

2021年度 統計データ分析コンペティション

統計数理賞 [大学生・一般の部]

若年女性の社会増減についての要因分析

三輪 俊太郎

(滋賀大学大学院データサイエンス研究科)

論文の概要

地方の人口減少は自然増減と社会増減の2つの側面があることから、女性の社会増減について145項目の変数を用いてランダムフォレスト回帰により若年女性の社会移動にかかわる要因を抽出し、産業面では特定産業の集積、産業面以外では婚姻数等が若年女性人口と相関していることを示している。

論文審査会コメント

先行研究も参考に分析を実施していて、独自の指標形成も評価できる。古典的計量実証と機械学習による探索とを連携させようとする努力をきちんと試みており、実証研究の未来の方向性を示した論文として評価できる。

若年女性の社会増減についての要因分析

三輪 俊太郎^{*1}

※1 滋賀大学データサイエンス研究科・宮崎県

1 問題と目的

人口減少は、平成末期の地方制度におけるキーワードである⁽¹⁾。まち・ひと・しごと創生長期ビジョン（令和元年改訂版）では、地方において地域社会の担い手が減少することにより、地域経済が縮小し、それが更に人口減少を加速させ負の連鎖に陥っていると指摘し、地方の未来に向け「活力ある地域社会」の実現と、「東京圏への一極集中」の是正を目指すと掲げている⁽²⁾。

人口減少には、死亡数と出生数の差による自然増減と、地域間の移動に伴う社会増減の2つの側面がある。地方の人口減少は、出生率の低下と地方から東京圏への人口移動とが複合して生じており、人口移動について東京圏への年齢階層別の転入超過数を見ると、15～29歳の若年者が大部分を占めている⁽²⁾。人口は限られた資源であり、その社会移動に関わる要因を確認することは、地方の持続可能性に向けた対策を行うにあたり重要であると考えられる。

若者の移動については、独立行政法人労働政策研究・研修機構の「移動する若者・しない若者」に詳しい。その中で林⁽³⁾は、若年女性の移動に着目しており、高い経済成長率を維持した時代は男性が都市部に集まっていたが、90年代から傾向が変わり、2000年以後は都市部の20～39歳の若年層において男性よりも女性の比率が高くなっていることを明らかにした。また、女性は男性よりも転入転出が少ないことから、都市部に集まり、都市部から出て行かない傾向にあると指摘している。

地方の人口減少は、自然増と社会増減の2つの側面があると先に述べたが、女性は出産を担うため、その両者に関わると言える。言い換えると、地方は都市部に比べ出生率が高い傾向にあるものの、社会移動により若年女性の総人口が少なければ、人口の持続可能性は乏しいことになる。また、林の指摘から、男性と女性とでは社会移動の傾向が異なっているため、若年女性のみを対象として社会移動の要因を探ることには大きな意義があると考えられる。

若年女性の社会移動要因の先行研究として、先の林⁽³⁾は、ジェンダー指数と20～39歳女性の都道府県間転入超過数とは相関があると指摘している。また、10歳階級別に都道府県間の女性の純転入率と社会・人口統計体系の各種指標との関係を見た田村ら⁽⁴⁾は、20歳代と30歳代とは指標との相関関係が大きく異なっており、20歳代では給与や物価などの経済的な要因と純転入率とに大きな正の相関があるものの、30歳代では経済的な要因との相関が見られず、婚姻率や教育費といった家庭の要因に相関が見られることを明らかにしている。ただし、どちらの研究も、女性の社会移動に関わる要因を網羅できていないこと、1時点のみを対象とした分析であることが研究上の限界であると考えられる。

そこで本研究では、若年女性の社会移動に関わる要因について、網羅的かつ複数年検討し、地方の持続可能性に向けた施策立案に資することを目的とする。

2 分析方針と方法

2.1 分析方針と用いるデータ

本研究では、若年女性の社会移動に関わる要因について、できる限り網羅的に検討することから、都道府県レベルを対象とした。その理由は、公的統計の範囲が都道府県と市町村とでは大きく異なり、都道府県レベルの統計の方が非常に多いためである。

また、目的変数である若年女性の人口として、国勢調査における各都道府県の30～34歳人口を扱うこととし、2005年・2010年・2015年の統計を分析の対象とすることとした。その理由は、以下の3点である。①若年者の社会移動は、15～29歳が大部分であり、30～34歳の人口はその結果として捉えることができるため。②2005年・2010年・2015年の対象年において、5歳階級を扱うことにより各調査年でコホートが重複せず、コホートの影響を取り除けるため。③国勢調査を用いる理由は、住民基本台帳があくまで住民票上の移動を把握するものであり、住民票を動かさずに移動しているケースについては社会移動を把握できないため。

要因を網羅的に検討するにあたり、因果連鎖分析⁽⁵⁾を用いた。具体的には、若年女性の社会移動には、社会移動に結びつく直接的な行動として「進学」「就職」「Uターン」「Iターン」があると想定し、それぞれに関わる要因をできる限り挙げ、後の変数選択における方針とした（因果連鎖分析グラフは省略）。

続いて、因果連鎖分析に基づき、本研究で用いる変数を選択した。選択した統計の一覧は表1の通りである。なお、人口や世帯数などの規模により、現数値では都道府県間の比較が困難であると考えられる変数については、人口千人あたりにするなどのスケーリングを行った。

変数は代理変数を含め、できる限り網羅するように選択した。例えば、「交通手段の利便性」を直接的に示す統計はないため、交通手段の選択肢が少ない地方はガソリンの消費が多くなると考え、ガソリン消費量を代理変数とした。また、因果連鎖分析内には要因として挙げられたものの、統計がない、対象年に調査がされていない等の理由から扱えなかったものもある（「愛着」「地域の情報発信」など）。

結果として、47都道府県×3時点×145変数のパネルデータとなった。データの取得には提供されたSSDSEの他、e-stat⁽⁶⁾、統計ダッシュボード⁽⁷⁾、RESAS⁽⁸⁾を用いた。

表1 使用した統計一覧

分野	変数名	2005年 (N=47)		2010年 (N=47)		2015年 (N=47)		スクーリング	出典
		平均	標準偏差	平均	標準偏差	平均	標準偏差		
人口・世帯	30 34歳女性人口	35.07	3.75	30.34	2.42	26.72	2.26	人口千人当	国勢調査 (総務省)
	1 一般世帯数_総数家族類型	368.03	28.25	386.26	29.76	401.88	28.31		
	3 一般世帯数_父子母子世帯	66.78	12.27	80.20	15.79	75.51	14.31		
	4 一般世帯数_3世代世帯	115.70	50.93	96.93	44.71	78.18	37.08		千世帯数当
	5 共働き世帯数	297.52	48.89	274.04	43.53	274.01	41.69		
	6 婚姻件数	5.22	0.53	5.06	0.56	4.65	0.50	人口千人当	SSDSE
	7 離婚件数	1.98	0.24	1.89	0.22	1.73	0.20	人口千人当	SSDSE
	8 平均婚姻年齢_初婚の妻	27.69	0.41	28.53	0.38	29.15	0.38		人口動態統計 (厚生労働省)
福祉・社会保障	9 現に保護を受けた生活保護被保護実世帯数【世帯】	18.70	9.00	23.70	10.69	26.59	11.52	千世帯数当	被保護者調査 (厚生労働省)
	10 要保護_要保護児童生徒数合計	9.37	4.15	11.19	4.25	11.09	3.95	人口千人当	要保護及び準要保護 児童生徒数 (文部科学省)
自然環境	11 年平均気温	15.20	2.44	15.79	2.24	15.86	2.15		SSDSE
	12 年間降水日数	112.91	37.32	126.00	29.95	124.43	19.88		過去の気象データ (気象庁)
	13 年間雪日数	36.13	32.89	32.00	33.14	33.17	31.15		
	14 年間日照時間	1924.79	241.44	1873.68	216.51	1917.07	154.47		全国_都道府県市町村 別面積調(国土交通 省)
15 可住地面積	126.45	72.96	129.71	76.57	133.83	80.43	人口千人当		
生活時間	16 通勤通学の平均時間_女性有業者	30.66	6.42	31.51	6.71	34.40	6.32		注1
	17 育児の平均時間_女性有業者	15.43	3.23	17.57	3.15	20.85	4.23		注1
	18 育児の平均時間_女性無業者	30.19	7.91	28.40	6.30	25.81	7.76		注1
	19 3次活動の平均時間_女性有業者	309.17	21.25	301.45	15.95	290.23	19.45		
居住	20 映画館、劇場、音楽会場等の数	0.04	0.01	0.04	0.01	0.04	0.01		衛生行政報告例(厚生 労働省)、社会生 活統計指標(総務省)
	21 ガソリン販売量	502.73	100.07	469.37	85.58	441.44	87.42		統計でみる都道府県 のすがた(総務省)
教育	22 大学・短大数	0.01	0.00	0.01	0.00	0.01	0.00		注3
	23 専修学校数	0.03	0.01	0.03	0.01	0.03	0.01		注4
	24 大学・短大学生数	16.21	9.25	15.98	9.31	16.04	9.15	人口千人当	SSDSE
	25 専修学校生徒数	4.71	2.70	4.06	1.96	4.20	1.94		
	26 女性の高等学校卒業者のうち進学者数	2.33	0.28	2.32	0.23	2.34	0.25		学校基本調査(文部 科学省)
	27 女性の高等学校卒業者のうち就職者数	0.88	0.34	0.69	0.27	0.74	0.26		
	28 幼稚園定員数	18.69	3.99	18.12	3.99	15.44	4.30		統計ダッシュボード
	29 保育所等定員数	19.16	5.68	18.88	5.28	21.79	5.71		SSDSE
	30 教育費_都道府県財政	99115.63	14936.25	95907.15	14593.67	99194.63	16761.16		地方教育費調査(文 部科学省)
	31 女性の地元就職進学者数	4.76	2.62	2.79	1.24	3.31	1.76		学校基本調査(文部 科学省)*RESASか ら取得
32 女性の流入大学・短大進学者数	2.05	1.14	2.02	1.12	2.03	1.14		注5	
33 女性の流出大学・短大進学者数	2.56	0.28	2.51	0.24	2.50	0.28		注6	
34 教育関係費	43229.45	7533.44	44046.13	6681.05	38176.53	6830.44		注7	
家計	35 可処分所得_二人以上の世帯のうち_勤労者世帯	445816.55	60673.23	429705.28	42148.73	430136.43	42602.93		家計調査(総務省)
	36 国内銀行預金残高人口1人当たり	323.28	148.92	361.31	165.63	420.78	202.94		社会生活統計指標 (総務省)
	37 郵便貯金残高人口1人当たり	154.98	26.57	128.51	20.35	128.26	19.76		
健康・医療	38 一般診療所数	0.77	0.12	0.79	0.12	0.81	0.12	人口千人当	SSDSE
労働	39 所定内給与月額_女	206.51	18.00	212.34	18.43	224.53	20.32		賃金構造基本統計調 査(厚生労働省)
	40 月間有効求人数_一般	135.02	31.74	89.89	15.93	140.20	27.03		SSDSE
	41 完全失業者数	30.37	5.99	32.48	5.03	20.74	2.69		
	42 非労働力人口女	224.97	13.31	220.05	15.35	217.86	15.78		
	43 就業者数女	207.62	12.47	204.60	11.04	210.34	11.84		
	44 就業者数女_家族従事者	22.73	6.73	17.96	5.67	15.55	5.30		注8
	45 就業者数女_A_農業_林業	13.89	7.80	10.52	6.30	10.03	6.07		
	46 就業者数女_B_漁業	0.55	0.62	0.45	0.51	0.40	0.47		
	47 就業者数女_D_建設業	6.50	0.91	5.54	0.89	5.82	0.91		
	48 就業者数女_E_製造業	29.42	9.62	25.47	8.50	24.73	8.65		
49 就業者数女_F_電気_ガス_熱供給_水道業	0.30	0.07	0.31	0.10	0.32	0.12			
50 就業者数女_G_情報通信業	2.28	1.43	2.15	1.99	2.06	2.02			
51 就業者数女_H_運輸業_郵便業	3.65	1.04	3.97	1.12	3.88	1.20			
52 就業者数女_I_卸売業_小売業	41.60	2.27	38.45	3.15	37.85	2.88			
53 就業者数女_J_金融業_保険業	5.72	0.82	5.74	1.36	5.60	1.26			
54 就業者数女_K_不動産業_物品賃貸業	2.49	0.98	2.55	1.03	2.89	1.00			
55 就業者数女_L_学術研究_専門_技術サービス業	3.96	1.10	3.84	1.36	4.22	1.47			
56 就業者数女_M_宿泊業_飲食サービス業	17.61	1.78	16.89	1.85	16.48	1.76			
57 就業者数女_N_生活関連サービス業_娯楽業	10.93	0.94	10.40	1.08	10.17	1.08			
58 就業者数女_O_教育_学習支援業	11.81	1.14	11.79	1.11	12.32	1.19	人口千人当	国勢調査 (総務省)	
59 就業者数女_P_医療_福祉	35.06	5.52	40.39	6.45	46.51	7.26			
60 就業者数女_Q_複合サービス業	2.18	0.54	1.69	0.56	1.95	0.60			
61 就業者数女_R_サービス業他に分類されないもの	12.56	1.97	9.33	1.69	10.07	1.82			
62 就業者数女_S_公務他に分類されるものを除く	4.41	0.84	4.40	0.85	4.90	0.96			
63 就業者数女_T_分類不能の産業	2.66	1.44	8.46	4.45	7.76	3.96			
64 従業者数女_A_管理的職業従事者	1.29	0.23	1.48	0.23	1.80	0.25			
65 従業者数女_C_専門的_技術的職業従事者	30.58	3.10	33.06	3.17	36.85	3.50			
66 従業者数女_D_事務従事者	51.77	5.35	49.30	4.41	51.95	3.72			
67 従業者数女_D_販売従事者	28.34	1.36	26.23	1.22	25.49	1.34			
68 従業者数女_E_サービス職業従事者	36.72	2.78	38.07	3.43	39.67	3.82			
69 従業者数女_F_保安職業従事者	0.42	0.12	0.46	0.13	0.53	0.14			
70 従業者数女_G_農林漁業従事者	13.85	7.72	10.29	6.15	9.70	5.88			
71 従業者数女_H_生産工程従事者	26.20	8.17	22.33	6.66	20.75	6.44			
72 従業者数女_I_輸送_機械運転従事者	0.51	0.14	0.44	0.11	0.47	0.11			
73 従業者数女_J_建設_採掘従事者	0.69	0.22	0.44	0.15	0.47	0.14			
74 従業者数女_K_運搬_清掃_包装等従事者	14.71	1.84	13.97	1.85	14.87	1.90			
75 従業者数女_L_分類不能の職業	2.54	1.34	8.28	4.23	7.47	3.76			

表1 使用した統計一覧(続き)

分野	変数名	2005年 (N=47)		2010年 (N=47)		2015年 (N=47)		スケーリング	出典
		平均	標準偏差	平均	標準偏差	平均	標準偏差		
76	H#指数 業種小分類 女性	0.04	0.01	0.05	0.01	0.05	0.01		注9
77	特化係数 就業者数女 A 農業 林業	1.41	0.76	1.46	0.84	1.46	0.84		注9
78	特化係数 就業者数女 B 漁業	1.30	1.46	1.31	1.50	1.33	1.57		注9
79	特化係数 就業者数女 C 建設業	1.02	0.11	1.02	0.12	1.02	0.13		注9
80	特化係数 就業者数女 D 製造業	1.06	0.31	1.09	0.33	1.09	0.34		注9
81	特化係数 就業者数女 E 電気 ガス 熱供給 水道業	1.04	0.23	1.03	0.29	1.02	0.34		注9
82	特化係数 就業者数女 G 情報通信業	0.67	0.45	0.60	0.48	0.58	0.51		注9
83	特化係数 就業者数女 H 運輸業 郵便業	0.85	0.26	0.86	0.28	0.85	0.31		注9
84	特化係数 就業者数女 I 卸売業 小売業	0.98	0.06	0.98	0.06	0.98	0.06		注9
85	特化係数 就業者数女 J 金融業 保険業	0.91	0.15	0.88	0.16	0.89	0.16		注9
86	特化係数 就業者数女 K 不動産業 物品賃貸業	0.75	0.33	0.74	0.29	0.77	0.27		注9
87	特化係数 就業者数女 L 学術研究 専門 技術サービス業	0.82	0.26	0.79	0.24	0.79	0.25		注9
88	特化係数 就業者数女 M 宿泊業 飲食サービス業	0.98	0.11	1.01	0.09	1.00	0.09		注9
89	特化係数 就業者数女 N 生活関連サービス業 娯楽業	0.99	0.08	1.00	0.09	0.99	0.09		注9
90	特化係数 就業者数女 O 教育 学習支援業	1.00	0.13	1.01	0.11	1.01	0.10		注9
91	特化係数 就業者数女 P 医療 福祉	1.06	0.15	1.08	0.15	1.08	0.14		注9
92	特化係数 就業者数女 Q 複合サービス業	1.15	0.24	1.28	0.38	1.23	0.33		注9
93	特化係数 就業者数女 R サービス業中に分類されないもの	0.88	0.17	0.90	0.16	0.90	0.16		注9
94	特化係数 就業者数女 S 公務員に分類されるものを除く	1.08	0.21	1.09	0.20	1.10	0.20		注9
95	特化係数 就業者数女 T 分類不能の産業	0.71	0.42	0.73	0.38	0.73	0.38		注9
96	H#指数 職種小分類 女性	0.07	0.01	0.07	0.01	0.07	0.01		注9
97	特化係数 従業者数女 A 管理的職業従事者	0.92	0.18	0.48	0.49	0.96	0.16		注9
98	特化係数 従業者数女 B 専門的 技術的職業従事者	1.01	0.10	0.60	0.44	1.02	0.08		注9
99	特化係数 従業者数女 C 事務従事者	0.92	0.12	0.60	0.36	0.94	0.09		注9
100	特化係数 従業者数女 D 販売従事者	0.97	0.07	0.56	0.42	0.97	0.06		注9
101	特化係数 従業者数女 E サービス職業従事者	1.00	0.07	0.62	0.43	1.04	0.06		注9
102	特化係数 従業者数女 F 保安職業従事者	0.92	0.27	0.47	0.50	0.94	0.26		注9
103	特化係数 従業者数女 G 農林漁業従事者	1.41	0.76	0.76	0.94	1.45	0.85		注9
104	特化係数 従業者数女 H 生産工程従事者	1.09	0.30	0.62	0.55	1.12	0.31		注9
105	特化係数 従業者数女 I 輸送 機械運転従事者	0.98	0.26	0.48	0.51	0.97	0.23		注9
106	特化係数 従業者数女 J 建設 採掘従事者	1.12	0.32	0.56	0.61	1.09	0.31		注9
107	特化係数 従業者数女 K 運搬 荷扱 包装等従事者	0.98	0.11	0.54	0.48	0.99	0.11		注9
108	特化係数 従業者数女 L 分類不能の職業	0.72	0.41	0.42	0.50	0.73	0.41		注9
109	従業者数 民営	401.78	54.66	422.83	53.56	427.24	51.69		注10
110	従業者数 民営300人以上	40.67	20.50	45.84	23.66	47.26	24.77	人口千人当	事業所・企業統計及 び経済センサス(経 済産業省)
111	民営事業所数	46.56	5.63	48.23	5.26	46.46	4.97		注10
112	民営事業所数 従業者300人以上	0.07	0.03	0.07	0.03	0.08	0.03		注10
113	1人当たり県民所得	2807.51	585.35	2648.34	473.48	2874.53	499.82		注11
114	県内総生産額 農林業	1619.47	439.10	1872.97	486.57	2088.02	621.84		注11
115	県内総生産額 水産業	4810.18	3895.20	4978.58	3441.98	6072.48	3875.29		注11
116	県内総生産額 鉱業	15831.53	10819.01	13951.08	6413.95	14992.04	6733.89		注11
117	県内総生産額 製造業	10216.07	2907.86	10554.86	2920.89	11881.03	2992.49		注11
118	県内総生産額 建設業	4965.17	1282.36	5105.23	1050.08	6116.21	1514.57		注11
119	県内総生産額 電気 ガス 水道 廃棄物処理業	52264.63	16575.61	53261.30	17811.28	55373.87	24760.60	当該産業 就業者 千人当	県民経済計算 (内閣府)
120	県内総生産額 卸売 小売業	5239.08	2974.71	5108.30	1764.99	5679.24	1814.41		注11
121	県内総生産額 金融 保険業	17710.31	6258.43	14073.82	2101.31	14469.45	2211.20		注11
122	県内総生産額 不動産業	60588.02	15249.22	67139.03	14723.86	64178.30	14308.58		注11
123	県内総生産額 運輸 郵便業	8159.64	1776.19	7735.72	1217.51	8909.25	1622.16		注11
124	県内総生産額 情報通信業	17494.09	5132.79	19827.30	3542.73	19956.33	3377.23		注11
125	県内総生産額 公務	11310.58	2662.79	11430.00	1509.39	11327.61	1724.35		注11
126	特化係数 県内総生産額 農業	1.72	1.15	1.70	1.15	1.71	1.29		注11
127	特化係数 県内総生産額 林業	1.99	1.88	1.98	1.92	2.06	2.12		注11
128	特化係数 県内総生産額 水産業	1.76	1.89	1.73	1.87	1.72	1.95		注11
129	特化係数 県内総生産額 鉱業	1.46	1.25	1.40	1.87	1.43	1.85		注11
130	特化係数 県内総生産額 製造業	1.05	0.41	1.09	0.41	1.08	0.42		注11
131	特化係数 県内総生産額 建設業	1.12	0.23	1.11	0.26	1.13	0.41		注11
132	特化係数 県内総生産額 電気 ガス 水道 廃棄物処理業	1.20	0.62	1.16	0.60	1.09	0.40		注11
133	特化係数 県内総生産額 卸売 小売業	0.80	0.21	0.82	0.19	0.83	0.19		注11
134	特化係数 県内総生産額 金融 保険業	0.83	0.19	0.83	0.19	0.83	0.20		注11
135	特化係数 県内総生産額 不動産業	0.98	0.17	0.96	0.16	0.96	0.16		注11
136	特化係数 県内総生産額 運輸 郵便業	0.98	0.20	0.97	0.18	0.97	0.19		注11
137	特化係数 県内総生産額 情報通信業	0.68	0.28	0.69	0.27	0.67	0.29		注11
138	特化係数 県内総生産額 公務	1.17	0.35	1.17	0.34	1.18	0.35		注11
139	特化係数 県内総生産額 第2次産業	1.07	0.31	1.09	0.30	1.09	0.30		注11
140	特化係数 県内総生産額 第3次産業	0.96	0.11	0.96	0.10	0.95	0.11		注11
141	GGI政治	0.03	0.02	0.05	0.04	0.07	0.04		注11
142	GGI経済	0.46	0.03	0.50	0.03	0.58	0.03		注11
143	GGI教育	0.98	0.03	0.99	0.03	1.00	0.03		注11
144	GGI健康	1.00	0.01	0.99	0.01	0.99	0.01		注11
145	GGI総合	0.62	0.01	0.63	0.01	0.66	0.01		注11

注1 2006年、2011年、2016年の統計値をそれぞれ2005年、2010年、2015年の値とした。

注2 「常設映画館数」「劇場、音楽会場等の数」の合計。2005年の北海道の「劇場～」値が欠測のため、2010年の値で補間した。

注3 「大学数」「短期大学数」の合計。

注4 「大学学生数」「短期大学学生数」の合計。

注5 「流入大学進学者」「流入短大進学者」の合計

注6 「流出大学進学者」「流出短大進学者」の合計

注7 「教育費」「教育施設費」の合算

注8 2015年の山梨県の漁業(女)に最低値の「10」で補間。

注9 マスキング等の箇所は0で補間。

注10 2006年、2009年、2012年、2014年、2016年のデータを用いて、2005年、2010年、2015年データを線形補間した。

注11 名目値。2005年の値(H23年基準)をH17年基準での2005年→2006年の増減割合で推計。また、基準改訂時に産業分類が変更となり、H23年基準の「8 宿泊・飲食サービス業」「12 専門・科学技術、業務支援サービス業」「14 教育」「15 保健衛生・社会事業」「16 その他のサービス」は対応なしのため削除。

2.2 独自に作成した指標

以下の変数については、独自に作成を行った。

①特化係数（各業種の県民総生産、各業種の就業者数、各職種の従業者数）

特化係数とは、地域における特定の産業の相対的な集積度を示す指標であり、地方の産業構造を簡易に把握することができる⁽⁹⁾。算出方法は以下の通りである。

$$\text{地域における産業Aの付加価値額特化係数} = \frac{\text{地域における産業Aの付加価値額の比率}}{\text{全国における産業Aの付加価値額の比率}}$$

②ハーフィンダール・ハーシュマン指数（各業種の就業者数、各職種の従業者数）

地域の産業・雇用の多様性（あるいは集中度）を示す指標として、ハーフィンダール・ハーシュマン指数（HHI）を作成した⁽¹⁰⁾。HHIは各分野の比率を2乗したものの総和であり、値は0から1の範囲を取り、0に近いほど産業・雇用の多様性が高いことを示す。算出方法は以下の通りである。

$$\text{当該地域の業種HHI} = \sum_{i=1}^n C_i^2$$

C_i ：当該地域のi番目の業種の比率，n：業種数

③ジェンダーギャップ指数

各都道府県における男女間の不均衡を示す指標として、世界経済フォーラムの算出方法に基づき、ジェンダーギャップ指数（GGI）を作成した⁽¹¹⁾。GGIは経済分野、教育分野、健康分野、政治分野の4つの指標と、その4指標を平均した総合指標がある。表2はGGIの算出方法である。

表2 ジェンダーギャップ指数の算出方法

分野	項目	算出方法	出典
経済	労働力の男女比	女性就業者数(15～64歳)÷男性就業者数(15～64歳)	国勢調査（総務省）
	所得の男女格差	「きまって支給する給与×12月+年間賞与とその他特別給与種」の男女比 (該当する統計がなかったため、算出しなかった。)	賃金構造基本統計調査（厚生労働省）
	同一労働、同一賃金		
	管理的職業従事者の男女比	管理的職業従事者数（女性）÷管理的職業従事者数（男性）	国勢調査（総務省）
	専門・技術的職業の男女比	専門的・技術的職業従事者数（女性）÷専門的・技術的職業従事者数（男性）	国勢調査（総務省）
教育	識字率の男女比		
	初等教育就学率の男女比	(日本において地域による男女差はないと考え、算出しなかった。)	
	中等教育就学率の男女比	(高校生、専修学校高等課程、特別支援学校(H17年度は盲学校・聾学校・養護学校)高等部、中等教育学校後期課程の男女別学生数÷15-17歳の人口)の男女比	学校基本調査(文部科学省)、国勢調査(総務省)
	高等教育就学率の男女比	(大学・大学院生、短大生、専修学校専門課程の男女別学生数÷15歳以上の人口)の男女比	学校基本調査(文部科学省)、国勢調査(総務省)
健康	出生時の男女比	(出生数(女)÷出生数(男))÷0.944	人口動態調査(厚生労働省)
	健康寿命の男女比	(健康寿命(女)÷健康寿命(男))÷1.06	引用文献(12)
政治	都道府県議会議員の男女比	女性議員数÷男性議員数	地方公共団体の議会の議員及び長の所属党派別人員調査(総務省)
	都道府県知事の在任年数の男女比(直近50年間)	女性知事の任期÷男性知事の任期	wikipediaより確認

※各指標は、世界経済フォーラムの算出方法に基づき、それぞれの項目の値を、標準偏差に応じた重み付けをした上で合計し算出した。

各指標 = \sum (各項目の値×各項目の重み)

各項目の重み = (0.01÷各項目の標準偏差) × 指標内の項目の(0.01÷各項目の標準偏差)の合計

2.3 分析方法

本研究の分析方法として、ランダムフォレスト回帰によるSHAP値を求めることとした。

(1) ランダムフォレスト⁽¹³⁾とは

ランダムフォレストとは、決定木を元とした機械学習手法であり、算出される説明変数の重要度により機械学習手法の中では結果の解釈性が高いこと、サンプルサイズより説明変数が多くとも安定した結果が出せること、多重共線性や変数の分布形状を問わずに分析可能であることが主な特徴である。

(2) SHAPとは

SHAP(SHapley Additive exPlanations)⁽¹⁴⁾とは、機械学習で得られた予測モデルの説明性や解釈性を高めるための指標であり、予測結果について説明変数の貢献度(予測への寄与度)単位に分解したものである。

(3) 時系列性を取り除く固定効果変換

本研究では3時点のパネルデータを扱うが、ランダムフォレストにそのままデータを入力すると、データには都道府県による異質性や時点による異質性が混在することになる。

そこで、パネルデータ分析を参考に、都道府県の固定効果変換を行うこととした。具体的には、都道府県ごとに異なるが時間を通じて一定である変数をダミー項に加え、各変数の値は以下の式の通り、各時点における全都道府県の平均値を控除した。

$$y_{it} - \bar{y}_t = \beta(x_{it} - \bar{x}_t) + \varepsilon_{it} - \bar{\varepsilon}_t$$

y_{it} ：目的変数， β ：パラメータ， x_{it} ：説明変数， ε_{it} ：誤差項， i ：都道府県， t ：時点

この変換により、時点による異質性はキャンセルアウトされる。なお、この変換は関数形として線形を仮定していること、ランダム効果はないと前提している点に留意する必要がある。

3 結果

scikit_learn パッケージの RandomForestRegressor を使い、各都道府県の 30~34 歳女性人口を目的変数、それ以外の変数を説明変数とするランダムフォレスト回帰を行った。

パラメータチューニング及び訓練後のモデルによる予測にあたり、141 件 (47 都道府県×3) のサンプルサイズを訓練データ (サンプルサイズの 64%)、バリデーションデータ (同 16%)、テストデータ (同 20%) の 3 つに分け、テストデータは予測にのみ用いた。パラメータチューニングには optuna を使い、kfold 法 3 分割の交差検証により、目的関数として RMSE (二乗平均平方根誤差) を最小化するようにパラメータチューニングを行った。テストデータによる予測と実測値による決定係数は 0.774 であった (図 2)。

図 2 予測値と実測値のプロット

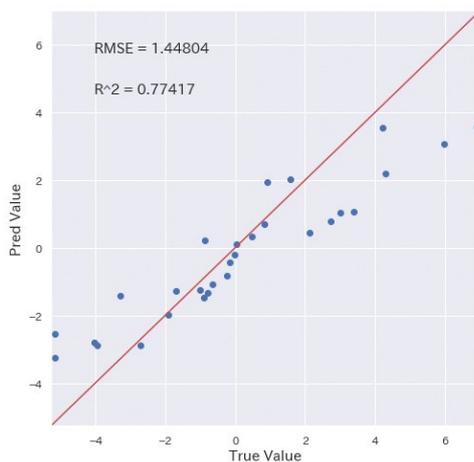


図 3 SHAP 値のサマリ

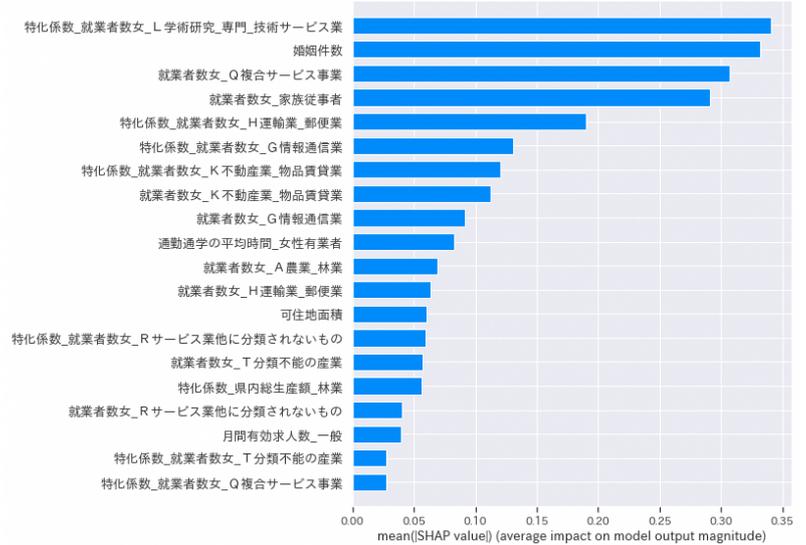
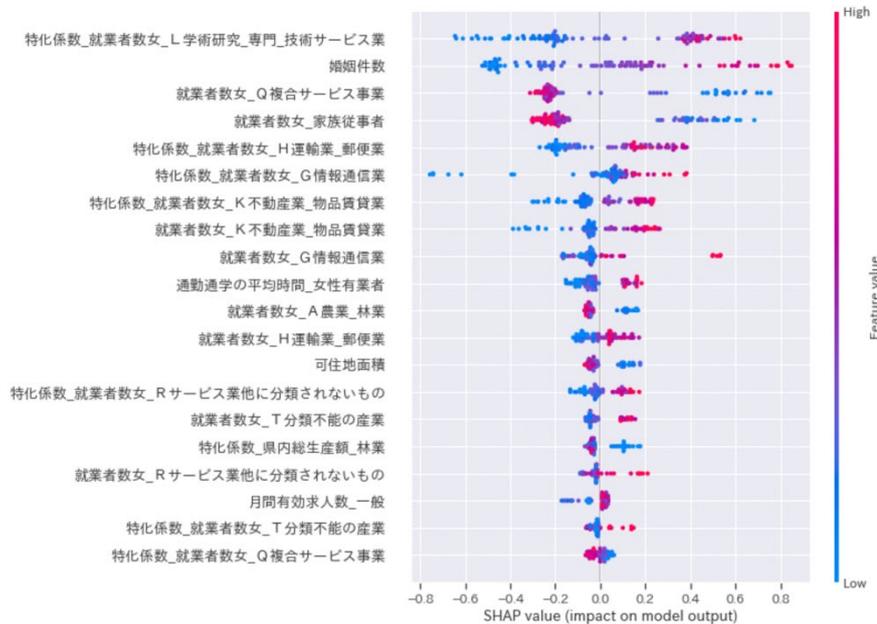


図 4 SHAP 値のサマリプロット



その後、訓練後のモデルと訓練データから、SHAP 値を算出した。

図 3 の SHAP 値は、予測における各説明変数の貢献度を棒グラフで一元的に示したものであり、図 4 の SHAP 値のプロットは、各説明変数と予測値との相関を示したものである。図 4 のグラフの横軸は予測値の正負のどちらの方向に寄与しているかを、プロットの色は各説明変数の値の高低を示しており、例えば一番上の「特化係数_就業者数女_L 学術研究_専門_技術サービス業」は、値が高いほど若年女性人口の予測値も高く、正の相関関係にあることを示している。

全体として、各都道府県の産業構造の違いが反映されており、「学術研究・専門・技術サービス業」「運輸業・郵便業」「情報通信業」「不動産業・物品賃貸業」の就業者比率が全国より高いと、若年女性人口の予測値

も高いことが示された。一方で、「複合サービス事業」や「家族従事者」は若年女性人口の予測値との負の相関関係を示していた。

また、婚姻件数も貢献度が二番目と高く、若年女性人口の予測値との正の相関関係にあった。

4 考察

4.1 産業面について

SHAP 値の算出により、若年女性の人口は業種別就業者数との関係性が強いことが示された。そのうち、正の相関が示された業種は、専門・科学技術・業務支援サービス業、運輸・郵便業、情報通信業、不動産業・物品賃貸業などであり、負の相関が示されたのは複合サービス事業や農業・林業であった。

日本全体の傾向との比較をするため、国民経済計算上の各業種の 2005 年から 2015 年の推移を確認した（表 3）。なお、本研究で主に用いた国勢調査上の業種と、国民経済計算上の業種とが名称含め必ずしも一致しない点に御留意いただきたい。

表 3 国民経済計算上の 2005 年～2015 年の増減率（％）

業種（経済活動別）	付加価値額（寄与度）	就業者数（寄与度） ※男女計・全年齢
専門・科学技術、業務支援サービス業	28.3% (1.7%)	23.3% (2.0%)
運輸・郵便業	5.1% (0.3%)	-1.8% (-0.1%)
情報通信業	-0.8% (0.0%)	5.4% (0.1%)
うち(1) 通信・放送業	10.4% (0.2%)	記載なし
うち(2) 情報サービス・映像音声文字情報制作業	-8.6% (-0.3%)	
不動産業	9.8% (1.1%)	12.2% (0.2%)
うち(1) 住宅賃貸業	4.3% (0.4%)	記載なし
うち(2) その他の不動産業	46.9% (0.7%)	
(参考)保健衛生・社会事業	30.9% (1.8%)	40.7% (3.5%)
国内総生産	1.0% (1.0%)	0.0% (0.0%)

※複合サービス業は、対応する業種がないため記載していない。

日本全体の付加価値額や就業者数が 2005 年から 2015 年にかけてほぼ横ばいであるのに対し、業種別では違いが見られた。

専門・科学技術・業務支援サービス業については、付加価値額・就業者数ともに大きな伸びが見られた。この業種は具体的には学術的研究の他、法律・会計・デザインやコンサルティングなどを手がけるもので、専門性や資格が求められるホワイトカラーの業種であり、対事業所向けのサービスが主であると考えられる。専門・科学技術・業務支援サービス業の特化係数と若年女性人口との相関（SHAP 値）が見られたことを踏まえると、表 3 の付加価値額や就業者数の伸びは都市部に偏っており、その中で、若年女性が就業機会を見出しているものと考えられる。そこには、「手に職をつけたい」と言った長期的なキャリア志向が関係しているかもしれない。

運輸・郵便業や情報通信業、不動産業についても、都市部への集積という点では、専門・科学技術・業務支援サービス業と同じことが言えそうである。

表 3 には記載していないが、若年女性人口との負の相関（SHAP 値）が見られた複合サービス業は、具体的には郵便局や農業協同組合である。全国に比較的満遍なく広がっている業種と考えられるが、その分都市部への集積が考えにくく、若年女性人口との負の相関に繋がっていると考えられる。

興味深い点として、医療・福祉（保健衛生・社会事業）については付加価値額や就業者数が大きく増加しているものの、2015 年時点での医療・福祉就業者数と若年女性人口との相関（原数値、表等には未記載）は-0.49 と、やや大きい負の相関が見られる。全国的な伸びがある一方で、都市部に偏った若年女性就業者の伸びではないことを意味しており、地方における若年女性の就業の受け皿となっている可能性がある（もしくは、若年女性以外が就業を担っている可能性がある）。

地方において医療・福祉業を伸ばすことは、若年女性の人口増加につながるであろうか。筆者にはそう思わない。

表 4 は、2010 年国勢調査における東京都の、業種ごとの 20-24 歳、25-29 歳、30-34 歳の女性の転入超過数である。国勢調査では 5 年前の居住地を調査しており、10 年ごとの調査では業種や年齢 5 歳階級ごとの各都道府県への転入・転出数を得ることができる。

医療・福祉業は、30-34 歳では東京都から転出しているものの、20-24 歳、25-29 歳では東京都への転入超過にあり、3つの年齢階級を合計すると、東京都への大きな転入超の状況である。30-34 歳で都市部からの転出があるとしても、一部に過ぎないと言える。また、医療・福祉業は少子高齢化で一定の需要は見込まれるものの、専門・科学技術・業務支援サービス業のように、産業集積によって地域の人口以上に需要が増えることは考えにくい。専門・科学技術・業務支援サービス業は地域を超えてサービスを提供可能であるのに対し、医療・福祉業はそれが難しいためである。

以上のことから、特定の産業面の集積が若年女性人口と相関していると言えるが、そこに地方が産業面で都市部と対抗することは困難であろうと考える。より小さな面積で産業集積化を図るか、よりニッチな部分で、若年女性を惹きつけるしかないのではないだろうか。

表4 業種ごとの東京都への転入超過数（2010年国勢調査）

現住地（2010）	20-24歳女 東京都	25-29歳女 東京都	30-34歳女 東京都	合計 東京都
A 農業、林業	-80	-211	-206	-497
B 漁業	-1	-1	-3	-5
C 鉱業、採石業、砂利採取業	0	1	7	8
D 建設業	232	35	-89	178
E 製造業	596	-159	-156	281
F 電気・ガス・熱供給・水道業	3	-11	-4	-12
G 情報通信業	3492	3666	1210	8368
H 運輸業、郵便業	776	362	30	1168
I 卸売業、小売業	6521	514	-253	6782
J 金融業、保険業	594	1155	1026	2775
K 不動産業、物品賃貸業	566	340	96	1002
L 学術研究、専門・技術サービス業	932	1212	489	2633
M 宿泊業、飲食サービス業	7760	-157	644	6959
N 生活関連サービス業、娯楽業	3076	-383	-494	2199
O 教育、学習支援業	1579	-856	-327	396
P 医療、福祉	4898	1092	-1303	4687
Q 複合サービス事業	32	-58	-27	-53
R サービス業（他に分類されないもの）	1230	671	79	1980
S 公務（他に分類されるものを除く）	792	-168	-127	497
T 分類不能の産業	2138	1252	632	4022

表4の転入超過数の合計で大きい業種は、情報通信業である。昨今ではリモートワークの必要性が叫ばれているが、若年女性の増加に向けては、情報通信業などで完全リモートワークが可能な企業と地方とが連携し就業に繋げる、という方向性が考えられる。また、農業については3つの年齢階級の全てで東京都からの転出超過であり、女性にとって農業で働きやすい環境を整える、といった施策も考えられる。

4.2 産業面以外について

産業面以外では、婚姻数と若年女性人口との正の相関（SHAP 値）が見られた。地方における、婚姻数の増加に向けた取組は有効であると考えられる。

女性の家族従業者と若年女性人口とでは負の相関（SHAP 値）が見られた。産業面との関わりが強いと思われるが、外に出て働く生活スタイルが多い場所に若年女性が集まっていることが窺えた。

その他、教育や福祉、ジェンダーといった要因については、SHAP 上の貢献度は認められなかった。

ジェンダーギャップ指数（GGI）については、元々が開発途上国も含めた指標として考えられており、日本国内のみを対象とした場合、健康分野や教育分野で男女格差が出にくいという問題点がある。GGI は、STEM 教育の比率など、給与に関わる部分での男女格差を反映できていないとの指摘もある⁽¹⁵⁾。紙面には記載していないが、GGI は2005年から2015年にかけて各都道府県で上昇傾向にあり、細かく検討する価値はあるかもしれない。

4.3 本研究で得られた知見とその限界

本研究は、若年女性の社会移動に関わると考えられる要因について、可能な限り網羅し分析を行った点に特徴がある。その上で、優先すべき要因を抽出できた点に研究上の価値があるものとする。

研究上の限界としては、まず、方法論上の問題が挙げられる。ランダムフォレストはあくまで予測のために用いられる手法であり、因果関係を明らかにするものではない。専門・科学技術・業務支援サービス業」の就業者を増やせば、若年女性人口は増える、といったことは本研究からは言えず、言えるのはあくまでも関連性の強さである。この分野については、Causal Forest⁽¹⁶⁾などの因果推論が可能な機械学習手法が開発されてきているため、研究上の発展として因果関係を精緻に分析する方向性が考えられる。

その他、3時点でのパネルデータを扱うにあたり、都道府県の固定効果変換を行った上でランダムフォレストを行ったが、より適切な方法があるものと思われる。また、ランダムフォレストの結果は、訓練データやパラメータにより多少変動しうるものである。補足すると、筆者はランダムフォレストを3度行い、SHAP 値やその順位について僅かな変動はあったものの、抽出される変数はほぼ変わらなかった。

方法論以外では、そもそもの問題設定として、人口という量的な側面のみで地方の持続可能性を考えるべきか、という点がある。若年女性は自然増減と社会増減の両面に関わり、その人口は持続可能性に大きく寄与するが、一方で人材の質という点も地方の持続可能性に関わりうるものと思われる。女性（に限らないが）起業家が地域の産業を活性化させ、その結果として若年女性人口が増えたり、あるいは非常に優秀な女性が活躍の機会を求めて海外に転出する、といった点も考慮に入れるべきものとする。

5 引用文献

- (1) 勢一智子「地域社会の持続可能性について」（最終アクセス 2021年8月31日）.
https://www.soumu.go.jp/main_content/000562248.pdf
- (2) 内閣府(2019)。「まち・ひと・しごと創生長期ビジョン（令和元年改訂版）」.
<https://www.chisou.go.jp/sousei/info/pdf/r1-12-20-vision.pdf>

- (3) 林玲子(2016).「女性の活躍と人口移動」『移動する若者／移動しない若者 ―― 実態と問題を掘り下げる』, 独立行政法人労働政策研究・研修機構.
<https://www.jil.go.jp/kokunai/blt/backnumber/2016/05/018-033.pdf>
- (4) 田村一軌, 坂本博, 戴二彪 (2018).「日本における女性の地域間移動パターンと影響要因」『公益財団法人アジア成長研究所 Working Paper』.
- (5) AGC株式会社「ビッグデータ時代のビジネス課題設定に向けた独自手法「因果連鎖分析」を確立」(最終アクセス 2021年8月31日).
https://www.agc.com/news/detail/1200060_2148.html
- (6) e-Stat 政府統計の総合窓口(最終アクセス 2021年8月31日).
<https://www.e-stat.go.jp/>
- (7) 統計ダッシュボード(最終アクセス 2021年8月31日).
<https://dashboard.e-stat.go.jp/>
- (8) RESAS 地域経済分析システム(最終アクセス 2021年8月31日).
<https://resas.go.jp/#/13/13101>
- (9) 総務省統計局「まちづくりのための地域経済分析」(最終アクセス 2021年8月31日).
<https://www.stat.go.jp/dstart/point/lecture/04.html>
- (10) 公正取引委員会 用語解説(最終アクセス 2021年8月31日).
<https://www.jftc.go.jp/soshiki/kyotsukoukai/ruiseki/yougo.html>
- (11) World Economic Forum (2021).“Global Gender Gap Report 2021”.
<https://www.weforum.org/reports/ab6795a1-960c-42b2-b3d5-587eccda6023>
- (12) 厚生労働科学研究 健康寿命のページ(最終アクセス 2021年8月31日).
<http://toukei.umin.jp/kenkoujyummyou/>
- (13) Breiman, L. (2001).“Random Forests”, *Machine Learning* 45, p.5-32.
- (14) Lundberg, S. M. and Lee, S. -I. (2017).“A unified approach to interpreting model predictions.”
Advances in Neural Information Processing Systems, p.4768-4777.
- (15) 畠山勝太, ジェンダーギャップ指数とかいう使えない指標と女子教育の話(最終アクセス 2021年9月1日). https://note.com/shota_hatakeyama/n/nbb2cc99c0eae
- (16) Athey, S., & G. W., Imbens. (2016).“Recursive partitioning for heterogeneous causal effects.”
Proceedings of the National Academy of Sciences, 113(27), 7353-7360.