

高等学校の探究学習における推測統計に関する一考察 A Brief Study on Statistics Education in Inquiry Based Learning in High School

山口 恭正[†]
Yasumasa Yamaguchi

1. はじめに

現在、STEM・STEAM 教育をはじめとした科学技術教育や、中等教育における探究活動も盛んにおこなわれている。特に、高等学校における探究学習は科学技術教育の中でもとりわけ大学での研究活動と強い結びつきを持つ。特に、2002 年から始まったスーパーサイエンスハイスクール (SSH) 制度は現在も続く息の長い科学技術政策の一つである[1]。SSH 指定校のみならず、大学進学への進学先の選択に、高校時代の探究学習は大きな影響を与えているケースも見受けられる。近年では、高校生による探究学習による研究活動は、高大連携の努力もあり非常にレベルの高いものになり、理系のみならず人文系でも統計解析を行った社会科学の文脈からの量的な研究も行われている。

近年の情報科学や科学技術の発展に伴い、科学的妥当性と実証性が大学だけでなく高等学校でも求められるようになり[2]、それらを担保する学問として研究スキルとして統計学が挙げられる。例えば、SSH 指定校の成果発表に際して、外部審査員 (近隣の大学の研究者や OB であるケースが見られる) が高校生の研究成果に対して、「その結果は有意だったのですか？」と問う場面も散見される。

しかし、一方で統計学の使用には学会を中心に警笛が鳴らされている。とくに仮説検定の重要な指標とされる p 値の誤用や p 値ハッキング[3]と呼ばれるような行為が問題視されるようになった。科学的研究の重要なインフラとも言える統計学に関するこうした議論は、今後の学術研究のあり方に一定の影響を与えると考えられ、それは、高等学校における探究学習の文脈でも同様である[2]。

本論文では、そうした人文系を含めた探究活動における有意検定、すなわち推測統計に関して、高校生にどのようにそれを教授し実践させるべきかについて簡単な議論を行う。

2. 仮説検定に関する基礎情報

科学的研究では基本的に理論上、観測されたデータからは結果を直接的には議論せずに、仮説検定に基づく統計的推論を行いその結果を論じる。こうした、仮説検定を経て実際の物事がどうであるかを推測するのが推測統計や推計統計学と言われ、人文社会科学を含む多くの研究がこうしたプロセスを経て結果を議論している。

仮説検定においては、一般には棄却しうる仮説即ち帰無仮説を設定し、観測された値が帰無仮説の下では得られないような値になっているか、あるいは、帰無仮説の下でも観測される可能性が十分にあるような値であるかについて

の検証を行う。仮説検定に用いられる統計量を検定統計量と呼び、帰無仮説に整合的でないとされる検定統計量の範囲を、棄却域と呼ぶ。統計検定料の値が棄却域に入っているならば、帰無仮説は棄却され、観測データは統計的に有意と判断し結果を論じることができるようになる。

こうしたプロセスで検定の対象となるのは帰無仮説であるため、一般に仮説検定と称される。棄却域は 5%水準が設定され、帰無仮説のもので稀にしか生じない事象を定義する確率の値を有意水準と称する。また、データから結果がぎりぎり有意となるのは有意水準がいくらの時か、を示すのが p 値であり、限界水準や有意確立と呼ばれる。この値が有意水準よりも低ければ帰無仮説は棄却され、統計的な議論をする上での必要条件とされる「有意差あり」を示すこととなる[4]。

伝統的統計学ではこうした過程を経て、「帰無仮説が正しい場合に 5%以下の確率でしか生じない事象は『まれ』な事象で帰無仮説とは整合的でない。故に帰無仮説を棄却し対立する本来示したかった仮説 (対立仮説) を支持する」という論理で仮説検定を行っている[4]。しかし、有意水準は慣例で決められた値で理論的な根拠はなく、科学的研究はそれを盲目的に採用し統計的に有意だと「みなしている」に過ぎないとも言える。そんな仮説検定に対して近年警笛が鳴らされている。

3. 有意検定と ASA 声明

有意差があるということは、統計的な分析を行うための必要条件とされ、研究活動においては、それが実験の結果でも、調査の結果でも有意確立が有意水準を下回らないと結果としては論じられないというのが、伝統的統計学のある種の作法であった。しかしながら、こうした状況はいつしか、悪質な研究不正や誤用を生み出すことになってしまったと言われている。

たとえば、 p 値が 0.05 を下回るまで実験や調査を続けたり、そうした結果が出るように標本を選択したり、 p 値そのものには分析上意味がないにもかかわらず、それが最も小さい値を示す結果を採択するというような、 p 値ハッキングと呼ばれる行為、そして統計的有意性と実質的有意性の混同といった、 p 値の概念そのものの誤解や誤用が挙げられる。そんな状況からアメリカ統計学会 (American Statistical Association) が 2016 年に「統計的有意性と p 値に関する声明 (The ASA's Statement on p -Values)」を発表した[5]。この ASA 声明は日本計量生物学会の佐藤が ASA の許可を得て日本語訳を公開している[6]。また、 p 値に代わるアプローチとして、信頼区間やベイズファクターなどの指標などが紹介されている。

4. 探究学習における統計学の利用

さて、高等学校の学習指導要領の改訂により 2022 年 4 月に高校一年生になった生徒から、「統計的な推測」が新課

[†] 仙台大学 Sendai University

程として必要となる[7]。こうした流れは欧米の STEM 教育の系統を汲むものとも考えられ、科学技術教育の国際競争の影響とも言える。

しかし、それに先立つ形で既に高校生の探究学習においても、とりわけ人文系のトピックを科学的に扱おうとする際にも、今日では統計的な分析が求められている現状がある。現実的な課題として、高校生が大規模な調査や実験を行うのは非常に難しく、往々にして学内の学生や教員、あるいは保護者等や地域住民を対象にアンケートや実験等を行うこととなる。こうした際に、果たして仮説検定に必要なサンプル数が集められるかという点は、留意すべき事項の一つだろう。とりわけ、成果発表会等で外部の審査員や大学の教員等からは検定に関して指摘されるケースが見られるが、果たしてその指摘は、高校生の探究学習の範囲における研究規模で議論すべきか、一考の余地があるのではないだろうか。

また、統計的検定は先の 2 章で述べたようなプロセスをとるが、それが果たして高校生に理解できるのかという点も非常に重要な問題である。新課程において仮説検定が扱われるとは言え、実際に実験データや調査データを収集し分析をするフェーズにおいて、記述統計的な分析ではなく、帰無仮説を設定して検定を行うという、まわりくどいひと手間を限られた時間で理解し、活用できるのか、それを指導できるのかも留意すべきだろう。とくに、先に述べたような p 値が小さい方が良いというような表面的な理解にとどまってしまうケースや、帰無仮説と対立仮説の関係性を誤解してしまうケース、少し高度にはなるが統計的な有意性と実質的な有意性の誤解してしまうケースが予測される。

そもそも、高校生を含めて研究者たちが学術研究において用いている検定が果たして正しいのかという点も留意すべきだろう。例えば、2 群の平均比較をする際には t 検定が使われることが人文系の研究でも非常に多い。しかしながら、本来の t 検定とは正規性や等分散性といった前提の下で執り行われる検定であり、どんな場合でも t 検定が良いというものではない。また、人間社会系のテーマで探究活動を行う高校生が使用するような（もちろん学術研究でも使用される）、心理学等の尺度やリッカート尺度を用いたデータの解析は、既存の学術論文においても順序変数を連続変数であるとみなした分析を行うケースが散見される。しかし、こうした分析は、研究者によっては t 検定やピアソンの相関係数のようなパラメトリック系の分析ではなく、Wilcoxon の符号順位検定や Spearman の順位相関係数のようなノンパラメトリック検定を用いるべきだという主張もある。これは、尺度の項目間の幅が必ずしも一定とは限らないという考え方に基づく流派であるが、統計学を教えるという行為に置いて、どこまでの柔軟さを許容するかという議論は必要であろう。

現在、学術研究全般において研究インフラとしての統計学への見直しが叫ばれている。こうした中で、教科としての統計学ではなく、データ分析における実践的な統計学の利活用について、どの程度の精度やどのようなスタンスでデータに向き合うのかは、学術界を挙げて考えていく必要があるだろう。

5. さいごに

高校生の探究学習と推測統計にまつわる諸問題は、非常に複合的である。統計分析のエンドユーザーとしての活用を教授するにも、近年の統計学にまつわる議論から p 値ハッキングや望ましくない分析を誘発する恐れがある。しかし、理論的な事柄に関しては、探究学習という限られた時間でそれを教授するのは困難である上に、今後の学術研究において統計学をどのように扱うのか、答えは出ていない。たとえば、新しい統計分析の潮流の一つとしては効果量を明記した結果の検証が提唱されている。この効果量を絡めたアプローチは現在、混乱が見られる統計分析の中の一つの落としどころとして頻繁に利用されるが、高校生が抱えるサンプル数や仮説検定の難解さといった、統計学を活用するための障壁を根本的に打開できるものとは言い難い。

一方で、効果量とは別にベイズ統計を活用するという流れも提唱されている。これは従来の仮説検定ではなく、ベイズファクター (BF) によって支持する仮説を選択するというものだ。現在は JASP というソフトによって他のソフトウェアと同様の感覚でベイズ統計を分析に用いることができるようになっており、選択肢の一つとして考えられるだろう。しかし、ベイズ統計に関しても学術研究でどのように用いるかに関するセオリーが確立されていない現状がある。

科学技術と情報科学に支えられた現代社会において、研究スキルとしての統計学の活用は分野を問わず非常に重要であることは間違いない。しかしながら、その利活用を取り巻く現象は、必ずしも十分に整備されているとは言えない。探究学習においてどこまで統計的検定に重きを置くのか？統計学の原則にどこまで則した分析を行うべきなのか？という事柄は、一つのガイドラインを定めるべきなのかもしれない。大学での例えば心理学に関する統計教育に関しては、既に議論がされている[8]。しかし、統計学を用いた研究活動に関しては、現場の教員だけではどうしても難しい部分が出てくる。したがって、大学や研究機関で、実際に分析を行っている人間が有識者として各学校にアドバイスや出前授業といった形で、その活用方法を教授する必要があるだろう。

参考文献

- [1] 国立研究開発法人科学技術振興機構, “スーパーサイエンスハイスクール”, <https://www.jst.go.jp/cpse/ssh/> (2023.6.12 最終閲覧)
- [2] 村岡一幸, “探究活動基礎における統計学的手法の導入” 生物教育, 第 64 巻, 第 1 号 (2022)
- [3] Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD, “The Extent and Consequences of P-Hacking in Science.”, *PLoS Biol* 13(3): e1002106. <https://doi.org/10.1371/journal.pbio.1002106> (2015)
- [4] 南風原朝和, “心理統計学の基礎—総合的理解のために”, 2002
- [5] Wasserstein, Ronald L., and Lazar, Nicole A, “The ASA’s Statement on p -Values: Context, Process, and Purpose.” In *The American Statistician*, Vol.70, Issue2: (2016)
- [6] 佐藤俊哉, “統計的有意性と P 値に関する ASA 声明”, 日本計量生物学会ウェブサイト, <http://biometrics.gr.jp/news/all/ASA.pdf> (2017) (最終閲覧 2023.6.12)
- [7] 文部科学省, “高等学校学習指導要領 (平成 30 年告示)”, <https://www.mext.go.jp/content/> (2018) (最終閲覧 2023.6.12)
- [8] 堀裕亮, “心理学を専攻する学生への統計教育はどうあるべきか”, *心理学評論*, 60(3), (2017)