# 高等学校における統計の指導について

岸谷　正彦

## 1．本稿の目的

**1－1　本稿の目的は，以下の通りである。**高等学校数学科において，

1）統計を学習するときのさまざまな課題を明らかにすること

2）統計の学習指導を行う立場で科目の内容である「場合の数と確率」,「統計的な推測」,「微分積分」およびその他の関連科目内容を検討してみること

3）統計の授業の具体的な授業案を提示すること

である。

課題と検討については，改善して指導する方法を明らかにすること，同時に様々な実験と検証を行うことに重点を置いた授業例を提示することで，新しい統計指導方法を提案したいと思っている。

**1－2　今回の学習指導要領の改訂における統計関連の学習内容の特徴について，数学Ⅰ「データの分析」においては，以下の通りである。**

1）データの散らばり具合や傾向を数値化する方法を考察し，分散，標準偏差，散布図及び相関係数の意味やその用い方を理解すること

2）目的に応じて複数のデータを収集し，適切な統計量やグラフ，手法などを選択し，コンピュータなどの情報機器を用いるなどして，データを表やグラフに整理したり，分散や標準偏差などの基本的な統計量を求めたりして分析を行い，データの傾向を把握して事象の特徴を表現すること

3）具体的な事象において仮説検定の考え方を理解するとともに，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

また，数学B「（2）統計的な推測」においては，以下の通りである。

1）標本調査の考え方について理解を深めること

2）確率変数と確率分布について理解すること

3）二項分布と正規分布の性質や特徴について理解するとともに，確率分布や標本分布の特徴を，確率変数の平均，分散，標準偏差などを用いて考察すること

4）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

5）目的に応じて標本調査を設計し，収集したデータを基にコンピュータなどの情報機器を用いて処理するなどして，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

これらを踏まえて統計教育における新しい視点をまとめると，数学Ⅰ「データの分析」においては，ある課題に対して目的に応じてデータを収集し，データの傾向を把握し事象の特徴を表現すること。さらに仮説検定の考え方を理解し，判断の根拠を明確にする，あるいは批判的に考察することが求められている。数学B「（2）統計的な推測」においては，標本調査の方法と確率変数の考え，代表的な分布について理解し，標本の分布の特徴を数値で捉えること，区間推定や仮説検定の方法を理解する，標本から母集団の特徴を推測すること，その方法や結果についても批判的に考察することなどが求められている。

## ２．統計教育の意義とその課題

### ２－１．統計リテラシーの育成と統計教育の実際と課題

そもそもなぜ統計教育の充実がさけばれているのか。要約すれば以下であろう。

1）科学技術が発展する中で，統計情報が日常的に溢れる今日，自身の発言や行動自体が責任ある根拠の上に立ったものである必要があること

2）様々な分野での課題解決を図るための調査分析に対し，統計的な思考やスキルを持った人材の育成が欠かせないこと

3）生活上必須の統計的な素養は，生活とともに身につける必要があり，年少のころから意識して学ばせる必要があること

それに対し私たち教員は，統計教育を通してどのような力を身に付けさせようとしているのかとその場合の課題について述べてみたい。

まず様々な事柄に問題意識をもち，それを調べたいという動機を大切にする教育の実現を計りたい。なぜという疑問を大切にする授業が，実生活に結びついた統計学習を通して現実化する。これは，これまでの数学を修得していく学習とは異なる一面と言える。

次にそのような問題意識に対し，どのようにしてアプローチするのかを考えられる力が必要となる。とくに漠然とした問題に対し，どのような切り方が適切か，どのようにして根拠となりうるものを集めるのか，綿密な計画のもと疑問の解決につながる過程を生徒自

ら考え出せる教育が必要となる。

次はデータの収集についてである。今日データ情報は玉石混淆であり，その中から価値のある情報をいかに収集するかを学ばせたい。公的な機関が公表しているデータなども活用したい。それらのデータを１次資料として利用するとき，データを収集した人への尊重と敬意を払うことや得られた結果に対し慎重になることなどは教えたい。

次に得られたデータの処理の仕方が問題となる。データを正しく読みとり，それから何が得られるのか，それを見いだす能力の育成である。特に高校においては，科学的な方法でデータの処理や分析が可能にはなる。そのスキルの習得も必要となる。しかし，得られた結果を鵜呑みにせず，別の方法で確かめ，自らあるいは他者からの批判的な見方への寛容さも必要となる。

そして得られた結果をさらに利用すること，どのように表現するのかにも責任がある。得られた結果に対し，これまでの反省や今後の予測に活かせないか，新たな調査分析の糸口にもなろう。

高校では，確率統計の系統的な視点でみた学習カリキュラムが必要であると思われる。また小中を通した系統的な学習をうけての高校の学びが，これまでと大きく変わらないやり方では，得ることは少ないであろう。形式的な計算で仮説検定ができて正解かどうかの学習には実感がなく，生活に必要ないと言われかねない。

コンピュータの活用は不可欠である。統計分野においては，コンピュータを用いてシミュレーションができる環境が必要である。統計の本質の理解の部分は高度な数学が必要となるため，数学的になぜの部分がどうしても残るが，それは，シミュレーションなどの疑似体験でカーバーできるだろう。大量のデータを扱うにもコンピュータは必須である。

数学的な背景理解の難しさがある。「数列」の学習前の「データの分析」では，データの変換，分散や相関係数の本来の理解にやや困難を要するだろう。また，「場合の数と確率」においても，末節にこだわることなく，必要最小限の内容にとどめるべきであろう。微分積分の学習前の「統計的な推測」においても同様のことが言える。底eの指数関数，ある図形の面積で確率を表すこと，連続型確率変数の意味など，ともかく統計教育における数学的な背景も含めた理解は，かなり難しい場面もある。統計分析の本では，数学的な説明は避ける傾向があるが，それが主流ではいけない。数学的な裏付けがあってのきちんとした理解であることも忘れてはいけない。

これまでの数学の学習は，概念理解と具体的な計算の積み重ねであり，特に計算演習をより多く行うことが重視されていた。今後は，どこまでを認めて何に利用するのかの選択が必要になってくるであろう。自分の言葉で表現し，他者に伝える数学学習の本質を統計学習の中で体現させたい。

# ３．統計の学習を見据えた立場で既存の科目内容を検討してみること

## ３－１．「場合の数と確率」の学び

これまでの数Ａ「場合の数と確率」は，数え上げることを主眼において，不確定な事象を数理的にとらえる数学のよさを認識できるような指導に重点をおいていた。今回「場合の数と確率」には，不確実な事象の起こりやすさを判断したり，期待値を意思決定に活用したりする役割がさらに与えられている。すなわち「場合の数と確率」には，数Ｂの「統計的な推測」への結びつきを重視する指導が付け加わったと言える。

これを詳細に述べてみる。

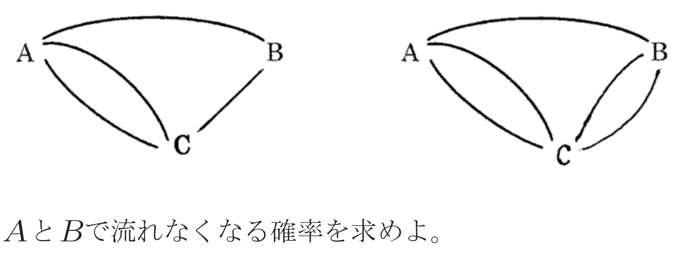
(1) 場合の数では，起こりうるすべての場合の数を計算してみることが求められる。確率では，個々の事象の確率を求めるだけでなく，考えられる事象全体の中での起こりやすさをより意識した指導が必要になる。さらに確率変数やその分布，同時分布を意識して確率を求めること，累積度数分布の活用なども視野に入れたい。

例１．いくつかのサイコロを投げて出た目の和や差の確率を求める問題。

問．３個のサイコロを投げたときの出た目の平均の値とその確率。４個の場合はどうなるか。【補足】単に，和が５になる確率だけでなく，すべての目の和の値に対する確率を求めること。確率が最も大きくなる目の和は，どのような場合かを予想し確かめるなど。

(2) 日常生活や社会の事象に題材を見つけて，確率を設定したりする場面も考えられる。コンピュータを使ったシミュレーションを利用して判断するということもあり得る。

例２．次の図のように$A$と$B$を高圧線がつながっている。さらに途中に，$C$を経由して行く線あり，合計4本の高圧線がつながっている（左図）。このうち何らかの事情でつないだ高圧線が流れなくなる確率は，互いに独立で$p$であるとする。




（ｉ）　左図で$A$と$B$で流れなくなる確率を求めよ。

（ii）　右図のようにもう一本$B$と$C$に線を増やした。このとき，流れなくなる確率は，どうなるか調べてみよう。

（iii）　流れなくなる関係が独立でない場合も考えてみよう。どのようにすれば，流れな

くなる確率がより小さくなるか？

(3) 確率では，試行の方法とくに非復元抽出，復元抽出，同時抽出による違い，試行が独立である場合と独立でない場合などにも，これまで以上に注意を払うべきである。また将来の離散的な分布の代表である二項分布と超幾何分布の例を意識して行った方がよい。

例３．赤玉３個，白玉２個入った袋から，３個の玉を次のように取り出す。

（ⅰ） １個ずつ非復元抽出で取り出す

（ⅱ） １個ずつ復元抽出で取り出す

（ⅲ） 同時に取り出す

それぞれ赤玉の出る回数Xについて，Xのとりうる値とそのときの確率を求めよ。

例４．白玉と赤玉あわせて30個入った袋がある。この袋から，10個の玉を取り出したとき，白玉が５個であった。

（ⅰ） 袋の中の白玉の個数は15個であると判断するには，どのような仮定が必要か？

（ⅱ） 袋の中の白玉の個数が15個であることがわかっているとき，10個を取り出して白玉が５個である確率と白玉が６個である確率を比較したとき，どちらの確率が大きいか。

（ⅲ） 袋の中の白玉の個数が15個であることがわかっているとき，10個を取り出して白玉が５個である確率と12個を取り出して，白玉が６個である確率を比較したとき，どちらの確率が大きいか。

(4) 条件付き確率の指導は，現行でも行われているが，トピック的な扱いではなく，本格的な扱いを期待したい。次の問題は，教科書などでよく原因の確率ということで取り上げられる問題である。

例５．ある病原菌の検査薬は，病原菌に感染しているのに誤って陰性と判断する確率が１％，感染していないのに誤って陽性と判断する確率が３％である。いま，全体の１％がこの病原菌に感染していることが分かっている集団から１つの個体を取り出すとき，陽性と判断されているのに，実際には病原菌に感染していない確率を求めよ。

【補足】 題意から感染している割合は $A=\frac{1}{100}$ ，残りの空欄の２カ所の情報が与えられると $G$ が求められて，条件 $G$ がついた確率が求められる。これから，この問題の構造が明らかであるだろう。（表１）

**表１**

| | 感染している | 感染していない | 計 |
|---|---|---|---|
| 陽性 | | $(1-A)\times\frac{3}{100}$ | $G$ |
| 陰性 | $A\times\frac{1}{100}$ | | |
| 計 | $A=\frac{1}{100}$ | $1-A$ | 1 |

**表２**

| | 感染している | 不明 | 感染していない | 計 |
|---|---|---|---|---|
| 陽性 | | | $e$ | |
| 陰性 | $c$ | $d$ | | |
| 計 | $a$ | | $b$ | 1 |

例６．以下において，集団を互いに排反である３つの事象に拡張してみる。（表２）

（ⅰ）　感染していないのに，陽性である確率を求めるとき，どのような情報があればよいか考えよ。

（ⅱ）　例えば，$a$，$b$を与えて，$c$，$d$，$e$の割合を提示すれば，様々な条件付き確率が求められることがわかる。

(5) 事象の独立につなげられる指導は大切である。独立に関しては，事象の独立，試行の独立，確率変数の独立の順で学ぶのがよく，数学Aで事象の独立と試行の独立を取り入れたい。また，事象の排反と独立の違いもきちんと理解させたい。

例７．袋の中に，１，２，３の数字が書かれた玉がそれぞれ３個，２個，１個入っている。この袋から１個ずつ玉を取り出すとき，１回目，２回目に出る数をそれぞれ$X$, $Y$とおく。このとき，和$X+Y$, 積$XY$の取りうる値をすべて求め，それぞれの確率を求めよ。

(i)　取り出した玉を袋に戻すとき

(ii)　取り出した玉を袋に戻さないとき

例８．$A$, $B$, $C$ の3人がゲームをする。１個のサイコロを投げて，１，２の目が出たら $A$ の勝ち，３，４の目が出たら$B$の勝ち，５，６の目が出たら$C$の勝ちになりコインを１枚もらう。このゲームを２回行い，変数 $X$ に $A$ が獲得するコインの枚数，変数 $Y$ にコインを獲得する人数とする。このとき，$X$ と $Y$ は独立であるかどうか調べよ。

## ３－２．「データの分析」

数学Ⅰ「データの分析」は，小中学校で学んだことを生かして，データの散らばり具合や傾向をグラフや散布図等で表し，ある数値でそれを表現し比較できること，とくに分散の導入，散布図において相関係数の意味と回帰直線を理解すること，それを用いて傾向を

読み取りができること，ある事象についての仮説を立て，それが起こりやすいかどうかに確率を用いて考えることができることが求められる。

詳細は以下である。

(1) 目的に対し課題を設定し計画をたてデータを収集し分析し結論を得る，統計的な問題解決の過程は，引き続き数学Ｉの主要テーマになろう。

例１．地方では，商店街が縮小している町も少なくない。その理由は何だろうか。人の移動，郊外型大規模店舗数の推移，小売りの店舗数やある業種に限った店舗数などの推移，考えうるデータを集め分析検討する。

(2) 数学Ｉでも仮説検定の考え方を実例から理解することが求められる。仮説に対し正しいかどうかの判断に確率を用いることや既存の分布に照らし合わせて確率を計算することで，それが起こりやすいかどうかの判断ができることも理解させたい。

例２．白玉と黒玉が大量に３：２の割合で入っている袋がある。これを１個ずつ何回か引いたが１個も白玉が出なかった。珍しい場合と考えられるのは，連続して何回黒玉が出るときか？【補足】「珍しいという主観を客観的に見るにはどうするか。どのような確率を求めればよいのか」も含めて考えさせる。

(3) 数学Ｉは必履修科目であり，その内容については，何を仮定して何が学習されていないのかを明確にしておかないと困難を感じる場面も出てくるであろう。例えば，場合の数や確率をどこまで学習しているかである。さらに平均や分散，共分散を扱うための和の記号の使用等は，本来積極的に利用したい。その結果データの１次変換による平均（仮平均），分散，相関係数の変換などは，その意味も含めて理解しやすいのではないか。

例３．予想からはじめる回帰直線の意味を探る。**指導案の提示参照**。

例４．エクセルを用いて，ある（$x, y$）のデータの散布図を描いてみせる。さらに，$u = -x+10$，$v = 5x$ で変換したときの（$u, v$）の散布図はどのようになるか予想させる。相関係数や回帰直線なども調べさせてみる。

### ３－３「統計的な推測」の学び

数学Ｂの「統計的な推測」の内容については，現行にはない検定が付け加わることで内容がより豊富になる。また数学Ｉで具体的な事象を用いた仮説検定の考え方を学習していることを考慮しなければならない。そのため，推定や検定を行う場面やその必然性に結び付けた授業がより必要になるだろう。公式に代入し即座に信頼区間を得ることや型通りの検定の手法をマスターすることだけに終わることなく，もとになる分布の概念や確率を用いて判断すること，なぜ推定や検定が行えるのかの考えが及ぶようにしなければいけな

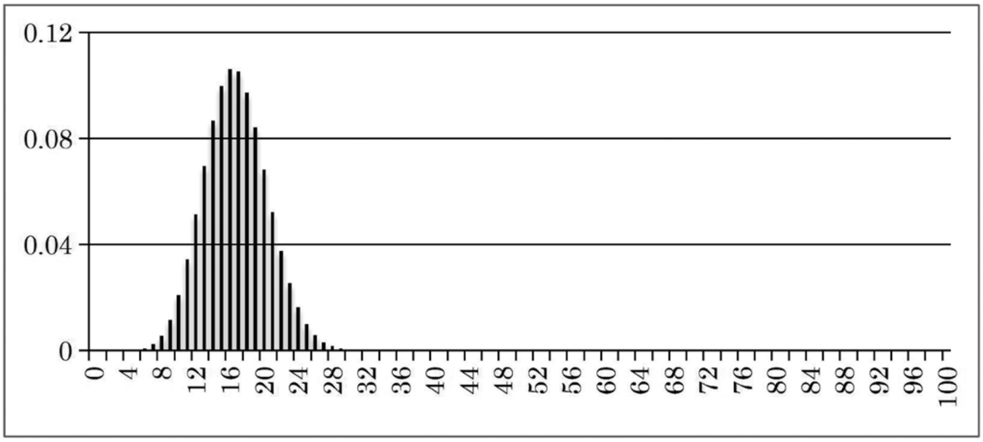
い。

詳細は以下である。

(1) さまざまな分布の理解がもっとも大切である。**指導案の提示参照**。

例１．離散型確率変数から連続型確率変数へは，二項分布から正規分布を作り出すことを例示する方が理解しやすい。サイコロを $n$ 回投げて，１の目が出る回数を確率変数にする確率分布を作ってみる。$n$ = 10，50，100 のように，サイコロの投げる回数を増やしていくと，分布のヒストグラムは，曲線状に近づくことが予想される。

下図は，サイコロを100回投げたときの１の目の出る回数と確率。




例２．２つの分布の和とは，何を意味しているのかの実験。

男子25人の身長の度数分布

| 身長（cm） | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 合計 |
|---|---|---|---|---|---|---|---|---|---|---|
| 度数（人） | 1 | 2 | 3 | 4 | 5 | 4 | 3 | 2 | 1 | 25 |

女子16人の身長の度数分布

| 身長（cm） | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 合計 |
|---|---|---|---|---|---|---|---|---|
| 度数（人） | 1 | 2 | 3 | 4 | 3 | 2 | 1 | 16 |

この２つの分布の和の分布を求めてみよう。分布の和とは何かをあらためて問う問題である。山型の曲線を２つ足すという意味ではないことを気づかせる。

(2) 母集団から標本の取り出し方と，母集団と標本の平均の分布の関係を様々な活動を通して考える。指導案の提示参照。

例３．５人の体重が66，68，70，72，74kgである集団を母集団とする。

（ｉ）　これから２人を非復元抽出で取り出し，その平均をとると20個のデータが得られる。その平均と分散を求めよ。

（ⅱ）これから２人を復元抽出で取り出し，その平均を求めると25個のデータが得られる。その平均と分散を求めよ。

これから分かることは，母集団の平均と標本の平均は等しくなり，標準偏差はより小さくなる。とくに，復元抽出の場合は，母集団の標準偏差と標本の大きさで，標本平均の標準偏差が定まることが分かる。

(3) 母平均の推定がどのような意味をもつのかを具体的に考えさせる必要がある。パラメータがわかっている母集団から，標本を取り出し，標本平均が求めた信頼区間にどの程度の確率で含まれるのかの実験を通して，信頼区間の意味を探ること。**指導案の提示参照**。

例４．１が３カ所，２が２カ所，３が１カ所付いたサイコロを５個同時に投げたとき，出た目の平均$X$の確率分布を作る。$E(X)=\frac{3}{5}$，$V(X)=\frac{1}{9}$であるから，95％の平均の信頼区間は，［1.37，1.96］である。これに対し，100回試行を行い，その平均が信頼区間に含まれる回数を求めてみよう。

(4) 検定の学習は，起こりうる確率が判断の基準になること，さらに有意水準の必要性などを押さえてから，仮説をたて検定するという手順の方が理解しやすい。また，母平均や母比率の推定は，もとになる母平均や母比率の分布から判断することをきちんと教えたい。また，数学Ⅰでは有意水準などの用語ではなく，「起こりうる確率は５％より大きいかどうかを調べよ。」でもよいのではないか。

例５． １個のサイコロを６回投げて，１の目が出る回数を$X$とするとき，$X$が５回以上である確率を求めよ。これから，このサイコロは正しいサイコロといえるか。

例６．硬貨の表の出る確率を $\frac{1}{2}$ であるとして，10回投げるとき，表が８回以上出る確率は，５％より小さいかどうかを調べよ。

### ３−４．「微分積分」やその他の内容の学び

統計の数学的な基礎の部分の理解があればより深く理解が及ぶであろう。ここでは，統計を学ぶ上での基本的な内容をまとめておく。

(1) 二項定理および二項係数は，単に二項展開することに目的があるわけではなく，多くの数え上げの際の基本道具であり，多項定理や一般二項展開への拡張などの源泉であること，また二項係数やパスカルの三角形のもつ面白さをもう少し伝えてもよいのではないか。そういう意味でも現行の数学Ⅱでの取り扱いは不適であり，場合の数と確率に関連させて教えられるものである。

例１．　$k{}_nC_k = n{}_{n-1}C_{k-1}$　，　$k(k-1){}_nC_k = n(n-1){}_{n-2}C_{k-2}$

などの式の変形も入れたい。二項分布の平均や分散の求め方はいろいろあるが，これを用いた証明が最も定義に則った方法である。

(2) 正規分布の確率密度関数のグラフの性質などは，数学Ⅲで積極的に取り入れるべきである。またガンマ関数，ベータ関数も大学入試問題で取り上げられることもあるが，そのグラフや性質も含めた指導は，興味を引くのではないか。この他，連続型確率変数を扱うとき，微分積分の基本がいかに大切かを思い知ることになる。関数，$y = xe^{-\frac{x^2}{2}}$，$y = x^2e^{-\frac{x^2}{2}}$ などの定積分を行うことで，正規分布の平均や分散も，一度は計算で求めさせたい。置換積分によって，正規分布をしている分布のある確率の値が標準正規分布表の値を読みとることで得られる仕組みにも興味を持たせたい。

例２．関数 $y = e^{-\frac{x^2}{2}}$ のグラフの概形，グラフの凹凸，変曲点を調べよ。

(3) 二項分布は，試行の回数 $n$ を大きくすると正規分布に近づいていくことは，さまざまなシミュレーション等を通して理解はされるが，正規分布の確率密度関数と結び付けるのは難しい。このようなとき，二項係数の近似値が求められることは，正規分布につながるきっかけにもなろう。$n!$ の近似式は，数学Ⅲの内容としても充分理解できる。

例３．$n \to \infty$ のとき，$n! \approx \sqrt{2n\pi}(\frac{n}{e})^n$ を用いて，${}_nC_r$ の近似値を与える式を求めよ。

また，${}_{2n}C_n$ の近似値を与える式を求めよ。

## ４．指導案の提示

ここでは，３つの授業例を取り上げる。

### ４－１．指導事例１

(1) 表題：回帰直線について考えよう。

(2) 本時の目的

①データを特徴づける直線の存在に気づき，どのような直線であるかを求めようとする。

②回帰直線の予測への活用とその妥当性を試してみる。

(3) 本時の指導のねらい

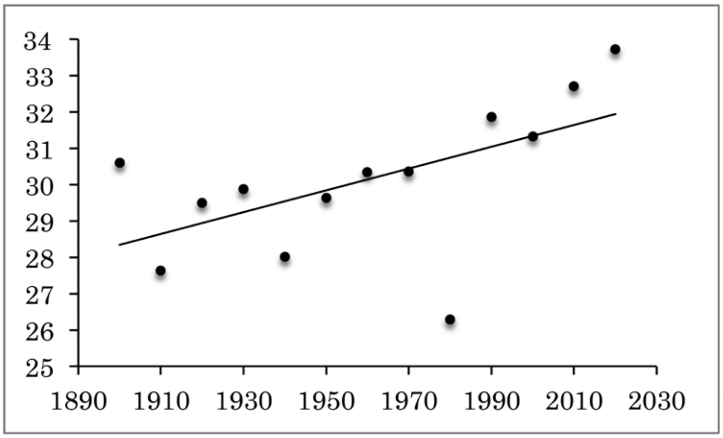
時系列データの散布図から，回帰直線の導入を図ってみる。少ないデータの場合から回帰直線の意味を知り，回帰直線の方程式を作ってみる。さらに散らばり具合という観

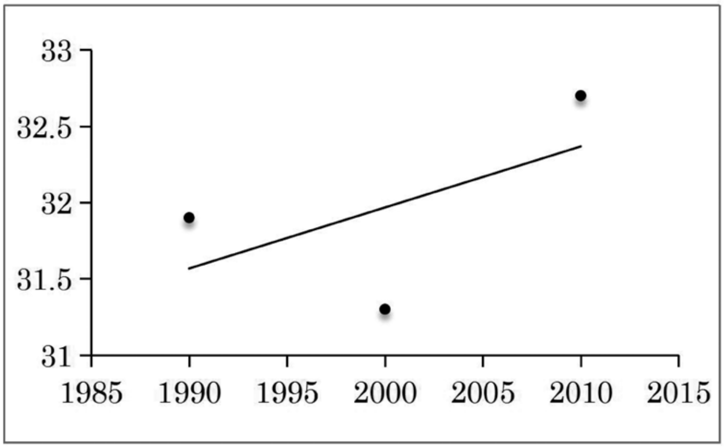
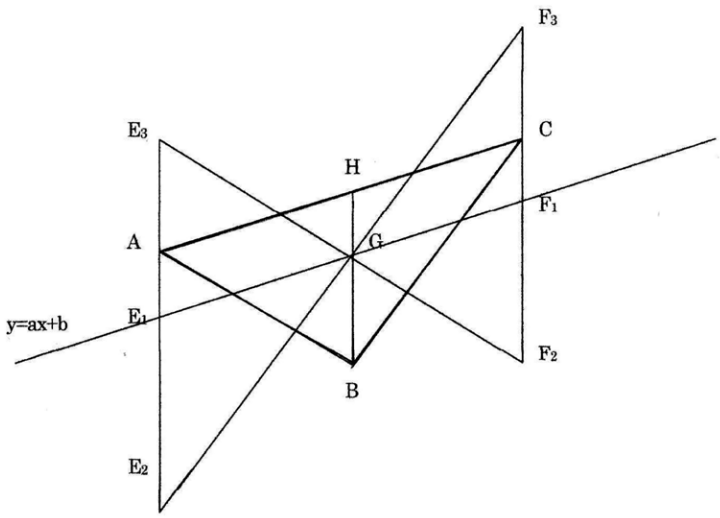
点から，回帰直線を考えていく。その直線は，予測や欠落したデータを補うことができることにも気づく。時系列データ以外の２次元データへも活用してみる。

(4) 本時の教材観：回帰直線の存在は，「百聞は一見にしかず」である。ただし，それがどのような方法で考えられるのかを考えさせることも有用であろう。今回は，過去のデータとして確定したものを用いて，予測や欠落した場合にも適用できることから，回帰直線の方程式がどのように得られるのか，分散や相関係数とどのように関係しているのかにも踏み込みたい。

(5) 学習計画：数学Ｉ「データの分析」の内容は，一通り学習した前提で設定してある。

(6) 本時の授業展開

| 学習活動（○：指導，●：指示，発問，□活動・予想される反応） | 指導上の留意点・評価 |
|---|---|
| ●次の散布図を見て下さい。<br><br><br>これは，横軸は1900年から2020年まで10年きざみ，縦軸は横浜のその年の８月の最高気温の平均をプロットしたものです。<br>●これから読みとれることは，何ですか？ | 事前に用意。<br>回帰直線は，すでに書き込んであるがないものを配布。<br>単位は，横軸は年度，縦軸は度<br>この直線を回帰直線ということにする。<br>傾き0.03<br>点（1890, 28.04）を通る。 |
| □点が右上がりであるから，最高気温が少しずつ高くなっている。 | 様々な直線が引かれる。 |
| ●これらの点の傾向を１本の直線で表してみるとどうなるか。実際書き込んでみよう。 | おおよそ右上がり。 |
| □それぞれの点の間を通るような直線かな？　右上がりである。<br>○今日はこの直線についてもう少し調べてみよう。 | 直線の上側，下側にある点の個数を調べる生徒もいる。 |

<table>
<tr><th>学習活動（○：指導，●：指示，発問，□活動・予想される反応）</th><th>指導上の留意点・評価</th></tr>
<tr><td>●次の図は，1990，2000，2010年の３つのデータから回帰直線を書きこんだものです。2020, 2030年の気温を予想してみよう。<br>□直線の傾きから2020年は，32.8℃，2030年は，33.2℃である。2020年は実際の気温の方が高い。データを多く使った方が正確な値が出ると思う。<br>
</td><td>1990，2000，2010年の場合，傾き0.04<br>点（2000, 32.0）を通る。<br>最初のデータを比べてみる。32.24°<br>1年刻みのデータがほしい。<br>A（1990，31.9）<br>B（2000，31.3）<br>C（2010，32.7）<br>である。</td></tr>
<tr><td>それは後で考えることにして，この直線の意味を考えてみよう。<br>●この直線は3点から見てどんな直線といえるか？<br>
<br>□直線$AC$に平行で，多分，重心$G$を通っている。</td><td>1990年と2010年のデータの点を結ぶ直線に平行で３点の重心を通ることが分かる。<br>これを図形として考える。三角形$ABC$の重心$G$を通る直線。<br>これから，ある2点に平行で中線を1：2に分ける点を通る直線を考えてみる。</td></tr>
</table>

| 学習活動（○：指導，●：指示，発問，□活動・予想される反応） | 指導上の留意点・評価 |
|---|---|
| □ $BC$ に平行で重心 $G$ を通る直線や $AB$ に平行で重心 $G$ を通る直線が回帰直線とは言わないのはなぜ？<br>● $y$ 座標に着目して，ここでも散らばりというものを考えてみよう。□① $AE_1$, $BG$ , $CF_1$ ② $AE_2$, $BG$ , $CF_3$ ③ $AE_3$ , $BG$, $CF_2$ のうち，散らばりが最も小さいのは，<br>□①の場合である。しかし偏差の和で比較するのはよくない。 | これを満たす他の２つの直線は回帰直線とは言わない。 |
| ●散らばり具合を表す分散は，どのように定めたか？<br>○偏差の平方の和の平均を分散といった。<br>□それでも①の場合が最小になると思う。<br>●4点目が入るとどうなるか？重心を通る？傾きは？ 2020年のデータを入れたものを見せる。<br>□ $x$ 座標，$y$ 座標の平均の座標（重心）は通る。傾きは変化した。<br>●この直線は，どのように定めたらよいか？ 一般に，$n$ 個の点（$x_1$ , $y_1$），（$x_2$ , $y_2$），……，（$x_n$ , $y_n$）に対しても同様に考えられるかな？いっしょに直線の式を作ってみよう。<br>□ある定数からの差の平方の和が最小になるときの直線を求めてみればよいか？ | 言い替えると，データの値とある定数との差の平方の和が最小になるとき，定数がデータの平均になる。<br>傾き0.068<br>（2005，32.4）を通る。<br>結果的に，$y$ 座標の残差の平方の和が最小になる直線を回帰直線と言われる。<br>変数 $X$ と $Y$ のデータの平均を $x$, $y$ 座標とする点を通る。 |
| ● $y_i - ax_i = b_i$ とおいて，偏差 $b_i - m$ の平方の和が最小になる $m$ は，$\bar{b}$ に等しくなるときであるから，$\bar{y} - a\bar{x} = \bar{b}$ であり，$\bar{b} = b$ となるように $a$ を定めよう。以下欄外に記述する。 | 時系列データではない例を出す。 |
| □傾きに，変数 $X$ と $Y$ のデータの分散，共分散や相関係数も出てきた。<br>○他の2次元のデータの散布図の回帰直線を求めてみよう。時系列データ以外でも使える。予想にも使える。 | 分散，共分散の値は与える。 |

【補足】一般に，$n$ 個のデータ $x_1$ , $x_2$ , ……, $x_n$ について，$\sum_{i=1}^{n}(x_i - m)^2$ が最小になるときの $m$ は，データの平均 $\bar{x}$ になるときである。いま，$n$ 個のデータ（$x_i$ , $y_i$）（$i = 1$，……，$n$）に対し，$y_i - ax_i = b_i$とおく。このとき，$\bar{y} - a\bar{x} = \bar{b}$ である。さらに，

$\sum_{i=1}^{n}(b_i-m)^2$ が最小になる $\boldsymbol{m}$ は $\boldsymbol{b}$ であり，$b=\ b$ となるように $\boldsymbol{a}$，$\boldsymbol{b}$ を定めるから，

$b=\bar{y}-a\bar{x}$ を用いて

$$
\begin{aligned}
\sum_{i=1}^{n}(b_i-m)^2 &= \sum_{i=1}^{n}(y_i-ax_i-\bar{y}+a\bar{x})^2 \\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2-2a\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+a^2\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= n\left({S_y}^2-2aS_{xy}+a^2{S_x}^2\right)
\end{aligned}
$$

は $a=\frac{S_{xy}}{{S_x}^2}$ のとき，最小値 $n{S_y}^2(1-{r_{xy}}^2)$ をとる。

ここで $r_{xy}=\frac{S_{xy}}{S_xS_y}$（相関係数）である。最小値 $\geqq 0$ から $-1\leqq r_{xy}\leqq 1$ も得られる。

## ４－２．指導事例２

(1) 表題：さまざまな分布について調べてみよう。

(2) 本時の目的

①身近な変量から得られるデータからどのような分布が得られるのかを試してみる。

②なぜ平均や分散が大切なのかに気づく。

③分布の意味を理解し，推定や検定に利用できることに気づく。

(3) 本時の指導のねらい

まず，様々な分布の作り方があることに興味を持たせる。母集団を与えて，いろいろな変量から作り出される分布を考えさせる。とくに標本平均の分布は，標本の大きさを大きくすると正規分布に近づいていくこと，それには分散が大きな役割を果たしていること，それらが確率を通して，推定や仮説検定に適用できることなどに具体例を通して感得できるようにする。また離散的な母集団の場合，母集団と標本の関係を求めさせる。最後に，正規分布をしている体重の分布からの偏差の分布や標準化された変量の分布はどのような分布になるかも調べさせる。

なお，この他の分布の例は，次のような場合を想定している。

(a) 母集団から取り出された２つ数の和の平均の分布，２つの数の差の分布，平方の和の分布，ある数からの差の２乗の分布，偏差の分布，標準化したデータの分布などである。また取り出した標本の大きさによって，分布がどのように変化するのかも確かめる。

(b) 二項分布をしているデータから同様の標本を取り出し，いろいろな分布を作る。

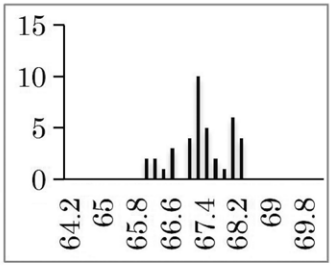
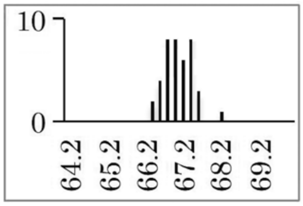
(c) 比率が既知のとき母集団から同様の標本を取り出し，いろいろな分布を作る。

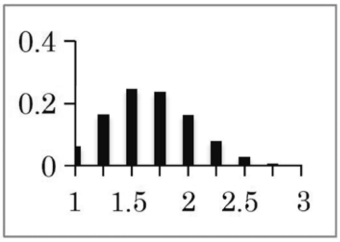
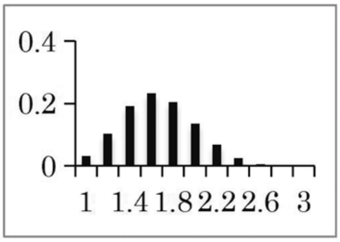
(d) いろいろな分布（例えば，度数分布）からも同様の標本を取り出し，いろいろな分布を作る。

(4) 本時の教材観：数Ｉでは，ヒストグラムで分布を作る程度であったが，確率からを得られる分布など，より多くの分布についても学ばせたい。データの出方にはある程度の規則があり，それを解析する方法として分布を調べることが大切であることに気づいてほしい。そのひとつの成果として中心極限定理が成り立つことを察知できるようにする。この他，同類の分布をしている分布の和の分布，正規分布に収束しない分布や誤差の分布なども教材にしたい。この授業では，(a) を対象にしている。また，エクセルで作成した一様分布から無作為に，１クラス40名が大きさ10個，20個の標本を持ち寄って集計したものを使っている。

(5) 学習計画：数学Ｂ「統計的な推測」は既習であるとする。

(6) 本時の授業展開

<table>
<tr><th>学習活動（○：指導，●：指示，発問，□活動・予想される反応）</th><th>指導上の留意点・評価</th></tr>
<tr><td>●Ｌサイズの卵は正味（殻なし）約60g，殻つきのままでは64g～70g未満である。ある養鶏場では，1日1000個産まれ，無作為に10個ずつのパックを作っている。重さのばらつきがないのか調べたい。また20個ずつのパックに詰めると，重さのばらつきはどうなるか？<br>□平均は，本来の平均に近づいてくる？<br>□20個の方がばらつきは大きくなる？いや小さくなる？<br>○どっちだと思うか？ 20個のばらつきの方が小さいか？<br>○確かめてみよう。<br>標本の大きさ10<br><br><br>標本の大きさ20<br><br><br>20個のばらつきの方が小さいのはなぜか？理由を説明してみよう。</td><td>ほぼ連続型一様分布をしている母集団から標本を取り出す。<br><br>作業40人分の資料を持ち寄り，集計したものを渡す。これをヒストグラムにしてみよう。<br>標本の大きさ10<br>平均67.2，分散0.38<br>標本の大きさ20<br>平均67.0，分散0.14</td></tr>
</table>

| 学習活動（○：指導，●：指示，発問，□活動・予想される反応） | 指導上の留意点・評価 |
|---|---|
| ●1の目が3カ所，2の目が2カ所，3の目が1カ所ついたサイコロをいくつか投げる。このとき，1個から5個投げたときの出た目の平均の値とその確率を求めよ。<br>□4個，5個の分布は以下である。横軸は平均，縦軸は確率を表す。 | 離散型分布でも試す。<br>手計算だと大変であるが，以前，確率だけは求めている。 |
| 4個投げる　　　　5個投げる<br><br><br><br> | 分布表もエクセルで作ってみる。<br>個数を増やすと徐々に正規分布に近づいていることが分かるか。 |
| ○このように，正規分布と思われる分布が出てきた。<br>●平均と分散はどうなっているか？具体的な数値で答えてみよう。 | |
| 1個のとき，$E(X)=\frac{5}{3}$，$V(X)=\frac{5}{9}$，<br>2個のとき，$E(X)=\frac{5}{3}$，$V(X)=\frac{5}{18}$，<br>3個のとき，$E(X)=\frac{5}{3}$，$V(X)=\frac{5}{27}$，<br>4個のとき，$E(X)=\frac{5}{3}$，$V(X)=\frac{5}{36}$，<br>5個のとき，$E(X)=\frac{5}{3}$，$V(X)=\frac{5}{45}$ | 標本平均の分散は，母集団分散の $\frac{1}{\sqrt{n}}$ である。 |
| ○分散の値が規則正しく小さくなっている。<br>●全国の高校生男子の身長から1000人の高校1年生の身長を無作為に取り出す。令和元年では，この母集団は平均168.3cm，標準偏差5.93cmの正規分布をしている。これから，いろいろな分布を作り出してみよう。 | 分布も正規分布に近づく。 |
| ○例．取り出したデータの差の分布，2乗の分布，10人分の平均の分布（既習）など | 偏差の平方も面白い。 |
| ●ここではある値からの偏差の散らばりを調べてみよう。とくに平均からの偏差のデータだとどうなるか？さらにそれを標準偏差でわったデータだとどうなるか？補足2参照。 | 誤差から出てくる分布を想定している。 |
| ○工場などで製品を作るとき，必ず正規の長さからの誤差が発生する。この誤差の分布が正規分布になることも分かる。 | 標準化されたデータの分布になる。<br>必ず誤差は発生する。 |

【補足１】ここでは，正規分布に近づける２つの事例を挙げた。

1) いくつかのサイコロの目の和の平均の分布を代表例とする，有限母集団から得られる起こりうる確率で表されている分布。二項分布の例。

2) 母集団から取り出した標本からなるデータで表される分布。誤差の例。

【補足２】1000 人の身長と平均との差の分布とそれを標準化したデータの分布

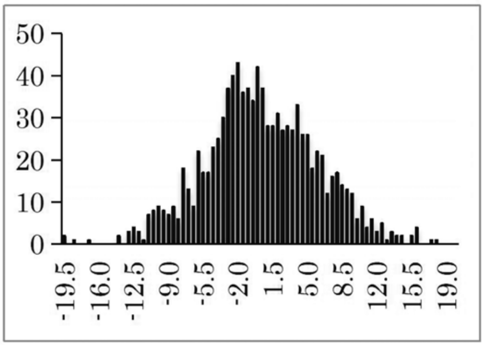



図　偏差の分布

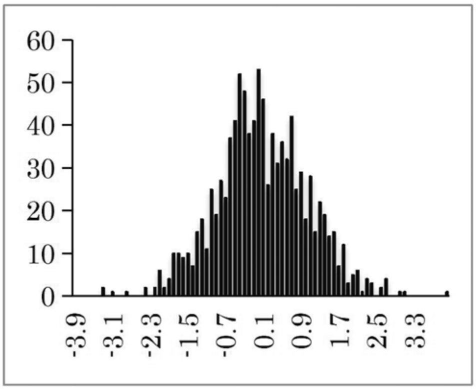



図　標準化した分布

## ４－３．指導事例３

(1) 表題：母集団の情報をどのようにして得るか，それはどれくらい確かな情報なのか？

(2) 本時の目的

①推定をするときの信頼区間の必要性を認識する。

②与えられた情報をどのように生かすかを議論しながらすすめていける。

③得られた結果に対する検討を行うことが大切であることを認識する。

(3) 本時の指導のねらい

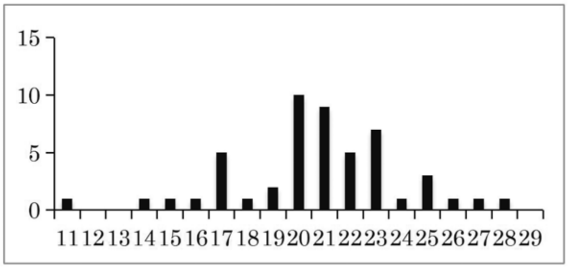
２色の玉が多く入った袋から，１個の玉を取り出すとき，どれだけのことが分かるのかを試してみることで，推定をするときの信頼区間の必要性を認識させる。その後，繰り返した試行から得られた情報は，どのような意味を持つのかを考えさせる。最後に，割合が定まっている資料からの実験を通して，結果に対して議論させる。

(4) 本時の教材観：高等学校で要求されている母集団の特性の推定は，母平均と母比率である。なぜたった１回の標本の実現値から，母集団の性質が推定できるのかは，つねにつきまとう疑問である。実際の実験を通して，信頼区間というある範囲を求める必要に気づくことで，今後の推定の場面でもその信頼区間の数値をどのように解釈するかの考えが定着するのではと思う。

(5) 学習計画：数学Ｂ「統計的な推測」は既習であるとする。

(6) 本時の授業展開

<table>
<tr><th>学習活動（○：指導，●：指示・発問，□活動・予想される反応）</th><th>指導上の留意点・評価</th></tr>
<tr><td>●赤玉と白玉が合わせて800個入っている袋から，無作為に50個の玉を取り出して白玉の個数を数えると19個だった。<br>●この袋に入っている白玉は，およそ何個と推測されるか？<br>□　$800:50=x:19$ から $x=\frac{800\times19}{50}=304$（個）</td><td>1回の取り出しで袋の中の白玉を予想する。</td></tr>
<tr><td>●19個取り出したのがたまたまであって，実際の白玉の個数はどれくらいかを調べるにはどうすればよいか？</td><td></td></tr>
<tr><td>□もっと取り出したらどうだろう？</td><td>情報を増やすことに気づく。</td></tr>
<tr><td>●実は，２回目に50個の玉を取り出して白玉の個数を数えると18個，同様に３回目は，20個，さらに合計30回取り出すと順に<br>19，18，20，15，21，22，17，20，18，19，20，21，23，18，19，<br>20，18，19，20，16，17，19，18，21，24，18，21，17，19，20　個<br>であった。白玉の個数は，何個あると推測されるか？</td><td>復元抽出で取り出している。事前に用意しておく。<br>取り出した順番は気になる？</td></tr>
<tr><td>□　$x$ を取り出した個数とみて $\frac{800\times x}{50}$ の30個の平均を求める。<br>○度数分布から平均を求めるのと同じだね。30個のデータの度数分布を作ってみよう。</td><td></td></tr>
<tr><td>
<table>
<tr><td>$x$</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>計</td></tr>
<tr><td>$f$</td><td>1</td><td>1</td><td>3</td><td>6</td><td>7</td><td>6</td><td>4</td><td>1</td><td>1</td><td>30</td></tr>
<tr><td>$xf$</td><td>15</td><td>16</td><td>51</td><td>108</td><td>133</td><td>120</td><td>84</td><td>22</td><td>23</td><td>572</td></tr>
</table>
</td><td>便宜上，横書きにしている。</td></tr>
<tr><td>これから，$\frac{800\times572}{50\times30}=305.1$　約305（個）</td><td></td></tr>
<tr><td>□30個のデータの平均を用いる点では，より優れてはいる。しかしこれだと，どれくらい確かなのかが分からない。</td><td>一理ある。</td></tr>
<tr><td>●度数分布表の散らばり具合が使えないか？</td><td></td></tr>
<tr><td>□散らばり具合でいうと18，19，20に全体の約60%がはいっているから，割合 $\frac{18.5}{50}\sim\frac{20.5}{50}$ から求めて，約　296個～328個</td><td>範囲で求めたことは評価できる。</td></tr>
<tr><td>●分散は使えないかな？</td><td></td></tr>
</table>

| 学習活動（○：指導，●：指示・発問，□活動・予想される反応） | 指導上の留意点・評価 |
|---|---|
| □度数分布から，取り出された白玉の個数の平均は $\frac{572}{30}=19.07$，標準偏差は1.73である。標本の大きさも30以上だから，標準偏差は，1.73で代用する。大きさ30の標本の平均値を信頼度95％で推定する。$19.07\pm1.96\frac{1.73}{\sqrt{30}}=19.07\pm0.619$ | |
| これから平均値の信頼区間は，［18.451，19.689］<br>よって，　個数は，［295.216，315.024］約　295個〜315個 | この方法は既習である。 |
| □毎回 $\frac{19.07}{50}$ の比率で30回取り出したとすると，次のように求めました。大きさ30の標本の標本比率は，$\frac{19.07}{50}$ として，信頼度95％で推定する。$\frac{19.07}{50}\pm1.96\sqrt{\frac{1}{30}\cdot\frac{19.07}{50}\cdot\frac{30.93}{50}}=\frac{19.07}{50}\pm1.96\cdot\frac{24.287}{50\sqrt{30}}$<br>$=0.3814\pm0.1738$　平均値の信頼区間は，［0.208，0.555］<br>よって，　個数は，［166.4，444］約　166個〜444個<br>○区間が広すぎるような気がします。 | 発表してもらう。<br>毎回の取り出し方が一定としたため，散らばりが小さくなり，より信頼区間が狭くなった。 |
| ●その原因は何だろう？<br>○やはり30個のデータの散らばりが大切なのだ。<br>●この他，白玉が何個あるかを調べる方法はないか？<br>□もっと多くデータを取り出すしかないか？信頼区間を狭くする。<br>●今度は実験してみよう。実際に赤玉320個，白玉480個から，大きさ50の標本を50回取り出してみる。どんなことが分かるか？<br>以下は，1回だけ取り出したときの白玉の個数の分布。 | 正しいかどうか不明の比率を用いたこと，標本の数が少ないことが原因。毎回の取り出し方を一定にしたのも原因。 |
|  | 意外に散らばっている。<br>本来の平均は320個であるが，平均は331個である。 |
| ○この場合の標本平均が20.7，標準偏差が3.21であり，これから母平均の95％の信頼区間を作ると，［19.81，21.59］<br>個数では，317個〜345個<br>□320個は入っている。 | 95％の確率で320個の場合を含んでいる。<br>含まないときもある。 |

## 5．まとめ

高校生に求める統計教育は，どのようなものが適切であろうか。その議論は今後も続いていくだろう。これまでの数学の学習や指導は，数学の演繹性を大切にすることであると言える。証明を最も大切なものとし，正しい答えを導かせるための指導であった。またそれが数学を理解する最良の方法であるとしてきた。統計はどうであろうか。推定や検定といっても，最終的に正しいことはわからないし，間違いのリスクさえ伴っている。しかし世の中は，様々な数値データに溢れ，その中で生きていくためには，帰納的により正しい道を選ぶしかない。コンピュータの発達により大量のデータから情報が得られるとしても，それを信じるかどうかは，自分次第なのである。正解などない。

高校で求められる統計教育は，答えのない教育そのものではないだろうか。信頼区間を求めるだけで終わりとせず，検定のための確率の値が求められたことで終わりとしてはいけない。形通りの問題は解けても，見たこともない問題に対し，どのようにアプローチしていけるのかの力が試されているのである。自分達の考えを持ち寄り，いろいろ調べ仲間と議論しながらより最善なものに近づいていく。そういう教育が生徒にとって今後を生きていく力になるのではないか。

### ［参考文献］

基本統計学［第４版］　宮川公男　有斐閣（2015年）
中学校学習指導要領（平成29年告示）解説　数学編　文部科学省（2017年）
高等学校学習指導要領（平成30年告示）解説　数学編　理数編　文部科学省（2018年）
数学教育３月号No.737，p84～p87そのパック詰め，不公平にならない？峰野宏祐　明治図書（2019年）