

《論文》

データサイエンス教育の現状と課題
—高等教育機関におけるデータサイエンス教育の方向性—

山口 和範
山口 誠一
門田 実

【要旨】 近年、高等教育機関におけるデータサイエンス教育の議論が盛んである。統計学部や学科が存在した海外はもとより、統計という名の付く学部学科がなかった国内においても、データサイエンス学部が複数の大学で開設された。さらに、国や産業界からは、数理・データサイエンス教育の必修化という要望も出されている。一部の大学では、全学生を対象に、データサイエンス関連の科目の必修化を進めている。一方、データサイエンスという領域やデータサイエンティストという呼称については、様々な定義や論点があることも事実である。立教大学では、2018年度から、これまでの社会調査や統計を基礎としたデータサイエンス副専攻を、全学生を対象にスタートさせている。本論文では、これまでの統計教育改革の流れを踏まえた上で、データサイエンス教育の現状と今後の課題について議論する。なお、ここで議論するデータサイエンス教育のターゲットは、リテラシーレベルの教育としての学生とする。

キーワード：統計教育、統計的思考力、コミュニケーション、課題解決

I はじめに

1990年代に入り、初等中等教育から高等教育に至るまで、海外での統計教育再編の動きが活発化した。その背景には、国家的に推進される科学技術振興政策があり、これが諸外国の学校教育の中で、統計教育の方法論として新しい枠組みであった“Statistical Thinking (統計的思考力)の育成”という柱をもたらした。先進各国は、人材や技術など「知の創造」をめぐる大競争時代に突入し、世界全体での持続的発展や自国の産業競争力の国際的優位性の獲得を目指し、科学技術・学術研究の戦略的な推進政策を推し進めた。とくに、1998年、全米研究会議が通称オドム・レポートを取りまとめ、数学と他分野および産業との連携の重要性を指摘して以降、米国科学財団(NSF)は、重点領域に数理科学を採用し、その中の重要テーマとして、“巨大データに関する数学的・統計的挑戦”、“不確実性の管理とモデリング”、“複雑な非線形システムのモデリング”を挙げた。このような振興政策を支える人材の育成のため、NSFが重点領域に数理科学を採用した時点でその具体的な目標として、

- ・数学および統計学の基盤研究の推進
- ・他の自然科学・工学との協働研究の推進
- ・学校教育における数理科学教育の推進

を掲げており、3点目の数理科学教育については、学校教育における主に数学や理科の中に反映され、前者2項との連動性から、数学や理科教育の中で統計教育の比重が増している。さらに、その中の統計教育においても単純に統計リテラシーを有しているだけでなく、数学以外の他の学問領域、また産業や市民生活の場における科学化(諸対象を科学的に探求するプロセス)を促進できる人材、いわゆる統計的課題解決型の思考力を持った研究者や生活者の育成をも強く意図されている。統計的思考力が科学技術発展の第3の腕と位置付けられ

の中で、統計科学の研究領域の中においても、統計的思考力自身の定義やその育成方法の研究が活発化した。大学における統計学入門教育の内容に関しては、1996年に米国統計学会（ASA）と全米数学協議会（MAA）の共同カリキュラム委員会がデータ分析の実践の要素を盛り込んだ統計教育の共同指針を発表し、翌1997年には the College Board が Advanced Placement テストに統計科目をこの指針の下に導入し、高校生の統計 AP テスト受験者数が毎年増加をたどり、10万人を超えるまでに至っている。この学校教育における統計教育の需要により、米国統計学会は、新しい枠組みでの統計教育のガイドラインと評価方法を初等中等教育から大学の統計入門コースまで体系的に作成・公開し、現場の教育を積極的に支援している。中でも、GAISE（Guidelines for Assessment and Instruction in Statistics Education）レポートでは、学校教育レベルと高等教育レベル、それぞれにおいて、教育の内容および評価についての指針を提示し、その後の統計教育の改革に大きな役割を果たしている。

一方英国では、王立統計学会（Royal Statistical Society）が統計教育センター（RSSCSE: Royal Statistical Society Centre for Statistical Education）を設立し、英国の統計局との協体制の下で、具体的な教材開発、大学における統計教育内容の指導と認証など統計教育のサポートを組織的に行っている。他の諸外国も同様の統計教育改革が進んでいる。

日本においても、日本統計学会が中心となり統計検定がスタートし、その受験者数も年々増加するなど、統計の学びに関する需要も高い。また、ビッグデータをキーワードとしたデータサイエンティストへの産業界からの関心もあり、高等教育機関における統計思考力養成への期待は高い。渡辺（2013）は、日本における統計教育改革の方向性を、問題解決のための重要なツールという視点で議論している。また、山口（2015）では、問題解決のために必要なリーダーシップ教育との関連で、統計教育の新たな枠組みについて議論を行い、その具体例として、立教大学における取組事例を紹介している。

このような統計教育改革の流れにさらに拍車をかけたのが、人工知能の社会実装の具体化の進展と、データを無尽蔵に生み出される有益な資源ととらえ、それを有効利用できる人材の必要性への気づきである。そこでは、統計という枠組みのみで捉えるのではなく、情報技術の領域等も広く含んだデータサイエンスという枠組みでの教育改革が求められている。本論文では、現状でのデータサイエンス教育の方向性をまとめ、今後の在り方について議論を行う。

II データサイエンス教育

1. データサイエンス教育の進展

データサイエンスというキーワードは、そのキーワードを用いる専門家の専門領域により重きを置く比重は変わるにせよ、統計学、情報技術、および、適用する専門領域の知識を活用して、問題解決にあたる方法論であると考えるのが一般的であろう。

データサイエンスは、科学や産業に革命を起こす新しいアプローチとして期待されている。これは、多くの研究分野のみならず社会の至るところで大規模で網羅的なデータが急速に蓄積されるようになり、データやその分析結果を利用することが可能になることと、利用

しないことが一種の機会損失であるという意識の高まりによると思われる。このことは、今後データを経済、社会及び日常生活の発展と新しい価値創造の源泉となるように有効利用できる人材が必要不可欠であると考えられている。

日本においては、2015年に情報・システム研究機構ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会の報告書『ビッグデータの利活用のための専門人材育成について』(北川 2015)で、「データから新しい価値を生み出すことができる「棟梁レベル」をはじめとするデータサイエンスの専門人材を育成することの重要性は言うまでもない。同時に、専門人材が生み出す成果の受け手となる国民全体のデータリテラシーを醸成することが、我が国においてデータサイエンス、AIを活かして超スマート社会を実現していくために必要不可欠だと考えられる。したがって棟梁レベルからリテラシーレベルまで裾野の広い人材育成が望まれる」と述べられている。

リテラシーレベルでの教育をここでの議論の対象とすると、データリテラシーとしては、データサイエンスやAIの不適切な利用や、個人データの利用から生じうる倫理的問題を認識できるようになることの重要性も広く言論されている。これは、その認識が今後のデータ駆動型社会を安全・安心に過ごすために必要不可欠なことであるからである。もちろんデータサイエンスの専門人材も、そのような側面を十分に認識した上で価値を創造することが強く期待される。

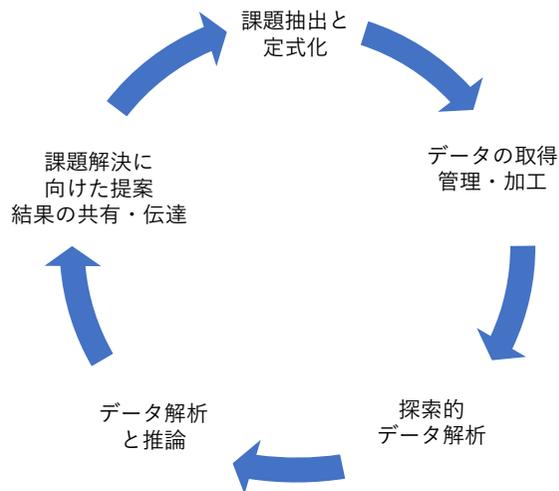
このような議論においては、統計教育における市民が身につけるべき内容の議論を行った Utts (2003) が十分に参考になる。ここでは、確率統計を学んだ市民が身につけておくべき内容として、「無作為化実験と観察研究の違い：交絡の理解」、「統計の有意であることの意味：大標本の場合」、「有意でないことの意味：小標本の場合」、「バイアスが生じる原因」、「偶然の意味：めったにないこともどこかでは起こる」、「因果の方向性と条件付き確率」、「バラツキの理解」の7点を挙げている。

データサイエンスにおけるデータの状況は、20世紀に発展した統計的推測法とは、一部その規模や形式において異なる点はあるものの、ここでの指摘の内容については、データサイエンスという呼称においても、少なくとも身につけておかないといけない統計学的側面のスキルであると考えられる。

一方で、データサイエンスは単なる統計的思考力を身につけることというより、それを課題解決に向けて組織やチームとしてデータに基づく知識発展のサイクルを回すことを強く意識する点が重要である。De Veaux et al. (2017) で指摘されているように、データサイエンスにおいては

課題抽出と定式化
データの取得・管理・加工
探索的データ解析
データ解析と推論
結果の共有・伝達、課題解決に向けた提案

というプロセスがあり、課題解決と同時にデータからの知識発見や新たな課題発見のためのサイクルが重要である(図表1)。



図表1：データサイエンスにおける課題解決のサイクル

ここで、図表1のようなデータサイエンスのサイクルを理解して実行する際、個人での実行というよりチームや組織での実行という意識が必要となることを強調しておきたい。

アメリカの The National Academies of Sciences, Engineering, and Medicine によるデータサイエンスの学部教育に関する報告書「Data Science for Undergraduates Opportunities and Options」では、学部教育において下記の10領域を重要なものとしてあげている。

- ・ Ethical problem solving 倫理に配慮した課題解決
- ・ Data description and visualization データの記述・可視化
- ・ Data management and curation データの取得・管理・加工
- ・ Statistical foundations 統計の基礎
- ・ Mathematical foundations 数学の基礎
- ・ Computational foundations コンピュータの基礎
- ・ Data modeling and assessment モデリングと評価
- ・ Domain-specific considerations 適用領域の知識
- ・ Communication and teamwork コミュニケーションとチームワーク
- ・ Workflow and reproducibility ワークフローと再現性

ここで、コミュニケーションとチームワークに注目したい。複雑化した課題に向き合う際に、1つの学問領域や成功者のこれまでの成功体験により、課題解決を図るのではなく、多様な人材からなる一チームでデータに基づく課題解決を目指す際に、求められるのがコミュニケーションとチームワークであり、言い換えればリーダーシップである。

このようなコミュニケーションとチームワーク、リーダーシップは、チームや組織での課題解決のためのスキルとして様々な領域で意識されてきた。

高等教育機関においてアカデミックな教育の重要性に加え、近年社会への人材輩出組織としての、人材育成としての教育の重要性が指摘されたのはそう古くはない。日本においては、キャリア教育の意識の高まりとともに、社会人基礎力や人間力というキーワードが大学教育の改善の方向性を示してきた。これは、学生が専門として学び身に付けた実力を、社会において実践する力ともいえよう。現在、高等教育機関において機能別分化が求められるなど、高等教育機関に求められる人材育成の内容が変化してきた。これまでのアカデミックな教育を行うという視点に加え、産業界等での働き手としてのプロフェッショナル教育の視点が重視されている。ひとつの工夫は、大学内に閉じた教育ではなく、社会と繋がりを持った学びの場の提供が求められており、産学連携による教育がある。これは、グローバル化への対応と共に、現代社会における高等教育機関につぎつけられた大きな課題である。データサイエンス教育重視への流れもその一環といえなくもない。

一方、統計家育成においては、米国をはじめとする各国で **Professional Statistician** の資格制が行われてきたが、その要件としてコミュニケーション力やリーダーシップスキルをあげる場合がある。統計の専門家だけで問題解決が図られるわけではなく、分野の専門家と協力して課題解決に当たる際の必要な能力として、コミュニケーション力やリーダーシップスキルが必要となる。リーダーシップスキル養成のためのプログラムは、ビジネススクールには設置されていることが多いが、米国の **University of North Carolina** の生物統計のコースでリーダーシッププログラムがスタートしている。さらに、**American Statistical Association (ASA)**でも、学会としてリーダーシップ教育の重要性を認識し、新たな取り組みをスタートさせている。

近年のデータサイエンス教育において、「ビッグデータ」のキーワードは重要な視点である。多種多様で巨大な規模のデータがスピードを伴い提供される環境で、その情報をどう生かし、問題解決にどう結び付けるかが大きな課題である。ここでの重要な視点は、これまでの統計教育に加えどのような内容を付加すべきかということであろう。Utts (2003) の指摘した市民が身につけるべき内容は、前述のとおり時代が変わっても、データに基づく意思決定を行う際には必ず必要となる事項である。とくに、確率についての常識や統計の有意性の正しい理解は、ビッグデータ時代こそ、その重要性が高まると思われる。

一方、椿 (2013) はマネージメントサイエンスにおける統計科学の役割を指摘するとともに、ビッグデータ時代に必要とされる専門家層の育成についての問題提起を行っている。統計学部や統計学科が存在しない日本の特殊事情を踏まえ、人材育成への対応が喫緊の課題という指摘である。佐々木 (2013) は、ビジネスにおけるデータ分析のプロフェッショナルの必要性と、そこに求められるビジネススキル、分析スキル、IT スキルのバランスの重要性を示すとともに、バランスのとれた人材の確保の困難性を指摘している。さらに、人材不足の具体的な解決策の方向性を示すとともに、立教大学経営学部での産学連携のデータサイエンティスト養成講座の試行の紹介も行っている。そこでは、統計の理論学習に重点を置くのではなく、実際のデータと分析のためのツールを使い、グループでのディスカッションを含むグループプロジェクトとしての実践を行っている。この点で重要なことは、個人での分析力の向上ということだけでなく、グループとして分析を行い意思決定につなげられるかという点であろう。米国の統計学会が認証しているプロフェッショナル・スタティスティシャン (P-STAT) の資格でも、統計分析の知識や技術力に加え、コミュニケーション力

やリーダーシップが求められている。ビッグデータ時代において、その傾向は、一段と高まると思われる。データの多様性や規模が拡大するほど、多様な専門性を持った専門家の協働の必要性が増大するであろう。チームとしての問題解決を図る上で、コミュニケーション力とリーダーシップは不可欠で、そこに分析力を組み合わせることが重要と考える。

2. データサイエンス教育における倫理教育

データサイエンス教育では、統計学に基づく分析手法を身につけるだけではなく、コンピュータで実データを処理するスキルやその際の注意点を習得することが重要である。また、ネットワークを介したグループでの共同作業により、データサイエンスの課題解決を導くことも想定され、ネットワークに繋がったコンピュータを扱うにあたり情報倫理の基礎を理解し、情報セキュリティへの理解も求められる。

情報倫理に関しては、知的財産に関する法律、個人情報保護に関する法律、不正アクセス行為の禁止等に関する法律などを含めた法規が情報社会で果している役割への理解、情報社会で生活する上でのマナー、モラル、倫理の意義や、法を遵守することの重要性の理解、さらに、現代情報社会では技術革新のスピードが非常に早く、法規や制度の対応が間に合わないケースがあること、そのために個人のモラルや倫理に基づく正しい対応が期待されることへの理解も含まれるであろう。

一方、データに関連する法律・規制も学ぶ必要がある。立教大学社会情報教育研究センターには、政府統計部会が設けられ、公的統計への理解を進める学びの場を提供しているが、個人情報保護法や統計法はデータに関する基本的な法規であり、後者では、公的統計が行政利用だけではなく、社会全体で利用される情報基盤として極めて重要であり、根拠に基づく（データに基づく）政策立案の基礎であることを学べるよう努めており、これもデータサイエンス教育における重要な要素である。

データサイエンスを応用し多種多様で大量のデータを分析した成果は、社会に多大なる便益をもたらす。その一方で、深刻な倫理的な問題を引き起こす可能性があることが指摘されている。例えば、個人情報の曝露や、プロファイリング等の自動処理による年齢・民族・性などで層別された特定のグループへの差別や人権侵害である。ここでは特に、倫理に配慮したデータ収集や利活用、匿名化されたデータであっても個人情報を曝露するリスクがあることへの理解を求める必要がある。

さらに、データの収集においても、倫理に配慮しなければならない。立教大学では社会調査士資格を取得するための科目設置が多く学部で行われており、その科目群のなかでは、調査における倫理教育が行われており、その意識は一定程度浸透していると思われる。ただ、人工知能の社会実装の場を中心として、データの活用における新たな倫理問題も台頭しており、その対応が大きな課題といえる。

3. 立教大学におけるデータサイエンス教育

立教大学では、2016年からグローバル教養副専攻がスタートし、社会情報教育研究センターが中心となり、2018年度からデータサイエンス副専攻がスタートした。このデータサイエンス副専攻の提案書では、「2013年6月に閣議決定された『日本再興戦略-JAPAN is BACK-』の「4. 世界最高水準のIT社会の実現、⑥産業競争力の源泉となるハイレベルな

IT 人材の育成・確保」で述べられている“IT やデータを活用して新たなイノベーションを生み出すことのできるハイレベルな IT 人材”の育成を目指す」とあり、「具体的には、データ活用力や IT 技術を身につけることで、メジャーとしての専門性をよりグローバルに活用できるための副専攻プログラムを目指す。新たなイノベーションの担い手としての専門性を重視しながら「データ活用力や IT 技術を身につける」副専攻を用意する」としている。このデータサイエンス副専攻では、指定された 16 単位の単位修得が求められる(図表 2)。

履修モデル:<テーマ>Data Science(データサイエンス)

系列	目的	科目		必要 単位数
第1系列	調査の仕組みを学ぶ	多彩な学び「社会調査入門」「社会調査の技法」 学部展開科目*1		4単位 以上
第2系列	データ活用の基礎的な力を身に付ける	(2-1)基礎科目 多彩な学び「データ分析入門」「データの科学」「数学の世界」 学部展開科目 *1	(2-2)先端科目 多彩な学び「多変量解析入門」「統計情報で社会・経済を診断する」 学部展開科目 *1	8単位 以上*2
第3系列	データ活用の発展的な力を身に付ける	(3-1)基礎科目 多彩な学び「Introduction to Statistics 1, 2」	(3-2)先端科目 2019年度開設科目はなし	
海外体験	海外を知る	Arts & Science Courseの海外体験に準ずる		認定
修了に必要な総単位数				16単位

*1 学部展開科目における履修条件(他学部履修の可否、先修規定等)は、科目設置学部の履修規定等に基づく。

*2 第2系列および第3系列の基礎科目(区分2-1、3-1)から4単位以上、第2系列および第3系列の先端科目(区分2-2、3-2)から4単位以上をそれぞれ修得すること。

1

図表 2 : 立教大学データサイエンス副専攻

展開されている科目としては、2019年度で、第1系列が10科目、基礎科目は61科目、先端科目38科目となっている。なお、英語での開講科目は、基礎科目で2科目が展開されており、今後数科目追加される予定である。また、7科目はオンデマンド科目として展開されている。もちろん、池袋と新座の両キャンパスそれぞれで、十分な科目が展開されており、副専攻修了はキャンパスを問わず可能となっている。

立教大学で実施しているデータサイエンス副専攻における具体的な科目構成であるが、社会調査や統計学、データ分析関連の科目群と、公的統計の利活用に関する科目などで構成されている。前節で述べた内容全般をカバーできる内容とはなっていない点やデータに基づく課題解決を含む実習系の科目がない点などの課題がある。しかし、グローバル教育センターが提供しているグローバルリーダーシッププログラム(GLP)との連携などが今後期待できる一方、倫理面の教育の強化は必須であると思われる。大規模データからの予測と個人情報との関連などをきちんと意識しながら、課題解決につなげられることが重要で、そのような内容の学びの場を、機械学習等の新たな分析手法の学びの場の提供と合わせて行う必要がある。

III まとめ

データサイエンス教育の重要性は、社会におけるデータの普及や活用の推進、さらにエビデンスに基づく意思決定が重視されるほど増すものと思われ、そのための人材育成が課題となる。日本においては、統計学部や統計学科がないことが問題として指摘されてきた。しかし、滋賀大学をはじめとするデータサイエンス学部の創設や、立教大学で実施しているデータサイエンス副専攻の実施は、今後大いに広まっていくと推察される。その中で、課題発見課題解決のための1つの方法論としてのデータサイエンスの位置づけはまだ定めてはいない。また、国が主導して議論が進んでいるデータサイエンスのモデルカリキュラムについての議論や実施もこれからである。大学におけるデータサイエンス教育は、今後広まっていくことに疑問の余地はないが、その教育の質保証や教員確保など課題も多い。特に、数理重視ではなく、課題解決につながる統計的思考力や一定のコンピュータスキル、さらにコミュニケーション力やリーダーシップスキルの育成と、必要となる倫理観の醸成が重要となるであろう。そのために、グループや組織として取り組む実践の場の提供が重要なポイントとなる。このことから、高等教育機関と産業界が協働してのデータサイエンス人材育成の取り組みが重要となるであろう。

参考文献

- 北川源四郎, 2015, 『ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会報告書』, 大学共同利用機関法人情報・システム研究機構.
- 佐々木宏, 2013, 「ビッグデータ・アナリティクスの組織適用とデータサイエンティスト」, 『経営システム』第3巻第4号, 237-241.
- 椿広計, 2013, 「ビッグデータ時代のアナリティクス—データの価値を増大させるヒトとコト—」『経営システム』第3巻第4号, 218-223.
- 山口和範, 2014, 「大学におけるデータサイエンス教育」『品質』44巻3号, 日本科学技術連盟.
- 山口和範, 大橋洸太郎, 大川内隆朗, 丹野清美, 2014, 「データ活用力の育成を意識した統計教育」『大学教育と情報』2014年度No.4, 32-35.
- 渡辺美智子, 2013, 「知識基盤社会における統計教育の新しい枠組み ~科学的探究・問題解決・意思決定に至る統計思考力~」『日本統計学会誌』第42巻第2号, 253-271.
- De Veaux, R. D. *et al*, 2017, Curriculum Guidelines for Undergraduate Programs in Data Science, *Annual Review of Statistics and Its Application*, 4, 15-30.
- Utts, J. 2003, What Educated Citizens Should Know about Statistics and Probability, *American Statistician*, 57(2): 74-79.

Summary

Challenges of Data Science Education
: The Direction of Data Science Education in Higher Education

Kazunori Yamaguchi
Seiichi Yamaguchi
Minoru Kadota

In recent years, there has been much discussion of data science education in higher education. Data science departments have been established at several universities in Japan, where there were no departments with the name “statistics.” In addition, the government and various industries have requested that math and data science education be compulsory. Some universities are making data science courses required subjects for all students. Rikkyo University began a data science minor program based on social research and statistics in 2018. However, there still exists various definitions and issues regarding the area of data science and data-scientists. In this paper, we discuss the current status of data science education and the future issues based on the recent trend of statistical education reform. The topic of data science education discussed here is regarding students receiving literacy level education.

Key words: statistical education, statistical thinking, communication skill, problem solving