

回帰分析の導入を目的とした Webアプリケーション教材の提案と授業実践

村上 綾菜† 伊藤 貴之† 山本 周††

お茶の水女子大学 人間文化創成科学研究科†

聖学院 中学校高等学校††

1.はじめに

技術のデジタル化が進み、デジタル機器が身近になるにつれて、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。コンピュータの使用に不慣れな初学者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。

本報告では、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。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。さらに本報告では、本教材を用いた授業実践についても述べる。

2. 情報Iとデータサイエンス教材

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、インタラクティブな学習ツールである。これらの教材の利点は、パソコンの操作が苦手な生徒でも学習に取り組むことができる。しかし、現在このような教材で高校生を対象とした事例はまだ少ない。2種類目は、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。

Proposal of Web Application for Introduction of Regression Analysis

Ayana Murakami†, Takayuki Itoh†, Shu Yamamoto††

†Ochanomizu University

†† Seigakuin Junior and High School

文部科学省が提示する高等学校情報科教員研修用教材では、Excelを活用した単回帰分析やPythonやRを用いた分析方法の例が紹介されている。こうした本格的な分析環境を用いる教材は、初学者には苦手意識が先行し、内容自体も難解に感じる可能性がある。一方で、専門的な分析手法が学べるという利点があり、一定以上のスキルを有する高校生には適している面がある。

以上より、我々は初学者がデータサイエンスの概念を学ぶ導入のために用いる教材には直感的に操作できる教材を活用するのが効果的であると仮定し、この仮定にもとづき本報告では、回帰分析の概念を学ぶ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

我々はPythonのWebアプリケーション作成パッケージであるStreamlitで本教材を実装した。Webアプリケーションとして実装することで、本教材使用者は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。本教材のメイン画面のスナップショットを図1に示す。



図1 教材メイン画面のスナップショット。

本教材は、単回帰分析、重回帰分析、可視化の3つの機能を有している。

3.1 使用したデータ

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。現時点で我々は、高校生男女の新体力テストのデータ[1](以後、体力測定データ)を適用している。体力測定データは、高校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。

3.2 本教材の機能：単回帰分析

図1に示した画面は、使用者が単回帰分析のために説明変数と目的関数を選択する画面である。画面上部のUIから説明変数と目的変数を選択したうえで、下のボタンをクリックすると単回帰分析が実行される。ここでは、単回帰分析による予測値を直接表示するのではなく、説明変数を横軸、目的変数を縦軸とする散布図と回帰直線を表示する。これにより、数学の一次関数と関連付けた単回帰分析の理解を促す。

4. 授業実践

我々は、2022年11月にお茶の水女子大学附属高等学校1年生(45分を2コマ) 3クラスと聖学院中学校・高等学校の高校3年生(50分を2コマ) 5クラスを対象にそれぞれ授業実践を行なった。

4.1 授業の流れ

本時の大きなゴールは、「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名であり立ち幅跳びの結果以外の種目の測定結果は提示する。

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、予測精度を向上させる工夫として「データのフィルタリング」と「重回帰分析」を取り上げ、本教材上での体験を通じ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

立ち幅跳びの結果を単回帰分析で予測するにあたり「自分が信頼できる説明変数は何か、また、なぜか。」と問いかけ、Google Formにて回答を集めた。その中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。

回答結果では、説明変数として選ばれた変数にばらつきが見られたが、その説明変数を選択した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(上述4つの回答のうち前半2つ)である。具体的には、上に示したが該当する。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種目の内容に着目した理由(上述4つの回答のうち後半2つ)である。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付きとして捉え、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。

5. まとめ

本報告では、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。我々は、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、本教材の使用者は、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。本報告では、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、今回の学びが活きると期待できる。

参考文献

- [1] 科学の道具箱 データライブラリ, (<https://rika-net.com/contents/cp0530/contents/04-03-01.html>).