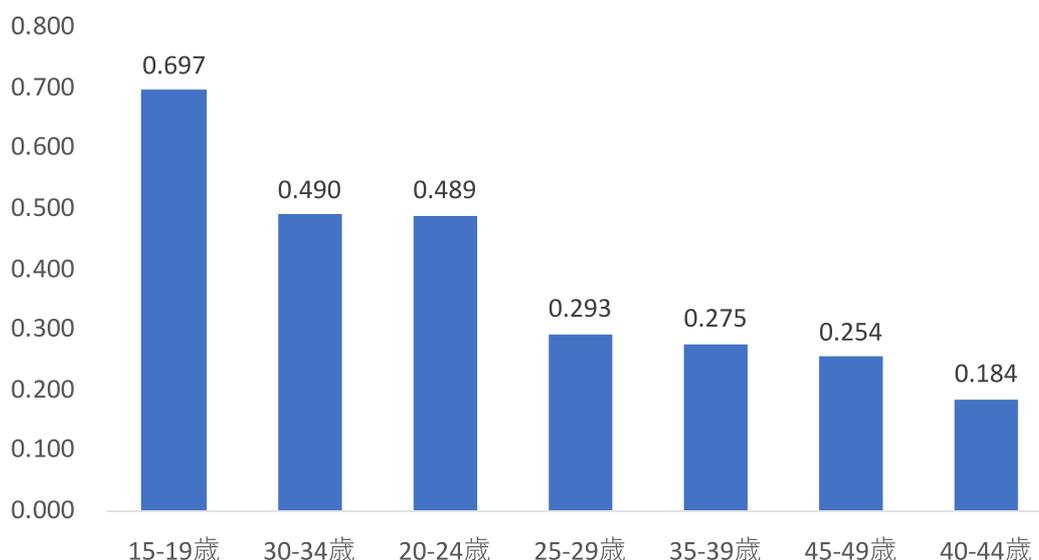


図6 ランダムフォレストによる変数重要度 (Importance)

機械学習アルゴリズムの1つであるランダムフォレストを用いて変数の重要度をみると、重回帰モデル同様、特に15-19歳の転入超過率の影響度が大きいことが分かる。

ここで、TFRを目的変数、人口1000人あたり学生数を説明変数として単回帰分析を行ったところ（表5）、1%水準で有意であった。回帰係数が負値であるため、人口あたりの学生数が増加することでTFRに下方圧力が発生することが示唆される。

表5 都道府県別TFRに係る単回帰分析結果

変数	回帰係数	標準誤差	t値
切片	1.907 **	0.119	16.016
人口1000人あたり学生数	-0.147 **	0.044	-3.354
サンプルサイズ	47		
決定係数 R ²	0.200		
補正決定係数 R ²	0.182		
標準誤差	0.126		

† p<0.1 * p<0.05 ** p<0.01

4.2 市区町村別TFR-Bによる京都府分析

次に京都府の市区町村別TFR-Bを目的変数、府内市区町村別5歳階級別女性転入超過率を特徴量（説明変数）とし、上記4.1と同様に重回帰モデル式を適用した。以下表5のとおり15-19歳、20-24歳、25-29歳及び30-34歳の女性の転入超過率で有意となった。

特徴量ゼロモデルでみると、1.42 (95% C.I. 1.39-1.45) となる。SSDSEの京都府2014-2017年の平均値は1.31であることから、一般的なTFRは低めに出る傾向が示唆される。

表6 京都府の TFR-B を用いた重回帰分析結果

変数	偏回帰係数	標準誤差	t 値	95% C.I. 下限	95% C.I. 上限	VIF
切片	1.420 **	0.017	82.780	1.386	1.454	
15-19 歳女性転入超過率	-0.021 **	0.006	-3.639	-0.032	-0.010	1.855
20-24 歳女性転入超過率	-0.007 *	0.003	-2.425	-0.014	-0.001	1.940
25-29 歳女性転入超過率	0.035 **	0.005	7.189	0.026	0.045	1.665
30-34 歳女性転入超過率	0.020 *	0.009	2.301	0.003	0.037	2.102
35-39 歳女性転入超過率	-0.002	0.014	-0.117	-0.030	0.026	1.774
40-44 歳女性転入超過率	-0.026	0.018	-1.438	-0.061	0.010	1.213
45-49 歳女性転入超過率	0.009	0.024	0.377	-0.038	0.056	1.368
サンプルサイズ	128 (32 市区町村 * 4 年分)					
決定係数 R ²	0.581					
補正決定係数 R ²	0.557					
標準誤差	0.152					

* p<0.05 ** p<0.01

※TFR の標準誤差が 0.1 以上である「笠置町」「和束町」「南山城村」「伊根町」については、外れ値として除去した。

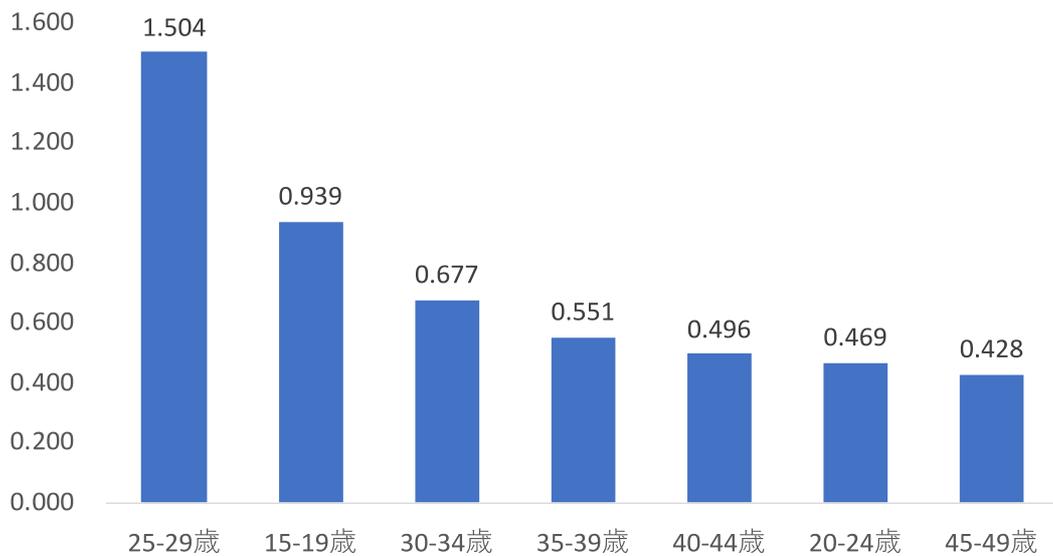


図7 京都府 TFR に係るランダムフォレストによる変数重要度 (Importance)

ランダムフォレストを用いて変数の重要度をみると、重回帰モデル同様、特に 25-29 歳及び 15-19 歳の転入超過率の影響度が大きいことが分かる。

第5章 結果の解釈

4章では、都道府県別及び市区町村別に分けて、TFR の分析を行った。その結果、共通して 15-19 歳及び 30-34 歳の女性の転入超過率が 5%未満の水準で有意であった。また、都道府県別と市区町村別でみた場合、有意である特徴量に差異があることも認められた。

15-19 歳の女性の転入超過は仮説のとおり TFR を薄めてしまう可能性があると考えられる。つまり、学生

を多く受け入れる京都府等の都道府県では、TFR が低めに出る傾向があるとみられる。

社会増減の要因を取り除いた特徴量ゼロモデルで考えると、京都府の TFR は、2014-2017 年の公表値平均 1.31 に対して、特徴量ゼロモデルは 1.42 (95% C.I. 1.39-1.45) と、ほぼ全国並 (全国:1.44) の水準であることが分かる。京都府が長きにわたり TFR が低い理由と、筆者の実感とのズレはここにあるのかもしれない。以上のことから TFR の公表値は必ずしも、「子育て環境の良さ」を示しているとは限らないと考えられる。したがって、TFR を目的変数にとるような分析を行うに際しては、TFR には攪乱要因が含まれるため「TFR の高低=子育て環境の良し悪し」と捉えることには注意が必要である。

30-34 歳の女性の転入超過については、まさに結婚し、出産を迎える女性の出産コア年齢であることから、当該年齢階級の女性の転入を促す施策または流出を防止する施策が、自治体別でみた場合の TFR の向上に資すると考えられる。ただし、このような施策について、自治体別で見た場合は有効かもしれないが、日本全体のマクロの視点からみた場合「パイを奪い合う」ことにも留意が必要である。全国値としての社会増減数は海外との転出入を除くと、ゼロだからである。我が国の TFR の向上に向けた根本的な体質改善には、やはり「未婚化」や「晩婚化」などにフォーカスした研究が引き続き必要となるであろう。

参考文献

(1) 内閣府：「平成 19 年版少子化社会白書」第 2 章第 1 節

<<https://www8.cao.go.jp/shoushi/shoushika/whitepaper/measures/w-2007/19webhonpen/index.html>>

(2) P.F.ドラッカー：「すでに起こった未来」ダイヤモンド社 (1994 年)

(3) 東北大学経済学研究科吉田研究室：「子ども人口時計」

<<https://sites.google.com/view/caestop/>>

(4) 松村波、熊野翔、川田瑛貴：「市区町村別でみる合計特殊出生率推移の特徴分析」2019 年度統計データコンペティション統計活用奨励賞 (大学生・一般の部)

(5) 小野恵子、宮内はじめ、白松俊、河口信夫、五十嵐康伸：「日本の全市町村における人口の自然増減の分布と説明要因」2018 年度統計データコンペティション特別賞 (大学生・一般の部)