

2020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]

階層ベイズモデルを用いた 学力に対する教育費の費用対効果推定

渡邊彰久・石川洸矢・近藤謙将
(東京工業大学大学院工学院経営工学系)

論文の概要

小中学生の学力について、状態空間モデルによる学力推定と階層ベイズモデルによる教育関連費用の効果推定を行い、学力に対する教育費の費用対効果が都道府県によって異なることを示している。

論文審査会コメント

実証に関わる問題意識や先行研究のレビューと用いている分析のバランスが良い論文となっている。

階層ベイズモデルを用いた 学力に対する教育費の費用対効果推定

渡邊彰久*¹・石川洸矢*¹・近藤謙将*¹

*¹：東京工業大学大学院 工学院 経営工学系

1. 研究テーマと目的

我が国の都道府県ごとの学力格差は改善されつつあるが、いまだに深刻な問題である。国立教育政策研究所が2007年度より実施している全国学力・学習状況調査^[1]の2017年度【小学生国語A】と【小学生算数A】の都道府県別平均正答率は図1のようになり、平均正答率が1番高い都道府県と1番低い都道府県の差を求めると、それぞれ、8%ポイント、9%ポイントとなる。この差を2007年度から2017年度まで求め、グラフの推移を図に表すと図2のようになる。図2を見ると、徐々に学力格差は改善されつつあるが、いまだに8%ポイントほどの差がある。我々はこの8%ポイントという数値を、決して小さくないものであると考えている。以上より、学力の都道府県格差は未だ深刻であることがわかる。

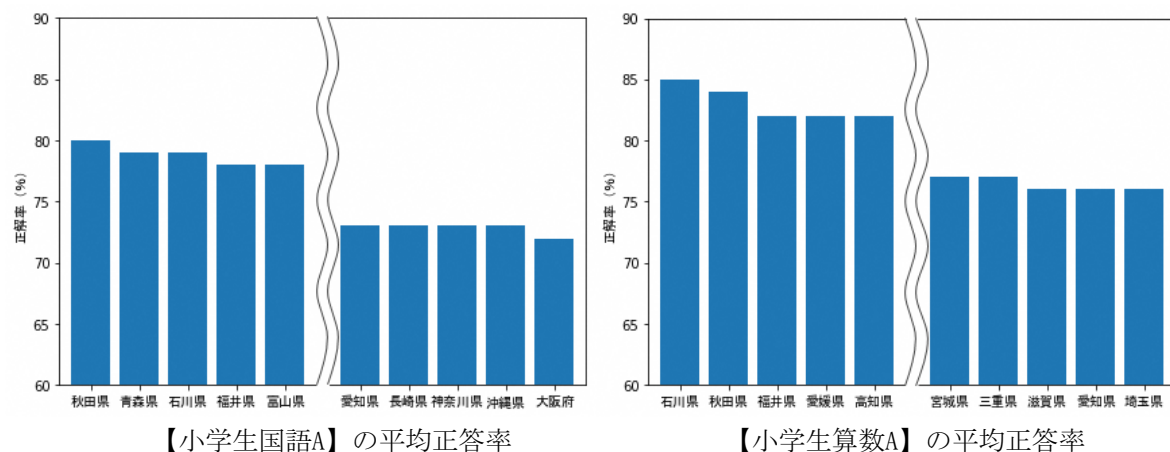


図1-1 2017年度全国学力・学習状況調査の平均正答率

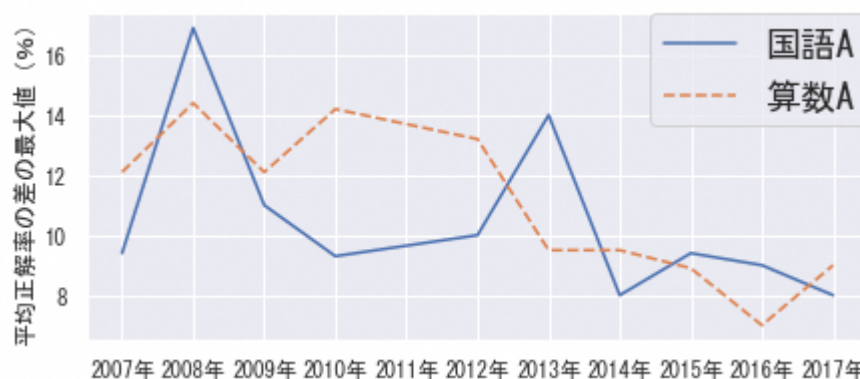


図1-2 平均正答率が1番高い都道府県から1番低い都道府県の差の推移 (2007年度～2017年度)

学力格差がもたらす影響として、例えば所得があげられる。階層・学歴・学力が所得にあたえる影響について松浦^[2]によると、銘柄大学の卒業生はその他の層に比べて年収が20.7%程度高いという事が示されている。このように学力格差が将来に与える影響は大きく、都道府県別の学力格差をさらに改善することは急務であると言える。

本稿では、分析①：各都道府県の本質的な学力の推定、分析②：各都道府県において教育関連費用が学力に対する費用対効果の調査を行った。また、調査結果をもとに、「小中学生の学力を向上させるためには、どの要因に注目すれば効率的か」という考察を行う。

2. 研究の方法と手順

本研究は、以下の2つの分析からなる。

2-1. 状態空間モデルを用いた各都道府県の真の学力の推定

テストの点数は本質的な学力だけでなく、問題の内容・テスト当日の天候など様々な要因に影響を受けると考えられる。また、各都道府県の本質的な学力が年単位で急激に変化することは考えにくく、前年と似た学力を持つのが通常であると考えられる。そのため、分析①ではテストの点数は真の学力に対して正規分布に従う誤差が乗ったものであると考え、状態空間モデルを用いた真の学力の推定を行った。

観測値として2007年度から2017年の計11年度で実施された、全国47都道府県の小・中学生を対象としたテストの正答率のデータを用いて算出した各年の都道府県の偏差値を用いた。

状態空間モデルでは、 y_t を $t(t = 2007, 2008, \dots, 2017)$ 年度における偏差値（観測値）、 α_t を $t(t = 2007, 2008, \dots, 2017)$ 年度における学力（状態）とし、以下のような状態空間モデルで真の学力（状態）の推定を行った。

$$y_t = \alpha_t + \epsilon_{y,t}, \epsilon_{y,t} \sim \text{Normal}(0, \sigma_y) \quad (1)$$

$$\alpha_t = \alpha_{t-1} + \epsilon_{\alpha,t}, \epsilon_{\alpha,t} \sim \text{Normal}(0, \sigma_\alpha) \quad (2)$$

方程式(1)、(2)はそれぞれ、観測方程式、状態方程式である。また、 $\epsilon_{y,t}$ は観測撓乱項、 $\epsilon_{\alpha,t}$ は状態撓乱項であり、各時点で独立な正規ホワイトノイズであるとする。この状態空間モデルによる推定を各都道府県に対して行い、都道府県・年度毎の真の学力を推定する。推定には、マルコフ連鎖モンテカルロ法（MCMC法）を用いた。

2.2. 階層ベイズモデルを用いた教育関連費用の効果推定

塾や家庭教師などへの平均支出が高い県は、学力が高くなると考えるのが自然であり、地方自治体が学校設備などに投資する事もまた、学力に好影響を与えることが想像できる。しかし、県民性、経済環境、教育政策などを考慮すれば、47都道府県で費用対効果が全く同様であると考えるのは、強すぎる仮定であるし、全く違うというのも強すぎる仮定であるように思える。そこで、階層ベイズモデルを導入し、費用対効果（回帰係数）は同じ事前分布に従うという緩やかな仮定を置くことで、先ほどの両極端な仮定の中庸を取るようなモデルを作成した。

目的変数として分析①で推定した2007年度から2017年度の各都道府県の真の学力を、説明変数として生徒・教師比率、教育費（二人以上の世帯）、教養娯楽費（二人以上の世帯）、義務教育費（自治体）を用いて分析を行った。

この分析では、 $Y_{p,t}$ を第 $p(p = 1, 2, \dots, 47)$ 都道府県の $t(t = 2007, 2008, \dots, 2017)$ 年度における真の学力（偏差値）を、 $X_{j,p,t}(j = 1, \dots, M)$ (M は説明変数の数)を説明変数として、以下の階層ベイズモデルによる回帰係数の事後分布の推定を行った。

$$Y_{p,t} = \beta_{p,0} + \sum_{j=1}^M \beta_{p,j} X_{j,p,t} + \epsilon_{p,t}, \epsilon_{p,t} \sim \text{Normal}(0, \sigma_{p,y})$$

$$\beta_{p,j} \sim \text{Normal}(\beta_j, \sigma_{\beta_j}), \quad (j = 0, 1, 2, \dots, M)$$

$$\beta_j \sim \text{Normal}(\beta, \sigma_\beta)$$

$$\sigma_{p,y} \sim \text{Uniform}(0, A_y)$$

$$\sigma_{\beta_j} \sim \text{Uniform}(0, A_\beta)$$

$\sigma_{p,y}, \sigma_{\beta_j}$ に対する事前分布には, $A_y, A_{\beta_j} = 100$ とした一様分布による無情報事前分布^[3]を, β_j に対する事前分布には, $\beta = 0, \sigma_{\beta} = 100$ とした正規分布による無情報事前分布を用いている. $\beta_{p,j}$ は都道府県の説明変数 j に対する回帰係数, $\epsilon_{p,t}$ は誤差項であり, p, t に関して互いに独立な正規分布に従っているものとする.

MCMC法によるサンプリングには, Python言語のライブラリであるPystan, Numpyroを用いた.

3. データ概要

ここでは本稿で用いたデータ項目とそれらを用いて作成した変数を示す. 本稿では各都道府県で義務教育を受けている小中学生の成績に注目しており, 教育関連費用が生徒の成績にどのような影響を及ぼすのかという点に着目した. 生徒の成績データとして2007年度より国立教育政策研究所が実施している全国学力・学習状況調査^[1]の成績データを利用している. 全国学力・学習状況調査は小学6年生と中学3年生を対象とした国語・算数(数学)のテストであり最新の令和元年度の調査結果^[4]によると小学校は全国19,600校中19,455校と約99%が参加しており, 中学校も95.6%が参加しており, 網羅性の高いデータである. 2011年度と2016年度の熊本県のデータは欠損値であったため, 線形補完を行った.

説明変数の計算に用いた各種指標については, 47都道府県の2007-2017年度のデータを, SSDSE-2020Bまたはe-Stat^[5]から入手した. 各指標の項目の説明, データの出出を以下の表3.1示す.

表3-1: 使用した教育関連指標一覧

項目名	項目の説明	出典
小学校児童数	学校教育法に規定する小学校の児童数	学校基本調査
中学校生徒数	学校教育法に規定する中学校の生徒数	
小学校教員数	学校教育法に規定する小学校の教員数	
中学校教員数	学校教育法に規定する中学校の教員数	
教育費 (二人以上の世帯)	授業料等, 教科書, 学習参考教材, 補習教育	家計調査
教養娯楽費 (二人以上の世帯)	教養娯楽用耐久財, 教養娯楽用品等	
小学校教育費 (自治体)	自治体が支出する小学生一人あたりの 教育費	地方教育費調査
中学校教育費 (自治体)	自治体が支出する中学生一人あたりの 教育費	

説明変数として, 表3-1の教育費(二人以上の世帯), 教養娯楽費(二人以上の世帯)と, 以下のよう
に求めた生徒・教員比率, 義務教育費(自治体)を用いた.

$$\text{生徒教員比率} = \frac{\text{小学校児童数} + \text{中学校生徒数}}{\text{小学校教員数} + \text{中学校教員数}}$$

$$\text{義務教育費} = \frac{\text{小学校児童数} \times \text{小学校教育費} + \text{中学校生徒数} \times \text{中学校教育費}}{\text{小学校児童数} + \text{中学校児童数}}$$

4. 分析結果と解釈

4.1. 分析①の結果と解釈: 状態空間モデルを用いた各都道府県の真の学力の推定

図4-1, 4-2 は, 各都道府県の偏差値の推移(オレンジ線)と状態空間モデルにより推定した真の学力(青線)をそれぞれプロットしたものである. また, 青い領域は真の学力が90%の確率で存在する区間である. 区間の図示の関係上, 図4-1, 図4-2の縦軸のスケールが異なることに注意されたい.



図4-1：偏差値（観測値）と推定された真の学力（状態）の推移（北海道, 東北, 関東, 中部）



図4-2：偏差値（観測値）と推定された真の学力（状態）の推移（近畿, 中国, 四国, 九州）



図4-3：偏差値（観測値）と推定された真の学力（状態）の推移（拡大）

図4-3は、図4-1, 4-2 から福岡県と埼玉県の例を拡大したものである。上述の通り、真の学力は1年単位で急激に変化することは少ないと考え、偏差値は真の学力にノイズが付加されて観測されているという仮定を置いている。図4-3の福岡県の結果から、偏差値（観測値）は比較的変動しているのに対して、推定された真の学力（状態）はほとんど変化しておらず、偏差値の変動はノイズによるものであると解釈できる。また、埼玉県の偏差値も福岡県と同様に各年で大きく変動しているが、真の学力の傾向は福岡県とは異なり下降トレンドである。このことから、埼玉県の偏差値の変動は、ノイズによるものだけでなく、真の学力の変動にも起因していると解釈できる。以上より、ノイズを除去した真の学力の変動を可視化する事が出来た。次節では、推定した真の学力を目的変数とした分析を行う。

4.2. 分析②の結果と解釈：階層ベイズモデルを用いた各種教育関連費用の効果推定

2章で示したモデルに対して、MCMC（マルコフ連鎖モンテカルロ）法を用いた事後分布からのサンプリングを行ったところ説明変数としていた生徒・教師比率と教育費（自治体）の相関係数が0.8と大きいことから事後分布の収束に関する指標であるRhatの値が不安定となった。そのため、生徒・教員比率を説明変数から除き、再度サンプリングを行った。

各回帰係数について、事後分布の両側5%区間に0が存在する場合は有意としている。各パラメータに対する結果は表4-1, 4-2にまとめている。表4-1は都道府県毎に各パラメータの平均点、5パーセンタイル点、95パーセンタイル点を、表4-2はパラメータの平均と標準偏差を示している。教育費（二人以上の世帯）は、全ての都道府県で有意な結果が得られなかった。原因として、成績の悪い子供ほど教育費用が高くなる場合が考えられ、教育費用の大きさが成績と線形の関係にない事が考えられる。教養娯楽費も同様である。それに対して、義務教育費に対しては複数の都道府県で有意な関係が見られたので、義務教育費の回帰係数が正で有意であった県の回帰係数の事後分布を図4-4に示している。図4-4は、横軸が教育費（自治体）に対する回帰係数であり、縦軸はカーネル密度推定による確率密度である。カーネル密度推定には、MCMC法により各回帰係数の事後分布から得られたサンプルを用いた。図に掲載されているのは、教育費（自治体）に対する回帰係数が有意であった都道府県である。この結果から、青森県、山形県、栃木県、福島県、高知県、佐賀県、沖縄県の7都道府県は、比較的教育費の費用対効果が高いと考えることができる。その中でも、特に沖縄県の回帰係数の平均は、他の倍以上であり沖縄県は教育費を効率的に学力に結び付けていることが確認できる。

5. 結果の考察と今後の展望

本稿の調査によって都道府県毎で教育費における学力に対する費用対効果が異なる事が明らかになった。まず、目的変数である学力の指標となる成績データから真の学力を推定する事によってその後の分析をより正確に行えるようにした。その後に真の学力を目的変数に、教育関連費用を説明変数とし、費用対効果が都道府県毎に異なっているという仮定の元でその調査をした。結果として、費用対効果は都道府県毎に異なっている事が確認できた。ここから費用対効果が高かった都道府県に対し、そこで行われている教育政策、教育費の割り当て、具体的な施策に注目する事で、学力に効果的な要因を特定する事が可能になると考えられる。本稿では具体的な教育政策の調査などは行わずに今後の課題とする。

本稿では全国学力・学習状況調査での成績データを元に考察をしてきたが、四谷大塚で開催されている全国統一テストと全国学力・学習状況調査での都道府県別の成績には大きな相関がないようである。^[6]一般的に学力を向上させる為には、全国学力・学習状況調査よりも進学塾などが主催している全国模試などが目的変数として適切である事が予想される。また、全国学力・学習状況調査は校則遵守率や読書時間などに強い相関を持っている。^[7]以上より学力のみを対象にするには、全国統一テストを目的変数にする事が望ましく、規律の高さなどにも焦点を置いている場合に全国学力・調査は有効である可能性が高い。ただし、全国学力・学習状況調査は他のテストに比べてかなり網羅的なデータである為、信頼性が高い事が予想される。また一方で、全国学力・学習状況調査では私立校の参加率は中学校共に約50%程度に収まっているため、私立校が多いと考えられる、東京都、神奈川県、大阪府

	$\beta_{p,0}$			$\beta_{p,1}$			$\beta_{p,2}$			$\beta_{p,3}$			$\sigma_{p,y}$		
	平均	5%	95%	平均	5%	95%	平均	5%	95%	平均	5%	95%	平均	5%	95%
北海道	50.00	34.40	66.37	-0.45	0.05	0.49	0.07	-1.67	1.78	-0.63	-7.08	5.52	50.09	6.05	95.45
青森県	35.03	33.89	36.10	-0.51	-0.03	0.34	-0.62	-1.70	0.46	3.53	2.41	4.72	0.95	0.61	1.47
岩手県	60.22	58.75	61.67	-0.31	0.08	0.46	-0.09	-1.25	1.05	-0.92	-2.35	0.43	0.98	0.62	1.56
宮城県	49.55	48.42	50.62	-0.41	0.04	0.46	-0.04	-1.25	1.21	-1.09	-1.79	-0.29	1.19	0.76	1.84
秋田県	46.80	45.48	48.10	-0.47	0.03	0.47	0.21	-0.95	1.33	-2.66	-6.24	1.01	2.61	1.74	3.87
山形県	74.13	69.50	78.80	-0.43	0.05	0.52	0.45	-1.01	2.10	4.43	0.92	7.91	3.45	2.29	5.09
福島県	55.39	52.07	58.42	-0.31	0.12	0.64	0.58	-0.86	2.21	-4.54	-7.37	-1.41	4.02	2.62	6.30
茨城県	46.59	45.06	47.91	-0.26	0.14	0.64	0.12	-1.05	1.28	-1.52	-2.87	-0.16	2.50	1.64	3.80
栃木県	52.47	51.85	53.07	-0.36	0.08	0.54	-0.82	-1.63	0.00	4.22	2.97	5.50	1.10	0.70	1.68
群馬県	44.36	43.10	45.87	-0.66	-0.05	0.37	0.59	-0.19	1.42	-3.84	-7.36	-0.25	1.66	1.02	2.39
埼玉県	51.09	49.79	52.34	-0.50	0.00	0.39	0.20	-0.60	0.98	-2.40	-4.45	-0.42	1.10	0.71	1.69
千葉県	45.59	36.15	53.86	-0.29	0.07	0.40	1.02	-0.16	2.12	1.02	-5.47	6.72	2.48	1.53	4.05
東京都	48.02	44.15	52.11	-0.40	0.06	0.53	-0.35	-1.43	0.73	-2.20	-6.86	2.74	2.10	1.34	3.30
神奈川県	60.61	57.34	63.42	-0.38	0.10	0.60	-0.16	-1.42	1.16	-4.58	-7.49	-1.40	1.74	1.12	2.72
新潟県	51.25	45.79	56.03	-0.36	0.05	0.45	-0.80	-1.96	0.35	1.39	-1.41	4.02	1.28	0.81	1.99
富山県	49.46	46.17	52.89	-0.60	-0.04	0.38	-0.23	-1.77	1.22	1.40	-1.39	4.10	3.13	2.05	4.85
石川県	65.30	63.94	66.56	-0.42	0.05	0.48	0.21	-1.03	1.48	0.18	-3.89	4.27	2.30	1.56	3.47
福井県	67.26	62.95	71.18	-0.40	0.09	0.60	-0.05	-1.77	1.58	0.74	-4.86	6.36	7.83	5.10	11.88
山梨県	74.70	73.57	75.75	-0.49	0.00	0.42	-0.80	-2.28	0.60	0.82	-1.41	3.18	1.50	0.91	2.46
長野県	46.57	45.15	47.95	-0.33	0.10	0.54	0.48	-0.78	1.81	-0.63	-3.08	1.89	1.62	1.03	2.50
岐阜県	48.91	48.10	49.68	-0.28	0.14	0.66	-0.51	-1.63	0.65	-2.39	-4.10	-0.57	1.39	0.87	2.18
静岡県	48.40	43.59	52.97	-0.21	0.15	0.61	1.12	-0.14	2.46	-6.20	-12.11	-0.59	1.84	1.11	2.98
愛知県	54.57	47.18	61.94	-0.34	0.09	0.55	-0.89	-2.90	0.68	-0.31	-5.34	4.79	3.68	2.25	5.85
三重県	56.28	50.30	62.01	-0.32	0.11	0.57	-0.49	-1.89	0.84	3.88	0.18	7.51	1.89	1.23	2.84
滋賀県	43.71	41.59	45.94	-0.25	0.13	0.61	0.32	-0.73	1.32	3.17	-1.22	8.38	1.74	1.13	2.66
京都府	40.45	36.54	44.49	-0.53	-0.01	0.39	0.60	-0.90	2.18	-1.42	-6.21	3.50	3.49	2.26	5.39
大阪府	54.31	53.19	55.44	-0.53	-0.06	0.29	0.18	-0.73	1.10	-1.60	-4.51	1.44	1.20	0.77	1.80
兵庫県	38.61	34.52	43.12	-0.44	0.02	0.41	-0.49	-1.99	0.95	0.75	-2.35	4.12	2.14	1.36	3.26
奈良県	50.40	48.78	52.05	-0.30	0.03	0.34	-0.64	-1.17	-0.07	-1.76	-3.89	0.40	0.64	0.41	0.98
和歌山県	50.92	46.70	54.89	-0.38	0.07	0.53	0.36	-0.81	1.52	-0.50	-5.97	4.64	3.55	2.38	5.47
鳥取県	42.31	38.68	46.18	-0.53	-0.02	0.38	0.03	-1.51	1.80	-0.02	-4.72	4.48	2.68	1.73	4.16
島根県	56.83	54.73	58.81	-0.42	0.04	0.46	0.49	-0.97	2.03	-3.28	-5.10	-1.47	2.04	1.37	3.05
岡山県	52.43	46.27	58.21	-0.36	0.09	0.54	0.28	-1.14	1.77	-2.89	-5.49	-0.22	2.36	1.51	3.68
広島県	42.89	41.39	44.47	-0.36	0.09	0.58	0.64	-0.58	1.93	-0.02	-2.46	2.48	2.27	1.49	3.39
山口県	57.78	55.59	59.78	-0.67	-0.10	0.30	-1.28	-2.67	0.21	1.43	-3.45	5.98	1.62	0.86	2.75
徳島県	51.46	49.72	53.13	-0.46	0.04	0.48	-0.44	-2.00	1.02	5.13	1.09	8.72	3.08	1.97	4.85
香川県	49.66	46.66	52.70	-0.62	-0.04	0.38	0.41	-1.13	1.93	-0.99	-3.43	1.40	2.51	1.64	3.86
愛媛県	56.67	55.36	58.04	-0.30	0.12	0.62	1.03	-0.06	2.14	-1.57	-3.63	0.54	2.13	1.38	3.21
高知県	54.06	51.77	56.47	-0.44	0.06	0.52	-0.09	-1.65	1.47	5.00	0.58	9.03	4.11	2.57	6.60
福岡県	32.83	24.21	42.77	-0.49	0.02	0.45	0.28	-1.01	1.62	2.30	-1.23	5.33	4.91	2.89	7.94
佐賀県	45.41	44.37	46.48	-0.24	-0.02	0.20	0.12	-0.22	0.40	1.58	0.54	2.63	0.38	0.22	0.63
長崎県	43.07	42.77	43.37	-0.07	0.21	0.52	0.12	-0.32	0.58	-1.87	-2.32	-1.46	0.36	0.22	0.55
熊本県	47.82	45.24	49.82	-0.50	0.03	0.46	0.66	-0.79	1.99	-1.97	-4.72	0.77	2.23	1.33	3.92
大分県	50.96	49.40	52.45	-0.47	0.01	0.41	0.62	-0.69	2.02	-1.43	-4.14	1.16	1.40	0.85	2.29
宮崎県	44.12	41.70	46.79	-0.42	0.03	0.43	-0.02	-1.39	1.30	3.84	-1.02	8.39	3.01	1.92	4.56
鹿児島県	46.13	44.00	48.35	-0.36	0.09	0.55	1.15	-0.22	2.43	-5.90	-8.62	-2.63	1.73	1.07	2.74
沖縄県	37.33	33.34	41.39	-0.30	0.13	0.63	1.37	-0.22	3.40	9.27	4.33	13.84	8.90	6.35	12.70

表4-1: 都道府県毎の各パラメータ

	平均	5%	95%
β_0	50.56	48.03	53.01
β_1	0.05	-0.15	0.26
β_2	0.10	-0.30	0.48
β_3	-0.14	-1.25	0.92
σ_{β_0}	9.44	7.77	11.44
σ_{β_1}	0.22	0.02	0.50
σ_{β_2}	0.97	0.51	1.45
σ_{β_3}	3.86	2.90	4.96

表4-2: 各パラメータの平均と標準偏差

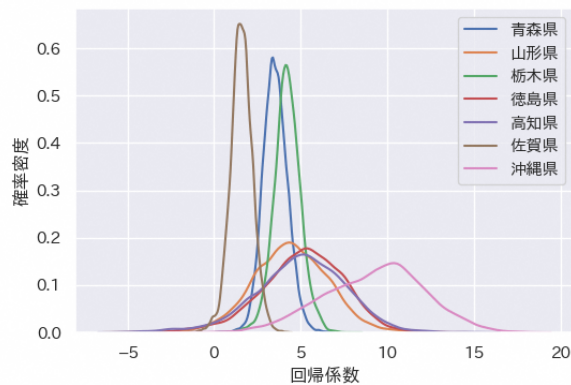


図4-4：義務教育費の回帰係数の事後分布（正に有意であった県）

などの都市部は多少の誤差がある事が予想される。^[4]

今後の展望として、教育関連費の費用対効果が都道府県毎に異なる事に注目し、各都道府県が行っている具体的な教育政策などに注目してより効果的な教育政策を特定する事が考えられるが、それには2つの問題点が存在する。まず1つに教育費の内訳や具体的な施策などのより細かいデータが必要になる点であり、これらのデータを集めるのは難しい事が予想される。次に教育政策などは数字で表せるものではなく、扱い方に工夫が必要である事である。この2点でこれより先の検証は一層難しくなる事が予想されるがまだ分析をアップデートする余地は残っていると考える。

参考文献

- [1] 国立教育政策研究所：教育課程研究センター「全国学力・学習状況調査」（平成19年度～平成29年度），（<https://www.nier.go.jp/kaiatsu/zenkokugakuryoku.html>）（2020年8月10日閲覧）。
- [2] 松浦司 階層・学歴・学力が所得にあたえる影響について，経済論叢，京都大学経済学会，Vol. 178, No. 3, pp. 302～321. (2006)
- [3] Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3), 515-534. (2006).
- [4] 国立教育政策研究所 全国学力・学習状況調査の結果(令和元年度)
（<https://www.nier.go.jp/19chousakekkahoukoku/19summary.pdf>）
- [5] e-Stat政府統計の総合窓口：地方教育費調査_都道府県別集計_教育分野別総教育費_第14表 教育分野別在学者・国民一人当たり及び一学級当たり経費（平成19年度～平成29年度）（2020年8月30日閲覧）。
- [6] 植松 康祐, 高橋 泰代 全国学力調査結果の統計的分析 国際研究論叢：大阪国際大学紀要, 30巻, 3号, 1-12, (2017).
- [7] 都道府県別統計とランキングで見る県民性（2020年8月30日閲覧）