

成績に関連する要素のデータ分析

慶應義塾大学 経済学部 長倉大輔研究会

河野裕輝・大丸和輝・田村啓太

要旨

本研究は OECD（経済協力開発機構）が 3 年に 1 度実施している世界の 15 歳を対象とした学習到達で調査データである PISA (Programme for International Student Assessment) データを主軸に国ごとのいくつかの説明変数を用いて、それらが成績にどのような影響を及ぼしているのかを線形混合効果モデルを用いて計量経済学的に研究を行った。PISA の数学スコアを被説明変数として設定して、各説明変数とどのように関係しているのかに注目し、その結果に対して考察を行った。

1. はじめに

近年、グローバル化と技術革新の進展に伴い、各国の教育システムは急速な変化を求められている。特に OECD が実施する国際学力到達度調査 (Programme for International Student Assessment: PISA) は、教育政策や学力格差の実態を国際的に比較するための主要なデータセットとして注目されている。PISA の結果は、学習者の知識や技能のみならず、家庭環境、学校環境、社会経済的背景など多層的な要因によって大きく左右されることが明らかにされてきた。

そこで本研究では、“Predicting and understanding student achievement using PISA 2018 data: A machine learning approach” (F. G. Özoğlu & S. H. Özoğlu, 2023) から PISA データを用いた、予測精度の高い機械学習モデルを選択した。この研究では 2018 年の PISA データを用いて、学生の学業成績を予測するための機械学習モデルを適用している。そのモデルに国ごとの日照時間、一人当たりの名目 GDP、ジニ係数といった独自の変数を追加し、できたモデルを線形混合効果モデルを用いて個人レベル、国・学校レベルの要因が成績に与える影響を定量的に検証することを本研究の目的とする。

2. 使用データ

本研究では、OECD が実施する国際学力調査 PISA の生徒個票データおよび各国・地域レベルのマクロ指標を用いて分析を行った。PISA は 15 歳を対象に、数学・読解・科学の 3 分野に関する学力および学習環境・社会経済的背景を測定する国際比較調査であり、その豊富な質問票変数を通じて個人・家庭・学校・社会の多層的要因を分析することが可能である。本研究では、学力（数学スコア）を被説明変数とし、個票レベルの 5 変数と地域レベルの 3 変数を説明変数として採用した。

生徒個人の教育達成に関連する要因として、経済・社会・文化的背景 (ESCS)、教師支援 (TEACHSUP)、家庭 ICT 環境 (ICTHOME)、ICT 関連資源 (ICTRES)、および学校内リスク認知 (SCHRISK) の 5 変数を用いた。これらは PISA における教育達成研究の中核的変数であり、社会経済的要因・心理的要因・教育環境の側面を包括的に反映する。

まず、ESCS (Index of Economic, Social and Cultural Status) は、親の職業地位・教育水準・家庭所有物を統合した家庭社会経済地位指数である (OECD, 2014)。

ESCS は学力の国際比較において最も強力な決定要因とされ、OECD（2013）は「ESCS が 1 標準偏差上昇するごとに数学得点が平均約 39 点上昇する」と報告している。そのため、教育の公平性を分析する上で ESCS は不可欠な変数である。

TEACHSUP（Teacher Support）は、生徒が教師からどの程度支援を受けていると感じているかを測る心理的指標である。教師支援は学習意欲や安心感を高め、学習成果を向上させる効果が報告されている（OECD, 2025）。したがって、TEACHSUP は家庭背景では説明しきれない心理的・関係的側面を補足する変数として位置づけられる。

SCHRISK（School Risk Perception）は、生徒が学校内で暴力やいじめ、安全上の不安をどの程度感じているかを測定する尺度である。OECD（2023）は、学校内の安全性や心理的安定感が学習集中や出席率に直接的な影響を与えると報告しており、リスク認知が高い学校ほど平均学力が低い傾向を示している。よって、SCHRISK は安全で安心な学習環境が学力に与える影響を把握するための重要な変数である。

個票レベルの変数に加え、各国・地域の構造的条件を反映するため、一人当たり GDP（GDP）、日照時間（SUNSHINE）、およびジニ係数（Gini）を地域レベルの説明変数として採用した。これらは経済水準・環境条件・分配構造といった異なる社会的な要素であり、個票変数のみでは説明しきれない地域差を補うものである。

まず、GDP は経済発展水準を示す代表的な指標であり、教育投資・施設整備・教員資源の充実度を反映する。OECD（2019）は PISA2018 の結果において「経済発展の水準が各国の学習到達度の約 28%を説明する」と報告しており、国・地域のマクロ経済条件が学力水準の基盤を形成していることを示している（OECD, 2019）。したがって、本研究では GDP を教育成果の構造的要因として導入した。

次に、SUNSHINE（日照時間）は環境・気候条件を表す外生的変数である。教室の昼光量に関する Heschong Mahone Group（1999）の研究では、最も採光が多い教室の生徒は数学で 20%、読解で 26%学習進度が速いと報告されている。この知見は、光環境や気候条件が学習・認知に影響を及ぼすことを示しており、本研究では広域の環境条件の代理指標として日照時間を採用する。

最後に、Gini（ジニ係数）は社会の所得格差を表す分配構造の指標である。Mazurek（2021）は PISA2018 の 75 か国を分析し、ジニ係数は統計的に有意でないと報告する一方、Sempé（2021）は校内レベルの所得不平等が数学スコアと有意に負の関連をもつことを示した。これらの知見は、社会全体の不平等が教育成果の分布に影響を及ぼす

可能性を示唆している。本研究では、ジニ係数を地域レベルの不平等構造の代理変数として採用する。

3. モデル設定

本研究では、OECD PISA データを基に、生徒の数学的リテラシー（PV1MATH）に影響を与える要因について検証を行う。PISA 2022 のデータは、「生徒」が「学校」に属し、その「学校」が「国」に属するという典型的な階層構造を有している。このデータ構造を統計的に正しく扱うため、本分析では線形混合効果モデルを採用した。分析には「lmer」関数（R の lme4 パッケージ）を用い、国・学校間の階層構造を考慮したモデルを推定した。

モデルの定式化は以下の通りである。

$$PV1MATH_{ij} = \beta_0 + \beta_1 ESCS_{ij} + \beta_2 TEACHSUP_{ij} + \beta_3 ICTHOME_{ij} + \beta_4 ICTRES_{ij} + \beta_5 SCHRISK_j + \beta_6 GDP_j + \beta_7 SUNSHINE_j + \beta_8 GINI_j + u_{0j} + u_{1j} ESCS_{ij} + \epsilon_{ij}$$

ここで、被説明変数である PV1MATH_{ij} は生徒 i の数学得点、説明変数である ESCS は社会経済的地位指数、TEACHSUP は教師による学習支援、ICTHOME は家庭での ICT 利用環境、ICTRES は学校の ICT 資源整備度、SCHRISK は学校のリスク水準、GDP は一人当たり GDP、SUNSHINE は日照時間、GINI は所得格差を示すジニ係数を表す。

分析対象は、国レベル変数（一人あたり GDP、ジニ係数、日照時間）の欠損値を除外した 67 カ国・地域、合計 535,432 人の生徒データである。生徒レベルの欠損値は多重代入法（MICE）によって補完し、モデルの計算を安定させるため、投入したすべての説明変数は標準化（平均 0、標準偏差 1）した。

国（CNTSCHID）をランダム効果として設定し、各変数の「傾き」そのものも国レベルで変動することを許容した。

4. 分析結果

4-1. 分析結果：固定効果

推定結果の REML 基準値は 1041025 であり、モデルの収束が確認された。ランダム効果の分散成分を見ると、学校間の切片分散が 2202.6、ESCS 傾きの分散が 132.1 であり、学校ごとに学力水準や社会経済的地位の効果が異なることが示された。残差分散は 4309.9 であった。

固定効果の推定結果を表に示す。

変数	推定値	標準誤差	t 値	効果の方向
(定数項)	436.34	0.99	438.85	－
ESCS (社会経済的地位)	15.17	0.39	38.92	正
TEACHSUP (教師支援)	2.58	0.24	10.95	正
ICTHOME (家庭 ICT)	0.18	0.21	0.89	不明確
ICTRES (学校 ICT 資源)	6.01	0.40	15.04	正
SCHRISK (学校リスク)	-4.62	0.24	-19.64	負
一人当たり GDP	17.50	1.42	12.34	正
日照時間	-15.64	1.27	-12.29	負
ジニ係数	-4.84	0.76	-6.33	負

表 固定効果の推定結果

分析結果から、社会経済的地位 (ESCS)、教師支援 (TEACHSUP)、および学校 ICT 資源 (ICTRES) は数学得点に対して統計的に有意な正の影響を持つことが明らかとなった。一方で、学校リスク水準 (SCHRISK)、所得格差 (Gini 係数)、および日照時間は負の影響を示し、学習環境や社会構造的要因が学力格差に寄与していることが示唆される。また、一人当たり GDP が高い国ほど得点が高い傾向にあり、マクロ経済的豊かさが教育成果に影響することが確認された。

本分析は、学校間差を考慮した階層的アプローチにより、個人属性と社会経済要因の両方が学力に及ぼす影響を定量的に明らかにした。特に、教育支援や ICT 環境の整備が学力向上に寄与する一方、社会的格差の存在が依然として学習成果に影響を与えていることが確認された。

4-2. 分析結果：先進国と途上国の比較

次に本研究では、国ごとの傾きの差異が、単なる地理的な「地域」（アジア、ヨーロッパ等）によるものか、あるいは「国の経済発展の段階（先進国か否か）」という要因に強く規定されているのではないか、という仮説を立て検証を行った。

本分析における「先進国」と「途上国」の分類は、既存の国際機関の定義を用いるのではなく、分析対象となった 67 カ国・地域のデータに基づいた相対的な分類を用いた。

具体的には、国レベル変数である「一人あたり GDP」の中央値（Median）を算出し、その中央値を上回る国々を「先進国グループ」、下回る国々を「途上国グループ」と定義した。この 2 つのグループと、地理的な「地域」のクロス集計によって、各変数の傾きの平均値を比較検証した。

分析の結果、「家庭背景指標」、「教師のサポート指標」、「家庭の ICT リソース指標」の 3 変数は、国の経済状況によって、その回帰係数（傾き）の平均値が明確に二分されることが判明した。

1. 「家庭背景指標」の効果

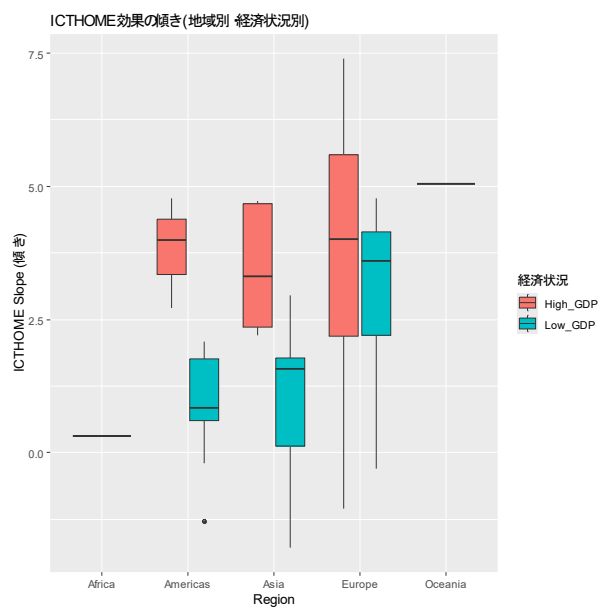
- 先進国グループ：平均傾き 41.3
- 途上国グループ：平均傾き 26.8

2. 「教師のサポート指標」の効果

- 先進国グループ：平均傾き 3.91
- 途上国グループ：平均傾き 0.906

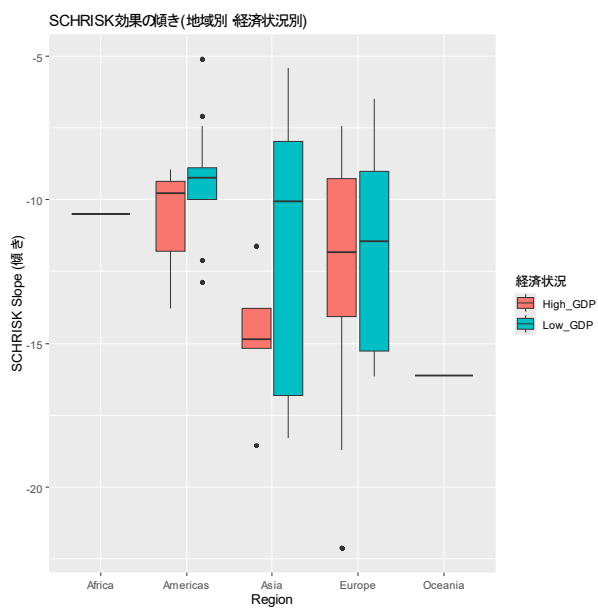
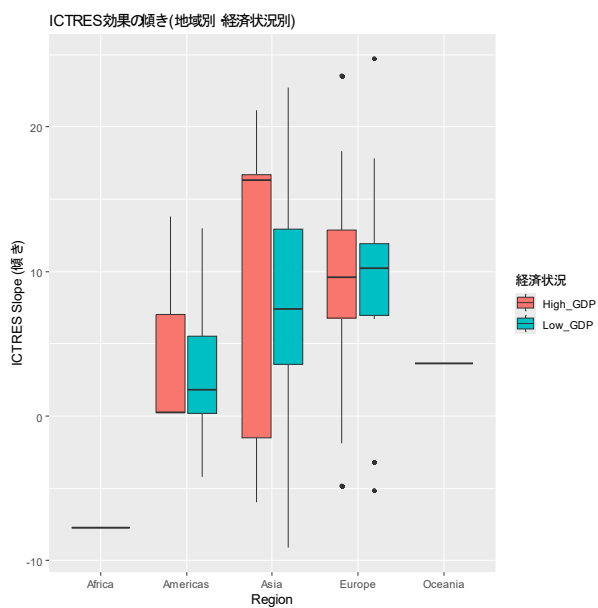
3. 「家庭の ICT リソース指標」の効果

- 先進国グループ：平均傾き 3.94
- 途上国グループ：平均傾き 1.56



一方で「学校の ICT リソース指標」は経済状況（GDP）では明確に二分されず、上記の 3 変数とは異なる傾向を示した。

ヨーロッパ・アジアは経済状況に関わらず、一貫して学力に強い相関がある。一方でアメリカ大陸とアフリカでは、この指標の効果が非常に弱い。



本分析は、学校間差を考慮した階層的アプローチにより、個人属性と社会経済要因の両方が学力に及ぼす影響を定量的に明らかにした。特に、教育支援や ICT 環境の整備が学力向上に寄与する一方、社会的格差の存在が依然として学習成果に影響を与えていることが確認された。

5. 考察

5-1: 考察：固定効果

本研究の分析結果から、学力格差の形成には個人レベルの社会経済的要因（ESCS）と学校・社会レベルの構造的要因の双方が強く関与していることが示された。特に、ESCS の係数が有意に正であったことは、家庭の経済的・文化的背景が学習成果に与える影響の大きさを示唆しており、教育機会の不平等が依然として主要な課題であることを裏付けるものである。

また、教師支援（TEACHSUP）および学校の ICT 資源（ICTRES）が有意な正の影響を示した点は注目に値する。これらは、学習支援の充実や教育環境のデジタル化が生徒の学力向上に寄与する可能性を示しており、教育政策における人的・物的資源の再配分の重要性を示唆する結果といえる。一方で、家庭での ICT 利用環境（ICTHOME）は有意な影響を持たなかったことから、ICT の「利用の質」が成果に結びついていない可能性が考えられる。

マクロ経済的な要因では、一人当たり GDP が学力に正の影響を及ぼす一方、ジニ系数（所得格差）は負の影響を示した。このことは、経済的豊かさが教育投資や学習環境を通じて成果に寄与する一方、社会的格差が学力格差を拡大させる二重構造を示している。また、日照時間の負の影響は、地域的气候条件が生活リズムや教育機会に間接的な影響を及ぼしている可能性を示唆しており、教育成果の地域的多様性を考慮した分析の必要性を示している。

総じて本研究は、教育成果を説明する際に「学校内の努力」だけでなく「社会経済的な部分」を同時に考慮する重要性を明確にした点で意義があるといえる。教育政策の立案においては、家庭背景に左右されない教育支援策の充実、および ICT を活用した教育の質的向上が今後の課題となる。

5-2. 考察：先進国と途上国の比較

分析結果から、国の経済発展の段階が、「リソース（家庭背景、教師、家庭 ICT）を学力に転換できるかどうかの前提条件」として機能していることを示唆している。

途上国グループでは、「教師のサポート指標」や「家庭の ICT リソース指標」が（指標上で）観測されても、それが学力向上に結びつく効果は先進国グループに比べて著しく低い（「教師のサポート指標」の平均傾きはほぼ 0 に近い）。これは、いわゆる情報格差が単なる「機器の保有」の問題ではなく、「学力向上に結びつく活用の格差」として現れていることを示している。

また、「学校の ICT リソース指標」の効果の差異は、「先進国か否か」という経済要因よりも、「地域」固有の要因によって強く規定されている。これは、ICT の導入方法、教員のスキル、カリキュラムとの連携など、各地域の「教育政策や活用のノウハウ」の違いが、経済状況以上に効果の差を生み出している可能性を示唆している。

以上から、PISA 2022 のデータを用いて、学力影響要因の効果（回帰係数）が国によって一様ではないことが実証された。

その変動性を説明する要因は一律ではなく、

1. 「家庭背景指標」や「教師のサポート指標」のように、国の「経済発展の段階」が効果の前提条件となる変数
2. 「学校の ICT リソース指標」のように、経済状況よりも「地域（政策）」が効果の差異を規定する変数

という、2 つの異なるパターンが存在することが明らかになった。

参考文献

- F. G. Özoğlu & S. H. Özoğlu, (2023). "Predicting and understanding student achievement using PISA 2018 data: A machine learning approach"
- Heschong Mahone Group. (1999). *Daylighting in Schools: An Investigation into the Relationship Between Daylighting and Human Performance*. California: Pacific Gas and Electric.
- Mazurek, J. (2021). *Inequality and students' PISA 2018 performance: A cross-country study*. Comparative Economic Research, 24(4), 7–28.
- OECD. (2013). *PISA 2012 Results: Excellence through Equity (Volume II)*. Paris: OECD Publishing.
- OECD. (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.
- OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: OECD Publishing.
- OECD. (2023). *PISA 2022 Results (Volume III): Supporting Students' Well-being*. Paris: OECD Publishing.
- OECD. (2025). *Teacher Support for Student Learning: Insights from PISA*. Paris: OECD Publishing.
- Sempé, L. (2021). *School-level inequality measurement based on categorical data: Evidence from PISA 2015*. Large-scale Assessments in Education, 9(1), 14