

推測統計の指導について

橋本三嗣・広島大学附属中・高等学校
〒734-0005 広島市南区翠一丁目1番1号
TEL.082-251-0192 FAX.082-252-0725
E-mail: mhashimo@hiroshima-u.ac.jp

1. はじめに

小学校のプログラミング教育は始まり、高等学校の教科書の統計に関わる部分は検定中である。AI 実装社会の到来でデータサイエンスの重要性が増し、データから意思決定・行動につながる統計的思考・判断・表現力を育成することが急務となっている。ICT の活用に関して学校教育への期待も強まり、文部科学省が進める GIGA スクール構想もその実現に向けて準備が進んでいる。具体的には学校で児童・生徒が1人1台の情報機器を自由に使える環境が整備されている。広島大学附属中・高等学校では、1995年に中高の授業で情報機器を活用するための情報館を設置し、授業開発に取り組んできた。2018年度よりスーパーサイエンスハイスクール事業（以降 SSH 事業と記す）に採択され、課題研究の活動を通して教員や生徒が ICT を活用する場面が増えた。現在では教職員全員がノート PC を貸与され、生徒が情報館以外の場所でも利用できるよう、ノート PC やタブレット PC の台数も増えた。その結果、課題研究以外の授業でも ICT を活用する授業が増えた。校内の Wifi 環境も整えている状況である。

このように学校内の設備面が充実する中、実際にどのように活用して指導を展開するかという問題が出てくる。最近では、感染症拡大による学校の休校や密を回避した指導を検討する中、情報過多にならないような適切な ICT の活用についても議論されている。

統計の指導に関しては、SSH 事業で学校設定科目を設定し、数学の教員が指導を担当してきた。当初は、理数に強い興味・関心を持つ生徒のためのカリキュラム開発であったが、最近では、数学と情報の協働による教科融合カリキュラムとして、「AS 統計科学」（科学者・技術者の文脈で必要な統計及びデータサイエンスの知識や技能を習得する）、「GS 社会と統計」（統計的な見方・考え方、概念を理解し、データやデータの分析を批判的に見る力を身に付ける）の指導実践に取り組んでいる。

また社会の変化に伴い、数学の捉え方が変わってきていることにも注目する必要がある。大学教育に関して、日本学術会議数理科学委員会数理科学分野の参照基準検討分科会（2013）は数理科学を「数学を中心とし、数学から生まれた統計学や応用数理などの分野と、数学教育や数学史など数学と他の学問分野との境界分野を合わせた学問分野」とした上で、「これからの時代の市民にとって、数理科学的な事象の把握・処理の能力は欠かせない。市民が正しい判断を行うためには、データに基づき物事を量的に把握することが必要不可欠であるが、そのような能力の涵養において、数理科学教育（算数・数学教育）が果たす役割は大きい。」と指摘している。この数学の捉え方の変化は、初等・中等教育における算数・数学の授業に影響を与え、次期学習指導要領では、中学校数学で「箱ひげ図」を指導し、高等学校数学 I で「仮説検定の考え方」を扱う等、統計の内容が充実している。現行では数学 B の内容である「期待値」も数学 A に移行し、「確率」の学習との接続ができるようになっている。

本稿は、SSH 事業の学校設定科目の指導実践から、数学 B の区間推定の指導例を紹介する。

2. 教材について

古語辞典の見出しの総数について推測する活動は、正規分布を用いた区間推定の活用として扱うが、中学校数学3の「標本調査」で学習した点推定を用いて考察することができる。生徒に既習内容とのつながりを意識させ、正規分布を用いることで何ができるようになったのかを理解させたい。

母平均を母集団の一部分である標本から推定する場合には、推定値を求めるだけでなく、推定値の誤差を明らかにすることが大切になるが、標本誤差の評価については発展的な内容となるので、ここでは扱わないようにする。無作為抽出による実験を通して、 n の値を大きくすると、信頼区間の幅が狭くなること、標本

の標準偏差の平均が一定の値に近づくこと等を確認することができる。複数のグループに分かれることで、推定の結果を比較できる。

本稿では、見出し約 23000 語収録と記載された古語辞典を用いた指導展開例を示す。無作為性を保証するために、調べるページの決定には乱数を用いることにする。調べるページ数は、制限時間内に可能な限りとする。学習指導案は表 1 の通りである。

表 1 学習指導案

時間	生徒の学習活動	授業者の指導・支援
導入 5 分	1 問題解決の見通しを立てる。	・古語辞典を示し、「見出しの総数を母平均の区間推定を利用して求める」という問題を提示する。 ・学習課題を、標本調査の計画を立てて実施すること、収集したデータから区間推定を行い、その構成要素の総数について推測することに設定する。
	課題 1 標本調査の計画を立てて、古語辞典の見出しの総数を調べよう。	
展開 10 分	2 乱数さいを利用したデータ収集の計画を立てる。	・無作為抽出の方法を確認し、乱数さいを利用してどのように調べるページを決めるかを話し合い、ルールを決める。
15 分	3 グループでデータ収集を行う。	・標本の大きさ n の値が大きい方が望ましいが、15 分間で調査できるページのデータを収集する。そのためグループにより調べるページ数が異なる。
	課題 2 収集したデータの標本平均から母平均を区間推定し、見出しの総数について推測しよう。	
10 分	4 収集したデータの標本平均から母平均の 95% の信頼区間を求める。	・ノート PC の表計算ソフトを利用させる。 ・各グループで求めた母平均の信頼区間を黒板に書かせる。
	5 求めた母平均の 95% の信頼区間から見出しの総数について推測する。	・古語辞典のページ数が 1423 であることから、母平均の 95% の信頼区間の各辺を 1423 倍して値の範囲を求める。
5 分	6 記載されている値と比較し、その違いについて考察する。	・見出し約 23000 と記載されていることを伝え、推測した結果と比較する。
まとめ 5 分	7 本時の学習を振り返る。	・学習を振り返り、標本調査の方法や結果をまとめる。

3. 授業の実際（2017 年 9 月実施）

(1) 課題 1：標本調査の計画を立てて、古語辞典の見出しの総数を調べよう。

データ収集前に、その収集方法を確認した。グループによって方法が異ならないように、ルール作りをすることが大切であると考えたためである。授業者と生徒の対話は次の通りである。

生徒 A この辞典は 13～1435 ページまで見出しがあり、古語辞典が全部で 10 冊しかないことも考えるとグループで分担して全ページの見出しを数えるのは無理だな。

生徒 B そうそう。時間内に終わらないよ。

生徒 C 乱数さいを利用するみたいだけど、千の桁まで決めるにはさいころが 4 ついるね。

授業者 どうやってページを決めますか？

生徒 C 4 つの乱数さい（図 1）で桁数を決めておいて、千の位は偶数を 0，奇数を 1 にすればいい。辞典にないページは無視ということ。



図 1 乱数さい

生徒 B そうしよう。

授業者 乱数さいを使って同じページに決まったらどうしますか？

生徒（複数） 2 回目以降は無視しよう。

生徒に活動させる場合には、特にルールの徹底に注意しなければならない。途中でルールを変更することで、それまでに収集したデータが使用できなくなる場合があるからである。また、データ収集の時間が長すぎると、結果を考察する時間が足りなくなり、1 時間で何を学んだかという印象が薄まってしまう危険性がある。乱数さいを使って乱数を発生させるときは、恣意性を排除し、無作為抽出を生徒に意識させるためにも、乱数さいを本やノート等に当てて転がすなどの配慮が必要である。

15 分程度の時間があれば、大体 40～60 ページの見出しの数を調べることができる。複数のグループに分けて活動させたため、時間の限り数えることにすると、各グループの標本数は異なる。

次に課題 2 を設定し、母平均を区間推定する。計算には、図 2 のように表計算ソフトを利用する。ノート PC 等の情報機器は、グループに 1 台ずつ準備した。

(2) 課題 2：収集したデータの標本平均から母平均を区間推定し、見出しの総数について推測しよう。

48	10	
49	18	
50	16	←見出しの数
n	"=COUNT(B2:B51)"	
平均 $\bar{x}$	"=AVERAGE(B2:B51)"	
標準偏差 s	"=STDEVP(B2:B51)"	
$\bar{x} - 1.96s / \sqrt{n}$	"=B53-1.96*B54/SQRT(B52)"	
$\bar{x} + 1.96s / \sqrt{n}$	"=B53+1.96*B54/SQRT(B52)"	

図 2 表計算ソフトの画面

実際に 10 のグループで、収集したデータから母平均 m の 95% の信頼区間を求めると、次の表 2 のように整理することができた。

表 2 各グループの求めた母平均 m の信頼区間

グループ	母平均 m の信頼区間 (95%)	グループ	母平均 m の信頼区間 (95%)
1 (50)	$14.0 \leq m \leq 16.4$	6 (61)	$13.8 \leq m \leq 16.0$
2 (44)	$13.6 \leq m \leq 16.1$	7 (58)	$14.2 \leq m \leq 16.4$
3 (48)	$15.2 \leq m \leq 17.4$	8 (47)	$15.0 \leq m \leq 17.2$
4 (45)	$15.4 \leq m \leq 16.7$	9 (63)	$13.5 \leq m \leq 16.2$
5 (40)	$11.4 \leq m \leq 16.7$	10 (55)	$14.0 \leq m \leq 16.6$

(グループの () はデータ数を表している、小数第 2 位を四捨五入)

この結果から、見出しの総数について、次のような生徒の発言を引き出すことができた。

授業者 見出しの総数について推測するにはどんな計算をすればよいですか？

生徒 D 古語辞典は全部で 1423 ページあるから、 m の不等式の各辺を 1423 倍すればいい。

授業者 母平均 m の 95% の信頼区間は、“1 ページあたりの見出しの数の範囲”を表しているため、1423 倍すると、見出しの総数について推測できます。計算してみましょう。

生徒 D 私のグループ 3 は $21629.6 \leq 1423m \leq 24760.2$ となります。幅が 3130.6 もあります。

授業者 幅を小さくするにはどのようにすればよいですか？

生徒 E 調べるページの数を増やすとよいです。

授業者 そうですね。母平均 m の信頼区間から、区間の幅は $2 \times 1.96 \cdot s / \sqrt{n}$ となります。 n の値を大きくすると、 s の値の変動は一般には小さいので、区間の幅は小さくなります。例えば、データ数が最も少ないグループ 5 は区間の幅が一番長いですね。

見出しの総数について推測した後に、見出し約 23000 と古語辞典に書かれていることを伝え、生徒から次のような反応が出た。

授業者 実はこの古語辞典の見出しは約 23000 だそうです。

生徒 D すごい。23000 は私のグループ 3 で計算した不等式の範囲に入っています。

生徒 E 私のグループ 8 も入っています。

授業者 皆さんのグループで 23000 は不等式の範囲に入っていますか？

生徒 E 僕たちのグループ 2 は、 $19352.8 \leq 1423m \leq 22910.3$ となり、23000 は不等式の範囲に入っていない。しかし、約 23000 なので僕たちのグループの不等式は全くの外れではないと思います。

授業者 見出しの総数について、不等式の範囲を図に表して比較してみましょう。(図 3)

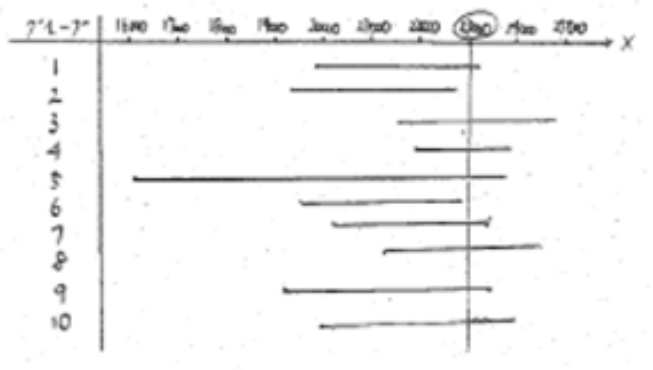


図 3 区間推定から推測した見出しの総数

生徒 F グループ 2 と 6 で 23000 が不等式の範囲の右側に外れています。

生徒 E (図 3 から) 標本が変わると不等式の範囲が変化することがわかります。また全てのグループで不等式の範囲は 23000 より左側によっています。

生徒 F これらの結果から見出しの総数は約 23000 といっても、実際は 23000 より少ないのかな。全てグループで 22000 は不等式の範囲に入っているしね。

授業者 区間推定することで、見出しの総数について不等式の範囲で考え、その結果の状況の比較から統計的には 23000 より少ない可能性があるという話になりました。

このように、結果を比較、検討することを通して、信頼度に関する理解を深め、標本調査の結果を批判的に考察することができた。また母平均を区間推定すると

きに標本平均と標本偏差を使用しているため、誤差を含めて推測していることも確認することができた。生徒は、母平均の真の値を知りたいとの要望が出たため、授業者が数えた数値を教えると盛り上がった。

(3) 授業の感想

授業後に生徒に「本時の活動の意義をどう感じましたか。」と聞いていると、次のような反応があった。

- ・予想した値の範囲に実際の値が入っているのがすごいと思った。また、入っていないグループの結果から何故かを考えるのが楽しかった。
- ・予想から、公表されている実際の値は本当かと思うようになった。確かめる方法を知ることができたのがよかった。
- ・ n の値を大きくすると、信頼区間の幅が狭くなること、標本の標準偏差の平均が一定の値に近づくことが分かった。
- ・見出し約 23000 語収録と記載された古語辞典の「約」という表現が絶妙だと思った。安易に信じてはいけないと思った。
- ・各グループの結果を表や図に整理すると、分かりやすくなったと思った。

4. まとめ・今後の課題

本稿では、高校数学で推測統計の内容をどのように指導するのかについて、区間推定の教材を基に指導例を紹介した。SSH 事業の学校設定科目では、高度な内容を含むため、教材や扱いの中にはそのまま数学の授業で使用できることとそうでないことがある。古語辞典の見出しの総数を推測するという課題は、1 時間で問題解決的な活動が展開できるものであり、すべての生徒にとって取り組みやすい設定になっているといえる。また表計算ソフトを利用する場合、ある程度の経験も必要となる。標準偏差を求めるなどの練習などは、数学 I の「データの分析」や情報の授業に期待できる。スマートフォンの普及等で、タッチパネルの操作には慣れているが、キーボードでの入力を苦手とする生徒も少なからずいるのが現状である。様々な授業の適切な場面で ICT を活用していくことが重要である。

また推測統計には、文字を含む数式が複数出てくる。式やグラフを丁寧に扱い過ぎると、内容を限られた授業時間内に収めるために、一方的に授業者が話す時間が増え、生徒の理解が進まない危険性がある。例えば、

正規分布の式には、自然対数 e (≈ 2.7) が出てくる。このように、既習内容だけでは説明できない箇所もある。そこで授業者は、式やグラフに関して、時には要点のみを示すことにして、その活用を扱うことも大切であると考ええる。数学の授業における指導の工夫が求められているとも言える。

今後は、問題解決的な推測統計の指導例を増やすとともに、授業を通して考え方をどのように育成するか、ICT をどのように活用するかについて検討したい。

5. 参考文献・参考 URL

- (1) 日本学術会議数理科学委員会数理科学分野の参照基準検討分科会 (2013), 報告 大学教育の分野別質保証のための教育課程編成上の参照基準 数理科学分野.
<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-h130918.pdf> 最終閲覧日 : (2021/02/19)
- (2) 日本学術会議 (2020), 提言 新学習指導要領下での算数・数学教育の円滑な実施に向けた緊急提言 : 統計教育の実効性の向上に焦点を当てて.
<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t228-4.pdf> 最終閲覧日 : (2021/02/19)
- (3) 文部科学省 (2019), 高等学校学習指導要領 (平成 30 年告示) 解説 数学編 理数編.
- (4) 渡辺美智子 (2011) 「科学的探究・問題解決・意思決定のプロセスを通して育成する統計的思考力」, 科学教育研究, 35(2), pp.71-83.
- (5) 渡辺美智子 (2014) 「不確実性の数理と統計的問題解決力の育成一次期学習指導要領の改訂に向けて」, 日本数学教育学会誌, 第 96 巻, 第 1 号, pp.33-37.