

2025 年度 統計データ分析コンペティション

論文審査会 総評

令和 7 年 10 月 18 日 審査委員長

第 8 回の統計データ分析コンペティションの受賞論文を選考できたこと、コンペティションを支えていただいている関係者に審査委員会を代表して感謝申し上げたい。

今年は、昨年エントリー件数などが低かった大学生・一般の部については、エントリー件数 175 件（前年 139 件）、論文提出件数 93 件（前年 62 件）、高校生の部のエントリー件数が 163 件（前年 170 件）、論文提出件数が 123 件（前年 124 件）となった。データ分析の結果を政策提言につなげ論文化するという難易度の高いコンペティションのため、エントリー数と提出論文件数には後述するように乖離がある。

しかし、このコンペティションの基本データである SSDSE の年間ダウンロードは年間 20 万件を超え、夏休み前のダウンロード数は多くなっている。

2025 年 9 月 10 日には、SSDSE の作成と普及並びに統計データ分析コンペティションの開催普及について、（独）統計センターが、（一社）日本統計学会統計教育賞を受賞した。受賞については、主催団体・審査員一同喜びを共有するとともに、これも全国の高校・大学関係者への浸透とご支援があればこそあり、SSDSE の充実、コンペティションの発展に努める所存である。

高校生の部

一次審査通過論文を眺める限り、具体的な社会問題を中心に多様なテーマに高校生の問題意識を緻密な仮説検証型や類型化や尺度化を駆使した重層的なデータ解析で深堀し政策提案まで行っていることは頗もしい限りである。

大学生・一般の部でよくみられる実証分析にチャレンジするなど若干の背伸び感は否めない。しかし、今後都道府県別、市区町村別データの限界を超えて、ミクロデータ解析による実証に挑み、将来の日本のデータサイエンスをけん引してくれる人材に育つことを心から期待する。例年同様、実証専門家からみれば膨大なデータのハンドリングの不備はあるし、こういう場面でこの手法を使うのかといった問題は指摘できようが、むしろこのコンペティションの趣旨にそった全力投球の論文が多く寄せられたことには、指導に当たられた教員にも敬意を表したい。

今回、本コンペティションの大学生・一般の部の受賞者が、地域のデータサイエンスの指導者となり高校生の部の受賞者を育てたということも、このコンペティションの将来に大きく期待できることと感じた。

一方、これだけの論文のレベルになった際、例えば重回帰分析の第一段階として相関分析を実施し相関の高いものを候補変数にするとか、多重共線性の背景にあるのが、因果か潜在因子かといった判断もなく VIF を適用するといった、本来のデータ解析戦略では少し疑問のある方法論が高校のトップ

クラスの標準になりつつあることについては、むしろ高校のハイレベルのデータ解析のガイドブックなどの整備が必要なのではないかと思った。もちろん、このような批判を高校生の部の総評に入れること自体が高校生論文の急速なレベルアップによることは言うまでもない。

大学生・一般の部

大学生・一般の部も、練りこまれた実証分析が投稿され、読みごたえが十分なものが多かった。高校生に比べて分野の専門性が高く、先行研究に基づく仮説形成を前提にするため、高校生的な仮説形成の自由度は失われがちなのはやむをえない。仮説が絞り込まれた分、特定の政策効果に絞り込んだ Rubin 流の因果推論や Granger 因果性の検討などは、高校生にはなかなかできない実証である。

また、昨年くらいから大学でも、個人の執筆論文ではなく、高校同様グループでの活動が多くなってきていることも大学におけるデータサイエンス教育のるべき姿を示すものとして頗もしく感じる。

なお一点だけ投稿時に注意いただきたいのだが、本コンペティションに論文投稿される場合にそれが、どこかで発表されたものを再投稿することは、通常の学術論文と同じように評価対象とできないことである。実際受賞論文は月刊「統計」に校閲掲載されるので、著作権上の問題もあり学会誌などの厳格な運用になっていることをご理解いただき、自ら投稿時に判断いただければ幸甚である。

以上