

データをインタラクティブに操作できる データサイエンス教育用ウェブ補助教材の開発

三浦 元喜^{1,a)} 中田 豊久^{2,b)}

概要：近年高等学校や大学等において、データサイエンス教育が盛んに行われている。学習者の主体的な学習を支援するため、これまでオンライン教材や演習環境、Web アプリケーションが開発されてきたが、データを修正して結果を確認するには分析操作をやり直す必要があった。我々は学習者がマウス操作でデータを修正すると、その結果を即座に反映するインタラクティブなウェブ補助教材を Processing 言語を用いて開発した。操作による影響や結果がリアルタイムに視覚化されるため、学習者の興味を惹きつけるとともに、視覚化や分析手法の概要を短時間で理解したり、試行錯誤を促進したりする効果が期待できる。データサイエンス系の大学の講義で実践を行い、学生のコメントを収集することで、インタラクティブなウェブ教材を設計するうえでの知見を得ることができた。

キーワード：Web 教材、なめらかなインタフェース、箱ひげ図、正規分布、偏相関、重回帰

Development of Web Applications for Data Science Education that can Manipulate Data Interactively

MOTOKI MIURA^{1,a)} TOYOHISA NAKADA^{2,b)}

1. はじめに

近年高等学校や大学等において、データサイエンス教育が盛んに行われている。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 [1] では、Excel を活用した単回帰分析が紹介されている。また大学においても、近年データサイエンスを学べることを特色とした学部や学科、研究科が多数創設されている。これらのことから、機械学習や AI に関する技術や知識を身につけることに関する社会の期待は高いといえる。

横山らは Jupyter Notebook を複数人で利用できる環境

を教育利用にカスタマイズした CoursewareHub を利用して、オンライン授業でのグループワークを可能にする「グループ作業場所メディア」を実装している [2]。Python が動作する汎用のプログラミング環境であるため、プログラムを入力すれば様々なデータ分析が可能である。しかしデータを選択したり修正したりするにはプログラムの修正が必要となる。

村上らは高校生および大学学部生を対象とした回帰分析初学者のためのインタラクティブな Web アプリケーション教材を開発し、授業実践を行っている [3]。説明変数やデータ可視化手法を選択し、手軽に分析を試せる工夫がされているが、データを修正する方法は提供されていない。岸本らは、高等学校の情報 II における重回帰分析までの学習を可能にする学習教材を提案している [4]。村上らの教材と同様に、変数を選択して散布図行列を表示する機能や、重回帰分析を行った結果の式を確認したり、重回帰分析の散布図を表示したりすることができ、高校生が具体的

¹ 千葉工業大学 工学部 情報通信システム工学科
Chiba Institute of Technology, Narashino, Chiba 275-0016, Japan

² 新潟国際情報大学 経営情報学部 情報システム学科
Niigata University of International and Information Studies,
Niigata 950-2292, Japan

a) motoki.miura@p.chibakoudai.jp

b) nakada@nuis.ac.jp

なデータをもとに分析作業を行いやすくする工夫がされている。我々は重回帰分析を直接扱う教材については今回作成していないが、2つの説明変数と1つの目的変数のあいだの相互の相関と偏相関の関係をインタラクティブに学ぶ教材を作成している。才田らは自然科学分野の観測データを含むオープンデータの可視化と簡易な解析が可能な Web アプリケーションを開発している [5]。データの可視化と確認のためにマウスオーバーによるデータ表示やズーム、パンなどのインタラクティブなグラフ機能を提供している。ただし、データの修正に関してはプロット処理短縮のためのデータの間引きを行う機能や欠損の確認を行う機能はあるが、データを直接編集して結果をインタラクティブに確認することはできない。

統計教育推進委員会が公開している統計活用授業のための Web サイト [6] では、箱ひげ図や散布図、正規分布曲線を表示する Excel シートをダウンロードすることができる。この Excel シートの数値を変更することで、図やグラフとしての表現を確認することができる。前田は、機械学習に関して動かしながら仕組みを学べる表計算ソフトを用いた学習教材を提案している [7]。モデルの学習過程をブラックボックス化せず、学習者が確認できる工夫がされている。これらの学習教材では表計算ソフトを用いるため、データの数値はキーボードで入力して変更する必要がある。

我々は外れ値や極端な例を含めた多様なデータについて試行を繰り返し、結果や振る舞いを確認することが、各種手法の特徴を理解するのに役立つと考えた。そのため、Excel よりもより簡潔な操作で学習者がデータを修正でき、データ修正結果や振る舞いを連続的に観測できるインタラクティブなウェブ補助教材を Processing 言語を用いて開発した。学習者が修正したいデータをマウスで選択しドラッグすると、その結果が即座に反映される。またデータをランダムに生成することで多様な場合について確認できる。これらの補助教材を用いたデータサイエンス系の講義において、受講者に演習の一環としてこれらの教材によって理解できることや学べること、勘違いしやすい点や改善点を自由記述形式で回答してもらった。

2. 開発したインタラクティブウェブ補助教材

ここでは、我々が開発したデータサイエンス教育用インタラクティブウェブ補助教材の設計方針と、作成した補助教材の詳細について述べる。

2.1 設計方針

上で述べたように、我々は外れ値や極端な例を含めた多角的なデータについて試行を繰り返し、結果や振る舞いを確認することが重要と考えている。この目的を達成するため、補助教材を作成するにあたっては以下の設計方針を挙げた。

- (1) **多様なデータ生成+直接修正** データをランダムに生成できることで、多様で多角的なデータについて試行できること。また生成したデータを容易に修正できること。修正は選択的に行えること。
 - (2) **連続的な修正操作と結果の観測** 試行結果を即座に確認できること。連続的なデータ修正操作に対しては、結果や振る舞いも連続的に観測できるようにすること。なるべくグラフや図などを用いて視覚的に観測できるようにすること。
 - (3) **試したくなる工夫** データの操作性を高めたうえで、見栄えを高めたり、操作に対応した効果音をつけることで、学習者の興味を喚起する。
 - (4) **わかりやすさ** なるべくシンプルにし、説明がなくても操作できるようにする。データの修正・削除操作については統一的なインタフェースを提供する。
- なお今回作成したのは補助教材であるため、ひとつのアプリケーションで手法の詳細をすべて完結的に理解できることを目的としていない。教科書や他の教材と併用してもらい、相互に補完しあうことを想定している。

2.2 補助教材

前述した設計方針に基づき、我々はこれまでに (1) 箱ひげ図、(2) 正規分布、(3) 単回帰分析、(4) ロジスティック回帰分析、(5) k-近傍法、(6) 偏相関 (重回帰分析) を扱う補助教材を開発した。以下では、それぞれの補助教材アプリについて説明する。

2.2.1 箱ひげ図アプリ

箱ひげ図アプリは、箱ひげ図の描画方法と、それに関連する四分位数の意味を理解するための補助教材である。図 1 に箱ひげ図アプリの画面を示す。学習者はマウスドラッグでデータを表す点を左右に移動したり、左クリックで点を追加したり、右クリックで点を削除したりすることができる。点を修正すると、箱ひげ図と四分位数の表示が即座に更新される。これらの操作により、学習者は中央値を含む四分位数や四分位範囲の意味を確認することができる。また箱ひげ図における箱やひげの長さ、外れ値を表す丸の意味を体験的に理解することができる。点をランダムに生成する方法は複数用意しており、[r] キーによるランダム生成のほか、数字キーによる点の数を指定したランダム生成ができる。またランダム生成キーを押す際のマウスカーソルの X 座標と Y 座標はそれぞれ生成される乱数の平均と分散に反映されるようにしており、学習者が多様なランダムデータを簡潔な操作で作成できるよう工夫している。点にマウスホバーすると値 (ここでは単純に点の X 座標) を表示する。[c] キーを押すと初期状態に復帰する。図 1 のキャプションに、箱ひげ図アプリの URL を示す。

2.2.2 正規分布アプリ

正規分布アプリは、データを「ある母集団から抽出され

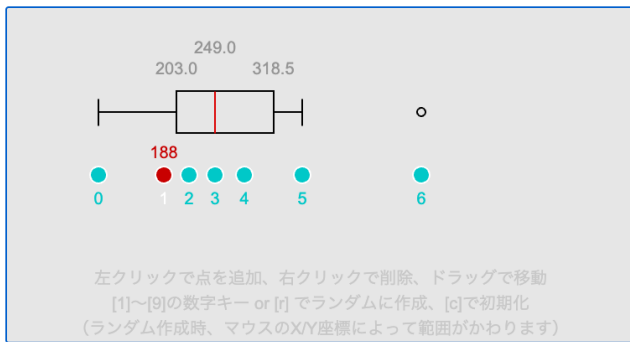


図 1 箱ひげ図アプリ <https://pjs.istlab.info/pjs/run/551>

たサンプル」とみなすとき、その母集団の正規分布を推測し、グラフとして表示する補助教材である。図 2 に正規分布アプリの画面を示す。箱ひげ図アプリと同様に、[r] キーを押すか、数字キー (2~9) を押すとランダムなデータ (点群) を生成する。画面下の点を追加・移動・削除することによって、正規分布を示すグラフが更新される。またマウスカーソルを $y > 0$ の領域に持っていくと、 $-\infty$ からマウスカーソル位置までの部分を色付ける。これにより、上側・下側確率の意味や、パーセント点の意識をもたせることをねらいとしている。

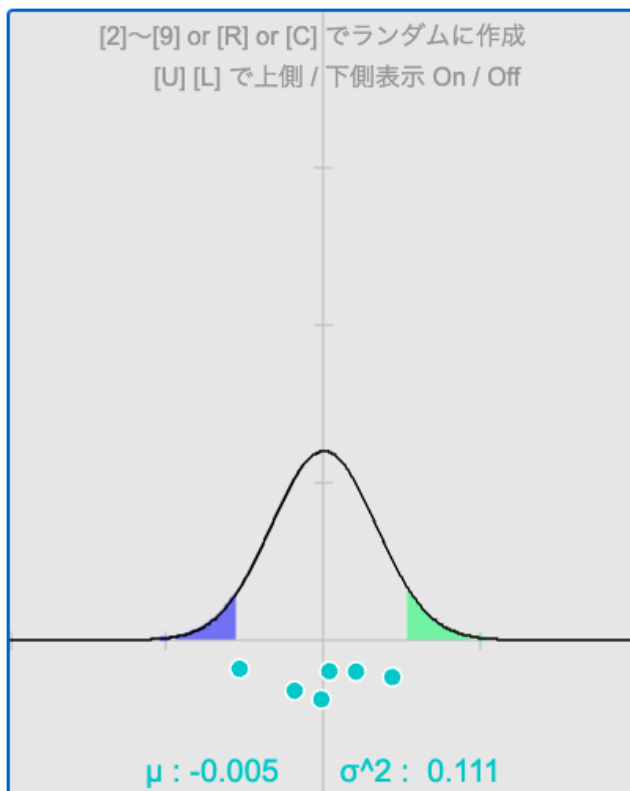


図 2 正規分布アプリ <https://pjs.istlab.info/pjs/run/554>

2.2.3 単回帰アプリ

単回帰アプリは、点で示されたデータの相関係数や決定係数を表示するとともに、線形単回帰分析の結果を式のパラメータ数値とグラフで示す補助教材である。図 3 に単回

帰アプリ画面を示す。箱ひげ図、正規分布アプリと同様、学習者はマウスドラッグで点を移動したり、左クリックで点を追加したり、右クリックで点を削除することができる。こちらも点を追加・移動・削除することによって、回帰直線の描画と数値が逐次更新される。そのため、学習者は回帰直線の意味や特性、相関係数、決定係数について体験的に理解することができる。

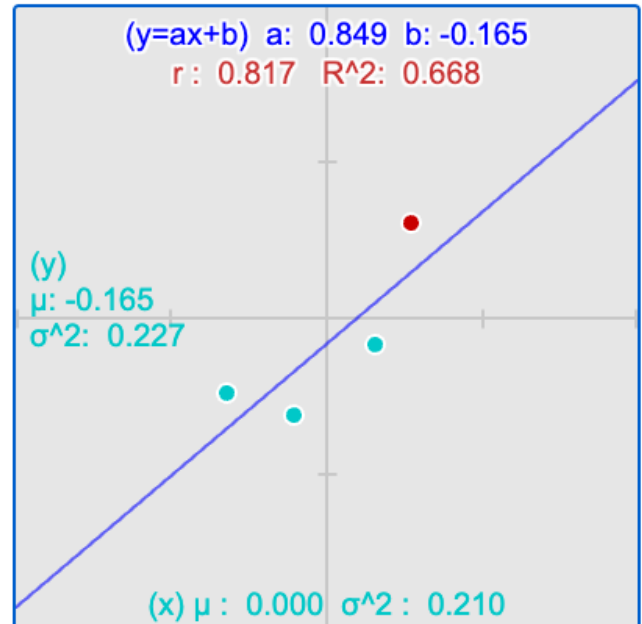


図 3 単回帰アプリ <https://pjs.istlab.info/pjs/run/555>

2.2.4 ロジスティック回帰アプリ

ロジスティック回帰アプリ (図 4) は、 y 座標が 0 または 1 のどちらかをとるデータ (点群) が与えられたときに、データにもっともあてはまるロジスティック関数のグラフと、その式を表示する補助教材である。学習者は前述のアプリと同様の操作で点の移動・追加・削除を行うことができる。学習者は点を操作することで、確率の推測値を示すロジスティック関数の振る舞いがどのように変化するかを知ることができる。

2.2.5 k-近傍法アプリ

k-近傍法アプリ (図 5) は、k-近傍法 (k-nearest neighbor method) の動作を理解するための補助教材である。初期状態では 2 色の点群 (色はクラスを示す) がランダムに配置されている。学習者がマウスで左クリックし、候補点を追加すると、設定されたパラメータ k (初期状態は 3) に従い、候補点に近い k 個の点を探索し、線分で表示する。また多数決によって推測された候補点のクラス候補が、点の外周の色で示される。この状態で、候補点およびデータを表す点はいずれも移動・削除ができる。Enter キーをおすと候補点のクラスが確定し、線分が削除される。また奇数の数字キーを押すと k を設定できる。学習者は候補点やデータ点を追加したり、移動したり、 k を変更したりしな

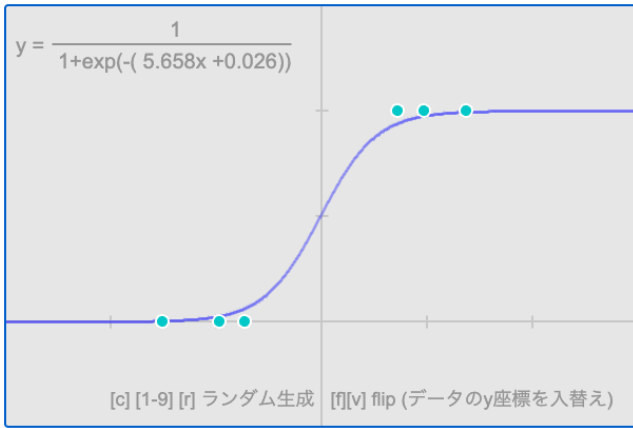


図 4 ロジスティック回帰アプリ
<https://pjs.istlab.info/pjs/run/559>

から k 近傍法の振る舞いを体験的に学ぶことができる。なお実際のアルゴリズムでクラスを推測する場合にはデータ点を移動することはないが、今回はアプリとしての自由度を高めるために機能を実装している。

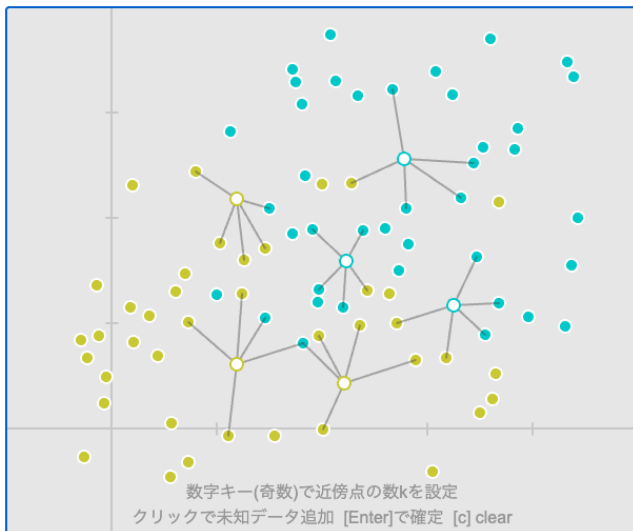


図 5 k-近傍法アプリ <https://pjs.istlab.info/pjs/run/560>

2.2.6 偏相関アプリ

偏相関アプリは、重回帰分析の場面において変数同士の相関係数と標準偏差が、偏相関係数や偏回帰係数にどのように影響しているかを確認することができる補助教材である。図 6 に偏相関アプリ画面を示す。学習者は数値（相関係数 r および標準偏差 σ ）の右のボタンを押したり、数値上を左右方向にマウスドラッグしたりすることによって数値を変更できる。変更した数値は、以下に示す式によって、偏相関係数（相）と偏回帰係数（傾）に反映される。

$a-y, b-y, a-b$ の相関係数がそれぞれ r_{ay}, r_{by}, r_{ab} であるとき、 b と y の偏相関係数は

$$\frac{r_{by} - (r_{ay} \cdot r_{ab})}{\sqrt{1 - r_{ay}^2} \cdot \sqrt{1 - r_{ab}^2}}$$

となる。

また、 b と y の偏回帰係数（偏回帰直線の傾き）は

$$\frac{r_{by} - (r_{ay} \cdot r_{ab})}{1 - r_{ab}^2} \cdot \frac{\sigma_y}{\sigma_b}$$

となる。

図 6 に示すように、説明変数間の相関が高い状態に設定することで、多重共線性が生じているときの重回帰分析の傾向や状況（偏回帰係数の絶対値が比較的大きくなる）についても観測できる。なお現状の偏相関アプリのみ他の補助教材と異なりデータを点群で扱っておらず、操作性が大きく異なる。今後は点群による統一的な操作を偏相関アプリに適用することも検討している。

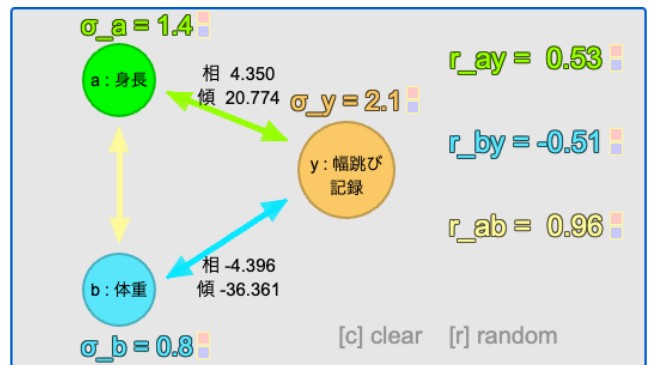


図 6 偏相関アプリ <https://pjs.istlab.info/pjs/run/558>

3. 授業実践

2023 年度前期に開講された、第一著者が担当するデータサイエンス系の学部講義において、開発したインタラクティブ Web 教材を補助教材として使用した。

2023 年 5 月 22 日の中間考査時に、以下の 5 項目 (Q1-1, Q1-2, Q2-1, Q2-2, Q3) のアンケートを 4 点分の演習として課し、自由記述で回答してもらった。期限は 1 週間後に設定した。学生は中間考査で余った時間に回答したり、家に帰ってから回答したりした。アンケートでは、箱ひげ図アプリと正規分布アプリの 2 つのみについて質問した。中間考査時点では単回帰アプリも講義で取り扱っていたが、正規分布アプリと重複する部分が多いと考え今回は除外した。ロジスティック回帰アプリ以降については中間考査の時点ではまだ講義で取り扱っていなかった。

受講生は 17 名で、うち 12 名が大学 3 年生、5 名が大学 4 年生であった。

- **理解できること (Q1-1, Q2-1)**：箱ひげ図/正規分布教材の操作や実演によって理解できることを、できるだけ多く書いてください。箇条書きでも結構です。
- **勘違いしやすい点 (Q1-2, Q2-2)**：箱ひげ図/正規分布教材を使用するうえで、勘違いしやすい点や、間違えて理解してしまいそうな点を書いてください。(アプリケーション教材以外と対比して書いてもよいです)
- **改善点 (Q3)**：(2 つの教材を通じて) それぞれの教

- 箱ひげ図からデータが外れた時のデータの存在について分かりづらい
- 一番最初の数値ナンバが0であるため、標本数を勘違いしやすい

(Q2-2) 正規分布アプリで勘違いしやすい点

- 使用方法を理解するまでは何を表すものであるかがわかりにくい
- 初学者が利用すると仮定するなら、横軸と縦軸が何を表しているのか示した方がいいと感じた。
- 青色と緑色で表されている内容がわかりにくい
- 左右対称でないものも正規分布であると考えてしまう。
- 平均値と中央値と最頻値のイメージがつきにくい。

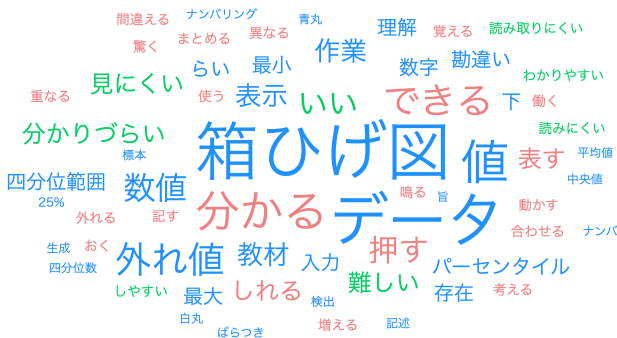


図 10 Q2-1「箱ひげ図教材で勘違いしやすい点」への回答文から生成したワードクラウド (頻度)

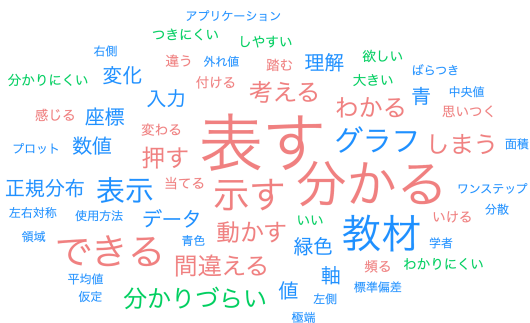


図 11 Q2-2「正規分布教材で勘違いしやすい点」への回答文から生成したワードクラウド (頻度)

箱ひげ図教材については、点のナンバリングと実際の数値との対応に関する懸念やわかりにくさに関する記述があった。正規分布教材については、座標や数値、軸の説明があったほうがよいという意見や、青色や緑色で塗っている意味がわかりにくい点が挙げられていた。

さいごに、改善点 (Q3) への回答文を、ULTMT で 5 行に要約した結果を示す。また図 12 に、回答文から生成したワードクラウドを示す。

(Q3) 改善点

- 教材と共に使用方法や動かす際のポイントを記すとより効果的に使用できると感じた
- もう少し文字の説明があったり、その説明が見やすかったりすると分かりやすい教材になると思いました。
- 数値や座標をキーボードで入力できたら具体案で考えやすくなるのではないかと思います。
- 確率密度関数をホバーしたときに積分した値が表示されると、確率密度関数に対する理解が深まると思います。
- 箱ひげ図の教材の下にパーセンタイル値の求め方の式があるとありがたいと思います。

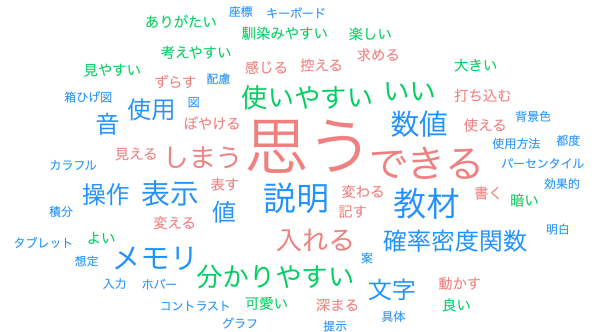


図 12 Q3「改善点」への回答文から生成したワードクラウド (頻度)

表示・入力方法や、操作方法の説明が不足していることに関しての言及があった。またパーセンタイルの求め方を具体的に示すことや正規分布教材で積分値を表示してほしいという要望があった。今回はシンプルさを重視してなるべく文字を減らしたため、軸の説明や目盛を省略していたが、具体的な数値を入力・確認したいといった要望もあるので、今後検討したい。

4. 議論

ここでは提案するインタラクティブウェブ教材の設計思想と、利活用にあたって注意すべき点について述べる。

本稿で紹介したインタラクティブウェブ教材では、基本的にマウスによってデータを直接修正することを許容しつつ、その修正結果を逐次反映される画面描画 (グラフや図などの視覚要素および計算結果としての数値) によって確認できるようにする、という基本思想に基づいて設計している。これによって、学習者はデータを気軽に修正し、計算結果を確認することができる。しかしながら、この「データを修正し、反映された結果を確認する」という手続きにおいて、以下の四つの注意すべき点があると考えている。

(1) 因果関係に関する誤解を招きやすい 今回示した正規分布補助教材では、画面下部の標本データを表す「点群」を操作することで、画面上部の確率密度関数のグラフが変化する。しかし、この補助教材では標本データが正規

分布に従うことと、標本データから推測した母平均・母分散が正しいことを前提として確率密度関数を定義し、そのグラフを表示しているに過ぎない。これと類似した話として、信頼区間の解釈において「標本から推定された信頼区間の中に95%の確率で母平均が入る」という誤解が多いことが挙げられる。推測統計学では「母集団の特性は変化しないもの」として扱われる。そのため正規分布補助教材に関しては「母集団の確率分布が、画面にプロットされた確率密度関数（のグラフ）でモデル化できるとき、母集団から抽出した標本のひとつの例を点群で示している」という解釈を妨げる可能性があるため注意が必要である。こうした誤解を避けるためには、確率密度関数のグラフをマウスで操作し確定操作をしたらいくつかの標本例を表示することや、母平均・母分散入力モードと、標本生成モードを分けるといった工夫が考えられる。

(2) **データの取り扱いに関する意識の低下** 今回の講義では示さなかったが、t検定をはじめとする統計的検定を視覚的に表現するアプリケーションも考えられる。統計的検定においては、実験データを修正することはない。そのため、データを修正することで結果をインタラクティブに確認するといった活動に慣れることで、データの取り扱いに関する意識が低下するおそれがある。

(3) **視覚表現に対する誤認識** アプリの設計段階で意図していない些細な表現に対して、学習者がその意味を過度に推測したり、表現に誤った意味づけをしてしまうことが考えられる。例えば、正規分布アプリでは画面下部のデータを表す「点群」の重なりを抑制し操作しやすくするため、y座標に自由度をもたせているが、学習者はy座標の違いになにか意味があるのか勘繰ってしまうことが予想される。

(4) **原理の軽視、設計意図の不理解** 動作として確認することで、理解していないことも理解したつもりになってしまうおそれがある。たとえば、正規分布アプリでマウスをグラフ上に動かしたときに塗りつぶし表示される「領域」について、その意味や意図が理解されないケースが考えられる。これに関しては、意図を正しく伝える教材の工夫（たとえば領域の面積を表示することや、説明を追加すること）が必要であると考えられる。

5. まとめと今後の課題

外れ値や極端な例を含めた多様なデータについて試行を繰り返し、結果や振る舞いを確認する活動を促すため、簡潔な操作で学習者がデータを修正でき、結果や振る舞いを連続的に観測できるインタラクティブなウェブ補助教材をProcessing言語を用いて開発した。データに対する操作の自由度を高めるため、ランダムなデータを生成することに加えて、マウスでデータを選択的に修正できるようにした。

これにより、学習者は自らの興味に基づき、様々なデータがどのように結果に影響するのかを容易に調べることができ、一部の教材についてはデータサイエンス系の講義において試用し、受講者がこれらの教材をどのように理解していたかや改善点を自由記述形式のアンケートによって収集し知見を得ることができた。

本来のデータ分析において、取り扱うデータを気軽に修正する行為は一般的にはタブーとされている。しかし、タブーであることを認識し注意を喚起したうえで、視覚化手法や分析手法を理解したり仕組みを把握したりする場面においてデータを修正して観測するというアプローチは一定の教育的効果が見込めるため今後も継続的に議論・検討すべきと考えている。我々が提案する体験型の補助教材およびアプローチは、従来の数学的な計算手法の表現・理解や、実際のデータを処理するプログラミングスキルに置き換わるものではなく、相互補完的に利用することで学習効果を高めたり理解を深めたりすることをねらっている。これらの検討および成果物が、今後のデータサイエンス教育の拡充に貢献できれば幸いである。

謝辞

アンケートに回答してくれた受講生のみなさんに感謝いたします。本研究の一部はJSPS 科研費 JP22K12319 の支援によるものです。

参考文献

- [1] 文部科学省：高等学校情報科「情報 I」教員研修用教材（本編），https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/1416756.htm（2023年6月14日確認）（2020）。
- [2] 横山重俊，浜元信州，桑田喜隆，長久 勝，中川晋吾，政谷好伸，竹房あつ子，合田憲人：オンライン授業における Jupyter Notebook を活用した演習環境提供，情報処理学会研究報告 教育学習支援情報システム（CLE），Vol. 2020-CLE-31，No. 2，pp. 1-4（2020）。
- [3] 村上綾菜，伊藤貴之：帰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，第 15 回データ工学と情報マネジメントに関するフォーラム（DEIM Forum），日本デジタル教科書学会（2023）。
- [4] 岸本有生，本多佑希，漆原宏丞，兼宗 進：高等学校における重回帰分析までの学習を可能にする学習教材の提案，情報処理学会研究報告 コンピュータと教育（CE），Vol. 2023-CE-170，No. 6，pp. 1-5（2023）。
- [5] 才田聡子，柏田元輝，外村慶明，尾花由紀，北村健太郎，古賀崇了：データサイエンス教育の題材としてのオープンデータ可視化 Web アプリケーションの開発，情報教育シンポジウム論文集，Vol. 2019，pp. 17-23（2019）。
- [6] 統計教育推進委員会：データで学ぶ統計活用授業のための教材サイト，<https://estat.sci.kagoshima-u.ac.jp/data/>（2023年6月9日確認）。
- [7] 前田篤彦：表計算ソフトを用いた機械学習教材に関する一検討，情報処理学会研究報告 コンピュータと教育（CE），Vol. 2023-CE-168，No. 22，pp. 1-8（2023）。