数学Iにおける「仮説検定の考え方」の指導

東京学芸大学附属高等学校紀要　57　pp.27～40，2019

数学Ⅰにおける「仮説検定の考え方」の指導

Teaching Hypothesis Testing in High School Mathematics I

数学科
大谷　晋　　青山　久美子　　井上　哲明　　荻原　洋介　　佐藤　亮太
祖慶　良謙　　田中　満城子　　野島　淳司　　吉岡　雄一

< 要旨 >
本校生徒の探究活動の質を上げるため、また、次期学習指導要領で数学Ⅰの「データの分析」導入される「仮説検定の考え方」への対応のため、数学Ⅰの時間と SSH 探究の時間を利用して、「仮説検定の考え方」の授業を行った。また、これまで多くの学校で採用されていなかった数学Ｂの「統計的な推測」が次期学習指導要領では多くの学校で採用される見込みである。それに対応するため、数学Ⅰで「区間推定」の授業を行った。評価問題を採点した結果では、高校1年生でも「仮説検定の考え方」は十分理解することができた。しかし、探究活動等現実場面で活用できるようになったかということに関しては課題があると考えられる。

< キーワード > 仮説検定　仮説検定の考え方　区間推定　探究活動

1　はじめに

近年、統計教育の重要性が強調されている。前回の次期学習指導要領の改訂では、数学Ⅰに「データの分析」が入った。そして、さらに次期学習指導要領では、数学Ⅰの「データの分析」に、「ア(ウ)具体的な事象において仮説検定の考え方を理解すること。」、「イ(ウ)不確実な事象の起こりやすさに着目し、主張の妥当性について、実験などを通して判断したり、批判的に考察したりすること。」という項目が入った。仮説検定は、これまでは数学Ｂの中で、確率変数などを学習した後で扱っていたが、次の改訂では、数学Ⅰで仮説検定の考え方を扱うことになる。

また、数学Ｂはこれまで、(1)確率分布と統計的な推測、(2)数列、(3)ベクトルで構成されており、内容を適宜選択することになっていたので、多くの学校では数列とベクトルを選択していたと考えられる。逆に言えば、「確率分布と統計的な推測」を学習していた高校生は少なかったといえるだろう。次期改訂により、数学Ｂの内容は(1)数列、(2)統計的な推測、(3)数学と社会生活の３つとなり、適宜選択するとなったことから、「統計的な推測」を選択する学校が増えることになる。

数学Ⅰは必履修科目であるから、少なくとも全ての高校生は「データの分析」を学習し、さらにこれまでよりかなり多くの生徒が「統計的な推測」まで学習することになる。その意味では、次期学習指導要領は、統計教育が重要であるという社会の要請に応えたものであると言ってよい。しかし、教えようによっては、ただ統計嫌いを作ってしまうだけで、生きて働く知識とはならないだろう。生徒が統計的探究プロセスを働かせて、学習するような授業が展開される必要がある。

本校では、1，2年次にそれぞれ１単位 SSH 探究が設置され、１年生は探究活動の基礎を学習し、２年生では各自がテーマを設定して１年間かけてそのテーマを探究することになっている。２年生の探究活動について、生徒の発表を見ると、実験などをやった場合に、その統計処理が十分にできているとは言い難いという現状があった。例えば、実験群と対照群を作って実験して平均値の比較から単純に結論を導き出していて、有意性の検定などが行われていないなどである。

以上のような本校の SSH 探究の現状を打破すべく、また、学習指導要領の改訂も見据えて、１年生に「仮説検定の考え方」の授業を実施することを試みた。ここでは、平成 30 年度、令和元年度に実施した本校の公開教育研究大会の記録を中心に記述することとする。

2　仮説検定の授業

学習指導要領解説（平成 30 年告示）では、仮説検定の考え方について以下のように記述されている。

例えば，「ある新素材の枕を使用した 30 人のうち 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

然に起こりえ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 回投げるという試行を繰り返す。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「起こりえないこと」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。この「起こりえないこと」かどうかの基準として, 平均から 2s（s は標準偏差）あるいは 3s 離れた値を用いることが考えられる。この考え方を数学的に精緻化していくと,「帰無仮説 : 新素材の枕はよく眠れる効果がなかった」を確率分布を用いて検定する「数学 B」の内容につながる。

ここでは、コインを 30 回投げるという試行を繰り返し、その結果から 24 枚以上表になる事の起こりやすさを判断するということになっている。数学 A の「確率」を学習するという前提がないので、実験等から起こりやすさを判断することになるのであろう。しかし、本校では、数学Ⅰと並行して数学 A を学習しているので、「確率」の知識は前提とできる。また、探究活動で生徒が使うことを考えると、本校の１年生全体が学習する内容としては、正規分布と仮定できる場合の起こりやすさについて判断できることが必要だと考えた。また、3 時間程度という限られた時間であるため、教える内容は「仮説検定」ではなく、あくまで「仮説検定の考え方」であるということとした。つまり、「帰無仮説」、「有意水準」などについては深入りをせず、起こりやすさの程度を判断できるようになることを目標とした。

「仮説検定の考え方」　学習指導案

本授業の要旨

試験の結果において、学年の平均値とクラスの平均値が異なる場合に、それをどのように評価し判断したらよいかを考えさせる授業である。

１．研究主題との関わり

⑴　本校で育てたい生徒像と，本単元で育てたい「資質・能力」との関係

本授業は、本校で重視する５つの「資質・能力」、「A 課題を発見する力」、「B 科学的なプロセスで問題解決する力」、「C 発信する力」、「D 展望・計画をもつ力」、「D 展望・計画をもつ力」のうち、A,B,C の育成に関連する。自ら、課題を発見して、これまでに学習したことなどを用いて、科学的なプロセスを踏んで問題を解決し、自分の考えを他者に分かりやすく論理的に発信しようとする姿勢を育成したい。

⑵　育てたい「資質・能力」を評価する方法

評価はテストによる（評価問題は別紙）

２．対　象　　　１年

３．単元名　　　統計的判断（検定の考え）

４．単元の目標

試験の結果において、学年の平均値とクラスの平均値が異なる場合に、適切な方法を用いて適切に判断できるようになる。起こった事柄が、偶然によると考えのが妥当か、偶然とは考えられないとするか。

５．単元設定の理由

⑴　生徒たちの実態および本単元に至るまでの学習

数学Ⅰ「データの分析」、数学 A「場合の数と確率」（確率は、現在学習中）

⑵　教材の特性と授業者の手立て

コンピュータを利用して、シミュレーションをできるようにしておく。

６．指導計画

⑴　単元計画（土曜探究は２，３時間目）

０時間目　正規分布について（身長、実力テストの得点、ポアンカレのパン屋）

１時間目　数学の試験でクラス平均点が、学年平均点より高いことをどう評価するのか

２時間目　母集団からサイズ n の標本を無作為抽出した時に、標本の平均値と標準偏差はどのようになるか（検定の方法の説明）

３時間目　探究における定量的なデータの分析

7. 指導計画

⑴　単元計画

数	内容	題材
1	ヒストグラム，5 数要約	10 点満点のテストの 20 人の得点
2	箱ひげ図	〃
3	偏差，分散，標準偏差，偏差値	〃
4	正規分布，データの変換	〃
5	データ収集。代表値，箱ひげ図での分析	10 秒選手権
6	代表値，箱ひげ図での分析	〃
	データの数を増やすと正規分布に近い	
7	散布図	都道府県別人口，高校数，平均気温，降雪量など
8	共分散	5 点の散布図
9	相関係数	〃
10	PC 使用での分析①	Excel，Geogebra
11	PC 使用での分析② レポート作成	〃
12	相関係数の特徴	新体力テスト
13	PC 使用での分析③ レポート作成・提出	Excel，Geogebra
14	データの分析演習	仮平均，データの変換
15	正規分布の利用	身長，実力テスト，正規分布表，ポアンカレ
16	レポート発表	

①１時間目のねらい

試験の結果において、学年の平均値とクラスの平均値が異なる場合に、適切な方法を用いて適切に判断できるようになる。

②１時間目の授業展開

時間	学習の流れと生徒の活動	教員の指導と手立て
導入 5 分	数学の試験で 1 年生の平均点は 75 点で、C 組の平均点は 80 点でした。C 組は優秀なのでしょうか？	「C 組は本当に優秀なのか、それとも、今回の結果は偶然に過ぎないのか？」を考えるように方向づける。
展開 1 25 分	グループに分かれて検討する。 （生徒たちは、最初は 80 点を取った C 組の平均的な生徒が全体の分布の中で、どの程度優れているかを考えると予想される。偏差値を計算する生徒もいると予想される。）	ばらつきについて聞かれたら、標準偏差の数値（学年の標準偏差は 17）を答える。他のクラスの平均点はわからないとして教えない。 1 人が 80 点をとることと、クラス全体の平均点が 80 点であることを同じように考えて良いかを投げかける。
展開 2 20 分	コンピュータを利用して、学年全体 342 名のデータから、C 組の人数の 43 のデータを無作為抽出してみて、得点の平均値がどのようになるか観察させ、平均値が 80（80 以上）になることは偶然と考えてよいかどうかを考えさせる。	3420 個のデータから 43 個のデータを無作為抽出するようなシミュレーションができる電子ファイルをあらかじめ準備しておく。 2 人で 1 台程度のコンピュータを利用して、操作する。

⑵　本時の学習（２と３/３ 時間目）

①本時のねらい

探究活動において何らかの主張を行うためには，多くの場合定量的なデータによって根拠を示すことが不可欠であると言えます。本講座では探究活動におけるデータの活用法の基礎を学ぶとともに，具体的なデータを使った演習を行い，自身の探究活動における定量的なデータの分析の計画を適切に立てることができるようになることを目標とします。

【キーワード】定量的，グラフ，分布の比較，相関関係と因果関係，推測統計

【パフォーマンス課題】自ら探究活動に資する定量的なデータの分析の計画を立てる

②２，３時間目の授業展開

時間	学習の流れと生徒の活動	教員の指導と手立て
0	【復習】 前時は，342 名のデータから、43 名程度のデータを無作為抽出してみて、得点の平均値がどのようになるかを調べた。どうなりましたか？ サンプルの平均の平均は，約 75 点。理論上は母平均と同じ。 分布は？正規分布の形 →サンプルの平均値はばらつきが少ない どれぐらいばらつきが少ない？ → 1/5 から 1/10。理論上は $1/\sqrt{43}$。$17.4\times\frac{1}{\sqrt{43}}\fallingdotseq 2.65$ 上記のことを知っていれば,「43 人の平均 80 点がどれぐらいの確率で起こるか」を，サンプルを取らなくても，計算できる。 （自力解決（数分）） $$\frac{80-75}{2.65}\fallingdotseq 1.89$$ 正規分布表より，43 人の平均 80 点以上である確率は, $$0.5-0.4706=0.0294=2.9\%$$ この結果から，平均が 80 点以上である確率は 2.9% と低い。よって，平均が 80 点であることは偶然ではなく，意味がありそうである。これを,「このクラスの平均が 80 点であることは有意に高い」という。	スライド エクセルで実演 データ数 43 ではなく データ数 10 など 正規分布表配布。 正規分布表の見方を確認する 語「有意に」の使い方
	何％で区切るか？ 一般的には，５％や１％。文脈による。 基準を有意水準という。 以下，有意水準を５％とする。	
20	【まとめ】 もとのデータ（母集団）の分布がどんな分布でも，サンプル（標本）の平均の分布は，標本数が多ければ，正規分布の形になる。 その場合， 標本の平均の平均は，母集団の平均 標本の平均の標準偏差は，母集団の標準偏差 $\times 1/\sqrt{\text{標本の大きさ}}$	

25	【練習問題】 25 人の平均値が 80 点であった。この 80 点は有意に高いか？ （自力解決（数分）） 25 人の平均の平均は，75 点 25 人の平均の標準偏差は，$17.4\times\frac{1}{\sqrt{25}}=3.48$ $$\frac{80-75}{3.48}=1.44$$ 正規分布表より，25 人の平均 80 点以上である確率は， $$0.5-0.4251=0.0749=7.5\%$$ この結果から，25 人の平均値が 80 点であることは，有意に高いとは言えない。	
30	【問１】お菓子の容量の表示は本当なのか？ （母分散未知のため標本の分散で読み替える。理系文系問わず） A 君は 100g 入りとして売られている大袋のポテトチップスに，本当に 100g 入っているのか疑問に思った。そこでこのポテトチップス 50 袋を購入してその中身の重さを調べた。すると，50 袋について中身の重さの平均は 99.80 g，標準偏差は 1.50g であった。この結果からこのポテトチップスの内容量は 100g よりも有意に少ないと言えるか。	
	（自力解決）	
37	【前問との違い】 母平均は 100 g 母集団の標準偏差はわからないが，標本の大きさが 50 と大きいので標本の標準偏差 1.50g と同じとして考えてよい。	前問との違いは，母集団がない。 母集団をどうするか？ 標本の大きさが 30 以上は大きいと知られている。
45	（自力解決） 母集団から 50 個の標本をとった平均値を X とする。繰り返し標本をとるとき，X の平均値は，母平均と同じで 100 X の平均値の標準偏差は $$1.5\times\frac{1}{\sqrt{50}}\fallingdotseq 0.212$$ このとき， $$\frac{99.8-100}{0.212}\fallingdotseq -0.9434$$ 正規分布表より 99.80g 以下である確率は $$0.5-0.3264=0.1736$$ よって無作為にこのポテトチップス 50 袋を購入したときに，その内容量が 99.80g 以下であることは 17.4％程度の確率で起こる。このことから，ポテトチップスの内容量は 100g よりも有意に少ないとは言えない。	
0	【実験】 【問２】植物にマイクロ波を当てると成長が早くなる？ 「植物にマイクロ波を当てると成長が早くなる」という仮説を検証するための実験をおこなった。植物 A の 50 個の種子から出た芽のうち 10 個にはマイクロ波を当て，他の 40 個には当てずに育てた（その他の生育条件は同じとした)。1 週間後，マイクロ波を当てずに育てたもののうち，きちんと育たなかったもの 5 個を除いた 35 個の大きさの平均は 8.8 cm，標準偏差は 1.2cm であった。一方でマイクロ波を当てて育てたもののうち，きちんと育たなかったもの 2 個を除いた 8 個の平均は 9.4 cm，標準偏差は 1.5 cm であった。 この結果から，植物 A はマイクロ波を当てると成長が早くなると結論づけてよさそうか。 （自力解決）	母集団の情報がないので標本の情報で読み替える。理系

5	母集団についての情報はないが，マイクロ波を当てずに育てたもののうちきちんと育たなかったものを抜いた 35 個についての情報を母集団に関する情報として読み替える。つまり母集団の平均は 8.8 cm，標準偏差は 1.2 cm と考える。 （自力解決）	母集団をどうするか？
10	条件をかえて得られたデータは 8 個なので，母集団から取った 8 個の平均値を X とする。繰り返し標本をとるとき， X の平均は，マイクロ波を当てなかったときの平均と同じで 9.4 cm であると考える。 X の標準偏差は $$1.2 \times \frac{1}{\sqrt{8}} \fallingdotseq 0.424$$ このとき， $$\frac{9.4-8.8}{0.424} \fallingdotseq 1.42$$ 正規分布表より 9.4cm 以上である確率は $$0.5-0.4222=0.0778$$ つまり，取った 8 個の平均が 9.4 cm 以上あるということは 7.8%の確率で起こり，これは十分に小さい確率であるとは言えない（たまたま起こるということが十分にあり得る）。すなわち，この実験結果だけでは植物 A にマイクロ波を当てると成長が早くなると結論づけることは適当ではない。 結論づけるためにはどうすればよい？ →データの数を増やすことが必要。 平均，標準偏差が同じであれば，いくつのデータが必要か？ $$\frac{9.4-8.8}{1.2 \times \frac{1}{\sqrt{n}}} > 1.64$$ $$\sqrt{n} > 3.3$$ $$n > 10.89$$	 式と結果のみを表示
20	【アンケート】 【問 3】学生は早く働きたくない？ 現在の学生の持っている就職に対する意識を調査するために，無差別に選出した学生（中高大生）80 名を対象に実施。 「早くはらたきたいと思うか？」 結果：Yes：45%（36 人），No：55%（44 人） このことから，「学生の半数以上が早く働きたいと思わないという現状がある。」と言ってよいか？ （自力解決）	問 3 は時間あればやる
25	よいと思う人？ よくないと思う人？なぜよくないと思う？ →思わない現状や思う現状がない場合，Yes か No かは 1/2。 そう仮定すると，44 人が No である確率は？ （自力解決）	

30	$${}_{80}\mathrm{C}_{44}*\left(\frac{1}{2}\right)^{80}=0.060$$ 44 人以上が No である確率は？ $$\left({}_{80}\mathrm{C}_{44}+{}_{80}\mathrm{C}_{45}+\cdots+{}_{80}\mathrm{C}_{80}\right)*\left(\frac{1}{2}\right)^{80}=0.217$$ No が 44 人以上になる確率は 21.7% であるので，これは十分に小さい確率であるとは言えない。(たまたま起こるということが十分にあり得る)。すなわち，このアンケート結果だけでは「学生の半数以上が早く働きたいと思わないという現状がある」と結論づけることは適当ではない。	
34	どうすれば良い？ →聞く人数を増やす。→何人？ 400 人中 55%（220 人）が No の場合 $$\left({}_{400}\mathrm{C}_{220}+{}_{400}\mathrm{C}_{221}+\cdots+{}_{400}\mathrm{C}_{400}\right)*\left(\frac{1}{2}\right)^{400}=0.026$$ No が 220 人以上になる確率 2.6% であるので，これは十分に小さい確率である。このことから，学生の半数以上が早く働きたいと思わないという現状がある」と言って良い。	
35	【問 4　評価問題】附高生は忙しい？ ある調査によると全国の高校生 2500 人に対して「あなたは忙しいか」という質問をおこなったところ,「はい」と答えた生徒は 1750 人（70%),「いいえ」と答えた生徒は 750 人（30%）であった。 同じ調査を学芸大附属高校の 1 年生 342 人に対しておこなったところ,「はい」と答えた生徒は 256 人（≒ 75%),「いいえ」と答えた生徒は 86 人（≒ 25%）であった。 このアンケートの結果から，学芸大附属高校の生徒は忙しい（忙しいと感じている）生徒の割合が有意に高いと判断してよいか。	母分散を求めるのに二項分布的な計算が必要。文系 複雑な計算はしなくてよい。 電卓で可能なら計算。 式を立てるだけでもよい
	方法 1 「はい」を 1,「いいえ」を 0 として考えると， 母集団の平均は 0.70 母集団の分散は $$\frac{(1-0.7)^2\times 1750+(0-0.7)^2\times 750}{2500}=0.21$$ 母集団の標準偏差は $\sqrt{0.21}\fallingdotseq 0.458$ ―――以上は言うが，以下は言わない。 母集団から 330 個の標本をとった平均値を X とする。繰り返し標本をとるとき， X の平均は，母平均と同じで 0.7 X の標準偏差は $$0.458\times\frac{1}{\sqrt{342}}=0.0248$$ このとき， $$\frac{0.75-0.7}{0.0248}\fallingdotseq 2.02$$ 正規分布表より「はい」の割合が 75% 以上である確率は $$0.5-0.4783=0.0217$$	

	つまり全国の高校生から無作為に 342 人を抽出したとき，「はい」の割合が 75%以上になる確率は 2.2%程度しかない。このことから，学芸大附属高校の（少なくとも 1 年生は）忙しいと感じている生徒の割合が有意に高いと言える。 方法２ 附高生も全国の高校生と同じく 0.7 の割合で「はい」と答えるとすると 256 人が「はい」と答える確率は， $${}_{342}\mathrm{C}_{256} \times 0.7^{256} \times 0.3^{342-256} = 0.0068$$ 256 人以上が「はい」である確率は？ $${}_{342}\mathrm{C}_{256} \times 0.7^{256} \times 0.3^{342-256} + {}_{342}\mathrm{C}_{257} \times 0.7^{257} \times 0.3^{342-257} + \cdots + {}_{342}\mathrm{C}_{342} \times 0.7^{342} \times 0.3^{342-342} = 0.0272$$ つまり全国の高校生から無作為に 342 人を抽出したとき，「はい」の割合が 75%以上になる確率は 2.7%程度しかない。このことから，学芸大附属高校の（少なくとも 1 年生は）忙しいと感じている生徒の割合が有意に高いと言える。	計算にはコンピュータ等が必要
	探究するにあたり 実験やアンケートのデータ数に注意してほしい 論文や本を読む際，統計の知識があればもっとわかる。	

(3)　評価基準（ルーブリック）

	4	3	2	1
２つのデータの間に有意な差があるかどうかを適切に判断できる。	２つのデータの間に有意な差があるかどうかを適切な方法を用いて判断できるととともに，それを明確な論理で説明している。	２つのデータの間に有意な差があるかどうかを適切な方法を用いて判断しようとする。	２つのデータの間に有意な差があるかどうかを，確率を用いて判断しようとする。	２つのデータの間に有意な差があるかどうかを，分布から確率を求めることによって判断できることを理解している。
参考：パフォーマンス課題に対する取組の例	標本平均が 0.75 以上となる確率（256人以上が yes と答える確率）を明瞭な論理の記述をもとに適切に求め，それを根拠にして妥当な結論を導いている。	標本平均が 0.75 以上となる確率（256人以上が yes と答える確率）を適切な方法で求めようとしているが，(論理的な説明が不十分である or 計算などに誤りがある)。	標本平均が 0.75 以上となる確率（256人以上が yes と答える確率）を求めようとしているが，方法に間違いが見られる（標準偏差を適切な方法で求めていない等)。	評価問題には適切に取り組めていないが，問１～３に対して確率，分布，標準偏差，有意などの用語を適切に用いてまとめることができている。

この授業は、平成 30 年度の公開教育研究大会において大谷晋が最初の 1 時間を実施し、残りの 2 時間は SSH 探究の時間に実施した。そして、令和元年もほぼ同じ内容の授業を継続して、数学Ⅰの 1 時間と SSH 探究の 2 時間で実施した。基本的には、今後も同じように実施する予定である。

また、令和元年の公開教育研究大会では、井上が「区間推定」の授業を高校 1 年生に実施した。これは、数学Ｂの「統計的な推測」に相当する内容である。ただし、２年次の探究活動で利用することを考えると、1 年生のうちに授業をしておいた方が望ましいと考えられる。次にその学習指導案を掲載する。

「区間推定」学習指導案

本授業の要旨

新学習指導要領において、検定の考えが数学Ⅰで取りあげられるなど、統計的分野が大きく取り扱われることになる。現在は数学Ｂの中で選択できる内容の一つとして位置付けられているが、実際に授業で扱うことはほとんどない。社会現象等を自ら設定した仮説のもとで、数量的に分析する力の育成がこれからはますます必要とされる中、そのための授業をどう展開できるかを考えたい。

１． 研究主題との関わり

⑴ 本校で育てたい生徒像と，本単元で育てたい「資質・能力」との関係

本校で育てたい生徒像の中で、②適切な情報収集・分析能力と課題発見能力がある。この単元はまさに、その能力を育てるために適切な題材であると言える。評価基準においては、特に「科学的プロセスで問題解決する力」として、統計的手法を用いて現実的な問題場面を解決していく積極的姿勢を育て、その考え方を身につけさせたい。

⑵ 育てたい「資質・能力」を評価する方法

授業の中で、現実問題を解決するための統計量を理解し、それを用いて実際に問題を分析して解決することができたかどうかをみたい。グループにより問題解決活動を行うので、協働的活動によりお互いの考えを共有しながら、最適な解決方法を一緒に見出せるかもみたい。

２．対　象　１年Ｇ組（男子 21 名､女子 21 名計 42 名）

３．単元名　データの分析－推定

４．単元の目標

①具体的な事象において仮説検定の考え方を理解すること（数Ⅰ新学習指導要領より）

②統計的推測について、数学的活動を通してその有用性を認識するとともに、次の事項が身につけることができるように指導する。

ア．正規分布を用いた区間推定及び仮説検定の方法を理解すること

イ．目的に応じて標本調査を設計し、収集したデータを基に PC など用いて、母集団の特徴や傾向を推測し判断するとともに、標本調査の方法や結果を批判的に考察すること。

（数Ｂ新学習指導要領より抜粋）

５．単元設定の理由

⑴ 生徒たちの実態および本単元に至るまでの学習

全体的に大人しいクラスであるが、数学を苦手としている生徒は多い。本時まで、数学Ⅰのデータの分析を学習してきており、その継続として、確率分布（二項分布と正規分布）、正規分布を用いた確率の計算を学習してきている。

⑵ 教材の特性と授業者の手立て

先にも述べたが、この教材は数学Ｂに含まれており、扱う時間は確保できていない。学習指導要領改訂に伴い、数学Ⅰで取り扱うこととなった。仮説を立てて事象を分析したり、数量的に可能性を検証したりすることは、実社会では必要な力であるので、ここでしっかりと身につけさせたい。

PC 等をツールとして用いて、確率分布について学習し､正規分布についても基本的な事項は事前に学習した。

６．指導計画

⑴ 単元計画（全 20 時間）

①代表値・ヒストグラム

②箱ひげ図によるデータの分析

③分散と標準偏差

④相関係数

⑤回帰直線

⑥二項展開と二項分布 , 重複試行の確率（数Ａ）

⑦確率分布（二項分布・正規分布）

⑧正規分布での確率計算

⑨標本調査

⑵ 本時の学習（20 時間目）

①本時のねらい

選挙の出口調査を例に、標本比率から母比率を推定して、当落の可能性を予想する。どのように母比率を推定していけばよいかを考えさせたい。

②本時の授業展開

時間	学習の流れと生徒の活動	教員の指導と手立て
導入	問題場面の提示：ある選挙区でのある標本（100）のＡ氏への投票数が 60 であったとき、その選挙区でのＡ氏への得票率を推定したい。	・この選挙区でＡ氏の得票率が 0.5 を超えているかどうかを推定する方法を探ることを目標にする。 ・標本比率が変数であるという考え方ができるように促したい。
展開	1 標本比率の分布を理解する。 2 母比率が 0.5 であると仮定すると、実現比率 0.6 はどの程度で起こりうるかを計算してみる。 →母比率をいろいろと変えて、同じ計算をしてみる。 3 この結果から、母比率を推定する方法を考える。	・得票数が二項分布に従うことを理解させる。 ・二項分布が正規分布に近似できることを確認する。 ・得票率が正規分布に従うことを確認する。 ・正規分布での計算を促す。 ・信頼度の導入 ・グループでの作業になる。巡視しながら、生徒の理解をみる。 ・母比率を仮定した標本比率の確率分布において、確率 95% で出現する比率の範囲と実現比率の関係を考える。
結論	標本比率を確率変数とみることが重要であり、標本比率をある確率で含む母比率の範囲が推定するものであることを確認する。	・実現比率からみて、母比率の存在する範囲を考えてみる。 ・母比率≠標本比率であること ・従来の方法で求める結果との比較（PC） ・信頼度をいろいろと変えてみて、その意味を再確認する。

⑶評価基準（ルーブリック）

	4	3	2	1
A 課題を発見する力	実際に、課題から自分の問題を特定している。	そのために、何をする必要があるかを理解している。	推定するものが何かを理解している。	何が問題であるかを理解できない。
B 科学的なプロセスで問題解決する力	仮説を設定し、母比率の範囲を求めることができる。	母比率を推定するための仮説を設定する。	標本比率を確率変数であること理解している。	標本比率を答えとし、そこから先には進まない。
C 発信する力				
D 展望・計画をもつ力	設定した計画を実行し、振り返りを行うことにより、よりよい活動へ繋げる。	自らの考えで計画を練り、実行しようとする。	教師の示唆により、計画を作ろうとする。	計画を作ろうとしない。
E 関係を構築する力，協働する力	他者と考え方を批判的に検討し、よりよい提案を行うことができる。	自分の考え方を発表し、他者と比較検討している。	他者との協働作業に参加している。	他者との協働作業に参加しない。

7．主な参考文献および資料

・高等学校　学習指導要領（平成 30 年度告示）文部科学省
・確率・統計　小松勇作編　旺文社
・高校からの統計・データサイエンス活用―上級編―総務省

資料

1　授業の構成

テ ー マ：ある選挙区で A 氏の当落を標本抽出により推定する。

問題場面：選挙において、出馬した A 氏の当落を予想するために、ある選挙区において、投票を終えた有権者を無作為に 100 人選び、A 氏に投票したか否かを調査した。その結果、60 名の人が支持していることが分かった。この結果から A 氏の当落を予想せよ。

場面　問題の理解

母集団の支持者数が過半数であれば、この選挙区で有利な状況であることとがわかる。他の選挙区でも同じことを調査すれば A 氏の当落が予想できる。各選挙区での A 氏の得票率が予想できれば良い。

予想　標本比率が 0.6 であるから、この選挙区の母比率もおよそ 0.6 とみなしてもよいのではないか？

→標本数がもっと多ければ、標本比率で点推定しても良いと思われるが、この数ではどうだろうか？

→区間で推定する必要性の理解

場面　標本比率を確率変数と考える。

この比率 p はどんな確率分布に従うと考えられるか？

→二項分布から、正規分布へ近似する過程を確認する

場面　母比率 p の推定：推定値の範囲を求める。

方法を考える。標本比率（実現比率）が 0.6 ということは有利であると考える生徒は多いはず。

→推定するときには、信頼度というものが必要になる。もしも、信頼度 100% とすると、比率の分布からもわかるように母比率がいくつであっても実現比率 0.6 は出現するので、母比率の推定はできないことになる。そこで、この信頼度（出現する確率）を 95% にして考えてみることを促す。

→母比率を仮定して考える。母比率が 0.5 であると仮定し、この仮定のもとで、確率 95% で出現する比率の値の範囲を求めてみよう。生徒と一緒に、計算してみる。

場面　グループ活動

母比率をいろいろと変えて、確率 95% で実現比率の生起する可能性を計算して、求めた結果について、グループごとに話し合う。母比率はどの範囲にあると考えられるかについて議論し、その理由についても話し合ってみる。

場面　グループ活動の共有

上記の場面での結果を発表してもらう。

場面　母比率の範囲を推定する値の範囲の予想

母比率を、0.52 , 0.55 , 0.6 , 0.65 , 0.7 .0.45 など生徒が予想に必要なものを設定して計算して範囲を推定する。一般的に、母比率 p を計算する生徒のグループもいるかもしれない。

場面　母比率の推定予想の集約

生徒から発表してもらう。

場面　母比率の推定：

標本比率は、母比率を p とすると、$N\left(p,\ \sqrt{\frac{p(1-p)}{100}}^{2}\right)$ に従うので、95% の信頼度で推定するならば、実現比率 0.6 について，$-1.96 \leq \frac{0.6-p}{\sqrt{\frac{p(1-p)}{100}}} \leq 1.96$ が成り立つから、この解を求めれば良い。

→二次不等式への変形も可能であるが､時間がなければ、ここは PC で算出する。

$0.502021<p<0.690583$ がその解となり、

$P(0.502021<p<0.690583)=0.95$ を意味する。

結論　この選挙区では、A 氏の得票率は 50% を超えていると考えられる。

（注）　99% の信頼度では、母比率の推定値は、
$0.4712<p<0.7163$ となる。
90% の信頼度では、母比率の推定値は、
$0.517551<p<0.677148$ となる。

50% では、$0.5668<p<0.6323$ となる。

（注）

⑴ 一方、教科書の方法によると、s を標本比率から算出する偏差として、

$$P\left(\overline{x_1} - 1.96 \times \frac{s}{\sqrt{m}} \leq m \leq \overline{x_1} - 1.96 \times \frac{s}{\sqrt{m}}\right) = 0.95$$

（信頼度 95%）を用いる。

これより、$0.504<p<0.696$ を推定値と考えることができる。

⑵ちなみに、母比率 p のときの、偏差 $\frac{\sqrt{p(1-p)}}{10}$ に着目すると、これは、$p=1/2$ のときに最大値 0.05 をとるこ

とがわかる。

⑶母比率の推定値の最小値が 0.5 を超えていれば、過半数の支持があると確率 95% で判断できるので、その部分だけに着目して考察すれば良い。

⑼母比率は標本比率と一致するものではない。

標本の取り方によって、比率は変化する。標本比率は確率変数である。

それは、どのような確率分布に従うのか？

⑴母比率 p を仮定して、標本を抽出したとき、標本比率は N(p, p(1-p)/100) の正規分布に従うことを確認する。確率分布のモデルを作る。

⑵標本比率から母比率を推定するとき、ある幅を持って考える。

→ 標本比率＝母比率ではないことの理解

→ A ＜母比率＜ B であること。

この A,B の値を推定したい。

→母比率を 100% の確率で定めるのがベスト。

母比率 0.5 が 100% で定まるとは、「そのときの標本比率の確率分布で確率 100% で出現する比率の範囲に実現比率 0.6 が含まれている」ことをいう。

仮に、母比率を 0.5 とすると、標本比率の確率分布は $N(0.5, 0.05^2)$ に従うので、この分布で確率 100% で出現する比率を求めると、それは実は何でもよく、実現比率 0.6 は当然、その中に含まれる。母比率をどう仮定しようが、そのときの標本比率の確率分布において、買う率 100% で出現する比率は何でも良いことにかわりはない。つまり、実現比率 0.6 は含まれるわけである。

結論として、100% で母比率を推定することはできないということである。

→そこで、多少の危険性を覚悟して 95% で推定してみよう。95% の信頼度で求めるということである。このとき、母比率はどんな範囲に定まるだろうか？

p= 0.5 から始めて、確率 95% で出現する比率の範囲を求めてみよう。

①母比率から標本比率の確率分布を作る。

②母比率を 0.5 に仮定したとき、信頼度 95% で出現しうる比率 p の範囲を求めよう。

→$-1.96 < \frac{p-0.5}{\frac{\sqrt{0.25}}{10}} < 1.96$　つまり、

$$0.5 - 1.96 \times 0.05 < P < 0.5 + 1.96 \times 0.05$$

よって、0.402<p<0.598　実現比率はこの中には含まれない。

③母比率を 0.55 に仮定したとき、信頼度 95% で出現する比率を求めよう。

→$-1.96 < \frac{p-0.55}{\frac{9.95}{200}} < 1.96$　つまり、

$$0.55 - 1.96 \times \frac{9.95}{200} < P < 0.55 + 1.96 \times \frac{9.95}{200}$$

よって、0.4525<p<0.6475 実現比率はこの中に含まれる。

つまり、母比率 0.55 は確率 95% で起こりうることを示している。

実現比率 0.6 から考えて、この母比率 0.55 は確率的に十分起こりうる値と判断してよい。

④母比率を 0.52

→$-1.96 < \frac{p-0.52}{0.05} < 1.96$　つまり、

$$0.52 - 1.96 \times 0.05 < P < 0.52 + 1.96 \times 0.05$$

よって、0.422<p<0.618 実現比率はこの中に含まれる。

⑤母比率を 0.6

→$-1.96 < \frac{p-0.6}{0.049} < 1.96$　つまり、

$$0.6 - 1.96 \times 0.049 < P < 0.6 + 1.96 \times 0.049$$

よって、0.504<p<0.696 実現比率は含まれる。

⑥母比率を 0.65

→$-1.96 < \frac{p-0.65}{0.048} < 1.96$　つまり、

$$0.65 - 1.96 \times 0.048 < P < 0.65 + 1.96 \times 0.048$$

よって、0.556<p<0.744 実現比率は含まれる。

⑦母比率を 0.7

→$-1.96 < \frac{p-0.7}{0.046} < 1.96$　つまり、

$$0.7 - 1.96 \times 0.046 < P < 0.7 + 1.96 \times 0.046$$

よって、0.61<p<0.79　実現比率は含まれない。

⑧母比率を 0.45

→$-1.96 < \frac{p-0.45}{0.04975} < 1.96$　つまり、

$$0.45 - 1.96 \times 0.04975 < P < 0.45 + 1.96 \times 0.04975$$

よって、0.353<p<0.545 実現比率は含まれない。

以上より、信頼度（確率）95% で、母比率は 0.52 より大きく、0.7 よりは小さいことが推定できる。

⑶母比率の推定

母比率を p とするとき、実現比率 0.6 が信頼度（確率）95% で出現するとする。

このとき、$-1.96 < \frac{0.6-p}{\frac{\sqrt{p(1-p)}}{10}} < 1.96$ が成り立つ。

これを解いて、0.502021<p<0.690583 となる。

母比率の最小値が 0.5 を超えているので、過半数を超えることが確率 98% で言える。

ちなみに、信頼度を 99% に設定すると、

$$-2.58 < \frac{0.6-p}{\sqrt{\frac{p(1-p)}{10}}} < 2.58$$

これを解いて、0.471214<p<0.716304

よって、99% で考えると、過半数を超えているとは言えないことになる。

信頼度を 90% に設定すると、

$$-1.65 < \frac{0.6-p}{\sqrt{\frac{p(1-p)}{10}}} < 1.65$$

これを解いて、0.517551<p<0.677148

よって、信頼度（確率）90% でよければ、過半数を超えると予想できる。

3　授業を実施して

3－1「仮説検定の考え方」の授業について

数学Ⅰを学習中の生徒（今回の授業を受ける生徒は、「データの分析」は学習が終わり、数学Ａの「場合の数と確率」を学習中）が持っている知識や技能は限られているため、この段階で仮説検定を扱うことは難しく、また制約も多いと考えられる。その中で、平均値の違いをどう捉えるべきかという課題に生徒を取り組ませた。

最初は、起こりやすさの程度をコンピュータのシミュレーションを使用して判断させた。母平均と標本平均、母分散と標本分散の関係を数学的に高校１年生に理解させるのは非常に困難であるので、コンピュータのシミュレーションで体感させ、教員から説明した。その次の段階では、正規分布を仮定して、正規分布表を用いて、起こりやすさの程度を判断できるように指導した。

評価問題から見ると、ルーブリックに示したように 0,1,2,3,4 の４点満点で評価したが､平均値が3.5であったので、生徒は比較的よく理解して、問題を解くことができた｡十分満足できる数値であると考えられる｡しかし、この評価問題に正しく解答できたことが、生徒に永続的な理解ができて、時間が経過しても、その理解が続くとは単純には期待できないであろう（それを期待するには、学習時間が３時間というのは少なすぎると思われる)。

3－2「区間推定」の授業について

時間がなくて最後まで終わらなかったものの、生徒は推定の方法についてはかなり理解できたと考えられる。

不等式 $-1.96 \leq \frac{0.6-p}{\sqrt{\frac{p(1-p)}{100}}} \leq 1.96$ を解くのではなく、仮説検定の考え方を含みつつ､母比率をいろいろ変化させて、信頼度（確率）95% で実現比率の 0.6 を含むかどうかを計算するという展開であった。

不等式 $-1.96 \leq \frac{0.6-p}{\sqrt{\frac{p(1-p)}{100}}} \leq 1.96$ を手計算で高校１年生に解かせることは場合分けなどがかなり煩雑になるので、コンピュータに計算させる方法も考えられたが、母比率を変化させて何通りか計算することによって、自分が何をしているかを把握しやすかったと思われる。

今回の問題場面が選挙の得票率に関することであったので、公民科の授業と連携した教科横断的な授業とすることも考えられる。

3－3 まとめ

「仮説検定の考え方」､「区間推定」どちらの授業も、授業の中での生徒の理解度は比較的よかったが、この授業を受けたから、生徒が独力で仮説検定や区間推定ができるようになったわけではない。探究活動などで実際に活用するためには、指導する教員の助力が相当に必要であろう。そのため、令和元年の 10 月には、探究活動を指導する教員（本校の全教員）を対象に､「仮説検定の考え方」を 50 分程度の研修会を実施した。これにより、２年生の探究活動で仮説検定が必要な場合には、担当教員がアドバイスをすることができると期待できる｡ただ、実際には実験の計画段階から検討してデータの数などを計画する必要があるので難しい面もある。実験によっては、データ数を増やすことができないものもある。計画段階から、実験後のデータ処理も見据えて計画を立てるような指導が必要になってくるであろう。今後の探究活動では、この点に留意して指導する必要があるだろう。

また、統計教育という観点で見ると、統計リテラシーはこれらの授業で格段に上がったと言えるだろう。しかし、和の記号Σ等を学習していないなど数学の道具の少ない高校１年生に「仮説検定の考え方」を指導するには限界があることも事実である。例えば、数学のテストの得点のように母集団がはっきりとわかる場合でも、母平均と標本平均、母分散と標本分散の関係を理解することは非常に難しい。まして、母数が未知である場合の数学的な扱いはさらに難しい。しかし、探究活動などの現実的な場面で活用するには、このような難しい場面を避けることはおそらくできないであろう｡今回の取り組みで、単純に平均値の違いだけで判断してはいけないというような統計リテラシーが１年生全員にある程度は身についたと考えられる。今後は２年次のそれぞれの探究活動で必要となる統計を個別に指導して、実験や調査から妥当な方法で結論を導くように指導することが求められる。