



連載



情報の授業をしよう！

本コーナー「情報の授業をしよう！」は、小学校や中学校で情報活用能力を育む内容を授業で教えている先生や、高校で情報科を教えている先生が、「自分はこの内容はこういう風に教えている」というノウハウを紹介するものです。情報のさまざまな内容について、他人にどうやって分かって

もらうか、という工夫やアイデアは、読者の皆様にもきっと役立つことと思います。そして「自分も教え方の工夫を紹介したい」と思われた場合は、こちらにご連絡ください。

(E-mail : editj@ipsj.or.jp)

情報教育の視点による統計教育 —「データ駆動型探究」の普及を目指して—



伊藤大貴 | 大分県立大分舞鶴高等学校

データサイエンス教育の充実を目指して

統計教育の必要性

高大接続の観点から見る統計教育

Society 5.0に向けたさまざまな取り組みが進む中、学校教育では、データをもとに行動や意思決定を行う「データ駆動型」の学習活動を中心的に担う教科として、高校「情報I」および「情報II」の統計・データサイエンスに関する学習の重要性が示されている。そのような中、2020年度より大学の初年次教育においても、「数理・データサイエンス・AI」に関する講義内容が位置づけられた。これは、大学生の数理・データサイエンス・AIへの関心を高め、適切に理解し活用する基礎的な能力や、課題を解決するための実践的な能力を育成するため、数理・データサイエンス・AIに関する知識および技術について体系的な教育を行うことを目的に設定されている。これらのことから、高大接続を見据えた統計に関する学習は重要性を増しており、さらなる知見の集積が求められる。さらに、探究活動や課題研究における「データの収集・分析・活用」の観点から照らし合わせても、高等学校普

通教科「情報」の位置づけは重要であると考えられる。実際に、学習指導要領解説情報編を見ても、データの活用に関する内容が充実しており、これからの情報教育には「統計やデータサイエンス」に関するリテラシーへの期待が含まれていることが分かる。

SSH 指定校としての役割

イノベーション人材を育成するための統計教育

本校大分県立大分舞鶴高等学校（以下舞鶴高校）は、2005年より「スーパーサイエンスハイスクール（SSH）」指定校として、さまざまな取り組みを行ってきた。SSH第IV期では、「高次の探究力を兼ね備えた Society 5.0の社会をけん引するイノベーション創出人材の育成」を研究開発課題として掲げ、STEAM教育、統計、プログラミングに特化した学びや、理数科による高度な探究活動を行っている。その中でも、筆者は情報科の教員として、1年次の「データサイエンス」や2年次の「SSH国際情報」等の授業科目を担当し、データ駆動型探究に活かすための統計教育に注力している。データ駆動型探究とは、データをもとに、新たな知見の追究や、課題解決の手法

を探索することで意思決定を行う探究活動であると捉えている。舞鶴高校では、情報科目における学びの中で、データを取り扱うスキルや考え方を育成し、探究活動に活かすことや、多くの学校にそのノウハウを普及させることをねらいとしている。たとえば、1年次の「データサイエンス」ではスキルや考え方を学び、2年次の「SSH国際情報」では、オープンデータを用いた探究を行うことで、学んだ力を活かす場面を設定している。このカリキュラムにより、2年次に行われる「課題研究」と並行してデータを活用した探究を行うため、データ駆動型探究を促進することができると考えている(表-1)。

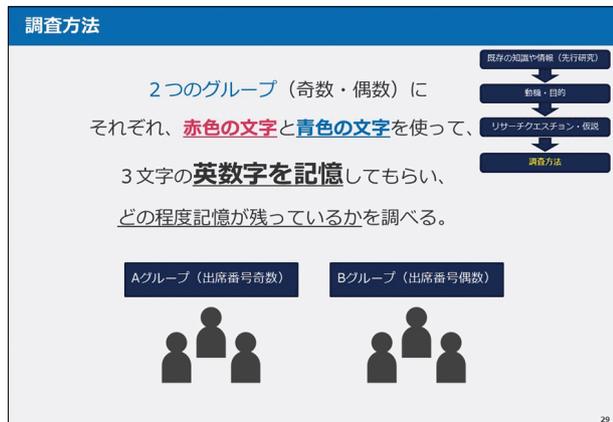
また、身近な題材や自分たちを被験者にした検証データを取り扱うことで、「分析してみたい!」と思えるような教材を開発することを意識している。たとえば、t検定を学ぶ授業では、「赤色と青色どちらが記憶に残りやすいか」を検証するようにしており、ワクワクするような題材を取り入れている(図-1)。

さらに、科学としての活動も意識し、「変数」の概念だけでなく、「リサーチクエスト」や「仮説

■表-1 データ駆動型探究を意識したカリキュラム

	単元	内容
1年 データサイエンス	ICT / Excel 演習	情報社会やICTの活用, Excelの活用方法について学ぶ
	Data Science 演習	身近なデータを用いて、統計的な手法を用いた探究活動を行う
2年 SSH国際情報	課題研究のための統計学	多変量データを題材にした統計手法について学ぶ
	Global Data Science 探究(普通科のみ)	オープンデータ(Kaggle)を統計的に処理し、問題解決の糸口を探る

※統計に関連する単元のみ抜粋



■図-1 t検定(実践)の授業スライド

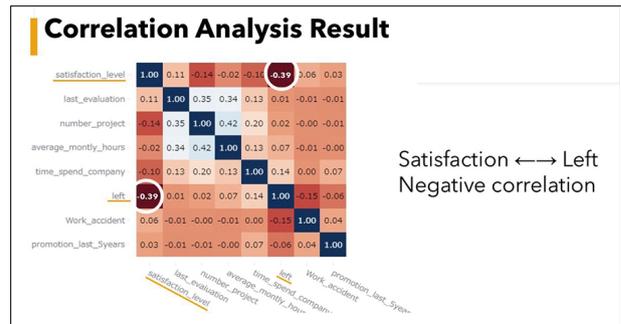
などの取り扱いにも留意している。

オープンデータの活用

2年次の「SSH国際情報」では、表-1のとおり「Global Data Science 探究」という単元を設定している。この単元では、オープンデータを活用して分析する活動を行う。班を編成し、Kaggleというオンラインプラットフォームを利用している。Kaggleには、無料で利用できる豊富なデータセットが提供されており、生徒の興味関心に応じたデータを取得することが可能である。生徒自身では、収集することが難しいようなデータが豊富にあり、実世界のデータをもとに学びや探究を行う機会を得ることができる。実際に、生徒はKaggleから得られたデータを分析しながら、問題解決の糸口を探り、その成果を発表する場として、台湾の生徒に英語で発表を行う場面を設けている。図-2は生徒が実際に取得したデータセットであり、図-3は取得したデータセットをもとに後述する「easyStat」を用いて相関分析を行った結果を、スライドに記載したものである(図-2, 図-3)。なお、「Global Data Science 探究」は普通科のみ実施している単元であり、理数科は、理数科課題研究にてデータ駆動型探究を行っている。

satisfaction_level	last_evaluation	average_monthly_hours	time_spent_company	Work_accident	left	promotion_last_5years	salary
0.38	0.53	157	3	0	1	0	low
0.8	0.86	262	6	0	1	0	medium
0.11	0.88	272	4	0	1	0	medium
0.72	0.87	223	5	0	1	0	low
0.37	0.52	159	3	0	1	0	low

■図-2 生徒が分析に使用したデータ(Kaggle)



■図-3 生徒による相関分析の結果

このように、オープンデータを活用することで、データに対する興味関心を高めることができるだけでなく、サンプル数等の信頼性について考える機会を提供することができる。また、「生徒がデータを選ぶ」ことにより、受け身になることなく、主体的に分析や探究活動に取り組む様子が見受けられた。

統計指導における問題点

さまざまな制約

実際に統計を指導する際に、さまざまな制約が浮上する。情報科を担当する教員であれば恐らく納得していただけたと思うが、主に以下の2点である。

- ①操作スキル
- ②PCのスペック

①については、さまざまな要因が考えられる。たとえば、1人1台端末が導入されたことによる「パソコン離れ」である。GIGAスクール構想による環境整備に伴い、従来のコンピュータ教室の役割が再評価され、一部の自治体ではコンピュータ教室廃止の動きが出てきている。そのような中、文部科学省は2022年12月19日に「GIGAスクール構想に基づく1人1台端末環境下でのコンピュータ教室の在り方について」という通知を发出し、コンピュータ教室の廃止に慎重な姿勢を示しているが、筆者の周辺の公立小中学校では、コンピュータ教室を廃止している学校は少ない。また、スマートフォンでできることも多くなり、パソコンを持たない家庭も増えているのではないだろうか。このような環境的な要因により、子どもがパソコンに触れる機会が減少していると考えられる。そのため、高校に入学し、「初めてパソコンに触ります」といった生徒も増えてきていることから、操作スキルに大きなばらつきが生じているように感じている。②については、授業で使用するPCのスペックによって、表計算ソフト等を使用した演習がスムーズにできないことが挙げられる。実際、筆者がいままで勤務した2つの学校では、起動に5分かかる、ブラウザや表計算ソフトの立ち上げに1分かかるなどといったハードウェアのスペックがボトルネックになることも多々起こっ

ている。さらに、一番困るのは、学年の全生徒が一斉に探究活動を行う際に、「コンピュータ教室に入りきれない」といった問題も挙げられる。そこで、環境的制約を極限まで低減することができるようなWebアプリケーションの開発を行った。

Webアプリケーションの開発

ブラウザ統計「easyStat」

統計指導における問題点を払拭し、データ駆動型探究を促進するために、環境的な要因に影響を受けない統計分析Webアプリケーションの開発を行った。本アプリケーションは、1年次のデータサイエンスの「Data Science 演習」にて、統計的仮説検定の後に取り扱うこととした。実際の活用場面では、授業用学校生活アンケート（睡眠時間や部活動の有無等の多変量のデータセット）を使用し、分析を繰り返すことで、それまでの授業で学んだ統計的な知識・技能の定着を図った。2年次の「課題研究のための統計学」や「Global Data Science 探究」においても使用することで、並行して行われる探究活動に活かせるよう工夫を行った。アプリケーション開発における主な条件は以下のとおりである。

- ①1人1台端末で軽快に動作することができる
- ②データの加工に関する操作を最小限に抑える
- ③タッチ操作のみで操作することができる
- ④解釈の補助機能を実装する

①に関しては、文部科学省「GIGAスクール構想の実現標準仕様書（2020）」を参考に、検討を行った結果、インストールが不要なブラウザベースでのWebアプリケーションが望ましいと判断した。そのため、Webアプリケーションとして公開することで、コンピュータ教室でなくとも1人1台端末を活用したデータ分析が可能になり、物理的な環境制限を減らすことができると考えられる。②および③に関しては、統計処理の阻害要因の1つである統計ソフトの操作性の複雑さを低減するために、高校生の持つICT操作スキルを考慮し、タッチ操作のみで動作可能なシンプルなUIを実装した(図-4)。具体的には、分析するデータのExcelファイルをアップロードするだけで分析

が可能になるものとした。実際に、情報授業で統計を教える際に、Excelの操作を指導することになるが、生徒の操作レベルのばらつきは非常に大きく、複雑な処理になればなるほど、技能を指導する教員の負担が大きくなってしまう。しかし、本アプリケーションを使えば、ファイルをアップロードし、クリックやタップを行うだけで、簡単に可視化や検定を行うことができるため、技能指導や習得の時間を大幅に減らすことができ、人間にしかできない「考察」や「意思決定」に時間を割くことができるようになる。

④については、統計初学者のつまずきを考慮し、有意水準の判定から解釈の補助を行う機能を実装した(図-5)。これにより、検定処理の知識がない生徒や指導に不安を持つ教員でも、容易に分析や指導を行うことができると考えられる。さらに、「情報I」「情報II」で取り扱う統計的な分析および探究活動で想

定し得る手法である、箱ひげ図、散布図(行列含む)、相関分析、t検定、分散分析、回帰分析等を実装した。ブラウザ機械学習「easyAutoML」

「数理・データサイエンス・AI」の「AI」の理解において、統計は重要な位置を占めており、高校「情報I」および「情報II」における統計・データサイエンスに関する学習の重要性が示されている。中でも機械学習は、統計の考え方を基盤にしてデータのパターンや法則性を見つけ出し、それをもとに予測や意思決定を行う技術である。高校「情報I」および「情報II」で学ぶ統計・データサイエンスの内容は、これらの基礎的な概念や手法を理解するための重要なステップとして位置づけられている。

たとえば、回帰分析や相関係数の理解は、データ間の関係性を探るための基本的なスキルである。また、データの可視化や確率分布の概念は、機械学習モデルの結果を正しく解釈し、意味ある洞察を引き出すために欠かせない要素であり、これらの知識を活用することで、生徒は単なるデータ分析にとどまらず、モデルの構築や評価に関する初歩的な理解を深めることができる。

さらに、機械学習の実践においては、大量のデータから自動的に学習し、汎化能力を持つモデルを構築する手法が重要になるが、その背後にあるアルゴリズムや統計的な理解があることで、ブラックボックス的に結果を信じるのではなく、理論に基づいた批判的な思考を養うことが可能となる。

このように、高校での統計・データサイエンスの学習は、単に数値を扱う技術を身につけるだけでなく、AIが社会に及ぼす影響や応用分野についても深く考えるための土台を築くことにつながる。

しかし、統計とAIの関係性について、実践的に学ぶ教材は存在するが、どの教材においても、敷居が高い。また、機械学習の細かなステップを学ぶには時間がかかるだけでなく、「情報II」を開設している学校が少ないことから、多くの生徒が学ぶことは現段階では困難である。

そこで、統計とAIの関係性について学ぶために、身近なデータで簡単に機械学習を行うことのできる



■図-4 「easyStat」の操作画面(例:相関分析)
参考: https://huggingface.co/spaces/itou-daiki/easy_stat_demo

解釈の補助

【勉強時間】と【スマホ時間】には中程度の負の相関がある($r=-0.51$)

【勉強時間】と【睡眠時間】にはほとんど相関がない($r=-0.04$)

【スマホ時間】と【睡眠時間】にはほとんど相関がない($r=-0.14$)

■図-5 解釈の補助機能の例

Webアプリケーションの開発を行った。主な条件は以下のとおりである。

- ①タブレット PC で軽快に動作することができる
- ②データの加工に関する操作を最小限に抑える
- ③タッチ操作のみで操作することができる
- ④機械学習のステップを理解できる

①②③に関しては、「easyStat」と同様、環境的要因に依存しない配慮を行った。④については、一般的な機械学習の流れの理解を促すために、ステップごとにボタンを実装し、使用者の理解を補助する工夫を行った。また、モデルのダウンロードや、モデルの可視化機能の実装も行った (図-6, 図-7, 図-8)。

これらのWebアプリケーションを開発することで、AIの基礎である統計や機械学習を学びながら探究を行うことができる。実際に、本校では、これらのWebアプリケーションを使って探究を行っている。また、探究の成果として、さまざまなイベントや学会を通じた成果の発表を行うようになった。イベントや学会に参加することで、専門家から助言を得るだけでなく、ほかの探究に触れることで、さらなる発展が期待できる。

統計を通した学びの変革

探究には「主体性」が重要であり、教員はあくまでもファシリテーションが役目である。生徒が自走するまでの支援が必要であるが、科学的な根拠を抽出するための統計および機械学習を用いた分析には、

今回開発したWebアプリケーションによって、探究を深めることができると考えられる。さまざまな教材やWebアプリケーションを活用した「学びの変革」によって、社会を変えようとする生徒が増えることを願っている。

情報は、探究だけでなく、各教科での学びをつなげる教科であり、今後の子どもたちにとって重要な位置づけである。「共通テスト」にとらわれることなく、この科目の素晴らしさを伝えていくことが、情報科を担当する教員の役目ではないだろうか。

(2025年1月6日受付)



伊藤大貴 (正会員)
itou-daiki@oen.ed.jp

大分県立大分舞鶴高等学校 情報科教諭。大分大学大学院教育学研究科修了後、大分県内の公立学校に勤務。統計・プログラミング教育に従事。科学部情報班の顧問として情報学研究や情報オリンピックの指導に尽力している。

チューニング結果

チューニング前

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.161	0.1503	0.3876	0.7817	0.2118	0.0792
1	0.1631	0.1494	0.3865	0.7961	0.2114	0.0796
2	0.2078	0.1997	0.4468	0.7188	0.2645	0.0737
3	0.1504	0.1318	0.3631	0.7744	0.2206	0.0531
4	0.1378	0.1257	0.3546	0.8219	0.197	0.0713
Mean	0.164	0.1514	0.3877	0.7786	0.221	0.0714
Std	0.0237	0.026	0.0323	0.034	0.023	0.0097

チューニング後

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.1465	0.1414	0.3761	0.7945	0.2036	0.073
1	0.1464	0.1437	0.379	0.8039	0.2054	0.0712
2	0.1833	0.1841	0.4291	0.7407	0.2565	0.0574
3	0.1389	0.1295	0.3598	0.7784	0.2176	0.0471
4	0.1232	0.1224	0.3499	0.8265	0.1935	0.0604
Mean	0.1477	0.1442	0.3788	0.7888	0.2153	0.0618
Std	0.0197	0.0214	0.0273	0.0286	0.022	0.0095

モデルのトレーニングが完了しました!

図-7 「easyAutoML」の操作画面 (チューニング)

モデル比較結果

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
et	Extra Trees Regressor	0.149	0.1521	0.3888	0.7778	0.2187	0.0661	0.2
rf	Random Forest Regressor	0.1503	0.1542	0.3918	0.7747	0.22	0.0691	0.196
knn	K Neighbors Regressor	0.1527	0.1729	0.4144	0.7459	0.2313	0.0742	0.042
xgboost	Extreme Gradient Boosting	0.1559	0.1828	0.4239	0.7349	0.2355	0.0744	0.308
gbr	Gradient Boosting Regressor	0.1699	0.1866	0.4307	0.7259	0.2393	0.0893	0.13
dt	Decision Tree Regressor	0.1857	0.2948	0.5408	0.5609	0.3001	0.1046	0.012
lightgbm	Light Gradient Boosting Machine	0.1944	0.1836	0.4274	0.7324	0.2375	0.1033	0.082
huber	Huber Regressor	0.2072	0.1588	0.3981	0.7671	0.2241	0.1159	0.046
br	Bayesian Ridge	0.2177	0.1618	0.4018	0.763	0.2275	0.1215	0.046
ridge	Ridge Regression	0.2225	0.1648	0.4055	0.7587	0.2294	0.1269	0.082

図-6 「easyAutoML」の操作画面 (モデル比較)

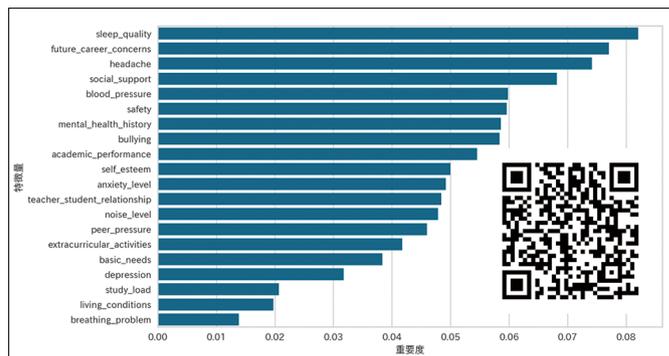


図-8 「easyAutoML」の操作画面 (可視化・評価)