

見解

大学における数理・データサイエンス・AI 教育
の中での統計科学の教育について



令和5年（2023年）9月26日

日本学術会議

数理科学委員会

数理統計学分科会

この見解は、日本学術会議数理科学委員会数理統計学分科会の審議結果を取りまとめ公表するものである。

日本学術会議数理科学委員会数理統計学分科会

| | | | |
|------|-------|--------|--|
| 委員長 | 竹村 彰通 | (連携会員) | 滋賀大学学長 |
| 副委員長 | 栗木 哲 | (連携会員) | 大学共同利用機関法人情報・システム研究機構 統計数理研究所統計思考院長・教授 |
| 幹事 | 青嶋 誠 | (連携会員) | 筑波大学数理物質系教授 |
| 幹事 | 渡辺美智子 | (連携会員) | 立正大学データサイエンス学部教授 |
| | 朝日 弓未 | (連携会員) | 東京理科大学経営学部経営学科教授 |
| | 上田 修功 | (連携会員) | 理化学研究所革新知能統合研究センター副センター長、NTT コミュニケーション科学基礎研究所客員フェロー |
| | 北川源四郎 | (連携会員) | 大学共同利用機関法人情報・システム研究機構 統計数理研究所名誉教授、総合研究大学院大学 名誉教授 |
| | 小林 正人 | (連携会員) | 明治学院大学経済学部教授 |
| | 西郷 浩 | (連携会員) | 早稲田大学政治経済学術院教授 |
| | 佐藤 美佳 | (連携会員) | 筑波大学システム情報系教授 |
| | 下津 克己 | (連携会員) | 東京大学大学院経済学研究科教授 |
| | 椿 広計 | (連携会員) | 大学共同利用機関法人情報・システム研究機構 理事、情報・システム研究機構統計数理研究所 長・名誉教授 |
| | 椿 美智子 | (連携会員) | 東京理科大学経営学部経営学科教授・経営学部 長・大学院経営学研究科長 |
| | 中西 寛子 | (連携会員) | 大学共同利用機関法人情報・システム研究機構 統計数理研究所特任教授 |
| | 中村 佳正 | (連携会員) | 大阪成蹊大学学長 |
| | 福重 元嗣 | (連携会員) | 大阪大学大学院経済学研究科教授 |
| | 松井 知子 | (連携会員) | 大学共同利用機関法人情報・システム研究機構 統計数理研究所研究主幹・教授 |
| | 南 美穂子 | (連携会員) | 慶應義塾大学理工学部教授 |
| | 山下 智志 | (連携会員) | 大学共同利用機関法人情報・システム研究機構 統計数理研究所副所長・教授 |
| | 山田 秀 | (連携会員) | 慶應義塾大学理工学部教授 |
| | 吉瀬 章子 | (連携会員) | 筑波大学システム情報系長 |
| | 美添 泰人 | (連携会員) | 一般社団法人新情報センター会長、青山学院大 学名誉教授 |

本見解の作成にあたり、以下の職員が事務を担当した。

| | | |
|----|-------|-----------------------------|
| 事務 | 松室 寛治 | 参事官（審議第二担当）（令和4年7月まで） |
| | 佐々木 亨 | 参事官（審議第二担当）（令和4年8月から） |
| | 高橋 直也 | 参事官（審議第二担当）付参事官補佐（令和5年3月まで） |

| | | |
|----|----|------------------------------|
| 柳原 | 情子 | 参事官（審議第二担当）付参事官補佐（令和5年4月から） |
| 稲元 | 祥吾 | 参事官（審議第二担当）付審議専門職付（令和5年1月まで） |
| 近藤 | 慈恩 | 参事官（審議第二担当）付審議専門職付（令和5年2月から） |

要 旨

1 作成の背景

統計科学をめぐる状況は、ここ 10 年ほどで急速に変化している。ビッグデータ時代は大きく進展し、特に深層学習に代表される AI(人工知能)技術の急速な発展と共に、数理・データサイエンス・AI 分野の重要性が強く認識されるようになった。そのような変化の中で統計科学をどのように位置付けるかは、統計科学の今後の発展にとって重要である。本分科会では、2014 年 8 月に提言「ビッグデータ時代における統計科学教育・研究の推進について」を公表したが、この間の急激な変化を踏まえ、本見解では、2014 年の提言をフォローアップし、統計科学の役割を再考し、今後の統計科学の教育・研究についての提案を示す。

2 現状及び問題点

アメリカをはじめとする諸外国では、2010 年頃からそれ以前と比較して統計科学の教育・研究の進展が加速した。これは、AI 技術が注目を集める中でも、統計科学の基礎的な重要性が認識されているためであると考えられる。一方、日本では最近になってようやく数理・データサイエンス・AI 分野の幅広い教育が政府の方針として取り上げられるようになってきた。データサイエンス系の学部の新設も相次いでいる。また、大学生の教育に加えて、既に社会人となった人々の再教育も我が国にとって重要な課題となっている。さらに、初等・中等教育においてもさらなる充実が必要である。このように、統計科学の教育の強化が求められる中で、統計科学を適切に教えることのできる教員の不足が深刻な課題となっており、早急な解決が必要である。

3 見解の内容

(1) 数理・データサイエンス・AI 分野の理論的基礎としての統計科学の位置付け

ビッグデータの処理において、統計科学の方法論は様々な手法の理論的基礎を与えるものであり、数理・データサイエンス・AI 分野の教育においても、データの正しい解釈のためには、様々な手法をツールとして教育するだけでなく、手法の理論的基礎も含めた教育が重要である。

(2) 数理・データサイエンス・AI 分野の再教育(リスキリング)の推進

社会人に対する数理・データサイエンス・AI 分野のリスキリングの推進が我が国にとって急務となっており、大学の果たす役割も大きい。

(3) 学士課程及び大学院教育が必要とする統計教員の育成

学士課程及び大学院教育が必要とする統計教員が極めて不足しており、教員の育成が喫緊の課題であり、統計エキスパート人材育成事業を大幅に拡大すべきである。

(4) 初等・中等教育における教材、ソフトウェア、デジタル環境の整備と統計教育のさらなる充実

数理・データサイエンス・AI 分野の大学教育の前提として、初等・中等教育における教材、ソフトウェア、デジタル環境を整備し、統計教育のさらなる充実を図る必要がある。

目 次

| | | |
|---|---|----|
| 1 | はじめに..... | 1 |
| 2 | 統計科学の役割 | 2 |
| | (1) 統計科学の特性 | 2 |
| | (2) ビッグデータと AI の時代における統計科学..... | 4 |
| | (3) 統計科学の産業応用 | 7 |
| | (4) データ倫理の問題 | 10 |
| 3 | 国際的動向..... | 11 |
| | (1) 高等教育の動向 | 11 |
| | (2) 初等・中等教育段階における統計の位置付け | 12 |
| 4 | 日本におけるこれまでの統計科学教育改善の動き | 14 |
| | (1) データサイエンス系学部等の新設や大学院の動き | 14 |
| | (2) 数理・データサイエンス・AI 教育の動きと認定制度..... | 14 |
| | (3) 学習指導要領の改訂 | 15 |
| | (4) 統計教育連携ネットワークと統計検定..... | 16 |
| 5 | 日本における統計科学の教育・研究体制の現状と課題..... | 18 |
| | (1) デジタル社会が求める専門人材..... | 18 |
| | (2) 大学院における統計科学教育と統計エキスパート人材育成..... | 18 |
| | (3) 統計科学分野の研究体制 | 19 |
| 6 | 統計科学教育・研究推進への提案..... | 20 |
| | <参考文献>..... | 21 |
| | <参考資料1>審議経過 | 35 |
| | <参考資料2>公開シンポジウム「数理・データサイエンス・AI 時代における統計科学 の教育及び研究について」 | 36 |
| | <付録>..... | 38 |
| | (1) 図表..... | 38 |
| | (2) 諸科学におけるビッグデータの統計解析 | 42 |
| | (3) 諸外国の大学におけるデータサイエンス教育の例..... | 44 |
| | (4) 諸外国の教科書と日本の指導要領..... | 47 |

| | | |
|-----|---|----|
| (5) | 数理・データサイエンス・AI 教育強化拠点コンソーシアムのモデルカリキュラム..... | 49 |
| (6) | 統計関連の学会と科学研究費の状況..... | 49 |

1 はじめに

ここ 10 年ほどで、統計科学をめぐる状況は大きく変化した。本分科会では、2014 年 8 月に提言「ビッグデータ時代における統計科学教育・研究の推進について」[1] (以下「前提言」と言う)を発出し、ビッグデータ時代に求められる統計科学の教育・研究体制について提案を行った。その後も、ビッグデータ時代は大きく進展し、特に AI 技術の急速な発展と共に、数理・データサイエンス・AI 分野の重要性が強く認識され、全ての大学生がこの分野を学ぶべきであるとされるようになった。本見解では、2014 年の提言をフォローアップし、統計科学の役割を再考し、今後の統計科学の教育・研究についての提案を示す。

統計科学は、数理・データサイエンス・AI 分野において数学に基づく理論的基礎を与えるものであり、この分野の手法の様々な他分野への応用が進むほど、その重要性が増すものである。理論的基礎をおろそかにして、応用のみを進めることには注意が必要である。特に、コンピュータのユーザインターフェースが進歩し、便利なツールが使えるようになってきたため、大学の教育においてもツールの使い方という観点から教育がなされることが多くなっている。応用の進展自体は望ましいものであるが、一方で、ツールの使い方という観点のみから教育が進められると、得られた結果の正しい解釈ができないといった問題点も生じる。特に、最近の AI 手法は、データにあてはめられるモデルが非常に複雑になり、モデルがブラックボックスになるという問題点が指摘されている。

この点からも、統計科学の役割や特質を明らかにし、それを大学教育に活かすことは重要である。2 章では特に、データの収集法とそれに基づくデータ生成モデルの構築や、因果推論の重要性などを論じる。3 章では、統計科学の国際的動向についてまとめる。詳しい情報が得られるアメリカやイギリスの状況については、いくつかの大学での教育を取り上げてその動向を論じる。また、初等・中等教育段階における統計の位置付けについても調査結果を示す。

4 章では、国内の状況について、データサイエンス系学部新設の動き、数理・データサイエンス・AI 教育強化拠点コンソーシアムを中心とした全国展開、学習指導要領の改訂による中等教育の改善などについて述べる。

5 章では、このような改善の中で今後さらに解決すべき問題として、デジタル人材の育成及び統計教員の養成などについて述べる。そして、最後の 6 章では、以下の 4 点について提案する。

- 提案 1 数理・データサイエンス・AI 分野の理論的基礎としての統計科学の位置付け
- 提案 2 数理・データサイエンス・AI 分野の再教育(リスキリング)の推進
- 提案 3 学士課程及び大学院教育が必要とする統計教員の育成
- 提案 4 初等・中等教育における教材、ソフトウェア、デジタル環境の整備と統計教育のさらなる充実

2 統計科学の役割

本章では、統計科学の特性、統計科学の最近の変化、統計科学の産業応用、データ倫理などの様々な観点から、統計科学が数理・データサイエンス・AI 分野の理論的基礎であることを論じる。本章の議論は、6章の提案1の根拠を与えるものである。

(1) 統計科学の特性

統計科学は、分析目的に応じてデータを収集し、データの整理・加工・推論などの統計分析により知見を得て、現象の理解・予測・意思決定を行うための方法を提供する科学である[2], [3]。広義には、それらを支援するための周辺領域、例えば、データの保存方法[4]、公的な統計機構の在り方や産業分類などの統計基準の作成[5]、データを扱う上での倫理[1]などに関する研究も含まれる。さらに、近年では、観察・測定 of 技術進歩により、より多様で広い範囲のデータが利用できるようになった。これにより、データの収集方法とその分析方法との対応関係を明らかにする統計科学の重要性が増している。

データと分析者との接点において発生する諸問題を扱うというその性質から、統計科学はあらゆる分野で必要とされる方法論として機能する。実際、近年興隆の著しいデータサイエンスにおいても、統計科学は重要な構成要素の1つと位置付けられている。すなわち、データサイエンティストには、1)複雑なデータをコンピュータで扱える情報処理の能力、2)データを分析するための数理及び統計科学の知識、3)データを処理・分析して得られた結果から価値を創造する力、の3つの能力が必要とされる[6]。特に、数理科学や統計科学の知識を欠いた情報処理能力と価値創造力の2つだけの結び付きは「危険地帯」(danger zone)であると警告されている[7]。

他方で、統計科学が活用される種々の科学分野には、それぞれ固有の論理と、それを扱う学問的・技術的な方法があり、それらが相異なる場合も多い。これらの個別の科学分野に共通して適用し得る方法を統計科学が提供できるのか、できるとすればなぜそれが可能なのかが問われなければならない[8]。これらの問いに答えることは容易ではない。少なくとも、種々の科学分野にはそれらに適応した統計科学的手法が求められる。このことは、計量政治学や計量経済学、計量生物学などがそれぞれ独自の発展を遂げていることにも表れている。「不備のない、周到に設計された無作為化実験のデータがあるのであれば、計量経済学という個別の領域が存在する余地はない」[9]という叙述は、個別の科学分野が固有のデータ分析方法を必要とすることを端的に表している。

しかし、同時に、個別の分野で必要とされるデータ分析の手法は統計科学と無関係ではありえない。その一例として、社会科学で重視されるようになった因果推論を挙げる。政治学では、初学者を対象とした教科書において、因果推論の重要性が強調されるようになっている(例えば、[10])。経済学においても、最近の計量経済学の教科書では因果推論が重視されている(例えば、[11])。このことは、これらの分野において政策効果の測定が望まれることによる要請である。ただし、相関と因果とが見分けられるためには、多くの前提条件が必要とされる。それらの条件は、統計科学的な研究によって明らかにされる。例えば、社会科学で用いられる因果推論の枠組みの1つは、統計科学における欠測データ処理を援用することによって与えられた[12]、[13]。かつて同時方程式モデル

の推定方法として考案された操作変数法を因果推論の分析用具として利用するときにも、前提条件の成否に統計科学的な検証は欠かせない[14]。

最近では、ビッグデータが利用可能となり、データ駆動型の研究という表現もよく聞かれるようになった。これは、特定の仮説を前提とせずに、データから諸科学にとって有用と思われる特徴を引き出す接近法を指す。ただし、類似の接近法は過去にも存在し、探索的データ解析[15]やデータマイニング(data mining)[16]などがこれにあたる。しかし、観察データ、すなわち、統計科学的に注意深く設計された収集方法によらない方法で得られたデータに基づいてデータ駆動型の研究を進めるときには、特に注意が必要である。なぜなら、交絡因子のために見かけ上の相関が発生する可能性などがあるからである[6]。最終的には、確認的データ解析を施すなど統計科学的に厳密な手法を併用するか、少なくとも、分析結果にどのような偏りが生じ得るのかを統計科学の見地から検証しなければならない[17]。

統計科学においては、古くからデータの収集方法と分析手法との関係が研究されていた。実際、無作為化比較実験や実験計画法、標本調査法など、データの収集方法を専門に研究する分野が存在する[18]。技術進歩によって容易に入手できるデータが急増した結果、データ収集に関する前提条件を無視した分析手法を手軽な標本(convenience sample)に適用することを戒める統計科学の役割が一層増している[14]。

統計科学は、データを分析する手法を諸科学に提供する役割を果たしてきた。科学の重要な目的の1つは、現実を抽象してその中から本質的と見なし得る関係をモデル化する[8]ことであるため、統計科学では入力される情報と出力される情報との間の対応が明瞭であるデータ生成モデルⁱが設定されることが多い。これに対して、情報学やAIの分野において考案された手法には、予測モデルに基づくアルゴリズムを主体とする傾向が強い。最近の大規模な予測モデルの性能の向上は顕著であり、様々な場面で有用である。しかし、一般にモデルを複雑にすればデータへのあてはまりは向上するため、予測モデルは複雑となり、入力と出力の対応関係を理解するという目的にはデータ生成モデルの方が適当なことが多い[19]。

因果推論やデータ駆動型の研究は、統計科学が諸科学に方法論を提供し得る例の一部に過ぎない。そのような統計科学の役割は今後も変わらないと予想される。日本学術会議の多くの委員会からの提言の中でも、統計科学の重要性が指摘されている。それらの提言については付録表1にまとめているⁱⁱ。

統計科学の貢献は、公的統計にも及ぶ。統計委員会[20]は、各府省が Evidence-Based Policy Making (EBPM)の推進をはじめとする統計の利活用の促進のため取り組むこととしている(p. 30)。同時に、毎月勤労統計調査や建設工事受注動態統計調査の事案において、調査の設計や集計方法の変更が本来作られるべき統計と整合しないことが見逃された点を重視して、幹部職員が主体的・積極的に統計作成プロセスの適切なマネジメントに取り組むことを求めている(p. 31)。その実現のための柱の1つとして、人材を含む統計リソースの確保と有効活用を挙げている。しかし、現状において、専門的人材は不足しており、また、専門的な知識や経験を備えた職員が有効活用される仕組みが整っているとは

言えない(p.38)。また、いわゆるオルタナティブデータの利用にも統計科学の方法論が必要とされる。2022年12月に提出した本分科会の中長期戦略[21]で提案した科学行政官、すなわち、行政機関において自ら博士号を有し、異分野の科学者を束ね分野横断的な研究をプロデュースできる人材の育成も重要な課題である。

近年では、公的統計だけでなく、政府全体でデータの利活用を促進することが提唱されており、そのために、統計科学を含むデータサイエンスの能力を備えた人材の育成が強調されている。例えば、[22]にはオープンデータや仮名加工情報制度などをはじめとするデータの利活用とそれを担う人材の育成が謳われている。この点は[23]でさらに詳細に検討されている。[23]でも触れられている統計科学に関連する教育の在り方については、[24]-[26]に具体的な政府の取組が述べられている。

(2) ビッグデータとAIの時代における統計科学

前提言[1]では、ビッグデータの利活用の重要性と統計科学が果たす役割について述べたが、前提言時点と大きく異なる点として、AI手法、特に深層学習(deep learning)の発展がある。以下の①では、計算技術の発展に伴う統計科学の変化を振り返る。また、②では、深層学習と統計科学手法の類似点と相違点、補完的關係、得失等について論じる。なお、AIの定義に関しては様々な議論があるが、本節では深層学習を中心とする第3波のAIブームで注目されている手法を意味するものとする。本節以外で「AI教育」などと言う際には、AI戦略2022[27]等と同様により広い意味で用いる。

① 計算技術の発展に伴う統計科学の変化

コンピュータの飛躍的な発達、統計科学に大きな変化をもたらした[17]。その変化をデータの収集及びデータの分析の2つの側面から議論する。

ア 収集可能なデータの変化

コンピュータと共にネットワークやセンサーなどの技術が発達した結果、我々が入手できるデータの規模と種類は急激に増大した。構造化データに分類される従来からあった時系列データや横断面データにしても、秒またはそれよりも短い間隔で記録された高頻度データや、1つの個体が膨大な種類の観察値を伴う高次元データが作られるようになった。また、極値分布など、いわゆるロングテールの分布の研究も進んでいる。これらは新しい統計科学的な分析手法を必要とするような大きな変化である[28]、[29]。

従来とは異なる形の非構造化データも容易に入手できるようになった。その代表的な例は、テキストデータや音声データ、画像データ、並びに生成AIによる同種のデータである。これらのデータについては、原データから分析しやすいデータ(特徴量)をどのように作成するかも研究の対象となる[6]。大量かつ複雑な構造を持つデータの総体はビッグデータと呼ばれるようになった。

イ 利用可能な分析方法の変化

コンピュータをはじめとする技術革新は、統計データの分析手法にも進歩をもたらしている。20世紀前半における統計科学は、次元の低いパラメータによって規定される統計モデルを強く志向し、数学的な厳密性が重視された[17]。このことは、サイズが小さく、次元もそれほど高くない標本データから有用な結論を導かなければならなかったという事情にもよる[17]。

1970年代以降、コンピュータを活用した計算機集約的(computer-intensive)な手法が開発されるようになった。その中には、回帰決定木分析やEMアルゴリズム、ジャックナイフ法やブートストラップ法などのリサンプリング法、交差検証法、マルコフ連鎖モンテカルロ法によるベイズ統計学的推測、などが含まれる[17]。注意すべき点は、これらの手法の開発においても、データが生成されるモデルを設定し、設定されたモデルの下で手法の理論的妥当性の研究が重視されたことであるⁱⁱⁱ。

1990年代後半から現在まで、統計データの規模の拡大や構造の複雑化に合わせて、コンピュータを活用した新しい手法が次々と開発されている。これらには、大規模仮説検定とFalse Discovery Rate(FDR)、スパース推定とLasso、ランダムフォレストとブースティング、ニューラルネットワークと深層学習、サポートベクトルマシンとカーネル法、などが含まれる[17]。これらの新しい手法の多くは、予測モデルに基づくアルゴリズムを主体とするものである。

入力から出力までの仕組みが複雑で、両者の対応関係を理解することが困難なモデルはブラックボックスと呼ばれる。逆に、両者の対応関係が明瞭であるモデルはホワイトボックスと呼ばれる。一般に、データへのあてはまりを向上させるにはモデルを複雑にしなければならなくなる。このため、予測を目的としたモデルはブラックボックスとなりやすい。実際、ニューラルネットワークをはじめとする深層学習のモデルにおいてはチューニングが必要なパラメータが莫大な個数となる[30], [31], [32], [33]。これに対して、データの生成モデルから出発する場合には、入力と出力との対応が明瞭である場合が多い。統計データを分析する手法を諸科学に提供する役割を果たしてきた統計科学においては、生成モデルが設定されることが多い。生成モデルの設定を基本とする統計科学の研究者の中には、予測モデル手法の数理的な基礎を明らかにすべきだと主張する者が少なくない[17]。例えば、ブースティングの考え方を早い段階で実現したAdaboostは、ロジスティック回帰の観点からも解釈できる。また、深層学習やAIの分野においても、予測モデル手法の数理的な基礎付けを重視する研究者もいる(例えば、[30], [34])。

以上に述べたように、歴史的には、統計的モデルは生成モデルから機械学習を用いた予測モデルへその範囲を広げていったと理解される。しかしながら、予測モデルの考え方は既に1970年代に統計数理研究所の赤池弘次により提唱された情報量規準統計学に見られる。赤池は、限られたデータと客観的知識、経験的知識を有機的に利用し、統計モデルの構築を通して、仮説の提案と検証を繰り返し真理に近づいていくことが統計的推論であるとしている。その際に役立つ手法が、赤池情報量規準(AIC)で

ある[35]。この立場によれば、統計モデルは正しいものではなくあくまでも現実を近似するものであること、また、その選択は目的にも依存することとされる。これは、当時は非常に革新的な考え方であったが、徐々に受け入れられ、その後の赤池自身のベイズ統計の積極的利用と相まって、現時点では知識情報処理の基本的考え方として確立している[36]。

統計科学がビッグデータの活用に本質的である分野は多い。付録(2)で地震、データ同化、天文等の分野における統計科学の役割の具体例を論じている。

② 深層学習と統計科学

生体の神経回路網を模した人工ニューラルネットワークは、1957年のパーセプトロンに端を発する。これは1層のニューラルネットワークと呼ぶべきものであるが、その多層化と最適化の方法（誤差逆伝搬法：バックプロパゲーション）も早い段階で考案され、理論的研究や実用化研究も行われていた。2000年代になり、計算機の進化を背景とし、層数を大幅に増やし適切に学習された多層（深層）ニューラルネットワークが、従来の常識を越えた高精度なパターン認識能力を持つことが見いだされ、それが現在の深層学習の興隆に繋がっている。

その後、実用研究が飛躍的に進む一方で、深層学習の様々な得失が明らかとなりつつある。例えば、深層ニューラルネットワークは優れた関数近似性能を持つ一方で、現時点ではなぜそのような高パフォーマンスを持つのか分からない。加えて、出力結果の解釈ができないブラックボックスであり、信頼性、安定性、再現性が不十分、などの問題点が指摘されている。以下では、それらと統計科学のアプローチの関わりについて現状と展望を論じる。なお、この分野の研究については、日本の若手統計科学研究者が精力的に取り組んで成果を上げている（例えば、[37], [38], [39]）。深層学習以外の機械学習の分野でも、近年はその信頼性、再現性を向上する必要があるという認識の下で、因果推論や選択的推論の研究が盛んである[40]。また、最近では大規模言語モデルに基づく統合的なAIシステムの説明可能性にも注目が集まっている。

ア ブラックボックスに対する統計的アプローチ

機械学習における教師あり学習結果は、解釈困難なことが多い。深層学習は、その最たるものである。例えそれがどんなに良い性能を持っていても、内部がブラックボックスであっては、医療行為や人事評価など人間に対する意思決定に使うことは倫理上の問題が起こり得る。また、自動運転などで事故が発生した場合の法的問題も複雑になる。それに対処するアプローチの1つとして、ブラックボックス自体は所与として、その動作を統計科学的手法で理解しようという試みがある。

PD (partial dependence) プロット[41]は、予測値の各引数（予測変数）の依存性（主効果）と、予測変数のペアの間の低次の交互作用効果を可視化して理解するためのものである。これにより、例えば病気のリスクと危険因子の間に明らかに逆転した構造が含まれているか、消費者信用リスクの予測に社会的に公正でない因子が関与し

ているかを確認できる。PDプロットを改良したものにALE(accumulated local effects)プロット[42]がある。

ブラックボックスモデルでも仮定できるような非常に弱い前提の下でも成り立つ統計推測手順の研究も行われており、Conformal Prediction[43]はデータが交換可能(順番を変えても確率的性質が変わらない)という仮定のみで、1期先オンライン予測を信頼係数付きで与える手順を提供する。

イ 深層学習とノンパラメトリック回帰・関数近似

深層学習の性能を評価する方法として、与えられた入力に対して目的の識別や予測を持った機械学習関数にどのくらい近づけることができるか、すなわち関数近似性能を測る方法がある。これは統計科学では汎化誤差の問題として枠組みが設定され、歴史的に多くの研究がある。深層ニューラルネットワークはこのノンパラメトリック回帰モデルの一種と見なすことができる。

統計科学におけるノンパラメトリック回帰モデルの研究は、1960年代のNadaraya-Watson推定量に端を発するカーネル回帰法、その再生核ヒルベルト空間での取扱い、回帰関数の滑らかさと汎化誤差の関係(Stoneの定理)など数理統計学の中心的な研究課題の1つであった。深層ニューラルネットワークを統計学の枠組みのノンパラメトリック回帰と捉えて、達成し得る汎化誤差の限界を求めたのはM. ImaizumiとK. Fukumizu[44]であり、通常の回帰では達成できない精度を深層学習では達成できることを示した。

深層学習の内部には立ち入らず、関数近似のためのツールとして統計学データ解析の手順に組み込む立場もある。統計学の基本的な手法である並べ替え検定は、帰無仮説の下でのデータを再現するための手続きと見ることができる。ロマノラはそのようなデータをノックオフデータと呼び、そのデータを深層学習で生成するアプローチを提案した[45]。生成ネットワークと識別ネットワークから構成される敵対的学習(GAN)による事後分布乱数生成も同様の例である[42],[46]。

(3) 統計科学の産業応用

産業界は、統計科学に通じる統計家やデータサイエンティストを必要としている。このことは、アメリカの雇用統計[47]に表れている。アメリカの職業分類には、独立した項目として統計家(15-2041)とデータサイエンティスト(15-2051)が設けられている。前者の定義は、「利用可能な情報を提供するために、データを収集・前処理・解釈・要約するための数理または統計科学的な理論または方法を開発または適用する」職業、後者の定義は、「データ指向型プログラム言語や視覚化のためのソフトウェアを使って、原データを有意な情報に変換するための技術や分析的なアプリケーションを開発または実行する」職業である[48],[49]。どちらもデータを利用して結果を導く点で共通しているが、強いて言えば、前者が統計科学の理論研究を志向するのに対して、後者は理論の実践的適用に重きを置くという違いがある。2021年5月には、前者の就業者が31,370人、後者

のそれが 105,980 人であった（付録表 2）。さらに、データサイエンティストの雇用は 2021 年から 2031 年にかけて 35.8%増加すると予想されており、統計家の雇用は 32.7%増加すると予想されている[47]。

一方、日本については、日本標準職業分類には、統計家もデータサイエンティストも分類項目となっていない。そのこと自体、社会で必要とされる統計科学に関連した人材が少ないことを表しているとも言える。内閣官房教育未来創造会議[50]によれば、国際経営開発研究所(International Institute for Management Development:IMD)の試算により、日本のデジタル競争力は 28 位（米国 1 位、韓国 12 位、英国 14 位、中国 15 位、ドイツ 18 位）となっている。特に、デジタル／技術スキルやビッグデータの分析と活用、国際経験等の項目において、他国に比べて遅れを取っている状況にある。日本経済団体連合会は、提言[51]を公表し、リテラシーレベルの数理・データサイエンス・AI 教育プログラムの開発及び必修化と、産学が共同してこれらの教育プログラムを実践的な内容にすることを求めている。

産業界からの統計科学に関するニーズは、デジタルトランスフォーメーション(DX)に関する各種の議論からも読み取ることができる（例えば、[52], [53], [54]）。これらの提言書、報告書などにおいて、共通的に論じられているのは、データで一企業を超えて顧客までを含む幅広いプロセスを連動させること、情報処理技術の積極的な活用による幅広い視点でのデータ収集を可能にすることに加え、収集されるデータを統合的に活用できる人材の育成である。このような人材を育成するには、大学での教育に加え、社会人再教育(リスキリング)を拡充する必要がある。日本では、大学の教育の対象はこれまで大学生や大学院生に限られてきたが、認証制度の構築や社会人向け講習など、社会人のリスキリングにも大学が貢献すべき時代となっている。リスキリングの重要性については、4 章(4)及び 5 章(1)でも述べる。

① 製造業における統計科学の必要性

統計科学が製品、サービスによる顧客満足獲得に果たす役割は大きく、この原点は、戦後から始まる製造現場における狭い意味での品質管理である。現在では、製造現場だけでなく、企画、開発、設計、生産、提供、販売、流通、アフターサービス、保守など、各段階における組織的な改善活動の推進のために、基礎的な統計的手法であるヒストグラム、散布図、管理図などの活用が積極的に推進されている。このことは、一般財団法人日本規格協会、一般財団法人日本科学技術連盟が主催し、一般社団法人日本品質管理学会が認定する品質管理検定(QC 検定)の受験者数からも確認できる。QC 検定は 1 級から 4 級で構成され、3 級、4 級の出題は基礎的な統計的手法とその応用を主とし、2021 年末で累積受験者数は 80 万人を超えている[55], [56]。また、職場での改善を統計的手法も活用しながら推進するために、QC サークルを代表とする小集団活動が幅広く企業で導入されている。例えば、一般財団法人日本科学技術連盟に登録されている QC サークルは、2022 年 12 月末で 6 万を、総メンバー数は 56 万を超えている[57], [58]。このように、改善活動などの基盤となる統計科学について、その重要性が

広く理解されている。今後は、IoT 技術を利用したスマートファクトリーなどの建設も進み、製造業における統計科学の手法はますます重要になると考えられる。国際的にはシックスシグマ推進の動きもある[59]。

企業の競争力の源泉たる製品、サービスによる高度な顧客満足獲得のためには、企画から保守の各段階において、基礎的な統計科学的アプローチでは不十分でより深い応用が必要となる。例えば、開発、設計段階における実験計画法の活用など、新たな方法が開発されている[60], [61]。また、企画段階であるマーケティングについては以下で述べる。近年の DX により、企画段階から保守までの一連のプロセスをデータで繋ぎつつあり、製品、サービスの品質という視点から、統計科学的アプローチの必要性は広く認識されている[54]。また、新たな物性を持つ材料の開発のためのマテリアルズ・インフォマティックスの技術も進展している。

② 非製造業における統計科学の必要性と人材育成

非製造業のうち、ファイナンス分野でリスクを科学的に定義したのはマーコビッツ[62]であり、複数の資産で構成されたポートフォリオのリスク計量化と意思決定のための最適化手法を確立した。続いて、マックファーデン[63]が非集計行動モデルとマイクロ経済学の消費者行動理論との関係を数学的に示した。この理論は、現在もマーケティングモデルの主流となっている。さらに、1973 年に確率過程モデルを応用した統計学的発見があり、これが数理ファイナンス（金融工学）という分野を作った。現在も派生商品の理論価格に使われるブラック＝ショールズ方程式[64]である。このモデルは、伊藤清が 1940 年代に提案した伊藤積分を金融市場に応用したものである。また、伊藤積分の応用として企業価値や企業の倒産確率を計算するマートンモデル[65]が提案された。以上の 3 つの業績は、それぞれノーベル経済学賞を受賞した。

数理ファイナンスの実用化においては、統計科学的な考察の不足による挫折があった。まずショールズとマートンが設立した投資会社 Long-Term Capital Management (LTCM) の 1998 年の破綻が挙げられる^{iv}。より大きな挫折は、2008 年のサブプライムローン問題と 2009 年のリーマンショックであり、高度な数学的仮定の下に成り立つ統計モデルも、実務家はその統計科学的前提に対して熟知しなければ逆にリスクを増大させてしまうことが教訓として残された。

このような問題に対応するため、先進国各国はリスク計量化と経営判断のための統計モデルの標準化を図り規制を強化した[66],[67]。しかし、これらの規制も民間銀行に統計学的知識を有した実務家が必要であるが、現状は十分とは言いがたい。

一方、マーケティングにおいては、消費者行動に関するビッグデータの活用が急速に進んでいる。現状では、膨大なデータを 1 社で保有し多くの統計技術者を有している GAFA のようなプラットフォーマーが圧倒的に有利な存在であり、我が国は大きく遅れを取っている。従来の消費者行動モデルの構築技術だけでなく、オルタナティブデータの利用に際してより高度な統計科学的理解が必須となり、データはあるが分析できない状態が続いている。ファイナンス分野においてもマーケティング分野において

も、我が国が米国などの先進諸国と競うためには、高度なデータベースと統計科学の知識を持つ人材を、質、量を共に充実させることが必要である[68]。

(4) データ倫理の問題

数理・データサイエンス・AI に支援されたデータに基づく意思決定や行動を適切なものとするための倫理教育が必要であり、その前提となるのが、国際統計協会倫理宣言[69]にもある「自身のグループの利益の最適化と一致すると共に、他のグループを犠牲にしてはならない」という原則である。倫理上正しい行動ができなくなるリスクが生じる場合として、次の2つが重要である。

第一は、数理・データサイエンス・AI の入力情報となるデータが偏ることで、行動が不正化するリスクである。意図的に偏ったデータを収集することは、倫理的に許容されないことは勿論である。しかし、統計的データ収集計画の裏付けのないビッグデータには、収集法に起因する人為的とは言えない偏りもあり、それが決定を偏らせる一因となることを教育する必要がある。なお、人を対象とした実験や調査では、結果の社会貢献が期待できても対象となるグループの人権尊重が必須である[70]。

第二は、分析や推論技法が妥当でないため、不適切な行動が推奨されるリスクである。ベイズリスク最適化決定、モデル選択を応用した決定、不確かさを考慮し信頼区間や仮説検定を用いた決定は、いずれも数理的には正当性があるが、それらから得られる結論は一般には一致しない。アメリカ統計学会が、P値偏重への注意[71]を發した仮説検定方式に限っても、多重性調整を行うか否かで、決定は一致しない。したがって、データからの意思決定を支える分析や推論方法の適用範囲や使用上の留意点なども理解し、自ら適切な方法の選択ができる能力の涵養は、倫理的な観点からも必要である。逆に、社会的影響や特定のコミュニティへの損害を与え得る結果が導かれる場合には、用いられた分析や推論が、どのように判断を導いたかをデータと共にオープンにする等の説明責任が生じる。特に、深層学習等の AI 技術では、最適化アルゴリズムにより探索されたモデルがブラックボックス化しがちであり、結果がどういう意味でどこまで妥当かが見えにくくなっている。これに関して、提言「科学的知見の創出に関する可視化」[72]は、Stanford AI 100[73]を引用し、「AI の推論メカニズムを適切に可視化することにより、AI の説明責任や透明性を高める必要がある」としている。

3 国際的動向

数理・データサイエンス・AI 教育の中での統計科学の教育については、近時、国際的にも注目されている。新型コロナウイルスの感染拡大が社会の大きな変化を生み、テレワークやオンライン教育をはじめとする新しい生活様式が一般化する中で、我が国の DX 実装の遅れや研究力の低下が指摘されている。本章では、大学における統計教育の国際的動向及び海外の初等・中等教育段階における統計の位置付けについて述べ、日本の遅れを指摘する。

(1) 高等教育の動向

米国統計学会 (American Statistical Association:ASA) による冊子「Statistical Education of Teachers」[74]によれば、データ駆動型社会において、統計リテラシーは、統計分析を行う研究者だけではなく、データに基づいて日常の意思決定を行うために、一般の市民にとっても不可欠な能力になりつつあるとしている。例えば、時事問題に関するメディア報道の追跡、財務上の意思決定、または健康リスクの評価など、統計情報を処理する能力は、現代社会を生き抜くために重要なものであるとして、主に数学教育や統計教育に携わる者を対象に統計教育のガイドラインを示している。また、統計的思考と数学的思考、さらに、計算的思考の違いを明確に示した上で、統計教育上の教育・学習における留意点を示している。

このような視点から、近年の世界の統計学に関連した学科を眺めるとき、2つの考え方があり、その1つは、近年の AI 技術に代表される計算的思考と統計的思考の融合であり、その顕著な例として、米国の統計関連学部の教育と研究の動向がある。もう1つは、数学に基づき統計的思考を捉え、近年の AI 技術に代表される計算的思考の技術も、その枠組みの中で解明しようとする立場である。その例として、英国ケンブリッジ大学、あるいは王立統計学会 (Royal Statistical Society:RSS) の認定大学の制度がある。以下では、主として前者について米国の動きを述べる。後者については、付録(3)に示す。

前提言[1]の2014年と現時点を比べたときの米国の動きとして、次の2点が挙げられる。(ア)統計関連(統計学、計量生物学科、データサイエンス学科)の教育、研究は順調に発展しており、学位(学士、修士、博士)授与数、学位授与大学数が順調に増加していること、(イ)統計学とコンピュータサイエンスが融合する形のデータサイエンスの教育、研究の施設やプログラムが色々な大学で立ち上がったこと。以下、この2点について説明する。

(ア) 統計関連の学位に関するデータは、米国統計学会のホームページ [75]が詳しい。2021年の統計学学位授与数は5,291(学士)、4,211(修士)、467(博士)、同じ年の計量生物学位授与数は49(学士)、917(修士)、222(博士)である。経年推移のグラフは[76]に示されている。男女比はほぼ半々である。

また、2021年の統計学学位授与大学数は205(学士)、174(修士)、85(博士)、同じ年の計量生物学位授与大学数は12(学士)、86(修士)、52(博士)であった。また、

データサイエンス関係学位授与大学数は 85 (修士) である[76]。なお、修士課程については、ノースカロライナ州立大学の高等分析研究所によるホームページも詳しい[77]。このように、日本との差は未だ拡大しつつある。

(イ) 深層学習を中心とする機械学習の急速な発展と相まって、米国ではデータサイエンスの教育、研究施設やプログラムが色々な形で立ち上がりつつある。2018 年 3 月 26 日に、コロンビア大学データサイエンス研究所主催で“Data Science Leadership Summit”と題する集会が開催された。その報告[78]には、現在の全米の各大学の概観が書かれており、ここでそれを簡単に紹介する。

この集会では、最初にデータサイエンスとはいかなるものかが議論された。John Tukey が予言していたような統計学の進化形[19]か、新しい分野であるかどうかについては意見の一致を見なかった。一方で、コンピュータサイエンスと統計学の概念と技術がデータサイエンスの中核であるという点ではコンセンサスが得られた。

データサイエンスが既存学問と異なる点は、その手法があらゆる分野、職業、部門に適用することができるという学際性である。このために大学の教育に取り込む場合、大学のどこに位置付けるか大学によりまちまちである。付録(3)では、具体例としてアメリカの 3 つの大学及び中国とイギリスの例を詳しく紹介している。

(2) 初等・中等教育段階における統計の位置付け

日本では「数学的活動の一層の充実」や「統計教育の一層の充実」などを特色とする高等学校学習指導要領が 2018 年 3 月に告示され、2022 年から年次進行により適用されている中で、統計教育については、高等学校だけではなく、小学校及び中学校においても充実させる方向になってきた。しかしながら、欧米及び中国をはじめとするアジア諸国では、既に 1990 年代初頭より政府及び関連学術団体が協力し、科学技術推進と 21 世紀型スキルの養成に向けた教育改革の柱に統計教育を位置付け、初等中等から高等教育基礎に至る教育内容と方法を問題解決的な統計実践のための教育に移行させてきた。さらに、2010 年代に入ると、ビッグデータを視野に入れたデータサイエンス教育のカリキュラムの構築から Information and Communication Technology (ICT) を活用した教材、ソフトウェア等、デジタル環境の整備に大きな予算を投入している。このような国際的な展開と比較すると、統計教育の実践と環境整備における日本の立ち遅れは著しく、さらなる充実が必要である。この点を 6 章で提案 4 とする。

以下では、海外の初等・中等教育段階における統計・データサイエンスの基礎教育の展開の経緯と特徴を紹介する。

1990 年代に入り、数学教育との関連では、イギリスの Mathematical Sciences Education Board[79]、アメリカの National Council of Teachers of Mathematics[80], [81]、OECD による生徒の学習到達度調査(PISA)などで、統計と確率、不確実性の数理の領域は従来にも増して相当重要な位置を与えるべきであると勧告している。さらに、統計教育の方法論に関する改革レポート[82]が公表されて以降、単に統計の知識を教えるだけではなく、統

計を実際の文脈に沿って活用する力、いわゆる統計的思考力の育成により重点を置くことが世界の主流となっている。同時に、科学的な探究能力の育成が重視され、アメリカの科学教育スタンダード[83]では、物理、地学、生命科学、科学技術の社会への応用等の領域に加えて、科学的探究(Scientific Inquiry)が独立の領域として設けられている。

さらに、ビッグデータの時代を迎え、米国では 2011 年に全米共通コアカリキュラムの数学において、高等学校での統計内容が大学初年次相当に匹敵するほど大幅に拡充され、2012 年には、全米学術研究協議会(NRC)が次世代科学教育スタンダード(Next Generation Science Standard)を発表し、ビッグデータの収集から処理、分析、可視化等を意識した科学的な探究学習が組み入れられ、多くの州で取り上げられている^v。2019 年には、オーストラリア、カナダ、イギリス、ドイツ、オランダ、ニュージーランド、米国の統計学会、コンピュータ科学の専門学会とその領域の科学者、及び初等・中等教育の教員、カリキュラム・教育専門家が参画し、International Data Science in Schools Project (IDSSP)が発足、高等学校第2、3学年向けのデータサイエンスカリキュラムを公表している。

以上のように、海外では、統計・データサイエンスの実践の基礎を生活や社会と接続した文脈の中で体験学習を通して学ぶ環境整備が教育のデジタル化と伴走しながら進められてきた。それが教科書にも反映されている。付録(4)で諸外国の教科書と日本の指導要領を比較している。また、かねてより統計教育に注力してきたニュージーランドの取組も参考になる^{vi}。

4 日本におけるこれまでの統計科学教育改善の動き

本章では、統計科学教育改善の動きとして、(1)データサイエンス系学部等の新設や大学院の動き、(2)数理・データサイエンス・AI 教育の動きと認定制度、(3)学習指導要領の改訂、(4)統計教育連携ネットワーク (JINSE) と統計検定の動向の4点について述べる。

(1) データサイエンス系学部等の新設や大学院の動き

近年、データサイエンス領域における人材需要が大きな高まりを見せ、データサイエンスを銘打つ学部学科、また、学べる専門のコースを設置する大学が急速に増えてきている。データサイエンス領域では、①数学・統計学、②情報学・データ処理・分析・解析、③専門分野のバランスが重要であるという認識の下、これらの内容が学べるカリキュラムが準備されている[7]。

データサイエンスが学べる大学として、国公立大学では、北海道大学、東京大学、滋賀大学、広島大学、横浜市立大学などが先行している。私立大学では、武蔵野大学、中央大学、東京理科大学、立正大学、早稲田大学などが挙げられる。学士課程と同様、データサイエンスを学べる大学院も増加傾向にある^{vii}。

このように、データサイエンス領域の学部等を設置している大学、大学院の数は増えているものの、海外と比較すると依然として少ない。総務省の「平成 26 年版 情報通信白書」にある McKinsey 社の調査によると、統計学や機械学習に関する高等訓練の経験を有し、データ分析に係る才能を有する大学卒業生の数は、日本は平成 20 年 (2008 年) 単年で 3,400 人であり、かつ、平成 16 年 (2004 年) から平成 20 年 (2008 年) までの 5 年間、日本におけるデータ分析の才能を有する人材が減少傾向にあったとしている ([84], [85])。付録図 1 から理解できることは、日本は IT 系企業の成長が著しいアメリカや中国、インドとの比較においてかなり低い水準と言える。

さらに、一般社団法人データサイエンティスト協会 (2019 年 11 月) 「データサイエンティストの採用に関するアンケート調査結果」 ([86]) が示した付録図 2 から分かるように、データサイエンティストがいる国内企業は、全体の 29% という実態である。データサイエンティストになるためには、データ解析やその結果をビジネスに応用する方法などを学ぶ必要がある。しかし、アメリカのように専門的な養成コースがある大学が少なく、大学や大学院を卒業してすぐに即戦力として働くのは非常に難しい。

以上により、データ分析を学べる教育機関が日本でいかに少ないかが分かる。

(2) 数理・データサイエンス・AI 教育の動きと認定制度

2017 年度以降、数理・データサイエンス・AI 教育強化拠点コンソーシアム [87]^{viii} が「大学の数理・データサイエンス教育強化方策について」 [88] 及び AI 戦略 2019^{ix} [89] 等の提言を踏まえ、データサイエンス教育の全国の大学への普及・展開に向けた活動を実施している。約 200 校からなる本コンソーシアムは、文理を問わず全国全ての高等教育機関の学生が、数理・データサイエンス・AI を習得できるような教育体制の構築・普及を目指している。本コンソーシアムは、リテラシーレベル及び応用基礎レベルのモデルカリ

キュラムを公表している。これらについては付録(5)にまとめている。

本コンソーシアムが実施した複数の「調査報告書」[90]-[93]から、日本の数理・データサイエンス教育状況を知ることができる。ここでは、2018年、2020年、2022年に行った3回の「数理・データサイエンス・AI教育現状調査」[93],[91],[92]を基に、一般教養（教養教育）段階での数理・データサイエンス・AI導入について述べる。

一般教育（教養教育）段階における「数理」教育の実施状況については、「全学または学部単位で導入」と回答した大学が3回の調査において、70%→79%→82%に増加した。また、一般教育（教養教育）段階における「データサイエンス・AI」教育の実施状況については、53%→65%→81%に増加した。今後の導入予定も含めると、全学導入は国公立大学全体で83%になることが3回目の調査から分かった。また、全学必修科目の導入割合の増加も見られる。これらのことから、数理・データサイエンス・AI教育の必要性について、多くの大学が理解していると考えられる。

前節でも述べたデータサイエンス系学部等の新設など、データサイエンス・AI教育の拡充の機運は高まっているものの、数理・データサイエンス・AI教育を担当できる教員の不足が指摘され、3回目の調査[92]では、「数理教育は約53%、データサイエンス・AI教育は約62%が不足」と回答している。早急に本問題を解決する必要がある、それについては5章で触れる。

文部科学省にて、令和3年度から「認定教育プログラム（リテラシーレベル）」、令和4年度から「認定教育プログラム（応用基礎レベル）」の認定が行われた[94]。なお、本認定制度は、内閣府、文部科学省及び経済産業省が創設したものであり、数理・データサイエンス・AIに関する基礎的な能力及び実践的な能力の向上を図る機会の拡大に資することを目的としている[95],[96]。

「認定教育プログラム（リテラシーレベル）」の認定は、令和3年度の第1回と第2回を合わせて78件、令和4年度は139件であった。一方、令和4年度の「認定教育プログラム（応用基礎レベル）プラス」は68件であった。さらに、認定（前年度認定を含む）された教育プログラムの中から、先導的で独自の工夫・特色を有するものを「認定教育プログラム（リテラシーレベル）プラス」として7件、「認定教育プログラム（応用基礎レベル）プラス」として9件選定した。これにより、本制度における「認定教育プログラム（リテラシーレベル）」は217件、「認定教育プログラム（リテラシーレベル）プラス」は18件となった[97]。

(3) 学習指導要領の改訂

平成10・11年に改訂された小中学校、高等学校の学習指導要領で統計に関する内容が大幅に削除されたという歴史がある。そのため、約10年にわたる間、日本は国際社会の流れに遅れることとなった。平成20・21年に改訂公布された小中学校、高等学校の学習指導要領は「生きる力」を副題とし、思考力・判断力・表現力などの育成を重視したものであった[98]。その流れの一環として、統計に関する指導内容の充実が図られた。この変化は、小中学校、高等学校の統計教育に大いに影響を与え、問題解決力や資料活用

力を養うことの重要性を社会に示した。この学習指導要領において特筆すべきことは、高等学校における数学科の必修科目数学Ⅰに「データの分析」という内容が新たに設けられたことである[99]。

平成 29・30・31 年改訂学習指導要領では「生きる力学びの、その先へ」と副題を変更した[100]。また、「理数教育」において、「観察、実験などによる科学的に探究する学習活動や、データを分析し、課題を解決するための統計教育を充実します」と明記された[101]。これを受け、必修科目の数学Ⅰの「データの分析」に「仮説検定の考え方」が、選択科目の数学Ⅱの「統計的な推測」に「正規分布を用いた区間推定」に加えて「仮説検定」が新たに入る。これらの変移を付録表 3 に示す[102], [103], [104]。

この動きに対して、日本学術会議数理科学委員会数学教育分科会が統計教育を踏まえた提言[105]を発表した。そこでは、「統計教育の実効性を高めること」として「データを活用し、意思決定に繋がる問題解決の方法として、算数・数学科での統計的な方法、考え方を体得させるべきである」としている。このように、高等学校の数学の中で統計教育が充実してきている。

高等学校の情報では、前高等学校学習指導要領（平成 21 年改訂）の「社会と情報」及び「情報の科学」の 2 科目からの選択必修を改め、高校生全員の共通必修科目としての「情報Ⅰ」を設けた。さらに、問題の発見・解決に向けて、情報システムや多様なデータを適切かつ効果的に活用する力やコンテンツを創造する力を育む「情報Ⅰ」の発展的な選択科目としての「情報Ⅱ」を設けた[106]。今回の改訂では、統計的手法を活用し、データの分析と表現に関する学習を充実するなどの改善が行われているため、統計学との関わりが深く影響力を持つ。

例えば、「情報Ⅰ」の「情報通信ネットワークとデータの活用」の内容については、「数学Ⅰ」の「データの分析」の内容と関連付けて扱うこと[107]、「情報Ⅱ」の「情報とデータサイエンス」の内容については、「数学Ⅱ」の「統計的な推測」の内容と関連付けて扱うことなどが提案されている[108]。

文部科学省から、「令和 7 年度大学入学者選抜に係る大学入学共通テスト実施大綱の予告」[108]が出され、大学入試センターは、プログラミングを含む「情報Ⅰ」を令和 7 年度の共通テストから出題すると発表した[109]。既に試作問題も発表されている[110]。本試作問題には、「データに関する基本知識」、「箱ひげ図」や「散布図」の読み取りや「回帰直線」に関する問題が含まれており、統計学との関係性が重要視されている。

大学教育については、日本学術会議の分野別参照基準の策定が一定の役割を果たした⁵。

(4) 統計教育連携ネットワークと統計検定

2012 年から 2016 年まで実施した文部科学省の補助事業により「統計教育大学間連携ネットワーク」(Japanese Inter-university Network for Statistical Education: JINSE) が設立された。JINSE では、統計科学教育の参照基準を拡充し、それに則った大学における統計科学教育の標準的カリキュラム体系の策定、その体系に基づく標準的コンテンツの作成、達成度評価の整備を目標とする産業界等による外部評価を含めた PDCA サイクルを意

識した活動を行った。本事業の終了後の2017年度からは、JINSEが開発・蓄積してきた資産を教育関係者に提供するために、参加対象機関を大学のみならず、全国の教育機関等に拡大し、「統計教育連携ネットワーク」の運用を開始した^{xi}。

統計教育連携ネットワークは、一般社団法人日本統計学会が設立した「一般財団法人統計質保証推進協会」が運営している[111]。本協会では、統計教育の改善の動きを背景として、日本統計学会が中心となり、教育の成果を評価する仕組みとして2011年に発足した「統計検定」の運営も行っている[112]。

現在の統計検定は、統計分野の質保証として、初等・中等教育の学習指導要領に準拠する形の3級及び4級、大学基礎課程で習得すべき大学学士力を質的保証する手段としての2級、基礎的講義に引き続いて学ぶ応用的な統計学の諸手法の習得の確認のための準1級、大学院水準に相当する1級がある。また、公的あるいは民間を問わず、各調査機関で必要とされる内容を問う、統計調査士及び専門統計調査士がある。

また、近年のデータサイエンスを重視した社会の動きに合わせ、統計検定データサイエンス基礎 (DS 基礎)、データサイエンス発展 (DS 発展)、データサイエンスエキスパートが開始された[113]。

2016年8月より、コンピュータ上で実施するCBT (Computer Based Testing) 方式を2級及び3級に導入した。2019年4月には、統計調査士についてもCBT方式を実施した。2020年のCOVID-19の感染拡大のため、2021年の紙媒体による実施を最後に、1級を除く全ての種別をCBT方式のみの実施とした。この間の受験者数の推移は付録図3に示す通りである^{xii}。このように、毎年順調に受験者を伸ばし、2022年度では25,000名以上の申込者があった。統計検定は、教育機関で利用されるだけでなく、企業においてもデータサイエンスの業務に就く際の指標とされている^{xiii}。社会人のリスクリングの観点からも統計検定の果たしている役割は大きい。

5 日本における統計科学の教育・研究体制の現状と課題

4章では、近年の統計科学教育に関する改善点や動向について述べたが、未だ解決していない課題がある。ここでは、この点をデジタル社会が求める専門人材、大学及び大学院における統計科学教育と統計エキスパート人材育成、統計科学分野の研究体制の現状に分けてまとめる。

(1) デジタル社会が求める専門人材

文部科学省行政情報化推進委員会が2022年10月に最終決定した『デジタル社会の実現に向けた重点計画』に基づき「文部科学省における中長期的な計画」[114]が公表された。課題として、「教育データの効果的な活用の推進について」が言及され、計画目標においてもこの点を重視している。

さらに、データ社会の実現に向けた主な取組事項として、1) 規制の一括見直しへの対応、2) デジタル庁が整備する共通機能の活用、3) 人材育成が挙げられ、人材育成として「文部科学省サイバーセキュリティ・IT人材確保・育成計画」に基づき、IT人材の確保・育成を行うとしている。

より広く社会全体で求める専門人材については、「デジタル庁における中長期的な計画」[115]に記載されている。特に、「デジタル人材の育成・確保」の項目[116]では「全ての国民が、それぞれのライフステージに応じて必要となるICTスキルを習得する環境を整備すると共に、社会のそれぞれの立場で求められる人材の確保・育成を図ることにより、目指すべきデジタル社会の着実な実現を図る」というスローガンが示されている。また、「デジタル人材の育成・確保に向けて」[117]では、現状と課題として、「デジタル人材が質・量共に充実しているとは言いがたく、人材全体の底上げや裾野の広がり、専門人材の育成・確保、都市圏への偏在解消等を同時に進めることが求められる」とし、対応の方針として「デジタル推進人材の育成」、(2022年度からの5年間で230万人)が取り上げられている。この230万人計画には、付録表4にあるように、文部科学省との連携により、数理・データサイエンス・AI教育の推進(2025年に応用基礎25万人/年)、リカレント教育の推進(受講者1,000人/年)が含まれている。

ここで示されている数理・データサイエンス・AI教育の推進には、4章(2)で紹介した「数理・データサイエンス・AI教育強化拠点コンソーシアム」や次節で示す「統計エキスパート人材育成」が大きく関係している。

2章(3)や本節で見てきたように、社会人のリスキリングは重要な課題であり、6章の提案2とする。

(2) 大学院における統計科学教育と統計エキスパート人材育成

Society 5.0の実現を目指す社会において、研究のDXやAIに必要なスキルを有する「統計エキスパート」の最上位に位置する「大学統計教員」を育成することが重視されている。日本では、最近のデータサイエンス系の学部開設まで、統計学部・学科が存在せず、統計教員の育成体制が極めて脆弱であった。この問題を解決するため、統計学を用いた

他分野との融合領域の研究振興を図ると共に、育成されたエキスパート人材が大学等で核となり、統計学や融合領域に係る教育や活動の普及・展開を行う体制（統計エキスパート人材育成エコシステム）を構築することを目的とした文部科学省の公募事業「令和3年度統計エキスパート人材育成プロジェクト」が2021年7月より開始された[118]。

本プロジェクト推進のために、統計数理研究所を「中核機関」とする「統計エキスパート人材育成コンソーシアム」が設立された。2022年1月に統計数理研究所は大学統計教員育成センターを設置した[119]。中核機関のほかに、協働して事業を実施する全国の大学等を「参画機関」、円滑な事業の遂行に向けて協力する機関を「協力機関」として構成され、基盤組織が構築された[120]。2022年11月現在、参画機関は25機関、協力機関は4機関となっている。本研修は、1期から3期の研修期間が設けられ、各期間は原則2年間である。

参画機関では、研修を修了した大学教員が中核となって機関内に「統計エキスパート育成システム」を構築し、機関所属の大学院生を「統計エキスパート」に育成する。これにより、5年間の事業期間中に少なくとも30名の大学統計教員を育成することを目指し、これらの育成された大学統計教員が所属する各機関においてさらに人材を育成することにより、5年間の事業期間も含め10年間で約500名の統計エキスパートを育成することを目指している。

500名程度では、必要とされている統計エキスパートには到底足りない状況であり、本プロジェクト終了後も、多くの大学機関や研究機関が連携した同様の事業が必要である。特に、統計エキスパート人材育成事業は有効であり、この事業を大幅に拡大すべきである。

統計教員の不足については4章(2)でも触れたが、喫緊の課題であり、6章の提案3とする。

(3) 統計科学分野の研究体制

4章(1)で言及したように、最近になってデータサイエンスに関する学部や大学院が設置され、さらに設置等が検討されている状況であるが、それまで日本の大学には統計学部・学科が存在しなかったために、統計科学の研究者は依然として様々な大学の学部に分散している現状であり、日本の統計科学の中心的な研究機関は統計数理研究所である。統計数理研究所は、1944年設立という長い伝統を持つが、1985年に大学共同利用機関として改組され、大学共同利用機関法人情報・システム研究機構統計数理研究所として共同利用を推進している[121]。公募型の共同利用の件数は、当初の62件から121件（2022年10月現在）に拡大し、その裾野は着実に広がっている。また特別共同利用研究員の受入れ制度を持ち、全国の国公立大学の大学院学生を受け入れて指導を行っている[122]。

このように、統計数理研究所が統計科学研究の中心であるが、最近のビッグデータの時代を迎え、十分な体制とは言えない。諸科学の共通言語としての統計科学研究の発展を進めるには、様々な分野の研究所に統計科学の専門家を配置し、分野融合的な研究を推進することが重要である。また、開設が相次いでいるデータサイエンス系の学部や大

学院においても、理論的基礎としての統計科学の教育を充実すると共に、統計科学の手法を用いた応用研究を進める必要がある。この点については、本分科会の中長期戦略[21]においてもデータサイエンスの学際的な研究・教育拠点の形成を提案した。

統計関連の学会及び科学研究費の状況については付録(6)で論じる。

6 統計科学教育・研究推進への提案

これまでの考察に基づき、本見解では我が国における統計科学教育・研究の今後の推進に向けて、以下の4項目の提案を行う。

提案1 数理・データサイエンス・AI分野の理論的基礎としての統計科学の位置付け

2章で述べたように、ビッグデータの処理において統計科学の方法論は様々な手法の理論的基礎を与えるものであり、数理・データサイエンス・AI分野の教育においても、データの正しい解釈のためには、様々な手法をツールとして教育するだけでなく、手法の理論的基礎も含めた教育が重要である。そのために、各大学等は統計関連学会連合による「統計学分野の教育課程編成上の参照基準」[123]を1つの基準として用いることができる。

提案2 数理・データサイエンス・AI分野の再教育(リスキリング)の推進

3章及び4章の現状分析を基に、5章で述べたように、社会人に対する数理・データサイエンス・AI分野のリスキリングの推進が我が国にとって急務となっており、大学の果たす役割も大きい。

提案3 学士課程及び大学院教育が必要とする統計教員の育成

同じく3章及び4章の現状分析を基に、5章で述べたように、学士課程及び大学院教育が必要とする統計教員が極めて不足しており、教員の育成が喫緊の課題であり、統計エキスパート人材育成事業を大幅に拡大すべきである。

提案4 初等・中等教育における教材、ソフトウェア、デジタル環境の整備と統計教育のさらなる充実

3章(2)で述べたように、数理・データサイエンス・AI分野の大学教育を広く行う前提として、初等・中等教育における教材、ソフトウェア、デジタル環境を整備し、統計教育のさらなる充実を図る必要がある。

<参考文献>

- [1] 日本学術会議数理科学委員会数理統計学分科会、提言「ビッグデータ時代における統計科学教育・研究の推進について」、2014年8月20日。
<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t197-1.pdf>
(2023年3月3日参照)
- [2] 竹内啓、『数理統計学』、東洋経済新報社、1963年。
- [3] 久保川達也・国友直人、『統計学』、東京大学出版会、2016年。
- [4] 松田芳郎、『データの理論』、岩波書店、1978年。
- [5] 工藤弘安、『入門 統計学—官庁統計の作成と利用—』、全国統計協会連合
1997年。
- [6] 竹村彰通、『データサイエンス入門』、岩波書店、2018年。
- [7] D. Conway, *The Data Science Venn Diagram*, 2013.
<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
(2023年3月3日参照)。
- [8] 竹内啓、『社会科学における数と量』、東京大学出版会、1971年。
- [9] Z. Griliches, *Economic Data Issues*, *Handbook of Econometrics*, Z. Griliches and M.
D. Intriligator, eds, 1st ed, Elsevier, 1986, pp. 1465-1514.
- [10] K. Imai, *Quantitative Social Science: An Introduction*. Princeton University
Press, 2017.
- [11] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's*
Companion. Princeton University Press, 2009.
- [12] G. W. Imbens and D. B. Rubin, *Causal Inference for Statistics, Social and*
Biomedical Sciences: An Introduction. Cambridge University Press, 2015.
- [13] 星野崇宏、『因果推論・選択バイアス・データ融合』、岩波書店、2009年。
- [14] D. A. Freedman, *Statistical Models and Causal Inference: A Dialogue with the*
Social Sciences. Cambridge University Press, 2010.
- [15] J. W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.
- [16] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, *Knowledge Discovery and Data*
Mining: Towards a Unifying Framework, *KDD-96 Proceedings*, 1996.
<https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf>
(2023年3月3日参照)
- [17] B. Efron and T. Hastie, *Computer Age Statistical Inference: Algorithms, Evidence,*
and Data Science. Cambridge University Press, 2016.
- [18] C. Wu and M. E. Thompson, *Sampling Theory and Practice*. Springer Nature, 2020.
- [19] D. Donoho, 50 Years of Data Science, *Journal of Computational and Graphical*
Statistics, 26(4), 745-766, 2017.
doi: 10.1080/10618600.2017.1384734.
- [20] 統計委員会、「公的統計の整備に関する基本的な計画（案）」、2023年2月。

- https://www.soumu.go.jp/main_content/000862730.pdf
(2023年3月3日参照)
- [21] 日本学術会議数理科学委員会数理統計学分会、『異分野・社会との連携のための共通言語「データサイエンス」の学際的な研究・教育拠点の形成』、2022年3月.
- [22] 内閣府、「経済財政運営と改革の基本方針2020 について」、2020年7月17日.
https://www5.cao.go.jp/keizai-shimon/kaigi/cabinet/honebuto/2020/2020_basicpolicies_ja.pdf
(2023年7月2日参照)
- [23] デジタル庁データ戦略推進ワーキンググループ（第4回）、「データ戦略の推進状況」、2022年9月6日.
https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/b565c818-75f4-4990-9125-dd43af8362ba/afe23c36/20220906_meeting_data_strategy_outline_01.pdf
(2023年7月2日参照)
- [24] 文部科学省、『令和3年度 文部科学白書』、2021年.
https://www.mext.go.jp/b_menu/hakusho/html/hpab202001/1420041_00010.htm
(2023年7月2日参照)
- [25] 文部科学省高等教育局、「デジタル人材育成等に資する取組（政府予算関連）」、2022年12月26日.
https://www.meti.go.jp/shingikai/mono_info_service/digital_suishin/pdf/002_01_00.pdf
(2023年7月2日参照)
- [26] 文部科学省高等教育局、「高等教育段階におけるデジタル人材育成の取組について」、2022年9月29日.
https://www.meti.go.jp/shingikai/mono_info_service/digital_suishin/pdf/001_02_01.pdf
(2023年7月2日参照)
- [27] 統合イノベーション戦略推進会議「AI戦略 2022」2022年4月22日.
https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf
(2023年3月6日参照)
- [28] 林高樹・佐藤彰洋、『金融市場の高頻度データ分析—データ処理・モデリング・実証分析—』、朝倉書店、2016年.
- [29] 青嶋誠・矢田和善、『高次元の統計学』、共立出版、2019年.
- [30] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*. The MIT Press, 2016.
- [31] 岡谷貴之、「深層学習」第2版講談社、2022年.
- [32] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*, 1st ed., Springer Nature, 2018.
- [33] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed.,

- Pearson, 2020.
- [34] O. Calin, *Deep Learning Architectures: A Mathematical Approach*, 1st ed., Springer Nature, 2020.
- [35] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, *Proceedings of the 2nd International Symposium on Information Theory*, B. N. Petrov and F. Csaki, eds., Budapest: Akademiai Kiado, 1973, pp. 267-281.
- [36] 赤池弘次・甘利俊一・北川源四郎・樺島祥介・下平英寿、「赤池情報量規準AIC—モデリング・予測・知識発見」共立出版、2007年.
- [37] 今泉允聡、日本統計学会誌第50巻2号p. 257-283「深層学習の原理解析：汎化誤差の側面から」、2021年.
doi: <https://doi.org/10.11329/jjssj.50.257>
- [38] 鈴木大慈、日本統計学会誌第50巻2号p. 229-256「深層学習の統計理論」、2021年.
doi: <https://doi.org/10.11329/jjssj.50.229>
- [39] 園田翔、日本統計学会誌第50巻2号p. 285-316「ニューラルネットの関数解析的方法と無限次元零空間」、2021年.
doi: <https://doi.org/10.11329/jjssj.50.285>
- [40] 杉山将、「機械学習分野における研究・人材育成の動向と課題」（日本学術会議、公開シンポジウム「数理・データサイエンス・AI時代における統計科学の教育及び研究について」報告資料）、2023年2月17日.
- [41] J. H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5) 2001.
doi: 10.1214/aos/1013203451.
- [42] D. W. Apley and J. Zhu, Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models, December 2016.
arXiv:1612.08468
- [43] G. Shafer and V. Vovk, A Tutorial on Conformal Prediction, *Journal of Machine Learning Research*, 9, pp. 371-421, 2008.
<http://www.jmlr.org/papers/v9/shafer08a.html>
- [44] M. Imaizumi and K. Fukumizu, Advantage of Deep Neural Networks for Estimating Functions with Singularity on Hypersurfaces, 2020.
<https://doi.org/10.48550/arXiv.2011.02256>
- [45] Y. Romano, M. Sesia, and E. Candès, Deep Knockoffs, *Journal of the American Statistical Association*, 115(532), pp. 1861-1872, 2020.
doi: 10.1080/01621459.2019.1660174.
- [46] 福水健次、「深層生成モデルによる統計的推論」（大学共同利用機関法人情報・システム研究機構統計数理研究所、「統計数理研究所オープンハウス”統計数理研究所は創立75周年を迎えます”」報告資料）、2019年6月5日.
<https://www.ism.ac.jp/openhouse/2019/>

- (2023年3月12日参照).
- [47] U. S. Bureau of Labor Statistics, Projections Overview and Highlights, 2021-31, November 2022. <https://www.bls.gov/opub/mlr/2022/article/projections-overview-and-highlights-2021-31.htm>
(2023年3月3日参照).
- [48] U. S. Bureau of Labor Statistics, Occupational Employment and Wages: 15-2041 Statisticians, May, 2021. [https://www.bls.gov/oes/current/oes152041.htm#\(1\)](https://www.bls.gov/oes/current/oes152041.htm#(1))
(2023年3月3日参照).
- [49] U. S. Bureau of Labor Statistics, Occupational Employment and Wages: 15-2051 Data Scientists, May 2021. <https://www.bls.gov/oes/current/oes152051.htm>
(2023年3月3日参照).
- [50] 内閣官房教育未来創造会議、提言「我が国の未来をけん引する大学等と社会の在り方について（第一次提言）」、2022年5月10日.
<https://www.cas.go.jp/jp/seisaku/kyouikumirai/teigen.html>
(2023年3月3日参照)
- [51] 一般社団法人日本経済団体連合会、提言「新しい時代に対応した大学教育改革の推進－主体的な学修を通じた多様な人材の育成に向けて－」2022年1月18日.
<https://www.keidanren.or.jp/policy/2022/003.html>
(2023年3月3日参照)
- [52] 一般社団法人日本経済団体連合会、提言「Digital Transformation (DX)～価値の協創で未来をひらく～」、2020年5月19日.
<https://www.keidanren.or.jp/policy/2020/038.html>
(2023年3月3日参照)
- [53] 経済産業省デジタル産業の創出に向けた研究会、報告書「DXレポート2.1 (DXレポート2 追補版)」、2021年8月31日.
<https://www.meti.go.jp/press/2021/08/20210831005/20210831005.html>
(2023年3月3日参照)
- [54] 一般財団法人日本科学技術連盟、第111回品質管理シンポジウム「データ駆動型社会における顧客価値創造と組織能力（品質保証）向上」、2021年6月3日.
https://www.juse.jp/qcs/archive/111qcs_pamphlet.pdf
(2023年3月3日参照)
- [55] 一般財団法人日本規格協会グループ、「品質管理検定(QC検定)とは」.
https://webdesk.jsa.or.jp/common/W10K0500/index/qc/qc_qc1/
(2023年3月3日参照)
- [56] 一般財団法人日本規格協会グループ品質管理検定センター、「受検者データ」、2021年.
https://webdesk.jsa.or.jp/pdf/qc/md_4742.pdf
(2023年3月3日参照)

- [57] 一般財団法人日本科学技術連盟、「QCサークル活動（小集団改善活動）」。
<https://www.juse.or.jp/business/qc/index.html#>
 (2023年3月3日参照)
- [58] 一般財団法人日本科学技術連盟QCサークル本部、「集計表ダウンロード」。
https://juse-qcch.juse.or.jp/Qcc_Honbu/Qccircle/Spreadsheet
 (2023年3月3日参照)
- [59] 石山一雄、「国際及び国内産業界における統計的手法を活用した問題解決手法・シックスシグマ(Six Sigma)の現状」(日本学術会議 公開シンポジウム「数理・データサイエンス・AI時代における統計科学の教育及び研究について」報告資料)、2023年2月17日。
- [60] 一般社団法人日本品質管理学会、「学会の活動」。
<https://jsqc.org/activity/>
 (2023年3月3日参照)
- [61] 一般社団法人日本品質管理学会、「研究実績の一覧」。
<https://jsqc.org/researchachievements/>
 (2023年3月3日参照)
- [62] H. Markowitz, Portfolio Selection, *Journal of Finance*, 7(1), pp. 77-91, 1952.
 doi: 10.1111/j.1540-6261.1952.tb01525.x.
- [63] D. McFadden, Conditional Logit Analysis of Qualitative Choice Behavior, *Frontiers in Econometrics*, in: P. Zarembka, ed., Academic Press, 1973, pp. 105-142.
- [64] F. Black and M. Scholes, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81(3), pp. 637-654, 1973.
https://www.cs.princeton.edu/courses/archive/fall109/cos323/papers/black_scholes73.pdf
- [65] R. C. Merton, On The Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance*, 29(2), pp. 449-470, 1974.
 doi: 10.1111/j.1540-6261.1974.tb03058.x.
- [66] バーゼル銀行監督委員会、「バーゼルI」、1988年。
- [67] バーゼル銀行監督委員会、「バーゼルII」、2004年。
- [68] 一般社団法人日本経済団体連合会、提言「今後の採用と大学教育に関する提案」、2018年12月4日。
<https://www.keidanren.or.jp/policy/2018/113.html>
 (2023年3月3日参照)
- [69] International Institute of Statistics, ISI Declaration on Professional Ethics, October 20, 2010.
<https://isi-web.org/isi-declaration-professional-ethics-0>
 (2023年8月14日参照)
- [70] World Medical Association, WMA Declaration of Helsinki – Ethical Principles for

- Medical Research Involving Human Subjects, 2013.
<https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
(2023年3月3日参照)
- [71] R. L. Wasserstein and N. A. Lazar, The ASA Statement on p-Values: Context, Process, and Purpose, *The American Statistician*, 70(2), pp. 129-133, 2016.
doi: 10.1080/00031305.2016.1154108.
- [72] 日本学術会議総合工学委員会・機械工学委員会合同計算科学シミュレーションと工学設計分科会、提言「科学的知見の創出に資する可視化に向けて」、2017年8月8日。
<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t247-5.pdf>
(2023年3月3日参照)
- [73] P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyanakrishnan, E. Kamar, S. Kraus, K. Leyton-Brown, D. Parkes, W. Press, A-L. Saxenian, J. Shah, M. Tambe, and A. Teller, Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel, Stanford University, Stanford, CA, September 2016.
<https://ai100.stanford.edu/2016-report>
(2023年3月3日参照)
- [74] C. A. Franklin, A. E. Bargagliotti, C. A. Case, G. D. Kader, R. L. Scheaffer, and D. A. Spangler, Statistical Education of Teachers.
<https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf>
(2023年3月3日参照)
- [75] The American Statistical Association, Statistics and Biostatistics Degree Data.
<https://www.amstat.org/education/statistics-and-biostatistics-degree-data>
(2023年3月3日参照)
- [76] The American Statistical Association, Statistics, Biostatistics Degree Growth Continued in 2021, *Amstat News*, December 1, 2021.
<https://magazine.amstat.org/blog/2022/12/01/statsbiostatsdegree/>
(2023年3月3日参照)
- [77] N. C. S. U. Institute of Advanced Analytics, Graduate Degree Programs in Analytics and Data Science, North Carolina State University.
https://analytics.ncsu.edu/?page_id=4184
(2023年3月3日参照).
- [78] J. M. Wing, V. P. Janeja, T. Kloefkorn, and L. C. Erickson, Data Science Leadership Summit Summary Report, September 2018.
<https://dl.acm.org/citation.cfm?id=3293458>

(2023年3月3日参照)

- [79] N. R. Council, *Reshaping School Mathematics: A Philosophy and Framework for Curriculum* Washington, DC: The National Academies Press, 1990.
doi: 10.17226/1498.
- [80] Commission on Standards for School Mathematics and National Council of Teachers of Mathematics. Commission on Standards for School Mathematics, *Curriculum and Evaluation Standards for School Mathematics*. 1989.
- [81] National Council of Teachers of Mathematics, *Principles and Standards for School Mathematics*. 2000.
- [82] G. Cobb, Teaching Statistics, in: L. A. Steen, ed., *Heeding the Call for Change: Suggestions for Curricular Actions*, Washington: The Mathematical Association of America, 1992, pp. 3-34.
<http://www.statlit.org/pdf/1992-Steen-MAA-Heeding-Call-For-Change.pdf>
(2023年3月3日参照)
- [83] N. R. C. Mathematical Sciences Education Board, *Reshaping School Mathematics: A Philosophy and Framework for Curriculum: Perspectives on School Mathematics*. Natl Academy Press, 1990.
<https://nap.nationalacademies.org/catalog/1498/reshaping-school-mathematics-a-philosophy-and-framework-for-curriculum>
(2023年3月3日参照)
- [84] 総務省、平成26年度版情報通信白書第1部第4節「本格的なデータ活用社会の到来」、2014年。
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/pdf/n3400000.pdf>
(2023年8月14日参照)
- [85] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, Big data: The next frontier for innovation, competition, and productivity, May 2011.
<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
(2023年3月3日参照)
- [86] 一般社団法人データサイエンティスト協会、「データサイエンティストの採用に関するアンケート調査結果」、2019年11月。
https://www.datascientist.or.jp/common/docs/c-research_2019.pdf
(2023年3月3日参照)
- [87] 数理・データサイエンス・AI教育強化拠点コンソーシアム公式ウェブサイト。
<http://www.mi.u-tokyo.ac.jp/consortium/>
(2023年3月3日参照)
- [88] 数理及びデータサイエンス教育の強化に関する懇談会、「大学の数理・データサイエ

- ンス教育強化方策について」、2016年12月。
https://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/_icsFiles/afieldfile/2016/12/21/1380788_01.pdf
(2023年3月3日参照)
- [89] 統合イノベーション戦略推進会議、「A I 戦略 2019 ～人・産業・地域・政府全てにA I～」、2019年6月。
<https://www8.cao.go.jp/cstp/ai/aistratagy2019.pdf>
(2023年3月3日参照)
- [90] 数理・データサイエンス教育強化拠点コンソーシアム、「数理・データサイエンス教育状況調査報告書」、2019年1月31日。
<http://www.mi.u-tokyo.ac.jp/consortium/pdf/report02.pdf>
(2023年3月3日参照)
- [91] 数理・データサイエンス教育強化拠点コンソーシアム、「数理・データサイエンス・AI 教育 現状調査 (第2回) の概要」。
<http://www.mi.u-tokyo.ac.jp/consortium/pdf/report03.pdf>
(2023年3月3日参照)
- [92] 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教育強化拠点コンソーシアムニュースレター第17号p. 6 - 9 「第3回数理・データサイエンス・AI 教育の現状調査結果 (速報)」、2023年2月。
<http://www.mi.u-tokyo.ac.jp/consortium/pdf/newsLetter17.pdf>
(2023年3月3日参照)
- [93] 数理・データサイエンス教育強化拠点コンソーシアム、「数理・データサイエンス教育状況調査の概要」、2019年1月31日。
<http://www.mi.u-tokyo.ac.jp/consortium/pdf/report01.pdf>
(2023年3月3日参照)
- [94] 文部科学省、「数理・データサイエンス・AI教育」。
https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm
(2023年3月3日参照)
- [95] 文部科学省、「数理・データサイエンス・A I 教育プログラム認定制度 (リテラシーレベル)」。https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm
(2023年3月3日参照)
- [96] 文部科学省、「数理・データサイエンス・A I 教育プログラム認定制度 (応用基礎レベル)」。https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002_00003.htm
(2023年3月3日参照)
- [97] 文部科学省、『「数理・データサイエンス・AI教育プログラム (リテラシーレベル・応用基礎レベル)」の認定・選定結果について』、2022年8月24日。
https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/1413155_00011.htm
(2023年3月3日参照)

- [98] 文部科学省、『学習指導要領「生きる力」』。
https://www.mext.go.jp/a_menu/shotou/new-cs/idea/1304378.htm
(2023年3月3日参照)
- [99] 文部科学省、「高等学校学習指導要領解説 数学編」、2008年。
- [100] 文部科学省、「平成29・30・31年改訂学習指導要領（本文、解説）」。
https://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm
(2023年3月3日参照)
- [101] 文部科学省、「生きる力 学びの、その先へ 新しい学習指導要領」。
https://www.mext.go.jp/content/1413516_001_1_100002629.pdf
(2023年3月3日参照)
- [102] 文部科学省、高等学校学習指導要領（平成11年3月）「第2章 普通教育に関する各教科 第4節 数学」。
https://www.mext.go.jp/a_menu/shotou/cs/1320155.htm
(2023年3月3日参照)
- [103] 文部科学省、「高等学校学習指導要領解説 数学編」、2018年。
- [104] 文部科学省、「高等学校学習指導要領解説 数学編 理数編」、2019年。
- [105] 日本学術会議数理科学委員会数学教育分科会、提言「新学習指導要領下での算数・数学教育の円滑な 実施に向けた緊急提言：統計教育の実効性の向上に焦点を当てて」、2020年8月4日。
<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t293-2.pdf>
(2023年3月3日参照)
- [106] 文部科学省、『高等学校情報科「情報Ⅰ」教員研修用教材（本編）』。
https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/1416756.htm
(2023年3月3日参照)
- [107] 文部科学省、『高等学校情報科「情報Ⅱ」教員研修用教材（本編）』。
https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_00742.html
(2023年3月3日参照)
- [108] 文部科学省、「令和7年度大学入学者選抜に係る大学入学共通テスト実施大綱の予告」、2021年7月30日。
https://www.mext.go.jp/content/20210729-mxt_daigakuc02-000005144_2.pdf
(2023年3月3日参照)
- [109] 独立行政法人大学入試センター、「令和7年度大学入学共通テストの問題作成の方向性及び試作問題等について」、2022年11月9日。
https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7ikou/
(2023年3月3日参照)
- [110] 独立行政法人大学入試センター、「令和7年度試験の問題作成の方向性、試作問題等」、2022年11月9日。
https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7ikou/r7mondai.html

- (2023年3月3日参照)
- [111] 一般財団法人統計質保証推進協会公式ウェブサイト.
<https://qajss.org/>
(2023年3月5日参照)
- [112] 一般財団法人統計質保証推進協会、「統計検定公式ウェブサイト」.
<https://www.toukei-kentei.jp/>
(2023年3月5日参照)
- [113] 一般財団法人統計質保証推進協会、統計検定「検定種別」.
<https://www.toukei-kentei.jp/exam/>
(2023年3月5日参照)
- [114] 文部科学省、『「デジタル社会の実現に向けた重点計画」に基づく「文部科学省における中長期的な計画」』、2022年8月31日.
https://www.mext.go.jp/a_menu/mext_02004.html
(2023年3月5日参照)
- [115] デジタル庁、「デジタル社会の実現に向けた重点計画」、2023年6月9日.
<https://www.digital.go.jp/policies/priority-policy-program/>
(2023年3月5日参照)
- [116] デジタル庁、「デジタル人材の育成・確保」.
https://www.digital.go.jp/policies/digital_human_resources/
(2023年3月5日参照)
- [117] デジタル田園都市国家構想担当大臣 若宮健嗣、「デジタル人材の育成・確保に向けて」、2022年2月4日.
https://www.cas.go.jp/jp/seisaku/digital_denen/dai3/siryou7.pdf
(2023年3月5日参照)
- [118] 文部科学省、「令和3年度統計エキスパート人材育成プロジェクトの公募について」、2021年4月2日.
https://www.mext.go.jp/b_menu/boshu/detail/1421775_00006.htm
(2023年3月5日参照)
- [119] 統計エキスパート人材育成コンソーシアム、「育成センターの活動」.
<https://stat-expert.ism.ac.jp/core/activity/>
(2023年3月5日参照)
- [120] 統計数理研究所 統計エキスパート人材育成コンソーシアム「統計エキスパートがデータサイエンスの未来を切り開く」.
<https://stat-expert.ism.ac.jp/>
(2023年3月5日参照)
- [121] 大学共同利用法人情報・システム研究機構統計数理研究所、「共同利用案内」.
https://www.ism.ac.jp/kyodo/index_j.html
(2023年3月5日参照)

- [122] 大学共同利用法人情報・システム研究機構統計数理研究所、2023年度特別共同利用研究院の受け入れ「特別共同利用研究員の受入れ制度について」.
<https://www.ism.ac.jp/kyodo/tokubetsu-kyodo/2023/index.html>
 (2023年3月5日参照)
- [123] 統計関連学会連合 理事会・統計関連学会連合 統計教育推進委員会と統計教育大学間連携ネットワーク 質保証委員会、「統計学の各分野における教育課程編成上の参照基準」、2014年8月1日.
<http://jfssa.jp/ReferenceStandard2.pdf>
 (2023年3月5日参照)
- [124] 数理・データサイエンス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～」、2020年4月.
http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf
 (2023年3月5日参照)
- [125] 東京大学・JST CREST、「次世代地震計測と最先端ベイズ統計学との融合によるインテリジェント地震波動解析」.
<https://www.eri.u-tokyo.ac.jp/project/iSeisBayes/>
 (2023年3月5日参照)
- [126] 文部科学省、「情報科学を活用した地震調査研究プロジェクト (STAR-Eプロジェクト)」.
https://www.mext.go.jp/a_menu/kaihatu/jishin/projects/
 (2023年3月5日参照)
- [127] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Society, Series B*, 58(1), pp. 267-288, 1996.
 doi: 10.1111/j.2517-6161.1996.tb02080.x.
- [128] 尾形良彦、「統計数理」第67巻2号p. 215-228、「統計地震学の発展と地震活動予測：個人的経験と展望」、2019年.
<https://www.ism.ac.jp/editsec/toukei/pdf/67-2-215.pdf>
- [129] 上野玄太、「統計数理」第70巻2号p. 129-267、「特集データ同化の方法」、2022年.
<https://www.ism.ac.jp/editsec/toukei/tokeisuri-70j.html#N02>
- [130] JST CREST、吉田直紀「広域撮像探査観測のビッグデータ分析による統計計算宇宙物理学」.
<https://www.jst.go.jp/kisoken/crest/project/44/14532369.html>
 (2023年3月5日参照)
- [131] 吉田直紀 (研究代表者)、『戦略的創造研究推進事業 CREST 研究領域「科学的発見・社会的課題解決に向けた 各分野のビッグデータ利活用推進のための次世代 アプリケーション技術の創出・高度化」 研究課題「広域撮像探査観測のビッグデータ分析による統計計算宇宙物理学」 研究終了報告書』、2020年3月.
[31](https://www.jst.go.jp/kisoken/crest/evaluation/s-houkoku/sh-</p>
</div>
<div data-bbox=)

- r01/JST_1111080_14532369_2019_PER.pdf
(2023年3月5日参照)
- [132] EHT-Japan、国際プロジェクト「イベント・ホライズン・テレスコープ (EHT)」、2019年4月1日。
<https://www.miz.nao.ac.jp/eht-j/c/pr/pr20190410>
(2023年3月5日参照)
- [133] EHT-Japan、国際プロジェクト「イベント・ホライズン・テレスコープ (EHT)」、2019年4月10日。
<https://www.miz.nao.ac.jp/eht-j/c/pr/pr20190410/c04>
(2023年3月5日参照)
- [134] 国立天文台、「(プレスキット：トップ) 史上初、ブラックホールの撮影に成功」、2019年4月10日。
<https://www.nao.ac.jp/news/sp/20190410-eh/>
(2023年3月5日参照)
- [135] 国立天文台、「イベント・ホライズン・テレスコープ ファクトシート」、2019年4月。
<https://www.nao.ac.jp/news/sp/20190410-eh/factsheet.pdf>
(2023年3月5日参照)
- [136] 成田憲保 (東京大学先進科学研究機構教授) ウェブサイト。
<https://gendai.media/list/author/norionarita>
(2023年3月10日参照)
- [137] CREST・さきがけ、研究領域[情報計測]「情報計測オンラインセミナーシリーズー数理・情報科学×計測科学の高度融合による新展開ー」。
<https://measurement-informatics-seminars.jp/>
(2023年3月5日参照)
- [138] Association for Computing Machinery, ACM-IMS Interdisciplinary Summit on the Foundations of Data Science, June 15, 2019.
<https://www.acm.org/data-science-summit>
(2023年3月5日参照)
- [139] Association for Computing Machinery, US, ACM-IMS Interdisciplinary Summit on the Foundations of Data Science, (June 15, 2019).
<https://www.youtube.com/watch?v=HRCqkbt91xU&t=341s>
(2023年3月5日参照)
- [140] Dietrich College of Humanities and Social Sciences, Statistics & Data Science, *Carnegie Mellon University*.
<https://www.cmu.edu/dietrich/statistics-datascience/index.html>
(2023年3月5日参照)
- [141] The Grainger College of Engineering, Computer Science, University of Illinois Urbana-Champaign, B.S. in Statistics & Computer Science.

- <https://cs.illinois.edu/academics/undergraduate/degree-program-options/bs-statistics-computer-science>
(2023年3月5日参照)
- [142] MIT Institute for Data, Systems, and Society.
<https://idss.mit.edu/>
(2023年3月5日参照)
- [143] The Chinese University of Hong Kong.
<https://www.cuhk.edu.hk/english/index.html>
(2023年3月5日参照)
- [144] R. Porkess, *The Future of Statistics in our Schools and Colleges*, London: The Royal Statistical Society and the Actuarial Profession, 2012.
<https://rss.org.uk/RSS/media/News-and-publications/Publications/Reports%20and%20guides/rss-reports-future-statistics-schools-colleges-roger-porkess-2012.pdf>
(2023年3月5日参照)
- [147] The Royal Statistical Society, Accreditation scheme.
<https://rss.org.uk/membership/professional-development/accreditation-scheme/>
(2023年3月5日参照)
- [146] The Royal Statistical Society, The RSS Accreditation and Quality Mark Schemes: A Guide for Accredited Partners, April 2021.
<https://rss.org.uk/RSS/media/File-library/Membership/Prof%20Dev/RSS-Accreditation-and-Quality-Mark-Guidelines-0421.pdf>
(2023年3月5日参照)
- [147] University of Cambridge, Department of Pure Mathematics and Mathematical Statistics.
<https://www.dpmms.cam.ac.uk/>
(2023年3月5日参照)
- [148] Research Excellence Framework, REF 2021.
<https://www.ref.ac.uk/>
(2023年3月5日参照)
- [149] Faculty of Mathematics, University of Cambridge, PhD in Mathematics of Information.
<https://www.maths.cam.ac.uk/postgrad/phd-mathematics-information/cmi>
(2023年3月5日参照)
- [150] 公益財団法人教科書研究センター委託研究、西村圭一（研究者代表）「算数・数学の教科書の世界的潮流に関する調査研究」、2022年7月。
<https://textbook-rc.or.jp/wp-content/uploads/2022/10/80160bf7b38e4d1fd6ab6aa113f8bd3e.pdf>

(2023年3月5日参照)

- [151] 「日本統計学会会報」194号p. 3 - 7、飯塚誠也 (運営委員長)・小森理 (実行委員長)・黒田正博 (プログラム委員長)「2022年統計関連学会連合大会の報告」、2023年1月31日.

<https://www.jss.gr.jp/wp-content/uploads/K194.pdf>

(2023年3月5日参照)

- [152] 独立行政法人日本学術振興会、「審査区分表等」.

https://www.jsps.go.jp/j-grantsinaid/02_koubo/shinsakubun.html

(2023年3月5日参照)

- [153] P. Hall, *The Bootstrap and Edgeworth Expansion*. Springer, 1992.

- [154] J. Shao and D. Tu, *The Jackknife and Bootstrap*. Springer, 1995.

- [155] 統計教育連携ネットワーク (JINSE) 公式ウェブサイト.

<https://jinse.jp/>

(2023年3月5日参照)

<参考資料1>審議経過

令和3年

- 1月21日 数理科学委員会数理統計学分科会（第1回）（遠隔会議）
役員の選出、今期の活動方針について

令和4年

- 3月7日 数理科学委員会数理統計学分科会（第2回）（遠隔会議）
意思の表出を作成することについて
数理統計学分科会から報告あるいは見解を作成することとした。
- 9月24日 数理科学委員会数理統計学分科会（第3回）（遠隔会議）
学術会議からの助言と見解の作成について
4月1日に当分科会より提出した申出書、それに対する学術会議からの助言及び調査結果について検討し見解を作成することとした。

令和5年

- 2月17日 公開シンポジウム「数理・データサイエンス・AI時代における統計科学の教育及び研究について」の開催
- 2月17日 数理科学委員会数理統計学分科会（第4回）（対面会議）
原稿の案の検討
全体的な内容、全体構成、用語定義について検討した。
- 3月28日 数理科学委員会数理統計学分科会（第5回）（メール審議）
見解（案）「大学における数理・データサイエンス・AI教育の中での統計学の教育について」の承認について
3月30日付で承認した。
- 9月21日 科学的助言等対応委員会承認
見解「大学における数理・データサイエンス・AI教育の中での統計学の教育について」

＜参考資料2＞公開シンポジウム「数理・データサイエンス・AI時代における統計科学の教育及び研究について」

同シンポジウムの概要は以下の通りである。

日時：令和5年（2023年）2月17日（金）13：30～15：30

開催地：ハイブリッド開催（日本学術会議講堂、オンライン）

プログラム：

総合司会 渡辺 美智子（日本学術会議連携会員、立正大学データサイエンス学部教授）

13:30 開会挨拶及び趣旨説明

竹村 彰通（日本学術会議連携会員、滋賀大学学長）

13:45 講演Ⅰ『機械学習分野における研究・人材育成の動向と課題』

杉山 将（理化学研究所 革新知能統合研究センター センター長、東京大学 大学院新領域創成科学研究科複雑理工学専攻 教授）

14:15 講演Ⅱ『国際および国内産業界における統計的手法を活用した問題解決手法・シックスシグマ（Six Sigma）の現状～統計プロフェッショナル育成と国際認証に関する課題と展望を踏まえて～』

石山 一雄（ISO TC69 統計的方法の応用/SC7 シックスシグマ日本代表委員、ISO TC69/SC7・SC8 国内合同委員会副査、中国質量協会シックスシグマ管理推進工作委員会 専門家委員）

14:45 総合討論

司会 青嶋 誠（日本学術会議連携会員、筑波大学数理物質系教授）

1 石山 一雄

2 戸谷 圭子（日本学術会議連携会員、明治大学専門職大学院グローバルビジネス研究科教授）

3 中谷 多哉子（放送大学情報コース教授）

4 竹村 彰通

15:25 閉会挨拶

栗木 哲（日本学術会議連携会員、情報・システム研究機構統計数理研究所教授）

主催：日本学術会議数理科学委員会数理統計学分科会、数理科学委員会数学教育分科会、数理科学委員会数学分科会

共催：統計関連学会連合、特定非営利活動法人横断型基幹科学技術研究団体連合、一般財団法人統計質保証推進協会

後援：一般社団法人日本統計学会、応用統計学会、一般社団法人人工知能学会、一般社団法人情報処理学会、日本行動計量学会、一般社団法人日本数学会、公益財団法人日本数学教育学会、一般社団法人日本経済学会、一般社団法人社会情報学会、一般社団法人経営情報学会、一般社団法人日本品質管理学会、大学共同利用機関法人情報・システム研究機構統計数理研究所

ホームページ：<https://www.scj.go.jp/ja/event/2023/336-s-0217.html>

<付録>

(1) 図表

表 1 日本学術会議における提言と統計科学との関係 (2012 年以降)

| 分科会 | 年・月・日 | 提言 | 統計科学との関係 |
|---------------------------------------|-----------|---|---|
| ゲノムコホート研究体制検討分科会 | 2012.8.8 | ヒト生命情報統合研究の拠点構築—国民の健康の礎となる大規模コホート研究— | 多種多様なデータと疾患との関連を統計学的に解析する新しい理論の構築 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t155-1.pdf |
| ゲノムコホート研究体制検討分科会 | 2013.7.26 | 100 万人ゲノムコホート研究の実施に向けて | 多種多様なデータと疾患との関連を統計学的に解析する新しい理論の構築 (上と同じ) |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t176-1.pdf |
| 情報学委員会 E・サイエンス・データ中心科学分科会 | 2014.9.11 | ビッグデータ時代に対応する人材の育成 | 恒久的なデータ解析部門の設置 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t198-2.pdf |
| 社会学委員会社会統計調査アーカイブ分科会 | 2017.9.19 | 社会調査をめぐる環境変化と問題解決に向けて | 正確な調査を実施するための方法論の検証 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t248-7.pdf |
| 総合工学委員会・機械工学委員会合同計算科学シミュレーションと工学設計分科会 | 2017.8.8 | 科学的知見の創出に資する可視化に向けて | 可視化されたデータから設定された仮説の検証 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t247-5.pdf |
| 科学的エビデンスに基づく「スポーツの価値」の普及の在り方に関する委員会 | 2020.6.18 | 科学的エビデンスを主体としたスポーツの在り方 | スポーツにおけるデータの統計科学的な活用 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t290-5.pdf |
| 化学委員会化学企画分科会 | 2020.7.7 | 化学・情報科学の融合による新化学創成に向けて | 膨大な化学関連データから新たな情報や新たな発想を引き出すための統計科学を含む AI 技術の活用 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t292-1.pdf |
| 社会学委員会 Web 調査の課題に関する検討分科会 | 2020.7.10 | Web 調査の有効な学術的活用を目指して | Web 調査の利用に関する統計科学的な検証 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t292-3.pdf |
| 経済学委員会数量的経済・政策分析分科会 | 2020.9.18 | 行政記録情報の活用に向けて | 行政記録情報の活用に関する統計理論的・統計制度的な検討 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t297-3.pdf |
| 経営学委員会経営学大学院における認証評価の国際通用性に関する分科会 | 2020.9.29 | わが国の経営学大学院における教育研究の国際通用性のある質保証に向けて | 大学院教育におけるデータ整備 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t298-4.pdf |
| 心理学・教育学委員会・情報学委員会合同教育データ利活用分科会 | 2020.9.30 | 教育のデジタル化を踏まえた学習データの利活用に関する提言—エビデンスに基づく教育に向けて— | 学習データの統計科学的な利活用 |
| | | | https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t299-1.pdf |

出典：数理科学委員会数理統計学分科会で作成

表2 統計家とデータサイエンティスト(DS)の就業者数と学歴構成比(2021年5月)

| 就業者数 | 統計家 | DS |
|----------------------|--------|---------|
| 全産業 | 31,370 | 105,980 |
| 科学的研究開発サービス | 5,230 | 5,840 |
| 政府（公立の学校・病院・郵便局を除く） | 4,680 | 7,330 |
| 大学、専門学校 | 2,390 | 2,190 |
| 経営的・科学的・技術的コンサルティング | 2,050 | 7,270 |
| コンピューターシステム設計と関連サービス | 2,030 | 16,620 |
| 会社および企業経営 | 1,480 | 12,570 |
| 信用仲介および関連サービス | 340 | 5,690 |
| 学歴（構成比%） | | |
| 総数 | 100 | 100 |
| 博士号取得者 | 14 | 1 |
| 修士号取得者 | 35 | 14 |
| 学士号取得者 | 37 | 55 |
| その他 | 14 | 30 |

出典：資料US Department of Labor, Careeronestop (<https://www.careeronestop.org/>) の一部を Data Tools を使って集計し、数理科学委員会数理統計学分科会で作成



図1 データ分析の訓練を受けた大学卒業生の数（2008年 単位：千人）

出典：[85]

Q. 貴社にはデータサイエンティストは何名いらっしゃいますか。

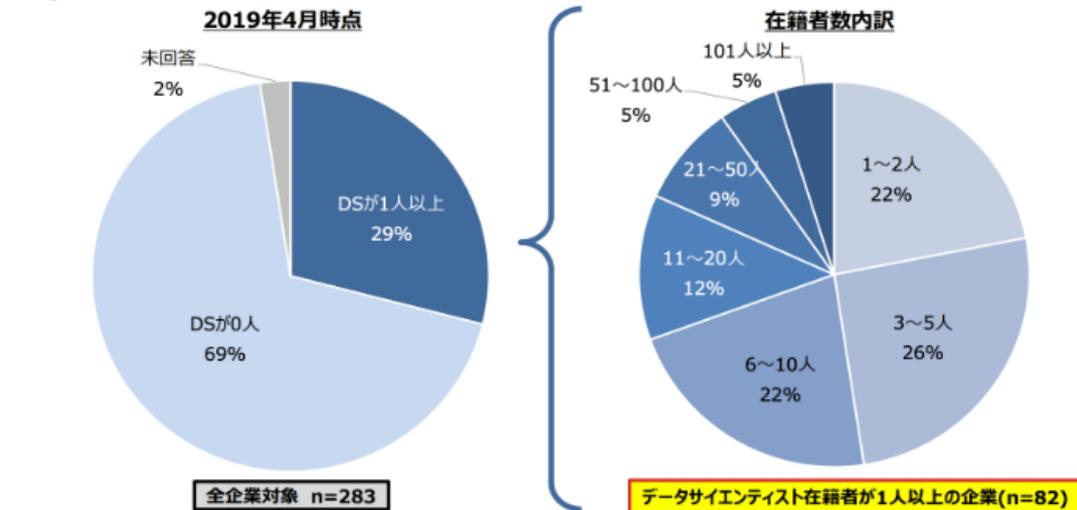


図2 データサイエンティストの在籍実態

出典：[86]

表3 学習指導要領（高等学校）の変移

| 平成 29・30・31 年改訂 | 平成 20・21 年改訂 | 平成 10・11 年改訂 |
|---|---|---|
| 数学 I (4)データの分析 データの散らばり ・分散，標準偏差 データの相関 ・散布図，相関係数 仮説検定の考え方 | 数学 I (4)データの分析 ア データの散らばり イ データの相関 | 数学B (3)統計とコンピュータ ア 資料の整理 度数分布表、相関図 イ 資料の分析 代表値、分散、標準偏差、 相関係数 |
| 数学B (2)統計的な推測 確率分布 ・確率変数と確率分布 * 確率変数の平均，分散，標準 偏差 ・二項分布 正規分布 ・連続型確率変数 ・正規分布 統計的な推測 ・母集団と標本 ・統計的な推測の考え * 区間推定，仮説検定 | 数学B (1)確率分布と統計的な推測 ア 確率分布 (ア) 確率変数と確率分布 (イ) 二項分布 イ 正規分布 ウ 統計的な推測 (ア) 母集団と標本 (イ) 統計的な推測の考え | 数学C (3)確率分布 ア 確率の計算 イ 確率分布 (ア) 確率変数と確率分布 (イ) 二項分布 (4)統計処理 ア 正規分布 (ア) 連続型確率変数 (イ) 正規分布 イ 統計的な推測 (ア) 母集団と標本 (イ) 統計的な推測の考え |

出典：[98]-[104]

表4 高等教育機関等におけるデジタル人材の育成確保【文部科学省】

| |
|--|
| <ul style="list-style-type: none"> • 数理・データサイエンス・AI教育の推進（2025年に応用基礎25万人/年） 各大学等の成果を全国へ普及・展開させるためのコンソーシアム活動等への支援を通じ、大学等において、文理を問わず数理・データサイエンス・AIを応用する力を持った人材の育成を加速 |
| <ul style="list-style-type: none"> • リカレント教育の推進（受講者1,000人/年） 大学・専門学校等が自治体や企業等と連携してDX等成長分野に関してリテラシーレベルの能力取得・リスキリングを実施するプログラムを支援。 |

出典：[117]



図3 統計検定の申込者数と受験者数

出典：統計検定センターから提供を受けたデータ

表5 リテラシーレベルの教育の基本的考え方

| |
|---|
| <p>数理・データサイエンス・AIを活用することの「楽しさ」や「学ぶことの意義」を重点的に教え、学生に好奇心や関心を高く持ってもらう魅力的かつ特色ある教育を行う。数理・データサイエンス・AIを活用することが「好き」な人材を育成し、それが自分・他人を含めて、次の学修への意欲、動機付けになるような「学びの相乗効果」を生み出すことを狙う。</p> |
| <p>各大学・高専においてカリキュラムを実施するにあたっては、各大学・高専の教育目的、分野の特性、個々の学生の学習歴や習熟度合い等に応じて、本モデルカリキュラムの中から適切かつ柔軟に選択・抽出し、有機性を考慮した教育を行う。</p> |
| <p>実データ、実課題を用いた演習など、社会での実例を題材に数理・データサイエンス・AIを活用することを通じ、現実の課題と適切な活用法を学ぶことをカリキュラムに取り入れる。</p> |
| <p>リテラシーレベルの教育では「分かりやすさ」を重視した教育を実施する。</p> |

出典：[124]

表6 リテラシーレベル モデルカリキュラムの構成

| | | | | | | | | | | | |
|-----------------------|---|-----------------------|--------------------|----------------------|----------------------|-------------------|---------------------|----------------|----------------------|----------------------|--|
| 導入 | <p>1. 社会におけるデータ・AI利活用</p> <table border="1"> <tr> <td>1-1. 社会で起きている変化</td> <td>1-2. 社会で活用されているデータ</td> </tr> <tr> <td>1-3. データ・AIの活用領域</td> <td>1-4. データ・AI利活用のための技術</td> </tr> <tr> <td>1-5. データ・AI利活用の現場</td> <td>1-6. データ・AI利活用の最新動向</td> </tr> </table> | 1-1. 社会で起きている変化 | 1-2. 社会で活用されているデータ | 1-3. データ・AIの活用領域 | 1-4. データ・AI利活用のための技術 | 1-5. データ・AI利活用の現場 | 1-6. データ・AI利活用の最新動向 | | | | |
| 1-1. 社会で起きている変化 | 1-2. 社会で活用されているデータ | | | | | | | | | | |
| 1-3. データ・AIの活用領域 | 1-4. データ・AI利活用のための技術 | | | | | | | | | | |
| 1-5. データ・AI利活用の現場 | 1-6. データ・AI利活用の最新動向 | | | | | | | | | | |
| 基礎 | <p>2. データリテラシー</p> <table border="1"> <tr> <td>2-1. データを読む</td> <td>2-2. データを説明する</td> </tr> <tr> <td>2-3. データを扱う</td> <td></td> </tr> </table> | 2-1. データを読む | 2-2. データを説明する | 2-3. データを扱う | | | | | | | |
| 2-1. データを読む | 2-2. データを説明する | | | | | | | | | | |
| 2-3. データを扱う | | | | | | | | | | | |
| 心得 | <p>3. データ・AI利活用における留意事項</p> <table border="1"> <tr> <td>3-1. データ・AIを扱う上での留意事項</td> <td>3-2. データを守る上での留意事項</td> </tr> </table> | 3-1. データ・AIを扱う上での留意事項 | 3-2. データを守る上での留意事項 | | | | | | | | |
| 3-1. データ・AIを扱う上での留意事項 | 3-2. データを守る上での留意事項 | | | | | | | | | | |
| 選択 | <p>4. オプション</p> <table border="1"> <tr> <td>4-1. 統計および数理基礎</td> <td>4-2. アルゴリズム基礎</td> </tr> <tr> <td>4-3. データ構造とプログラミング基礎</td> <td>4-4. 時系列データ解析</td> </tr> <tr> <td>4-5. テキスト解析</td> <td>4-6. 画像解析</td> </tr> <tr> <td>4-7. データハンドリング</td> <td>4-8. データ活用実践（教師あり学習）</td> </tr> <tr> <td>4-9. データ活用実践（教師なし学習）</td> <td></td> </tr> </table> | 4-1. 統計および数理基礎 | 4-2. アルゴリズム基礎 | 4-3. データ構造とプログラミング基礎 | 4-4. 時系列データ解析 | 4-5. テキスト解析 | 4-6. 画像解析 | 4-7. データハンドリング | 4-8. データ活用実践（教師あり学習） | 4-9. データ活用実践（教師なし学習） | |
| 4-1. 統計および数理基礎 | 4-2. アルゴリズム基礎 | | | | | | | | | | |
| 4-3. データ構造とプログラミング基礎 | 4-4. 時系列データ解析 | | | | | | | | | | |
| 4-5. テキスト解析 | 4-6. 画像解析 | | | | | | | | | | |
| 4-7. データハンドリング | 4-8. データ活用実践（教師あり学習） | | | | | | | | | | |
| 4-9. データ活用実践（教師なし学習） | | | | | | | | | | | |

6

出典：[124]

(2) 諸科学におけるビッグデータの統計解析

ここでは、いくつかの自然科学分野においてビッグデータの統計解析が進んでいる例を挙げる。

ア 地震データの統計解析

近年の観測機器の性能向上や観測体制の整備により地震データの収集量の増大も著しく、現時点では我が国だけでも 2000 点以上の観測点で高精度地震計測データが 1 秒間 200 回の頻度で常時収集されている。これに加え、建物、電気・ガス等の IoT データ、スマートフォンのセンサデータを統合的に活用する地震ビッグデータベース構築が試

みられている。このような地震ビッグデータ時代を見据え、スパースモデリング、機械学習、データ同化、ベイズ統計、ベイズ最適化の地震データへの活用を目的の1つとしたCREST 課題[125]やSTAR-E プロジェクト[126]が数年前より開始している。スパースモデリングと p-値統合法を用いて、衛星測位データからの地震のゆっくり滑り自動検出手法の開発などが行われている。ここでスパースモデリングとは、1996年に統計学者 R. Tibshirani [127]が提案した Lasso に端を発するデータ解析、データ圧縮、計測情報処理の総称である。

歴史的にも、統計モデルに基づく地震データ解析は、我が国の貢献が大きい。統計数理研究所の尾形良彦が提案した ETAS モデルは地震を（マーク付）点過程モデルと見なしたとき、その強度関数を地震の履歴が反映される形でモデル化したもので、余震のリスクを精確に見積もることができる（[128]）。前述の STAR-E プロジェクトは ETAS モデルの改良、近代化も目的の1つとしている。情報量規準や残差解析を通して、モデルが現実を大きなバイアス無く反映するかどうかを確認できることは、統計的方法の顕著な利点である。

イ 気象学・海洋学

気象学、海洋学で観測されるデータは典型的なビッグデータである。そこではデータ同化（Data Assimilation）と呼ばれる、観測データを用いてシミュレーションを改善する方法論が中核技術として用いられている。必要なシミュレーションも膨大な計算量となるため、利用可能な計算資源の中でパフォーマンスを上げることが追求されている。データ同化は統計的方法の持つ汎用性の性格から、気象学、海洋学以外の広範な分野で適用されるようになってきている。

統計的方法をベースに発展を遂げているデータ同化手法に、逐次型データ同化がある。時系列解析で伝統的に使われてきた状態空間モデルを介してベイズ更新則（カルマンフィルタ）が適用可能となり、数値的に逐次計算が容易となる。逐次型データ同化はこの考えを高度に複雑化したものである。まず、物理モデルと観測過程を状態変数モデルで記述する。モデルは非ガウス、非線形であり、多くの未知パラメータを含む。その際に (a) 未知パラメータを状態変数として組み込み逐次推定する（自己組織化）、(b) 状態変数の分布を、解析的ではなく多くの実現値の集合（アンサンブル）として記述する、(c) ベイズ更新則をサンプリングとして記述する（粒子フィルタ）、という工夫を取り込むことにより、未知のパラメータを含む状態空間モデルによるシミュレーションを実行することができる。

逐次型データ同化モデルは、その状態変数の意味合いが明確で、その点で機械学習のいわゆるブラックボックスモデルとは大きく異なる。しかしながら特に粒子フィルタを取り込んだデータ同化の方法は、計算量の観点で現在のところ現業段階には至っておらず、さらなる実用化研究が進められている[129]。

より大規模なデータに適用可能な4次元変分法による非逐次型データ同化法に関しても、欠点とされた不確実性評価が可能となるアルゴリズムが開発された[126], [127]。

これは常に分布や誤差を考える統計的方法の利点である。

ウ 天文学における統計的手法

天文学においても、観測機器の性能向上が著しい。例えば、すばる望遠鏡の Hyper-Supreme Camera は1スナップショットで1ギガピクセルの画像が取得され、銀河カタログの作成には数百テラバイトのデータが必要となる。そのために、従前のデータ解析法だけではなく、このような天文ビッグデータに対応する新たな方法論が必要とされている。例えば CREST[130], [131]ではそのような方法論を「統計計算宇宙物理学」と命名し、ビッグデータ解析のための手法を開発している。開発された手法の一例として、AUCブースト法やpAUC法がある。これは統計科学の基本的ツールであるROC曲線の下側面積であるAUCと機械学習手法であるブースティングを融合させたアルゴリズムで、新星候補天体抽出に用いられる。また宇宙モデルに対して物質分布を高速計算するエミュレータの開発では、主成分分析、ガウス過程、ノンパラメトリック平滑化といった統計計算が用いられている。

国際天文プロジェクト「イベント・ホライズン・テレスコープ (EHT)」は、M87 銀河の中心の巨大ブラックホールのブラックホールシャドウを撮影することに成功した。データは、チリ、スペイン、南極、ハワイ、メキシコ、米国アリゾナの8つの望遠鏡を用いて取得され、容量は合計3500テラバイト(3.5ペタバイト)であった。本プロジェクトは、データの取得、装置開発、画像解析、理論・シミュレーション解析の多岐にわたるものであるが、その中の画像解析のプロジェクトでは、観測データを解析してシャドウの画像を得た。その際には日本チームが開発したスパースモデリングが本質的な役割を果たしている[132]-[135]。

また、太陽系外における地球型惑星の探索では、多色撮像データに対して、ガウス過程回帰によるモデリングやレプリカ交換モンテカルロ法による惑星の個数推定が行われ、地球型惑星や新しいタイプの惑星が発見された[136]。

そのほか、ここに紹介できなかったライフサイエンス、マテリアルインフォマティクス、ビジネス、ネットワーク分析における事例は[137]が参考になる。

(3) 諸外国の大学におけるデータサイエンス教育の例

ここでは、諸外国におけるデータサイエンス教育の様々な形態について、アメリカの3つの大学、中国、英国における現状について説明する。

代表的なケースは以下の4通りである。

- (a) データサイエンス学科など新しいユニット創設
- (b) 統計学科、コンピュータサイエンス学科など既存の組織の再利用
- (c) 学術ユニットにも属さない新しい組織(研究所、センター)の創設
- (d) 学内の複数の学術ユニットと共同で新しい組織を創設

また、教育プログラムは概ね以下のように類別される。

- 全ての学部生にプログラミングと確率・統計学を教えるためのスキルセットを提供する。

例えば、カリフォルニア大学バークレー校の“data8”やコロンビア大学の“Data: Past, Present, Future”がその例である。

- データサイエンスを専門としない学生向けには機械学習による上級データサイエンスコースを用意する。
- 既存のコンピュータサイエンスと統計学を組み合わせるデータサイエンス専攻を創設する。
- 新しい専攻を作る代わりに、既存の専攻（コンピュータサイエンスまたは統計学）に専門分野を設ける。
- 生物学や経済学などの既存の分野別専攻にデータサイエンス分野を追加する。
- データサイエンス副専攻を創設する。
- データサイエンスと他の学問分野との複合専攻を創設する。

報告書[78]は、このような多様性は好ましいものであり、今後は組織的な連携が必要であるとしている。また報告書[78]の集会については、[138]、[139]でも紹介されている。

次に、アメリカのカーネギーメロン大学、イリノイ大学、マサチューセッツ工科大学、中国、英国における統計学教育の現状について述べる。

① カーネギーメロン大学 (Carnegie Mellon University) ([140])

ディートリヒ人文社会カレッジ(Dietrich College of Humanities and Social Sciences)に統計及びデータサイエンス学科(Department of Statistics & Data Science)がある。統計学科の中にデータサイエンスが設置された例である。

その規模は、550人以上の学部生、100名以上の大学院生、50名以上の教職員、14スタッフである。また学部卒学生の進路は、就職71%、進学24%、未定4%、主な就職先はIT企業、金融、コンサルティング企業のソフトウェアエンジニア、データサイエンティスト、アナリスト等で、就職先の年収平均90000USD中央値85000USDとのことである。

この学科で取得できる学位(学部)は、Bachelor of Science in Statistics(統計学学士)、Minor in Statistics(統計学副専攻)、Bachelor of Science in Economics and Statistics(経済と統計学学士)、Bachelor of Science in Statistics and Machine Learning(統計と機械学習学学士)、Mathematical Sciences(数学学学士)である。

授業科目は、StatCore(統計コア)、StatML(統計と機械学習コース)、EconStat(経済統計トラック)、StatMath(統計・数理科学トラック)、StatNeuro(統計・神経科学トラック)の5コースに分類される。

本大学のコンピュータサイエンス学部(School of Computer Science)の機械学習学科(Machine Learning Department)は機械学習全米の拠点の1つでありU.S.News Best Colledge Rankingsでもしばしば1位を獲得する有名校である。教育プログラムも機械学習学科との力強い連携が見て取れる。

現在行われている研究は、統計学の理論と方法論の研究として、計算金融/応用確率モデリング/時系列予測/国勢調査と調査方法/推論の基礎/最適化とアルゴリズム/ノン

パラメトリック法/ネットワーク/統計的機械学習/高次元統計/トポロジー及び機能データ分析/クラスタリングと教師なしメソッド/ベイジアン推論等があり、また学際的研究領域として、天体統計学/バイオインフォマティクス/生物統計学/教育&教育データ分析/公共政策と社会科学/神経科学/プライバシーとサイバーセキュリティ/法と正義が研究されている。

② イリノイ大学([141])

グレインジャー工科大学 (The Grainger College of Engineering) のコンピュータサイエンス学科に教育プログラム「統計とコンピュータサイエンス」が用意されている。既存のコンピュータサイエンスにデータサイエンスが用意された事例である。

イリノイ大学のコンピュータサイエンスは全米トップ5と言われている。歴史的には、教育研究機関が自前で開発して所有した最初のコンピュータ ILLIAC は有名である。また最初の web ブラウザである Mosaic も本大学で開発された。

取得できる学位 (学士) は、コンピュータサイエンス、数学とコンピュータサイエンス、統計学とコンピュータサイエンスである。授業科目は、統計学、並びに統計とコンピュータサイエンスの2つに分類されている。

③ マサチューセッツ工科大学[142]

MIT IDSS (Institute for Data, Systems, and Society:データ・システム・社会研究所) は、統計学とデータサイエンス、情報及び意思決定システム、社会科学の共通部分の研究、教育のための学部から独立した研究所である。

アカデミックプログラムとして、社会工学システム博士課程 (SES)、技術と政策プログラム (TPP)、統計学・データサイエンスの副専攻、学際的博士課程プログラム、並びにオンライン教育プログラムが用意されている。

研究は、統計とデータサイエンス、情報と意思決定理論 (システムと制御、最適化とゲーム理論、ネットワーク) 並びに社会問題のモデル化、意思決定理論、公共政策設計など社会科学に関するものである。

④ 香港中文大学 (Chinese University of Hong Kong) [143]

中国の統計分野でトップレベルとされる本大学においても、統計的思考と計算的思考とを融合させ、統計学をより広い立場から俯瞰して捉えた教育がなされている。同大学では、まず、数学の学部とは別に独立して統計学部があり、同学部は Statistics、Risk Management Science、Quantitative Finance and Risk Management Science の3つのプログラムから成り立っている。また、Statistics と Risk Management Science の大学院プログラムがある。全てのプログラムを通じて、理論と応用の調和を目指し、基盤となる統計学の理論からビジネス、ファイナンス、医学、品質管理、リスク管理等、多様な応用分野にも通ずる多岐にわたる学問を教授している。

⑤ 英国 Royal Statistical Society (RSS)

以上に述べた教育機関とはやや異なり、数学に基づき統計的思考を捉え、近年の AI 技術に代表される計算的思考の技術も、その枠組みの中で解明しようとする立場から教育し研究を行う教育機関もある。Royal Statistical Society (RSS)による冊子「The Future of Statistics in our Schools and Colleges」[144]によれば、国の教育政策として、全ての学生が基本的な統計学の実用的な知識を備えていることを保証する必要があるとし、これには、必要な関連する数学的能力と、それが日常生活にどのように影響するかを理解することが含まれているとしている。さらに、数学は、統計の中心的な役割を果たすとしている。RSS では、教育機関とのパートナーシップの強化を目的として、RSS 認定大学 (RSS Accredited University) としての認証を行っている。さらに、プログラムやモジュール/トレーニングコースについても認証を行い、認証されたプログラムの卒業生に Graduate Statistician (GradStat) の資格を与え、モジュール/コースに RSS Quality Mark を与えるなど、教育・学習の高い質を保証するための活動を行っている[145], [146]。

⑥ ケンブリッジ大学

ケンブリッジ大学における Department of Pure Mathematics and Mathematical Statistics [147]も、RSS と同様の立場にある教育理念と研究理念を持っている。英国の大学研究評価指標である REF (Research Excellence Framework) 2021[148]の結果によれば、同大学の数理科学 (Mathematical Sciences) 分野における総合評価は、独創性、意義、厳密さにおいて世界をリードする質とされている four star レベルの割合が全体の 71%であり、平均の 48%から比べると 1.5 倍程度高く、高い研究力があると評価されている。この評価は、評価軸として、アウトプットの質、インパクト、研究環境の 3つの要素が設定されているが、同大学の評価結果はどの評価軸においても平均を大きく上回っている。特に、社会的インパクトにおいて、影響範囲と重要性の点で顕著なインパクトがあるとされ、その最高レベルの評価の割合が 77.8%であり、平均の 55.4%と比較しても、同大学の数理科学分野が社会的に多大な影響力を持つと評価されていることが分かる。

同大学の Department of Pure Mathematics and Mathematical Statistics の教育理念は、数学がデータサイエンスの現在及び将来の傾向を理解するための重要な能力であるとし、数理に基づく統計学を教授することにある。最新の機械学習アルゴリズムは、後のコースで、または統計的洞察の適用と同時に導入されている。また、Mathematics of Information の PhD プログラム[149]は、統計の真の進歩、さらにはデータサイエンスのより広範な進歩には、数学内での共同作業が必要であるという考えに基づいて構築されている。

(4) 諸外国の教科書と日本の指導要領

比較的最近の資料である「算数・数学の教科書の世界的潮流に関する調査研究」[150]

においても、確率と統計に関する各国の扱いに関して、次の特徴を指摘している。

- 1) 高校生用教科書では、いわゆる文系進学者には統計を中心とした内容で基礎と発展を含め、理系進学者には、統計の基礎と微積・行列など代数を含めた、2種類が発行されている。
- 2) 話し合いや議論を促す問題や医療・健康など現実の探究課題が豊富に掲載され、その中で統計の有用性と活用の意義を小学校の段階から繰り返し学ぶ構成となっている。
- 3) 現実事象と関連付けした統計的確率を早い段階から多く扱い、不確実性や予想の確度の指標として、日常会話の用語との連動も併せて学ぶ構成となっている。

特徴として、イギリスでは批判的な分析を求める問題が掲載され、批判的な分析を行うことにより推測を行い、結論として問題を解決するための提案をさせている。アメリカではアクティブラーニングが重視され、問題状況の協調的探究を行うと共に、クラス全体でその探究における数学的アイディアや原理を分析、抽象化し、さらにそれらを適用する活動が取り入れられている。ニュージーランドでは、PPDAC サイクル（統計的探究サイクル）の単元で様々な分析事例が紹介され、分析の結果のまとめ方やレポートの読み方に関する内容も豊富である。

一方、日本の中学校、高等学校における学習指導要領で提示されている統計的な内容は、次の通りであるが、例示されているデータ量も少なく現実の社会課題との結び付きが適切に示されていない。

中学校「資料の活用」

- ・ヒストグラムや代表値の必要性と意味及び応用
- ・確率の必要性と意味、簡単な場合に確率を用いて不確定な事象を捉える
- ・標本調査の必要性と意味、簡単な例で母集団の傾向を捉える

高等学校「データの分析」

- ・データの散らばり（四分位範囲、分散、標準偏差などの意味と活用）
- ・データの相関（散布図、相関係数の意味と応用）

高等学校「確率分布と統計的な推測」

- ・確率変数と確率分布、不確定な事象の考察への活用（二項分布、正規分布）
- ・統計的な推測（標本調査の考え方と母集団の統計的な推測）

このように、日本の中等教育においても統計的手法に関する最小限の主題は含まれているが、海外の統計やデータサイエンス教育は、専門学術団体や研究者・企業との連携を踏まえ、さらに統計的な思考を強化する方向で展開されており、日本でも、統計教育のさらなる充実が求められる。

(5) 数理・データサイエンス・AI 教育強化拠点コンソーシアムのモデルカリキュラム

数理・データサイエンス・AI 教育強化拠点コンソーシアムは、その活動の1つとして、分野を問わず、全ての大学生・高専生を対象にした、リテラシーレベルの教育の基本的考え方、学修目標・スキルセット、教育方法等について、意見募集を経て、「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～」[124]をまとめた。そこでは、「リテラシーレベルの教育の基本的考え方」「リテラシーレベルの学修目標・スキルセット」「リテラシーレベルの教育方法」「リテラシーレベルの実施展開に向けた取組」について述べられている。

リテラシーレベルの教育の基本的考え方は、付録表5に示す4つの考え方に基づきカリキュラムを実施することを期待している。

学習目標を「今後のデジタル社会において、数理・データサイエンス・AI を日常の生活、仕事等の場で使いこなすことができる基礎的素養を主体的に身に付けること。そして、学修した数理・データサイエンス・AI に関する知識・技能を基に、これらを扱う際には、人間中心の適切な判断ができ、不安なく自らの意志でAI等の恩恵を享受し、これらを説明し、活用できるようになること」とし、モデルカリキュラム(付録表6)は、コア学習項目として位置付けている「導入」「基礎」「心得」と、学生の学習歴や習熟度合いに応じて適宜選択する「選択」の4つの学修項目を設けた。

教育方法は、「導入」「基礎」「心得」にわかれている。「導入」では、動画を使った反転学習、講義ではデータ・AI 活用領域の広がりや技術概要の解説を行う、学生がデータ・AI 利活用事例を調査し発表するグループワークを行い、一方通行の講義ではなく、ディスカッションが起こる講義が望ましいとしている。「基礎」では、実際に学生自身がデータ利活用プロセスの一部を体験できること、必要に応じてフォローアップ講義を準備することが望ましい。「心得」では、データ駆動型社会のリスクを自分ごととして考えること、データ・AI が引き起こす課題についてグループディスカッション等を行うことが望ましい。選択部分でも、各大学・高専の特徴に応じて適切なテーマを設定し、実データ(あるいは模擬データ)を用いた講義を行うことが望ましいとしている。

(6) 統計関連の学会と科学研究費の状況

日本で最大の統計関連学会連合大会は、年々講演件数と参加人数が増加し、2022年度の講演件数は302件、大会参加登録者数は968名であった([151])。本大会は、統計関連6学会(応用統計学会、日本計算機統計学会、日本計量生物学会、日本行動計量学会、日本統計学会、日本分類学会)からなる統計関連学会連合により実施され、統計科学の普及・発展を目的とし、種々の共同事業を推進している。統計科学の必要性は認識されているものの、各学会の学会員に関しては順調に増加しているとは言い難い状況である。

さらに、科学研究費補助金(科研費)に関しては、「統計」という単語が含まれるものが、「経済学、経営学及びその関連分野」の小区分に「経済統計関連」、「情報科学、情報工学及びその関連分野」の小区分に「統計科学関連」として置かれているだけである[152]。このような動きは統計科学の研究者にとって不自由なものであり、研究機関、科学研究

費など研究の基本となるものを早急に整える必要がある。

脚注:

ⁱ ここで言う生成モデルは、変量間の確定的な関係や確率的な誤差の入り方を、変数間の順序を考慮した上で数式を用いて表した数理モデルのことを言う。最近注目を浴びている深層学習手法を用いた画像や文章の生成モデルとは異なる意味である。

ⁱⁱ 本見解の作成作業中において社会学委員会社会統計調査アーカイヴ分科会及び総合工学委員会・機械工学委員会合同計算科学シミュレーションと工学設計分科会とも意見交換を行ったが、それらの見解でも統計科学の重要性が論じられている。情報学委員会情報学教育分科会及び数理科学委員会数学教育分科会からも有益なコメントを頂いた。

ⁱⁱⁱ 例えば、今や標準的な手法の1つとなったブートストラップ法についても、初期の頃は、どのような条件の下で当該の手法が使用できるのか、必要な修正が何かを明らかにすることが研究の中心であった[153], [154]。

^{iv} LTCMは数理ファイナンスの技法を使って投資収益を上げることを目的に作られたファンドであったが、統計モデルのパラメータ推計のための仮定が強すぎたことが原因で、それが成立しない局面（ロシア債務危機）ではモデルは機能しなかった。

^v 特にカリフォルニア州では、2010年に全米科学財団、カリフォルニア州教育委員会、カリフォルニア州立大学統計学部が連携して教育内容を革新し、既に中高校生は、SNS、モバイルセンサー等で「データの洪水」の世界を生きている。中高生がデータの生産者かつ統計情報の賢い消費者として、ビッグデータや多くのオープンデータソースを理解するためにカリフォルニア州は必要な知識と実践技術を学ぶ中高生向けコース「データサイエンス入門」Mobilizeプロジェクトを製作し、既に州の多くの高校が1年間のこのコースを組み入れた授業を2014年から展開している。Mobilizeプロジェクトでは、生徒が自らセンシング技術を通して大量にデータを収集する仕組み、そのデータを整理して分析・可視化を行うソフトウェア等、授業支援教材や授業そのものをデジタル化しており、教員・教科書型知識供与が中心である日本のデータ教育とは一線を画している。

^{vi} ニュージーランドは、PISA2003の数学リテラシー「不確実性とデータ」領域で上位5%の得点が1位を獲得したほか、PISA2006の「科学リテラシー」試験の上位5%の得点2位、IGCE-Aレベルの数学2005で1位、IGCSE数学2006で1位となるなど、高い成績をおさめてきた。2009年には、科目「数学」の名称を「数学と統計」に変更し、さらなる統計教育の充実を中心としたカリキュラム改革を行った。義務教育であるレベル6までは、レベルごとに「数と代数」、「幾何と測定」、「統計」の3領域でカリキュラムが構成されている。レベルが上がるごとに「数と代数」が少なくなり、統計が多くなっている。また、レベル7と8は進路に合わせて「数学」と「統計」の別科目を選択できる。特に「統計」科目内容に関しては、ニュージーランド統計局、統計学会、統計教育研究者が中心となって適切に学習内容と方法、教材開発、教師研修支援の充実が行われており、日本でも参考とすべきである。

^{vii} 2022年時点で次の研究科等が挙げられる。横浜市立大学データサイエンス研究科、滋賀大学データサイエンス研究科、京都大学臨床統計家育成コース、順天堂大学医学研究科データサイエンスコース、武蔵野大学データサイエンス専攻（修士課程）、東京情報大学データサイエンス研究室、立教大学人工知能科学研究科、中央大学ビジネスデータサイエンス専攻

^{viii} 2022年3月までは「数理・データサイエンス教育強化拠点コンソーシアム」で、2022年4月に現在の名前に変更された。

^{ix} AI戦略2019は、日本の大学における数理・データサイエンス・AI教育に大きな影響を与

えたが、その後も改訂されている。AI 戦略 2022(2022 年 4 月 22 日 統合イノベーション戦略推進会議) [27] ではパンデミックや大規模災害等の差し迫った危機への対処などが盛り込まれている。

^x 日本学術会議は 2010 年に「大学教育の分野別質保証の在り方について」を取りまとめ、さらに引き続きいくつかの分野に関して参照基準の策定を進めてきた。統計学に関しては、2010 年 8 月に統計関連学会連合理事会、同統計教育推進委員会によって、報告書「統計学分野の教育課程編成上の参照基準」 [123] が公表された。統計関連学会連合がこの報告書を作成した理由は、日本学術会議に独立した「統計学」分野が存在せず、「大学教育の分野別質保証の在り方について」で示された 30 分野にも含まれていないためであった。日本学術会議より公表されている数理科学分野や情報学分野の参照基準も統計学の教育と深い関連がある。

^{xi} 統計教育連携ネットワーク(拡大版 JINSE)は、連携学会(6 学会: 応用統計学会 日本計算機統計学会 日本計量生物学会 日本行動計量学会 日本統計学会 日本分類学会)、連携団体(7 団体: 日本アクチュアリー会 日本銀行調査統計局 日本経済団体連合会 日本製薬工業協会 日本統計協会 日本マーケティング・リサーチ協会 統計数理研究所)をはじめとして、組織会員、個人会員などが参画し、海外統計教育講演会などの活動を行っている [155]。

^{xii} 2020 年度は紙媒体の試験は実施されていない。また、CBT 試験会場のいくつかは閉鎖された。2021 年度の実施においても COVID-19 の感染の影響を受けている。

^{xiii} 本データは統計検定センターより提供を受けた。