

統計教材のパフォーマンス課題とその評価

広島大学附属中・高等学校
喜田英昭



本校のSSHについて

国立大学法人広島大学附属高等学校 第4期SSH

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

AS (Advanced Science) コース

第2学年: ASサイエンス・コミュニケーション

第2学年: AS統計科学

社会に開かれた科学技術を
先導する人材の基盤となる能力

SAGAcity

(賢明な判断が下
せる) 洞察力, 深い
知性, 先見の明

3年間の「課題研究」

第3学年:
AS科学探究Ⅱ

第2学年:
AS科学探究Ⅰ

第1学年:
総合科学

「課題研究」を中核とした学校設定教科「SAGAs(探す)」

第2学年:
GS総合科学探究Ⅰ

第3学年:
GS総合科学探究Ⅱ

GS (General Science) コース

第2学年: GSクリティカル・コミュニケーション

第2学年: GS社会と統計

Scientific

S

「科学する」
教科融合授業

Academic

A

高大連携・
接続システム

Global

G

海外連携校との
課題研究

Autonomous

A

広大メソッド
(指導・評価)

学校設定教科「SAGAs(探す)」

国立大学法人広島大学附属高等学校 第4期SSH

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

新・学校設定教科「SAGAs(探す)」

第2学年より、普通科5クラスのうち、
1クラスをASコース、4クラスをGSコースとする

	AS (Advanced Science) コース	GS (General Science) コース	SSHの4つの柱
3年	AS科学探究Ⅱ (1単位)	GS総合科学探究Ⅱ (1単位)	
2年	AS科学探究Ⅰ (2単位)	GS総合科学探究Ⅰ (2単位)	
	AS統計科学 (1単位)	GS社会と統計 (1単位)	
	ASサイエンス・コミュニケーション (1単位)	GSクリティカル・コミュニケーション (1単位)	
1年	総合科学 (2単位)		

「AS統計科学」の学習目標

1. 標本の関数である統計量をもとに母集団の特徴を推測する**推測統計学の手法を学び**，具体的な課題に対して推測統計の手法を活用できるようにする。
2. 推測統計の手法を用いて，自然科学，社会科学等における**現実的な課題に対して問題解決を行うこと**で，データを分析，解釈，表現し，**データに基づく判断や意思決定ができる**データリテラシーを育む。

「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」

- 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**。

研究仮説

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。

「AS統計科学」の学習内容

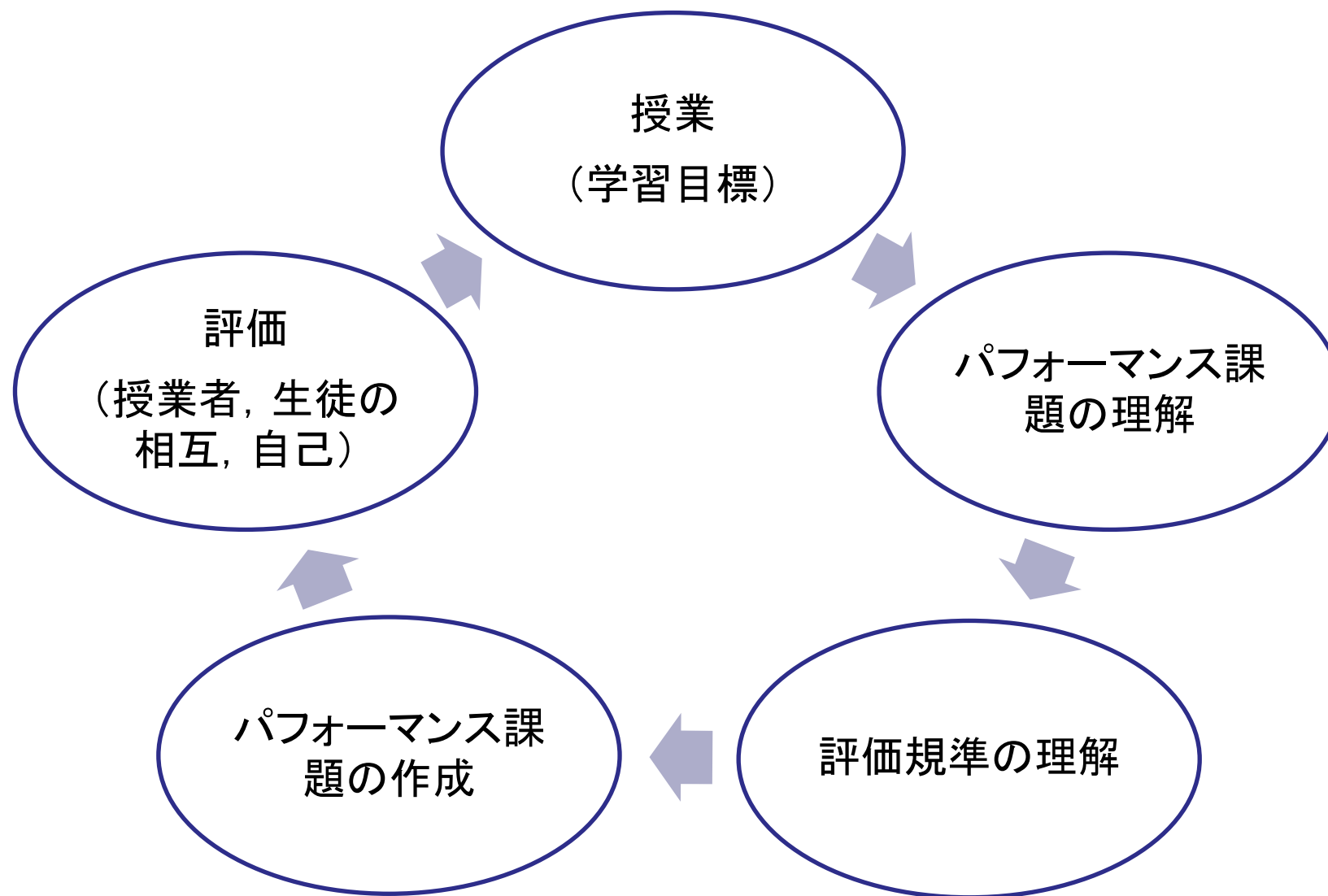
1学期	2学期	3学期
推測統計と確率分布の考え方 推定法	統計的検定 分散分析	回帰分析

- ・教科書は使用せず(教科書を特定していない), 毎回ワークシートを配布する。
- ・通常教室で授業を行い, 自分のスマートフォン等の使用は認めている。
- ・パフォーマンス課題は単元が終わった後に提示し2週間程度の作成期間をとっている。

パフォーマンス課題の目的

- 総括的評価（5段階評価）のため
- 学習目標の理解のため

学習目標の理解のためのパフォーマンス課題



パフォーマンス課題の内容

第1回 条件付き確率(7月)

第2回 二項分布と正規分布(7月)

第3回 平均値の区間推定(9月)

第4回 分割表の検定(12月)

第5回 二元配置分散分析(2月)

パフォーマンス課題の内容

第1回 条件付き確率(7月)

第2回 二項分布と正規分布(7月)

第3回 平均値の区間推定(9月)

第4回 分割表の検定(12月)

第5回 二元配置分散分析(2月)

第1回 条件付き確率

【課題】 PCR検査等でのウィ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。**

第1回 条件付き確率

【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。

A評価の例

(2) (1)より、検査陽性者の半数は感染していない。また、割合が高いものの陰性の人の中にも感染者がいることになる。「誰でもPCR検査を受けられるようにする」と、その分罹患率は下がるのでより適合率が下がり、大量の感染していない人も病院に殺到してすぐに医療崩壊が起きてしまうので良い方法とは言えないと思う。実際に厚生労働省がやったように、感染している可能性が高い人に検査をさせることにより検査精度が飛躍的に上がるので実際の対応には良い点がたくさんあったのではないかと考えた。

S評価の例

(2) (i) 適合率への影響

「誰でもPCR検査を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

(i) で計算した通り、罹患率が下がると、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、たり、感染したばかりでウイルスが十分に無か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人件がかかる。

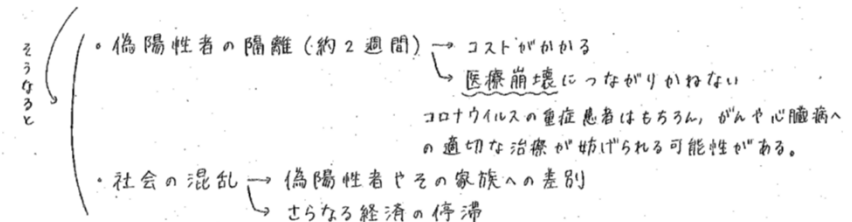
a) PCR検査を増やすと?

(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てしまい、感染していない確信による意識の大幅な低下がおりやすくなり、集団感染の一因に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、咳エチケット、マスクの着用や手洗いの励行などの予防、不要な集会・活動の自粛をし、感染リスクを減らすことの方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。

各評価の人数

評価	S	A	B	C
授業担当者	16	24	0	0
別の教員 (数学)	12	23	5	0

- 「多角的な考察」を, 「幅広い考察」ととらえ色々な分野から考察をしているが, 統計的に「深い考察」になっていない。
- 授業者は実際の授業の様子がわかり, 「こうあってほしい」という願望があるため, 低い評価(Bなど)をつけにくい。

パフォーマンス課題の内容

第1回 条件付き確率(7月)

第2回 二項分布と正規分布(7月)

第3回 平均値の区間推定(9月)

第4回 独立性の検定(12月)

第5回 二元配置分散分析(2月)

第4回 独立性の検定

【課題】 ある疾病Aについての予防接種の効果を調べるために、40人を無作為抽出し調査したところ、次のようなデータが得られた。

	予防接種を受けた人	予防接種を受けなかった人	計
疾病 A に罹患した人	6	8	14
疾病 A に罹患しなかった人	16	10	26
	22	18	40

(1) 疾病Aに罹患することと予防接種の間に関係はあるだろうか。

次に、もっと多くのデータを得るために、新たに200名を無作為抽出し、調査したところ、次のようなデータが得られた。

	予防接種を受けた人	予防接種を受けなかった人	計
疾病 A に罹患した人	28	42	70
疾病 A に罹患しなかった人	77	53	130
	105	95	200

(2) 疾病Aに罹患することと予防接種の間に関係はあるだろうか。

(3) (1),(2)の問題から、疾病Aについての予防接種の有効性について、統計的手法を用いて判断することについて自分の考えを科学的側面、社会的側面から述べよ。

第4回 独立性の検定

【評価基準】

S: 分割表の検定について、正しく検定することができる。さらに、予防接種の有効性について科学的、社会的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: 分割表の検定について、正しく検定することができる。さらに、予防接種の有効性について科学的、社会的側面から考察できている。

B: 分割表の検定について、正しく検定することができる。

C: 分割表の検定について、正しく検定することができていない。

B評価の例

(3) 11) では関係がなかったけど、12) では関係があった。
つまり母数を増やすことによって予防接種と罹患するとの
関係があるかたなりかが変わった。インフルエンサウイルス
や新型コロナウイルスもリリキーンが効果があるかたなり
かを調べる必要がある。その時に使用するデータも
とるだけ多くする必要がある。
また、予防接種をうける私たちも、予防接種に期待し
ます。手洗いうがいを怠らないうにしなければならなし。

B評価の例

(3) 今回、統計で予防接種の有効性を判断した結果が正反對のものになってしまった。
この結果からは、予防接種の有効性を統計的手法で判断するのは良い方法とは言えない。しかし、一般の人が予防接種の有効性を知ろうと思った時、統計のデータは分かりやすく、数値で示されるので、社会的には良い方法だと思った。一方で、罹患することと予防接種を受けることとの関係をどちらの結果でも出すことができる、メディアなどが使うときに都合の良い方の結果のデータを選んで使うことができちゃうので、注意も必要だと思う。
また、予防接種の有効性は子ども、大人で異なると思うので、年代別にした標準を使って結果を出すのも良いのではないかと考えた。

A評価の例

(3)(1), (2)の問題から, 予防接種の有効性について, 統計的手法を用いて判断することについて自分の考えを科学的側面, 社会的側面から述べよ。

今回の課題(1)と(2)を比べたとき, (1)の古の抽出数が少なかったため, (1)と(2)で結果が異なるものとなつたのだと考えられる。そもそも目的が, 「予防接種の効果を調べるため」とあるので, 抽出数40の時点で少なすぎる。また, 抽出数の大小によって得られる結論が変わる手法は, どれだけ多くの抽出数をとれば良いのかあやふやであり, また, Excelでの情報の使われ方について, Excelに都合の良い方の結果が使われかねない。

今回の課題のように, 目的に対して明らかに抽出数が少ない場合の結論が使われないようにするためにも, 分割表の検定の結果を用いる際には, 結果だけでなく, 抽出数も一緒に示すべきだと考える。

S評価の例

(3) (1)と(2)の問題の違いはデータの大きさのみであり、対象の疾病、予防接種はまったく同じである。しかし、データの大きさによって検定の結果は変わった。統計学的にデータの大きい(2)を信用すべきなのは言うまでもないが、問題なのは検定結果が望まれた結果でなかったときに、そのことを社会に公表できるかということだと考える。

社会に対して成功という成果をあげたいがために検定結果を公表せず、望まれる検定結果になるまで、データを大きくするのは、もちろん今回の問題のように成功することもあるが、いたずらにコストだけがかさんでいくリスクもある。

この問題を防ぐには、社会がデータが大きくなればなるほど精度が増すが、絶対ではない統計の性質を理解し、研究者と社会の両者が納得する予算を確保し、研究者が安心して結果を公表できる環境をつくる必要がある。また、研究者も自身の研究の内容はもちろん、統計の性質などを社会に対して説明する必要もある。

S評価の例

「それでは、予防接種の有効性について判断する際、分割表の検定を使うべきではないのだろうか？ また、どのような対策をとるべきなのだろうか？」

先程、分割表の検定では、 n の大きさによって結果が左右されてしまう、と述べた。インターネットで調べてみると、確かに、標本が十分に大きくない際は、正規分布への近似があまり良好ではないようだ。このような問題への対処として、正規分布への近似の改善を意図したイエーツの連続修正の χ^2 検定や、フィッシャーの直接確率計算が用いられることもあるようだ⁴⁾が、元の χ^2 の検定の方がより正確であり、 $\frac{n \times f_{ij}}{n} \geq 5$ のときは、 χ^2 分布の検定が推奨されている。今回の(1)(2)も分割表の検定を用いた判断は適確だと言える。また、 χ^2 分布の検定は計算が容易であることから、実際の医療現場（今ではコロナウイルスのワクチンが1日でも早く使えるように努力している方々がいる）でも役に立ちそうである。よって、分割表の検定も χ^2 検定で行うことは社会的な面においても有効であるようだ。では、標本の大きさの影響を踏まえたうえで、どのような対策をとるべきだろうか？ 私が提案したいことは主に2つある。1つ目は、 χ^2 検定の特徴をしっかりと把握しておくことだ。結果だけに着目せず、標本の大きさやデータの様相も理解しておきたい。2つ目は、標本の大きさをあらかじめ考えておくことだ。これは、より有利な結果になるように操作することを防げる。

各評価の人数

評価	S	A	B	C
授業担当者	10	26	4	0
自己評価	12	18	10	0

- 統計的な考察が現実にはどのように影響を与えているか、統計部分の考察が弱い。
- 生徒同士の相互評価後に自己評価を行ったので、自分を低く判断しB評価をつける生徒が見受けられた。

今後の課題

- 学習目標を達成できるようなパフォーマンス課題とその評価規準を設定する
- 「統計的な考察」を意識した評価基準の設定
- 高校生の「統計」は、何を目標とし、どこまで学べばよいか