

2022年度 統計データ分析コンペティション

論文審査会 総評

2018年に始まり今回で5回目となるコンペティションを振り返る中で、高校生の部、大学生・一般の部に、多くの努力が捧げられたことがわかる。そして、考えさせられる実証研究が多々寄せられたことは、審査員一同驚きと共に、喜びを感じている。前回に危惧した機械学習の適用というモダンなアプローチによる探索的研究は今回もあるが、その結果を解釈する試みも萌芽しており、次世代研究の将来についても十分期待できるものである。

例年同様、論文審査会では入賞した作品についても様々な誤りや問題点があることが指摘されたが、それらは応募論文の主旨を残す意味で必要最小限の修正を除いて手を加えていない。総務大臣賞などを受賞した8本の論文については、2023年1月号から月刊誌「統計」に掲載されるまでには、問題点は十分解消されるものとなる。しかし一方で、今回の受賞論文の中には、月刊誌「統計」の限られた誌面では、生徒・学生の思考や検討の流れを十分表現できないのではと思われるほどの力作も多々あった。もし余裕があれば、生徒・学生の自由な発想にあふれる原論文と、私たち審査委員会校閲後の少し整理しすぎてしまった論文とを対比されると良い。

ここに掲げた渾身の統計データ分析、しかも様々な試行錯誤を示す生徒・学生オリジナルな論文が、多くの教員・生徒の目に触れ、これらの論文自体が教材となり、新たな統計・データサイエンス教育の規範を育むこととなることを審査員一同期待する。

高校生の部

第5回のコンペティションでは、高校における統計教育が教育現場に浸透してきていることを実感すると共に、高校生の水準をはるかに超えた審査員らが若いころの実証研究の水準に迫るか超えると思われる論文も登場した。もちろん、通常の研究者の論文とは異なり、高校生らしい統計活用や試行錯誤のプロセスも示していることは、高校教員の探究教育活動の指導にとって大いに参考になると思われる。関心のある問題に真摯に取り組み、SSDSE以外のデータの収集に費やしたであろう努力も評価できる活動が多かった。昨年度からの傾向でもあるが、論文構成も先行研究からの問題導出などがしっかりしており、大学生・一般の部

に劣らない読みごたえのあるものも増えてきた。重回帰分析などが高校のデータ分析活動ではもう当たり前になってきたこと、また大学・一般の部以上にモデル当てはめの問題を示す外れ値などについての考察も精緻になり、更にモデルを進化させようという試みがあったことは強く印象に残っている。

入賞を逃した論文についても、論文審査会は学校現場の活動も含めて高く評価する。論文執筆に当たった生徒は勿論、その指導に当たられた先生方に深甚の敬意を表する。

大学生・一般の部

大学生・一般の部の上位は、昨年度ほど機械学習系の論文が多くを占めるという状況とはならなかった。指導教員の有無によって、高校生の論文に比べて記述や構成にばらつきがみられたが、それ以上に仮説探索的研究と仮説検証的研究との2極化は進んでいる。もちろん、このどちらが良いといった判断はできない。テーマの多様性も好ましく、大臣賞論文のようにパンデミックに対して膨大な分析を試み、かつオリジナリティの高い部分もある力作もあった。少し意外だったのはパネルデータ分析のしっかりした実証研究が複数みられたことである。率直に申し上げて、これらを審査の上で差をつけるのには苦勞した。分析技術で差別化できないといった中で政策的提言の価値の評点が高かった論文が上位に来たという事情も否めない。

大学における数理・データサイエンス・AI教育の進展に伴い、高校生に比べて統計実証分析あるいは探索的データ分析の技術水準は高くなっていると共に、安心してみていただけるものとなっている。逆に、高校生が行っているモデル当てはめの診断（回帰診断、外れ値の考察）や試行錯誤などはあまりなくて、論文らしい論文が多くなってきた。贅沢を言えば、高校生のような個別データに対するこだわりのようなことが、平均的にはもう少しあっても良いのではないかと考える。