

高等学校における  
データサイエンス教育教材の研究

岸本 有生

大阪電気通信大学大学院 博士後期課程  
工学研究科 制御機械工学専攻 DM18B001



# 目次

|              |  |           |
|--------------|--|-----------|
| <b>第 1 章</b> | <b>序論</b>  | <b>5</b>  |
| 1.1          | 研究目的と概要 . . . . .  | 5         |
| 1.2          | 本論文の構成 . . . . .   | 6         |
| <b>第 2 章</b> | <b>学校教育で利用可能なデータ分析ツールに求められる要件</b>                              | <b>7</b>  |
| 2.1          | 高等学校におけるデータサイエンス教育の概要 . . . . .                                | 7         |
| 2.2          | データサイエンス教育に関する先行研究 . . . . .                                   | 11        |
| 2.3          | 本論文の位置付け . . . . .   | 16        |
| <b>第 3 章</b> | <b>データ分析ツール Connect DB の設計と実装</b>                              | <b>18</b> |
| 3.1          | Connect DB の概要 . . . . .                                       | 18        |
| 3.2          | さまざまな環境で動作する Web アプリケーションの設計と実装 . . . . .                      | 26        |
| 3.3          | データベースの設計と実装 . . . . .   | 28        |
| 3.4          | データの分析機能 . . . . .   | 30        |
| 3.5          | データの収集機能 . . . . .   | 32        |
| 3.6          | 操作ログの保存 . . . . .  | 35        |
| 3.7          | 情報の授業との対応 . . . . .  | 36        |
| 3.8          | 情報の教員の感想 . . . . .   | 38        |
| <b>第 4 章</b> | <b>Connect DB: An Online Learning System for Data Analysis</b> | <b>40</b> |
| 4.1          | Introduction . . . . .   | 40        |
| 4.2          | Some problems for using data analysis in class . . . . .       | 40        |
| 4.3          | Design of data analysis learning system . . . . .              | 41        |
| 4.4          | Evaluation of Connect DB . . . . .                             | 42        |
| 4.5          | Conclusion . . . . .   | 49        |
| <b>第 5 章</b> | <b>高等学校において重回帰分析を扱うデータ分析学習ツールと授業の提案</b>                        | <b>50</b> |
| 5.1          | はじめに . . . . .   | 50        |
| 5.2          | 共通教科「情報」におけるデータ活用の学習 . . . . .                                 | 50        |
| 5.3          | 重回帰分析を学習する上での課題 . . . . .                                      | 51        |
| 5.4          | Connect DB への追加機能の設計と開発 . . . . .                              | 52        |

---

|              |  |           |
|--------------|--|-----------|
| 5.5          | 重回帰分析を扱う授業案の検討 . . . . .               | 57        |
| 5.6          | 大学生を対象とした授業実践 . . . . .                | 58        |
| 5.7          | 考察 . . . . .                           | 64        |
| 5.8          | おわりに . . . . .                         | 65        |
| <b>第 6 章</b> | <b>スマートフォンのセンサを活用した計測データ分析教材の提案と実践</b> | <b>66</b> |
| 6.1          | はじめに . . . . .                         | 66        |
| 6.2          | 工業科におけるデータ分析 . . . . .                 | 66        |
| 6.3          | Connect DB への追加機能の設計と開発 . . . . .      | 67        |
| 6.4          | スマートフォンを活用したデータ分析の授業実践 . . . . .       | 68        |
| 6.5          | ワークシートの記述内容の分析 . . . . .               | 70        |
| 6.6          | 考察 . . . . .                           | 73        |
| 6.7          | まとめ . . . . .                          | 75        |
| <b>第 7 章</b> | <b>結論</b>                              | <b>76</b> |
|              | <b>謝辞</b>                              | <b>78</b> |
|              | <b>参考文献</b>                            | <b>79</b> |

# 第1章

## 序論

### 1.1 研究目的と概要

近年、情報技術の進化によりサイバー空間 (仮想空間) とフィジカル空間 (現実空間) が融合し、Society 5.0 と呼ばれる新たな社会が現実のものとなりつつある [1]。この社会では、ビッグデータを基にした AI(人工知能) やロボットが人間の作業を支援し、誰もが快適で充実した社会生活が実現されることを目指している。そのためには、データを活用する能力を持ったデータサイエンティストの育成が求められている。

2020 年度から小学校から大学までの教育にデータサイエンスの授業が段階的に導入され、その注目が高まっている。特に高等学校の共通教科「情報」では、カリキュラム改編に伴い「社会と情報」「情報の科学」の2科目が「情報 I」に統合されるとともに、新たに「情報 II」が新設され、データサイエンス教育が本格的に導入された [2]。具体的には、必修科目である情報 I では「(4) 情報通信ネットワークとデータ活用」に、選択科目である情報 II では「(3) 情報とデータサイエンス」に、データサイエンスの内容が組み込まれている。また、大学では数理・AI・データサイエンスのリテラシーレベル、応用レベル、エキスパートレベルの学習項目が提示されている [3]。

このような教育課程の変更に伴い、情報 I は 2025 年度の入学生を対象に大学共通テストの出題科目として採用されることが決定した。大学共通テストのサンプル問題では、サッカーの強いチームと弱いチームの違いを散布図行列を利用して分析し、その特徴を述べる例が出題されている [4]。このような問題は、生徒にとって身近な現象をデータ分析し、問題解決の能力を養うことを目指している。

一方、高等学校の情報科ではこれまでデータサイエンス教育が行われておらず、初めて授業の中でデータサイエンスの内容を扱う教員が多い。そのため、文部科学省は高等学校の情報の教員研修用資料を公開した。その中でデータサイエンスについては「簡単なデータ操作ではなく、大量のデータをプログラミングで処理、分析する」と定義され、表計算ソフト、R、Python、Tableau を活用したデータサイエンス教育の具体的な事例が紹介されている [5] [6] [7]。これまでの「社会と情報」では表計算ソフトが扱われていたことが多かったため、出版された情報 I の教科書でも表計算ソフトが取り上げられていることが多い。情報 II の教科書では、データ分析に加えて機械学習も扱うため、プログラミング言語である R や Python の事例が紹介されている。このように、複数の分析ツールを組み合わせた授業展開が必要となるが、高等学校においては PC の操作にまだ慣れていない生徒も多く、限られた授業時間の中で能力を養うことは容易ではないという課題が存在する。

また、データサイエンス教育は国外でも進んでおり、国際的な計算機学会である ACM のカリキュラムやアメリカのデータサイエンスのカリキュラムでは、授業で利用しやすい分析ツールや、生徒にとって身近なデータセットの活用が重要であることが強調されている [8] [9]。こうした国際的な傾向を踏まえると、データサイエンス教育を成功させるには、生徒が現実の問題に対して効果的なデータ分析を行えるようなツールが必要であると考えた。

そこで本論文では、高等学校の情報科や大学の授業において利用できるデータ分析ツール「Connect DB」を提案し、設計と実装を紹介した後に、高等学校と大学で実施した授業での教育について報告する。

## 1.2 本論文の構成

本論文は、高等学校におけるデータサイエンス教育の現状と課題を踏まえ、高等学校の授業に適した学習教材の検討と、提案する教材の学習効果を実践を通して明らかにする。

2章では、高等学校の共通教科「情報」における、情報 I の「(4) 情報通信ネットワークとデータ活用」と、情報 II の「(3) 情報とデータサイエンス」の学習について議論する。これまでに利用されてきた分析ツールと授業利用における課題について整理し、これらを解決するために求められている分析ツールの機能について検討する。

3章では、2章で明らかになった課題に対処するために開発されたデータ分析学習システム「Connect DB」を報告する。この章では Connect DB のシステム的な機能に加え、技術的な工夫点についても説明する。

4章では、Connect DB の授業利用について検討する。特に、分析ツールの操作方法が難しいことが授業内容を難しくしている点について調査を行った。まず、授業で多く利用されている分析ツールのひとつである表計算ソフトとの操作回数の比較を行い、大学1年生が利用した際のアンケート結果から使いやすさを分析した。その結果、本システムを利用することで、初心者がツールを利用する際の認知的な負荷を軽減できる可能性が示唆された。なお、本章は論文投稿時の内容を残す意味から、英語での記述とした。

5章では、Connect DB に実装された機能が情報 II の重回帰分析に適用できるかを検証している。ここでは、大学2年生を対象に行った調査から、散布図行列を利用して目的変数と説明変数の関係を理解し、モデルを最適化するために重要な説明変数を選択する方法を理解していることが示唆された。

6章では、Connect DB が授業準備の難しさを克服できるかに焦点を当て、センサを利用したデータ計測に注目した。従来ではセンサやマイコンの購入などの負担が大きかった。そこで、スマートフォンの内蔵センサ (加速度センサ) を活用できるシステムを構築することで、センサやマイコンの購入を必要とせず、データ計測の授業が行えることを確認した。

## 第2章

# 学校教育で利用可能なデータ分析ツールに求められる要件

### 2.1 高等学校におけるデータサイエンス教育の概要

平成29年および30年の告示において、学習指導要領の方向性が中央教育審議会答申に基づいて示された[10]。社会生活などの様々な場面において必要なデータを収集して分析し、その傾向を踏まえて課題を解決したり、意思決定をしたりすることが求められている。このような能力を育成するため、高等学校情報科などとの関連も図りつつ、小・中・高等学校教育を通じて統計的な内容等の改善が検討されていることが記述されている。

小・中学校の算数、数学では、簡単なグラフの読み方や標本から母集団を推定する統計の基礎を学習する。高等学校では、数学で統計の理論を学習し、情報科でデータ分析ツールを用いて実社会における問題解決に焦点が当てられている。大学では分野を問わず全ての学生が統計のリテラシーレベルを学び、応用レベルやエキスパートレベルの内容も提供されている。

高等学校の共通教科「情報」では、「社会と情報」と「情報の科学」が、必修科目である「情報Ⅰ」に統合され、新たに選択科目である「情報Ⅱ」が新設された。図2.1に、学習指導要領の「情報Ⅰ」の「(4) 情報通信ネットワークとデータの活用」からの抜粋を示す。ここに記されている「知識及び技能」「思考力、判断力、表現力等」の観点からみると、問題解決にはデータの収集から始まり、データの整理・分析を経て結論を導く過程が重要であり、これらの能力を身につけることが求められている。

次に、情報Ⅰ及び情報Ⅱにおける学習内容を調査したところ、表2.1のようにデータサイエンスの内容が本格的に組み込まれていることがわかった。これまでの「社会と情報」の科目では、表計算ソフトのスキルに焦点を当てた情報リテラシーの学習であったのに対し、情報Ⅰのデータ活用の学習では、同様に表計算ソフトを利用しつつも、データの特徴を読み取り考察する能力が強調されている。

#### 2.1.1 情報Ⅰのデータサイエンスの学習内容

情報Ⅰの「(4) 情報通信ネットワークとデータ活用」には、データサイエンスの教育が組み込まれている。ここでは、情報Ⅰの教科書から読み取れる具体的なデータサイエンスの学習について説明する。

- ア 次のような知識及び技能を身に付けること。
- (ア) 情報通信ネットワークの仕組みや構成要素, プロトコルの役割及び情報セキュリティを確保するための方法や技術について理解すること。
  - (イ) データを蓄積, 管理, 提供する方法, 情報通信ネットワークを介して情報システムがサービスを提供する仕組みと特徴について理解すること。
  - (ウ) データを表現, 蓄積するための表し方と, データを収集, 整理, 分析する方法について理解し技能を身に付けること。
- イ 次のような思考力, 判断力, 表現力等を身に付けること。
- (ア) 目的や状況に応じて, 情報通信ネットワークにおける必要な構成要素を選択するとともに, 情報セキュリティを確保する方法について考えること。
  - (イ) 情報システムが提供するサービスの効果的な活用について考えること。
  - (ウ) データの収集, 整理, 分析及び結果の表現の方法を適切に選択し, 実行し, 評価し改善すること。

図 2.1: 情報 I で扱われるデータ活用 (学習指導要領 [2] より)

表 2.1: 情報 I, 情報 II における学習内容一覧

| 科目    | 教科書で扱われる学習内容の例  |
|-------|---|
| 情報 I  | 質的・量的データ、欠損値、外れ値、ヒストグラム、クロス集計、分散、標準偏差、散布図、相関係数、単回帰分析、箱ひげ図、四分位数、検定、テキストマイニング |
| 情報 II | ビッグデータ、バイアス、クレンジング、重回帰分析<br>ロジスティック回帰、決定木、k 近傍法、k 平均法                       |

## データの種類

データは、量的データと質的データに大別される。例えば、「身長」のような連続的なデータである比例尺度は、具体的な数値（センチメートル）で表される。一方で、「年齢」のような離散的なデータは間隔尺度と呼ばれ整数値で表される。他にも、「順位」や「血液型」などを表す順序尺度や名義尺度が存在する。これにより、主観的な感覚で判断するのではなく、客観的にデータを観察し理解できる能力が養われる。

## データの収集

データの収集には、総務省統計局の e-Stat [11] や国や自治体が公開しているオープンデータが利用される。

また、データベースの学習も不可欠である。ここでは RDBMS のデータベースを紹介し、データベースにアクセスする方法 (選択、射影、結合) を学習する。教育用に開発されたツールとしては、長瀧らが開発したデータベース学習ソフト「sAccess」を利用した、コンビニの POS データから必要なものだけを抽出して分析を行う実習例が紹介されている [12]。

### データの整理

データの整理方法も扱われる。データには、欠損値、外れ値、異常値が存在する可能性がある。具体的な手法として、表を観察して欠損値を見つけたり、箱ひげ図から外れ値や異常値を検出する実習が挙げられている。また、外れ値や異常値の有無による平均値や中央値の変動を観察することで、平均値は大きく影響をうけて中央値はあまり影響を受けないなど、それぞれの代表値の特性を理解することができる。

### データの分析

データの分析方法は大きく6種類ある。そのほとんどが記述統計であり、一部は推測統計の学習内容である。

1つ目は、量的データの可視化手法に焦点をあて、例えば、棒グラフや散布図などを用いて数値の傾向や特徴を視覚的に表現する。ただし、値の変化に注目したい場合は折れ線グラフ、大きさに注目したい場合は棒グラフ、値の割合に注目したい場合は円グラフや帯グラフといったように、データの種類に合わせて可視化するものが変化する。このように、データの可視化は何通りもの方法が存在し、適切なグラフを切り替えながら分析できることが望ましい。

2つ目は、代表値（合計値、平均値、中央値、最頻値、分散、標準偏差）の理解が求められる。代表値は、データの傾向を1つのデータで表す指標であるが、データが変化することによって、代表値がどれだけ変動するかは想像しにくい。そのため、可視化されたグラフと併用することでデータの特性を理解しやすくなる。

3つ目は、質的データを分析する方法（度数分布、クロス集計）の理解が求められる。例えば、「血液型」を取り上げ、度数分布を作成して血液型の割合を確認したり、また、クロス集計を用いて血液型と性別の関係を調査できる。ここで、質的データが量的データとして変換されるため、棒グラフ、円グラフ、帯グラフで比較が行える。

4つ目は、散布図と相関係数から2つの関係が互いに相関しているかを学習する。データ上では相関が見られる場合でも、「擬似相関」が存在することから、データの種類を検討する必要がある。さらに、単回帰分析に進み、回帰直線を算出し、その決定係数からデータをどれだけ正確に表現できているかを評価する能力も身に付けることが求められる。

5つ目は、t検定やz検定のような推測統計を学習する。例えば、購入した100gのお菓子10個程度を標本とし、その重さを計測して母集団の平均を推定することや、Aクラスの成績とBクラスの成績の平均に有意な差があるかどうかを調査することができる。

6つ目は、テキストデータから分析を行うテキストマイニングを学習する。例えば、太宰治の「走れメロス」をテキストマイニングしたサンプルデータを利用して、言葉の出現頻度を視覚化することや言葉の関係図から文章の主題を把握することが可能である。

## 2.1.2 情報Ⅱのデータサイエンスの学習内容

情報Ⅱの「(3) 情報とデータサイエンス」には、データサイエンス教育が組み込まれており、情報Ⅰよりもやや高度な内容が含まれている。主な特徴は、膨大で多様なビッグデータや機械学習を取り

扱うことであるが、情報 I から引き続き記述統計や推測統計に関する内容も含まれている。ここでは、情報 I との変更点を中心に各項目を説明する。

### データの収集

データの収集においては、バイアスの問題が扱われている。具体的には、実験や調査において実際に観測したものとそうでないものの差が選択バイアスであることや、実験や調査の方法が違うものの差が情報バイアスであることについての理解が求められる。これらのバイアスを最小限に抑えつつ公平性を保つ実験の実施が必要である。また、Web API を利用して XML、JSON 形式のデータを取り扱う方法も載っている。

### データの整形

データの整形では、異なる複数のサイトからデータを収集した表記揺れに対処する方法が説明されている。具体的には、収集したデータの名称が異なるが意味は同じである場合、それを統一することで機械的な処理が容易になるといった事例が挙げられている。さらに、試験結果に対して 1(合格) 及び 0(不合格) に変換するダミーデータの生成についても触れられている。

### 欠損値と外れ値

外れ値の取り扱いにおいては、実際に計算から外れ値であるかを判断する方法について説明されている。また、欠損値に対する処理方法についてグループで議論を通じて、平均値や中央値の代入、データの削除などが検討されるべきであることが記述されている。

### 機械学習

機械学習においては、教師あり学習と教師なし学習の 2 つの種類が紹介されている。

教師あり学習については、4 種類程度のアルゴリズムが扱われている。重回帰分析は、複数の説明変数を用いて目的変数を予測する。例えば、住宅価格の予測など、数値を予測する場面で利用できる。ロジスティック回帰は、主に分類問題に適した手法であり、1 と 0 の二値を予測する。例えば、合格・不合格、買い・売りのような二分類問題に利用され、データの整形で学習したダミーデータにより 1 と 0 に分類する。決定木は、質問に対して Yes/No を入力していく形で学習する。これにより、データを分岐して最終的な目的変数を予測できる。分類問題に有用で、例えば性格診断のような分類が可能だが、適切な質問を考えることが重要である。分類に偏りが生じないような工夫が求められる。k 近傍法は類似度の高いデータをグループ化する教師ありの機械学習の手法である。入力データ集合から異常値を持つデータを発見する異常検知に利用される。

教師なし学習では、k 平均法などのアルゴリズムが扱われている。k 平均法はクラスタリングの最も簡単な手法である。k 近傍法では、ラベルデータを用いているが、k 平均法は、数値の集合として扱いグループ化する。例えば、まだ未知で分類できていない顧客データから、どのようなグループに属しているかを知るのに利用できる。

## 2.2 データサイエンス教育に関する先行研究

### 2.2.1 一般的に利用される分析ツール

#### 表計算ソフト

表計算ソフト [13] [14] は、データの集計やグラフの可視化が行えるソフトウェアであり、代表的なものに Microsoft Excel や Google スプレッドシートがある。情報 I の教科書 12 冊中、9 冊がこれらのソフトウェアを採用している。基本的には、情報 I で取り扱う記述統計や推測統計が利用でき、アドインをインストールすれば重回帰分析までの利用が可能となる。

表計算ソフトは、グラフを可視化したい場合は「グラフの挿入」を利用し、データの集計を行いたい場合は「ピボットテーブル」を利用し、代表値の算出や検定を行いたい場合は、キーボードを用いて「関数」を入力する。ここでは、図 2.2 のように、気温とアイスクリームの売り上げ数の相関を調査する例を見る。まず、分析したいデータをマウスで選択する。次に、挿入タブから散布図を選択する。質的データの集計が必要な場合は、左にあるピボットテーブルを利用する。出力された散布図は 0 を基点に描かれているため、グラフの表示範囲を変更して詳細に観察する。

表計算ソフトは高等学校の教員も使い慣れているといったメリットがあるが、統計に特化したソフトウェアでないことから操作は簡単とは言えない。さらに、ビッグデータのような大量のデータの分析に対応していない。

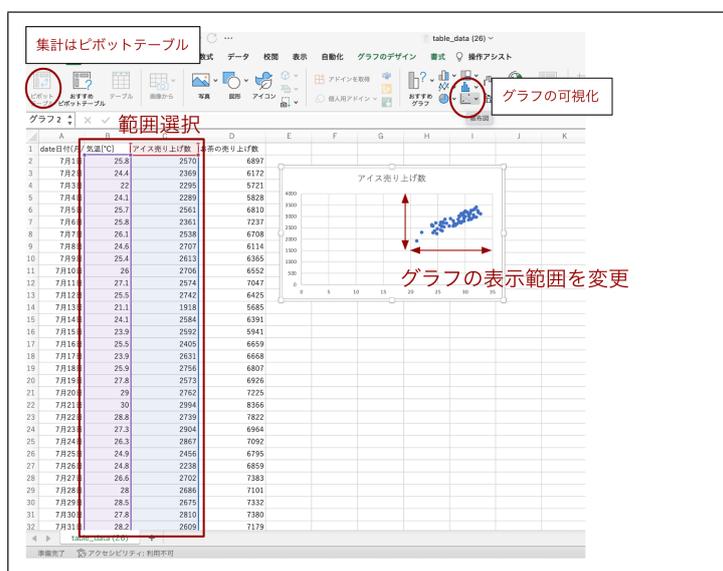


図 2.2: 表計算ソフトの画面例 (Microsoft Excel)

## R

統計に特化したプログラミング言語である R [15] は、シンプルに統計処理を記述できる特徴があり、授業での利用が期待されている。1991 年に、Ross Ihaka と Robert Gentleman によって開発され、オープンソースで誰もが無料で利用できる。統合開発環境である R Studio [16] が用意されてお

り、プログラムを入力する部分とグラフを可視化する部分が分かれている。この言語はデータ処理において、配列の考え方を知らなくても利用できることや、パッケージより機能をカスタマイズできることが利点である。ここでは、表計算ソフトと同様に気温とアイスクリームの売り上げ数の相関を調査する例を見る。R Studioは、図2.3のように、左上に表、左下にプログラムを記述するコンソール画面、右上に変数の値を確認する画面、右下にグラフが表示される。プログラムを実行するとき、右上のような変数の確認ができる画面があると、実際にデータが読み込めているかを確認することができるためデバッグしやすい。図2.4には、そのソースコードを載せている。まず、`read.csv`にてCSVファイルを読み込む。読み込んだものを表に表示するにはViewを利用する。`vx,vy`に気温とアイスクリームの売り上げの値を代入し、`plot`で散布図、`cor`で相関係数を計算している。このように、Rはプログラミング言語であることから、プログラミングの概念や文法を学習しなければならない。

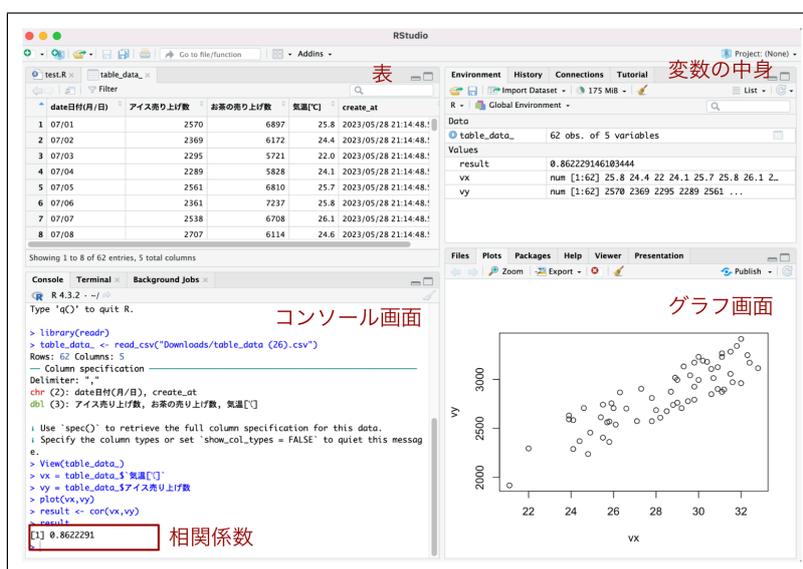


図 2.3: R Studio の画面例

```

01: library(readr)
02: table_data_ <- read_csv("table_data.csv")
03: View(table_data_)
04: vx = table_data_.$'気温 [°C]'
05: vy = table_data_.$アイス売り上げ数
06: plot(vx,vy)
07: result <- cor(vx,vy)
08: result
  
```

図 2.4: 散布図と相関係数を出力するプログラム例 (R)

## Python

Python [17] は、1991年に Guido van Rossum によって開発されたプログラミング言語で、動的型付けシステムにより C 言語よりもプログラムが書きやすい。実際に、情報の教科書では、プログラミングの学習に Python を採用していることが多い。

Python の利点は、豊富なライブラリに統計処理のパッケージ (matplotlib、pandas、scikit-learn) が含まれていることである。matplotlib は、グラフを可視化するパッケージである。scikit-learn は、回帰直線やロジスティック回帰などの機械学習が利用できるパッケージである。pandas は、CSV や JSON 形式のファイルから読み込んだデータを表形式 (DataFrame) で扱うことができるパッケージである。これにより、matplotlib や scikit-learn にデータを入力する手間が軽減できたり、データの集計、代表値の算出が行える。これらを入れることで R と同様に統計処理が行え、機械学習の導入に利用できる。また、学習環境も揃っていることが利点として挙げられる。例えば、Google のアカウントがあれば、Google Colaboratory 上で Python の学習が可能である [18]。さらに、長島らは Bit Arrow にて Python の統計処理用ライブラリを用意しており、どちらもブラウザから利用できるという利点がある [19]。

ここで、表計算ソフトと同様に Python で散布図と相関係数を出力する例を示す。図 2.5 は、BitArrow での開発画面である。ブラウザの右側にはプログラムを記述するエディタがある。実行された結果はダイアログとして出力され、そこにはグラフなどが表示できる。BitArrow は素材管理ができ、CSV ファイルをクラウド上に保存して利用することができる。図 2.6 には、プログラム例を示している。パッケージを使用するためには import 文が必要である。このプログラムでは、matplotlib と pandas が利用されている。まず、pd.read\_csv で、CSV ファイルを DataFrame で変数に代入する。それを利用して plt.scatter で散布図を表示し、df.corr で相関係数を計算している。このように、Python は R と同様にプログラミングの概念や文法を理解する必要がある。

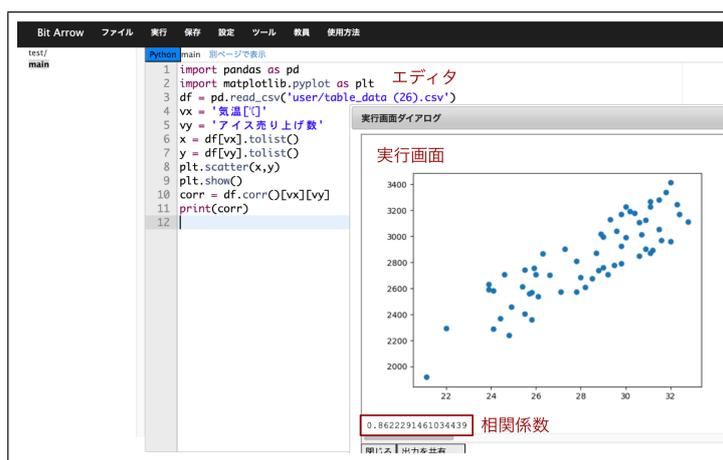


図 2.5: Python の開発環境例 (BitArrow)

```

01: import pandas as pd
02: import matplotlib.pyplot as plt
03: df = pd.read_csv('user/table_data.csv')
04: vx = '気温 [°C]',
05: vy = 'アイス売り上げ数'
06: plt.scatter(df[vx].tolist(),df[vy].tolist())
07: plt.show()
08: result = df.corr()[vx][vy]
09: print(result)

```

図 2.6: 散布図と相関係数を出力するプログラム例 (Python)

## BI ツール

BI ツールは統計処理に特化したツールである。有名なソフトウェアでは IBM 社の IBM SPSS Statistics [20], SAS Institute 社の SAS [21], Tableau 社の Tableau [22] 等がある。どのソフトウェアも有料のソフトウェアである。ここでは、Tableau を例に説明する。BI ツールは、ビッグデータと呼ばれる表計算ソフトでは処理しきれないほどの大量のデータ分析が可能である。基本的には、図 2.7 のように分析したいデータをマウスでドラッグ & ドロップで分析することができ、直感的でわかりやすい。また、データベースとの連携も可能で、データベースの中身が変更された場合に、自動的に反映してくれる機能を持つ。

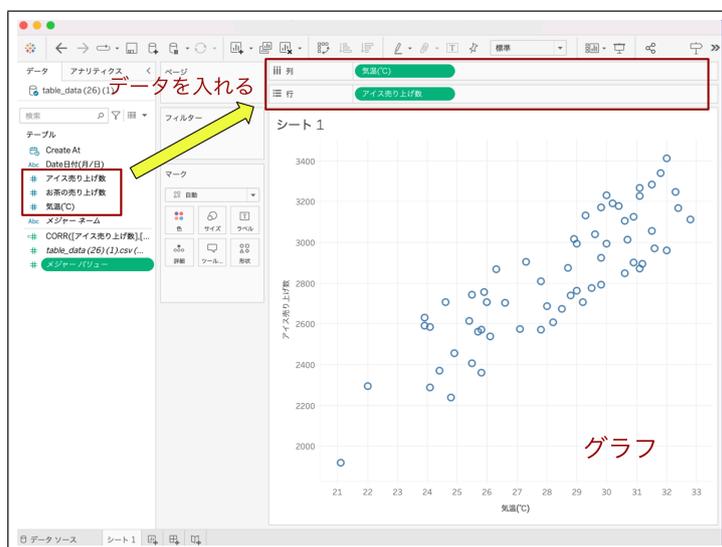


図 2.7: BI ツール (Tableau)

### 2.2.2 教育手法に関する先行研究

この節では、統計教育の手法について、先行研究をもとに説明する。藤田は、日本の学校教育におけるデータサイエンス教育に何が必要であるかを、イギリスで行われている教育の視点から報告して

いる [23]。そこでは、データ分析の指導のフレームワークとアプローチの方法が記述されている。

まず、データ分析のフレームワークとして、Wild & Pfannkuch らが提唱した PPDAC サイクル [24] が挙げられる。これは、Problem - Plan - Data - Analysis - Conclusion のサイクルからなる統計的探究のプロセスであり、統計的な問題解決を行うのに使用する。これは、日本でも林らがこのフレームワークを軸として指導内容を構築し、探究活動の授業での利用の報告がある [25]。

次に、データ分析のアプローチ方法として、3種類の報告がある。1つ目は、異なるデータセットを比較することである。これは、記述統計に該当する部分であり、ヒストグラムや箱ひげ図を可視化することで、何が違うのかを中央値や平均値を利用しながら理解に役立てることができる [26]。2つ目は、統計的モデリングを利用して、現実のデータのメカニズムの理解を深めることである。川上は、統計的モデリングの指導や学習に関しての国際的動向を調査しており、「現実の世界」と「数学の世界」を行き来しながら、統計的モデリング過程を学習に位置づけたり、指導の中で統計的モデリングの考え方を強調したりすることが求められていると報告している [27]。3つ目は、ICTなどのテクノロジーを利用することである。テクノロジーの利用により、必要以上に複雑な計算を簡略化でき、多くの時間を適切な分析方法の選択や解釈に当てることができる。これまで、統計学を教育で利用するには、その計算手法に焦点を当てて授業が組み立てられてきたことが課題としてあり、これからのデータサイエンス教育にはデータ分析ツールの利用が不可欠である。

### 2.2.3 分析ツールに関する先行研究

ここでは、教育用に開発されている分析ツールの先行研究について紹介する。

eStat [28] は Jung Jin Lee らが開発した。このツールは、従来の教育向け分析ツールが特定の用途に限定されがちなか中、各学習レベルに対応し、小学校から大学までの幅広い教育に適用可能である。その特徴的な点として教育段階によって機能を制限するといった制御が可能である。また、学習内容に適したデータセットを提供するためのサイトも公開されている [29]。

TinkerPlot [30] は Jane Watson らが開発した。この分析ツールは、ドットプロットから棒グラフをはじめとする多様なグラフを作成可能である。そのため、生徒はドットプロットを数えることから始め、他のグラフへと表現ができる。このツールを通して、これまでになかった解釈が可能なグラフを自由に作成できる。この考え方を継承した CODAP [31] は、Web サイトにアクセスすれば利用できる。

jamovi [32] は Murat Doğan Şahin らが開発した。この分析ツールは、SPSS のような GUI とマウスの操作だけでデータ分析が行えるという特徴を持っている。その内部では R が動いているためその計算結果も保証されている。また、他のソフトウェアと違って、開発者以外がパッケージを開発できるといった拡張性をもった信念を貫いている。

これらの開発された分析ツールは、表示したグラフの種類によって必要な代表値だけが表示できる機能を持っていることから、統計の基礎的な知識を支援していることが特徴である。以上のことから、教育用として開発されたものには、簡単なユーザインターフェースや操作性などを実現するための技術的特性や、操作時の速度 (操作回数が少ない)、効率的であるなどを実現するための使用特性が求められる [33]。

## 2.3 本論文の位置付け

本章では、高等学校の情報科におけるデータサイエンス教育に焦点を当てて述べた。数学科では統計学の理論を学習する一方で、情報科では統計ツールを活用して問題解決に取り組んでいる。この実現には、データの収集から始まりデータ分析を通して結論を導き出すまでの過程が不可欠である。そして、これまで利用されてきた分析ツールを授業で活用するには、4つの課題が明確になった。

最初の課題は、データ分析が行える環境構築である。一般的に高等学校のPCにインストールされている表計算ソフトのように、どこでも利用できるような環境が求められる。これは、先行研究でも挙げられていた分析ツールのほとんどがWebアプリケーションであることを考えるとその形態が望ましい。次に、高等学校でデータサイエンスを扱う時間数には限りがあり、短い時間で学習が行える支援機能が求められる。これは、分析ツールの操作の容易性の実現が望ましい。次に、授業の準備の困難さを克服することが求められる。生徒が授業でデータを分析するためには、データファイルを使用するが、従来は教員自身がサンプルデータセットを作成し、受講する生徒全員にデータファイルを配布する必要があった。また、センサからデータを計測するにはマイクロビットなどの教材を購入する必要があった。最後に、実際の授業での利用が可能な安定したシステム運用が求められる。研究を目的として作られたソフトウェアは、通常特化したアプリケーションであることが多い。

本論文では、これらの課題に対処するために、データ分析学習システム Connect DB を開発した。まず、Webアプリケーションで動作し、スマートフォンやタブレット端末でも利用できるUIを設計することで、どこでも利用できるような環境を構築した。次に、短時間で分析が可能になるように、学習者の認知負荷を軽減するような支援機能を実装した。具体的には、試行錯誤を促す手法として新しいタブに分析結果を追加し、必要に応じてそのタブを削除する機能を組み込んだ。さらに、多くの分析方法から選ばせない手法として、データの種類によって分析できる方法だけを提示する機能を導入した。また、情報IIで学習する重回帰分析の結果の理解が深められるような視覚的な工夫を行った。

次に、データベースを提供することで、教員がサンプルデータセットをアップロードできるようにした。そのデータベースには、効果的な学習が行えるサンプルデータセットも提供しており、教員がサンプルデータセットを作成する手間を軽減した。例えば、身長サンプルデータは、e-Statで公開されている16歳の平均身長と標準偏差を活用して、正規分布に似た形状の架空のデータを生成した。身長は性別ごとにその平均値と標準偏差が異なることから、身長だけに注目するとヒストグラムが少し広がった形となるが、性別でデータを分けてからヒストグラムを観察すると正規分布に近い形状を確認できる工夫をした。他にも、気温とアイスの売上数とお茶の売上数を利用することで、アイスの売上数とお茶の売上数の関係は、疑似相関であることを考察できる工夫をしている。情報IIの重回帰分析においても、目的変数に影響がある説明変数とそうでない説明変数を準備することで、学習者に結果を考察させる工夫をした。また、入力フォームを作成することでデータの収集を行うこともできる。

最後に、センサからデータを計測するために、スマートフォンから本システムを利用することで加速度センサ、ジャイロセンサ、GPSを利用できるようにした。これにより、センサの購入を必要と

せずに計測の授業を実現した。

次章以降では、これらの4つの課題を解決するために開発されたデータ分析学習システム「Connect DB」の技術的な内容を説明する。そして、「短い時間で学習が行えること」「情報Ⅱの重回帰分析の結果の理解ができること」「授業の準備の困難さを克服すること」について授業実践からその学習効果を検証することで議論する。

## 第3章

# データ分析ツール Connect DB の設計と実装

### 3.1 Connect DB の概要

#### 3.1.1 システム構成

Connect DB は、教育用のデータ分析学習システムである。図 3.1 に Connect DB のシステム構成を示す。Google Chrome、Microsoft Edge、Apple Safari 等の Web ブラウザに対応したレスポンシブデザインを実装することで、PC、タブレット、スマートフォンで動作するように設計した。Connect DB を Web アプリケーションとして実装することで、教員が教室の PC にソフトウェアをインストールする作業が不要になり、「データ分析が行える環境」を構築することができた。

システムは、Python と Flask フレームワーク<sup>\*1</sup>を使用して Web アプリケーションとして実装した。統計処理の実装には Python の Numpy<sup>\*2</sup>、Pandas<sup>\*3</sup>、scikit-learn<sup>\*4</sup> ライブラリを使用した。画面レイアウトには Bootstrap<sup>\*5</sup>を使用した。グラフ描画には D3.js<sup>\*6</sup> ライブラリを使用した。サーバ OS には Ubuntu<sup>\*7</sup> を、Web サーバには Nginx<sup>\*8</sup> を使用した。

Web アプリケーションを構築する場合、Python の Flask 以外にも、node.js の Express や PHP の Cake PHP など考えられる。しかし、本システムが Python の Flask を採用した理由は、マイクロフレームワークという軽量で機能を最小限に抑えているため、他のフレームワークと比べて動作が軽いからである。さらに、Python の機械学習のパッケージである scikit-learn を組み合わせることで、機械学習を効果的に導入できると考えたからである。このライブラリは信頼性が高く、プログラムを迅速に実装する上で確実な手段である。本システムでは、特に重回帰分析においてこのライブラリを活用している。

---

\*1 Flask: <https://flask.palletsprojects.com/en/3.0.x/>

\*2 Numpy: <https://github.com/numpy/numpy>

\*3 Pandas: <https://pandas.pydata.org/>

\*4 scikit-learn: <https://github.com/scikit-learn/scikit-learn>

\*5 Bootstrap: <https://getbootstrap.com/>

\*6 D3.js: <https://d3js.org/>

\*7 Ubuntu: <https://jp.ubuntu.com/>

\*8 Nginx: <https://nginx.org/en/>

データベースは MongoDB<sup>\*9</sup>を使用した。MongoDB は表形式のデータを扱う関係データベースとは異なり、JSON 等の木構造データを効率良く扱うことのできる NoSQL データベースの一種である。また、データ検索等の問い合わせ処理を SQL でなく JSON 形式で指定でき、大量のデータを高速に処理できる特徴を持つ。データベースには、関連した ID だけがアクセスできる仕組みが必要であり、Flask のログインやセッション管理機能 (flask\_login) を利用している。セッションデータは、利用者がログインを行うことで Flask によって暗号化され、Web ブラウザの Cookie に保存される。利用者がログアウトすれば Cookie のセッションデータが削除される仕組みとなっている。

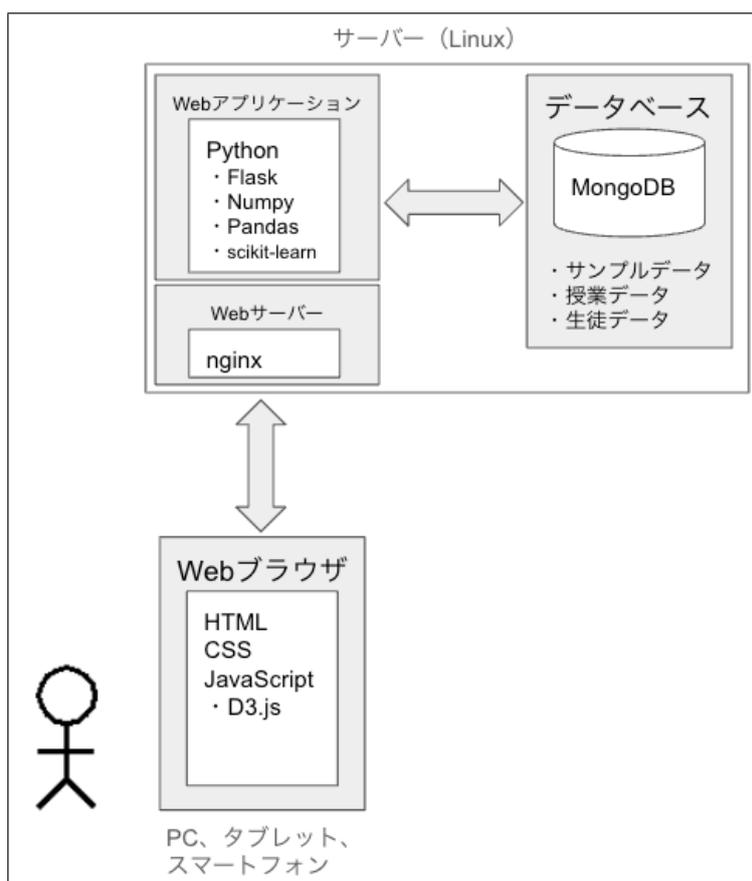


図 3.1: Connect DB のシステム構成

### 3.1.2 データ分析機能

表 3.1 に、本システムがサポートする統計処理機能を示す。これらの関数は、「情報 I」の教科書で扱われている内容に準拠している。

具体的には、「定性・定量データ」、「欠損値・外れ値」、「ヒストグラム」、「分散・標準偏差」、「散布図・相関係数」、「回帰分析」、「箱ひげ図・四分位数」「t 検定」に対応しており、「情報 I」の学習を幅広くカバーすることができる。また、「情報 II」に対応するために「散布図行列」「重回帰分析」の内

<sup>\*9</sup> MongoDB: <https://www.mongodb.com/ja-jp>

容も実装した。将来的には「テキストマイニング」や情報 II の「機械学習」のサポートも予定されている。

グラフの描画は、点の上にマウスを載せることで数値が読み取れるようにするために C3.js のライブラリを採用している。その一方、箱ひげ図や散布図行列は C3.js に実装されていないことから、その元となるライブラリである D3.js を採用して描画を行った。また、重回帰分析は Python の scikit-learn を利用してサーバ側で処理をしている。実際に 100 人程度で授業を行ったところサーバが止まることはなく動作したことを確認できている。

表 3.1: Connect DB のデータ分析機能

| 項目   | 機能       |
|------|----------|
| 統計処理 | 度数分布     |
|      | クロス集計    |
| 代表値  | 合計値      |
|      | 平均値      |
|      | 四分位数     |
| 予測   | 単回帰分析    |
|      | 重回帰分析    |
| 検定   | t 検定     |
| 可視化  | 折れ線グラフ   |
|      | 棒グラフ     |
|      | 積み上げ棒グラフ |
|      | 円グラフ     |
|      | 帯グラフ     |
|      | ヒストグラム   |
|      | 箱ひげ図     |
|      | 散布図      |
|      | 散布図行列    |

### 3.1.3 学習支援機能

データ分析の学習を支援するために、以下の機能を実装した。これらの機能により、生徒がデータ処理の学習に集中できるよう使いやすいツールになり、短い時間で学習が行えるはずである。

#### 統計の学習を容易にする工夫

本システムでは、統計的な考え方を身につけやすくするために、以下のような機能を実装した。まず、グラフ表示時に代表値を計算して表示するようにした。グラフと代表値の関係は表 3.2 の通りである。これは、情報 I や II の教科書を参考に必要なものを調査した。代表値はすべてを表示するのではなく、表示するグラフに必要なものだけが表示される。例えば、箱ひげ図を表示する場合は、四分位数を計算する。ヒストグラムを表示する場合は、分散と標準偏差を計算する。散布図を表示する

表 3.2: グラフと代表値の関係

| 機能     | 代表値          |
|--------|--------------|
| 単回帰分析  | 単回帰分析の式、決定係数 |
| 重回帰分析  | 重回帰分析の式、決定係数 |
| ヒストグラム | 分散、標準偏差      |
| 箱ひげ図   | 四分位数         |
| 散布図    | 相関係数         |
| 散布図行列  | 相関行列         |
| t 検定   | t 値、p 値      |

場合は、相関係数を計算する。散布図行列を表示する場合は、相関行列を計算する。単回帰分析や重回帰分析の場合は、直線の回帰式と決定係数を計算する。

次に、図 3.2 のように外れ値や異常値を可視化し、表中に赤文字で表示することで注意を喚起する機能を実装した。これは四分位偏差法を用いており、四分位偏差の 1.5 倍以上の値が第 1 四分位数や第 3 四分位数から離れている場合に外れ値候補と判定する。

| ■身長[cm] | ■靴のサイズ[cm] |
|---------|------------|
| 165     | 55         |
| 177     | 30         |
| 177     | 28         |
| 167     | 28         |
| 176     | 28         |

図 3.2: 外れ値候補の強調表示

### 操作の容易性

本システムでは、統計処理を PC のマウス操作またはタブレットのタッチスクリーンで完結させるようにした。先行研究の分析ツールの多くは、SPSS のような BI ツールの操作性のものが一般的で、主にマウスをドラッグ & ドロップする。これはマウスでの操作に関しては問題ないが、スマートフォンやタブレットでの利用を難しくしていた。基本的な操作は図 3.3 のように「データの選択」と「分析方法の選択」の 2 つの操作で行える。これらの操作は全てスクリーンをタッチするだけ実行できるため、マウスを利用できない状況でもデータ分析が可能である。

データの並べ替え、データの種類の自動判別、データの削除などの機能も搭載されている。処理を行う際は、「入れ替え」ボタンで列を入れ替え、表示されるグラフの縦軸と横軸を入れ替えることができる。次に、図 3.4 のように各列のデータ型と、「型」ボタンを押したときに表示されるソートボ



図 3.3: 分析の手順

タンを示す。図の左の画面は初期状態であり、型ボタンを押すことで右の画面に切り替えることができる。並べ替えボタン（▼▲）を押すと、データが昇順または降順に並べ替えられる。選択したデータを表から削除したい場合は、削除ボタンを押す。

データ型は「時刻」「文字列」「数値」の3つが表示されるが、初期状態では各列の最も一般的なデータ型が表示されており、これは変更することもできる。本機能の実装方法は、取得した全てのデータを調査し、文字列のデータであっても数値として扱える場合は、適切に数値に変換する。また、時刻を文字列として扱えば、グラフの描画時に横軸の間隔が異なる場合でも等間隔で表示されてしまう課題があった。この課題に対処するため、「2020/10/27」は2020年10月17日の時刻に変換し、「10:10:10」は10時10分10秒の時刻に変換している。時刻の変換には、javascriptのライブラリである day.js を利用し、文字列を適切な時刻に変換した。



図 3.4: 列ごとのデータ型の表示例

### データ処理方法の提案

生徒が選択したデータに応じて、適切な分析方法を提示する機能を実装した。この機能により、列名にあるデータの種類（時刻、文字列、数値）にチェックを入れると、データに合わせてプルダウンメニューに表示される処理方法の候補が変わる。図 3.5 に数値の列と文字列の列の表示例を示す。データ型が数値の場合は、「散布図、ヒストグラム、箱ひげ図」が候補として表示され、テキスト型の場合は、「度数分布、クロス集計」が候補として表示され、データ型に応じた処理方法を選択すること

ができる。この機能は、データベースにデータの種類とそれに対応する分析方法を保存することで実現している。

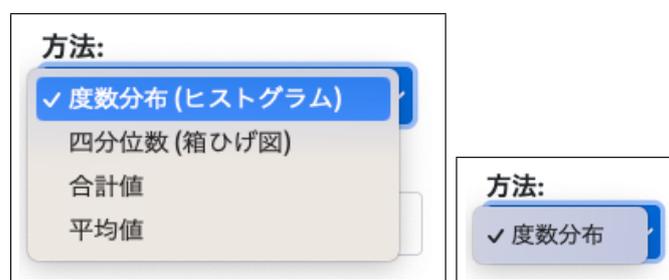


図 3.5: 列のデータ型に応じた処理方法のメニュー表示例

### 学習の試行錯誤を可能にする機能

データ分析処理では、「加工処理→集計処理→グラフ表示処理」を行う。生徒が処理の結果を確認する際に、期待する結果が得られていなければ Undo 機能で途中の処理に戻り、やり直すことができる。そこで、処理結果をタブ形式で保存できる機能を実装した。タブの右側にある「× ボタン」を押すと、結果を元に戻すことができる。この機能により、生徒が自分自身で様々なグラフを表示し、適切な処理やグラフを確かめながら学習を進めることができる。本システムでは、タブのデータはオブジェクト指向で管理するようにした。タブが生成されると新しいオブジェクトにデータをディープコピーで生成し、分析前のデータとは区別される。

### 3.1.4 授業準備を支援する機能

ここでは「授業準備の困難さ」を解決するために実装した機能について説明する。

#### 学習用サンプルデータセットの提供

情報の教科書ではページ数の都合上、数十件程度の表しか記載されていない。そこで、多くのデータを分析できるように数百件のデータ分析学習用にサンプルデータセットを用意し、授業で利用できるようにした。表 3.3 に、用意されたデータとそれに対応する統計処理を示す。データの特徴やグラフの意味を理解するためには、完全にランダムなデータでは難しいため、特徴が理解しやすいように加工している。例えば、「生徒のデータ」のサンプルデータセットを利用した場合、身長と靴のサイズには正の相関があるが、身長と学力には相関がないことを示すデータが作成されている。他にも、「文化祭の売上履歴」については「男性」と「女性」の「焼きそば」と「たこ焼き」の購入履歴が記録されているが、男性はたこ焼きを多く購入し、女性は焼きそばを多く購入するように作成している。

#### 分析データの入力と保存

生徒が自らデータを登録し分析できるように、端末画面からデータを登録する機能と、登録したデータを Connect DB に蓄積する機能を実装した。これらの機能で扱うデータは、授業利用 ID（以下、クラス ID）を用いて管理する。教員はユーザ登録を行うことで授業ごとのクラス ID を作成する

表 3.3: 用意したサンプルデータセット

| サンプルデータセットの名前     | 学習内容   |
|-------------------|--|
| 購入履歴              | 度数分布   |
| 文化祭の売上履歴          | クロス集計  |
| 気温とアイスクリームとお茶の売上げ | 相関係数<br>単回帰分析                                    |
| 身長                | ヒストグラム   |
| 気温                | 箱ひげ図   |
| 身長と靴のサイズ          | 外れ値、異常値  |
| 生徒のデータ            | 度数分布<br>クロス集計<br>相関関係<br>単回帰分析<br>ヒストグラム<br>箱ひげ図 |
| 家賃                | 散布図行列<br>重回帰分析                                   |

ことができる。これらの機能により、生徒が収集したデータを共有することを前提とした蓄積・収集機能を実装することができた。

### データ蓄積機能

教師が学習データをシステムに登録・蓄積できる機能を実装した。データの登録は CSV ファイルを読み込むか、後述するデータ収集機能で行う。国や自治体が公開しているオープンデータを利用した学習も可能になっている。

生徒はクラス ID を入力することで、教師が用意したデータを利用することができる。用意されたデータは読み取り専用データとして生徒に提供されるため、画面上でデータを修正・削除するなどの操作を行っても、元のデータに影響を与えることはない。

### 生徒からの分析データの収集

アンケート方式で生徒からデータを収集する機能を実装した。これにより、データの収集と分析が同じソフトで行えるようになった。図 3.6 に入力フォームの編集画面を示す。教師はこの画面から設問を作成する。入力項目の種類は、「数値」、「テキスト」、「日付」、「選択肢」、「画像ファイル」から選択することができる。入力フォームを作成したら、URL を生徒に配布する。生徒が画面から入力した回答を学習用データとして利用できる。

この機能により、例えば「身長、靴のサイズ、睡眠時間」を入力させる授業を行い、入力されたデータを分析することができる。

データの個数：  
2

名前：  
身長[cm]  
Number

名前：  
靴のサイズ[cm]  
Number

送信

図 3.6: データ入力フォームの編集画面

### 3.1.5 Connect DB を利用したデータ分析の例

本システムを用いたデータ分析の一連の流れとして、サンプルデータセットから文化祭の売上履歴データを用いた分析例を示す。まず、使用するデータを選択し、読み込む。読み込まれたデータは、「性別」と「購入したもの」が表形式で出力される (図 3.7)。

データの解析や削除は、それぞれの行と列の左側にあるチェックボックスで行う。データを削除する場合は、各行の左側にあるチェックボックスにチェックを入れ、「データ削除」ボタンを押す。このとき削除されるデータはブラウザ上のローカルデータであり、データベースの内容は削除されない。データを解析するには、各列から使用するデータのチェックボックスにチェックを入れ、「方法」から分析方法を選択し、「表示」ボタンを押す。

方法：  
度数分布

表示

<< 1/5 >>

スクロールしてテーブルを見てね

データ削除 入れ替え 型

| <input type="checkbox"/> | <input type="checkbox"/> create_at | <input type="checkbox"/> 性別 | <input checked="" type="checkbox"/> 購入したもの |
|--------------------------|------------------------------------|-----------------------------|--|
| <input type="checkbox"/> | 2021/06/27 13:01:31.351            | 男性                          | たこ焼き                                       |
| <input type="checkbox"/> | 2021/06/27 13:01:31.364            | 男性                          | たこ焼き                                       |
| <input type="checkbox"/> | 2021/06/27 13:01:31.365            | 女性                          | たこ焼き                                       |
| <input type="checkbox"/> | 2021/06/27 13:01:31.366            | 女性                          | 焼きそば                                       |

図 3.7: 文化祭の売上履歴データの表示例

図 3.8 は「購入したもの」の度数分布処理の結果を示している。集計結果の表が新しいタブに出力



図 3.8: 分析ごとのタブの追加 (度数分布)

され、分析前は定性的なデータでしかなかったものが、分析後には定量的なデータとして評価できる。この処理の結果をさらにグラフ化したい場合は、使用するデータにチェックを入れ、「方法」からグラフ化したい方法を選択し、「表示」ボタンを押すことで同様にグラフ化することができる。

## 3.2 さまざまな環境で動作する Web アプリケーションの設計と実装

本システムでは、サーバー上に分析すべきデータを保存できるデータベースを提供している。これにより、筆者らが考案した効果的なサンプルデータセットの提供や、教員が授業用のデータセットを簡単に共有できるような環境を整えた。例えば、Google スプレッドシートでは共有したデータの URL を伝えるだけで、簡単に生徒がデータを利用できる。しかし、全ての高等学校が Google のアカウントを所持しているわけではないことから授業の準備が難しいという課題があげられる。それに対して、Connect DB では分析する側はアカウントの登録を必要としないようにした。共有方法に関しては、Google スプレッドシートと同様に、“<https://cdb.eplang.jp/database?key=6BDK2UB0F00L3YO>”の URL と API キー (6BDK2UB0F00L3YO) を伝えるだけであり、この API キーがデータの読み込みに役に立っている。さらに、この API キーを活用すれば、図 3.9 のような Python のプログラムからでもデータを読み込むことが可能である。これにより、Connect DB に保存したデータを Connect DB だけにとどまらずに Python で分析するといった授業展開も可能である。また、利用者ごとにデータベースの保存先を区別するために、ID を登録するための Sign Up 機能や、ID を識別するための Login 機能も実装している。

### 3.2.1 ファイル構造

本システムのファイル構造は図 3.10 の通りである。本システムはデータベースを扱う model、Web ページの制御や描画を扱う views から成り立っている。

まず、model は ID の管理と実際のデータの読み書きを担当している。利用者が Connect DB にログインすると、利用者の ID に関連するテーブル名が表示され、各テーブル名には関連付けられた

```

import requests
import json
key="6BDK2UB0F00L3Y0"
r=requests.get('https://cdb.eplang.jp/api/get?key='+key)
datas = r.json()
print(json.dumps(datas))

```

図 3.9: API キーを利用してデータベースのデータを読み込むプログラム例 (Python)

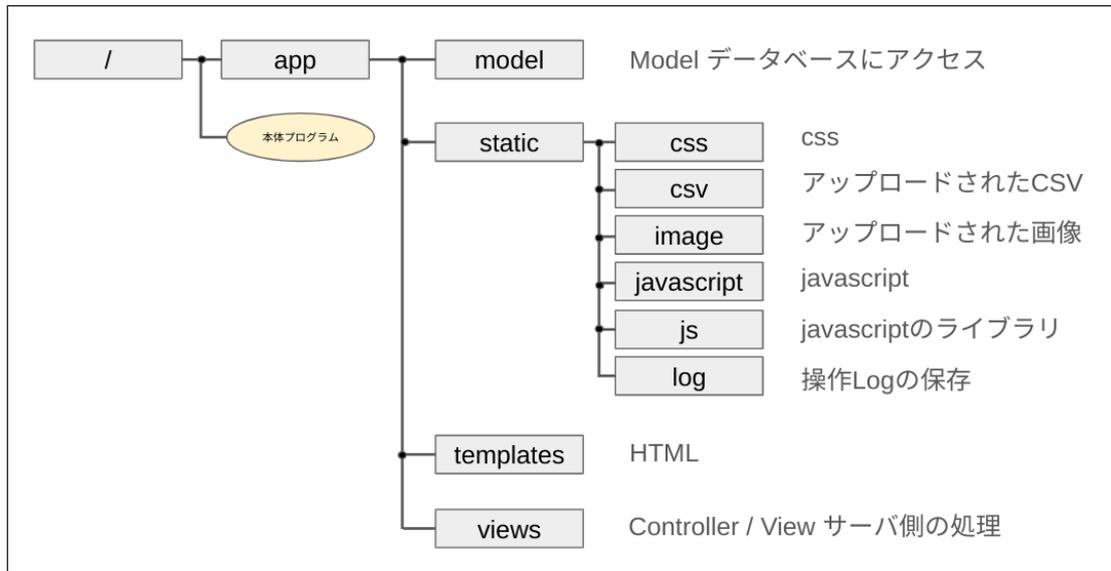


図 3.10: サーバー上でのファイル構造

API キーが付与される。この API キーを利用して、紐付けられた ID とテーブル名の場所に、実際のデータを読み書きする。

次に、views はサーバ側の処理を担当している。例えば、利用者が Connect DB の Web サイトの Top ページに訪れると、ログイン済みであれば自動的にダッシュボードのページにリダイレクトされる。逆に、未ログインの場合はダッシュボードのページに訪れても Top ページにリダイレクトされる。このように、セッションの有無によってページの遷移を制御している。

次に、template は HTML が保存されている。ここでは、テンプレートエンジン Jinja2 を利用し、サーバ側から得たデータを HTML に埋め込むことができる。例えば、図 3.11 のようにサーバ側で `render_template()` を使って `dashborad.html` にデータを埋め込む。図 3.12 には `dashboard.html` のテンプレートの一部を示している。“`{{ }}`”で囲むことで `render_template()` から受け取った `contents` データを利用できる。このテンプレートエンジンは `for` などの制御構文にも対応している。この `dashboard.html` の場合、データベースから ID に対応したデータが見つかった場合は、繰り返し文を使って表を生成している。一方、何も見つからなかった場合は「登録されていません」と表示される。

最後に、static は javascript や css が保存されており、これらのプログラムは HTML から読み込まれている。これらを利用することで、動的な HTML の書き換えが行われページの内容が変化して

いく。また、利用者がアップロードした画像や CSV も static に保存されるだけでなく、開発者側が利用目的を把握するために、分析の操作 Log を保存する機能も実装した。

```
⋮
datas = データベースから ID に対応したテーブルを検索する
contents = []
for data in datas:
    contents.append({"table":data['table'], "full-key":data['key'],
                    "read-key":data['read-key'], "form-key":data['form-key']})
return render_template('dashboard.html', contents=contents)
```

図 3.11: HTML にデータを送るためのプログラム例

### 3.3 データベースの設計と実装

本システムのデータベースは、大量のデータを扱える NoSQL である MongoDB を採用している。MongoDB は、表形式にとらわれない柔軟なデータ保存に向いている。

データベースの設計図を図 3.13 に示す。各 ID にデータベースが用意され、そのデータベース内にはテーブル (コレクション) が作成される。この設計により、異なる ID であれば同じテーブル名でも衝突することはない。ここでは、どのような流れでデータベースにアクセスするのかについて、システムの管理 (login) の ID 管理用テーブルと API 管理用テーブルの 2 つを紹介しながら説明する。

#### ID 管理用テーブル

ID 管理用テーブルは、ID を登録した際に追記される。必要なデータは、「ID 名」「email アドレス」「Password」「授業利用の有無」である。ここでの Password はハッシュ化されたものが保存されているため、外部にパスワードが漏れても元のパスワードを解読することはできない。

ID には、個人利用と授業利用 (クラス ID) の 2 つの利用方法がある。授業利用ではセキュリティがやや緩くなるが、その代わりに授業支援機能が提供される。この機能により、生徒は ID 登録なしでデータベースを利用し、クラス内でデータを共有できる。

生徒はクラス ID を使って生徒用ログイン画面からユーザ名を入力してログインする。ユーザ名は半角英数字であれば何でもよく、データベースへのアクセスの識別に使用される。これにより、先生が公開したデータを分析する際や、他の Web サイトから取得したオープンデータをデータベースに保存し、クラスメイトとディスカッションを行う際に活用できる。

ただし、クラス ID の場合、クラスで ID 名が共通していることから、図 3.13 の設計図では、データベース名が固定されてしまう。その結果、クラス内のテーブル名が重複してしまう問題があった。そこで、生徒が作成したテーブル名には、テーブル名の後ろにハイフン (-) をつけて、ユーザ名を付与する形にした。例えば、ユーザ名「1」がテーブル名「身長データ」を作成した場合、「身長データ-1」というテーブル名になり、テーブル名の衝突を回避することで問題を解決した。

```
⋮
{% for content in contents %}
  <tr>
    <td>
      <a href="{{url_for('database.table',key=content.read-key)}}">
        {{content.table}}</a>
    </td>
    <td>
      <div class="btn-group dropup">
        <button type="button" class="btn dropdown-toggle"
data-toggle="dropdown" aria-haspopup="true" aria-expanded="false">
          </button>
        <div class="dropdown-menu dropdown-menu-right">
          <a class="dropdown-item" href="
{{url_for('menu.form_input',key=content.form-key)}}">入力</a>
          <a class="dropdown-item" href="
{{url_for('database.table',key=content.full-key)}}">データ編集</a>
        </div>
      </div>
    </td>
  </tr>
{% else %}
  <tr>
    <td>登録されていません</td>
    <td></td>
  </tr>
{% endfor %}
```

図 3.12: Dashboard.html の一部

## API 管理用テーブル

API 管理用テーブルは、ダッシュボード画面でテーブルを作成した際に追記される。必要なデータは「テーブル名」「ID 名」「Full-key」「Read-key」「Form-key」「共有の有無」である。「共有の有無」は授業利用の場合のみ使用する。

ここで注目すべきデータは「key」である。これはデータベースにアクセスできる半角英数字 15 文字の API キーであり、他のテーブルの API キーとの重複を避けるようにランダムで生成する。データベースに対して読み込み・書き込み・削除を行いたい場合は「Full-key」を使用するが、このキーはデータの書き込みや削除ができてしまうため、他の利用者には教えないように注意深く取り扱う必

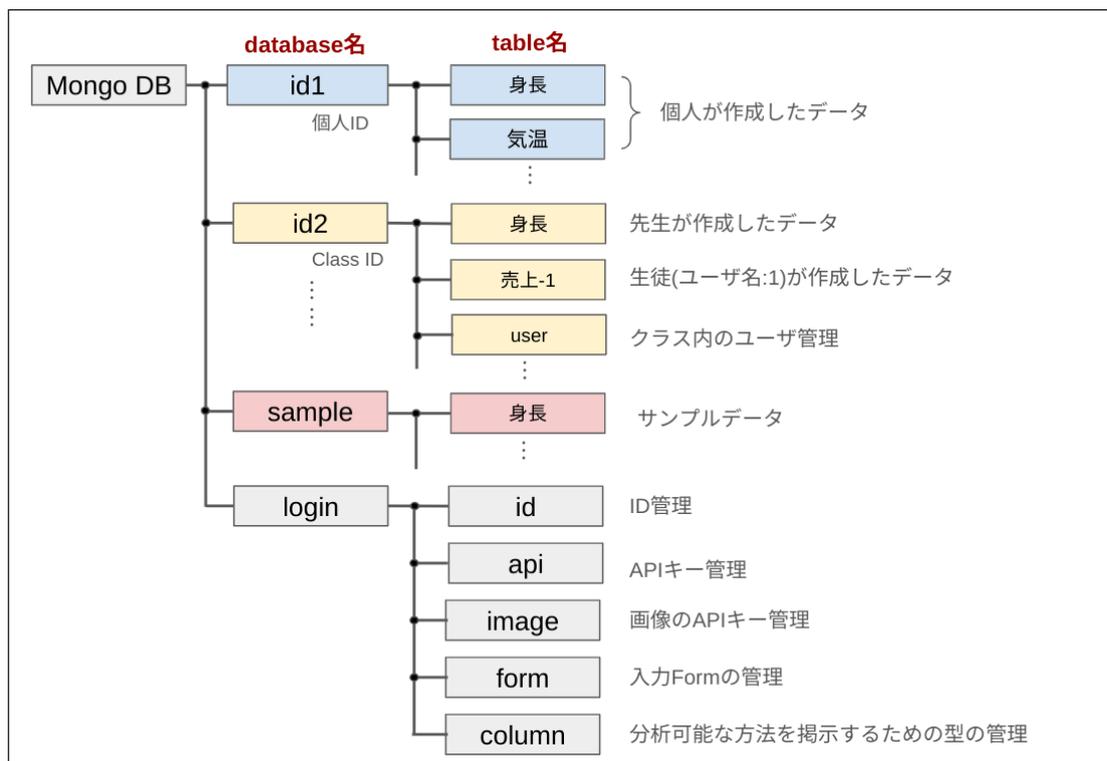


図 3.13: データベースの構成図

要がある。その代わりとして、データの書き込みや削除ができない読み込み専用の「Read-key」を使用することもできる。また、本システムでは Google Forms のような、データを入力してもらうことでデータを収集できる機能も用意している。この機能を利用したい場合は、データの書き込みが可能な「Form-key」を使用する。

API キーの動作原理は次の通りとなる。まず、利用者が入力した API キーを API 管理用テーブルを通して検索し、一致するキーを見つける。一致するキーが見つければ、そのデータにはテーブル名と ID 名が記載されている。その後、ID 名とテーブル名を利用して、実際のデータにアクセスする。データを読み込みたい場合は GET 通信を用い、データを書き込みたい場合は POST 通信を用いる。そして、データを削除したい場合は DELETE 通信を用いる。削除は、データを追記するときに MongoDB 内で自動的に作成される重複のない id を識別子として用いるため、誤って似たデータが削除されることはない。また、ダッシュボード画面からテーブル自体を削除した場合は、API キーのデータだけでなく、実際のデータも削除されるようになっている。

### 3.4 データの分析機能

利用者がデータ分析のページに遷移すると、API キーを利用してデータベースにアクセスする。これにより、API キーが一致したテーブルの JSON 形式のデータがデータベースから取得できる。ここでは、「短い時間で学習が行えること」を実現するために分析ツールの操作の支援機能である「タブの実装」「分析掲示機能の実装」と、情報 II の学習に必要な機能である「散布図行列の実装」「重回帰分析の実装」について述べる。

### 3.4.1 分析すると追加されるタブの実装

新しいタブの追加には、Bootstrap のタブ機能を利用しており、分析を行うたびに HTML のコンテンツが動的に変化する。新しいタブを追加したい場合は、`.nav-tabs` クラスに `<li></li>` (リストアイテム) を追加し、`.tab-content` クラスに分析結果を追加する。`.nav-tabs` クラスにリストアイテムを追加するのは 1 行で実装できるが、`.tab-content` クラスに分析結果を追加するにはかなりの実装が必要となる。そこで、HTML のテンプレートエンジン `jQuery Templates` を採用した。このテンプレートエンジンは、`Jinja2` と同様に HTML の生成を行う前に変数を割り当てることができるだけでなく、HTML の生成をクライアント側だけで実行できる。

### 3.4.2 データ型に応じた分析提示機能の実装

分析対象となるデータの型から、表示可能なグラフだけを提示する機能は、データベースのシステム管理用 (`login`) に、`column` テーブルを作成することで実装した。このテーブルでは、分析対象となるカラムの型の個数と使用可能な分析方法が保存されており、表 3.4 のように「時刻の個数」「数値の個数」「文字列の個数」「画像の個数」に対して「分析方法」が明記されている。例えば、折れ線グラフは「時刻 1 個と数値 1 個以上」や「数値 2 個以上」を選んだ場合に表示され、一方でクロス集計は「文字列 2 個」を選んだ場合にのみ表示される。

### 3.4.3 散布図行列の実装

各変数の関係を理解するためには散布図行列が必要である。一般の表計算ソフトではこの機能が提供されていないため、`R` や `Python` を利用する必要がある。ただし、プログラミング言語がまだ理解していない学習者にとっては難しいと感じられることもある。

そこで、`Connect DB` ではこの問題を解決するために散布図行列を表示する機能を実装した。ラベルデータを使用すると、名前に対して色分けされた散布図行列が表示される機能も実装している。この機能により、データ上では相関しているようなものでも、ラベル分けすると相関していないなどの特徴を知ることができる。

### 3.4.4 重回帰分析の実装

重回帰分析を行うためには、表計算ソフトではアドインのインストールが必要で、説明変数を変更しながら分析を行う際には、列データを入れ替える手間がかかっていた。そこで、`Connect DB` では他の分析と同じく、カラムにチェックを入れて簡単に重回帰分析が行える。本システムでは、サーバ側の `scikit-learn` を使用して重回帰分析が実行されているが、授業では説明変数を変更しながら繰り返して分析できるほど短い時間で分析が可能であった。さらに、本システムでは予測した結果の精度を調査するために、横軸に回帰式から算出した予測値、縦軸に実際の値を使った散布図で描画することとした。これにより、図 3.15(a) のように直線的であると予測の精度が高く、それとは逆に図 3.15(b) のように散布図が広がっていれば予測の精度が低いことが理解できる。

表 3.4: 分析対象の型と分析方法の対応

| 型                                   | 分析方法     |
|-------------------------------------|----------|
| 時刻 1 個、数値 1 個以上<br>数値 2 個以上         | 折れ線グラフ   |
| 文字列 1 個、数値 1 個以上<br>時刻 1 個、数値 1 個以上 | 棒グラフ     |
| 文字列 1 個、数値 2 個以上<br>数値 2 個以上        | 積み上げ棒グラフ |
| 文字列 1 個、数値 1 個以上                    | 円グラフ     |
| 文字列 1 個、数値 2 個以上                    | 帯グラフ     |
| 数値 1 個以上<br>文字列 1 個、数値 1 個          | 箱ひげ図     |
| 数値 1 個以上<br>文字列 1 個、数値 1 個          | ヒストグラム   |
| 文字列 1 個                             | 度数分布     |
| 文字列 2 個                             | クロス集計    |
| 数値 2 個                              | 散布図      |
| 数値 3 個以上<br>文字列 1 個、数値 3 個以上        | 散布図行列    |
| 数値 2 個                              | 単回帰分析    |
| 数値 3 個以上                            | 重回帰分析    |
| 数値 1 個以上<br>文字列 1 個、数値 1 個以上        | 平均値      |
| 数値 1 個以上<br>文字列 1 個、数値 1 個以上        | 合計値      |
| 数値 2 個                              | t 検定     |
| 数値 2 個<br>文字列 1 個、数値 2 個            | 地図表示     |
| 画像 1 個                              | 画像表示     |

### 3.5 データの収集機能

ここでは、「授業準備の困難さ」の課題を克服するために実装した「データ入力機能の実装」「スマートフォンのセンサ計測機能の実装」について述べる。

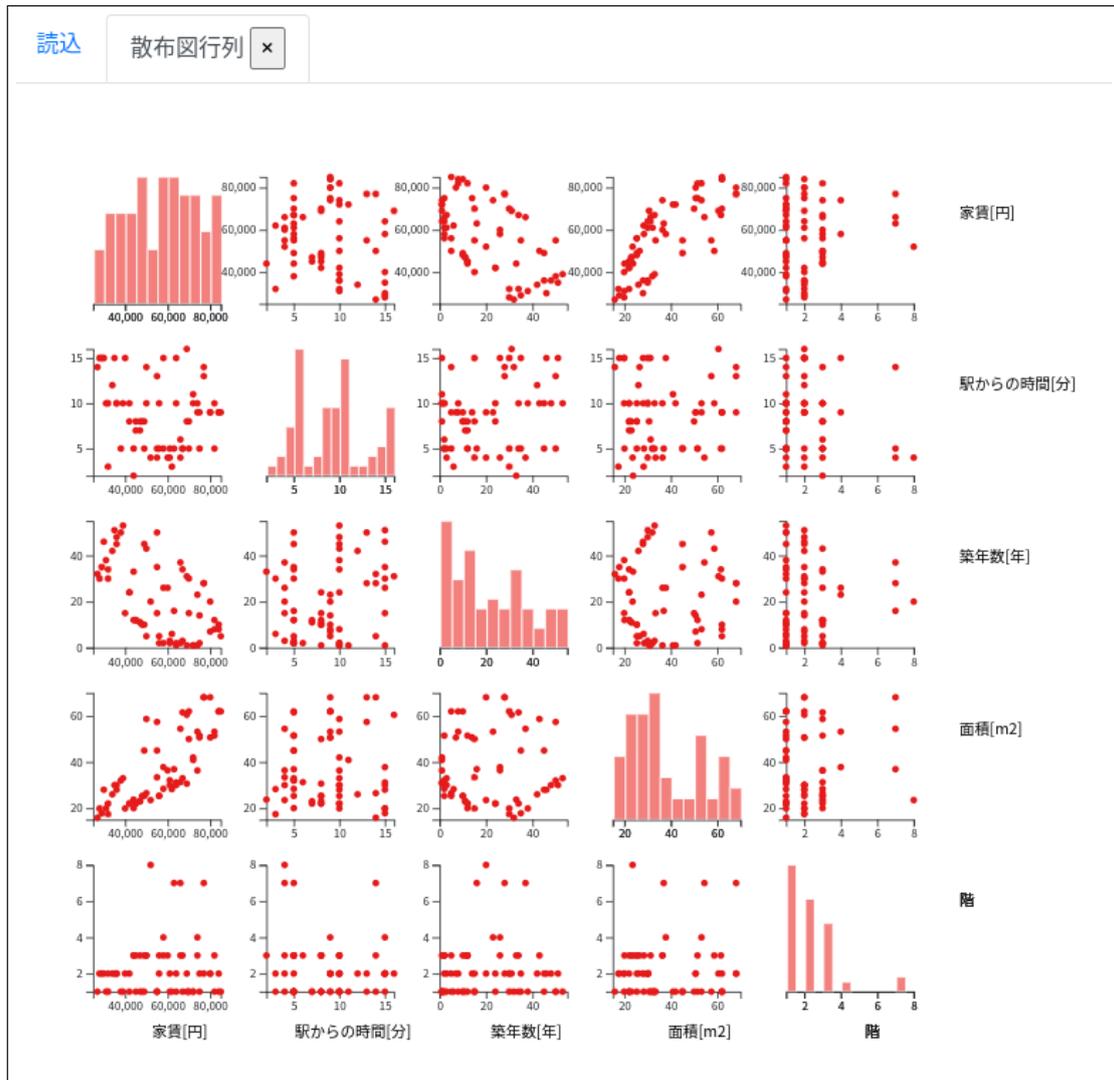


図 3.14: 散布図行列

### 3.5.1 データ入力機能の実装

本システムは、データベースが利用できることを活かしており、Google Forms のようにデータを簡単に入力し、その後データの分析が行えることができる。シンプルな入力システムでありながら、他のソフトを必要とせず利用できるため、利便性が向上した。質問事項は、データベースのシステム管理 (logiin) の edit テーブルに保存される。このテーブルにも API キーが発行されており質問事項を読み込むことができる。入力できるデータの種類の、「数値」「文字列」「選択肢」「日付」「時刻」「画像」である。

数値の入力時には、利用者が誤って文字列を入力した場合、システムが異なるデータ型であることを教示する機能が実装されている。また、文字列の場合、データの集計において表記揺れが課題となることがある。このため、予め用意した選択肢から選ぶ方法も提供している。さらに、スマートフォンやタブレットで撮影した画像のファイルサイズが大きいため、サーバの負担が大きくなる。そこで、予め適切なサイズに変更して最適化している。

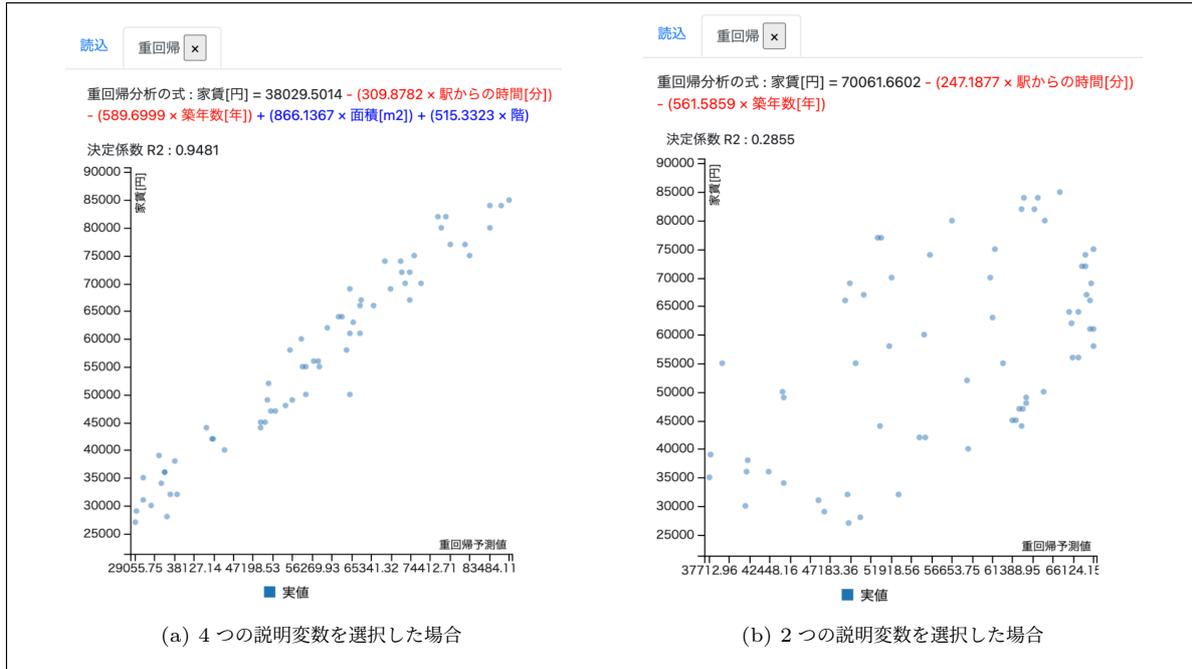


図 3.15: 重回帰分析の散布図

### 3.5.2 スマートフォンのセンサ計測機能の実装

利用者に身近なスマートフォンに搭載されている加速度センサ、ジャイロセンサ、GPS を活用した体験を提供するため、これらのデータを取得できるようにした。

スマートフォンの Web ブラウザを通じて本システムにアクセスすることで、加速度センサ、ジャイロセンサ、GPS を活用できる。システムでは、約 50ms ごとに加速度センサやジャイロセンサの計測が行われ、そのデータがブラウザ内に保存される。利用者は、保存された計測データに対して、「データベースにアップロードする」または「そのままブラウザで分析する」のいずれかが選択できる。データベースにアップロードを選択すると、後で計測データにアクセスできるようになる。一方で、そのままブラウザで分析を選択すると、直ちに計測データを視覚化することができる。

スマートフォンの加速度センサを取得するプログラム例を、図 3.16 に示す。なお、iOS と Android のような OS の違いや機種の違いによっては加速度センサの軸の向きが反転することがある。そこで、利用者が混乱しないようにどの端末でも同じ軸になるように工夫した。筆者が所有している iPhone と Android の端末では、この工夫により軸が同じ向きになっていることを確認した。

加速度センサとジャイロセンサのデータは約 50ms 秒の間隔で取得しており、そのまま折れ線グラフとして表示できる。一方で GPS は、加速度センサやジャイロセンサと同様の間隔でデータを取得すると位置情報の誤差が目立ってしまう。そこで、GPS を利用する場合は 15 秒の間隔で取得するようにした。また、図 3.17 のように地図が表示できるように leaflet のライブラリを利用している。

```
⋮  
window.addEventListener("devicemotion", function(e){  
    aX = e.accelerationIncludingGravity.x;  
    aY = e.accelerationIncludingGravity.y;  
    aZ = e.accelerationIncludingGravity.z;  
    if (navigator.userAgent.indexOf("iPhone") > 0 ||  
        navigator.userAgent.indexOf("iPad") > 0){  
        aX *= -1;  
        aY *= -1;  
        aZ *= -1;  
    }  
});
```

図 3.16: 加速度センサのデータ取得プログラム



図 3.17: 地図の描画 (leaflet)

### 3.6 操作ログの保存

利用者がこのアプリケーションをどのように利用しているかを開発者側が把握するため、操作ログを保存する機能を実装している。クライアント側のみで動作する分析機能でも、サーバ側に対して行われる分析内容を送信している。

保存されるログは、「時間」「ID 名」「ユーザ名 (クラス ID のみ)」「固有識別子」「操作」である。ログインしていない場合、ID 名が存在しないため、「Anonymous」という ID 名で保存される。同一人物であるかを識別するために、Cookie に 10 文字のランダムな英数字を保存し、それを固有識別子として利用している。

主な操作は、どの対象に対してどの分析が行われたかを記録している。これらのデータを利用する

ことで、分析がどの程度の速度で行われているのかや、どのような分析が多く利用されているかを評価することが可能である。

2023年12月時点において利用状況を確認した結果、本システムのID登録数は約40名程度で、これは教員が登録を行った人数である。そのうち6名は本システムを頻繁に使用していることがログから確認できた。また、本システムは登録無しでも利用でき、教員・生徒合わせて約2300名が活用している。2023年、IDを登録せずに利用している者の分析の実行回数を調査したところ約14万回の実行を確認できた。IDを登録せずに利用している人数が約2300名であることを考慮すると、一人当たり約60回の分析が行われている。また、2023年9月から10月にかけての分析の実行回数は約5万回と多いことから、高等学校ではこの時期にデータ分析の授業が重点的に行われていることがうかがえる。

これらの数値を整理すると、システムの利用状況が明確になる。

### 3.7 情報の授業との対応

ここでは、情報の授業で表計算ソフトや Connect DB を利用した場合の流れについて説明する。

#### 欠損値

入力しなかったデータは欠損値として取り扱われる。このような値は授業では取り扱わずに計算することが多い。そのため、表計算ソフトや Connect DB では、総数にカウントせずに合計や平均などの代表値を計算する。また、対応のある t 検定においては、欠損値が片方の群に存在する場合、検定を行うことができないため検定対象から省かれる。

#### 外れ値

授業では外れ値を確認する際、表計算ソフトや Connect DB を使用してデータを図 3.18(a) のように散布図などを表示することで確認する。しかし、オープンデータのような実際のデータを扱う実習では、外れ値が含まれており、授業の中で生徒がデータ内の外れ値を見つけられるための工夫が必要である。Connect DB には図 3.18(b) のように外れ値を赤色で強調する機能が備わっている。一方で、表計算ソフトはそのような機能が搭載されていないため、実際のデータを取り扱う際に外れ値を見逃す可能性がある。

#### 散布図行列

授業では、2つ以上のデータの間関係を調べる際には、散布図行列が有用である。しかし、一般的な表計算ソフトには散布図行列を表示する機能は備わっていないため、データ数が  $n$  の場合、約  $2$  の  $n$  乗通りの図 3.19(a) のような散布図を描く必要がある。これは、非常に手間がかかり、縦軸と横軸の調整も難しい。このままでは、生徒がデータを選択する際に謝る可能性も高まり、実用的ではない。一方、Connect DB には散布図行列の機能が搭載されている。さらに、図 3.19(b) のように質的データを選択した場合には、色を変えて散布図行列を描いてくれるため、大学共通テストのサンプル問題

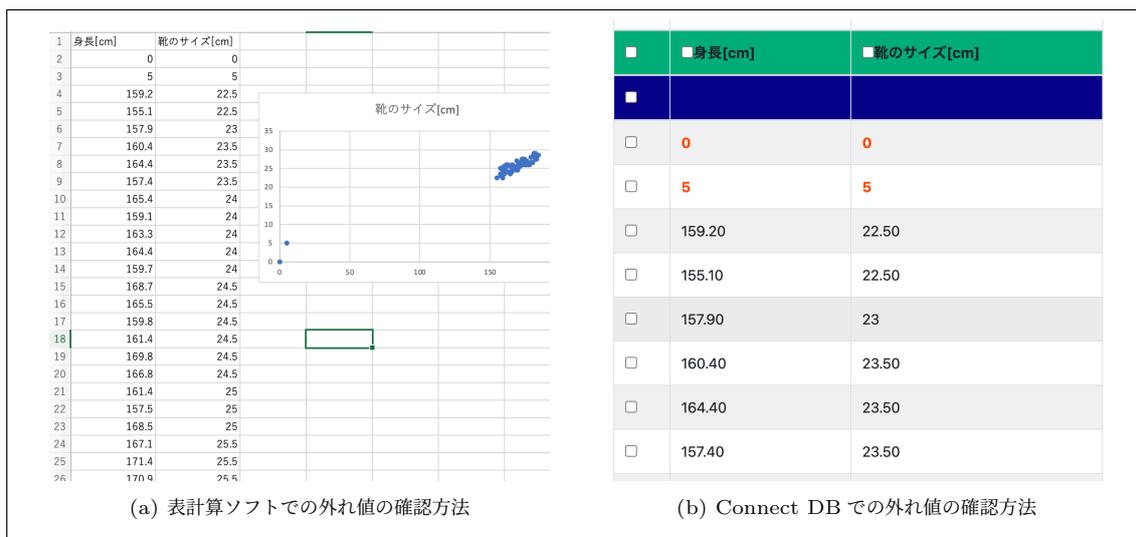


図 3.18: 授業における外れ値の確認 (表計算ソフト、Connect DB)

のような設問にも対応できる。

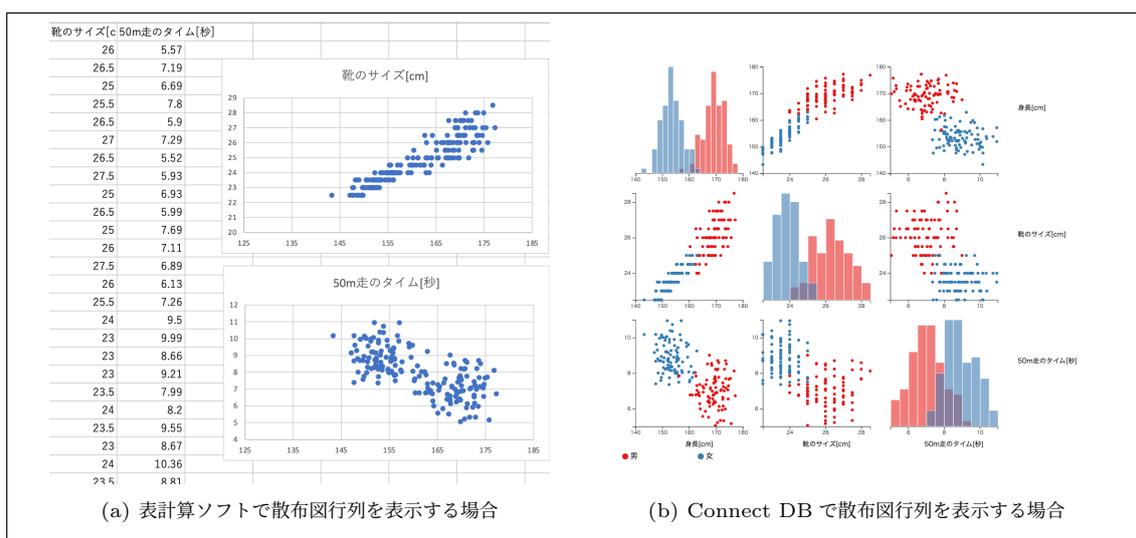


図 3.19: 授業における散布図行列の表示 (表計算ソフトと Connect DB)

### 重回帰分析

授業で表計算ソフトを使用した重回帰分析を行う際、まず図 3.20(a) のようにアドインのインストールが必要である。ただし、学校の備え付けの PC が約 40 台ある場合、この作業が教員にとって負担となりえる可能性がある。また、教員が管理者権限を持っていない場合はインストールできず、結果として重回帰分析ができない状況も考えられる。一方で、Connect DB を使用すれば重回帰分析が可能であり、偏回帰係数や決定係数の値を取得できる。さらに、図 3.20(b) のように予測値と実際の値との散布図を通じて精度を視覚的に確認することもできる。

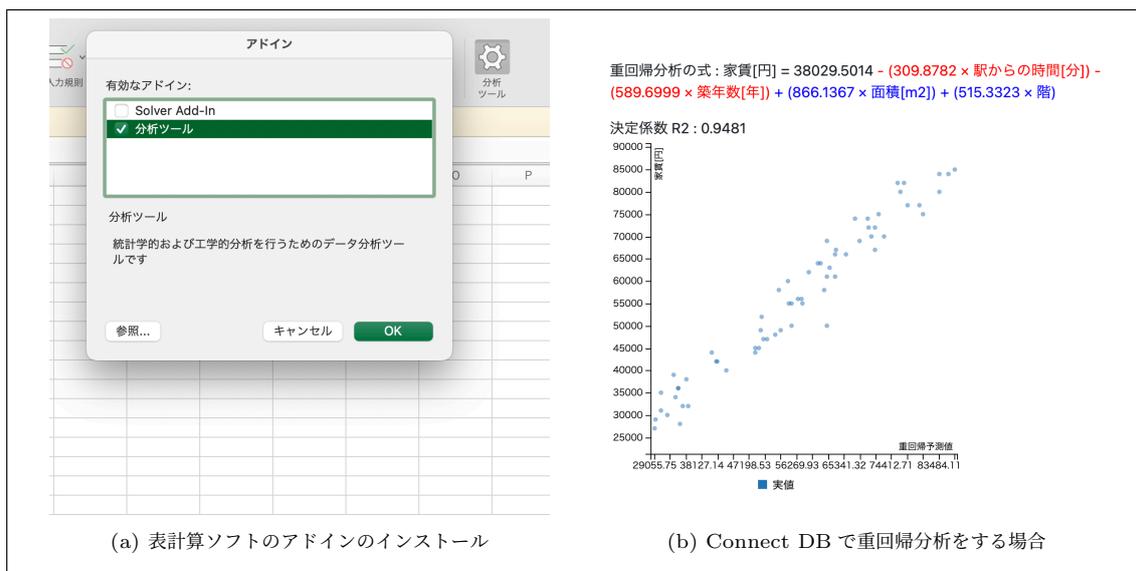


図 3.20: 授業における重回帰分析の利用 (表計算ソフトと Connect DB)

## t 検定

授業で表計算ソフトを使用した t 検定を行う際、図 3.21(a) のように「TTEST」関数を使用する。まず、TTEST の 1 つ目と 2 つ目の引数には対象となるデータを選択する。そして、片側検定なら「1」、両側検定なら「2」を 3 つ目の引数として入力する。最後に、対をなすデータの場合は「1」、等分散の 2 標本を対象にする場合は「2」、非等分散の 2 標本を対象にする場合は「3」を 4 つ目の引数として入力する。授業では、生徒は分析の条件を設定するために対応表を確認するが、条件が数値であるために生徒が正確な条件を入力できたかを確認することが難しい可能性がある。一方で、Connect DB では図 3.21(b) のように分析の条件を選択肢から選択できるようにしている。また、分析の条件を後から確認しやすくするために、出力結果にどの条件を選択したかを表示させる工夫をしている。

## テキストマイニング

テキストマイニングの機能は、表計算ソフトや ConnectDB には搭載されていない。「AI テキストマイニング」の Web サイト\*<sup>10</sup>での実習が可能であり、教科書でも採用されている。

### 3.8 情報の教員の感想

情報の教員 64 名に対して、情報 I の授業が始まる前に Connect DB を活用してもらい有効性の評価を行った。その結果を図 3.22 に示す。

図 3.22 左側の箱ひげ図は、授業開始前に収集したアンケートであり、データ分析の授業に対する自信を 5 段階評価で尋ねた結果である。授業に対する自信の平均値は 2.47 で、おおよそ中間から低

\*<sup>10</sup> AI テキストマイニング: <https://textmining.userlocal.jp/>



図 3.21: 授業における t 検定の利用 (表計算ソフトと Connect DB)

い自信が示された。アンケートのコメントには、「データ活用に便利なツールを知りたい」といった分析ツールに対する期待があった。

図 3.22 右側の箱ひげ図は、提案システムを活用した後で、再び教員に対してこのツールの授業利用の有効性を 5 段階で訪ねた結果である。平均値は 4.17 と、本提案システムを授業で利用することに対しての有効性が高いことが示された。アンケートのコメントには、「提案システムが簡単にグラフを作成できること」や「試行錯誤が促進されること」が特に好評であった。これにより、教員にとって操作が簡単で試行錯誤が促進されやすい分析ツールであることが示唆された。

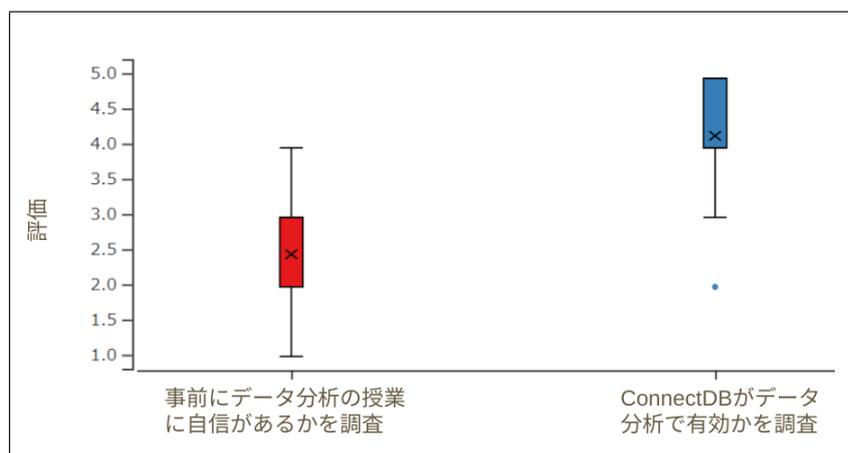


図 3.22: 高等学校の情報科教員によるアンケート結果 (64 名、5 段階評価)

## 第 4 章

# Connect DB: An Online Learning System for Data Analysis

### 4.1 Introduction

In the previous Chapter, we developed the data analysis learning system ‘Connect DB’ to realize the analysis tools needed in class.

This Chapter focuses on learning problems on the student side and lesson preparation problems on the teacher side. We consider functions to solve them. We have found that one of the learning problems on the student side is the difficult operation of the analysis tool. Therefore, we will compare the number of operations using the proposed system and spreadsheet software(Microsoft Excel), and clarify the reasons for the easy to operation of the proposed system through the class.

### 4.2 Some problems for using data analysis in class

When considering the lessons on data analysis, we found that there are some problems for students in learning and for teachers in preparing the lesson.

The problems for the class are shown in (Problem 1) and (Problem 2).

(Problem 1) The learning cost is high because of the mix of learning statistics and learning how to operate the system. This fact was reported in a statistics lesson that yielded survey results showing that the tools were helpful [53], and we surmised that there was a mix of content to learn and learning how to operate the tools.

(Problem 2) In order for students to develop the skills described in the guidelines [2], we realized that it would be ideal for students to be able to display and judge different graphs for themselves. Also reported is that students in statistics classes are required to develop the ability to judge and explain data according to their purpose [54].

The problems for teaching preparation are shown in (Problem 3) - (Problem 5).

- (Problem 3) The textbooks [41–52] actually used in the lesson include examples of surveys being collected and discussed in groups by several people, so there are opportunities to use this in the lesson. In such lessons, it is desirable to create opportunities to share the collected data so that discussions can take place. For example, data can be easily be collected using Google Form. However, if everyone is to analyze the data individually, copies of the collected data will be distributed. However, if the data collected is to be added to, the data will need to be redistributed.
- (Problem 4) Teachers can't freely install software on the terminals used in high schools because of security and maintenance requirements. In this regard, it has been pointed out that teachers in the schools that actually use the software are unable to install the software they want [19].
- (Problem 5) There are a few practical examples in textbooks and some items are only introduced. Therefore, teachers need to prepare practical training data. However, it is not easy to prepare data sets for learning. The content of the data set is reported to be something that high school students are familiar with in order to aid their thinking [55].

This paper proposes a data analysis and learning system aimed at solving these problems.

### 4.3 Design of data analysis learning system

This chapter considers the needs for the following functional requirements based on the tasks and problems described in Chapter 4.2.

To solve Problem 1 that 'learning costs are high because the learning of statistical concepts and analysis tool operations are mixed,' it is necessary to greatly reduce the burden of learning operations for letting students focus on learning statistics.

We considered it necessary to have functions for 'displaying information on representative values corresponding to the analysis', 'displaying outliers and abnormal values that are difficult to notice', 'When students analyze data, they can only select the necessary analysis method from a lot of methods available' and 'They can only operate with a left mouse click.' In addition, these outliers and abnormal values are derived using the quantile deviation method. For example, EZR [56] is a typical example of software that can be operated with only a left mouse click. This is useful because it enables analysis without having to learn difficult commands in R and many papers using it have been reported.

To solve Problem 2 it is necessary to enable students to independently look at various graphs and judge the results of the analysis. The undo function is reported to be suitable for incorporating trial and error learning [57]. However, the undo function has to be executed several times to return to the original state. Therefore, the results of data analysis are displayed separately for each tab, and tabs that are deemed unnecessary can be deleted to return to the state before analysis. It is hoped that this will facilitate trial and error for students through the undo

function.

Problem 3 requires making it easy to share data among students. When other students answer the questions created in the lessons, these answers are stored in this system's database. Data collected on this systems can be accessed as read-only data within the class without any special settings. This data is always retrieved in its most up-to-date state and does not need to be redistributed, even if data is appended.

Problems 4 depends on the installation of software in the learning environment. In recent years, students in Japan have their own terminal. This system is made as a website on the Internet, it can be used from high schools and universities via a web browser.\*<sup>1</sup>

Problem 5 is related to the difficulty in preparation of sample data sets. In order to enable learning through practical training regardless of the type of textbook, it is necessary to provide a data set suitable for high school study items.

## 4.4 Evaluation of Connect DB

In this chapter, it is verified through (1) a comparison of the number of operations and (2) practical training in data analysis for university students that this system enables students to concentrate on learning data utilisation due to its easy-to-learn operation format. Microsoft Excel (hereinafter referred to as 'Excel'), a spreadsheet software used for data analysis, was used as a comparison target for this system since Excel has been frequently used in schools as an analysis tool, such as being introduced in 'Informatics I' textbooks. The comparison was considered appropriate because Excel is a tool that is used frequently in schools, such as being introduced in 'Informatics I' textbooks. In addition, Microsoft Excel is also useful learn on its own and an online version is available that requires no installation.

(1) We considered that the fewer the number of operations required when drawing graphs and formatting data for data analysis, the easier it would be to learn the operations, so we compared the number of operations required when drawing graphs and formatting data using this system and Excel.

(2) Practical training of data analysis using Excel and Connect DB is conducted for university students taking information literacy courses, and the operability of the tools used is compared using a questionnaire after the training. The reason for targeting first-year university students rather than high school students for this evaluation is that they are from a generation that did not learn data analysis in high school. In addition, the operation log recorded in the system will clarify whether the learners were able to engage in learning data utilisation. Comparisons will be made by formatting and visualizing graphs using common data in the practical training.

---

\*<sup>1</sup> [https://cdb\\_wiki.eplang.jp](https://cdb_wiki.eplang.jp)

### 4.4.1 Verification by comparison of the number of operations

#### Verification method

This Chapter assesses whether the system is easy to learn. Therefore, the number of operations for drawing graphs and processing data during data analysis is compared. It is assumed that the system is easy for learners because the fewer the number of mouse and keyboard operations, the less they have to learn.

This time, we will compare the number of operations for ‘histogram’, ‘cross tabulation’, ‘box plots’, ‘scatter plots and correlation coefficient’ and ‘regression analysis’, which are often used for data analysis in Japanese upper secondary schools. However, since Excel offers a variety of paths to reach the desired analysis, the easiest method is selected. the pivot tables were used for ‘histogram’ and ‘crosstabulation’, and graphs display were used for ‘box plots’, ‘scatter plots and correlation coefficient’ and ‘regression analysis.’

The number of operations was counted in the three categories of ‘mouse clicks’, ‘mouse dragging’ and ‘input from keyboard’, and the total was calculated. For ‘mouse clicks’, one left mouse click was counted as one operation. For example, checking a check box or pressing a button is counted as one click. Mouse dragging was counted as one mouse drag or drag and drop. For example, the action of selecting a range of cells or putting data into a pivot table field is counted as one click. ‘Keyboard input’ was counted once for each input point. For example, entering a function in one cell counted as one time.

#### Verification results and discussion

The results of the comparison of the number of operations between this system and Excel are shown in Table 4.1. It was found that this system was able to perform the graphing and data processing necessary for data analysis with fewer operations than Excel in all the items treated in this study. In particular, in the creation of ‘histograms’, the number of operations for this system was 3 compared to 19 for Excel. This is because this system can be operated only with a ‘mouse click’ and automatically divides the histogram into 10 parts and tabulates them in the system, whereas Excel requires a range of data to be specified when tabulating, resulting in more ‘mouse clicks’ and ‘input from the keyboard’ operations.

From these results, it was found that the system’s features, such as ‘selecting data for analysis using check boxes in columns’ and the function ‘automatically displaying representative values required for analysis’, enable the drawing of graphs and data processing required for data analysis with fewer operations than in Excel.

Table 4.1: Comparison of operation counts of Connect DB and Excel.

|                   | Connect | Excel |      |          |       |
|-------------------|---------|-------|------|----------|-------|
|                   | DB      | Click | Drag | Keyboard | Total |
| Histogram         | 3       | 13    | 4    | 2        | 19    |
| Cross tabulations | 3       | 3     | 4    | 0        | 7     |
| Box plots         | 5       | 4     | 3    | 5        | 12    |
| Scatter plots     | 5       | 5     | 3    | 1        | 9     |
| Simple regression | 6       | 8     | 1    | 0        | 9     |
| Total             | 22      | 33    | 15   | 8        | 56    |

Table 4.2: Contents of classes.

| session | used tool  | learning contents   |
|---------|------------|---|
| 1       | Excel      | cross-tabulations(pivot table)<br>displayed graph(bar graph/ pie graph/ band graph) |
| 2       | Connect DB | cross-tabulations<br>displayed graph(bar graph/ pie graph/ band graph)              |

#### 4.4.2 Verification through practice

##### Practical training content

To confirm the operability of the system and its potential for classroom use, practical training was conducted as part of a lecture for 214 university students taking an information literacy course. The practical training was conducted in two sessions (90 minutes x 2): the first session was conducted using Excel, and the second session was conducted using Connect DB. Many of the students had used Excel before but Connect DB was new to them. Although students are more familiar with Excel than with Connect DB, it is possible that they understood the learning content better the second time around rather than the first time.

In the course of the lecture, the first session dealt with blood type data and weather data in order to learn how to draw graphs using spreadsheet software, while the second session dealt with blood group data, birth count data and average temperature data in order to experience data analysis using open data.

As the data handled were not exactly the same, it was decided to make a comparison of the content conducted using common data in both panels. Table 4.2 shows the content of the classes conducted using the common data. Table 4.3 shows the blood group data from 50 Japanese and 50 French people were used.

In the practical training, cross-tabulations were first performed on the data in Table 4.3 to create a tally table by nationality and blood type. The data were then visualised and analyzed graphically for ‘Number of Japanese by blood type’, ‘Percentage of Japanese by blood type’ and

Table 4.3: Example data used in the class.

| No  | nationality | blood type |
|-----|-------------|------------|
| 1   | Japan       | B          |
| 2   | Japan       | A          |
| 3   | France      | O          |
| ⋮   | ⋮           | ⋮          |
| 100 | Japan       | B          |

|            |  |
|------------|--|
| Question 1 | Which did you find easier to use when tabulating data and drawing graphs, the Excel you used or Connect DB?<br>[ options ]<br>· Connect DB was easier to use.<br>· Connect DB was rather easier to use.<br>· Excel was rather easier to use.<br>· Excel was easier to use. |
| Question 2 | Please feel free to describe why you think so.   |

Fig. 4.1: Contents of the questionnaire.

‘Percentage of Japanese and French people by blood type.’ A bar chart was drawn for ‘Number of Japanese by blood type’, a pie chart for ‘Proportion of Japanese by blood type’ and a band chart for ‘Proportion of Japanese and French people by blood type.’

#### Analysis of questionnaire results

A questionnaire was sent to the students after the second practical session. There were 198 responses to the questionnaire. This time, 181 responses from students who attended both sessions were analyzed.

Fig. 4.1 shows the questionnaire questions. Question 1 was answered in a multiple choice format (four-question method), while Question 2 was answered in an open-ended format. Question 1 asks whether Excel or the proposed system is easier to use. The reason for adopting the four-question method for this question is that in Japan there is an unconscious tendency to give a neutral evaluation. Question 2 asks the reason for the choice made in Question 1.

The aggregate results of the questionnaire responses are shown in Fig. 4.2. Aggregating the responses to Question 1, 75 respondents (41.4%) said that Connect DB was easier to use, 60 respondents (33.1%) said that Connect DB was rather easier to use, 31 respondents (17.1%) said that Excel was rather easier to use and 15 respondents (8.3%) said that Excel was easier to use. This shows that 76.6% of the respondents stated that the proposed system was easier to

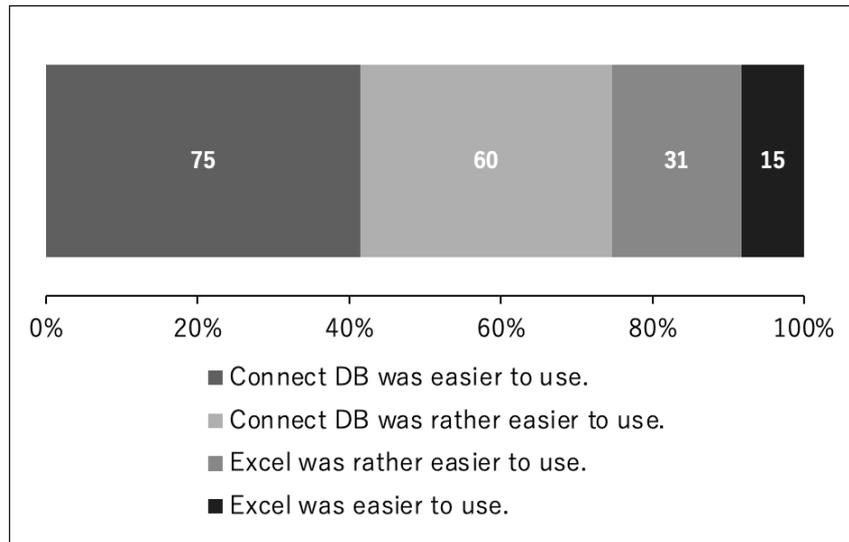


Fig. 4.2: Results of the questionnaire.

Table 4.4: Classification of reasons for responses. (N=181, multiple choice)

| options                              | operations         |             |           |            | functions           |                |         | others |
|--------------------------------------|--------------------|-------------|-----------|------------|---------------------|----------------|---------|--------|
|                                      | easy to understand | easy to use | procedure | experience | recommended methods | display Graphs | freedom |        |
| Connect DB was easier to use.        | 17                 | 34          | 11        | 1          | 8                   | 17             | 0       | 4      |
| Connect DB was rather easier to use. | 13                 | 23          | 13        | 0          | 4                   | 19             | 1       | 4      |
| Excel was rather easier to use.      | 3                  | 0           | 1         | 14         | 0                   | 6              | 9       | 3      |
| Excel was easier to use.             | 1                  | 1           | 0         | 8          | 0                   | 1              | 5       | 2      |

use and 25.4% of the respondents stated that Excel was easier to use.

Table 4.4 shows the resulting totals for the reasons for the answers for each option described in Question 2. Within the framework of the Technology Acceptance Model (TAM) [58], ‘operations’ corresponds to ‘Perceived Easy of Use’, and ‘functions’ corresponds to ‘Perceived Usefulness.’ Classification was done by checking the descriptions and classifying descriptions relating to the easy of operation into ‘easy to understand’, descriptions relating to the easy of operation into ‘easy to use’, descriptions relating to the number of operations and operation time into ‘procedure’ and descriptions relating to familiarity with the operation into ‘experience.’ Descriptions relating to the system’s function of ‘presenting recommended analysis methods based on the type of data to be analyzed’ were classified as ‘recommended methods.’ Furthermore, descriptions relating to graph drawing were classified as ‘display graphs’, descriptions relating to the degree of freedom of operation were classified as ‘freedom’, and those that did not fit into these categories were classified as ‘other’ and tabulated. If a description fell into more than one category, it was added to all of them.

The most common reason given for answering that ‘Connect DB is easier to use’ was easy of operation, such as ‘easy of importing external files’ and ‘no need to select a range of data’, with

34 responses, accounting for 37.0% of the reasons given for answering that ‘Connect DB is easier to use.’ In addition, 17 (18.5%) of the responses related to easy of operation and graph drawing, such as ‘operation is simple and easy to understand’, 11 (12.0%) related to fewer procedures when analyzing data, such as ‘work time is shorter compared to Excel’, and 8 (12.0%) related to the graph drawing suggestion function, such as ‘it suggests appropriate graphs.’ The function to suggest graphs to be drawn, such as ‘it suggests appropriate graphs’, was described in 8 cases (8.7%).

Similarly, 23 (30.0%) of the respondents stated that Connect DB was easier to use, 19 (24.7%) stated that it was easier to draw graphs, 13 (16.9%) stated that it had fewer steps and was easier to understand when analyzing data, and 4 (4.7%) stated that it was easier to draw graphs. The number of respondents who stated that they would rather use Excel for data analysis was 13 (16.9%), and the number of respondents who stated that they would rather use Excel for data analysis was 13 (16.9%).

As for the reason for answering that ‘Excel is rather easier to use’, 14 (38.9%) of the responses related to familiarity with the operation, such as ‘I am used to using Excel more’, and 9 (25.0%) related to customizability, such as ‘I can change the graphs to make them easier to read’ and ‘I can set detailed settings.’ There were also a small number of statements such as ‘Excel can save graphs’, ‘data and graphs can be viewed at the same time’ and ‘uploading analysis data is time-consuming’. As for the reason for answering that ‘Excel is easier to use’, 8 (44.4%) stated that it was related to familiarity with the operation, and 5 (27.8%) stated that it was related to the customizability.

This indicates that the respondents stated that Connect DB was easier to use than Excel when practicing data analysis, and that the system’s features such as ‘easy of operation’ and ‘few operating procedures’, as well as the ‘ability to suggest graphs’, influenced the easy of use of the tool when analyzing data. The fact that Excel was selected for the reason of ‘familiarity with the operation’ suggests that more students may respond that the proposed system is easy to use after using it several times and becoming familiar with it.

Regarding the relationship between easy of operation and learning, among the three cognitive loadings in the cognitive load theory [59], the intrinsic (related to educational content) is considered common regardless of the tool, since the educational content of data processing does not change. In the proposed tool, the results of each operation are displayed on a tabbed screen, and the results can be easily restored by deleting the tab, so it was possible to learn through trial and error by considering appropriate processing while looking at the data. As for extraneous (other than that), in Excel, because it is a general-purpose tool, one has to select the ‘Insert’ menu to perform the ‘cross tabulation’ process and needs to understand terms such as ‘pivot table’, ‘specify row names’, ‘specify column names’, etc. In the proposed tool, data processing operations such as ‘check column names’ and ‘cross tabulation’ are displayed the same as seen on the buttons, etc., thus reducing the cognitive load.

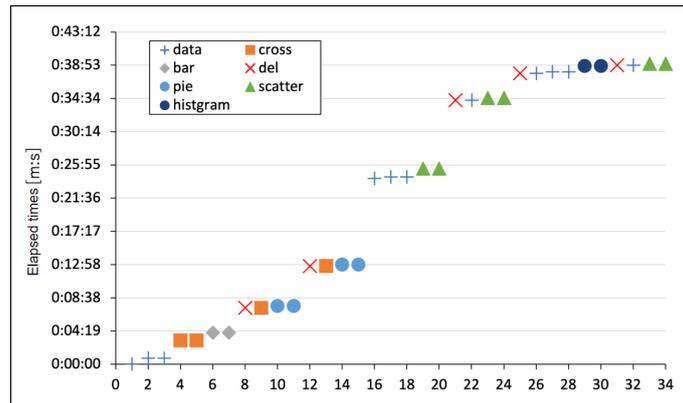


Fig. 4.3: Operational history of a learner who answered that Connect DB was easier to use.

### Analysis of learning activities by operation logs

An analysis of the Connect DB operation log was conducted to check whether the learners were able to engage in learning to utilise the data. The operation logs record the statistical processing, graph drawing, etc. selected for each analysis operation by the learner. There were a total of 16,124 analysis operation logs recorded in the system, with an average of 75.3 operations per person.

Fig. 4.3 shows the operations of one learner who answered that Connect DB was easier to use. The vertical axis shows the time elapsed since the first operation and the horizontal axis shows the number of operations. The dots plotted on the graph indicate that some kind of operation was performed.

While the statistical processing and graphing operations were carried out as the practical training progressed, several operations of ‘reading data’, ‘histogram’, ‘deleting tabs’ and ‘scatter plots’ were carried out in 1 or 2 seconds before and after 00:38:53. This indicates that the learner is able to analyze the data with short trial and error intervals.

Fig. 4.4 shows the operation log of one learner who stated that Excel was easier to use. A total of 89 operations were performed. The operations were carried out in accordance with the progress of the practical training, and the learner also tried various analysis methods such as box plots, averages and histograms. In particular, 30 operations were performed between 0:43:12 and 0:47:31, a period of less than four minutes, and a check of the contents suggests that they were trying out box-and-whisker diagrams, averages and histograms, and comparing the differences in the way the data looked depending on the graphs they were drawing.

This suggests that learners were able to use Connect DB and learn while experimenting with different analysis methods and graph drawings in a short time.

### Student Survey Descriptions

After the class, 167 students responded to the “What I learned in this class” survey question; 100 students commented on Connect DB. Examples of specific responses include, “I learned that



## 第 5 章

# 高等学校において重回帰分析を扱うデータ分析学習ツールと授業の提案

### 5.1 はじめに

前章では、学習者にとって分析ツールの操作が難しい可能性がある点を解決するため、データ分析学習システム Connect DB の操作の容易性を明らかにした。

本章では、情報 II において重回帰分析やロジスティック回帰、クラスタリングといった機械学習の内容が扱われることに焦点を当てる。授業では、学習に適した分析ツールやデータセットが必要となるが [8] [9]、機械学習の内容に活用可能な学習ツールやデータセットは実践報告多くないといった課題がある。これを克服するため、本研究では重回帰分析に着目し、高校生が授業の中で重回帰分析を扱うデータ分析学習ツールの開発を行い、その特徴を学習できるデータセットを検討する。大学生を対象とした教育実践から得られた理解度をもとに、提案した機能やデータセットの実用性を明らかにする。

### 5.2 共通教科「情報」におけるデータ活用の学習

高等学校の共通教科「情報」では、平成 30 年に告示された学習指導要領解説 (情報編) [2] において必履修科目として「情報 I」、選択科目として「情報 II」が設置された。データ活用の内容は、主に情報 I では「(4) 情報通信ネットワークとデータの活用」、情報 II では「(3) 情報とデータサイエンス」にて扱われる。情報 I の教科書 12 冊と情報 II の教科書 3 冊で扱われる内容とその割合を表 5.1 に示す。情報 I ではデータ分析に関する基本的な内容を中心に扱い、情報 II ではデータのモデル化等の発展的な内容を扱う。

高等学校におけるデータ活用の実践としては、アンケート調査を活用した授業実践 [36] や実践を通じた全校での教科横断によるカリキュラム開発 [37]、機械学習を題材とした実践 [60] 等の報告がある。共通テストサンプル問題 [4] においては、強いサッカーチームと弱いサッカーチームの違いの分析に散布図行列が扱われている。これらの多くは情報 I における内容であり、情報 II は今年度教科書が発刊されたため、今後実践が行われていくことになる。そのため、情報 II で活用可能な学習ツールや実践事例が求められている状況である。

そこで本研究では、情報 II の全ての教科書で扱われ、量的データの予測モデルの基本である重回帰分析に着目した。重回帰分析は、ある事象に対して複数の説明変数から影響を与える変数を予測できる分析手法である。情報 II の教科書では、50m 走のタイムを立ち幅跳び、ハンドボール投げ、握力、上体起こしの記録から予測する例 [61]、各都道府県の高等学校数を総人口と中学校数から予測する例 [62]、アイスの売上個数を平均気温と平均湿度から予測する例 [63] が例示されている。

そこで、本研究では、重回帰分析を学習できるような授業案とデータセットを検討し、それを実現する分析ツールの開発を行う。

Table 5.1: 情報 I/II の教科書で扱う内容

| 項目        | 情報 I(N=12) | 情報 II(N=3) |
|-----------|------------|------------|
| 質的・量的データ  | 12(100.0%) | 3(100.0%)  |
| 欠損値/外れ値   | 12(100.0%) | 3(100.0%)  |
| 散布図/相関関係  | 12(100.0%) | 3(100.0%)  |
| ヒストグラム    | 11(91.7%)  | 1(33.3%)   |
| 分散/標準偏差   | 10(91.7%)  | 1(33.3%)   |
| テキストマイニング | 10(91.7%)  | 0(0.0%)    |
| 箱ひげ図/四分位数 | 8(66.7%)   | 1(33.3%)   |
| 単回帰分析     | 7(58.3%)   | 2(66.7%)   |
| 仮説検定      | 6(50.0%)   | 1(33.3%)   |
| クロス集計     | 6(50.0%)   | 0(0.0%)    |
| 重回帰分析     | 0(0.0%)    | 3(100.0%)  |
| クラスタリング   | 0(0.0%)    | 3(100.0%)  |
| ロジスティック回帰 | 0(0.0%)    | 2(66.7%)   |

### 5.3 重回帰分析を学習する上での課題

本章では単回帰分析と重回帰分析の違いについて述べた後、重回帰分析を学習する上での課題を整理する。単回帰分析は、図 5.1 のような 1 つの説明変数から予測を行うモデルである。例えば、「気温が上がればアイスクリームが売れる」のように人間の感覚に合った説明を行うことができる。予測には回帰直線を用いて散布図に対し最小二乗法を利用して残差が最も少なくなるように一次関数の直線を引くことで求めることができる。直線の式を導出するための計算式は高校生にとっては簡単ではないが、散布図を見せながら説明することで、「散布図の中心辺りに直線を引けばよい」という感覚で理解することが可能である。

一方、重回帰分析は図 5.2 のように複数のデータを合成して予測を行う。例えば、不動産物件の家賃を築年数や面積などの複数の説明変数から予測できる。複数の変数を組み合わせる必要のある予測は、人間の感覚だけでは理解しにくい。なぜなら、説明変数が 2 つであった場合、目的変数と合わせて 3 次元のグラフとして可視化することができるが、それぞれの説明変数は単位や数値のスケールも異なるため、生徒はそこからデータの特徴を読み取りにくいからである。さらに、説明変数が 3 つ以

上になるとグラフとして可視化すらできなくなってしまう。このように大学生が学習する上でも簡単ではないものを、高校生が活用できることが求められる。

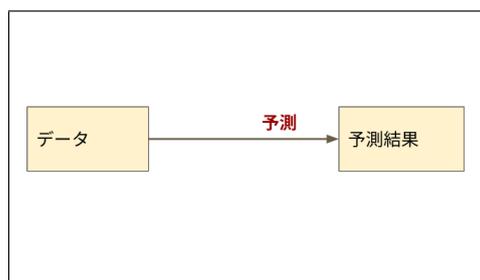


Fig. 5.1: 単回帰分析モデルのイメージ図

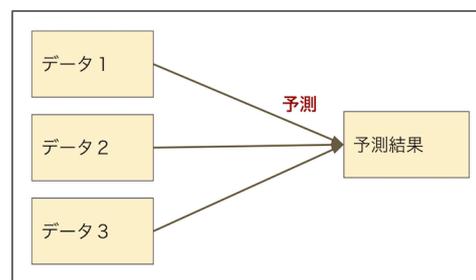


Fig. 5.2: 重回帰分析モデルのイメージ図

高校生が情報 II の授業で重回帰分析を学習する際、一般的に教科書に記載されている表計算ソフトが利用される。生徒は、教科書の手順に従って重回帰分析を行い、図 5.3 の「重決定 R<sup>2</sup>」や「補正 R<sup>2</sup>」から決定係数、「係数」から偏回帰係数の数値を取得する。生徒は、授業の中で説明変数を変更しながら決定係数からモデルの評価を行うが、数値の変化にどのような影響があるのかを想像することは難しい。そこで、後述する Connect DB で採用した図 5.7(a)(b) のような散布図を利用することで、説明変数による決定係数の値が目的変数の値をどの程度予測できているかを視覚的に確認することができるように工夫した。

また、重回帰分析の処理は、内部でどのような計算が行われているかを想像することは難しい。そこで、散布図を利用して各説明変数の目的変数との相関を調査することで、重回帰分析との関係を考察する学習が有効だと考えた。しかし、多くのデータの散布図を表示する散布図行列は、教科書に載っている表計算ソフトは対応していないため、散布図行列と同様の比較をするためには多くの散布図を表示する必要がある、操作が煩雑になってしまう問題があった。代替手段として考えられる R や Python では、散布図行列を表示する機能が備わっているが、高校生が授業の中で扱うには難易度が高いという問題がある。

そこで、本研究では高校生が重回帰分析を学習する上で「自分の経験と実際のデータを分析したことから得られた特徴の相違に気づく」ことや「説明変数を複数にすることで予測モデルの精度が単回帰分析よりも高くなる」こと、「複数の説明変数が目的変数に与える影響を知る」こと、「影響が少ない変数を除いてモデルを最適化できる」ことを理解して、他のデータに応用できる力を身につけさせる授業案を検討する。そして、それを実現するためにデータ分析学習ツール Connect DB に散布図行列や重回帰分析の機能を追加した。

## 5.4 Connect DB への追加機能の設計と開発

### 5.4.1 Connect DB の概要

我々はオンライン上でデータを整形・可視化する Connect DB の開発を進めてきた。Connect DB は、学習者が容易な操作でデータ分析の学習に注力できるように、各種統計処理に関するデータの加工やグラフ描画を支援するデータ分析学習ツールである。これまでに、大学や高等学校での実践 [64]

| 概要        |            |            |            |            |            |            |
|-----------|------------|------------|------------|------------|------------|------------|
| 重回帰統計     |            |            |            |            |            |            |
| 重相関 R     | 0.97532295 |            |            |            |            |            |
| 重決定 R2    | 0.95125485 |            |            |            |            |            |
| 補正 R2     | 0.94800517 |            |            |            |            |            |
| 標準誤差      | 3809.55784 |            |            |            |            |            |
| 観測数       | 65         |            |            |            |            |            |
| 分散分析表     |            |            |            |            |            |            |
|           | 自由度        | 変動         | 分散         | 測された分散     | 有意 F       |            |
| 回帰        | 4          | 1.6993E+10 | 4248209036 | 292.722925 | 1.2833E-38 |            |
| 残差        | 60         | 870763854  | 14512730.9 |            |            |            |
| 合計        | 64         | 1.7864E+10 |            |            |            |            |
|           | 係数         | 標準誤差       | t          | P-値        | 下限 95%     | 上限 95%     |
| 切片        | 38029.5014 | 1792.91546 | 21.2109841 | 1.436E-29  | 34443.1365 | 41615.8663 |
| 駅からの時間[分] | -309.87825 | 130.714248 | -2.3706539 | 0.02098651 | -571.34568 | -48.410822 |
| 築年数[年]    | -589.69991 | 30.4998053 | -19.334547 | 1.8508E-27 | -650.7086  | -528.69121 |
| 面積[m2]    | 866.136694 | 31.1139832 | 27.8375382 | 5.2695E-36 | 803.899461 | 928.373927 |
| 階         | 515.332274 | 307.803533 | 1.67422469 | 0.0992921  | -100.36646 | 1131.03101 |

Fig. 5.3: 表計算ソフトでの重回帰分析の実行画面

がある。図 5.4 に Connect DB の学習画面例を示す。この例では、架空の生徒データを使って所属クラブ別に身長四分位数を求め、箱ひげ図を作成した。箱ひげ図の上にある「読込、削除、四分位数、箱ひげ」と書かれたタブは、箱ひげ図を描くまでに行った処理を表している。箱ひげ図の下にある表は、四分位数を表し、一番下の表が分析に使用したデータを表している。

画面に表示された表形式のデータから分析に必要なカラムにチェックを付けて、プルダウンメニューから具体的な分析方法を選択する。処理を行うと新たなタブが生成され、その中に分析結果が表示される。タブを切り替えることで、分析前のデータや途中の分析結果を確認できる。

ここで、Connect DB がデータ活用の学習に提供する支援機能について、具体的な例を紹介しながら説明する。

まず、このツールは高等学校の授業利用を想定している。教科書で採用されている表計算ソフトを利用する際、「グラフ表示」、「ピボットテーブル」、「関数」など、分析の目的によって操作方法が異なっている。そのため、分析ツールの操作に慣れていない高校生は、手順書を見ながら操作を行うため時間がかかってしまう可能性があった。この課題に対処するために、データの種類に応じて、適切な分析方法を提示することにし、分析ツールを扱いやすくした。次に、教科書にはグラフと代表値がどちらも表示されている。表計算ソフトでは、これらの結果を出力する際に別々の操作を行っていたため、分析に慣れていない高校生なら代表値を出力し忘れる可能性があった。Connect DB は、表示したいグラフと関連した代表値を同時に表示させることで、学習を進めやすくした。最後に、表計算ソフトでは、図 5.5(a) のような複数の操作で散布図と相関係数を出力する。表計算ソフトの Undo は操作ごとに行われるため、分析前の状態に戻すためには 4 回以上の Undo 操作が必要になってしまう。Connect DB では、図 5.5(b) のように Undo が分析単位で行われる。出力結果を、「×ボタン」をクリックすることで、1 回の Undo 操作で分析前の状態に戻すことができる。また、Undo 操作で戻った状態では、分析対象としてチェックしたデータ名や実行した途中経過が残っているため、生徒は少ない操作で「同じデータを対象に異なった分析方法を試す」、「異なったデータを対象に同じ分析方法を試す」といった分析を行うことが可能になった。

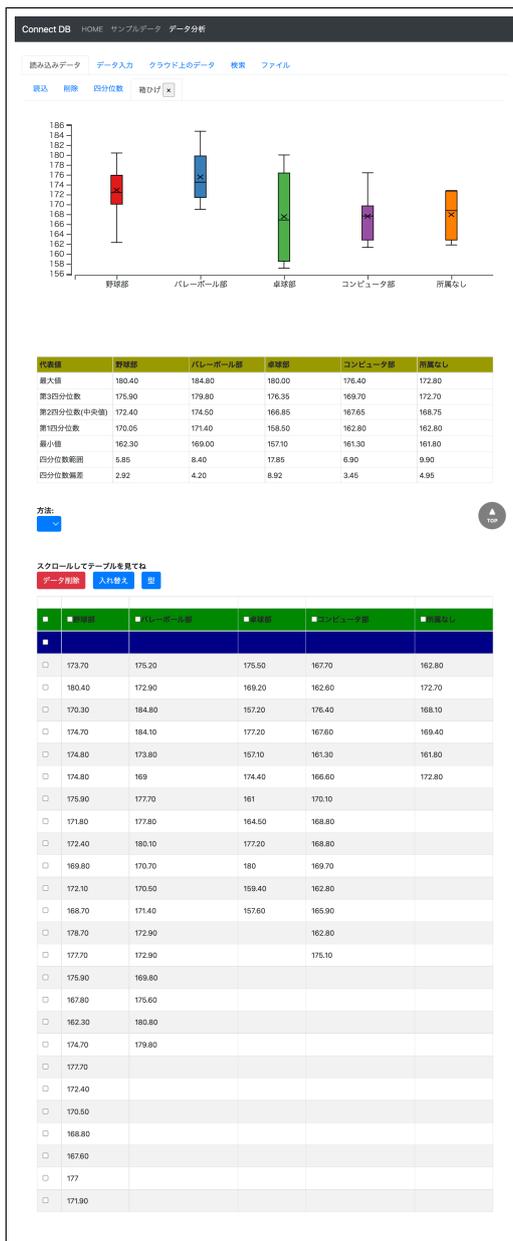


Fig. 5.4: Connect DB の学習画面例

### 5.4.2 重回帰分析機能の検討

本節では Connect DB に追加する機能を検討する。重回帰分析を学習する上で、目的変数に対して説明変数が与える影響を考察し、仮説を立てる必要がある。そこで、実際のデータを使って各説明変数の散布図やヒストグラムを描画して観察することが有用な手段だと考える。ただし、説明変数の数が多い場合は、それだけツールの操作回数が多くなってしまいうことから、説明変数ごとの散布図とヒストグラムを同時に観察できる散布図行列が描画できるとよいと考えた。例えば、説明変数の中に相互に独立していない多重共線性のある要素があったとしても、生徒たちは、散布図行列から依存関係を見つけることができる。ただし、初めてデータ分析を行う場合、高校生にとって図 5.6 のよ

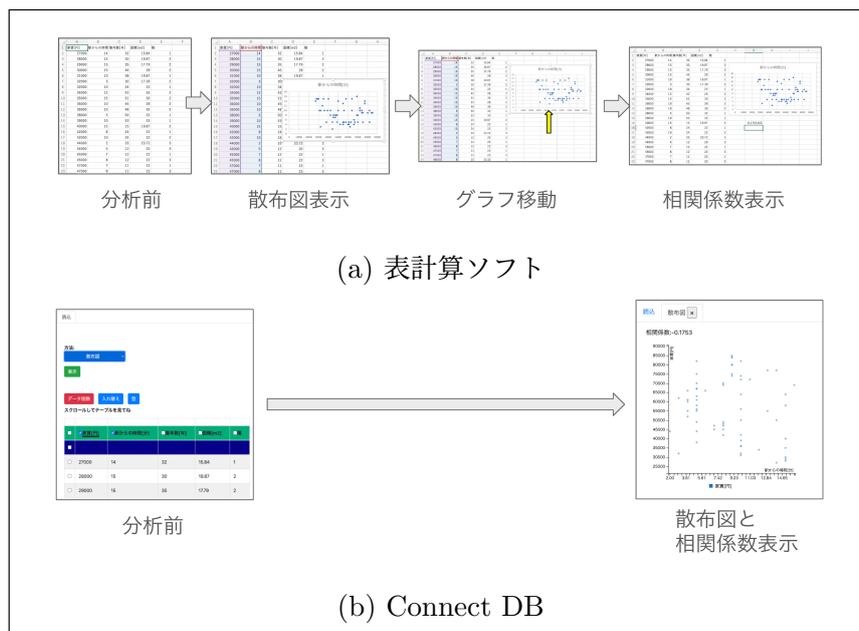


Fig. 5.5: 表計算ソフトと Connect DB の分析手順の比較

うな散布図行列は情報量が多すぎるため、その読み方がわからないことも考えられる。その場合、1つ1つの散布図を観察するところから始めることは有効である。複数の散布図を表示するためには、Connect DBのUndo機能を利用すれば、分析前の状態に戻せるため「異なったデータを対象に同じ分析方法を試す」といった活動が行いやすい。

重回帰分析の学習では、予測の精度を表す決定係数と偏回帰係数を示す式(回帰式)を導くことができる。散布図行列などを利用して十分に目的変数に対する説明変数との関係を考察した後であれば、重回帰分析の予測から説明変数の影響力を評価できるはずである。さらに、重回帰分析のモデルがデータにどれだけ適合しているかを評価するために、「横軸に回帰式に説明変数を代入した目的変数の予測値と、縦軸に実際の値を持った散布図(Observed-Predicted Plot)」を利用すればよいと考えた[65]。ここでは、図5.7(a)のような4つの説明変数を選択し、重回帰分析を行った場合、決定係数が高く、散布図も直線に近い形となる。それとは逆に説明変数を2つ選択した場合、決定係数が低く散布図が広がった形になっている。このように、決定係数の高さが視覚的にわかる工夫を行った。

また、データ分析の学習では、学習に使用するデータも重要となる。使用するデータは、その背景等もイメージしやすくなるように生徒にとって身近な題材だとよい。

これらのことから、次の3つの機能を実装することにした。

- (1) 散布図行列を表示する機能
- (2) 重回帰分析を行う機能
- (3) 学習に適したサンプルデータの提供

(1)は散布図行列と相関係数を表示する機能である。各変数について散布図行列を描画すると共にそれぞれの相関係数を表で表示する。(2)は重回帰分析の結果を表示する機能である。決定係数と回帰式を表示し、「回帰式に各説明変数の値を代入した目的変数の予測値」と「目的変数の実際の値」を散布図で表示することで残差を目視で評価できる。(3)はサンプルデータを提供する機能である。生

徒にとって身近な題材かつ重回帰分析の学習に適したサンプルデータを提供する。

### 5.4.3 重回帰分析機能の実装

前節で述べた(1)~(3)を Connect DB に実装した。「(1) 散布図行列を表示する機能」については、分析に使用するデータから数値データを3カラム以上選択すると、図5.6のような散布図行列を表示できるようにした。散布図行列の下に表示されているカラム名は横軸、右に表示されているカラム名は縦軸である。同じカラムのグラフではヒストグラムを表示し、異なるカラムでは散布図を表示した。

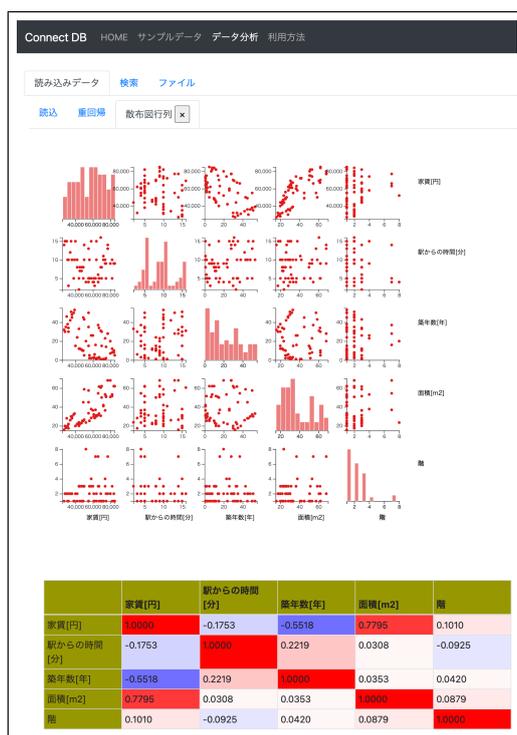


Fig. 5.6: Connect DB で描画した散布図行列

「(2) 重回帰分析を行う機能」については、数値データを3カラム以上選択することで重回帰分析を行えるようにした。このとき下線が引かれた一番左のカラム名が目的変数となり、それ以外が説明変数となる。カラムの位置を入れ替えることで、目的変数を変更することができる。もし選択したカラムが2つの場合は、自動的に単回帰分析を行う。

図5.7(a)はマンションの「家賃」、「駅までの時間」、「築年数」、「面積」、「階数」のデータについて家賃を目的変数として重回帰分析を行った結果である。横軸は重回帰分析の回帰式から各説明変数の値を代入し、目的変数を予測した値で、縦軸は目的変数の実際の値である。回帰式を知りたい場合は「回帰直線の式を表示」ボタンを押せば表示することができる。回帰式の表示は変数ごとの係数について、負の値であれば赤字、正の値であれば青字で表示することで正負の係数がひと目で分かるようにした。

「(3) サンプルデータの提供」については、「賃貸の家賃」、「50m 走のタイム」、「ノート PC の価格」

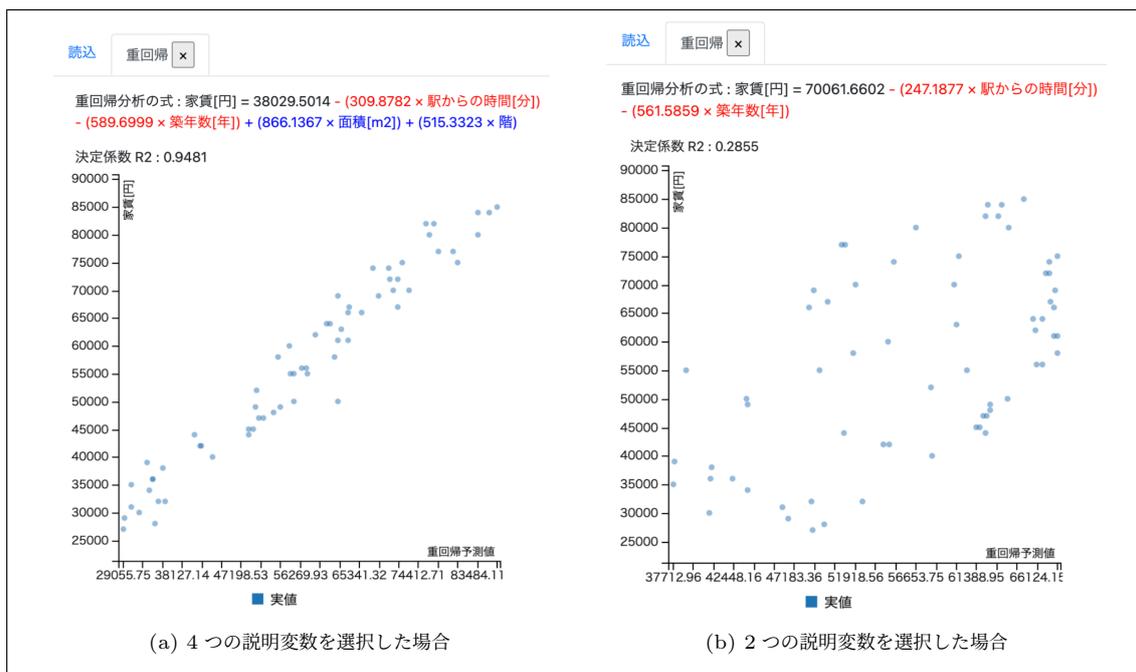


Fig. 5.7: Connect DB による重回帰分析の結果

の3種類のサンプルデータを用意した。

「賃貸の家賃」には、マンションの家賃、駅までの時間、築年数、面積、階数のデータが65件含まれる。RやPythonで利用できるデータセット「ボストンの住宅価格」や、情報Iの教員用教材の「中古住宅」でも使われているため[5]、高校生が想像できそうなデータを作成した。

「50m走のタイム」には、50m走のタイム、身長、靴のサイズのデータが180件含まれる。教科書にも身体能力が掲載されているため馴染み深いデータであると感じ作成した。これらのデータから50m走のタイムの予測を行う学習活動が可能である。

「ノートPCの価格」には、ノートパソコンの値段、CPUのスコア(PassMark)、メモリの容量、SSDの容量、画面の大きさ、重量のデータが49件含まれる。情報Iのコンピュータの仕組みとの連携で、各パーツの役割や値段を知るきっかけになると考え作成した。これらのデータから、コストパフォーマンスが高いノートパソコンを検討するといった学習活動が可能である。

## 5.5 重回帰分析を扱う授業案の検討

本章では、前章で実装した機能を用いて、授業案の検討を行う。その際、重回帰分析の学習目標である「自分の経験と実際のデータを分析したことから得られた特徴の相違に気づく」ことや「説明変数を複数にすることで予測モデルの精度が単回帰分析よりも高くなる」こと、「複数の説明変数がどの程度目的変数に影響を与えているかがわかる」こと、「影響が少ない変数を除いてモデルを最適化できる」ことを学習できるよう留意した。今回は「賃貸の家賃」のサンプルデータを題材として検討する。マンションを実際に借りた経験がある高校生は多くはないと思われるが、「新しい家に住みたい」「広い家に住みたい」「駅から近いところに住みたい」「泥棒が入ってこないような高い階に住みたい」などの複数の要望を満たす場合は、家賃が高くなる傾向にあることを想像できると考えた。授

業は情報IIの全3時間(50分×3)で設計した。学習者は情報Iでデータ分析を学習しており、ヒストグラムや箱ひげ図、散布図、単回帰分析について学習済みの高校生を想定する。授業は「(A) 仮説を立てる」、「(B) 単回帰分析による予測を行う」、「(C) 重回帰分析による予測を行う」の3段階にわけ、1時間目に(A)と(B)を、2、3時間目に(C)を扱う。

第1時間目では、学習者は自らが立てた仮説について単回帰分析を行い確認することで「1つの変数だけでは結果を正確に予測することは難しい」ことを体験する。

まず導入として、学習者に、自分が住む部屋を選ぶとしたらどの項目を重視するかを考えさせる。授業内容を自分ごととして捉えて、重視する説明変数は何であるかの仮説を立てさせた。次にサンプルデータ(表5.2)を提示し、実際のデータでは、どの説明変数を利用すると適切な家賃を予測することができるかを考えさせる。導入で考えた仮説が正しいかを単回帰分析を行い、確認する。例として図5.8に家賃と駅までの時間のデータを利用して単回帰分析を行った結果を示す。結果からは相関が無いように見える。一般的には駅から近いほうが家賃は高くなる傾向になる場合が多いが、他の説明変数にも影響されるため、この結果から正しく予測するには単回帰分析は妥当ではない。このように、自分の経験と実際の分析との違いを体験させることが重要である。他の説明変数で行う場合においては、図5.6のようにすべての説明変数を用いて散布図行列を描画し、その特徴を観察すればよい。例えば「築年数が増えると家賃が安くなる」ことや「面積が増えると家賃が高くなる」といった関係を読み取ることができる。このように、相関関係がありそうな「面積」の項目も発見できるが「面積が広くて安くても古いマンションに住みたいか」の質問を投げかけることで、学習者に「複数の説明変数を組み合わせる必要がある」ということに気づかせると共に、「1つの変数だけでは結果を正確に予測することは難しい」こと「自分の経験とは異なった分析結果を発見した」ことを体験させる。

第2時間目では、学習者は、重回帰分析を行い、説明変数から家賃の予測を行う(図5.7)。決定係数が単回帰分析よりも高くなることから、「説明変数を複数にすることで予測精度が高くなった」ことを学習する。分析を行った際に表示される回帰式の意味を読み取らせ、家賃を安くする項目を考えさせ、「駅からの時間」や「築年数」が増えると家賃が減ることがわかる。これらの活動により、「複数の説明変数が目的変数に与える影響は重回帰分析によってわかる」ことを学習する。

第3時間目では、学習者は「影響が少ない変数を除いてモデルを最適化できる」ことについて体験的に学習する。まず、重回帰分析の説明変数を増減し、再分析した際に表示される決定係数の変化を確認する。これにより学習者は説明変数を増やしていくごとに予測の精度がよくなっていくことに気づくと考える。その後、目的変数に影響が低い項目について考えさせる。これにより学習者は除いても問題のない説明変数があることに気づき、家賃の予測モデルの最適化を行うことができると考える。

## 5.6 大学生を対象とした授業実践

### 5.6.1 実践の概要

前節で検討した授業案で授業が可能なことを、情報IIの実施前に大学の授業で確認した。授業では「自分の経験と実際のデータを分析したことから得られた特徴の相違に気づく」ことや「説明変数を複数にすることで予測モデルの精度が高くなる」こと、「複数の説明変数がどの程度目的変数に影響

Table 5.2: 賃貸家賃のサンプルデータ

| No. | 家賃 (円) | 駅からの時間 (分) | 築年数 (年) | 面積 (m <sup>2</sup> ) | 階数 (階) |
|-----|--------|------------|---------|----------------------|--------|
| 1   | 69000  | 8          | 1       | 30.64                | 1      |
| 2   | 67000  | 5          | 34      | 61.56                | 3      |
| 3   | 62000  | 3          | 6       | 28.26                | 1      |
| ⋮   | ⋮      | ⋮          | ⋮       | ⋮                    | ⋮      |
| 65  | 85000  | 9          | 5       | 62                   | 1      |

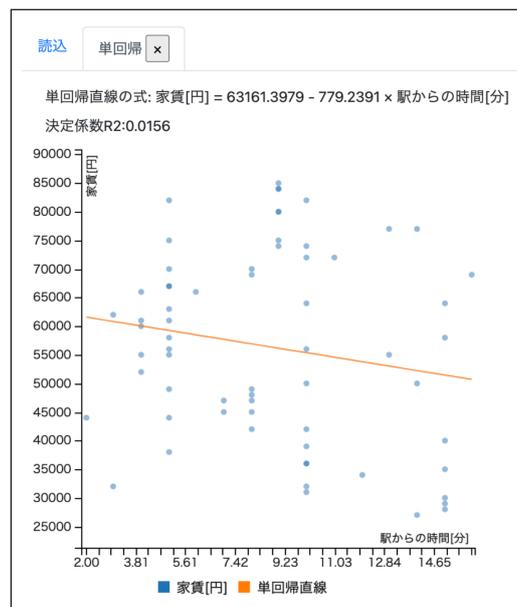


Fig. 5.8: 単回帰分析の例 (家賃と駅までの時間)

響を与えているかがわかる」こと「影響が少ない変数を除いてモデルを最適化できる」ことを学習できるかを確認する。検討した授業案による実践を工学部に在籍する大学2年生149名を対象に全2回(90分×2)行い、事前アンケートと授業内に実施した課題の記述内容から上記の重回帰分析の特徴について学ぶことができたかを検証する。

対象の学生はこれまでに単回帰分析や重回帰分析などについて学習していないため、データ分析の知識においては情報Iを学んでいない高校生と同程度であると考えられる。そのため、検討した授業案の内容に入る前に、重回帰分析と散布図行列を読み取るための事前知識としてヒストグラム、箱ひげ図、散布図、単回帰分析についての学習を1回目に行った。2回目に検討した授業案の内容を実施した。

### 5.6.2 実施した事前アンケートと各課題の概要

学生には1回目の授業の前に事前アンケートを実施した。これは、データ分析前の学生の考えを知るために行った。2回目の授業の最後に3つの課題を提示し、各自で取り組ませた後、最後に振り返りとして授業で学んだことを記入させた。事前アンケートと課題内容を表5.3に示す。事前アンケー

Table 5.3: 実施した事前アンケートと課題の内容

| 回 | 項目      | 問い  |
|---|---------|---|
| 1 | 事前アンケート | 自分が住む部屋を選ぶとしたら、どの項目を重視しますか？               |
| 2 | 課題 1    | 散布図行列から分析できるデータの特徴を箇条書きで答えなさい。<br>(自由記述)  |
|   | 課題 2    | 家賃が安くなる傾向がある説明変数を 2 つ選択しなさい。              |
|   | 課題 3    | 重回帰分析に利用しても影響力が低い説明変数を 2 つ選択しなさい。         |
|   | 振り返り    | 重回帰分析の授業を通して理解できたことを箇条書きで答えなさい。<br>(自由記述) |

トと課題 2,3 は選択式であり、提示した選択肢はすべて同一で「駅からの時間」「築年数」「面積」「階数」である。

事前アンケートは、部屋の値段を決めるのに重視する項目を、学生たちの経験から仮説を立てさせる目的で実施した。ここでは、部屋を借りる人のことを想定して、自らが重視する項目のうち上位 2 つを選ぶ。

3 つの課題では、学生が授業を通して散布図行列と重回帰分析の内容を読み取ることができるようになったかを確認する。

課題 1 では、散布図行列から目的変数と説明変数の関係を読み取れているかを確認する。解答は自由記述とし、データの特徴を箇条書きで記入する。ここでは、学生の経験とは関係なく、実際のデータから目的変数と説明変数との間に、単回帰分析や相関関係があるということを読み取れているかに着目し、「築年数が増えるに対して家賃が安くなる」ことや「面積が増えるに対して家賃が高くなる」ことをグラフから読み取ることができた記述を正答とする。

課題 2 では、重回帰分析の回帰式を正しく読み取ることができるかを確認する。ここでは説明変数の値が大きくなると、目的変数の値が増減のどちらであるかということを知ることが重要となる。回帰式は説明変数の係数について、負の値であれば赤字、正の値であれば青字で表示することで正負がひと目で分かるようにしているため、家賃の安さに影響を与える説明変数が「駅からの時間」と「築年数」の 2 つを選ぶことができると考えた。

課題 3 では、重回帰分析の説明変数を変更しながら何度も試行錯誤を繰り返し、決定係数の変化を読み取ることができるかを確認する。正しく決定係数の変化を読み取ることができていれば、選択しなくてもほとんど決定係数に影響を与えない説明変数である「駅からの時間」と「階数」の 2 つを選択できると考えた。振り返りは、主観的に考えて理解できたか否かを確認する目的で実施した。学生は本授業を通して理解できたことを箇条書きで記入した。

### 5.6.3 事前アンケートと各課題内容の分析

学生が回答した事前アンケートと各課題内容を分析し、授業により「自分の経験と実際のデータを分析したことから得られた特徴の相違に気づく」ことや「説明変数を複数にすることで予測モデルの精度が高くなる」こと、「複数の説明変数がどの程度目的変数に影響を与えているかがわかる」こと、「影響が少ない変数を除いてモデルを最適化できる」ことを学習できたかを確認する。

### 学生が賃貸物件で重視する項目の集計

学生が重視する項目について、事前アンケートの集計結果から確認する。図 5.9 に集計結果を示す。多く選択されていた順に「駅からの時間」が 119 件 (39.9%)、「面積」が 94 件 (31.5%)、「築年数」が 69 件 (23.2%)、「階数」が 16 件 (5.4%) であった。このことから、多くの学生は「駅からの時間」や「面積」が家賃と関係があると感覚的に考えていることがわかった。

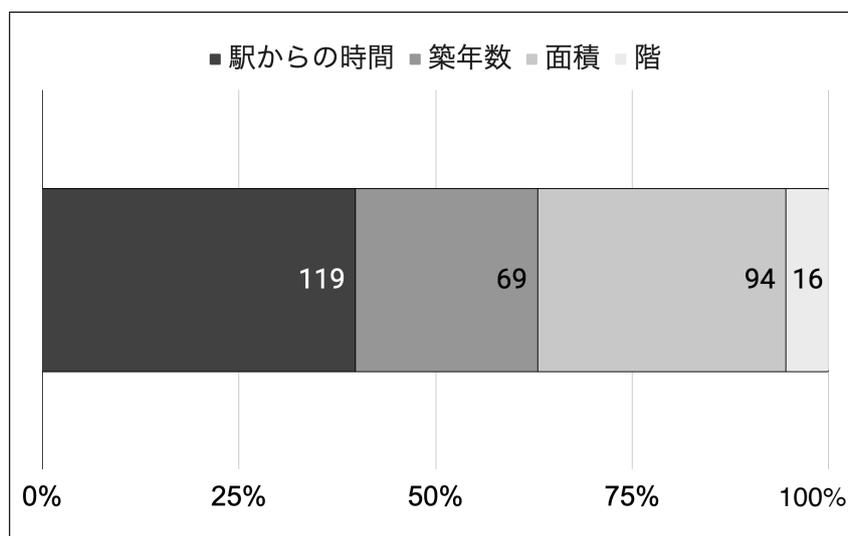


Fig. 5.9: 事前アンケートの結果 (N=149、複数回答可)

### 散布図行列の読み取りに関する検証

散布図行列から目的変数と説明変数の関係を考察できたかを課題 1 の解答結果から確認する。課題 1 の自由記述を 1 文ずつ取り出して分析を行う。ここでは、「自由に分析を行っても良い」と指示したため、学生たちから多様な回答を得られた。その全 809 件の記述内容を分析すると、学生の解答は表 5.4 の A~E の 5 つに分類することができた。A は築年数や面積が家賃に比例していることや相関関係があることを記述したものを分類し、件数は 238 件 (29.4%) だった。B は家賃との関係があまりない項目について記述しているものを分類し、件数は 102 件 (12.6%) だった。C はデータの多さや偏りについて記述したものを分類し、件数は 309 件 (38.2%) だった。D は自らの感覚を記述したものを分類し、件数は 107 件 (13.2%) だった。E は表示されたグラフの情報を記述したものを分類し、件数は 53 件 (6.6%) だった。表 5.4 に各分類の記述内容の件数と例を示す。件数をみると、学生たちは B のような「関係がない」という記述をするよりも、A や C のような「特徴がある」といった表現の方が記述しやすかったことが推測できる。D の記述は、事前アンケートでも多かった「駅からの時間が強く関係する」といった、学生の経験から起こる思い込みが記述されたことが推測できる。E の記述は、散布図行列の見方やデータの分析方法がわかっていないことが推測できる。

これらの分類の中から、散布図行列から単回帰分析や相関関係を踏まえた分類 A の特徴を述べることができた学生は、116 名 (77.7%) だった。このことから、約 8 割の学生が散布図行列から、目的変数と説明変数の関係を実際のデータから分析することができていることがわかる。

Table 5.4: 課題1の解答件数と記述内容 (N=809、自由記述)

| 分類          | 件数 (割合)        | 記述内容  |
|-------------|----------------|---|
| A<br>(相関あり) | 228<br>(29.4%) | ・ 築年数は小さい方が家賃が高い<br>・ 面積は大きい方が家賃が高い           |
| B<br>(相関なし) | 102<br>(12.6%) | ・ 駅からの近さは明確な差はない<br>・ 階に影響はあまりない              |
| C<br>(偏り)   | 309<br>(38.2%) | ・ 階が高い物件は少ない<br>・ 5~7階がない<br>・ 築年数は0の方に集中している |
| D<br>(感覚)   | 107<br>(13.2%) | ・ 広い方が良い<br>・ 駅からの近さの重要性<br>・ 家賃を気にしている人が多い   |
| E<br>(グラフ)  | 53<br>(6.6%)   | ・ 主に散布図が使われている                                |

### 重回帰分析の読み取りに関する検証

重回帰分析の回帰式を読み取ることができたかを課題2の解答結果から確認する。課題2で正しい解答である「駅からの時間」と「築年数」を選択できた学生は119名(79.9%)であった。そして、課題3より予測精度にあまり影響を与えない説明変数を読み取れたかを課題3の解答結果から確認する。課題3で正しい解答である「駅からの時間」と「階数」を選択できた学生は127名(85.2%)であった。課題2と3の両方で正しい選択肢を選んだ学生は104名(69.7%)だった。このことから、約7割の学生が重回帰分析を正しく読み取り、利用することができた。

### 課題1と課題2、3の正誤関係の検証

散布図行列と重回帰分析の両方を読み取ることができた学生数を確認すると共に、課題1と課題2、3の正誤に関係が見られるかを分析する。まず、課題1で正しく特徴を述べることができた学生を「正解者」、そうでない学生を「不正解者」に分けた。そして、それぞれを課題2、3の「両方とも正解」、「課題2のみ正解」、「課題3のみ正解」、「どちらも不正解」に分けて集計を行った。結果を表5.5と図5.10に示す。

課題1の正解者のうち、課題2、3が両方とも正解だった学生は91名(78.4%)だった。約8割の学生は散布図行列から目的変数と説明変数の相関関係を分析でき、複数の説明変数があったとしても、それぞれの影響力を考慮しながら利用することができている。

課題1の不正解者のうち、課題2、3が両方とも正解であった学生は13名(39.4%)で半数に満たなかった。それ以外は20名(60.6%)いることから、散布図行列から目的変数と説明変数の相関関係を分析できなかった学生は、説明変数を複数使用することで新たに見えてくる特徴を見つけられなかった可能性がある。

Table 5.5: 課題1と課題2、3の関係 (N=149)

| 分類      | 課題1       |           | 合計  |
|---------|-----------|-----------|-----|
|         | 正解者       | 不正解者      |     |
| 両方とも正解  | 91(78.4%) | 13(39.4%) | 104 |
| 課題2のみ正解 | 9(7.8%)   | 6(18.2%)  | 15  |
| 課題3のみ正解 | 14(12.1%) | 9(27.3%)  | 23  |
| どちらも不正解 | 2(1.7%)   | 5(15.2%)  | 7   |
| 合計      | 116       | 33        | 149 |

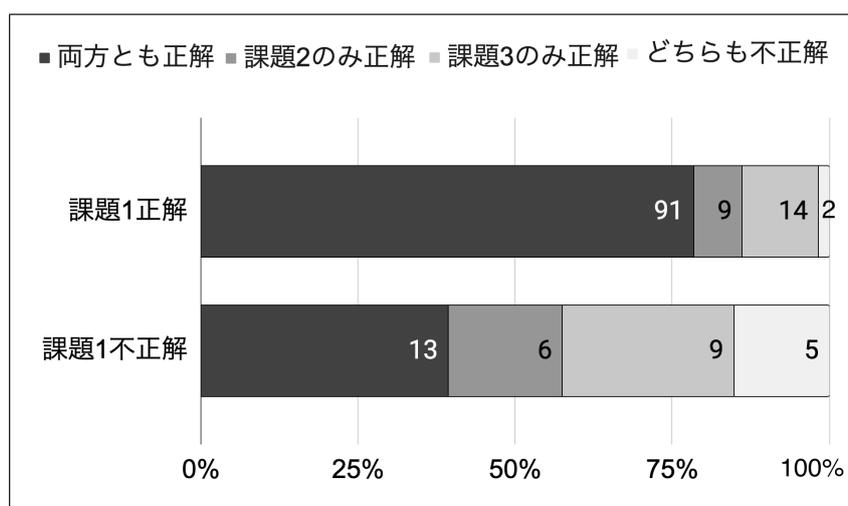


Fig. 5.10: 課題1と課題2、3の関係の比較 (N=149)

### 振り返りの分析

学生が記述した振り返りの内容を分析する。表 5.6 に記述内容の一部を示す。学生 A、B の解答からは、「複数のデータを組み合わせる」、「さらに正確にデータを予測できる」といった重回帰分析についての記述があった。さらに学生 C の解答からは、「影響力が小さい値は省いても良い」といったモデルの最適化に関する記述があった。そして、学生 D の解答からは、「予想を超えた結果が得られた」といった、自分の結果と実際のデータ分析との結果の相違についての記述があった。最後に学生 E のように、「データ同士の関係を考えることが大切」といったデータ同士の関係に印象が残っている学生もいた。

これらの記述から学生は授業を通して、「自分の経験と実際のデータを分析したことから得られた特徴の相違に気づく」ことや「説明変数を複数にすることで予測モデルの精度が高くなる」こと、「複数の説明変数がどの程度目的変数に影響を与えているかがわかる」こと、「影響が少ない変数を除いてモデルを最適化できる」ことを学ぶことができたことが示唆される。

Table 5.6: 振り返りの記述 (一部抜粋)

| 学生 | 記述内容                                  |
|----|---------------------------------------|
| A  | 複数のデータから、さらに正確なデータを予測できることを知った.       |
| B  | 複数のデータを組み合わせて分析する方法を学べた.              |
| C  | 重回帰分析では影響力の小さい値は省いても問題ないことがわかった.      |
| D  | 実際にデータを分析すると、自分の予想を超えた結果が得られることがわかった. |
| E  | データ同士の関係を考えることが大切だと知ることができた.          |

## 5.7 考察

学生が取り組んだ事前アンケートと3つの課題、振り返りの分析により、検討した授業案が「自分の経験と実際のデータを分析したことから得られた特徴の相違に気づく」ことや「説明変数を複数にすることで予測モデルの精度が高くなる」こと、「複数の説明変数がどの程度、目的変数に影響を与えているかがわかる」こと、「影響が少ない変数を除いてモデルを最適化できる」ことを学習できるものとなっているかを確認した。

散布図行列の特徴を読み取る課題1と重回帰分析の特徴を読み取る課題2、3を両方とも正解した学生は約8割いた。このことから、約8割の学生は今回の授業を通して「説明変数を複数にすることで予測モデルの精度が高くなる」こと、「複数の説明変数がどの程度、目的変数に影響を与えているかがわかる」ことを学習することができたと考えられる。初めて散布図行列を利用した学生にとっては、情報が多すぎため理解が難しかったと考えている。そこで授業では、1つ1つの散布図を表示して考察をさせていた。これは、Connect DBの分析単位のUndo機能を利用することで、「異なったデータを対象に同じ分析方法を試す」活動が行いやすくなり、短期間で効果的な学習が可能になったためと考えられる。

また、学生が記入した振り返りを確認すると、「重回帰分析では影響力の小さい値は省いても問題なかった。」との記述が見られることから、授業内で重回帰分析の説明変数を増減し、決定係数の変化を確認することと合わせて、影響が低い項目を考えたことを通して、「影響が少ない変数を除いてモデルを最適化できる」ことを学習することができたと考えられる。今回の授業では、重回帰分析のモデルがどれだけデータに適合しているかを評価する際に、Observed-Predicted Plotを利用した。この方法により、目的変数に対する説明変数による決定係数の値の影響を視覚的に読み取ることができたからだと考えられる。

今回、情報IIを想定して検討した全3回(50×3)の授業案を大学生が90分で取り組むことができた。これらの結果から、Connect DBに追加した機能は重回帰分析の意味の理解をサポートしており、高等学校等の授業で利用可能であることと、検討した授業案によってデータ分析を学んでいない大学生が重回帰分析を学習することが可能であることを確認できた。

## 5.8 おわりに

本研究では、情報 II で扱われる分析手法の 1 つである重回帰分析に着目し、授業の中で重回帰分析の意味の理解や活用方法をサポートした分析ツールの開発と重回帰分析の特徴が学習可能な授業案を検討した。

筆者らが開発した Connect DB に重回帰分析と散布図行列の機能を追加することで、重回帰分析を扱える分析ツールを実現した。また、情報 II において全 3 回の授業案 (50 分×3) を検討した。工学部の大学 2 年生 149 名を対象に検討した授業案による実践を実施し、学生が取り組んだ事前アンケートと 3 つの課題、振り返りから学習効果について検証を行った結果、約 8 割の学生が重回帰分析と散布図行列の特徴を読み取ることができた。また、実践は 90 分で取り組むことができた。

このことから、Connect DB に追加した機能は、授業の中で重回帰分析の意味の理解や活用方法をサポートしており、検討した授業案によって学習が可能であることを確認できた。今後は実際に高等学校で実践を行い、効果検証を行っていきたい。

## 第6章

# スマートフォンのセンサを活用した計測データ分析教材の提案と実践

### 6.1 はじめに

前章では、情報Ⅱの重回帰分析に活用可能な分析ツールが多くない課題を解決するために、提案システムに重回帰分析の特徴を理解できる機能とサンプルデータセットを提案した。

本章では、データの収集方法に焦点を当てる。データの収集にはオープンデータを利用することが多いが、センサからのデータを計測して収集する方法なども可能である。しかし、データ計測を行うには、センサを購入するという課題がある。それを克服するために、提案システムにスマートフォンの加速度センサ、ジャイロセンサ、GPSを読み取る機能を追加した。高校生を対象にした実践授業を通して、生徒が持参したスマートフォンを活用することで得られた利点を明らかにする。

### 6.2 工業科におけるデータ分析

高等学校の工業科では、平成30年に告示された次期学習指導要領解説(工業編)より共通科目である「工業数理基礎」と「情報技術基礎」が整理統合され「工業情報数理」が新設された。また、この科目は「情報Ⅰ」の代替科目として位置づけられており、学習内容の一つとして『工業に関わる事象の数値処理、実験データをグラフで可視化し、データの特徴を見出す学習』[66]が行われる。

これまでに高等学校ではセンサによる計測データを用いた分析を取り入れた学習として、マイコンボードにセンサを接続し計測したデータをサーバーに蓄積、そのデータをダウンロードし表計算ソフトにより分析を行った実践[67]やロボット教材に搭載したセンサの計測データを分析した実践[68][69]などが報告されている。これらの学習を行うためには、(1)GPIOが用意されたマイコンボード(計測機器)の準備、(2)計測プログラムの作成、(3)取得したデータの加工が必要となる。

(1)ではマイコンボードとセンサを適切に接続する必要がある、これには電気電子回路の基礎知識が必要となる。また、教員がマイコンボードや学習内容に応じたセンサ、配線用のケーブル等の準備や、それらのメンテナンスなどを行う必要がある、教員側に準備の手間がかかる。(2)ではプログラミングの基礎知識とともに使用するセンサによってはAD変換等の知識が必要となる。また学習者が意図しない実験結果を得たとき、誤りがプログラムなのか、実験方法なのかを判断するのが難しい場

合がある。(3)でもプログラミングの基礎知識が必要となり、計測したデータを取得するためのプログラムの作成や取得後のデータの加工が必要となる場合がある。このようにデータ分析の学習にセンサ等の計測データを用いる場合には、多くの前提知識が必要となり、学習者が取得したデータをグラフで可視化し、データの特徴を見出すデータ分析の学習に集中して取り組むのは容易ではない。

ブロックに内蔵されたセンサ値を Bluetooth 経由で取得できる MESH [70] と IFTTT [71] を組み合わせてデータを蓄積し、分析を行う実践 [72] も報告されている。MESH を利用することにより (1) と (2) が解決されるが、蓄積したデータの分析は別のソフトウェアを使用する必要があり、センサ値の計測からデータ分析までを同一環境で行える環境は少ない。

## 6.3 Connect DB への追加機能の設計と開発

### 6.3.1 追加する機能の検討

前章で述べた (1)~(3) の課題を解決するためには、身近にある端末を使用してセンサ値を取得でき、かつデータの加工や分析が可能となる環境が実現できれば、センサやマイコンボードの準備、計測プログラムの作成などの授業準備の負担を減らせることができると考えた。

(1) はスマートフォンやタブレット端末等に内蔵されたセンサを利用することで解決できる。(2) はプログラムを作成せずに計測データを取得できればよい。(3) は自動で計測データを取得し、データの整形やグラフ描画等のデータの可視化する機能を提供することで解決できる。

これまでに我々はオンライン上でデータを整形・可視化する Connect DB の開発を進めてきた。Connect DB は、学習者がデータ分析の学習に注力できるように度数分布や四分位数、クロス集計といった統計処理に関するデータの加工や各種グラフ描画を支援するデータ分析学習ツールである。図 5.4 に、架空の生徒データを使って所属クラブ別に身長四分位数を求め、箱ひげ図を作成した例を示す。

箱ひげ図の上にある「読込、削除、四分位数、箱ひげ」と書かれたタブは、箱ひげ図を描くまでに行った処理を表している。箱ひげ図の下にある表は、四分位数を表し、一番下の表が分析に使用したデータを表している。この Connect DB にスマートフォン等の端末に内蔵されたセンサ値を取得する機能とその値を蓄積する機能を追加することで (1)~(3) の課題が解決できると考えた。

### 6.3.2 機能の実装

Connect DB にスマートフォン等の端末に内蔵されたセンサ値を取得する機能とその値を蓄積する機能を追加した。対応するセンサー一覧を表 6.1 に示す。

表 6.1: 対応する端末内蔵センサ

| 対応センサ | 取得できる値             |
|-------|--------------------|
| ジャイロ  | alpha, beta, gamma |
| 加速度   | x, y, z            |
| GPS   | 緯度, 経度             |

スマートフォン等の端末内蔵のセンサ値の取得は、Connect DB 内の専用ページ (図 6.1) にアクセスすることで可能にした。まず、取得したいセンサをタップする。その後、「記録スタート」をタップすることで、計測を始める。「アップロード」をタップすることで、保持された計測データを Connect DB 上での分析が可能になる。

図 6.2 に取得した加速度センサの x, y, z の値を使って折れ線グラフを描画した例を示す。



図 6.1: 端末内蔵のセンサ値を計測する画面例



図 6.2: 加速度センサの計測値の折れ線グラフで表示例

## 6.4 スマートフォンを活用したデータ分析の授業実践

### 6.4.1 実施した授業の概要

Connect DB に追加した機能を使用して、スマートフォンに内蔵されたセンサにより計測を行い、その計測値を使用したデータ分析の学習が可能であることを確認するため、高等学校の工学科 2 年生 30 名を対象に実践を行った。生徒はこれまでに Connect DB を使用してデータの加工やグラフ描画を行った経験がある。他にも工業の実習の授業を通して、表やグラフや数式を利用して実験データを考察する学習をしている。しかし、今回の授業で取り扱う加速度センサのような計測データを行動の違いから考察することは、生徒にとって初めての体験である。加速度センサの知識については、加速度を計測していることや、ゲーム機のコントローラやスマートフォンに内蔵されていることを授業の中で紹介している。

今回実践した授業は、スマートフォン内蔵の加速度センサの知識の取得と、その活用方法として

表 6.2: 実施した授業の流れ

| 時限 | 時間 (分) | 内容                   |
|----|--------|----------------------|
| 1  | 10     | 導入                   |
|    | 20     | 実験 1: 3 軸加速度センサの理解   |
|    | 20     | 実験 2: 加速度センサを用いた動作分析 |
| 2  | 40     | 動作分析 (歩く、走る、オリジナル)   |
|    | 10     | まとめ                  |

「速さと加速度」から、行動の特徴を見出すことを目的とした。生徒は、加速度センサの値を可視化し、データの特徴から「歩く」「走る」の動作について分析した。教員は、動作の違いによるグラフの変化を、実際に読み取り考察できているかを判断し評価した。実施した授業は、工業の電子情報技術で2時間(50分×2)である。表 6.2 に授業の流れを示す。



図 6.3: 端末内蔵のセンサ値を取得する様子

まず、1時間目には加速度センサの値変化を体験することを目的に2つの実験を行った。1つ目の実験として、スマートフォンの持ち方によってどのように値が変わるかを考えた。生徒はスマートフォンの持ち方を図 6.3 の (a)~(f) のように変化させて、加速度センサの x、y、z 軸の値がどのように変化するかを読み取り、ワークシートに記入した。読み取った結果から、各軸がスマートフォンに対してどの方向に向いているのかを確認した。

次に2つ目の実験として、スマートフォンを上下に振ったときに加速度センサの値がどのように変化するかを考えた。取得した計測値を折れ線グラフで可視化することで、スマートフォンを振った時の値の変化を確認し、どのような特徴が読み取れるかを検討した。

2時間目は、生徒はスマートフォンの加速度センサの値から「歩く」や「走る」といった動作を区

別することができるかを考えた。

まず、生徒は動作を分析するための計測方法としてスマートフォンを「手に持つ」場合と「ポケットに入れる」場合のどちらが良いかを予想した。

次にスマートフォンで Connect DB にアクセスし、「歩く」と「走る」の動作時の加速度センサの計測を行った。計測から分析の流れは以下とおりある。

1. スマートフォンから Connect DB にアクセスする
2. 記録をスタートし、スマートフォンを手に持ち「歩く」または「走る」の動作を行う
3. 記録をストップし、データをグラフ化(可視化)する
4. グラフから動作の特徴を読み取り、考察をワークシートに記入する

図 6.4 は実際に生徒がスマートフォンを手に持って「歩く」動作と「走る」動作を行った際に得られた計測データを可視化したものである。

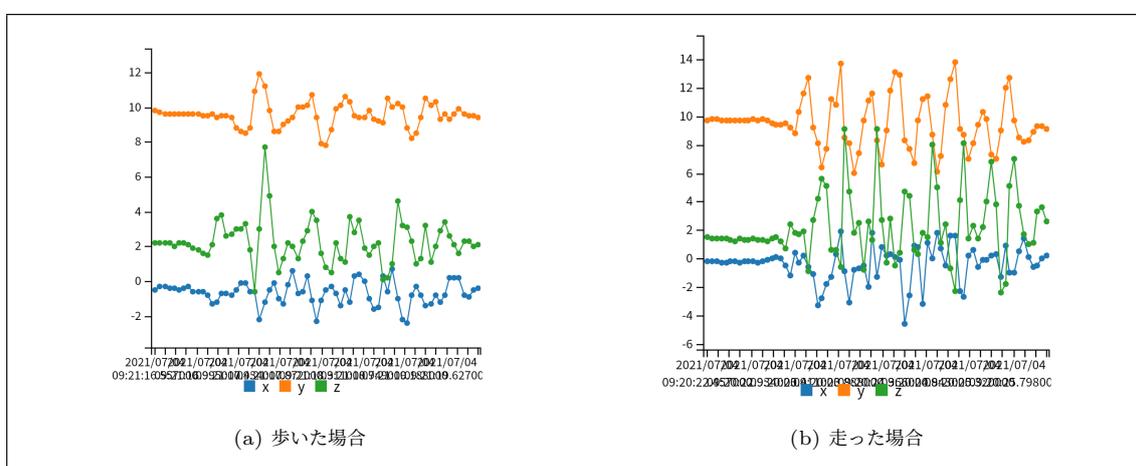


図 6.4: 加速度センサによる生徒の動作分析

## 6.4.2 実験結果の考察

このように生徒は Connect DB 上で計測した値を使ってグラフを描画し、その特徴について考察した。考察時には、値をグラフ化したときに読み取れる特徴等について分析を行い、それぞれワークシートに記述した。

「歩く」動作と「走る」動作の後には、「ジャンプする」や「しゃがむ」といったように各自で考えたオリジナル動作について計測を行い、分析を行った。ワークシートには計測した動作と値から読み取れる特徴について記述した。

## 6.5 ワークシートの記述内容の分析

生徒が提案システムを使用してスマートフォンで加速度センサの値を計測、データを可視化し、特徴を読み取ることで動作の分析を行うことができたかを、共通の動作(歩く/走る)とオリジナルの動

表 6.3: 軸についての生徒の記述の分類

| 分類項目       | 生徒の記述例   |
|------------|--|
| 3 軸の相対的な変化 | x、y 軸は共に上下に変化していたが、z 軸はあまり変化していなかった            |
| 3 軸の値の変化   | x 軸はプラスの数値にあり、y 軸はマイナスの値にあり、z 軸は 10 から 0 の間にある |
| 3 軸の交わり    | x、y、z 軸は交わらなかった。                               |
| 3 軸に変化なし   | x、y、z があまり変わりがなく 3 つとも穏やかな動きをしている              |
| 特定の 1 軸に着目 | x 軸がマイナスになることはなかった                             |

作について、ワークシートの自由記述項目より確認する。

### 6.5.1 共通動作における軸に関する記述内容の分析

「歩く」動作と「走る」動作を分析し、ワークシートの内容から生徒の考察の記述を確認した。「歩く」動作では 29 名 (96.7%)、「走る」動作では 19 名 (63.3%) の生徒は、動作の折れ線グラフから身体の動きと軸ごとの値の変化の関係性に触れた形で考察が記述されていた。軸の記述がない生徒は、「歩く」動作では 1 名 (3.3%)、「走る」動作では 11 名 (36.6%) だった。軸ごとの値の変化の記述を表 6.3 に示す項目に分類した。

表 6.3 で、「3 軸の相対的な変化」には、3 軸について相対的に比較し、グラフの変化について記述した考察を分類した。「3 軸の値の変化」には、値の変化に着目して記述された考察を分類した。「3 軸の交わり」には、3 軸の交わりの有無について記述された考察を分類した。「3 軸に変化なし」には、3 軸に変化がないと記述した考察を分類した。「特定の 1 軸に着目」には、特定の 1 軸の変化について着目して記述された考察を分類した。「軸の記述なし」には、軸についての記述はなくグラフの変化について記述された考察を分類した。

また、これらの分類について、「歩く」動作と「走る」動作ごとに集計した結果を表 6.4 に示す。「歩く」動作の考察では、「3 軸の相対的な変化の記述」が 10 名 (33.3%) と最も多く、次に「3 軸の値の変化」、「3 軸の交わり」が 6 名 (20.0%) だった。「歩く」の動作の記述例を表 6.5 の生徒 A、B に示す。これらの生徒は x、y 軸が変化して z 軸が変化しなかったことを報告している。特に生徒 B に関しては、変化する原因についても詳しく述べている。このように 29 名の生徒が身体の動きと軸ごとの値の変化の関係性に触れた形で考察を記述していたことから、多くの生徒は取得したデータを折れ線グラフに可視化し、グラフから 3 軸の値の変化について読み取ることができていることが示唆される。

「走る」動作の考察では、「3 軸の相対的な変化の記述」が 15 名 (50.0%) と最も多く、次に「軸の記述なし」が 11 名 (36.7%)、「3 軸の値の変化」が 4 名 (13.3%) だった。「歩く」動作と比較すると、「軸の記述なし」が 10 名増えている。この 10 名は動作の違いによる特徴を記述しており、これは「走る」動作の分析を「歩く」動作の後に行ったために、比較対象が生まれたことが要因であると考えられる。次節では、生徒の共通動作の比較に関する記述について分析を行い、生徒がどのような違いを発見したかについて確認する。

表 6.4: 考察内容の集計結果 (N=30)

| 生徒の記述内容   | 「歩く」動作    | 「走る」動作    |
|-----------|-----------|-----------|
| 3軸の相対的な変化 | 10(33.3%) | 15(50.0%) |
| 3軸の値の変化   | 6(20.0%)  | 4(13.3%)  |
| 3軸の交わり    | 6(20.0%)  | 0(0.0%)   |
| 3軸に変化なし   | 4(13.3%)  | 0(0.0%)   |
| 特定の1軸に着目  | 1(3.3%)   | 11(36.7%) |

### 6.5.2 共通の動作の比較に関する記述内容の分析

共通の動作(歩く/走る)の比較について分析する。「歩く」動作と「走る」動作を比較した生徒は、23名(76.7%)だった。このうち、13名(43.3%)が軸ごとの値の変化について分類した生徒であり、どの軸の振れ幅が大きいかを詳細に記述できていた。また10名(33.3%)は「軸の記述なし」に分類した生徒であるが、動作の違いから生まれる特徴について記述できていた。「走る」動作の記述例を表6.5の生徒C、Dに示す。生徒Cは「動作間隔が狭い」と周期について記述しており、4名(13.3%)の生徒に同様の記述が見られた。生徒Dは「振り幅が大きい」という部分に着目していることがわかる。これは、「歩く」動作との比較を行った全員が記述していた。このように、共通の動作の違いを「周期の間隔」や「振れ幅の大きさ」として読み取ることができた。また、「歩く」動作と「走る」動作の比較をできなかった生徒は7名(23.3%)であり、共通の動作のそれぞれのグラフ変化の様子の記述のみにとどまっていた。

このことから、生徒は1つの動作に関する分析を行う場合は身体の動きと軸ごとの値の変化に着目して考察を記述するが、分析対象を増やすことでグラフの比較を行う形でデータの特徴を考察する傾向があることが示唆された。今回の実践では「歩く」動作と「走る」動作を比較できる授業にしたことで、生徒が「走る」動作の分析時に「歩く」動作と比較した形でデータを考察し、「周期の間隔」や「振れ幅の大きさ」に気づくことができた。

### 6.5.3 生徒が考えたオリジナル動作の記述内容の分析

生徒が自ら考えた動作の計測を行い、得られた値の分析を行った際の考察の記述を確認した。生徒が考えた動作の一覧を表6に示す。身体動作を計測したのが14名、計測を行う端末を動かす形で計測したのが16名であった。26名(86.7%)の生徒は動作時の3軸の値の変化の特徴を捉え、詳しく考察することができていた。軸の記述がない生徒は、4名(13.3%)だった。3軸の変化を考察できたものは、生徒Dの記述のように軸の記述ができていたが、軸の記述がなかったものは、生徒Cのように軸の記述がなかったことが読み取れた。「振る」や「回す」といった単純な動作が多い中、発展的な内容にチャレンジした生徒もいた。記述例を表6.5の生徒E、Fに示す。

生徒Eはスマートフォンを手に持った状態で腕を動かして踊る動作時の加速度センサの値を計測した。z軸の値を観察することで手の傾きを検出できたことを示唆する記述がみられた。また四分位数についての記載があることから、折れ線グラフの描画だけでなく、自発的に箱ひげ図などの様々な分析を行ったことが読み取れる。

表 6.5: 生徒の考察記述例

| 生徒 | 項目         | 記述  |
|----|------------|---|
| A  | 歩く         | x 軸と y 軸は上下に変化したが、z 軸はあまり変化しなかった                  |
| B  | 歩く         | x 軸は足を前後に振るときに、y 軸は足を上げ下げするときに変化する                |
| C  | 走る         | 腕を振る速度が歩くときより速いため、動作間隔が狭い                         |
| D  | 走る         | 歩いたときと似ているが、歩いたときはマイナスだった y の値が 0 を超えるときがあった      |
| E  | オリジナル (回転) | y 軸を軸に回転させたときに x 軸の値がプラスからマイナスに変化し 360 度の回転を検出できた |
| F  | オリジナル (振る) | 上下に強く振ったときに y 軸は重力加速度が減少し、第 3 四分位数も -0.75 となった    |
| G  | 感想         | スマホ以外の加速度センサを搭載している電子機器と、どのような使われ方をしているかを調べたくなった  |
| H  | 感想         | スマホの画面が傾きで変わる理由がわかって楽しかったです                       |

生徒 F はスマートフォンを立てた状態で回転させた時のジャイロセンサの値を計測した。共通の実験では加速度センサの値を利用していたが、オリジナルの実験では、同様の方法でジャイロセンサの値を得ることができる機能があることを発見したことが読み取れる。計測した z 軸の数値から端末の回転を検出できることを示唆する記述がみられた。

このように生徒自身が「歩く」「走る」以外の動作を計画し、Connect DB を用いて計測を行い、得られた値を分析・観察して結論が述べられていることを確認できた。

#### 6.5.4 授業全体に関する記述内容の分析

授業全体の感想の記述例を表 6.5 の生徒 G、H に示す。生徒 G は、スマートフォン以外の電子機器でのセンサ利用に興味関心を持った。生徒 H は、普段スマートフォンの縦横を回転させると画面が回転することの原理に気付いた。このことから、生徒たちはセンサが身近な存在であり、自分たちの生活の中で活用されていることに気付くことができたと推測される。

## 6.6 考察

生徒が Connect DB を使用してスマートフォンで加速度センサの値を計測、データを可視化して特徴を読み取ることで動作の分析を行うことができたかをワークシートの記述により確認した。

生徒はスマートフォンから Connect DB にアクセスし、「歩く」動作と「走る」動作について計測を行い、グラフを描画することができた。授業の実施にあたっては、ほとんどの高校生がスマート

表 6.6: 生徒が計測した動作一覧 (N=30)

| 動作                | 件数 |
|-------------------|----|
| ジャンプする            | 6  |
| 端末を垂直方向に投げる       | 4  |
| 端末を上下に振る          | 3  |
| 端末を自由落下           |    |
| 踊る                | 2  |
| 端末を机の上で回す         |    |
| しゃがむ              | 1  |
| 腕立て伏せをする          |    |
| スイングをする           |    |
| その場で回る            |    |
| 端末をハンドルのように持ち回す   |    |
| 端末を包丁のように持ち千切りをする |    |
| 端末で扇ぐ             |    |
| 端末を回転させる          |    |
| 端末を倒す             |    |
| 端末を前後に振る          |    |

フォンを所持していることから、加速度センサやマイコンボードの準備、計測プログラムの作成等の準備は不要で学習を行うことができた。また、Connect DB へのアクセスについても QR コードを利用することで容易に行うことができた。これにより、生徒はグラフの観察に多くの時間を使うことができた。また、描画したグラフを観察することで、それぞれの動作時の値の変化や変化の特徴について分析することができた。

具体的には、「歩く」動作と「走る」動作はどちらも折れ線グラフで描画すると波のような波形を描くが、「走る」動作の方が振動が大きくなるといった様子を詳しく考察できていたことが確認できた。多くの生徒が値の変化に着目し、変化の特徴を読み取ることができていた一方で、計測値と身体の動きを関連付けて考察が記述されていたのは、7名(23.3%)だった。また、自由記述による生徒の感想にも、「特徴を書くのが難しかった」との記載があった。このことから、生徒はグラフから動作による値の変化等の特徴を読み取ることはできたが、その特徴を身体動作と結びつけて考察することは難しいことがわかった。

また、生徒が自ら考えた動作の計測では、Connect DB を使用し様々な動きによる値を計測していたことから、生徒が関心を持ち取り組めたことがわかる。

ワークシートに記述された考察についても、自らが行った動きと対応する形で3軸の値の変化について考察が書かれており、生徒は動作による値の特徴を読み取ることができたといえる。また折れ線グラフの描画だけでなく、様々な分析手法を用いて分析を行ったことが考察から読み取れた。

## 6.7 まとめ

本研究では、データ分析の授業でセンサ等の計測データを使用する際に課題となっていた点を解決するために、オンラインデータ分析学習ツール「Connect DB」にスマートフォンに内蔵されたセンサ値を取得し、取得したデータを蓄積できるようにした。

高校生を対象に授業実践を行い、「歩く」動作や「走る」動作等の身体動作を行ったときの加速度センサの値を計測し、分析する実習を行った。生徒が記述したワークシートの考察から Connect DB を使用することでスマートフォンからセンサ値を取得し、行動の特徴を読み解き、データの違いやその理由について詳細に述べる事ができていることを確認できたことから、提案システムが実際の授業で活用できることが分かった。

## 第7章

# 結論

本論文は、高等学校におけるデータサイエンス教育の現状と授業でデータ分析ツールを利用する場合の課題を踏まえて、授業に利用できるデータ分析ツールの開発の検討を行った。そこで、データ分析学習システム「Connect DB」の開発を行った。そして、提案システムが授業で利用可能であることを検証した。

3章では、提案システムの技術的な工夫だけでなく、情報の授業での利用も考慮し、一般的な表計算ソフトと Connect DB の授業展開の違いについて整理した。これにより、従来の分析ツールを使用する際に発生する可能性のある問題点を克服できることが明らかになった。例えば、表計算ソフトでは散布図行列の表示が難しいのに対し、提案システムでは容易に行える。また、提案システムは情報科を教えている教員にも有用であり、データ分析において高い効果が示唆された。

利用統計によると、本システムは教員・生徒合わせて約 2300 人が利用している。2023 年 9 月から 10 月にかけて、分析の実行回数が約 5 万件であり、高等学校の授業でも積極的に利用されていることが明らかになった。

4章では、生徒側の学習上の課題として、分析ツールの操作が難しい可能性があることを解決するために、提案システムの操作の容易性を評価した。ここでは、提案システムと従来使われてきた分析ツールである表計算ソフト (Microsoft Excel) との操作回数を比較することで操作回数が少ないことが示唆された。さらに、大学 1 年生を対象に授業実践を通してどちらのツールの方が使いやすいかを調査した。その結果、提案システムの方が使いやすいと評価され、その理由として「操作が容易である」、「操作の手順が少ない」、「分析可能なものだけを掲示してくれる」が役に立つとの意見があった。これにより、学習者の負担を軽減する機能を搭載していることが示唆された。

5章では、情報 II の機械学習の内容について焦点を当てた。従来は機械学習の意味を理解できる分析ツールや適切なデータセットは多くないといった課題があった。そこで、提案システムに重回帰分析を学習できる機能と適切なデータセットを検討した。大学 2 年生を対象に授業を実施した結果、「説明変数を複数にすることで予測モデルの精度が高くなること」、「複数の説明変数がどの程度、目的変数に影響を与えているかがわかること」、「影響が少ない変数を除いてモデルを最適化できること」を理解していた。これにより、提案システムに追加した重回帰分析の機能と適切なデータセットに実用性があったことを確認できた。

6章では、データの収集方法に焦点をあてた。収集方法としてセンサを利用したデータ計測が挙げ

られるが、これを実現するにはセンサを購入しなければいけない課題があった。そこで、提案システムにスマートフォンの加速度センサ、ジャイロセンサ、GPS を読み取る機能を実装した。工業科を設置する高等学校の生徒を対象に、生徒たちが持参したスマートフォンを利用して、加速度センサの値を計測しながら「歩く」、「走る」、「オリジナル」の動作を分析することができた。特に、スマートフォンはコードレスで動作することから、生徒たちはオリジナル動作として「ジャンプ」、「投げる」といった新しい動作を計測していることが確認できた。

# 謝辞

本論文をまとめるにあたり、大阪電気通信大学 兼宗進教授に御指導をいただきました。また、大阪電気通信大学 新関雅俊教授、中田亮生教授には、本論文の完成にあたり御教示、御助言をいただきました。感謝いたします。

本論文中の第3章から第6章の研究は、四天王寺大学助教 本多佑希先生、大阪電気通信大学博士後期課程 漆原宏丞氏にコメントをいただきながら教材開発を進めました。大阪電気通信大学講師 島袋舞子先生には、研究データの分析や論文の執筆について助言をいただきました。東京農工大学教授 並木美太郎先生、明星大学教授 長慎也先生、東京農工大学助教 長島和平先生、四天王寺大学教授 間辺広樹先生には、Connect DB のデータ管理機能やプログラミング学習環境 Bit Arrow との連携機能について貴重なアドバイスをいただきました。感謝いたします。

最後になりましたが、社会人大学院生ということで、平日は学校の職務に、夜間と休日は研究に、という生活を続けながら本論文を完成できたのは、常日頃より支援・応援いただきました職場の方々や家族のおかげです。深く感謝いたします。

## 参考文献

- [1] 内閣府: Society 5.0, 入手先 <[https://www8.cao.go.jp/cstp/society5\\_0/](https://www8.cao.go.jp/cstp/society5_0/)> (参照 2023/12/06).
- [2] 文部科学省: 高等学校学習指導要領 情報編 (平成 30 年告示) 解説, 入手先 <[https://www.mext.go.jp/a\\_menu/shotou/zyouhou/detail/mext\\_01831.html](https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_01831.html)> (参照 2023/05/10).
- [3] 数理・データサイエンス・AI 教育強化拠点コンソーシアム: “数理・データサイエンス・AI(リテラシーレベル) モデルカリキュラム ～データ思考の涵養～”, 入手先 <[http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\\_literacy.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\_literacy.pdf)> (参照 2023/08/14).
- [4] 大学入試センター: “サンプル問題「情報」(令和 3 年 3 月 24 日)”, 入手先 <[https://www.dnc.ac.jp/kyotsu/shiken\\_jouhou/r7/r7\\_kentoujoukyou/](https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7/r7_kentoujoukyou/)> (参照 2023/08/14).
- [5] 文部科学省: 高等学校情報科「情報 I」教員研修用教材「第 4 章 情報通信ネットワークとデータの活用・巻末」, 入手先 <[https://www.mext.go.jp/content/20200722-mxt\\_jogai02-100013300\\_006.pdf](https://www.mext.go.jp/content/20200722-mxt_jogai02-100013300_006.pdf)> (参照 2023/12/06).
- [6] 文部科学省: 高等学校情報科「情報 II」教員研修用教材「第 3 章 情報とデータサイエンス 前半」, 入手先 <[https://www.mext.go.jp/content/20200702-mxt\\_jogai01-000007843\\_004.pdf](https://www.mext.go.jp/content/20200702-mxt_jogai01-000007843_004.pdf)> (参照 2023/12/06).
- [7] 文部科学省: 高等学校情報科「情報 II」教員研修用教材「第 3 章 情報とデータサイエンス 後半」, 入手先 <[https://www.mext.go.jp/content/20200609-mxt\\_jogai01-000007843\\_007.pdf](https://www.mext.go.jp/content/20200609-mxt_jogai01-000007843_007.pdf)> (参照 2023/12/06).
- [8] ACM Data Science Task Force: Computing Competencies for Undergraduate Data Science Curricula, 入手先 <[https://www.acm.org/binaries/content/assets/education/curricula-recommendations/dstf\\_ccdsc2021.pdf](https://www.acm.org/binaries/content/assets/education/curricula-recommendations/dstf_ccdsc2021.pdf)> (参照 2023/08/14).
- [9] Bargagliotti, A., Franklin, C. Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D.A.: Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II (GAISEII), American Statistical Association (2020).
- [10] 文部科学省: 幼稚園教育要領、小・中学校学習指導要領等の改訂のポイント, 入手先 <[https://www.mext.go.jp/a\\_menu/shotou/new-cs/1384661.htm](https://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm)> (参照 2022/12/28).
- [11] e-Stat, 入手先 <<https://www.e-stat.go.jp/>> (参照 2023/12/12).
- [12] sAccess, 入手先 <<https://saccess.eplang.jp/#!/index.md>> (参照 2023/12/12).
- [13] Microsoft Excel, 入手先 <<https://www.microsoft.com/ja-jp/microsoft-365/excel>> (参照 2023/12/07).

- [14] Google スプレッドシート, 入手先 <[https://www.google.com/intl/ja\\_jp/sheets/about/](https://www.google.com/intl/ja_jp/sheets/about/)> (参照 2023/12/07).
- [15] R, 入手先 <<https://cran.r-project.org/>> (参照 2023/12/07).
- [16] R Studio, 入手先 <<https://posit.co/download/rstudio-desktop/>> (参照 2023/12/07).
- [17] Python, 入手先 <<https://www.python.org/>> (参照 2023/12/07).
- [18] Google Colaboratory, 入手先 <<https://colab.research.google.com>> (参照 2023/12/07).
- [19] 長島 和平, 長 慎也, 間辺 広樹, 兼宗 進, 並木 美太郎: Web ブラウザを用いたプログラミング学習支援環境 Bit Arrow の設計と評価, 情報処理学会論文誌教育とコンピュータ (TCE), 第 4 巻, 第 1 号, pp.57–69 (2018) .
- [20] IBM SPSS, 入手先 <<https://www.ibm.com/jp-ja/spss>> (参照 2023/12/07).
- [21] SAS, 入手先 <[https://www.sas.com/ja\\_jp/home.html](https://www.sas.com/ja_jp/home.html)> (参照 2023/12/07).
- [22] Tableau, 入手先 <<https://www.tableau.com/ja-jp>> (参照 2023/12/07).
- [23] 藤田 太郎: 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.44, No.2, pp.68–76 (2020).
- [24] Wild. C. ], Pfannkuch. M.: Statistical Thinking in Empirical Enquiry, International Statistical Review, Vol.67, No.3, pp.223–265 (1999).
- [25] 林 宏樹, 渡辺 博芳: スーパーサイエンスハイスクールにおける学校設定科目によるデータサイエンス教育のカリキュラム開発, 情報処理学会論文誌教育とコンピュータ (TCE) , 第 9 巻, 第 2 号, pp.19–28 (2023).
- [26] Watson, J., & English, L. : Eye color and the practice of statistics in Grade 6: Comparing two groups. Journal of Mathematical Behavior, 49, 35–60 (2018).
- [27] 川上 貴: 統計的モデリングの指導と学習に関する研究の国際的動向 – 日本の初等中等教育段階における統計教育の展望 –, 日本数学教育学会誌, 第 101 巻, 3 号, pp.15–27 (2019).
- [28] Jung Jin Lee, Jong Chan Lee: Developing Web-based Dynamic Graphical Software, eStat, for Statistics Education, 日本計算機統計学会大会論文集, 第 32 巻, pp.96–99 (2018).
- [29] eStat Statics and Data Science, 入手先 <<http://estat.me/estat/eLearning/en/index.html>> (参照 2022/12/28).
- [30] Watson, Jane, Donne, Julie: TinkerPlots as a Research Tool to Explore Student Understanding, Technology Innovations in Statistics Education, Vol.3, No.1 ,pp.1–35 (2009).
- [31] CODAP, 入手先 <<https://codap.concord.org/>> (参照 2023/12/07).
- [32] Murat Doğan Şahin, Eren Can Aybek2: Jamovi: An Easy to Use Statistical Software for the Social Scientists, International Journal of Assessment Tools in Education, Vol.6, No.4, pp.670–692(2019).
- [33] Sedigheh Abbasnasab Sardareh, Gavin T. L. Brown, Paul Denny: Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences, Four Contemporary Statistical Software Tools, Teaching Statistics, Vol.43, Issue S1, pp.S157–S172 (2021).
- [34] 文部科学省: 「教育の情報化に関する手引」(令和元年 12 月), 入手先 <<https://www.mext.go>.

- jp/a\_menu/shotou/zyouhou/detail/mext\_00117.html> (参照 2023/01/21).
- [35] 青山 和裕: 小学校算数の「データの活用」領域の授業におけるタブレット端末の利用について. 日本科学教育学会年会論文集, Vol.45, pp.117–118 (2021).
- [36] 稲垣 俊介: 情報の授業をしよう!: 高校生がスマートフォンの利用状況について当事者意識を持ってデータ分析をする実践 -情報Iの「データの活用」による「問題解決」の授業-, 情報処理, 第 63 卷, 第 9 号, pp.512–517 (2022).
- [37] 林 宏樹: 情報の授業をしよう!: 高等学校におけるデータサイエンスを基盤とした問題解決実践 -AI時代を切り拓くための創造力の育成を目指す-, 情報処理, 第 62 卷, 第 10 号, pp.560–565 (2021).
- [38] 朴 壽永, 新部 昭夫, 安江 紘幸, 井形 雅代, 山田 崇裕: 統計的検定学習支援用ウェブアプリケーション BuMoc の開発, 情報処理学会論文誌, Vol.61, No.5, pp.1111–1124 (2020).
- [39] 総務省統計局, 入手先 <<https://www.stat.go.jp/>> (参照 2022/12/28).
- [40] 日本統計学会 統計教育分科会: データで学ぶ! 統計・データサイエンス活用授業のための教材ダウンロードサイト, 入手先 <<https://statds.jp/data/>> (参照 2022-12-28).
- [41] 東京書籍: 新編情報 I (2022).
- [42] 東京書籍: 情報 I Step Forward!(2022).
- [43] 実教出版: 高校情報 I Python (2022).
- [44] 実教出版: 高校情報 I JavaScript (2022).
- [45] 実教出版: 最新情報 I (2022).
- [46] 実教出版: 図説情報 I (2022).
- [47] 開隆堂: 実践 情報 I (2022).
- [48] 数研出版: 高等学校 情報 I (2022).
- [49] 数研出版: 情報 I Next (2022).
- [50] 日本文教出版: 情報 I (2022).
- [51] 日本文教出版: 情報 I 図解と実習 (2022).
- [52] 第一学習社: 情報 I (2022).
- [53] 大西 俊弘: 情報科で「統計・データ処理」を教えよう, 日本科学教育学会 第 30 回年会論文集, pp.475–478 (2006).
- [54] 西仲 則博, 吉川 厚: 中学校教育における統計的思考力を育む授業実践, 日本科学教育学会, 第 35 卷, 第 2 号, pp.153–166 (2011).
- [55] 渡辺 美智子: 統計教育の新しい枠組み - 新しい学習指導要領で求められているもの -, 数学教育学会誌, 第 48 卷, 第 3–4 号, pp.39–51 (2007).
- [56] Kanda Y: Free statistical software: EZR (Easy R) on R commander, 入手先 <<https://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>> (参照 2023/06/24).
- [57] James E., Archer, JR., Fred B.: User Recovery and Reversal in Interactive Systems, ACM Transactions on Programming Languages and Systems, Vol.6, No.1, (1984).
- [58] Lee, Y., Kozar, K. A., Larsen, K. R.: The Technology Acceptance Model: Past, Present, and Future, Communications of the Association for Information Systems, Vol.12, Article

- 50, pp.752—780 (2003).
- [59] Sweller, J., Ayres, P., Kalyuga S.: CHAPTER TWO - Cognitive Load Theory, Psychology of Learning and Motivation, Academic Press, Vol. 55, pp.37–76 (2011).
- [60] 春日井優, 森本康彦: 機械学習を題材とした授業の検討と実践, 日本情報科教育学会誌, 第 11 巻, 第 1 号, pp.47–52 (2018).
- [61] 東京書籍: 情報 II (2023).
- [62] 実教出版: 情報 II (2023).
- [63] 日本文教出版: 情報 II (2023).
- [64] 岸本有生: 情報の授業をしよう!: 高等学校 (工業) でのスマートフォンを利用したデータ活用の授業, 情報処理, 第 63 巻, 第 2 号, pp.64–69 (2022).
- [65] Piñeiro, G., Susana P., Juan P. G., José M. P.: How to evaluate models: Observed vs. predicted or predicted vs. observed?, Ecological Modelling, Volume 216, pp.316–322 (2008).
- [66] 文部科学省: 高等学校学習指導要領解説 (工業編), pp.42, 入手先 <[https://www.mext.go.jp/a\\_menu/shotou/new-cs/1407074.htm](https://www.mext.go.jp/a_menu/shotou/new-cs/1407074.htm)> (参照 2022/08/16).
- [67] 間辺 広樹, 大村 基将, 林 康平, 兼宗 進: 情報科教育における IoT 学習環境の利用方法の検討, 情報処理学会情報教育シンポジウム論文集, 第 2016 号, pp.98–105 (2019).
- [68] 角樋 大地, 岸本 有生, 覺前 友哉, 小川 勝史: 情報教育におけるロボット教材を活用した IoT システム構築実習の授業実践報告, 情報処理学会情報教育シンポジウム論文集, 第 2019 号, pp.287–292 (2019).
- [69] 覺前 友哉, 角樋 大地, 岸本 有生, 西田 隆司, 小川 勝史: 情報教育における IoT システム構築実習支援用ロボット教材の開発, 情報処理学会情報教育シンポジウム論文集, 第 2019 号, pp.305–308 (2019).
- [70] MESH, 入手先 <<https://meshprj.com/jp/>> (参照 2022/08/16).
- [71] IFTTT, 入手先 <<https://ifttt.com/>> (参照 2022/08/16).
- [72] 木村 真人, 山本 利一, 在間 拓幹, 木村 僚, 工藤 雄司: IoT の仕組みやデータの効果的な活用を学習する計測・制御システムのプログラミング学習の授業実践と評価, 日本産業技術教育学会誌, 第 62 巻, 第 4 号, pp.349–356 (2020).

# 関連発表

## 主論文

### 論文誌

1. Tomonari Kishimoto, Yuki Honda, Kosuke Urushihara, Maiko Shimabuku, Susumu Kanemune: Connect DB: An Online Learning System for Data Analysis, *Computers in Education for Creative Learning, Journal of Information Processing*, 2024. (掲載予定)
2. 岸本 有生, 本多 佑希, 漆原 宏丞, 島袋 舞子, 兼宗 進: 高等学校において重回帰分析を扱うデータ分析学習ツールと授業の提案, *日本情報科教育学会学会誌*, 第 16 号, 2024. (掲載予定)
3. 岸本 有生, 島袋 舞子, 兼宗 進: スマートフォンのセンサを活用した計測データ分析教材の提案と実践, *日本産業技術教育学会誌*, 第 65 巻, 第 3 号, pp.197-204, 2023.

### 国際会議

1. Tomonari Kishimoto, Yuki Honda, Kosuke Urushihara, Susumu Kanemune: Connect DB: An Online Learning System for Data Analysis, *Association for the Advancement of Computing in Education (AACE)*, pp.532 - 537, 2021.

## 参考論文

### 解説記事

1. 岸本有生：情報の授業をしよう！：高等学校(工業)でのスマートフォンを利用したデータ活用の授業, 情報処理, 第 63 巻, 第 2 号, pp.64–69 (2022).

### 口頭発表

1. 岸本 有生, 本多 佑希, 兼宗 進: 計測データのクラウド保存と分析が可能な IoT 学習教材の提案, 情報処理学会情報教育シンポジウム論文集, 第 2020 号, pp.139–145 (2020).
2. 岸本 有生, 本多 佑希, 漆原 宏丞, 兼宗 進: スマートフォンの内蔵センサを用いたデータ分析教材の提案, 情報処理学会情報教育シンポジウム論文集, 第 2021 号, pp.159–163 (2021).
3. 岸本 有生, 本多 佑希, 漆原 宏丞, 島袋 舞子, 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, 研究報告コンピュータと教育 (CE), 2023-CE-170, pp.1–5 (2023).
4. 岸本 有生, 本多 佑希, 漆原 宏丞, 島袋 舞子, 兼宗 進: 高等学校で重回帰分析を学習できる教材の提案, 情報処理学会情報教育シンポジウム論文集, 第 2023 号, pp.81–88 (2023).