

演習等で使用するデータ教育用教材の作成

研究年度 平成 30 年度

研究期間 平成 30 年度

研究代表者名 石田 和彦

1. はじめに

現在、「データサイエンス教育」は、わが国の大学教育における最大の課題の 1 つと言ってもよい状況にある。現代の情報化社会が求める「高度なデータリテラシーを有する人材」の大幅な不足に対し、従来の、文系理系分離型の大学教育では対応が困難なことは明らかである。既に、いくつかの大学には「データサイエンス学部」が設置され、現在、計画や構想中のものも少なくない。さらに、文理融合型のデータサイエンス教育を拡充するためのモデル作りを目指した、文科省の補助事業も始まっている。

このような状況の中で、本学学生の平均的なデータリテラシーは、残念ながら極めて低いと言わざるを得ない。例えば、本学が重点的な教育課題として取り組んでいる「しまなび」でも、離島が抱える課題やそれに対する対応策の提案を、データ面からきちんと裏付けている報告は殆どみられない。たまに、報告にグラフや数字が用いられることはあるが、そのレベルは、正直に言えば、大半が小学生か、精々中学生レベルである。アンケート調査もしばしば行われるが、調査の基礎をまったく理解しないで考案されたものが多く、その信頼性には大きな疑問がある。現代の政策決定においては、「証拠に基づく」(Evidence Based) と言うことが強く求められ、その「証拠」は多くの場合データやその分析結果であることを考えると、これは大きな問題である。

また、「ビッグデータの時代」と言われ、企業の意味決定においても様々なデータの有効活用が大きく重要度を増す中で、筆者が所属する経営学部には、そもそもデータ教育に関する科目が配置されておらず、統計学の専任教員もいない。特に、グローバル人材の育成を謳う国際経営学科では、英語教育ばかりが重視され、国際的な会議・会合やグローバルなビジネス現場での最大の共通語は実は「データ」である（下手な英語で長々と説明するよりも、的確なグラフを 1 枚見せることができれば、それで話は終わる）ことがすっかり置き忘れられている。

こうした事態を改善するためには、学部・学科のカリキュラムの抜本的改訂、新たな科目や教員の配置といったことが必要であろうが、その実現は、仮に可能であったとしても長い時間を要する。その間、本学の学生がまともなデータ教育を受けることなく放置され、ますます進展するデータ社会において本学やその学生が次第に競争力を失っていくことを防止するために、まず、現在できることから少しでも手を付けようというのが本研究の発想である。

具体的には、主として演習の場を通じて、学生に、社会で必要とされる最低限のデ

一タリテラシーを身に着けさせることを目指し、そのための教材作成を行う。教材は、筆者が担当する演習科目（基礎演習＜2 年生＞を中心に、一部は教養セミナー＜1 年生＞、専門演習＜3 年生＞を含む）で実際に使用して学生にデータ教育を行いながら、その効果を見極め、必要なら改善を図っていく。

2. 研究内容について

筆者は、2013 年度より、当時在籍していた日本銀行から、「統計教育大学間連携ネットワーク」（略称 JINSE、青学大、東大、早大等 8 大学連携による文科省補助事業）に派遣され、そこで、データに基づく課題解決型人材育成のための統計教育の改善という仕事に従事した（本学着任後現在に至るまでも、ここで築いた JINSE＜文科省補助金終了後は、新 JINSE に移行＞や、その支援団体である日本統計学会、同統計教育部会等との協力関係は維持している）。その際の業務の 1 つとして取り組んだのが、経済学、経営学分野における統計教育の標準的なカリキュラム（「コア・カリキュラム」）の策定である。この「コア・カリキュラム」は、統計学の専任教員がいない大学での活用も想定して、カリキュラムの各項目（1 学期 15 回のシラバスの各回を想定）に、そこで使用する教材の実例が、その一部にはあるが付されている。

当初、本研究でも、これを出発点として、学生が実際に自分の手で、データ探し・収集、集計、分析、グラフ作成等の作業をしながらデータ処理の基礎を学ぶことが出来るような教材を、現実の経済・経営に関するデータを用いて多数作成し、それらをコア・カリキュラム表に則して体系的に整理することを考えていた。

実際、本研究着手と共にこうした作業を開始し、いくつか作成した教材（変化率グラフの作成、季節要因の発見、前年比と前期比の違い、等々）を実際に演習（主として、2 年生対象の基礎演習）で学生に課題として与え使用してみた。しかし、極めて基礎的な内容であるにもかかわらず、学生の理解度、適応度が余り高くないことが次第に判明してきた。

そこで、学生のデータリテラシーの現状を把握するために、2 年生を中心に、一部は 1 年生（教養セミナー）、3 年生（専門演習）にも、統計検定 4 級（中学数学「資料の活用」レベル）及び同 3 級（高校数学 I ＜全員必修＞「データの分析」レベル）の問題を解かせてみたところ、3 級はおろか 4 級でも、合格レベルに達する学生はごく少数であることが判明した。無論、演習の成績評価には関係させないことを断った上でのゼミ時間中での試験であるので、学生の「本気度」が低いこと等は考慮する必要があるのは言うまでもないが、試験後に行った解説への反応や、コメントシートへの感想等の記入内容をみても、多くの学生がこのレベルのことすらきちんと理解していないことは明らかであった（たまたま、筆者のゼミ所属学生のレベルが低かった＜サンプル・バイアス＞ということも考えられなくはないが、教養セミナー、国際経営学科基礎演習は、ともにランダムにクラス分けが行われていることを考えれば、その可能性は低いであろう）。

本学経営学部の学生は、基本的には全員、大学入試センター試験で「数学 I」を受験し、なおかつ 1 年次の必修科目として「統計学」を学んでいるはずであることを考えると、この結果は極めて憂慮すべき事態である。冒頭に述べたような、データの活用に関するしまなびの惨状も、ある意味当然の結果であると言えよう。このような結果を踏まえ、特に、経営学部においては、1 年生対象の必修科目「統計学」について、その教育内容の再検討・刷新が図られるべきと筆者は考えているが、取り敢えず、本研究では、こうした現状を所与として、学生に最低限のデータリテラシーを身に着けさせるため、より基礎的な教材の作成を試みることに方向転換した。

具体的には、総務省から提供されている『全国消費実態調査』の教育用疑似マイクロデータ（完全に現実のデータではないが、現実のデータに極めて近くなるように作成されている）を材料として使用し、さらに、データ数（サンプル数）も無理なく学生が扱えるように差し当たり 200 個に絞り込んだうえで、データセットを EXCEL ファイルで学生（2 年生の基礎演習ゼミ生）に提供し、度数分布表・ヒストグラムの作成、基本統計量（平均値、中央値、四分位数、分散、標準偏差、相関係数、等）の計算、等の作業をゼミの課題として行わせた。学生の作業後に、結果を演習の場で報告・発表させ、さらに当方より詳しい解説を加えるというプロセスを繰り返しながら、統計学の初歩的な概念を、現実のデータに則して改めて理解させるような教育を進めた。

なお、その際に重視したのは、平均値、分散等の数式の展開ではなく（例えば、統計学の教科書に標準的に登場する、「仮平均を用いた平均値の簡単な計算法」や「分散の簡単な計算法（ $\sum X^2 - \bar{X}^2$ の式）」等は、EXCEL 時代にはもう不要である）、EXCEL 等の表計算ソフトを用いて実際にデータからグラフを描いたり、統計量を実際に求めたうえで、その意味、解釈、利用方法等を理解させることである。基本的な数式を一度理解しておくことの重要性を否定するものではないが、そこは科目としての「統計学」の分担として貰いたい。また、演習でこうした課題・作業を繰り返す中で、学生の多くが、実は EXCEL 等の表計算ソフトの基本的な使い方を十分には身に着けていないことも判明したため、この面の指導も併せて行った。

3. 研究成果

以上のような試行錯誤を繰り返した結果、現時点で、以下のような教材が作成されている（カッコ内は、主に使用したデータ）。もし興味のある教員がいれば、教材としての提供も可能である。

- ・ データ分析の基礎
 - データの性質、全数調査と標本・サンプル調査の違い、公的統計・民間統計と統計法、等の基礎事項
- ・ データの在りかと探索・発見方法（その一例として、『国際調査』、『人口統計』、等を

探索)

- ・ 度数分布表、ヒストグラムの作成とその意味（総務省『全国消費実態調査』の教育用疑似マイクロデータ）
- 特に、左右非対称分布と対称分布（疑似乱数で作成）の比較、その意味
- ・ 平均値、中央値、四分位数、分散、標準偏差、等の基本統計量の計算（同）
- ・ 相関係数の計算（同、総務省『家計調査』）
- 特に、相関関係と因果関係の区別について理解させる
- ・ 時系列データ分析の基礎としての変化率（前年比、前期比、前期比年率、等）の計算（『GDP 統計』、その他、多数の実際のマクロ時系列データ）
- ・ 変化率グラフの書き方（同）
- ・ 時系列データの季節性の発見、その意味
- ・ 季節性を踏まえた分析方法（『GDP 統計』、『百貨店売上高』、等）
- ・ 季節調整済データと原系列の比較
- 統計的季節調整手法自体は高度過ぎて扱えないが、その意味だけは理解させる
- ・ 寄与度分解と要因分析（『GDP 統計』、『百貨店売上高』、等）

4. おわりに

2.で述べたような本学学生のレベルに合わせた軌道修正も加わったため、当初考えていた JINSE「コア・カリキュラム」と対比すると、30 年度の研究でカバーし実際に教材を作成できた項目は、全体の半分以下に止まった。それでも、3.に挙げたような項目を、現実のデータを用いて自分で作業・計算し、その方法や意味を理解した経験は、学生のデータリテラシーを大幅に向上させたものと考えている。

もし、31 年度も継続して学長裁量研究費を獲得することができれば、「コア・カリキュラム」に他の項目についての教材作成を進めるとともに、3.に挙げた現在作成済の項目についても、実際の演習科目の中で引続き使用し、学生の反応や理解度をみながらさらに改善して行くこととしたい。

さらに、こうした実践結果を、データサイエンスや統計教育に関する研究集会、ワークショップ等に提示し、他大学（データサイエンス教育先進校やモデル校だけでなく、本学と同じような悩みを抱える多数の大学を含む）との情報・意見交換も進めて行きたい。こうした研究成果を踏まえ、いずれかの段階で、本学データ教育の抜本的な改革を提案することを、本研究の最終目標としたい。

以上