

高校数学「仮説検定の考え方」のためのフィッシャーの反証主義 —条件付き確率の考え方との対比—

上ヶ谷 友佑

高等学校数学 I において「仮説検定の考え方」が扱われるようになり、2025 年 1 月、大学入学共通テストにおいても初めてこの内容が出題されることとなった。しかしながら、そこでの取り扱いは、統計学者 R. A. フィッシャーの反証主義的な実験計画法の趣旨に反しており、統計教育の在り方として不十分であるように思われる。そこで本稿では、「仮説検定の考え方」のあるべき取り扱いについての考察と、適切な教材の開発を目的とした考察を展開する。その結果、データを得る前の実験計画法を重視し、反証したい事柄の存在をストーリーの起点とするような文章題が、フィッシャーの仮説検定の考え方を指導する教材として必要であることを指摘した。また、具体的な教材案として、条件付き確率の考え方と対比して仮説検定の考え方の理解を深める教材を提案した。

Fisher's falsificationism for hypothesis testing in high school mathematics: Contrasting with conditional probability

Yusuke Uegatani

In the Japanese high school *Mathematics I* curriculum, hypothesis testing appeared for the first time in the January 2025 National Center Test for University Admissions. However, its current treatment seems insufficient for effective statistical education, as it contradicts the core of R. A. Fisher's falsificationist approach to experimental design. This paper examines how hypothesis testing should be properly incorporated and considers the development of suitable teaching materials. It emphasizes the importance of planning experiments before data collection and framing word problems by starting with the presence of what one seeks to refute. The paper also proposes teaching materials that contrast conditional probability with hypothesis testing to enhance students' understanding.

1. はじめに

高等学校の数学 I に「仮説検定の考え方」が導入されて数年が経った。また今年は、新課程で初めての大学入学共通テストが実施され、仮説検定の考え方を問う問題がどのように出題されるのかにも注目が集まった。しかしながら、教科書や共通テストにおける現状の「仮説検定の考え方」の扱い方は、推測統計を確立した統計学者 R. A. フィッシャー (1971/2013) の反証主義的な実験計画法の趣旨に反しており、統計教育の在り方として不十分であるように思われる。

そこで本稿では、現状の取り扱いに対する批判的考察と、現状の取り扱いの弱点をカバーするための教材を提案したい。この教材は、フィッシャーの仮説検定の考え方の根底にある反証主義的な実験計画法のエッセンスを反映し、かつ、仮説検定の考え方につきまとう誤解を払拭するために条件付き確率の考え方との対比を行う教材

である。日本では、「仮説検定の考え方」は必修科目の数学 I の学習内容だが、「確率」は選択科目の数学 A の学習内容であり、仮説検定の考え方を学習するすべての生徒が、高等学校水準での確率を学習しているとは限らない。つまり、彼らは中学校水準での確率しか学ばずに仮説検定の考え方を学んでいる可能性がある。しかし、仮説検定の考え方の指導については、数学 I と数学 A を並行履修している高等学校が比較的多いことに基づいて、数学 A で学習する確率の内容と関連させて実施する案も提案されている (小林, 2022)。本稿でもこうした先行研究と軌を一にし、数学 A で学習する条件付き確率と関連させた「仮説検定の考え方」の指導を企図する。

2. 日本の高等学校で指導する「仮説検定の考え方」とその問題点

学習指導要領解説 (文部科学省, 2018) には、次の問題

場面の例が挙げられている。

例えば、「ある新素材の枕を使用した 30 人のうちの 80%にあたる 24 人が以前よりよく眠れたと回答した」という結果に対して、新素材の枕を使用するとよく眠ることができるかと判断できるか、という問題に取り組ませることを考える。この問題を解決するために、この結果が偶然に起こり得た可能性はどのくらいあるのかを、コイン等を使った実験を多数回繰り返して考察する。つまり、以前よりよく眠れた場合とそうでない場合が起こる可能性が半々だとしたとき、24 人以上がよく眠れたと回答することがどの程度起こるかを考える。実験として、コインが表になった場合を「以前よりよく眠れた場合」とし、コインを 30 回投げるといって試行を繰り返す。実験結果を表やグラフなどに整理し、24 枚以上表になった回数の相対度数 p を「起こり得ないこと」の尺度として用いることで、「30 人中 24 人以上がよく眠れたと回答することが、無作為性（ランダムネス）だけで説明できる可能性は p しかないように思われる。」という、判断の根拠が得られたことになる。(p. 49)

学習指導要領解説には、これ以上の具体的な記述はない。仮説検定の細かい問題に触れずに済ませているという点で実に巧みな記述であるように思う。一方、実際にどのように仮説検定の考え方を扱うかについては、教科書会社ごとにばらつきがある。例えば大谷・石橋(2022)は、教科書会社間で仮説検定の考え方を扱う際の確率をどのような意味で用いているかにばらつきがある点を指摘している。そこで本稿では、教科書会社ごとの違いに言及しそれらを吟味するというよりは、仮説検定の考え方に関する共通の基盤として、令和 7 年度の大学入学共通テストの出題に着目したい。なぜなら、大学入学共通テストは、高等学校教育の成果として身に付けた、知識・技能や思考力・判断力・表現力等を問う問題が出題されると謳われており、どの教科書で学んだとしても身に付けることが期待されている力を問う問題が出題されていると考えられるからである。

令和 7 年度の大学入学共通テスト「数学 I、数学 A」における仮説検定の考え方に関する問題は、次の通りであった。

太郎さんが住む地域では、その地域に宿泊を促すためのキャンペーンとして、キャンペーン A、B が実施されている。

太郎さんは、キャンペーン A の方がよいと思っている人が多いといううわさを聞いた。このうわさの

とおり、キャンペーン A の方がよいと思っている人が多いといえるかどうかを確かめることにした。そこで、かたよりなく選んだ人たちに、キャンペーン A、B のどちらがよいかについて、二択のアンケートを行ったところ、アンケートに回答した 35 人のうち、23 人が「キャンペーン A の方がよい」と答えた。この結果から、一般にキャンペーン A の方がよいと思っている人が多いといえるかどうかを、次の方針で考えることにした。

方針

- “「キャンペーン A の方がよい」と回答する割合と「キャンペーン B の方がよい」と回答する割合は等しい”という仮説を立てる。
- この仮説のもとで、かたよりなく選ばれた 35 人のうち 23 人以上が「キャンペーン A の方がよい」と回答する確率が 5%未満であれば、その仮説は誤っていると判断し、5%以上であればその仮説は誤っているとは判断しない。

〔中略〕したがって、今回のアンケート結果からは、キャンペーン A の方がよいと思っている人が〔ネ〕(大学入試センター, 2025, pp. 22–23)¹⁾

問題全体としては、上の引用の〔ネ〕の空欄に、「多いといえる」を埋めさせる問題であった。事実上、これが、日本の高等学校数学における「仮説検定の考え方」を必要とする問題の代表例と考えてよいであろう。

この例を吟味するにあたっては、次の 2 点に注意が必要である。1 つは、反復試行に基づく確率分布の理論値の取り扱いがない点である。これは、高等学校数学 B の内容であることと、反復試行を学習する数学 A が、数学 I と独立に学ばれている可能性がある点への配慮である。そのため、数学 I ではコイントスなどによる経験的確率に基づいて帰無仮説の棄却を試みる。実際、上記の問題においても、引用した箇所以降に、コイントスによる実験結果が添えられている。ただし、冒頭で小林(2022)を引用して論じたように、実態としては、仮説検定の考え方を習う頃には反復試行を学習し終えるようなカリキュラムの高校も多いと考えられ、共通テストでの出題は難しいにしても、授業としては、仮説検定の考え方と同時に理論値を扱うこともできるであろう。

より重要なことは、2 つ目の注意点であり、第二種過誤が取り扱われていない点である。数学 B においてでさえ、「〔統計的推測の〕内容については理論的な取り扱いに深入りせず、具体的な例を工夫したりコンピュータなど

の情報機器を用いるなどして確率分布の考えや統計的な推測の考えを理解できるようにする」(文部科学省, 2018, p. 107, 括弧内筆者)と述べられているように、数学Ⅰでは、理論的な深入りを一層避ける配慮が求められている。そして巧みな記述ではあるが、「5%以上であればその仮説は誤っているとは判断しない」という方針を立てており、「その仮説は正しい」という判断はしないことになっているので、第二種過誤が生じない工夫がなされている。これは、第二種過誤に関する理論的深入りを防ぐ方便であると同時に、フィッシャーの仮説検定の考え方を採用していることを意味している。実際、検定仮説(=帰無仮説)と対立仮説のどちらの蓋然性が高いかを考えるネイマン・ピアソンの仮説検定の考え方を採用するのであれば(土居, 2010; 松原, 1989 参照)、検出力を高めるための標本の大きさを検討しなければならないが、日本の高等学校数学ではこの扱いがない。

こうした指導系列上の制約を念頭に置くとき、数学Ⅰにおいて仮説検定の考え方として身に付けることが期待されていることは、データが与えられた際の判断力である、と要約できるであろう。単元全体としては、課題学習の一環として、データの収集から分析までを取り扱うことがもちろん重要ではあるが、仮説検定の考え方という一点においては、この判断力が肝であるといえよう。しかし、その判断力に力点を置いた取り扱いには、大きく3つの問題がある。

第一の問題は、上述のような仮説検定の考え方の取り扱いが、疑わしい研究行為に繋がる危険性を有している点である。まず上記の問題では、35人中23人がキャンペーンAの方がよいと回答したというデータに対してどう判断すべきかが問われていた。被験者を35人と設定したり、有意水準を5%と設定したりする段階は不問に付されている。しかも、23人がキャンペーンAの方がよいと回答したという事実が示された後に、分析の方針が示されている。しかし、これではあたかも、データを得た後に分析方針を決定してよいかのようである。そして、これは統計的実践として問題がある。なぜなら、それは、いわゆるHARKing(Hypothesizing After the Results are Knownの略称から取られた造語)と呼ばれる、疑わしい研究行為の一種(Kerr, 1998)になりかねないからである。23人がキャンペーンAの方がよいと回答したという結果を見てから、帰無仮説であれ、対立仮説であれ、仮説を設定してはならない。データを得る前に、仮説を設定すべきである。

第二の問題は、「太郎さんは、キャンペーンAの方がよいと思っている人が多いといううわさを聞いた」(大学入試センター, 2025, p. 22)という導入で始まる共通テス

トの出題が、フィッシャー(1971/2013)の実験計画法の趣旨に反している点である。有意性検定の科学基礎論的背景と社会的含意の関係を論じた松原(1989)は、帰無仮説が棄却されるということ、対立仮説が採択されると転義するのが、ネイマン・ピアソン流の考え方であり、フィッシャーの考え方とは異なる点を指摘する。そして、フィッシャーの本来の考え方と、ポパーの反証主義との関連性が議論している。実際、フィッシャー(1971/2013)は、帰無仮説と実験の関係性について次のように述べる。

帰無仮説は、実験の過程で否定されることはあっても、決して証明されたり確立されたりすることはない、という点に注意すべきである。あらゆる実験は、事実に対して帰無仮説を否定する機会を与えるためにのみ存在するといえることができる。(p. 14)

つまり、仮説を裏付けるというデータという考え方と距離を取り、常に既有的仮説を反証する形でのみ科学を前進させる態度を取る。この考え方があるからこそ、フィッシャーの仮説検定の考え方は、第二種過誤を検討しなくて済むのである。

この観点を踏まえると、共通テストの出題は、「太郎さんは、キャンペーンAの方がよいと思っている人もキャンペーンBの方がよいと思っている人もどちらも同じくらいらしいといううわさを聞いた」という導入で始まるべきである。そうでなければ、フィッシャーの仮説検定の考え方を使おうと思ふべきではない。しかも、噂通り、キャンペーンAの方がよいと回答した人が多いのであれば、何の反証にもなっていない。

第三の問題は、共通テストの出題が、フィッシャーの仮説検定を採用しているように見えながら、ネイマン・ピアソンの考え方を部分的に採用するハイブリッド的な考え方に陥っている点である。土居(2010)は、フィッシャーの考え方とネイマン・ピアソンの考え方の差異をわかりやすく端的に解説した貴重な日本語文献の1つである。曰く、フィッシャーの仮説検定は、1回の実験においてどのように判断を下すかという科学的な帰納的推論を考えた理論である。科学的探究という文脈を前提として、基本的には単一の結果から性急な結論を導くことに禁欲的である。それに対して、ネイマン・ピアソンの仮説検定は、規格化された工業製品の品質管理を主たる応用領域として、実験をするごとに一定の確率での判断の過誤を許容しながら、検出力に配慮したサンプルの大きさでの実験を繰り返し、全体として判断の確からしさを高める理論である。そして土居(2010)は、仮説検定実践の問題点を指摘した先行研究を概観しながら、現在の社会科学においては、両者の特徴の違いを無視した理論的妥当性

を欠いたハイブリッド仮説検定が(統計の教科書においても)普及してしまっているという問題を整理する。共通テストの出題は、見事にこのハイブリッド仮説検定に陥っている。フィッシャーの仮説検定の考え方を採用するのであれば、最後の「ネ」の空欄に、キャンペーン Aの方がよいと思っている人が「多いといえる」と埋めさせる出題をしてはいけない。あくまでも帰無仮説が棄却されるという結論で踏みとどまらなければならない。

以上より、共通テストの出題に顕著に現れているように、現状の仮説検定の考え方は、HARKingに繋がりがかねない考え方であり、フィッシャーの考え方とネイマン・ピアソンの考え方のハイブリッド的な考え方になってしまっている。科学的な研究実践の指導としても、統計学的な理論の指導としても、適切でない出題である。

3. 仮説検定の指導に関する先行研究とそれらに対する批判的検討

本節では、仮説検定の指導に関する先行研究を3つ取り上げ、それらの成果に依拠するだけでは、前節で指摘したような問題点に対処できない点を指摘する。

(1) 背理法との関連

仮説検定を背理法と類似した論理構造として指導することは、有望なアプローチとして古くから提案されている(例えば、Reeves & Brewer, 1980)。背理法は、 $\neg P$ を仮定すると、矛盾(すなわち、起こり得ないこと、あるいは、確率0の事象)が生じるから、 $\neg\neg P$ 、すなわち、 P と結論付けることが合理的であるとする論法である。一方、仮説検定は、帰無仮説 H_0 を仮定すると、有意水準 α より小さい確率 p でしか得られないデータ x_1, x_2, \dots, x_n が生じたことになることから、帰無仮説 H_0 の否定、す

なわち、対立仮説 H_1 と結論付けることが合理的であるとする論法である。矛盾の観察から仮定の否定を採択するのか、ほとんど起こらない事象の観察から仮説の否定を採択するのかという点で、確かに論法の構造が類似している。

一方、Otani (2019) は、両者の論証構造を対比し、不変性と変動性の違い、矛盾と低確率の違い、仮定によって結論が指示されるかどうかの違い、結論が必然的かどうかの違いの4つの違いを指摘する。仮説検定を確率論的な証明であると捉えてしまうミスコンセプションはかねてより指摘されており (Castro Sotos et al., 2007; Falk & Greenbaum, 1995)²⁾、背理法との構造的関連性を扱いながらも仮説検定に特有の変動性を意識させるための教材開発も行われている(福田ら, 2018)。しかしながら、これらの工夫によって生徒達が仮説検定特有の変動性を意識するようになったとしても、HARKingの危険性やフィッシャーの仮説検定とネイマン・ピアソンの仮説検定の差異を指導することには繋がらない。単に仮説検定が証明ほどには信頼できないということが伝わるのみである。

(2) 条件付き確率との関連

Falk and Greenbaum (1995) は、仮説検定を巡る誤解を防ぐために、次のように述べる。

混同のリスクを低減する1つの安全な方法は、話し言葉的なショートカットなしで済まし、条件付き確率の言語を支持することであろう。(p. 88)

「『キャンペーン Aの方がよい』と回答する割合と『キャンペーン Bの方がよい』と回答する割合は等しい」という事象を H_0 、偏りなく選ばれた35人のうち23人以上が「キャンペーン Aの方がよい」と回答する事象を R とするとき、有意水準 0.05 と比較される確率は、 $P_{H_0}(R)$ で

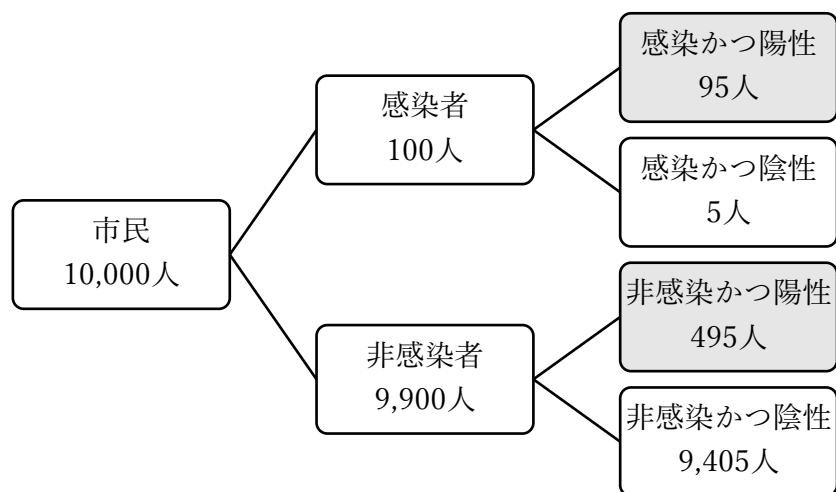


図1: 市民 10,000 人の判定結果

あって、 $P_R(H_0)$ ではない。確かに表現の正確性という観点では一考に値する。しかし、あくまでも「帰無仮説の正しい確率が有意水準未満である」という誤解が防げるだけで、仮説検定それ自身の意味を学ぶことへの寄与が大きいとは考えにくい。

また、Falk and Greenbaum (1995) は、仮説検定が確率論的証明にならないことを示す例として、検査の偽陽性の問題も挙げている。ここでは、Falk and Greenbaum (1995) が挙げた遺伝的カウンセリングの実例ではなく、数値を単純化した架空の感染症検査を例としてその要点を示そう。市民 10,000 人中、感染率が 10% の感染症について、感染者を正しく陽性判定する確率と非感染者を正しく陰性判定する確率がそれぞれ 95% の検査キットがある。この条件下で、この市民から無作為に一人選んで検査をするとき、陽性判定となるのに非感染者である確率は、感染かつ陽性の 95 人と非感染かつ陽性の 495 人の合計中に、非感染かつ陽性の 495 人が占める割合であるから、

$\frac{495}{95+495} \approx 0.839$ である (図 1 参照)。つまり、約 83.9% と大

部分が偽陽性である。ところが、この例に、仮説検定の考え方を重ねると、次のようになる。今、検査対象の市民 1 名について、帰無仮説を「この市民は感染していない」、対立仮説を「この市民は感染している」とし、この市民の検査結果が陽性であったとする。帰無仮説が正しいときに陽性判定となる確率は、検査キットの性能通り 5% しかないはずであるから、帰無仮説は棄却される。つまり、この市民が感染していないという仮説が誤りである (この市民は感染している) と判定されることになる。上で述べたように、大部分が偽陽性のはずであるから、この判断は不合理である。

条件付き確率の考え方と関連させて、こうした例を扱うことは、統計教育として一定の価値がある。しかし、仮説検定の考え方を初めて習う場面でこの例を同時に提示してしまつては、そもそも仮説検定の考え方の価値が理解できない恐れがある。扱ってもよいが、仮説検定の考え方がどんな場面で役に立つのかは、少なくとも同時に扱うべきである。また、このような条件付き確率との関連付けでは、HARKing の危険性やフィッシャーの仮説検定とネイマン・ピアソンの仮説検定の差異を指導することには繋がらない。もう一工夫、必要である。

(3) 情報科との関連

3 つ目の重要な先行研究は、情報科との関連である。大谷・古賀 (2024) は数学科と情報科の双方の教科書分析を通じて、数学科と情報科の統計教育が学問主義と実用主義で差別化されている点を示し、数学科におけるコンピュータ利用に関してジレンマを提起している。すなわち、

仮説検定の考え方を学ぶにあたって、数学科において学習指導要領解説で推奨されているようなコンピュータ利用を推進していくと、情報科で指導しているような実用主義的な統計教育に還元されてしまいかねないという危惧があるからである。数学科では、統計処理をコンピュータに任せてブラックボックスとして扱うのではなく、きちんと統計学を学問主義的な観点から指導しなければならないのだという。

この点は、極めて重要な指摘である。HARKing を犯したり、フィッシャーの仮説検定とネイマン・ピアソンの仮説検定を混同したりすることは、まさに理論を欠いた実践によって生じる問題であり、学問主義的な観点から統計の理論を学ばなければ防げない。しかし、現在の数学科の教科書に、これらに対処する工夫があるわけではない。学習指導要領の制約を直接受ける教科書では実現できない指導上の工夫が、個々の数学科教員に求められている。

3. 教材開発の条件

ここまでで検討した議論を踏まえて、教材開発の条件を明確化していこう。教材開発の条件は、共通テストの問題点から示唆される 3 条件と、先行研究の批判的検討から示唆される 3 条件とに大別される。

(1) 共通テストの問題点から示唆される 3 条件

共通テストの問題点から示唆される第一の条件は、HARKing 問題を回避するために、開発される教材は、データが既に得られている設定の問題にしていけないということである。これは、数学の文章題が、どのような時制で書かれており、読者に何を求める、どんなジャンルの文章なのかという議論 (Gerofsky, 1996; Ishibashi & Uegatani, 2022; Thomas & Gerofsky, 1997) とも関連する。もし仮説検定の考え方を指導する文章題のストーリーが、一定程度、そのまま生徒達の実生活・社会生活に役立つことを念頭に置くのであれば、データを取る前の段階から生徒達に考えさせなければならないであろう。実験計画の段階から生徒達に考えさせるカリキュラム構成上の役割を、課題学習にのみ負わせるのではなく、仮説検定の考え方を指導する文章題の中にも自然と位置付け、カリキュラム内のギャップを低減させるべきであろう。

第二の条件は、フィッシャーの仮説検定の考え方が反証主義的な考え方であるという点を正しく反映させることである。立証したい仮説があつて、仮説検定をするわけではない。反証したい仮説があつて、仮説検定をするのである。立証したい仮説があることと反証したい仮説があることを、同一視しないところが、フィッシャーの

仮説検定の考え方の核心部分である。理想的には、生徒達は、帰無仮説にも対立仮説にも中立的な態度が求められるところではあるが、それは検定を実施する際に事後的に中立的に振る舞うよう意識することであり、検定を実施しようと思ったその瞬間は、通常、何らかのバイアスのかかった動機がある。このバイアスの部分が、適正化されるべきである。

具体的には、きちんと帰無仮説と対立仮説の対立構造を、文章題のストーリー中に表現すべきである。先の共通テストの出題では、太郎さんが聞いた噂と、太郎さんが立証したいと思っている事柄が一致しており、対立構造が存在しない。太郎さんは、聞いた噂が正しくないのではないかと疑念を持ち、噂を反証しようと仮説検定を試みた、というストーリーになっているべきである。生徒達の解答が文章題の微妙な表現の違いに影響されることは、近年、注目が集まっている論点であり (Carotenuto et al., 2021; Ishibashi & Uegatani, 2022)、慎重な設計が求められる。

第三の条件は、きちんと「帰無仮説が棄却される」という結論で踏みとどまるべき点である。第二の条件が満たされていれば、この点は達成されやすい。そもそもの仮説検定の動機が反証であれば、欲張って自説の立証を高らかに宣言する必要はないからである。ただし、この第三の条件は、現状、実践的な推奨条件に留まることになるであろう。高等学校の教員としては、共通テストで出題歴がある以上、フィッシャーの仮説検定の考え方に沿っていないからと言って、帰無仮説の棄却と対立仮説の採択の同一視を指導しないわけにはいかないであろうからである。

(2) 先行研究の批判的検討から示唆される3条件

先行研究の批判的検討からも、3条件が示唆される。第一の条件は、背理法との関連ではなく、間接的アーギュメンテーションとの関連を扱うべきである。証明法としての背理法と発見法としての間接的アーギュメンテーションとの間には、構造的な差異があることが知られている (e.g., Antonini & Mariotti, 2008; Hakamata et al., 2023; 上ヶ谷・袴田・早田, 2021)。背理法は、偽であると目される仮定から演繹される命題と真であることが既知の命題が矛盾することを示さなければならない一方、間接的アーギュメンテーションは、何かを仮定したときに意図した数学の対象が構成できないことを述べるだけに留まるなど、背理法よりも「矛盾」の扱いが幅広い。Freudenthal (1973) が間接証明を発見法と呼んだ際は、数学教育研究において間接証明と間接的アーギュメンテーションの弁別が不十分であったが、現代的には、生徒達は発見法として自然と間接的アーギュメンテーションは構成し得

るが、背理法の証明を構成することには独特の困難があるとされるのである。

こうした背理法研究の成果を念頭に置くと、仮説検定の考え方を指導する上で関連性を重視すべきは、証明としての背理法ではなく、発見法としての間接的アーギュメンテーションである。真偽不明の言明があり、その真偽はすぐには明らかにならなさそうであるとなったならば、その言明の真偽を確かめるべくいろいろと検討を重ねる中で、その言明が真だとしたらどうか、偽だとしたらどうかを考えることは、知的探究として自然であり、基本でもある。間接的アーギュメンテーションは、反証したい言明が正しいとしたら、こんな言明が得られるとは考えられない、という論法であり、仮説検定は、反証したい言明が正しいとしたら、こんな実験結果が得られるとは思えない、という論法である。そして、この点を適切に指導するためには、HARKing への対処と合わせて、どのような実験結果であれば、どんな実験をするのかと、どんな基準で「こんな実験結果が得られるとは思えない」と見なすのかを、事前に決めることを強調する教材であるべきである。

第二の条件は、仮説検定の考え方の価値を明確化する形で条件付き確率との関連を扱うことである。Falk and Greenbaum (1995) が挙げたような、仮説検定の考え方が偽陽性を生んでしまう問題は、もちろん仮説検定の考え方の限界として重要ではある。しかしながら、生徒達に偽陽性の例をそのまま伝えることは、単に仮説検定の考え方の頼りなさを知らしめるだけである。生徒達の中から見れば、条件付き確率をちゃんと使えばいいではないか、となりかねない。したがって、条件付き確率との関連を扱うにあたって考えるべきことは、なぜ条件付き確率を用いて偽陽性の確率を計算するのではなく、わざわざ仮説検定の考え方に立脚するのか、という疑問に答え得る教材を開発することである。

このように考えたとき、偽陽性の問題を教材化していく上で着目すべきは、情報量の差である。偽陽性の問題は、「市民の感染率が10%の感染症」という情報がついていからこそ、陽性判定が偽陽性である確率が計算できるのである。仮説検定の考え方は、そうではない。判定が出る前の実験計画の段階で、情報がほとんどない状況下で採用する考え方である。したがって、仮説検定の考え方が有用な場面と条件付き確率の考え方が有用な場面とを対比的に扱うことが重要となる。

第三の条件は、情報科との関連を考えて、きちんと学問主義的な観点を重視した教材を開発することである。実用主義的に、得られたデータを解析するというのではなくて、どういうデータを得ようとするのかの実験計

画の部分に力点を置かなければ、学問主義的とは言えない。Devlin (2014) が、確率は事象に対して適用されるのではなく、事象の情報に対して適用されるのだと述べたように、確率論を背景に持つ統計学は、どんな情報が得られているときに用いる考え方なのかを明確化するべきで、そこまでを含めて統計学の理論だと見なすべきである。数学教育も、仮説検定の考え方が、どんな情報が得られているときに用いる考え方なのかを扱うべきである。こうした検討は、学校数学の内容としての「数学的方法学」(上ヶ谷・石橋・服部, 2021) を考える上でも重要な議論である。

(3) 教材の条件のまとめ

本節で議論した教材の条件を踏まえ、次節では、少なくとも以下の 6 条件を満たすような文章題を、教材として考案する。

- [C1-1] データが既に得られているような場面の文章題にしない。
- [C1-2] フィッシャーの仮説検定の考え方に沿って、反証したい事柄の存在をストーリーの起点にし、検定実施者との対立構造を明示化する。
- [C1-3] 【参考】帰無仮説が棄却されることを重視し、対立仮説の採択まで無理に進まない。
- [C2-1] 仮説検定の考え方は、間接的アーギュメンテーションとの関連を扱う程度で十分である。

る。

- [C2-2] 仮説検定の考え方の価値がわかるような形で条件付き確率の考え方を対比的に扱う。
- [C2-3] 学問主義的観点から、仮説検定の考え方がどんな情報が得られているときに用いる考え方なのかを扱う。

次節以降での便宜のため、ここでは [C1-1] ～ [C2-3] の記号を振った。ただし、条件 [C1-3] は、上でも述べていたように、共通テストの出題実績を踏まえると、必ず満たすべきとは言えない参考条件である。

4. 考案した教材

前節を踏まえて考案した教材「公称値の信用できないガチャ」が、図 2 の問題である。表 1・表 2 をあわせて与えることを想定した問題である。この教材は、生徒達が、数学 A で条件付き確率を、特に原因の確率を既に学習していることを前提とした教材である。以下では、前節で議論した教材の条件をどのように満たしているかを議論しよう。

問 1 は、条件 [C1-1] を踏まえて、データが得られる前の段階で判断基準をどうするかを問うている。表 1 より、レアアイテムの排出確率が $\frac{1}{3}$ のとき、レアアイテムの排

【問 1】

A さんは、運営会社によってガチャでのレアアイテムの排出確率が $\frac{1}{3}$ と謳われているスマホゲームを遊んでいる。しかし、実際の排出確率がこの公称値よりも低いのではないかと怪しんだ A さんは、排出確率が $\frac{1}{3}$ であるという公称値の正しさを仮説検定の考え方で検証することにした。ガチャを回してレアアイテムが出たかどうかを記録するという実験を 20 回行い、公称値が正しいという仮説が正しいときに 5%未満でしか起こらないような結果となった場合は、この仮説が正しくないと考えることにする。20 回中、レアアイテムの出た回数が何回以下であれば、この仮説が正しくないと考えることになるか？

【問 2】

A さんは、ガチャでのレアアイテムの排出確率が $\frac{1}{3}$ であるという B さんが開発したゲームを遊んでいる。実際の排出確率が B さんのいう値よりも低いのではないかと怪しんだ A さんが B さんを問い詰めたところ、B さんが言うには、 $\frac{1}{2}$ の確率で数値設定を誤っており、誤っていた場合には、 $\frac{1}{6}$ の確率でしかレアアイテムが排出されないという。A さんは、仮説検定の考え方をを使うつもりがないまま、ガチャを回してレアアイテムが出たかどうかを記録するということを 20 回行っており、レアアイテムが出たのは 20 回中 2 回だけであった。このことから言えることは何か？

図 2: 開発した教材「公称値の信用できないガチャ」

表 1: $\frac{1}{3}$ で起こる事象が 20 回中 n 回生じる確率のおおよその値

回数 n	0	1	2	3	4	5	6	7	8	9	10
確率	0.000	0.003	0.014	0.043	0.091	0.146	0.182	0.182	0.148	0.099	0.054
回数 n	10	11	12	13	14	15	16	17	18	19	20
確率	0.054	0.025	0.009	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000

表 2: $\frac{1}{6}$ で起こる事象が 20 回中 n 回生じる確率のおおよその値

回数 n	0	1	2	3	4	5	6	7	8	9	10
確率	0.026	0.104	0.198	0.238	0.202	0.129	0.065	0.026	0.008	0.002	0.000
回数 n	10	11	12	13	14	15	16	17	18	19	20
確率	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

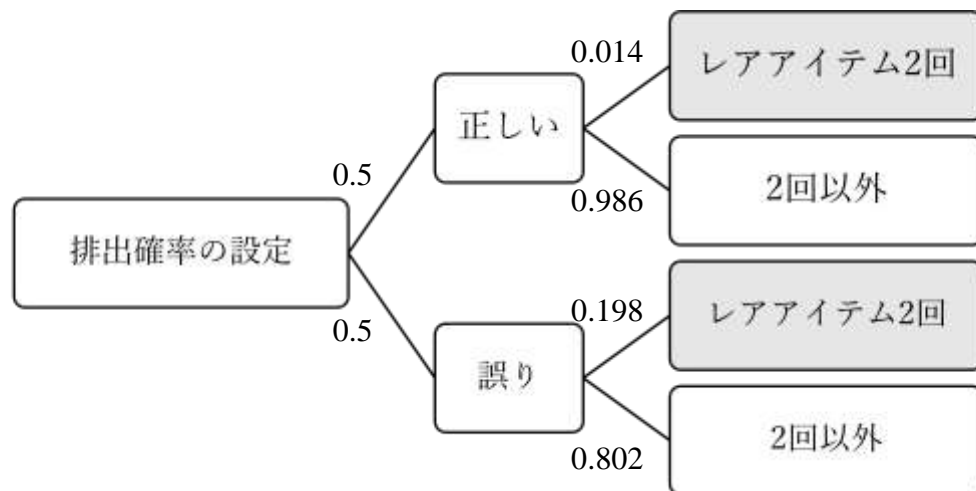


図 3: 問 2 のガチャを引いた結果

出が 20 回中 2 回以下である確率は約 0.017, 3 回以下である確率は約 0.060 であるから, 有意水準を 5% と考えれば, 問 1 の答えは「2 回以下」となる。「レアアイテムが 2 回出た」というような, 実験結果がわかった後の設定にしないことが重要である。

問 1 は, 条件 [C1-2] を踏まえて, 公称値が疑わしいということをストーリーの起点とした。検定実施者は, 反証したい仮説があることで仮説検定の考え方へと動機付けられており, ここに運営会社と A さんの間の対立構造

が見られる。

また, 条件 [C1-3] を踏まえて, 必ずしも対立仮説の採択にまで踏み込まなくても問題解決となる設定とした。20 回中 2 回以下しかレアアイテムが排出しなかったとしても, 運営会社は, 「その結果は偶然であり, 追試が必要である」と反論する余地がまだ残されている。問 1 は, 運営会社と A さんとの対立構造が顕在化されているため, この対立構造を意識させることで, 仮説検定の考え方を通じて断定的な結論を出してはいけないことを学ぶ

機会にすることができる。

問 1 は、条件 [C2-1] を踏まえて、証明法としての背理法ではなく、発見法としての間接的アーギュメンテーションに類似した構造の出題としている。文章題中でデータは得られていないため、背理法の矛盾に相当する「低確率な事象」はそもそも生じていない。単に疑わしいから実験して検証しようという設定を与えているだけであり、あくまでも発見法として、公称値の正しさを判断するための 1 つの材料を得ようとしているだけである。フィッシャーの仮説検定の考え方に従って、決定的な結論を出そうとせず、禁欲的な態度で挑む設定でもある。

問 2 は、条件 [C2-2] と [C2-3] を踏まえて設定した問いである。問 2 は、「仮説検定の考え方をを使うつもりがないまま」、「20 回中 2 回」という結果を既に得てしまった。HARKing を避けるため、データを得始めた当初の段階で仮説検定の考え方をを使うつもりがなかったのであれば、仮説検定を使うべきではない。そうでなければ、偶然有意差が出るまでデータを取り続けてよいことになってしまい、研究実践として不適切である。数学教育で扱う文章題も、仮説検定で検証すると宣言してから、検証を始めるという設定になっているべきである。

では、何を使うべきかといえば、条件付き確率である。問題設定が問 1 とよく似ているため、生徒達には見抜くことが難しいかもしれないが、原因の確率の問題である。そう、数学 A で学んだことを適切に活かすならば、我々は、結果を見てしまったならば、その結果が持つ情報量を加味した条件付き確率の計算によって意思決定をすべきなのである。数学 A では、そのように教えているはずであるのに、仮説検定の考え方を教える際にそのことが生きていないようでは、数学教育全体としては適切ではない。

問 2 では、レアアイテムが 20 回中 2 回出たということが確定しているのだから、図 3 を踏まえて、次の要領で、レアアイテムが 2 回出たときに排出確率が正しい条件付き確率を求めることができる。すなわち、求める確率は、排出確率が正しく、かつ、レアアイテムが 2 回出る確率 0.5×0.014 と、排出確率が誤っており、かつ、レアアイテムが 2 回出る確率 0.5×0.198 の合計の内、前者が占める割合であるから、 $\frac{0.5 \times 0.014}{0.5 \times 0.014 + 0.5 \times 0.198} \approx 0.066$ である。したがって、問 2 の設定において、排出確率が正しい可能性はわずか 6.6% であり、B さんは 93.4% の確率で排出確率の設定を誤っているといえる。

問 2 において、「このことから言えることは何か?」という出題にしているのは、データを既に得てしまった場面で何が主張可能かというのは、生徒達が将来直面し

得る場面であるからである。この問題場面においては、仮説検定の考え方をを使うのではなく、条件付き確率の考え方をを使うのであると指導したい。実験結果を見てからでは、仮説検定の考え方が活かない。

もちろん、実験結果を見てしまったとき、常に条件付き確率の考え方が使えるわけではないこともまた、同時に問 2 で指導したい。問 2 で、仮説検定の考え方よりも強力な形で条件付き確率によって B さんの誤りを判定できている理由は、B さんが設定を誤っている確率がいくらかかと、誤った際にどんな排出確率になるのかが明確だからである。通常、このような情報が得られている場面は稀であるから、リアリティの低い文脈を与えてしまっているという批判もあるかもしれないが、逆である。この文脈はリアリティが低いいため、仮説検定の考え方をきちんと使えるように、事前に実験計画をしっかりとしなければならぬということを、この問題解決を通じて指導するのである。リアリティが低いから学べないということはない (Ishibashi & Uegatani, 2022)。リアリティが低いからこそ学べることはあるはずである。

5. 考察

本節では、前節で提案した教材の意義をより一般的な視座から検討する。

(1) 提案された教材の発展可能性

本稿が提案した教材の条件は、様々な発展可能性を秘めている。特に、フィッシャーの仮説検定の考え方の動機を明確化している点は重要である。例えば、小林 (2022) は、仮説検定の考え方が生徒達から生み出されるためには、その仕掛けとして、データから得られた素朴な見解に疑いをかける「指導教員の問い」が必要であると指摘している。しかし、それもそのはずである。フィッシャーの仮説検定の考え方は、そもそも、データが得られた後に使いたいと思う考え方ではないからである。反証主義を念頭に置くならば、反証したい真偽不明の言明に出会ったその瞬間こそが、フィッシャーの仮説検定の考え方をを使いたくなるべき瞬間である。ある仮説を支持する者と、その仮説を反証したい者の両方がいる状況下で、帰無仮説の支持が合理的でないことを客観的に示す術が、フィッシャーの仮説検定なのである。

この方針で、本稿が提案した問 1 にさらなるストーリー性を付与するとすれば、例えば次のようになる。まず、運営会社はガチャでのレアアイテムの排出確率が $\frac{1}{3}$ だ

と主張している、という実態がある。それに対して、A さんはその排出確率は間違っていると疑義を呈する。そし

で中立的な立場から判定を下そうとしてくれている C さんに向かって、次のように主張する。「もしレアアイテムの排出確率が $\frac{1}{3}$ だとしたら、20 回ガチャを回して 2 回以下しかレアアイテムが出ないなんてことは 5%未満でしか起こらない。これから 20 回引いてみせるから、2 回以下しかレアアイテムが出なければ、運営会社の公称値を信じることでいいだろう？」と。このとき、運営会社は C さんへ次のように言うかもしれない。「A さんの提案する 5%では、偶然起こってしまうかもしれない！」と。これは、あり得る交渉である。そこで運営会社、A さん、C さんで協議して、有意水準を 1%に下げたり、実験回数を増やしたりすることを検討する。本稿が提案した教材は、数学 I の範囲で仮説検定の理論に深入りしないという制約で提案しているが、学習指導要領解説が課す制約を考えないのであれば、本来的には、この三者協議でどのような折り合いをつけるかを学ぶことが、仮説検定の考え方の学習では重要であろう。

(2) ハイブリッド仮説検定の問題点

フィッシャーの仮説検定の考え方を指導する上で教師が意識しておかなければならないことは、反証主義を採用する以上、一度、公称値に疑惑を持たれてしまったら、公称値の正しさを証明する術はない、ということである。20 回ガチャを回して 3 回以上レアアイテムが出たとしても、今回は嫌疑不十分であったと結論されるだけである。そのため、有意水準を何%にするかや、サンプルの大きさをどれくらいにするかは、当事者達が納得する形ならば比較的自由に決めることができる。

この点は、ネイマン・ピアソンの仮説検定の考え方と決定的に異なる。ネイマン・ピアソンの仮説検定の考え方では、有意水準をいたずらに低い値にすると、第二種過誤のリスクが高まる。有意水準、サンプルの大きさ、検出力は相互に関係を持つため、何回実験できるかやどれくらいの期間でどれくらい信頼性の高い結論を出す必要があるのかも含めて、実験計画を総合的に検討する必要がある。

そういうわけで、フィッシャーの仮説検定の考え方を扱う以上は、徹底して反証主義を採用すべきである。中途半端なハイブリッド仮説検定は、その考え方自体に科学的合理性が乏しい。このことは、中途半端なハイブリッド仮説検定の考え方のよさが生徒達に納得されるような、当該の考え方の「知的必要性」(Harel, 2013) が高い自然な文脈の教材を構成することは困難であることを示唆する。問題解決過程において学習する概念の意義を実感させるような指導を企図するにあたっては、まずは教材研究の段階で統計学的に一貫性の取れた仮説検定の考

え方を採用しておかなければならない。

(3) 無計画にデータを取ってしまった場合について

図 2 の問 2 が示唆するように、無計画にデータを収集してしまった場合は、仮説検定の考え方の使用は適切ではない。しかし、問 2 のように、排出確率がどの程度の確率で誤った数値になっているかや誤っていた場合の排出確率がいくらかといった情報は、より現実的な文脈においては明らかでない可能性が高い情報であるとも考えられ、実世界においては、そう都合よく条件付き確率の考え方が適用できる場合ばかりではないであろう。したがって、無計画にデータを収集してしまった場合にどうすべきかは、どんな情報がどの程度明らかであるか、あるいは、ある種の確率が正確にわからないまでも高いと見積もれるのか低いと見積もれるのかなどに応じて、あくまでも個別の具体的な場面に依拠して検討するしかない。

そして統計教育として重要なことは、無計画にデータを収集してしまった場合は、多くの場合、確率論を根拠にした推測統計が適用できなくなるということを指導することであろう。確率論的な推測ができなくても、データの分析自体は記述統計的に可能であり、何もできなくなるわけではない。ただ、データは、それがどんなデータであるかだけでなく、どうやって得たデータかがわかっている方が、そのデータからわかることが多くなるというだけである。フィッシャーの仮説検定の考え方が実験計画法であるという点を念頭に置いて指導をするのであれば、問 1 を通じて事前の実験計画をきちんと立てることの重要性を、問 2 を通じて既知の情報を最大限活用して推測することの重要性を、指導するとよいであろう。

(4) 推論主義の視座から見た教材の可能性

近年、数学教育研究において、Brandom (1994, 2000) の推論主義に基づく研究が、これまでの数学教育研究にはない教育観・学習観をもたらしている (例えば、Bakker & Derry, 2011; Seidouvy & Schindler, 2020; Uegatani et al., 2023; Uegatani & Otani, 2021, 2023; 上ヶ谷・大谷, 2019)。特に、数学教育研究において数学の脱中心化を図るべきであると論じた Uegatani et al. (2025) の指摘は、本稿が提案した教材の可能性を一層現実的なものにすると考えられる。彼らの指摘によれば、生徒達が日常生活や社会生活において数学的概念を活用するようになるためには、数学的概念の方を指導するばかりでは不十分である。日常生活や社会生活の中にある日常的概念が、その数学的概念をいかに呼び起こすかが重要なのであって、数学教育はむしろ日常的概念の指導にこそ力を入れるべきであるかもしれないのである。

このような視座に立つとき、データが既に得られてい

る状態で仮説検定の考え方を適用する文章題は、教材として不適切である。一般的には、そうならないように指導すべきなのであって、生徒達に仮説検定の考え方を使おうと思ってほしい瞬間が適切に含まれるような文章題の構成になっていなければならない。そうでなければ、どんなによく学んだ生徒達であったとしても、習った通りのタイミングでしか仮説検定の考え方を使おうと思わないであろう。その意味で、[C1-1] ～ [C1-3] の条件は、特に重要な教材の条件であるといえるであろう。

(5) 統計教育は数学教育か？

統計学を数学とは異なる形で指導すべきであるという議論は古くから存在する (例えば, Cobb & Moore, 1997)。実際, 本稿でも議論してきたように, フィッシャーの仮説検定の考え方は, 判断の論理としてよりはむしろ, 反証主義的な実験計画法として指導されるべきである。そして, 本稿が提起した教材以上に, さらに実践的な統計学を指導するとなつたならば, ますます数学の範疇を逸脱した実験計画法としての指導が必要となる。

例えば, フィッシャー (1971/2013) は, 実験計画としてどのような無作為化を施すかについても議論している。具体的な実験場面はこうである。[A]「ミルクを先に入れて作ったミルクティー」と [B]「紅茶を先に入れて作ったミルクティー」を試飲して弁別することを考える。そして, 「ある婦人は, それらについて一切の弁別能力を有さず, [A] と [B], どちらのミルクティーを飲んだのかをデータラメに答えている」ということを反証する実験を計画する。このとき, 8 杯中 4 杯に [A] を, 残り 4 杯に [B] を割り当て, それら 8 杯を婦人に提供する順序のみを無作為化するのか, 8 杯各々に [A] と [B] を無作為に割り当てるのか, その違いを議論する。前者の場合, 婦人は 8 杯の中から [A] だと思ふカップを 4 つ選べばよいから, 可能な選択肢は ${}_8C_4 (=70)$ 通りである。後者の場合, 婦人は 1 杯ごとに [A] と [B] のどちらかを選ばないといけないから, 可能な選択肢は $2^8 (=256)$ 通りである。どちらも 8 杯を利用する設定ではあるが, 一見すると, 後者の方が, 本当は帰無仮説が正しいのに偶然帰無仮説を棄却してしまうリスクが低くなるように思われる。しかし, この点についてフィッシャー (1971/2013) は次のように述べる。

多くの型の実験において, ここに述べたような構造の変化〔後者の 256 通りの実験計画を採用すること〕は明らかに有益であろう。しかしながら, 精神物理的実験の特殊な要求に対しては, この利益は断念するほうが多分よいであろう。というのは, 全部の茶わんが同じ処理を受けることが時には起こって, 予

期されない事情が被実験者を惑わすばかりでなく, 比較によって判断するという実際上の利益を被実験者から奪うことになるからである。(pp. 19-20, [] 内は筆者による補筆)

これは, 実験の内容が実験の構造に与え得る影響を議論しており, 数学の範囲を明らかに逸脱した (しかし, それでいて, 統計学的には重要な) 議論である。

そのため, 経験的な (empirical) 要素を含むという意味では, 確かに統計教育は, 代数教育や幾何教育などの数学教育とは異なる科学教育である。しかしながら, 本稿が同時に提起したように, 仮説検定の考え方がいつ使えるのかについて理解を深めるためには, 条件付き確率の考え方と対比的に学習することが有望だと考えられる。大谷・古賀 (2024) が数学科と情報科の教科書がそれぞれどのように統計的内容を扱っているかを分析していたように, 統計教育は, 様々な教科の融合領域であり, より多角的かつ適切に統計学を理解するために, 数学教育が果たさなければならない役割は依然として大きい。数学教育がどの程度どのように統計教育に資するべきかは, 今後も継続した研究が必要である。

6. 結論

本稿は, 令和 7 年度の大学入学共通テストにおける仮説検定の考え方に関する出題が, フィッシャーの反証主義的な実験計画法としての仮説検定の考え方を適切に反映していない点を指摘し, そうした誤りを克服する教材の条件および教材の具体例を提案した。現状のハイブリッド仮説検定は, 仮説検定の考え方を適切に反映せず, 研究実践上, なし崩し的に生じた分析手法であるから, そもそもどの部分をもって「仮説検定の考え方」と呼んでいるのが不明確である。今後も今回の共通テストでの出題を真似た問題の演習が盛んに行われるようになったならば, ともすれば, 日本全国で指導される「仮説検定の考え方」が, 「仮説検定の考え方」と呼ぶに値しない計算処理に陥ってしまう可能性がある。

本稿が提案した教材の条件 [C1-1] ～ [C2-3] と, それに基づく具体的な教材例 (図 2) は, この問題を克服し得る教材である。数学 I と数学 A を並行履修していることが教材の活用的前提となってしまうのはいるが, 仮説検定の考え方の適切な理解に条件付き確率が必要なのだとすれば, このことは, 条件付き確率の理解が覚束ない生徒に仮説検定の考え方を教えようという現状のカリキュラム編成には, 教材の系統性の観点から無理があることを示唆している, とも受け取れる。

また, 学習指導要領上でも, 数学 I は「データの分析」

という名称で内容が構成されているが、データありきで分析することを伺わせるこの名称は、フィッシャーにせよネイマン・ピアソンにせよ、本来の仮説検定の考え方に照らせばミスリーディングな名称である。確かに、まず大量のデータが目の前にあって、それをいかに活かすかを考えなければならないような時代なのかもしれない。教育には、そうした時代の要請に応える責任がある。しかし、数学教育には、実験計画法の観点から見た不適切なデータ収集や確率論の観点から見た不適切な統計的分析を要請してくるような現代社会に抗える力を、生徒達に授ける責任がある。いい加減な理屈でデータ分析をし、統計学の装いを纏うことで分析結果を権威付けするような力を、我々は統計的リテラシーと呼ぶべきではない。むしろそこに No と言える力こそ、統計的リテラシーと呼ぶべきである。仮説検定の考え方の誤解を巡る問題は世界的な問題であり、一朝一夕に解決するものではないけれど、数学教育はこの問題にきちんと向き合わなければならない。

一方、本稿は、教材の条件と教材の具体例を提案するに留まってしまった。実際にこの教材を授業実践でどのように運用するかについては、今後の課題としたい。また、本稿では第二種過誤が扱われていないことを根拠に、高等学校数学における仮説検定の考え方がフィッシャーの仮説検定の考え方であるという前提で議論を進めたが、第二種過誤を扱うことも含め、高等学校数学でネイマン・ピアソンの仮説検定の考え方を一貫して採用すべきだという見方もあり得る。高等学校数学として最適な仮説検定の考え方が何であるかも継続した検討課題としたい。

注

- 1) 原文において「方針」欄は枠で囲われている。引用の都合上、ここでは枠を付さなかった。なお、「数学 I、数学 A」だけでなく、「数学 I」においても同じ問題の出題がある。
- 2) ただし筆者は、仮説検定を確率論的証明だと考えてしまうミスコンセプションについては、より慎重な検証が必要だと考えている。なぜなら、「証明」という概念が正しく理解されていないという問題の方が、数学教育研究においてより広く知られた問題だからである (Education Committee of the European Mathematics Society, 2011)。つまり、仮説検定を確率論的証明だと考えてしまっているのは、「仮説検定」の誤解ではなく「証明」の誤解のせいかもしれない。

研究費助成

本稿は、JSPS 科研費 (課題番号: JP24K00422) の助成を受けて実施された研究成果の一部である。

利益相反の開示

本稿には、教科書分析を実施した先行研究の引用や、教科書内容に関する考察、教科書に掲載し得る教材への示唆が含まれる。本稿の著者は、数研出版株式会社が発行する中学校向けの文部科学省検定済教科書の著者の 1 人であるが、先行研究の批評や本研究の実施にあたって、客観性と公平性を確保するよう努めた。

参考文献

- Antonini, S., & Mariotti, M. A. (2008). Indirect proof: What is specific to this way of proving? *ZDM*, 40(3), 401–412.
- Bakker, A., & Derry, J. (2011). Lessons from Inferentialism for Statistics Education. *Mathematical Thinking and Learning*, 13(1–2), 5–26.
<https://doi.org/10.1080/10986065.2011.538293>
- Brandom, R. (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism*. Harvard University Press.
- Carotenuto, G., Di Martino, P., & Lemmi, M. (2021). Students' suspension of sense making in problem solving. *ZDM – Mathematics Education*, 817–830.
<https://doi.org/10.1007/s11858-020-01215-0>
- Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98–113.
<https://doi.org/10.1016/j.edurev.2007.04.001>
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, Statistics, and Teaching. *The American Mathematical Monthly*, 104(9), 801–823.
<https://doi.org/10.1080/00029890.1997.11990723>
- 大学入試センター (2025). 『令和 7 年度本試験 数学① [数学 I 数学 I・数学 A]』.
- Devlin, K. (2014). The most common misconception about probability? In *Probabilistic Thinking: Presenting plural perspectives* (pp. ix–xiii). Springer.
- 土居淳子 (2010). 「帰納的推論ツールとしての統計的仮説検定: 有意性検定論争と統計改革」. 京都光華女子

- 大学人間関係学会『年報人間関係学』, 13, 15–36.
- Education Committee of the European Mathematics Society. (2011). Do Theorems Admit Exceptions? Solid Findings in Mathematics Education on Empirical Proof Schemes. *Newsletter of the European Mathematical Society*, 82, 50–53.
- Falk, R., & Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, 5(1), 75–98. <https://doi.org/10.1177/0959354395051004>
- フィッシャー, R. A. (2013). 『実験計画法』(遠藤健児・鍋谷清治 訳; POD 版). 森北出版. (邦訳の初版は 1971 年, 底本の第八版は 1965 年出版)
- Freudenthal, H. (1973). *Mathematics as an educational task*. Reidel.
- 福田博人・大谷洋貴・岩崎秀樹 (2018). 「統計的検定の教授単元の開発研究: 背理法からの展開と区別に着目して」. 日本科学教育学会誌『科学教育研究』, 42(4), 335–349. <https://doi.org/10.14935/jssej.42.335>
- Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics*, 16(2), 36–45.
- Hakamata, R., Uegatani, Y., & Hayata, T. (2023). How do open-ended problems promote indirect argumentation? Conceptual replications in a Japanese secondary school. *The Journal of Mathematical Behavior*, 70, 101026. <https://doi.org/10.1016/j.jmathb.2022.101026>
- Harel, G. (2013). Intellectual Need. In K. R. Leatham (Ed.), *Vital Directions for Mathematics Education Research* (pp. 119–151). Springer New York. http://link.springer.com/chapter/10.1007/978-1-4614-6977-3_4
- Ishibashi, I., & Uegatani, Y. (2022). Cultural relevance of validation during mathematical modeling and word problem-solving: Reconceptualizing validation as an integration of possible fictional worlds. *The Journal of Mathematical Behavior*, 66, 100934. <https://doi.org/10.1016/j.jmathb.2022.100934>
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- 小林廉 (2022). 「「仮説検定の考え方」の学習指導に関する一考察」. 日本数学教育学会誌『数学教育』, 104(5), 16–25. https://doi.org/10.32296/jjsme.104.5_16
- 松原望. (1989). 「有意性検定の社会的含意」. 『科学基礎論研究』, 19(2), 59–62. https://doi.org/10.4288/kisoron1954.19.2_59
- 文部科学省 (2018). 『高等学校学習指導要領 (平成 30 年告示) 解説 数学編 理数編』.
- Otani, H. (2019). Comparing structures of statistical hypothesis testing with proof by contradiction. *Hiroshima Journal of Mathematics Education*, 12, 1–12. <https://doi.org/10.24529/hjme.1201>
- 大谷洋貴・石橋一昂 (2022). 「仮説検定の考え方における確率の意味に関する教科書分析」. 『日本科学教育学会年会論文集』, 46, 547–548. https://doi.org/10.14935/jssep.46.0_547
- 大谷洋貴・古賀峻也 (2024). 「高等学校数学 I と情報 I における統計教育の相違: 教科書分析を通して」. 日本数学教育学会『第 57 回 秋期研究大会発表集録』, 377–384.
- Reeves, C. A., & Brewer, J. K. (1980). Hypothesis Testing and Proof by Contradiction: An Analogy. *Teaching Statistics*, 2(2), 57–59. <https://doi.org/10.1111/j.1467-9639.1980.tb00387.x>
- Seidouvy, A., & Schindler, M. (2020). An inferentialist account of students' collaboration in mathematics education. *Mathematics Education Research Journal*, 32, 411–431. <https://doi.org/10.1007/s13394-019-00267-0>
- Thomas, R., & Gerofsky, S. (1997). An exchange about word problems. *For the Learning of Mathematics*, 17(2), 21–23.
- 上ヶ谷友佑・袴田綾斗・早田透 (2021). 「「間接証明」の集合体モデル」. 全国数学教育学会誌『数学教育学研究』, 27(2), 33–50. https://doi.org/10.24529/jasme.27.2_33
- 上ヶ谷友佑・石橋一昂・服部裕一郎 (2021). 「学校数学の内容としての数学的方法学」. 全国数学教育学会 第 55 回研究発表会 発表資料. <https://ir.lib.hiroshima-u.ac.jp/00051584>
- 上ヶ谷友佑・大谷洋貴 (2019). 「数学教育における推論主義の可能性: 学力調査で求められる実践的知識としての統計的概念に関する批判的考察」. 全国数学教育学会誌『数学教育学研究』, 25(1), 67–76. https://doi.org/10.24529/jasme.25.1_67
- Uegatani, Y., & Otani, H. (2021). A new ontology of reasons for inferentialism: Redefining the notion of conceptualization and proposing an observer effect on assessment. *Mathematics Education Research Journal*, 33, 183–199. <https://doi.org/10.1007/s13394-019-00289-8>
- Uegatani, Y., & Otani, H. (2023). An inferentialist view of notions and concepts. *For the Learning of Mathematics*,

43(3), 2–6.

- Uegatani, Y., Otani, H., & Fujita, T. (2025). Decentralising mathematics: Mutual development of spontaneous and mathematical concepts via informal reasoning. *Educational Studies in Mathematics*, 118(2), 229–248. <https://doi.org/10.1007/s10649-024-10366-w>
- Uegatani, Y., Otani, H., Shirakawa, S., & Ito, R. (2023). Real and illusionary difficulties in conceptual learning in mathematics: Comparison between constructivist and inferentialist perspectives. *Mathematics Education Research Journal*. <https://doi.org/10.1007/s13394-023-00478-6>