

報告

# データサイエンスのための統計教育とその効果

—統計解析ソフトウェア R を用いた授業の報告—

## Statistical education for data science and its effect

-Report on lecture of statistical software R-

清水優祐\*

SHIMIZU, Yusuke\*

概要：数理的思考やデータ分析・活用能力を持ち、社会における様々な問題の解決・新しい課題の発見およびデータから価値を生み出すことができる、データサイエンス力のある人材が求められている。そこで、城西大学理学部数学科で行っている、統計解析ソフトウェア R を用いたデータサイエンスのための統計教育の授業内容と、その教育効果について報告する。授業アンケート結果より、統計的プログラミングへの関心の高さが判明した。今後は、データサイエンス教育のための体系的なカリキュラムの構成が必要である。

### 1. はじめに

2020 年度から小学校でプログラミング教育が必修化されるように、近年、プログラミング能力のある人材が求められている。さらに、AI（人工知能）に代表される第 4 次産業革命技術革新が産業や社会の在り方を大きく変えようとしている現在、ビッグデータの解析に必要な統計学の基礎理論を理解し、それをプログラミングで実装できる学生を養成する、体系的なデータサイエンス教育が大学には求められている。実際、データ処理・分析と価値創造を担うプロフェッショナルであるデータサイエンティストを専門で養成する“データサイエンス学部”が、2017 年 4 月に国内で初めて滋賀大学で誕生し、翌年の 2018 年 4 月には、横浜市立大学で開設された。2019 年 4 月には、武蔵野大学が私立大学で初めてデータサイエンス学部を立ち上げることが発表されている。その他の大学でも、統計教育に力を入れている学部が文理問わず存在する。文部科学省も、大学における数理・データサイエンス教育強化策を議論しており、本学においても、数理的思考やデータ分析・活用能力を持ち、社会における様々な問題の解決・新しい課題の発見及びデータから価値を生み出すことができる人材の育成に資する教育体制を、早急に構築する必要性があると言える。実際、理学部数学科生の就職先として、ICT（情報通信技術）系は教職と同様に人気であり、情報系科目の体系的なカリキュラムの構成は、第 4 次産業への変革を担う人材の輩出だけでなく、学習意欲のある入学者の確保にもつながると予想できる。そこで本報告書では、理学部数学科におけるデータサイエンスのための統計教育の一環として行っている、統計解析ソフトウェア R（以下 R と呼ぶ）を用いた授業内容の紹介と、その教育効果について述べる。2 章では、R を用いた授業内容を、シラバスを引用して簡単に説明する。3 章では、教育効果の確認のために用いた授業アンケートの集計結果を載せている。まとめと今後の展望は 4 章で述べる。

---

\* 城西大学理学部数学科

## 2. Rを用いた授業内容

2017年度より、理学部数学科2年生の前期に開講される、「数式処理による統計（実習含む）」の講義を担当している（注1）。坂戸キャンパスの清光会館303教室にて、学内情報教育環境を使用してパソコン実習を行っている。Rを用いたパソコン実習を通して、データ解析およびプログラミングの基礎を学び、データの整理・視覚化、シミュレーションなど、統計的分析の理解を深めることが本講義の目標である。ここで簡単にRについて説明する。R（言語）は、1996年頃に登場した、オープンソース・フリーソフトウェアの統計解析向けのプログラミング言語およびその開発実行環境のことである。Excelなどの他の統計ソフトのデータ読み込みに対応しており、C・C++・Fortranなどの外部プログラムと動的リンクして拡張することも可能である。また、他のユーザーが開発したRプログラム（パッケージ）がCRANというネットワークで配信されており、ダウンロード・インストール・アップグレードが容易にできる。統計学を超えて学問分野や業界を問わず、金融工学・時系列分析・機械学習・データマイニング・バイオインフォマティクスなど、柔軟なデータ解析や視覚化そして知識共有の需要に応えうるR言語の普及は、世界的な広がりを見せている。

ここでは主に、2017年度の授業内容を報告する。表1は、2017年度の講義スケジュールである。シラバスを抜粋しているのので、詳細はそちらを参照されたい（注2）。教科書は指定していないが、[1]を参考に講義スライドを作成し、プロジェクターを用いた授業を行った。1コマ目では、Rの起動・終了、スクリプトファイルの新規作成・開き方・保存などの基本操作、四則演算の実行方法、正弦関数や対数関数などのRに標準実装されている関数を用いた計算手法、変数への代入方法、ベクトルの生成・成分へのアクセス・結合・演算を行った。2コマ目では行列を扱った。具体的には、行列の生成・成分へのアクセス・演算、行列式の求め方、転置行列・単位行列・対角行列・逆行列の生成、固有値・固有ベクトルの計算手法を行った。3コマ目では、関数plot()を用いたデータの可視化を行った。4-6コマ目では、関数read.table()を用いてExcelデータの読み込みと、解析を行った。7, 8コマ目は、関数function()を用いて、関数の作成を行った。function()は、よく行う計算を関数化できる非常に便利なものである。9コマ目ではこれまでの内容に関する演習問題を用意し、各自実習を行った。10コマ目以降は応用編とし、主にプログラミングに関する内容を扱った。10, 11コマ目では、コードを繰り返して実行するfor文と、条件を分岐するif文・else文を扱った。for文はコードの簡略化や計算負荷の削減のために、if文は多様な場合分けを行う際に必要不可欠なものである。12コマ目では、forとif文に関する実習を各自行った。13, 14コマ目では、乱数の生成と数値シミュレーションの基礎内容を扱った。15コマ目は、期末レポート課題に取り組んだ。分からないところはまわりと積極的に相談し、ほとんどの学生がA以上の成績であった。アンケートによる授業効果の報告は次節で詳細を述べる。

注1 2018年度より、科目名が「コンピュータによる統計」に変更。

注2 [https://junavi.josai.ac.jp/camweb/slbsskyr.do?value\(nendo\)=2017&value\(tntoKyosyoid\)=13321&buttonName=searchKougi&methodname=kyoinSearch](https://junavi.josai.ac.jp/camweb/slbsskyr.do?value(nendo)=2017&value(tntoKyosyoid)=13321&buttonName=searchKougi&methodname=kyoinSearch)

表 1 授業スケジュール

コマ	内容	コマ	内容
1	Rの基本操作, ベクトルの計算	7, 8	関数の作成
2	行列の計算	9	演習
3	グラフの作成	10, 11	プログラミング入門, if文, else文, for文
4	1次元データ解析, 平均, 分散, 標準偏差	12	演習
5	2次元データ解析, 散布図, 相関係数	13, 14	シミュレーション入門, 乱数の発生
6	回帰分析	15	演習

### 3. 授業アンケート結果

この章では、本学で最終授業時に実施している授業に関するアンケートのうち、本授業への関心・取り組み・評価に関する結果を報告する。アンケートの回答者 58 人のデータを集計すると、図 1 の結果が得られた。いずれの問いも、「強くそう思う」、「そう思う」、「どちらともいえない」、「そうは思わない」、「全くそうは思わない」の 5 択で回答する質問であるが、Q1, Q2, Q5, Q6 は、「全くそうは思わない」と回答した者が 0 人であったため、表記を省略している。また Q5 は、「そうは思わない」と答えた者も 0 人であった。Q1 より、77% の学生がシラバスを見て興味を持ったことが分かり、プログラミングに興味を持っている意欲的な学生が多いことが判明した。この授業への取り組みとして、ノートを取るように努力したかを質問した Q2 では、95% の学生が「強くそう思う」または「そう思う」と回答した。高い数値になったのは、本授業では授業スライドを印刷したハンドアウトを配らず、R のスクリプトファイルにタイピングするよう指示したためである。授業のペースは落ちるが、受け身にならず、集中して授業に取り組めるよう工夫し、教育効果の向上を狙った。その効果が表れたのが Q3 である。Q3 では、77% の学生が、この授業で理解できない部分があったとき、友人と話し合ったと回答している。これより、能動的な学習を促すことができていると見て取れる。また、図には記載していないが、Q3 で、「どちらともいえない」、「そう思わない」、「全くそうは思わない」と回答した 11 人のうち、「担当教員に質問した」の質問には、「強くそう思う」が 1 人、「そう思う」が 3 人、「どちらともいえない」が 4 人、「そうは思わない」が 3 人、「全くそうは思わない」が 0 人であった。今後は TA をつけるなどし、課題が分からず一人で悩んでいる学生にも声かけを行っていきたい。Q4 から Q6 は、この授業の評価に関する問いである。内容の難易度についての質問である Q4 では、「強くそう思う」または「全くそうは思わない」と回答した学生が 7% おり、個人差があることが判明した。これについては、演習が早く終わった学生に対して難易度が高めの問題を用意するなどし、対応する必要がある。Q5 は授業内容が興味深いものであったかという質問で、90% の学生が興味深いと答えた。また、理解度に関する質問である Q6 によると、84% の学生が理解できたと回答した。Q4 から Q6 の結果より、興味を惹くような適度な難易度の授業になっていると言える。この他にも、「この授業を 10 点満点で採点してください」の問いの集計結果は、平均約 8.44 点、標準偏差約 1.68、最高点 10 点、最低点 2 点であり（無回答 2 名）、「この授業で良かったと思う点を書いてください」の問いには、

- ・微積や統計数学などのほかの科目の内容と絡めて教えてくれた。
- ・復習を毎回行っていて理解しやすかった。
- ・パソコンの使い方や作業に慣れることができた。
- ・スライドが見やすかった。
- ・定義から例題の流れが分かりやすかった。
- ・授業のペースが丁度よかった。
- ・プログラミングの内容と関連付けながら学べた。
- ・ノートをパソコンでとるので、タイピングの練習になった。
- ・生徒の理解具合を確認しつつ回ってきてくれるので、質問がしやすかった。
- ・解答もしっかり言ってくれた。

といったコメントがあった。また、「この授業で改善すべきだと思う点があれば述べてください」の問いには、

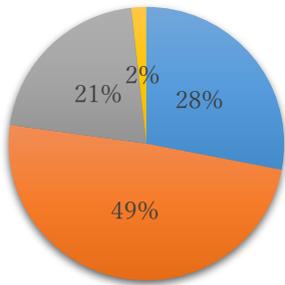
- ・授業スライドのプリントが欲しかった。
- ・演習の時、早く終わった人は授業の半分近くが待ち時間で何もすることがなかったので、 $+\alpha$ の課題を出すか、終わったら帰っていいとかしてもらえたらよかった。
- ・スライドを送るのが少し早いので、もう少し余裕が欲しかった。
- ・スクリーンの文字をもう少し大きくすると見やすくなると思った。

といったコメントがあった。

#### 4. まとめ・今後の展望

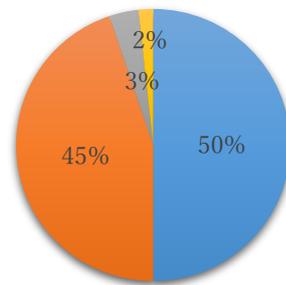
データサイエンスのための統計教育の一環として、Rを用いた授業を行い、アンケートによりその教育効果を調査した。授業内容は、Rの基本操作からデータの可視化、プログラミング・シミュレーションの入門的な事項を主に扱った。授業に関するアンケートより、統計的プログラミングに意欲的な学生が多い結果、理解度や満足度が非常に高くなる傾向があることが判明した。また、演習中に周りとは相談している様子が見受けられたことから、アクティブラーニングとしても機能していることが分かった。しかしながら、本授業の開講時期が統計学を学び始める2年生前期ということもあり、Rの基本操作と簡単なデータ解析がメインとなり、乱数を用いた数値シミュレーションといった応用的な内容があまり扱えないことが改善すべき点である。冒頭で述べたように、データサイエンス教育は、IoT (Internet of Things)、ビッグデータ、AI、ロボットに代表される第4次産業社会で活躍できる人材の輩出につながる。したがって、情報系科目の体系的なカリキュラムの構成を行う必要があると思われる。統計的プログラミングの枠組みでは、Rだけでなく、機械学習や深層学習を行う際に用いられるLinuxやPythonなどを授業に取り入れ、将来的にはTensorFlowを用いた画像認識や音声認識などを行っていきたい。

Q1 この授業を履修した動機は  
シラバスを見て興味を持ったから



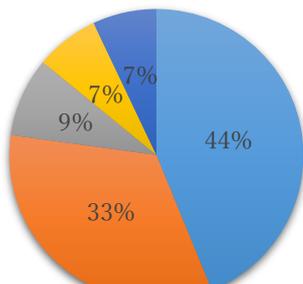
■ 強くそう思う ■ そう思う  
■ どちらともいえない ■ そうは思わない

Q2 この授業への取り組みについて  
ノートを取るように努力した



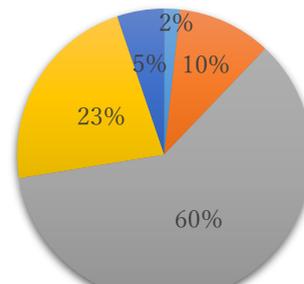
■ 強くそう思う ■ そう思う  
■ どちらともいえない ■ そうは思わない

Q3 この授業で理解できない部分があったとき  
友人と話し合った



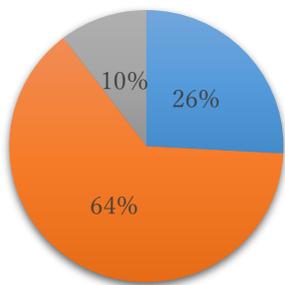
■ 強くそう思う ■ そう思う  
■ どちらともいえない ■ そうは思わない  
■ 全くそうは思わない

Q4 この授業について  
内容の程度が高すぎた



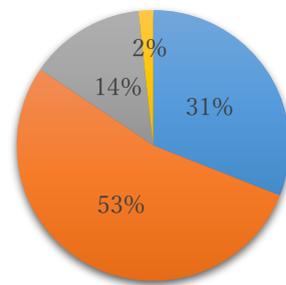
■ 強くそう思う ■ そう思う  
■ どちらともいえない ■ そうは思わない  
■ 全くそうは思わない

Q5. 授業内容について  
興味深い内容であったか



■ 強くそう思う ■ そう思う  
■ どちらともいえない

Q6 この授業に対する満足度について  
内容がよく理解できた (分かりやすかった)



■ 強くそう思う ■ そう思う  
■ どちらともいえない ■ そうは思わない

図1 アンケート結果

参考文献

- [1] 舟尾暢男, The R Tips 第2版 データ解析環境 R の基本技・グラフィック活用集, オーム社, 2019