

「統計的な推測」の授業実践について

さいたま市立浦和南高等学校 太田 敏之

＜要旨＞

新しく学習指導要領に追加された「統計的な推測」の章を今年度に初めて授業を行った。今回はその中で特に「正規分布」と「統計的な推測」の節について、理解度を深めるための3つの仮説をたてて授業実践を行った。本論ではそれらの授業実践について報告し、効果的な授業法について考察する。

1. 研究動機

学習指導要領が新しくなり、「統計的な推測」の章が新しく追加されたが、まだその指導法については研究が進んでいない。筆者は今年度2年次を担当し、初めて「統計的な推測」の章について授業を行った。そこで今回はその中で特に「正規分布」と「統計的な推測」の節についての授業実践についてまとめ、効果的な授業法について考察してみることにした。

2. 研究仮説

今回の研究で、以下の3つの研究仮説をたてた。

- ① 単元の学習前に、身近なものを題材とした探究的な課題を行うことで、生徒が単元に興味をもち、それによって授業への導入がスムーズになり、理解が深まるのではないかと。
- ② 学習内容の要点を、図を使うなどしてわかりやすくまとめて提示することによって、理解が深まるのではないかと。
- ③ 実験や現実事象をうまく取り入れてイメージさせることによって、理解が深まるのではないかと。

3. 「探究的な課題」の実践

単元の学習前に、生徒が興味をもち、授業への導入がスムーズになるように、2年次の夏休みに以下のような「身近なものを題材とした探究的な課題」を行った。

【課題】

コインを10枚投げたときに表が出る枚数について、実験をして考察しよう。

（実験方法）

- ① コイン（10円玉、100円玉など）を10枚用意し、それを同時に投げて表が出た枚数を数える。これを何セットか繰り返す。

（多いほどよいが、今回は最低で200セット、最高で1000セットまでとする。）

※さいころを10個投げて偶数が出た個数を数えるなど、確率が2分の1の事象であれば試行を工夫してもよい。

- ② 試行したセット数と表が出た枚数のデータをクラッシーのアンケートに入力する。
- ③ 表が出た枚数のデータを(例1)を参考にしてグラフにする。
- ④ 表が出た枚数のデータの平均(期待値)と分散を求める。

【考察】

コインを10枚投げたときに表が出る枚数について、自分の実験データやグラフを見て気づいたことを考察せよ。また、以下の(例1)は、私が100セット実験したものである。(例1)のデータと自分の実験データ(セット数が(例1)のデータより多い)を比べて、わかったことを考察して記述せよ。

(例1) 表が出た枚数をXとする。実験セット数100セット

X	0	1	2	3	4	5	6	7	8	9	10	計
回数	0	0	2	8	15	35	21	12	6	1	0	100
度数分布	$\frac{0}{100}$	$\frac{0}{100}$	$\frac{2}{100}$	$\frac{8}{100}$	$\frac{15}{100}$	$\frac{35}{100}$	$\frac{21}{100}$	$\frac{12}{100}$	$\frac{6}{100}$	$\frac{1}{100}$	$\frac{0}{100}$	1

※平均(期待値)

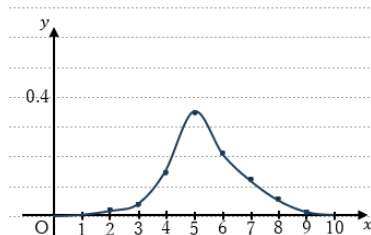
$$E(X) = 2 \cdot \frac{2}{100} + 3 \cdot \frac{8}{100} + \cdots + 9 \cdot \frac{1}{100} = \frac{530}{100} = 5.3$$

X^2 の平均

$$E(X^2) = 2^2 \cdot \frac{2}{100} + 3^2 \cdot \frac{8}{100} + \cdots + 9^2 \cdot \frac{1}{100} = \frac{3004}{100} = 30.04$$

分散 $(= (X^2 \text{の平均}) - (X \text{の平均})^2)$

$$V(X) = E(X^2) - \{E(X)\}^2 = 30.04 - (5.3)^2 = 1.95$$



以上の課題は、「正規分布のような形状のグラフに慣れさせる」ことをねらいとする。大数の法則により、(例1)の100セットに比べて、200セット…1000セットと増やしていくと、正規分布のようなグラフに近づくことがわかる。ただし、今回の実験は二項分布であり、コインを投げる枚数が10枚と少ないため、二項分布が正規分布にそれほど近似しないのであるが、コインを投げる枚数を増やすと二項分布が正規分布に近似していくことについては、後の授業の中で述べていくこととする。

(1) 生徒の実験データ

生徒の実験データを集めた合計60,960セットの実験結果は以下の通りである。

実験データ 試行セット数 60960 セット

X	0	1	2	3	4	5	6	7	8	9	10	計
回数	86	734	2820	7106	12357	14992	12478	6982	2752	684	59	60960
度数分布	0.001	0.012	0.046	0.117	0.203	0.246	0.205	0.115	0.045	0.011	0.001	1

※平均(期待値)

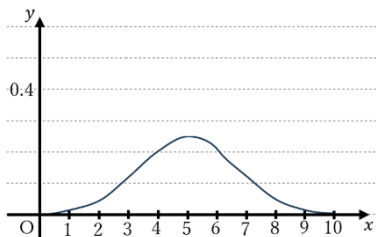
$$E(X) = 1 \cdot 0.001 + 2 \cdot 0.012 + \dots + 10 \cdot 0.001 = 4.995$$

X^2 の平均

$$E(X^2) = 1^2 \cdot 0.001 + 2^2 \cdot 0.012 + \dots + 10^2 \cdot 0.001 = 27.50$$

分散 $(= (X^2 \text{の平均}) - (X \text{の平均})^2)$

$$V(X) = E(X^2) - \{E(X)\}^2 = 27.50 - (4.995)^2 = 2.55$$



ここで理論的には、コインを10枚投げたときに表がでる枚数の分布は二項分布 $B(10, \frac{1}{2})$ に従うので、平均 $10 \times \frac{1}{2} = 5$ 、分散 $10 \times \frac{1}{2} \times \frac{1}{2} = 2.5$ である。

予備実験データ100セットの平均が5.3、分散が1.95に対して、生徒のデータを集めた60,960セットの平均が4.995、分散が2.55と理論値に近くなっていて、今回の実験を通して、大数の法則を実感させることができた。

次に「コインを10枚投げたときに表が7枚以上出る確率」について考えさせる。この確率を生徒の実験データからその確率を計算すると、

$$\frac{6982 + 2752 + 684 + 59}{60960} = \frac{10419}{60960} \approx 0.1709$$

となり、二項分布で計算した結果

$${}_{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + {}_{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2$$

$$+ {}_{10}C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^1 + {}_{10}C_{10} \left(\frac{1}{2}\right)^{10} = \frac{176}{1024} \approx 0.1719$$

と近い値となるが、正規分布で計算すると、

$$P(X \geq 7) = P\left(Z \geq \frac{7-5}{\sqrt{2.5}}\right) = P(Z \geq 1.26)$$

$$= 0.5 - u(1.26) \approx 0.10383$$

となり、コイン10枚投げたときでは、二項分布は正規分布にあまり近似しないことがわかる。

これに対して、「コインを100枚投げたときに表が56枚以上出る確率」について考えると、

二項分布で計算した結果は

$${}_{100}C_{56} \left(\frac{1}{2}\right)^{56} \left(\frac{1}{2}\right)^{44} + {}_{100}C_{57} \left(\frac{1}{2}\right)^{57} \left(\frac{1}{2}\right)^{43} + \dots + {}_{100}C_{100} \left(\frac{1}{2}\right)^{100} \approx 0.1356$$

この実験は二項分布 $B(100, \frac{1}{2})$ に従うことから、

平均 $100 \times \frac{1}{2} = 50$ 、分散 $100 \times \frac{1}{2} \times \frac{1}{2} = 25$ より、

正規分布で計算した結果は、

$$P(X \geq 56) = P\left(Z \geq \frac{56-50}{\sqrt{25}}\right) = P(Z \geq 1.2)$$

$$= 0.5 - u(1.2) \approx 0.11507$$

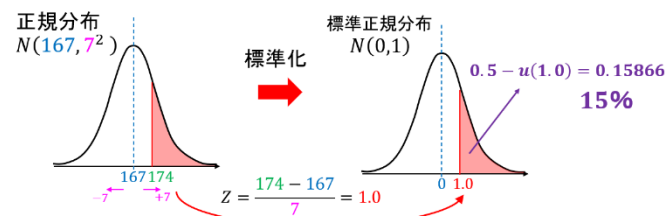
となり、コインを10枚投げたときより、コインを100枚投げたときの方が正規分布に近似していることがわかる。

4. 図をうまく利用したまとめ

次に、教科書の例題などを解くときに、数式だけではなく、以下のような図を利用して説明したので、紹介する。

(例題1) 正規分布の標準化

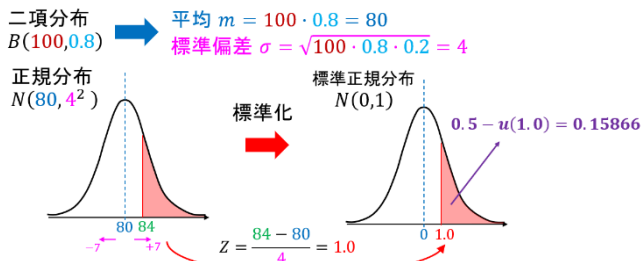
ある高校の男子の身長分布は平均167cm、標準偏差7cmの正規分布とみなせるといふ。身長が174cm以上の生徒はおよそ何%いるか。



この問題は上記のような正規分布の図を書き、標準化して解くとわかりやすい。

(例題2) 二項分布の正規分布による近似

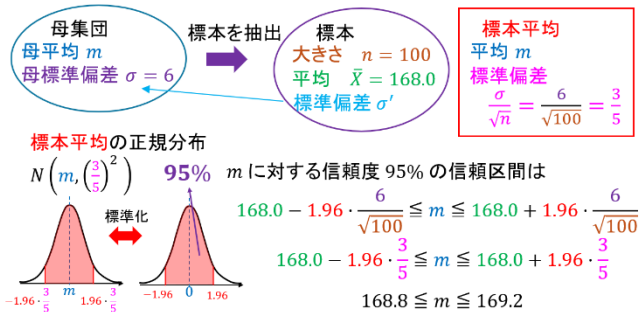
ある植物の種子の発芽率は80%である。花壇にこの種子を100粒まいたとき、84粒以上が発芽する確率を求めよ。



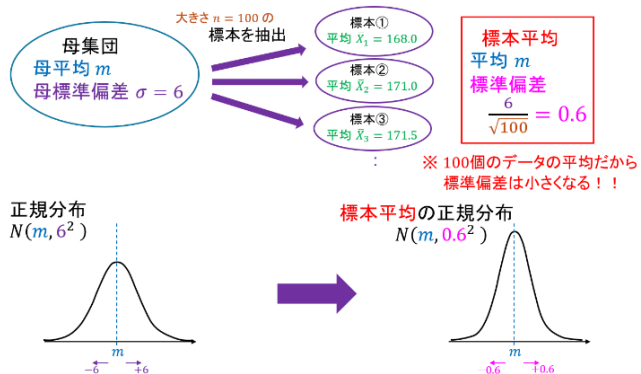
この問題は二項分布から平均と標準偏差を求めてから、正規分布の図を書き、標準化して解くとわかりやすい。

(例題3) 母平均の推定 (信頼区間)

ある県の17歳男子の100人を無作為に選んで調べたところ、身長が平均168cmであった。母標準偏差を6cmとして、この県の17歳男子全体の平均身長 m に対する信頼度95%の信頼区間を求めよ。

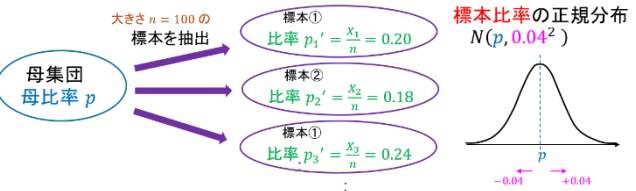
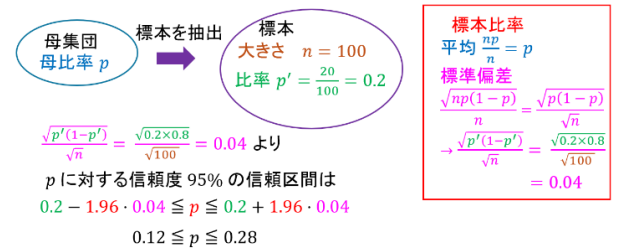


この問題は標本の標準偏差 σ ではなく、標本平均の標準偏差 $\frac{\sigma}{\sqrt{n}}$ を求めて用いる。信頼区間の幅を意識して計算するとわかりやすい。特に標本平均の標準偏差が、 n が大きければ大きいほど、平均をとることから、母集団の標準偏差より小さくなることをしっかりと説明することは重要である。ここはわかりづらいので、以下のような図や例を用いて丁寧に説明するとよい。



(例題4) 母比率の推定 (信頼区間)

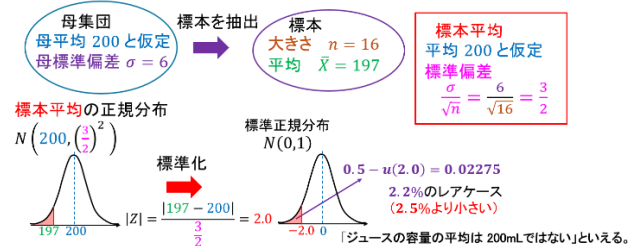
ある選挙区で、100人を無作為に選んで調べたところ、A党の支持者が20人であった。この選挙区におけるA党の支持率 p に対する信頼度95%の信頼区間を求めよ。



この問題は、標本の標準偏差が $\sqrt{np(1-p)}$ なので、母平均の推定と同じように標本比率の標準偏差を $\frac{\sqrt{np(1-p)}}{n}$ としてしまう誤答が多い。しかし、この問題は比率なので、標本比率の標準偏差は、母比率の標準偏差と近似することと、比率なので全体を n で割る必要があること、そして計算を簡単にするために、 p を推定するのにある標本①から比率をとって $p = p'$ と近似し、 $\frac{\sqrt{np'(1-p')}}{n} = \frac{\sqrt{p'(1-p')}}{\sqrt{n}}$ として信頼区間の幅を計算するところが難しい。

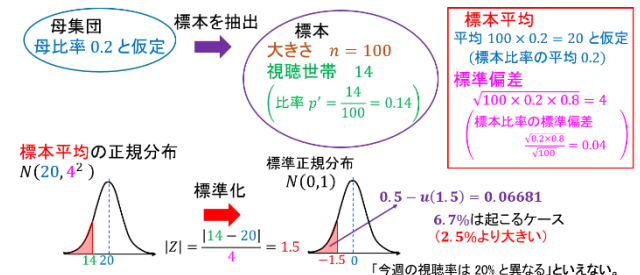
(例題5) 母平均の検定 (仮説検定)

ある会社では、ジュースを1本平均200mL、標準偏差6mLの正規分布に従うように製造している。16本を無作為抽出して調査したところ、容量の平均は197mLであった。この結果から、「ジュースの容量の平均は200mLではない」と判断できるか。有意水準5%で仮説検定せよ。



(例題6) 母比率の検定 (仮説検定)

ある番組の先週の視聴率は20%であった。無作為に抽出した100世帯のうち、この番組を今週視聴した世帯は14世帯であった。「今週の視聴率は20%と異なる」か。有意水準5%で仮説検定せよ。



これらの2つの例題のような仮説検定の問題も、正規分布の図を書いてイメージさせ、さらに実際に起きたことが偶然に起きるとしたらレアなケースなのか、よく起こりうるケースなのかを考えさせることが大切である。

5. 実験や現実事象からのイメージ

上記のいろいろな例題について、調査する数を変えて信頼度95%の信頼区間の幅の変化をイメージさせるなど、現実事象を題材にして理解を深める活動が必要である。以下は、前章で紹介した(例題5)について、調査する本数を変えて、このジュースの平均容量 m に対する信頼度95%の信頼区間を求めることで、信頼区間の幅の変化を考えさせるものである。

$n = 16$ のとき

$$197 - 1.96 \cdot \frac{6}{\sqrt{16}} \leq m \leq 197 + 1.96 \cdot \frac{6}{\sqrt{16}} \text{ より } 194.1 \leq m \leq 199.9$$

$n = 100$ のとき

$$197 - 1.96 \cdot \frac{6}{\sqrt{100}} \leq m \leq 197 + 1.96 \cdot \frac{6}{\sqrt{100}} \text{ より } 195.8 \leq m \leq 198.2$$

$n = 400$ のとき、

$$197 - 1.96 \cdot \frac{6}{\sqrt{400}} \leq m \leq 197 + 1.96 \cdot \frac{6}{\sqrt{400}} \text{ より } 196.4 \leq m \leq 197.6$$

実際に実験をしたデータを用いて考察するとよいのであるが、ジュースを400本買って調査するのは予算的にも大変なので、お菓子をを使って次のような実験をしてみた。

(実験)

お菓子のカラムーチョ100本を無作為抽出して重さを測って調査する。

(ただし母標準偏差は標本調査により0.33gと推定したとする。)

※これについては授業後に実験したので、授業では紹介していない。



母集団として、カラムーチョ300本で46gだったので、1本当たりの母平均は0.153gと考える。

カラムーチョ100本を取り出して重さを測る実験を3回行ったら、結果は以下の通りになった。

1回目 16.6gより 母平均(推定)は 0.166g

$$0.166 - 1.96 \cdot \frac{0.33}{\sqrt{100}} \leq m \leq 0.166 + 1.96 \cdot \frac{0.33}{\sqrt{100}} \text{ より } 0.101 \leq m \leq 0.231$$

2回目 18.4gより 母平均(推定)は 0.184g

$$0.184 - 1.96 \cdot \frac{0.33}{\sqrt{100}} \leq m \leq 0.184 + 1.96 \cdot \frac{0.33}{\sqrt{100}} \text{ より } 0.119 \leq m \leq 0.249$$

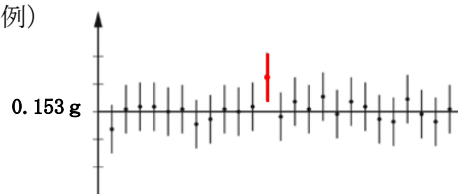
1回目 15.4gより 母平均(推定)は 0.154g

$$0.154 - 1.96 \cdot \frac{0.33}{\sqrt{100}} \leq m \leq 0.154 + 1.96 \cdot \frac{0.33}{\sqrt{100}} \text{ より } 0.089 \leq m \leq 0.219$$

この実験では信頼区間の幅が思ったよりせまくならず、思うような推定を行うことはできなかった。

実験回数を増やしていった、例のようなグラフをかいていくと推定できるのかもしれないが、もっと精度や高く、効率のよい実験がないかを考えてみたいと思う。

(例)



(例題7) 母比率の検定(仮説検定)

あるペンが書きやすいかどうか調査したところ、書きやすいと答えた人は77%であった。「このペンが書きやすいと思う人は80%と異なる」か。有意水準5%で仮説検定せよ。

$n = 100$ のとき、平均 $100 \times 0.8 = 80$ 標準偏差 $\sqrt{100 \times 0.8 \times 0.2} = 4$

$$Z = \frac{|100 \times 0.77 - 80|}{4} = 0.75 \text{ より } 0.5 - u(0.75) = 0.22663 \text{ (22.7\%)}$$

$n = 400$ のとき、平均 $400 \times 0.8 = 320$ 標準偏差 $\sqrt{400 \times 0.8 \times 0.2} = 8$

$$Z = \frac{|400 \times 0.77 - 320|}{8} = 1.50 \text{ より } 0.5 - u(1.50) = 0.06681 \text{ (6.7\%)}$$

$n = 900$ のとき、平均 $900 \times 0.8 = 720$ 標準偏差 $\sqrt{900 \times 0.8 \times 0.2} = 12$

$$Z = \frac{|900 \times 0.77 - 720|}{12} = 2.25 \text{ より } 0.5 - u(2.25) = 0.01222 \text{ (1.2\%)}$$

→ 2.5%より小さいので、異なるといえる。

この問題については、調査結果が同じ77%でも、調査した人数によって答えが違ってくる。調査した人数を100人、400人、900人と増やしていったとき、調査した人数が400人のときは6.7%で、2.5%より大きいので、「書きやすいと思う人が80%と異なるといえない」が、900人のときは1.2%で、2.5%より小さいことから、「書きやすいと思う人は80%と異なるといえる」ことがわかる。

6. まとめ

今回の授業実践において3つの仮説をたてた。1つ目の仮説については、単元の学習前に「コインを10枚投げたときに表が出る枚数を数える」という身近なものを題材とした探究的な課題を行うことで、正規分布の図のイメージが付きやすくなり、単元授業への導入がスムーズになった。2つ目の仮説については、学習内容の要点を、図を使うなどしてわかりやすくまとめて提示することによって、授業の中では理解が深まっているように感じた。ただし、定期試験での正答率を分析すると、まとめたものを提示するだけでなく、生徒が自分でまとめるような活動があった方がよかったのではないかと思った。3つ目の仮説については、教科書の例題も現実事象が多く、授業でもそれを扱っていったが、本論で扱ったように、現実事象に即して調査数などの構成要素を変化させることでどう変わるかなどをもっと考察させることで、理解がもっと深まっていくのではないかと思う。また、実験については効果的な実験が見つからなかったため、今後の教材研究の中で探していきたいと考える。