

「特集・心理統計学の新展開」について

岡田 謙介¹・持橋 大地² (オーガナイザー)

心理統計学は、人間の心と行動のメカニズムを探求する心理学において、データの収集、分析、解釈のために必要となる統計学的方法を研究する分野である。そのスコープに含まれる統計モデルは多岐にわたるが、とくに人の能力、態度、性格特性、精神的健康といった直接観測できない構成概念を、質問項目への解答・回答データに基づいて測定する方法を大きく発展させてきたことは、心理統計学を特徴づける要素である。経済現象の解明や予測のために用いる統計モデルを扱う研究分野は計量経済学(econometrics)であり、生物学的現象の定量的理解に関する統計モデルを扱う研究分野は計量生物学(biometrics, 生物測定学とも)である。これらと同様に、心理学的構成概念の測定のための理論と方法を提供する統計モデルを扱う研究分野は計量心理学(psychometrics, 心理測定学とも)と呼ばれる。

計量心理学の統計モデル研究は、Spearman (1904)による、各教科のテスト成績の背後に想定された一般知能因子を抽出する因子分析(factor analysis)法の開発に始まる。1950年代からは、問題項目への正答・誤答のようなカテゴリカルな観測データに基づき、解答者の能力や特性を推定する項目反応理論(item response theory)モデルが発展し、様々な教育テストに応用されるようになった。また1980年代からは、解答者の知識習得状態を詳細に分析し、個々の学習や成長の進行状況を評価し、ニーズに合わせたフィードバックを提供するための診断分類モデル(diagnostic classification model, または認知診断モデル cognitive diagnostic model)の研究が進展した。

因子分析、項目反応理論、診断分類モデルに代表される、心理学的構成概念の測定のための統計モデルは測定モデル(measurement model)と総称される。心理学的構成概念を適切に定義し、測定することは、心理学研究を行う上で重要な課題であり、その実現には信頼性と妥当性の高い測定尺度の開発が求められる。ここで信頼性は測定が一貫性のある結果をもたらす程度を、妥当性は測定が実際に意図した構成概念を捉えている程度を指す。本特集号が対象とするのは、心理統計学の中でもとくに、こうした計量心理学の測定モデルである。

測定モデルは、教育評価、臨床診断、人事選考といった場面での社会的影響力の大きなテストや、日常的な場面での形成的評価や医療的スクリーニングのためのテストに応用されて、現代の社会や教育、医療・公衆衛生を支えている。心理測定をとりまく近年の変化として、コンピュータで実施されるテストの大幅な増加が挙げられる。たとえば文部科学省 CBT (computer-based testing) システム MEXCBT(メクビット)は、2024年2月現在で公立小学校の80%超、公立中学校のほぼ全てに導入された。これによって、わが国でも義務教育段階から、コンピュータを用いて各種テストを実施することが当たり前になりつつある。自然言語処理や機械学習など、計量心理学の関連分野における技術の飛躍的進歩は、テストの作問や採点に大きな革新をもたらしている。一方で、デジタル技術や社会のネットワーク化の進展は、心理測

¹ 東京大学 大学院教育学研究科：〒113-0033 東京都文京区本郷 7-3-1

² 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

定における倫理やプライバシーといった新たな課題を浮き彫りにもしてきた。また、大規模言語モデルが知能と呼びうるような驚くべき性能を発揮することが明らかになった中で、各種言語モデルの「能力」をどのように測定・評価するかという新たな研究課題も生じており、テスト「解答者」は人だけでなく機械をも含むようになった。このように、測定モデル研究は、従来の枠組みを超えた新しい展開の時を迎えている。

こうした中で、本特集号では計量心理学の統計モデル研究における最近の発展を扱う7本の論文を掲載する。冒頭3篇は、情報技術と測定モデルの融合に焦点を当てたものである。加藤論文では、近年の技術的発展を踏まえた、コンピュータで実施されるテストの解答データを対象とする測定モデルの新たな展開が紹介され、議論される。江原論文では、問題項目が自然言語で記述されることを踏まえ、テキスト情報に基づいて推定される項目の困難度と項目反応理論との関わりが論じられる。また著者により開発された方法をテストデータに適用した結果が示される。渕本・植野論文では、e-Testingにおける項目の露出をコントロールしながら同程度の測定精度を実現するための方法が提案され、その性能を検証した数値実験結果が報告される。

続く4篇の論文は、測定モデルの重要な一側面に関する近年の研究動向をまとめたレビューを提供し、また実践例や数値実験結果を報告するものである。光永論文では、複数のテスト間の対応付けを行うための等化やリンクングの方法がレビューされ、学力調査における分析結果が報告される。国里・竹林論文では、米国国立衛生研究所(NIH)が主導する患者報告アウトカム尺度 PROMIS の尺度開発プロセスが解説され、わが国における課題が論じられる。山口論文では、診断分類モデルの各種推定法が網羅的にレビューされ、現代的な課題と今後の展望が論じられる。また佐宗・岡・宇佐美論文では、小サンプル状況下における推定法間の比較が数値実験によって行われ、議論される。

最後になるが、本特集号が編纂されるきっかけは、岡田が研究者として、持橋が領域アドバイザーとして参画する JST 戦略的創造研究推進事業(さきがけ)研究領域「信頼される AI の基盤技術」における出会いと議論にあった。同研究領域を発端として、このような貴重な機会に恵まれたことに、感謝の意を表したい。本特集号が心理統計学への理解や関心を広げ、その研究を一層前進させる一助となることを願ってやまない。

参 考 文 献

- Spearman, C. (1904). 'General intelligence,' objectively determined and measured, *The American Journal of Psychology*, **15**(2), 201–293, <https://doi.org/10.2307/1412107>.