

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

村上 綾菜[†] 伊藤 貴之[†]

[†] お茶の水女子大学人間文化創成科学研究科理学専攻情報科学コース 〒112-8610 東京都文京区大塚 2-1-1
E-mail: †{murakami.ayana,itot}@is.ocha.ac.jp

あらまし 高等学校における共通必修科目「情報」が始まり、政府が掲げる AI 戦略でも高校生の段階から AI・データサイエンスを学ぶことが求められている。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不慣れな学習者にとって難解であることが現状である。そこで我々は、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開発している。本報告では、開発した Web アプリケーション教材を提案するとともに、それをを用いた実践授業について紹介する。さらに、実践授業にて収集したアンケート結果と操作ログを用いた学習効果の評価について論じる。

キーワード 可視化, ユーザ支援, HCI, 教育, データサイエンス

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 [1] 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。

高等学校において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、高校生に適した学習内容の教材が必要である。そして学習の手段についても注意しなければならない。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。

本報告では、高校生および大学学部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おり、他の分析手法についてはまだ実装していない。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。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。

我々は、本教材の開発にあたり、以下の 3 点を主要な要件とした。ここで定めた要件の実現が本教材の大きな特徴ともいえる。

要件 1: 高校数学 I の学力の範囲内で単回帰分析の概念を理解し、活用できる。

要件 2: プログラミングを知らなくても、また複雑な操作を覚えなくても、生徒がオンライン教材を使いこなせる。

要件 3: 使用者が自分の経験を活かしながら楽しく回帰分析を学習できる。

我々は、この要件の達成を目指して本教材を開発すると同時に、想定使用者の中でも最も学年の低い高校生 1 年生が使用することを鑑み、本ツールに以下の 3 点の工夫を施すこととした。

1 点目として、**要件 1** のために、高校生 1 年生が学習済みの可視化手法のみをデータの視覚表現に用いる。現行の高等学校学習指導要領解説 [2] によると、データの可視化手法の中でも、中学校の数学では箱ひげ図とヒストグラムが、高校の数学 I では散布図が、それぞれ学習項目とされている。本教材ではこれらを視覚表現に用いる。また、単回帰分析での予測値の表示の際に、散布図と回帰直線を用いて可視化することで、予測値の算出方法の理解を促進する。

2 点目として、**要件 2** のために、全ての操作を Web ブラウザ上での GUI ベースで実装するとともに、GUI 部品を必要最低限にし、クリック操作のみではほぼ全ての操作を実現できるようにする。GUI を工夫するにより、生徒にとって苦手意識の可能性のあるキーボード入力を不要にする。この実装方針により、学習内容以外の点での負担を減らすことで、学習内容への集中を促す。

3 点目として、**要件 3** のために、高校生にとって身近なデータを題材とする。現時点での実装では新体力テストの測定値を題材とした。データの結果と自分の経験を結びつけることで、データが身近なものであることを理解し、同時にデータ分析も身近なものであることを理解することを期待する。

我々は本教材を、お茶の水女子大学附属高等学校の 1 年生の情報科の授業にて実習題材として使用した。本報告では、授業の際に取得した授業前および授業後のアンケート結果と操作ログを解析することで、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。具体的には、科学の道具箱 [3] や Bowland Japan が公開している教材 [4] があげられる。これらの教材の利点は、パソコンの操作を苦手とする生徒でも学習に取り組むことができ、ゲーム感覚で楽しく学ぶことができる点である。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 [5] 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 [6] 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。神部ら [7] は、R を用いた統計教育教材を採用しているが、機器の操作方法の習得へ意識が奪われている学生がかなり見受けられたと報告した。森山ら [8] 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、情報科の学習に対して自己効力感が十分に高まらなると主張している。まとめると、本格的な分析環境を用いる教材は、プログラミング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といえる。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我々は、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報告では、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。

インタラクティブな教材の使用に関して Najjar [9] は、これらを使うことで学習者はより早く学習し、さらにその学習態度もより優れているとまとめている。また Alonso ら [10] は、認知負荷理論の視点にもとづき、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。具体的には、新たな内容の学習は学習者にとって負担が大きいため、操作の難易度の低い教材を使うことで学習内容に集中できることが重要であると考えられる。以上のことから、初学者を対象とする学習教材は、簡易な操作で実行できるインタラクティブな教材であることが望ましいといえる。一方で藤田 [11] は、ボタン1つでグラフや表が表示されることから、その概念の表面的な理解にとどまる危険性を提起する。このことから、学習教材においては、インタラクティブな教材のメリットを十分生かしつつも、学習者が表面的な理解で終わらない工夫が必要であるといえる。表面的な理解に陥るのを避けるためには、Entwistle ら [12] が提唱する学習への「深いアプローチ」を実現する必要がある。この提唱の中にある6つの観点のうち、本報告では、「既存の知識・経験に考えを関連づける」という観点を、学習者にとって身近な

データを分析対象のデータに選ぶことにより実現する。また Wiggins ら [13] は、知識を他の知識・考え・経験と関連づけて一般化することで、深い学びに導くと提唱する。このことから、本報告ではデータ分析と数学を切り離すことなく、数学の知識と関連づけてデータの読み取りができる工夫を大切に、データ分析への深い学びを促す。

3 提案教材

本章では我々が開発した教材の概要と詳細について説明する。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 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。我々が開発した Web アプリケーション教材のメイン画面のスナップショットを図 1 に示す。本教材は主に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。それぞれの機能については第 3.3 節で詳細に述べる。



図 1 教材メイン画面のスナップショット。

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 [15] (以後、体力測定データ) を適用している。使用する体力測定データの統計情報を表 1, 2 に示す。体力測定データは、8つの種目の測定値と換算得点の2種類を有する。換算得点とは、新体力テストの実施要項 [16] に記載される1~10点の指標であり、性別および学年の属性別に種目ごとにそれぞれ各得点の定義域が定められている。本教材では、単回帰分析の理解のために散布図を用いて可視化するため、量的変数である測定値を採用する。

我々が体力測定データを選択した理由は主に2つある。まず、文部科学省が作成した情報 I の教員研修用教材 [5] における回

表 1 体力測定データの件数

	合計	女子	男子
高 1	318	182	136
高 2	314	173	141
高 3	316	181	135

表 2 体力測定データの測定値の統計情報

項目	平均(全体)	平均(女子)	平均(男子)
身長	162.6	157.4	169.3
体重	55.1	52.0	59.1
座高	87.7	85.3	90.9
握力	32.6	26.9	39.9
上体起こし	31.4	28.4	35.2
長座体前屈	55.1	53.5	57.1
反復横跳び	52.6	48.5	58.0
シャトルラン	73.6	57.3	94.8
50 m走	8.0	8.6	7.2
立ち幅跳び	204.2	180.5	235.1
ハンドボール投げ	21.5	16.1	28.7

帰分析の授業例においてまさに体力測定データが例示されているからである。一方で、研修用教材にて取り上げられる授業実践例は表計算ソフトウェア Excel を用いたデータ分析方法であり、学習方法の点において本報告の提案内容は大きく異なるものである。また、文部科学省が提示する資料と同じデータを使用することで、将来的に本教材を広く普及する場合でも理解を得られやすいことを期待する。次に、本教材が主に対象とする高校生および大学学部生にとって、体力測定データは身近であることが理由としてあげられる。新体力テストは平成 11 年よりほぼすべての公立小中高等学校で実施されている。本教材が想定する使用者は、高校生および大学学部生であり、これまでの学校生活の中で新体力テストの経験がある生徒および学生も多いことが予想される。言い換えれば、本教材の対象者の多くにとってデータ項目が既知であること、ひいては自分の経験と比較しながらデータを扱うことができると期待される。

表 2 に示すように、このデータでは男女で測定値に大きな差がある点が重要な特徴である。この特徴を把握して回帰分析に活かせるかがデータ活用能力の育成において重要な点の一つになる。

3.2 学習目標

本教材は、回帰分析の導入教材を想定する。対象ユーザは高校生や大学学部生をはじめとするデータ分析初学者である。本報告では、本教材を用いた授業実践も併せて実施した。第 1 章では教材そのものの目的となる要件 3 つを列挙したが、特に要件 1 をさらに細分化し、本教材を用いた授業実践の学習目標を以下のように定める。

目標 1: 単回帰分析を適切に活用して目的に沿った予測値を算出できる。

目標 2: 分析対象データの変数の特性に着目した分析ができる。

目標 3: 精度の高い予測のための方針について仮説を立ててそれを検証できる。

本研究では、ここに定めた目標が達成されるように授業設計と教材設計を目指す。

3.3 教材の画面

本教材においてユーザに課する作業は、「回帰分析に用いる変数の選択」と、分析対象データの理解を目的とした「データ可視化のグラフおよび変数選択」である。その結果として、ユーザに求める最終的な目標を「特定の種目のより精度の高い予測」であるとした。この作業を使用者が円滑に遂行できるように、本教材では以下の機能を実装する。

- 単回帰分析のための説明変数の選択
- 重回帰分析のための説明変数の複数選択
- データ可視化結果上のデータの詳細情報の確認
- データ可視化手法の選択
- それぞれの可視化手法における使用変数の選択
- 分析データの属性によるフィルタリング

ユーザは上記の機能を適切に使用しながら、回帰分析を用いた予測値の算出を目標にし、より精度の高い予測のための試行錯誤を繰り返すものとする。

本教材は以下の 3 つの画面で構成されている。なお、最後の「分析データの属性によるフィルタリング」機能については、第 3.4 節で詳しく述べる。

3.3.1 単回帰分析画面

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。画面上部のプルダウン式の UI から説明変数と目的変数を選択したうえで、「分析スタート」と書かれたボタンをクリックすると、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、図 1 中央に示すように、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。これには、第 1 の要件 1 に示した単回帰分析への理解を目的とする。回帰直線を用いて予測をすることで、数学の一次関数と関連付けた単回帰分析の理解を促す。

なお、マウスのカーソルを散布図上の点に重ねると、その点に対応するの座標値がカーソル付近に表示される。また、マウスカーソルを回帰直線上に重ねると、その点の x 座標における y 座標（予測値）をカーソル付近に表示する。散布図と合わせて回帰直線を表示することで、使用者が目的変数と説明変数の相関関係や、複数の数値を回帰直線に代入することによる予測値の妥当性の検証を参考にして、使用者が総合的に考えながら試行錯誤をすることを期待する。

本教材では、主に以下の 2 つの理由により、予測値を直接表示するだけでなく、散布図とあわせて回帰直線を表示しその可視化結果をもとに予測作業をさせている。

1 つ目の理由は、使用者が感覚的に単回帰分析における予測値の算出方法を学ぶためである。例えば、1 次関数である直線をもとに任意の x 座標から対応する y 座標を求める計算は、中学校の数学の時点で学習してある。本教材を用いて可視化された回帰直線をもとに予測値を算出する経験を通じて、生徒は体感的に、単回帰分析の予測値算出に 1 次関数の内容が使われて

いることを学ぶ。また、説明変数をいくつか試す過程の中で、説明変数と目的変数の相関が強いものほど予測値が分散しづらいことに気づくことを促すこともできる。こうした数学と情報にまたがる教科横断の内容は、座学だけで学ぶよりも、最初に視覚的な体験をしその後改めて学習するほうが深い理解につながると考える。

2つ目の理由は、複数の数値を回帰直線に代入することによる予測値の妥当性を検証する機会を与えるためである。本教材では体力測定データを使用しているため、自分の経験と予測値を紐づけて考えることができる。例えば、説明変数に身長、目的変数に立ち幅跳びを設定した場合、予測したいデータの予測値を見るだけでなく、自分の身長に対応する予測値をみて立ち幅跳びの予測値が妥当かを考えることができる。使用者に身近なデータを分析対象に用いることの強みは、自身の経験にもとづく試行錯誤ができることであるため、その強みを活かすためにも回帰直線を散布図上に表示し、任意の数値から回帰直線を用いて瞬時に予測値を算出できる状態は非常に大きな意味を持つ。

3.3.2 データ可視化画面

データ可視化画面で、分析対象データのユーザはテストデータのプロット位置と境界線の位置関係を確認することができる。その画面のスナップショットを図2に示す。また、単回帰分析画面と同様に、それぞれの可視結果上の点にマウスオーバーすることで、その点に対応する詳細の座標値を読むことができる。

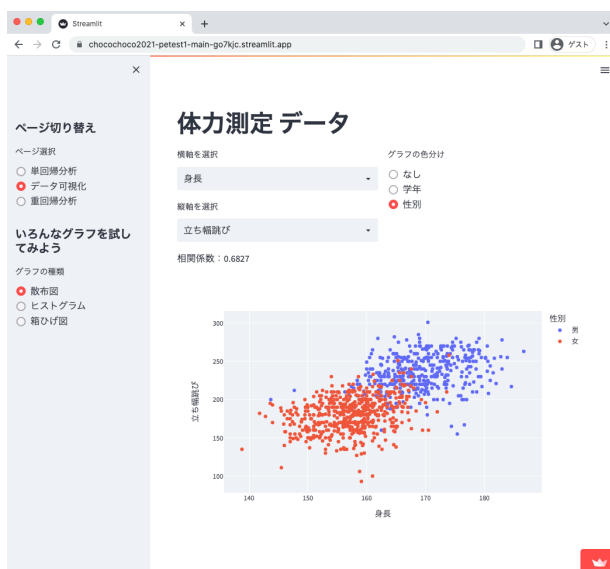


図2 教材のデータ可視化画面のスナップショット。

本教材では、対象とする学習者（特に高校1年生）の学力で十分に読み取ることができる可視化手法のみを採用する。現行の学習指導要領[2]において、散布図は高校の数学I、箱ひげ図とヒストグラムは中学校の数学の学習項目とされている。そこで、本教材で選択できる可視化手法を、散布図、箱ひげ図、ヒストグラムの3種類とする。

図2は、散布図を選択した場合の画面のスナップショットである。散布図では、画面上部に見えるUIを操作することによ

り、縦軸、横軸それぞれの変数、属性別のデータ色分けの有無を選択できる。また、画面上に選択した2変数の相関係数を表示する。

図3はそれぞれ、箱ひげ図を選択した場合の画面のスナップショットと、ヒストグラムを選択した場合の画面のスナップショットである。なお、紙面の都合によりどちらも、画面左部のサイドメニューは図から割愛してある。箱ひげ図では、任意の1つの変数を選択しそのデータ分布をみることができる。1つの変数につき箱ひげ図は2種類表示され、左が学年別の分布、右が性別別の分布を示す。図4に示すヒストグラムでも同様に、任意の1つの変数を選択しそのデータ分布をみることができる。

体力測定 データ

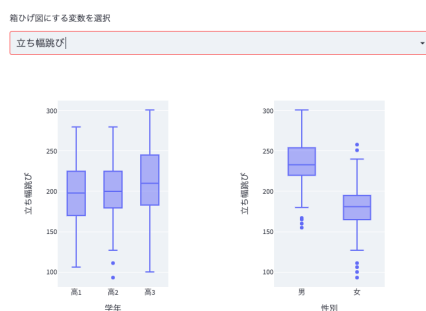


図3 箱ひげ図を選択した場合の画面のスナップショット。

体力測定 データ

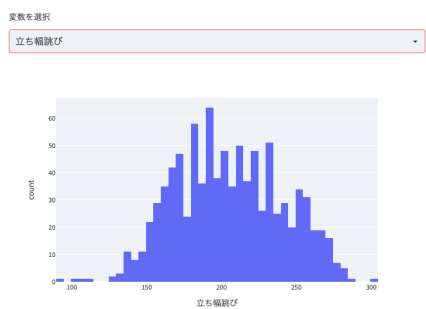


図4 ヒストグラムを選択した場合の画面のスナップショット。

本教材の使用者は、データ可視化機能を用いてデータの分布を確認し、データの特徴をつかむことができる。本教材を用いて授業する際には、このデータ可視化機能の活用方法次第で多様な授業展開を検討することができる。具体例を2つ例示する。

1つ目の展開例として、十分な授業時間がある場合には、分析目的に応じたデータ選択およびデータクレンジングの重要性の理解を促すことができる。データ分析においてデータの特徴を理解し、必要に応じてフィルタリング等のデータクレンジングをすることは非常に大切である。本教材はあくまでも回帰分析の導入を目的としているため、データクレンジングを実行することはないが、データの特徴や偏りの危険性を把握したうえで適切なデータの使用を意識づけることができる。

2つ目の展開例として、授業時間が限られている場合には、第3.3.1項に示した単回帰分析での変数選択において、選択理

由をデータ可視化結果をもとに説明させることができる。根拠にもとづいてデータ分析をすることは重要であり、複数の可視化結果を俯瞰的に見ることにより「情報活用能力[14]」の要素である「情報を複数の視点から捉えようとする」力を養うことができる。

3.3.3 重回帰分析画面

図5には、使用者が重回帰分析のために説明変数と目的関数を選択する画面を示す。UIの構成は、第3.3.1項に示した単回帰分析画面とほぼ同じだが、重回帰分析は複数の変数から1つの目的変数を予測するため、複数の説明変数をUI上で選択可能である点が単回帰分析画面と異なる。現時点での実装では、体力測定データの変数の中から任意の個数を説明変数として選択することができる。説明変数と目的変数を選択したうえで、「分析スタート」と書かれたボタンをクリックすると、重回帰分析が実行される。ここで分析結果の表示方法は、第3.3.1項の単回帰分析画面と大きく異なる。ここでは、重回帰分析によって算出した回帰式と予測値を直接数値で表示する。選択された3つ以上の説明変数を次元削減等を用いて可視化したとしても、本教材の想定使用者は次元削減等を適用した散布図を提示しても見慣れていない可能性が高いため、混乱を生じさせる可能性がある。本教材は、導入のための教材を想定しており、苦手意識を持つことを避けたいため、次元削減等を適用した散布図による可視化は割愛した。ただし、理解を促すために予測値の算出過程を文字を使って画面上に記述した。

目的変数に「立ち幅跳び」の測定値、説明変数に「身長」と「50m走」を選択した際の分析結果を図5に示す。このように、予測対象のデータを回帰直線に代入して予測値を算出するまでの途中式を記述することで、数学と情報の関係を意識しながら重回帰分析における予測値の算出方法を理解することを期待する。



図5 重回帰分析画面のスナップショット。

3.4 分析データのフィルタリング

本教材は、単回帰分析画面と重回帰分析画面でのみ、簡易的

なデータのフィルタリング機能を搭載している。図1および図5のサイドメニュー下部に注目されたい。タイプ1からタイプ3までラジオボタンで選択可能であり、デフォルトではタイプ1が選択されている。これは、分析に使用するデータを属性（現在使っているデータでは性別および学年）でフィルタリングするための機能である。現時点での実装でのフィルタリング方法は3種類で、フィルタリングなし、特定の性別のみのデータ、特定の性別かつ特定の学年のデータ、の3つから選択可能である。表2で示した通り、体力測定データは男女差の大きいデータである。そのため、男女混合での回帰分析によって特定の性別の予測をすると、大幅に予測値がずれることが事前実験によりわかっている。本教材は、あえてその特性を活かしてより良い予測精度のためにできることはないか生徒に考えさせ、その後、このデータフィルタリング機能を使って適切なデータを用いて予測した場合の予測値のズレを体験させることができる。

我々は、授業の展開に応じてサイドメニューに記載するボタン名を変更して授業を実施してきた。図1および図5に示す教材においては、タイプ1が男女合同データ、タイプ2が女子のみのデータ、タイプ3が高1女子のみのデータである。なお、タイプ2およびタイプ3は、授業対象者の属性と授業展開に応じて変更する可能性がある。「タイプ1」ではボタン名を伏せることも可能であり、あるいは「男女混合データ」のように明示することも可能である。本報告での実践授業では、生徒自身に予測精度を高める工夫を仮説として考えさせ、それを本教材のデータフィルタリング機能を使って検証するプロセスを経験させるために、ボタン名を伏せる形式を採用した。この点に関しては、第4章で詳しく述べる。

3.5 取得可能な操作ログ

本教材には使用者の操作ログを記録する機能がある。この機能では、使用者が所定の操作を行うごとに、その時刻と操作内容を記録する。操作ログが記録されるタイミングとその記録内容を表3に示す。なお、本教材上で起動時にユーザ名を設定でき、操作ログは一意に操作者を特定できる。本名や学籍番号に関係ないユーザ名を設定することで、受講者の匿名性を確保することもできる。

表3 操作ログの取得タイミングとその記録内容

画面	タイミング	取得内容
単回帰	単回帰分析実行時	分析手法, 説明変数名, 目的変数名 フィルタリングの有無
可視化	散布図描画時	可視化手法, x軸の変数 y軸の変数, 色分け
可視化	箱ひげ図描画時	可視化手法, 描画する変数
可視化	ヒストグラム描画時	可視化手法, 描画する変数
重回帰	単回帰分析実行時	分析手法, 説明変数名のリスト 目的変数名, フィルタリングの有無

4 本教材を用いた授業実践

本章では、本教材を用いて実施した授業実践の概要と詳細に

ついて述べる。我々は、2022年11月にお茶の水女子大学附属高等学校1年生を対象に授業実践をした。以下に概要を示す。

- 対象：お茶の水女子大学附属高等学校1年生(3クラス)
- 授業者：筆頭著者
- 授業形式：PC室にてクラスごとに実施
- 授業時間：45分授業2コマ(うち1クラスは、40分の短縮授業日の実施につき40分授業を2コマ)

生徒はまず、第2著者によるデータサイエンスの概論を座学形式で学習した。次に、その講義から数週間以内に本教材を用いた授業実践を実施した。本授業実践の位置づけは、座学形式の概論を体感するために手を動かす実習の時間であった。なお、生徒全員が当該年度に情報Iを履修中であったが、本授業の実施時点にて、情報Iのデータ分析は未習であった。

4.1 授業の流れ

本時の大きなゴールは、「ある女子生徒の立ち幅跳びの結果を予測する」ことである。なお、この女子生徒を本報告では予測対象生徒とよぶこととする。本時で使用したスライドの抜粋を図6に示す。

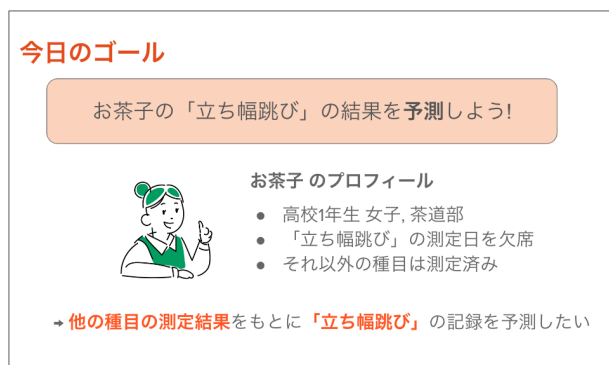


図6 本時で使用したスライド(一部抜粋).

本時の授業の展開を以下に示す。なお、下記展開の中での「予測」とは、「予測対象生徒の立ち幅跳びの結果を予測する」ことを指す。

- (1) 数件のデータをもとに予測
- (2) 本教材を用いて単回帰分析で予測
- (3) より高い予測精度にする工夫を考える
- (4) 本教材を用いて重回帰分析で予測。または、分析使用データをフィルタリングして予測。
- (5) まとめと他の学習項目への接続
 - 単回帰分析と重回帰分析
 - データバイアスの危険性
 - 判別分析
 - 画像分析 AI の実用例

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「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る、分散が小さめだったから。
- 50m走：散布図を見ると50m走一番ばらつきが少ないから。また、赤線のデータのところにデータが集まっており、予想が当たりやすいと思われる図になっているから。
- 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。
- 上体起こし：腹筋が強い→体幹がしっかりしている→体に軸がある→ブレずに跳べる
- 反復横跳び：同じく足の筋肉を使う種目だから。

回答結果から、説明変数として生徒が選ぶ変数にばらつきが見られた。一方で、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である。具体的には、上に示した6つの回答のうち前半3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、説明変数の分散を単回帰画面またはデータ可視化画面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種目の内容に着目した理由である。具体的には、上に示した6つの回答のうち後半3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[14]」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 学習効果の評価と考察

本章では、第4章で述べた授業実践の際に、授業前後で実施したアンケートと生徒の教材操作を記録した操作ログを分析することで、学習効果を評価する。

5.1 評価：アンケート比較

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4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4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

表4 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		

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。

5.2 評価：操作ログの分析

授業後のアンケートにおいて、自由記述形式で授業の感想を収集した。収集した感想をポジティブな意見を含むもの（グループ1）とそうでないもの（グループ2）に分け、それぞれのグループに属する生徒の操作ログを比較した。ここで、ポジティブな意見とは、「面白い」「すごい」「驚き」「自分で」「もっとしたい」「-しやすい」の5つの語句を含む意見とする。本稿では、本教材の3つの主要な機能である単回帰分析、重回帰分析、可視化の機能が使用された回数を生徒ごとに集計し、各グループの平均値を比較した。その結果を表5に示す。

表5 ポジティブな感想を持つユーザー群とポジティブな感想を持たないユーザー群の比較

機能	グループ	平均	標準偏差	p 値
単回帰分析	ポジティブな感想	14.8	10.7	0.760
	ポジティブではない感想	13.9	11.1	
重回帰分析	ポジティブな感想	13.0	16.7	0.175
	ポジティブではない感想	16.7	13.1	
可視化	ポジティブな感想	22.6	17.3	0.048
	ポジティブではない感想	13.7	18.2	

表5において、ポジティブな感想を残したグループとそうでないグループの2つのグループを比較すると、可視化機能の使用回数で有意な差が見られた。第4章に述べた授業実践において、可視化機能は使いたい人のみが使うような促しをしており、強制はしていない。しかし、表5に示す差が確認できたことより、ポジティブな感想を持った生徒は、それ以外の生徒と比較し、教材の可視化機能を多用していたといえる。これより、本教材における可視化機能のような、教材の主目的以外の任意の機能が生徒の探究心を刺激した可能性が考えられる。

6 まとめ・今後の課題

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、授業前後でのアンケートと操作ログを解析した。この結果、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。また、一部の生徒は、単回帰分析の精度向上のためにはデータの適切なフィルタリングや重回帰分析の活用等の重要性に気づ

くことができ、回帰分析ひいてはデータ分析に対して意識すべきマインドを体得したと考えられる。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、今回の学びが活けると期待できる。

今後の課題としては、重回帰分析のモデル選択をはじめとする発展的な内容への対応も検討したい。高等学校情報科の学習指導要領解説[17]に記載されている学習内容のうち、共通必修科目である情報Iの「情報通信ネットワークとデータの活用」においては、データの収集整理に適切な方法を選択し、実行・評価・改善する能力を身につけることが期待されている。本教材で次に求められることは、実行結果を評価し、改善するプロセスを養うことである。本授業実践における重回帰分析の体験の中で、体力測定データのほぼすべての変数を説明変数に設定し重回帰分析を実行する生徒が散見された。重回帰分析を活用する場面では、ある程度汎用的なモデルを目指すべきであり、多重共線性などにも気をつけなければならない。そのような観点からの機能は現時点での本教材ではまだ実装されていない。そこで、本教材を用いた学習後の次のステップとして、正しいモデル選択の学習を支援できる Web アプリケーションの開発を進めたい。

謝 辞

本教材を用いた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。

文 献

- [1] 首相官邸ホームページ, “AI 戦略 2019 人・産業・地域・政府全てに AI”, <https://www.kantei.go.jp/jp/singi/ai-senryaku/pdf/aistratagy2019.pdf>, 2019.
- [2] 文部科学省, “高等学校学習指導要領解説 数学編”, pp. 11-18, 2018.
- [3] 科学の道工具箱 データライブラリ, <https://rika-net.com/contents/cp0530/contents/07.html#>.
- [4] Bowland Japan, “ポーランド・ジャパンの教材紹介”, https://bowlandjapan.org/materials_jp.
- [5] 文部科学省, “高等学校情報科「情報I」教員研修用教材(本編)”, 第4章 情報通信ネットワークとデータの活用, pp. 184-191, 2019.
- [6] 文部科学省, “高等学校情報科「情報II」教員研修用教材(本編)”, 第3章 情報とデータサイエンス(前半), pp. 124-143, 2019.
- [7] 神部, 玉田, “プログラミング活用による統計教育への問題解決型教材の開発に向けて”, *Informatio: 江戸川大学の情報教育と環境*, pp. 29-32, 2020.
- [8] 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”, *兵庫教育大学 研究紀要*, 第57巻, pp. 65-75, 2020.
- [9] Najjar L.J., “Principles of educational multimedia user interface design”, *Human Factors*, vol. 40, no. 2, pp. 311-323, 1998.
- [10] Juan C. Castro-Alonso, Bjorn B. de Koning, Logan Fiorella, and Fred Paas, “Five strategies for optimizing instructional

materials: Instructor-and learner-managed cognitive load.”, Educational Psychology Review, vol. 33, no. 4, pp. 1379-1407, 2021.

- [11] 藤田, “学校教育におけるデータの分析教育: イギリスの視点から”, 科学教育研究, Vol. 15, No. 2, p. 73, 2020.
- [12] Noel Entwistle, Velda McCune, and Paul Walker, ”Conceptions, styles, and approaches within higher education: Analytic abstractions and everyday experience.”, In Perspectives on thinking, learning, and cognitive styles, pp. 103-136, 2014.
- [13] Grant Wiggins, and Jay McTighe, ”Understanding by design”, Ascd, 2004.
- [14] 文部科学省, ”学習の基盤となる資質・能力としての体系表例とカリキュラム・マネジメントモデルの活用情報活用能力の育成”, 2020.
- [15] 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>.
- [16] 文部科学省, ”新体力テスト実施要項 (12 歳~19 歳対象) ”,
- [17] 文部科学省, “高等学校学習指導要領解説 情報編”, pp. 35-40, 2018.