

# 回帰分析初学者のための Web アプリケーション教材の提案と授業実践

理学専攻 情報科学コース 2140672 村上 綾菜 (指導教員：伊藤 貴之)

## 1 はじめに

技術のデジタル化が進み、AI やデータサイエンスをはじめとする情報科学技術の重要性が急激に注目されている。この傾向は教育分野においても同様で、政府は全ての高等学校卒業生が「数理・データサイエンス・AI」に関する基礎的なリテラシーを習得することを今後の教育の目標の一つとして掲げている。また、2022 年度から共通必修科目の情報 I が始まり、データ分析やプログラミングがとりあげられている。データサイエンスおよび AI に関する教育を推進するにあたり、初学者に適した学習内容の教材が必要である。そして、コンピュータを使用することに不慣れな生徒が一定数いることを踏まえると、データサイエンスの教育においても高度な ICT スキルを前提としない教材の開発が求められる。

本研究では、高校生をはじめとするデータサイエンス初学者を対象としたインタラクティブなデータサイエンス教材の一事例を提案する。現時点で本教材では分析手法を回帰分析に絞っている。また、あくまでも本教材は初学者への回帰分析の導入を目的としており、回帰分析に関する専門性の習得を目的としていない。本教材を通じて初学者が回帰分析への理解を深める経験が、回帰分析以外のデータ分析手法の学習を潤滑にできると考えられる。さらに、本教材を用いた授業実践についても述べる。授業内容は、単回帰分析の体験から開始し、データ可視化の重要性の議論や重回帰分析の体験へ展開するものであった。授業実践で得られたアンケートの回答と操作ログを分析することで本教材の学習目標の達成度についても議論する。

## 2 関連研究

データサイエンス手法の学習を目的とした既存の学習教材は大きく 2 種類に分類できる。1 種類目は、Web サイトやアプリケーションを利用したインタラクティブな学習ツールである。インタラクティブな教材を使うことで学習者はより早く意欲的に学習できるとされている。Alonso ら [1] は、認知負荷理論の視点から、学習者にとって負担が大きい内容を扱う際には、操作の容易な教材を使うことで学習内容に集中できると提唱する。一方で、現在このようなインタラクティブな教材で高校生を対象とした事例はまだ少なく、充実しているとはいえない。

2 種類目として、分析ソフトやプログラミング言語を用いた本格的なデータ分析環境があげられる。文部科学省が提示する高等学校情報科「情報 I」教員研修用教材では、Excel を活用した単回帰分析が紹介されている。また、高等学校情報科「情報 II」教員研修用

教材においては、Python や R を用いた分析方法の例がソースコードとともに紹介されている。こうした本格的な分析環境を用いる教材の場合、初学者には苦手意識が先行し、データサイエンス自体も難解に感じる可能性がある。その一方で、実践レベルの専門的な分析手法が学べるという利点があり、一定以上の学力と PC スキルを有する高校生の学習レベルには適している面がある。

以上の議論を踏まえて、初学者がデータサイエンスの概念を習う導入のために用いる教材と、概念の習得後に専門的な分析スキルを習得するための教材を、明確に分ける必要があると考える。具体的には、前者にはインタラクティブで直感的に操作できる教材を用い、その後の実データの分析練習には表計算ソフトウェアやプログラミングを活用するのが効果的であると仮定する。この仮定にもとづいて本研究では、回帰分析の概念を習う導入のためのインタラクティブな教材を提案する。

## 3 提案手法

Python の Web アプリケーション作成パッケージである Streamlit を用いて本教材を実装した。本教材を Web アプリケーションにて実装した理由は、本教材使用者が事前のインストール作業等の煩雑な環境構築をすることなく使用できるようにするためである。本教材メイン画面のスナップショットを図 1 に示す。本教材は、単回帰分析、重回帰分析、可視化の 3 つの機能を有しており、画面左側には、機能の切り替えボタンが搭載される。単回帰分析および重回帰分析の機能においては、学習者は説明変数と目的変数をクリック操作で選択することにより、回帰分析を実行する。



図 1: 教材メイン画面のスナップショット。

### 3.1 使用データ

本教材にはあらかじめ分析に使用するデータが読み込まれているものとする。現時点では、高校生男女の新体力テストのデータ [2](以後、体力測定データ)を適用している。体力測定データは、高1年生から高校3年生の男女合計948人の8つの種目の測定値を有する。

## 4 授業実践

本研究では、2022年11月にお茶の水女子大学附属高等学校1年生3クラスにそれぞれ授業実践を行なった。授業時間は、およそ45分を2コマ連続で実施した。

### 4.1 授業の流れ

本時の活動目標は、「ある生徒の立ち幅跳びの結果を予測する」ことである。なお、ここで「ある生徒」とは、体力測定データの中の1名であり、授業内では立ち幅跳びの結果以外の種目の測定結果を提示する。

最初に、本教材を用いて立ち幅跳びの予測値を単回帰分析で算出したのちに、生徒の予測結果を比較する。次に、予測精度を向上させる工夫を問いかけ議論する。得られた回答の中から「データのフィルタリング」と「重回帰分析」を取り上げ、本教材上で体験して確かに予測精度が上がることを確認した。

### 4.2 生徒の様子

単回帰分析での予測においては、立ち幅跳びの結果を予測するにあたり「あなたが信頼できる説明変数は何か。また、なぜか。」と問いかけた。その回答を原文のまま抜粋して以下に示す。

- シャトルラン：データが回帰線にギュッと寄っていて、分散が小さめだったから。
- 身長：データ可視化で、身長が低い人ほど立ち幅跳びで跳べる距離が短く、高い人ほど長いという傾向が顕著に出ていたから。
- 握力：瞬発的な筋力を測るという点で関連するのではないかと思ったから。
- 反復横跳び：同じく足の筋肉を使う種目だから。

説明変数として生徒が選んだ変数にばらつきが見られた一方で、説明変数の選択理由の回答は大きく2つの視点にもとづくものが多かった。1つ目は、散布図から読み取れる数学的視点に着目した理由である。具体的には、上記4つの回答のうち前半2つが該当する。数学の専門用語を使ってないものもあるが、説明変数と目的変数の相関関係、説明変数の分散を本教材上の散布図から読み取っていることがわかる。2つ目は、体力測定データの特性つまり各種目の内容に着目した理由である。具体的には、上記4つの回答のうち後半2つが該当する。身近なデータであるからこそ、これまでの自分の経験と紐付けて考察できている。この力は、文部科学省が掲げる「情報活用能力」の「世の中の様々な事象を情報とその結びつきとして捉え、情報及び情報技術を適切かつ効果的に活用して、問題を発見・解決したり自分の考えを形成する力」に該当すると考えられる。

### 4.3 アンケートの評価

授業の前後でアンケートを実施し、データサイエンスに対する興味と、自分との関連性を調査した。各設問に対し、「全くそう思わない」を1、「とてもそう思う」を5として5段階の選択肢から回答した上で、その理由を自由記述欄に述べる。5段階の回答に対し、授業前後の変化を両側T検定にて有意性を検証し、ヘッジズの効果量も算出した。その結果を表1に示す。なお、Q1「データサイエンスに興味はあるか」Q2「データサイエンスは自分の将来の進路に関係あると思うか」である。

表1: 授業前後のアンケート回答比較

設問	平均値	標準偏差	p値	効果量
Q1	3.80	0.83	0.00002	0.54
	4.21	0.68		
Q2	3.92	0.80	0.00004	0.51
	4.30	0.67		

表1の結果より、本授業を経てデータサイエンスに対する興味をもつ生徒が増加したことがわかる。また自由記述の回答を授業前後で比較すると、授業後の回答のほうが、データサイエンスと自分の進路や生活の関係性について具体例を伴いより詳細に言及したものが多くみられた。以上のことより、本教材を用いた授業実践がデータサイエンスやAIに対する興味と理解を向上させたと考えられる。

## 5 まとめ

本研究では、データサイエンス初学者を対象とし、回帰分析を例題としたオンライン型のデータサイエンス教材を提案した。また、データサイエンス初学者である高校1年生を対象として、本教材を用いた授業を実施した。授業内での生徒の活動をみると、本教材の利用者は、データ分析を身近に感じ深く理解していた人が多いことがわかった。今回の実践授業では、回帰分析の導入教材に焦点を当てたが、今後生徒が本格的に表計算ソフトウェアやプログラミングを使ったデータ分析に挑戦する場合でも、今回の学びが活けると期待できる。

## 謝辞

本教材を用いた授業実践にご協力くださったお茶の水女子大学附属高等学校の先生方および生徒のみなさんに感謝の意を表します。

## 参考文献

- [1] J. C. Castro-Alonso, B. B. de Koning, L. Fiorella, F. Paas, "Five Strategies for Optimizing Instructional Materials: Instructor and Learner-managed Cognitive Load.", *Educational Psychology Review*, vol. 33, no. 4, pp. 1379-1407, 2021.
- [2] 科学の工具箱 データライブラリ, <https://rika-net.com/contents/cp0530/contents/04-03-01.html>.