

身近なデータを『サイエンス』してみよう ～附属中学校「大学研究室訪問」より～

たけむら けんじ
竹村 謙司

1. はじめに

奈良教育大学附属中学校では、「学問の世界にふれ、学びの方法を知る」ことを主題に、第2学年の生徒が大学教員の講義を受け、質疑や対話を通して学ぶことの意味やその方法について知り、学問の世界の深さと広がりを知るとともに、学ぶことから得られる充実感や喜びについて考えることを目標として、「大学研究室訪問」を実施している。その中の講義の一つである「身近なデータを『サイエンス』してみよう」では、小学校算数や中・高等学校数学の領域である「データの活用」「データの分析」で学ぶ考え方をを用いて日常の身近な事象をデータ化することにより、「バイアス（偏りや主観）」を取り除き、「ファクト（客観的な事実）」をとらえることをテーマとしている。今回は3つの身近なデータの活用や分析を通して、中学校数学科の統計分野から始まり高等学校や大学でのデータ分析につながる指導や研究について述べてみたい。

2. 『ヘイキン』とは？

「平均値」に関する【課題1】について考察した。

【課題1】

Aさんの数学と国語の得点は、どちらも66点でした。Aさんは数学と国語のどちらが得意だといえるでしょうか。

数学と国語の平均点はどちらも同じであることから、「平均値」以外のどの値を調べて比較すればよいのかを、現行の小学校学習指導要領算数編及び中学校学習指導要領数学編「Dデータの活用」領域に挙げられている用語から選択するという課題である（図1）。

得点を平均点と比べて「真ん中」より上位か下位かを判断することは、日常よく行う比較であるが、値の分布によっては順位が大きく異なってくる（図2）。

中学校2年で扱う「四分位範囲」や「箱ひげ図」は、データの分布（散らばりの度合い）を比較することに適している。今回のデータでは、上位25%の境界値である第3四分位数と得点を比較することで、平均点以外の値を用いて上位か下位かを考察することができる（図3）。

1. 『ヘイキン』とは？

Aさんの数学と国語の得点は、どちらも66点でした。Aさんは数学と国語のどちらが得意だといえるでしょうか。

Q1 次のどの値を調べて比較しますか？

【フォーム①】で回答

1. 平均値 小6
2. 中央値（メジアン） 小6
3. 最頻値（モード） 小6
4. 階級 小6
5. 相対度数 中1
6. 累積度数 中1
7. 四分位範囲 中2
8. 箱ひげ図 中2
9. 誤差 中3
10. 近似値 中3

数学	国語
78	66
48	58
67	32
59	58
77	82
8	31
69	62
14	24
49	47
41	37
66	66 A
61	58
30	34
65	60
28	42
99	75
28	41
21	33
4	33
85	46
20	38
12	37
11	45
79	60
74	49
61	64
77	69
39	44
56	26
43	39
77	61
84	70
12	15
11	40
30	42
48.1	48.1

ホントウに平均値を比べれば、どちらが得意か分かるの？

図1

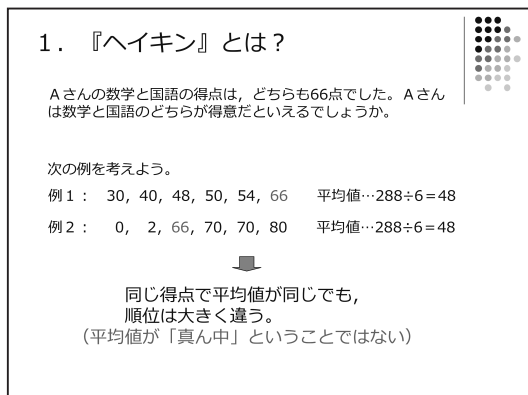


図 2

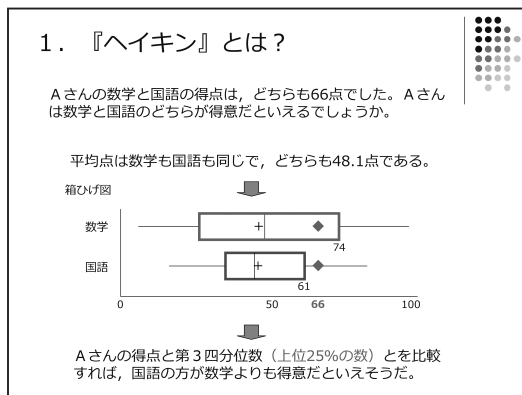


図 3

3. 数学ガール

書籍「数学ガール」を読んでいるかどうか、数学が好きであるかどうかという2つの分類について調査したクロス集計表に基づく仮説に関する【課題2】について考察した。

【課題2】

【仮説】 数学ガールを読んでいる人は数学好きである。

	読んでいる	読んでいない
数学が好き	10 人	2 人
数学が好きでない	10 人	18 人

【課題2】の仮説に対して、2つの検証を行った。【検証1】では、数学ガールを読んでいる人のうち数学が好きな人の割合が50%であることから、仮説の正しさは「五分五分」としている。日常的な感覚としては、「五分五分」だから仮説は正しいとも正しくないとも言えないという判断が多いと推測される(図4)。

【検証2】では、数学ガールを読んでいる人のうち数学が好きな人の割合とクラス全体のうち数学が好きな人の割合を比べることで、仮説は「正しい」としている(図5)。

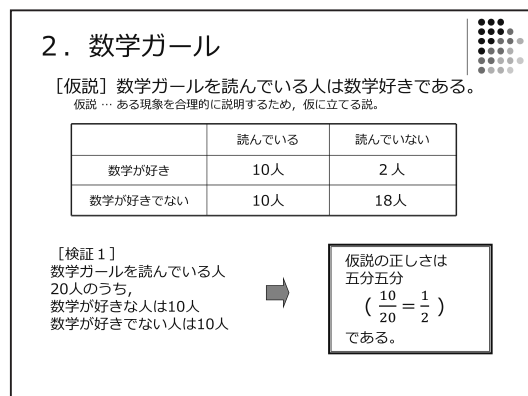


図 4

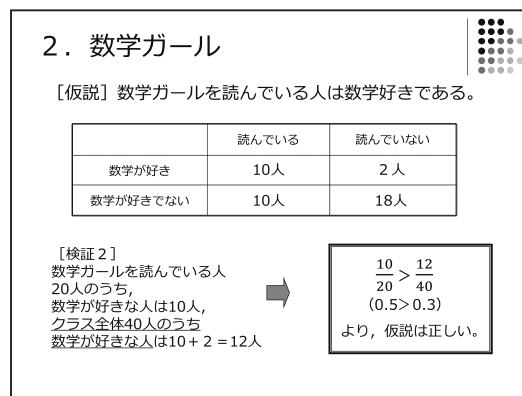


図 5

このことは、数学ガールを読んでいるという「条件」を付けたときと付けないときの割合を比較することの必要性を示しており、高等学校で学習する「条件付き確率」の考え方につながっている。また、[検証 1] の「五分五分」が曖昧な判断基準であることも示唆している（図 6）。

なお、2 つ以上の分類基準をもつクロス集計表において、分類基準間に関連があるかどうかを検定する統計的手法である χ^2 検定によるデータ分析を行うと、数学が「好き」か「好きでない」という結果には、「数学ガールを読んでいるかどうか」が影響していると認められる（ $\chi^2(1) = 7.619048$, $p=0.005775$ ）（図 7）。

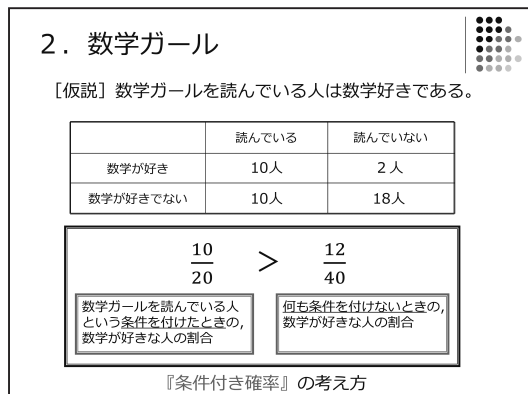


図 6

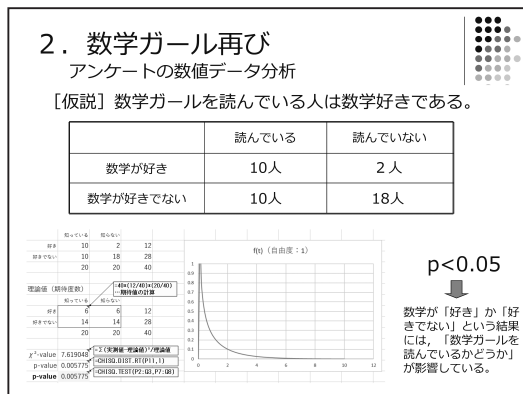


図 7

4. 附中進路ガイダンス

高校入試の合格率に関する【課題 3】について考察した。

【課題 3】

高校入試の合格率は 95% とします（5% の人が合格しない）。

ある模擬試験を受けたとき、

高校入試で合格する人が模試で D 判定以下となる確率は 5%

高校入試で合格しない人が模試で D 判定以下となる確率は 95%

であることが分かっています。

このとき、模試で D 判定以下だった人が高校入試に合格する確率は？

【課題 3】において、高校入試で合格する人が模試で D 判定以下となる確率は 5% と設定されていることから、模試で D 判定以下だった人が高校入試に合格する確率は高くないと推測する生徒が多い（図 8）。

「過去の経験」と「新たに得たデータ」をもとに、「推定したい事象」を予測する手法である「ベイズ推定」を活用すれば、D 判定以下の人も 50% の確率で合格することが示される。また、ベイズ推定の基本概念である「事前確率」「尤度」「事後確率」に基づく「ベイズの定理」は、小学校算数の単元である「割合」の考え方をを用いて説明することができる（図 9）。

ベイズ推定では、一つ前の事象で確認された事後確率を次の事象の事前確率として用いることで、事後確率を逐次更新していく「ベイズ更新」の考え方が用いられる。例えば、模試で 2

回連続 D 判定以下だった人が高校入試に合格する確率をベイズ更新により求めるなど、数学的活動としての取組も考えられる。

3. 附中進路ガイダンス

高校入試の合格率は95%とします（5%の人が合格しない）。
ある模擬試験を受けたとき、
高校入試で合格する人が模試でD判定以下となる確率は5%
高校入試で合格しない人が模試でD判定以下となる確率は95%
であることが分かっています。
このとき、模試でD判定以下だった人が高校入試に合格する確率は？

Q2 どのくらいの確率でしょうか？
① 10%以下 ② 20~30% ③ 40~50% ④ 60~70% ⑤ 80%以上

Q3 選んだ番号とその理由は？

【フォーム④】で回答
【フォーム⑤】で回答
グループ協議



図 8

3. 附中進路ガイダンス

高校入試の合格率は95%とします（5%の人が合格しない）。 → 「過去の経緯」（事前確率）
ある模擬試験を受けたとき、
高校入試で合格する人が模試でD判定以下となる確率は5% → 「新たに得たデータ」（尤度（ゆうど））
高校入試で合格しない人が模試でD判定以下となる確率は95%
であることが分かっています。
このとき、模試でD判定以下だった人が高校入試に合格する確率は？ → 「推定したい事象」（事後確率）

奈良県公立中学校
第3学年…約10,000人

① $10,000 \times 0.95 \times 0.05 = 475$ 人
② $10,000 \times 0.95 \times 0.95 = 9,025$ 人
③ $10,000 \times 0.05 \times 0.95 = 475$ 人
④ $10,000 \times 0.05 \times 0.05 = 25$ 人

求める確率は

$$\frac{\text{①}}{\text{①} + \text{③}} = \frac{475}{475 + 475} = \frac{1}{2} \quad (50\%)$$
 （ベイズ推定）

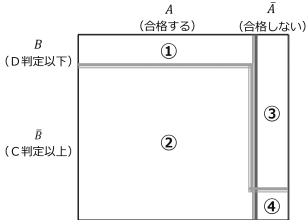


図 9

5. リフレクション

今回取り扱った3つの課題では、それぞれ「平均より上か下か」「五分五分」「D 判定以下」という日常に潜む「バイアス」に対して、「箱ひげ図」「条件付き確率」「ベイズ推定」という「データの活用」「データの分析」で学ぶ考え方を用いて、「ファクト」をとらえることができた。受講した生徒は、予想した考えとは異なる結果が統計的な手法を用いて明確に示されることに対して、驚きとともに興味をもって考察していた。「バイアス（偏りや主観）」を取り除き、「ファクト（客観的な事実）」をとらえることで、真理や目的に到達することができる。そのための『データサイエンス』の重要性について、拙稿より感じていただければ幸甚である。

本学教職大学院では、学校教育マネジメント科目として『教育のためのデータサイエンス』が開講されており、教育現場における多様かつ大量のデータの有用性に着目し、データサイエンスの手法によりデータを分析し、その結果を今後の教育活動に活用するための分析に関する基礎を習得することを目的としている。小学校算数や中・高等学校数学の領域である「データの活用」「データの分析」を通じて、現職の教員が現実の統計解析へのつながりを意識するための指導について、今後も研究を進めていきたい。

参考文献

- 結城浩（2007）：数学ガール．SB クリエイティブ株式会社
 涌井良幸（2022）：高校生からわかる統計解析．ベレ出版
 本丸諒（2018）：文系でも仕事に使える 統計学はじめての一步．株式会社かんき出版
 竹村謙司（2024）：教育のためのデータサイエンス．奈良教育大学大学院専門職学位課程
https://www.nara-edu.ac.jp/ADMIN/KYOUNU/SYLLABUS/syllabus/2024/detail/2024_11_K7040.html

（奈良教育大学 教授）