

統計教材のパフォーマンス課題とその評価

喜田英昭・広島大学附属中高等学校

〒734-0005 広島県広島市南区翠 1-1-1

TEL : 082-251-0192 FAX : 082-252-0725

E-mail : hkida@hiroshima-u.ac.jp

1. はじめに

平成 28 年 8 月に報告された「算数・数学ワーキンググループにおける審議のとりまとめ」において「小・中・高等学校を通じた統計教育のイメージ」として、高等学校は次のように述べられている。

・統計的に分析するための知識・技能を理解し、日常生活や社会生活、学習の場面等において問題を発見し、必要なデータを集め適切な統計的手法を用いて分析し、その結果に基づいて問題解決や意思決定につなげる。

・データの収集方法や統計的な分析結果などを批判的に考察する。

このように、高等学校における統計教育では「問題解決」、「意思決定」、「批判的な考察」が求められているが、では、これらの目的を果たすための学習活動とその評価をどのようにすればよいだろうか。

そこで、「統計的な知識・手法を用いて現実の課題を解決するパフォーマンス課題に取り組むことで、生徒の統計的な問題解決能力が育成される」と研究仮説を設定し、パフォーマンス課題の開発とその評価について実践研究を行う。本稿では、具体的なパフォーマンス課題と生徒の記述を述べ、この実践研究の意義と課題を報告する。

2. 「統計科学」の授業概要

本校はスーパーサイエンスハイスクール（SSH）に指定されており、研究開発課題「社会に開かれた

科学技術を先導する人材育成の起点となる科学教育カリキュラムの開発」に取り組んでいる。本校では、社会に開かれた科学技術（未来社会の創造に向けた科学技術イノベーション）を先導する人材の基盤をなす能力を「Sagacity」とし、この能力を育成するために「課題研究」を中核とした学校設定教科「SAGAs」を設定し、科学教育カリキュラムの開発を行っている。この研究開発では、特に次のことを目指している。

S : 科学的 (Scientific) : 課題研究の基盤となる融合カリキュラムの開発

A : 高度かつ専門的 (Academic) : 課題研究を活性化・高度化するための高大連携・接続システムの開発

G : 国際的 (Global) : 海外の理数重点校及び広島大学留学生との協働的課題研究の実施を内容とする定常的交流プログラムの開発

A : 主体的・自律的 (Autonomous) : 主体的・自律的な課題研究を展開するための「広大メソッド」の開発

この研究開発では、第 2 学年、第 3 学年に普通科 AS (Advanced Science) コース 1 クラス、普通科 GS (General Science) コース 4 クラスを設置し、学校設定教科「SAGAs」において、AS コースは「AS 統計科学」（第 2 学年、1 単位）、GS コースは「GS 社会と統計」（第 2 学年、1 単位）で統計に関わる内容を学習している。「AS 統計科学」のカリキュラムを以下に示す。

3. パフォーマンス課題とその結果

本年度は試行的に4回のパフォーマンス課題を実施した。単元の総括的評価として行い、約2週間の製作期間をとり、レポート形式で提出させた。本稿では第1回、第3回の課題の内容と主な生徒の解答を以下に示す。

第1回 条件付き確率と確率分布（2020年7月）

PCR検査等でのウイルスに感染しているかの検出検査を判断するとき、「感度」、「特異度」、「罹患率」、「適合率」という用語で説明されることがあります。（以下、説明略）現在のPCR検査は、感度は70%、特異度は90%程度であるといわれています。

- (1) 罹患率が①0.1%、②1%、③10%のそれぞれの場合について、適合率を求めよ。
- (2) 「誰でもPCR検査が受けられるようにする」ことについて、適合率を用いてこの手法の是非を論じなさい。
- (3) ウィルスに感染していないことを判断するのに、この検査を繰り返して行い、2回連続で「陰性」が出れば感染していないと判断されていました。この判断は正しかったのでしょうか。罹患率を定め、2回連続で「陰性」であったとき、感染していない確率を求め、判断の正しさを論じなさい。

この課題と同時に以下の評価基準も提示した。これを先に示したのは、単に数値を求めて終わるのではなく、得られた数値から何がいえるか、つまり、統計を用いて問題解決を行うことに価値づけをする

【評価基準】

- S: 罹患率と適合率の関係を正しく理解し、科学的、社会的な側面から多角的に考察し、具体的に解決策を提案することができる。
- A: 罹患率と適合率の関係を正しく理解し、科学的、社会的な側面から考察することができる。
- B: それぞれの罹患率における、適合率を正しく求めることができる。
- C: 基準Bに達していない。

ためである。

この課題では、PCR検査の拡大が求められていた当時、条件付き確率を用いて正しく検査の有効性を理解すること、そして、その結果について、「科学的、社会的な側面から多角的に考察すること」を求めた。生徒の解答は、言葉だけのものから数式やグラフを用いて議論しているものなど幅広くみうけられた。次に生徒S₁の(2)の解答を示す。

(i) 適合率への影響

「誰でもPCR検査が受けられるようにする」= 検査数(母集団)が増える
= 罹患率が下がる

- (1) で計算した通り、罹患率が下がる、適合率は下がる。
また、右下の式でも確認できる。

	罹患している	していない
陽性	真陽性	偽陽性
陰性	偽陰性	真陰性

$$\text{適合率} = \frac{\text{真陽性}}{\text{真陽性} + \text{偽陽性}}$$

適合率の低下がどう社会に影響を与えるのかについて注目する。

(ii) 社会・科学への影響

a) PCR検査の実態

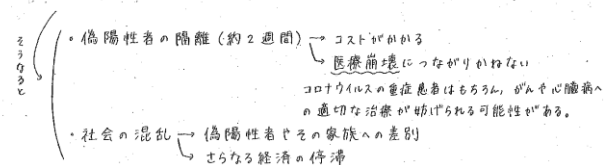
- ・鼻やのどを綿棒で拭いた液の中に新型コロナウイルスの遺伝子が存在するのを確認する。
- ・鼻やのどからうまく採取できない、感染したばかりでウイルスが十分に集まった場合などは検知されない。(などの理由で精度は低い)
- ・検査の結果が出るまで時間と金がかかる。

b) PCR検査を増やすと?

- (i) より適合率が下がる。

ii

偽陽性者(陽性と認定されるが、実際は感染していない者)が増える
→ 一方で、検査で陽性=感染者というのが国民の大半の認識と思われる



また、偽陰性者も検査数が増えることで増えしまい、感染していない確信による意識の大幅な低下がおこりやすくなり、集団感染の一因となってしまうかもしれない。

(iii) まとめ・提案

以上より、私はPCR検査を増やすことが必ずしも良い結果を導くとは思わない。医療崩壊や社会の混乱を招く可能性も考えると、検査を増やすより、検査キット、マスクの着用や手洗いの励行などの予防、不要な集まり活動の自粛をし、感染リスクを減らすことが適切であると考えられる。また、国民が正しく情報を理解できるように働きかけるべきだと思う。

この生徒は「誰でもPCR検査が受けられるようにする」といった主張に対し、計算式により罹患率と適合率の関係を示し、次に「適合率の低下がどう社会に影響を与えるのか」について社会的側面を含めて議論している。2020年7月の段階で「医療崩壊」への可能性も言及しており、確率という統計的

な手法を用いて社会問題への批判的な考察ができていると考えられる。

第3回 推定とその利用（2020年9月）

- (1) ある工場では、「製品の強度が正規分布にしたがって分布するものとして、製品の強度の平均値の95%信頼区間を求めるとき、その下限が100kg以上であること」という条件の製品の注文を受けた。そこで、いくつか作った試作品の中から5個を無作為抽出してそれらの強度を測定した結果、100, 105, 105, 110, 135(kg) という結果を得た。ところが、工場長は、データを100, 105, 105, 105, 105(kg) のようにわざと低い数値に書き直して注文主に提出した。工場長は、なぜこのように「良いデータをわざわざ悪く書き換える」ような不正を行なったのだろうか、その動機を工場長の立場になって具体的な数値を用いて説明せよ。
- (2) この事例は問題のために作成したもので、実際には、このような不正な改ざんはもちろんしてはならない。では、工場長はどのような行動をとるべきだったか。科学的側面、社会的側面にそって自分の考えを述べよ。なお、複数の考えがあれば、それも述べなさい。

この課題では、標本の中に外れ値のような極端な値が現れたときに、その標本から得られる信頼区間がどのようになるのかを通して、データに対する正しい向き合い方や、データのとり方の改善点を理解することを目的としている。この課題では、以下の評価基準を設定した（S、Aのみ示す）

【評価基準】

S: データ改ざんの問題を理解し、信頼区間を正しく求め、それを基に統計的に正しく議論することができる。さらに、改ざんなどの社会的な問題に対して科学的、社会的な側面から多角的に考察し、具体的な対策を提案することができる。

A: データ改ざんの問題を理解し、信頼区間を正しく求め、それを基に統計的に正しく議論することができる。さらに、改ざんなどの社会的な問題に対して科学的、社会的な側面から考察できている。

次に、生徒 S₂ の解答を以下に示す。

(2) 元のデータで求めた95%信頼区間の幅が大きすぎるのは、データの散らばりが大きいためである。しかし、実際に無作為抽出された製品は5個しかなく、その5個だけでは全体の傾向をつかむのはかなり難しい。したがって、製品の強度の95%信頼区間を正確に求めるために、工場長はより多くのデータを用いるべきだった。現に、元のデータでは135kgという数値が飛び抜けて大きく見受けられる。もしも、と多くのデータがあれば、この135kgを外れ値であるか否かを判定できたはずである。そして、外れ値に分類されたならば、平均や分散を算出して求められる95%信頼区間に大きな影響を与えてしまうので、このデータは除いて計算をするべきである。望ましい。また、このような注意を工場長が行っていたのは、工場長に統計についての正しい認識が不足していたからではないかと考えられる。もし正しい認識を十分に持っていれば、不正ではなく、上記のように計算方法の見直しをするはずだからである。工場長が不正を行おうと思ったのは、やはりもう少し安いデータで求めた95%信頼区間が条件に合致しなかったからだ。そのような状況を防ぐためには、あらかじめ製品の注文条件の中に、95%信頼区間を求めるために無作為抽出するデータの個数を含めておくことが良い解決方法ではないかと思う。なぜなら、95%信頼区間は確かに基準となる値だから、それを求めるデータの数を指定すれば、正しく求めることができるからだ。そして、正しく求められている基準に従って判断をすることは余りにも危うい。

S₂は「135kg」の扱いについて「多くのデータがあれば外れ値であるか否かを判断できる」とし、外れ値ならば「このデータは除いて計算するほうが望ましい」としている。さらに「注文の条件に無作為抽出するデータの個数を含める」という提案をしている。「正しく求められてない基準に従って判断することは余りにも危うい」という最後の言葉は、データを扱う研究者にとって最も大事な戒めであった。

4. まとめ

本稿で報告した生徒の解答は目的とする「統計的な手法を用いた多角的な考察」が出来ていたが、多くの生徒が、統計的な手法よりも社会的な側面の議論を多く行っていた。本来は統計的な意思決定を行うことが目的であるから、より統計的な手法について議論を深められるような課題設定が求められる。

参考文献

文部科学省中央教育審議会初冬中等教育分科会教育課程部会（2016）「算数・数学ワーキンググループにおける審議の取りまとめ」

https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/073/sonota/_icsFiles/afieldfile/2016/09/12/1376993.pdf（2021年2月2日確認）