

データサイエンス初学者向け Web アプリケーション教材の実践と分析

村上 綾菜* 伊藤 貴之*

概要. 近年、高校生の段階からデータサイエンスや AI の基礎を学ぶことが求められている。具体的には、高等学校における教科「情報」の開始や、政府による AI 戦略が挙げられる。しかし、実際の教育現場ではデータサイエンス初学者に適した教材が不足しており、表計算ソフトウェアやプログラミングに慣れていない学習者にとってはハードルが高いのが現状である。そこで本研究では、クリック操作のみで単回帰分析を学習できる Web アプリケーション教材を提案し授業実践を行う。また、予測精度を高める方法として、重回帰分析への展開もする。本稿では、昨年度の実践から得られた反省を生かし、単回帰分析から重回帰分析への学びの展開を目指した Web アプリケーション教材の開発と、それをを用いた実践授業についてまとめる。

1 はじめに

今年度より、教科「情報」が共通必修科目として本格的に始まり、事実上データ分析が必修となった。中でも学習すべき項目の1つとして、単回帰分析が挙げられている。しかし、既存の学習教材は汎用性が高い故に、操作が複雑であり初学者にとって煩雑である。そこで、本研究では単回帰分析に焦点を当てた初学者向けのデータサイエンス教材を提案する。

また、単回帰分析から重回帰分析への学びの展開を目指す。これは、昨年度の単回帰分析をテーマとした授業実践の中で、事後アンケートにて複数の変数を同時に予測に使用する必要性を説く生徒がいたことから、重回帰分析の体験も有益であると考えた。本稿では、今年度を実施した本教材を用いた授業の内容と、授業の前後で実施した生徒のアンケートの結果についてもまとめる。

2 関連研究

新たな学習項目は、学習者にとって負担が大きい。Alonso ら [3] は、認知負荷理論の視点に基づき、脳内のワーキングメモリの効率的な活用の重要性を述べている。また、森山ら [4] は、ICT に対する苦手意識が情報活用の実践力に対する意欲を減衰させてしまい、情報科の学習に対して自己効力感が十分に高まらないと述べる。初学者を対象とする学習教材は、簡易な操作で扱えるシステムにすることが求められる。しかし、文部科学省が公開した「情報 I」の教員研修用教材 [2] では、Microsoft Excel (以下、Excel) を用いた単回帰分析の学習を提案する。Excel は汎用的だが、複雑な操作が原因で、学習の狙いとは異なる点でつまづく可能性が非常に高い。そこで本研

究では、クリック操作のみで回帰分析を学習できる Web アプリケーション教材を提案する。

3 提案システム

本教材では、高校生の体力テストの種目別測定値 [1] のデータを使用する。以下に示す、主に3つの機能で構成される。

3.1 単回帰分析

図1に示す画面で、単回帰分析を行うことができる。ユーザは、クリック操作にて目的変数と説明変数をそれぞれ1つ選択できる。単回帰分析実行後は、選択した2変数を軸に持つ散布図と回帰直線が描画され、マウスホバーによってユーザは詳細情報を参照できる。

単回帰分析による予測

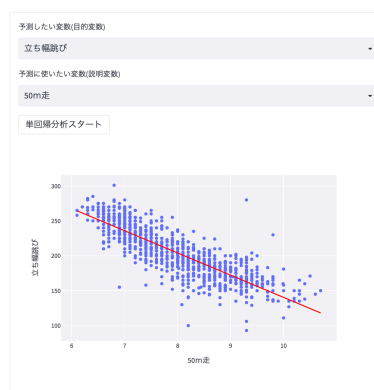


図 1. 単回帰分析機能の画面

3.2 データ可視化

使用データの中から任意の変数を選び、散布図、ヒストグラム、箱ひげ図の3種類のグラフで可視化

できる。また、散布図においては、データの属性に応じた色分けも可能である。例えば、図2に示す画面は、性別に応じて点群の色を分け、ユーザが指定した変数を軸とする散布図を表示する。複数の可視化結果を容易に選択や比較できることにより、学習者がデータを多角的に捉えることを期待する。



図 2. データ可視化機能の画面

3.3 重回帰分析

図3に示す画面で、重回帰分析を行うことができる。ユーザは、クリック操作にて1つの目的変数と複数の説明変数を選択できる。重回帰分析実行後は、選択した説明変数を用いた回帰式と予測値、決定係数が文字情報として表示される。



図 3. 重回帰分析機能の画面

3.4 使用データの選別

単回帰分析・重回帰分析の2つの機能を使用する際には、分析に使用するデータをフィルタリングできる。選択肢は、男女、性別別、性別及び学年別、の3種類である。

4 実践授業と評価

4.1 概要

2022年11月に実施した、本教材を用いた実践授業の概要は以下のとおりである。

- 対象者: お茶の水女子大学附属高等学校の1年生
- 授業時間: 98分(うち休憩8分含む)
- 回数: 各クラス1回ずつ3クラス実施

事前に座学形式の講義を通じて、単回帰分析の概要を含んだデータサイエンスに関する知識は学習済みの状態とし、本授業実践はその実践編にあたる。

単回帰分析を用いた予測を体験したのち、その予測精度を向上させる方法を考える。その過程で「重回帰分析」と「データの属性を考慮したデータ選び」を伝える。詳細は発表にて説明する。

4.2 評価: アンケート比較

本実践の前後でアンケートを実施し、前後での比較を行った。観点は興味関心と理解度の2つである。まず、データサイエンスに対する興味関心に関する回答の比較を図4に示す。

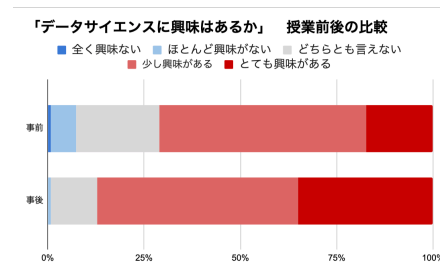


図 4. データサイエンスに対する興味の変化

授業を経て、データサイエンスに興味を示す生徒の割合が増加したことが読み取れる。また、自分の進路になぜデータサイエンスが関係あるかという質問に対する自由記述欄には、「データサイエンスは様々な場面で活用することができることが分かったから。」といったデータサイエンスの分野横断の性質に言及するものだけでなく、「AIの下した判断を鵜呑みにするのではなく、私たち人間が判断することが必要になるから。」のように、ITリテラシーの観点からも意識の変化を読み取ることができた。

5 まとめ

本稿では、データサイエンス初学者向けのWebアプリケーションを提案し、またそれを用いて実施した実践授業の詳細をまとめた。詳細なアンケートの分析結果は発表にて言及する。

アンケート結果からもわかるように、身近なデータを用いて体験的に学習することはデータサイエンスにおいても有用であることが確認できた。今後はログの分析も進めていきたい。

参考文献

- [1] 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>.
- [2] 高等学校情報科「情報 I」教員研修用教材. https://www.mext.go.jp/content/20200722-mxt_ogai02-100013300_06.pdf.
- [3] J. v. M. F.Paas. Five strategies for optimizing instructional materials: Instructor-and learner-managed cognitive load. Vol. 33, pp. 1379–1407. Springer, 2021.
- [4] 森山, 原田, 福井, 中尾, 小倉, 近澤, 山下. 高校生の ICT に対する苦手意識と情報活用実践力および自己効力感との関連性. pp. 65–75, 2020.