

再現可能性ははたして必要なのか：p 値問題から垣間見る科学研究の多様性

三中 信宏

国立研究開発法人 農業・食品産業技術総合研究機構 農業環境変動研究センター

統計分析における「p 値 (p -value)」が看過できない“誤用”を犯しているとのアメリカ統計学会の声明 (Wasserstein and Lazar 2016) は、研究分野の壁を越えて広く科学者コミュニティに反響をもたらしました (たとえば心理学の分野では友永他 2016)。長年にわたって主として農業実験分野での統計分析の現場に接する機会が多かった私個人の経験を振り返ると、確かにその声明に指摘されているような統計的データ解析の手法や基準の“誤用”はそれを意図するかしないかに関係なくさまざまな場面で数多く見られます。実験観察の目的が「5%レベル有意性の星」あるいは「p 値の小数点以下の 0 の個数」のみにあるとき、研究分野を問わず、さまざまな“不正”の手口が編み出されるのは不思議ではないでしょう。

統計的仮説検定に目を向けると、古典的な仮説検定の方法論は時代によって移り変わりがありました。たとえば、フィッシャーは対立仮説を設定せずに帰無仮説を検定しようとしたのですが、ネイマン-ピアソンは帰無仮説に対置する対立仮説を仮定したという根本的なちがひがあります。フィッシャーはデータの一意性を強調したのに対し、ネイマン-ピアソンは母集団からの無限回抽出を仮定した — 両者の溝は想像以上に深かったわけですが、80 年後の現在、その論争は未解決のまま放り出されていることはほとんど忘れ去られているようです。ネイマン-ピアソンの意思決定パラダイムが帰無仮説と対立仮説の命運を分ける絶対的な基準を置くのに対し、フィッシャーの尤度パラダイムは仮説間の証拠 (すなわち尤度) による相対的な重みづけをするだけで、仮説の受容や棄却の意思決定を伴いません。ここでの対立点は、統計的推論の科学哲学に大きく踏み込んでいます。

統計的データ解析は推論様式としての「アブダクション」のためのツールとみなすことができます。アブダクションは、データを説明するために立てられた仮説の“真偽”を問いません。むしろ、同一のデータを説明しようと競合する複数の対立仮説の間で、データを証拠とする相対的な“支持”の順位を踏まえ、その時点でもっともよい仮説を選び出すことを目標とします。つまり、対立する他の仮説とのデータ=証拠に基づく相対的比較が決定的であるということです。

科学という営為はけっして一枚岩ではありません。一方には、仮説の真偽が実験によって白黒をつけることができる実験系の科学もあります。他方には、歴史学や進化学のように、直接的な観察や実験がまったくできない歴史叙述科学もあります。実験科学ならば綿密な実験計画のもとに再現可能な結論を得ることはきっと可能でしょう。しかし、実験科学ではないタイプの科学については実験的な研究方法がもともと適用できないこともありえるでしょう。そのような場合でも、状況証拠に基づくアブダクションによってベストの仮説をそのつど選び出していくという道が残されています。統計学はそれぞれの科学のもつ性格や特性に応じてさまざまな形をとり得ると考えれば未来が拓けるのではないのでしょうか。

文献リスト

- 友永雅己・三浦麻子・針生悦子編 (2016). 特集〈心理学の再現可能性：我々はどこから来たのか 我々は何者か 我々はどこへ行くのか〉 心理学評論, 59(1): 1-141. <http://team1mile.com/sjpr59-1/>
- Wasserstein, R. L. and Lazar, N. A. (2016). The ASA's statement on p -values: Context, process, and purpose. *The American Statistician*, 70(2): 129-133 [佐藤俊哉訳 2017. 統計的有意性と P 値に関する ASA 声明. <http://biometrics.gr.jp/news/all/ASA.pdf>]