情報の授業で統計を扱う*

中西渉
学校法人名古屋学院 名古屋高等学校
〒 461–8676 名古屋市東区砂田橋 2–1–58
watayan@meigaku.ac.jp

高校の数学の教科書には統計分野があるにもかかわらず，大学入試でほとんど出題されないため，授業で扱われることはないに等しい．教科「情報」ができた今も統計にまともに向き合っていないことに変わりはない．筆者はこの状況に不満を抱いている．そこで，情報 B の「モデル化とシミュレーション」から派生して，簡単に統計的分野を扱ってみた．その実践について報告を行なう．

1 高校の授業における統計

1.1 数学

現行の高等学校の数学の指導要領 [1] では，数学で統計的な内容を扱うのは

数学 A 「場合の数と確率」. . . 期待値
数学 B 「統計とコンピュータ」. . . 標準偏差，相関係数
数学 C 「確率分布」. . . 標準偏差
「統計処理」. . . 正規分布，母平均の推定

とされている．しかし，数学 B・C のこれらの分野が大学入試で出題されることは極めて稀であり，出題されたとしても選択問題として回避できるため，その分野を履修しない高校生は多い．すなわち，標準偏差も知らないまま卒業する生徒も相当数いるということである．

1.2 情報

一方，情報の指導要領 [2] には「統計」という単語は一度しか登場しない．情報 C の「(3) 情報の収集・発信と個人と責任」の取り扱いについて

> 表計算ソフトウェアなどの簡単な統計分析機能やグラフ作成機能などを扱うようにする．

と触れているだけで，この「簡単な統計分析機能」がどの程度のことなのかは明示されていない．

いくつかの教科書では表計算ソフトの `AVERAGE`，`MAX`，`MIN`，`RANK`，`STDEVP` などの関数を使った例が紹介されているが，計算させた値の意味は説明していない．`MAX` や `MIN` なら説明は必要ないが，標準偏差 `STDEVP` はそういうわけにはいかない．たとえば [3] には「散らばり具合」という説明と分布図の例があるが，その定量的な意味に言及してはいない．

2 本校での実践

以下，本校の情報の授業での実践について述べる．教科書にない内容なので，教材はプリントを用意している．中心となるのは，標準偏差に意味づけを行なうことである．

2.1 値のバラツキと標本の大きさ

2.1.1 ビュッフォンの針

ビュッフォンの針というのは

> 等間隔に平行線を無数にひき，その間隔の半分の長さの針を無作為に落としたとき，針が平行線のどれかと共有点を持つ確率はどれだけか．

という問題である．もちろん積分によってその値を求めることはできるのだが，高 1 ではそこまで習っていないし，せっかくパソコンがあるのだからモンテカルロ法で近似値を求めることにする．

* Treating statistics in a class of information. Wataru Nakanishi (Nagoya High School)

授業では PEN [*1]用のプログラムを配布して何度か実行させ，確率の値（とその逆数）を生徒に記録させる（図 1）．「この確率の逆数は君たちがよく知ってるある数なんだけどね...何だと思う？」と尋ねると，何人かが「ひょっとして π ですか」と口にする．正解に気づくことにももちろん意味はあるが，それよりも生徒に注目してほしいのは値のバラツキが大きいことである．プログラムでは 1000 本も針をばらまいているのに，1 の位さえ保証されない．

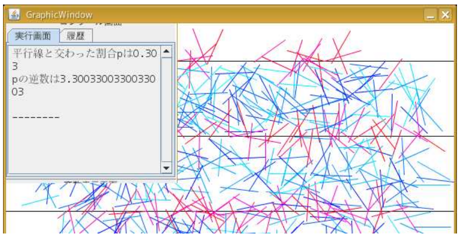


図 1　ビュッフォンの針

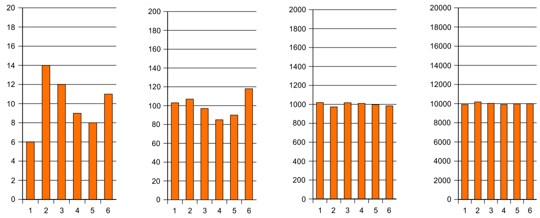
2.1.2　サイコロ

「バラツキが意外に大きい」という印象を強めるために，サイコロを何度も振って各目の出た回数を記録するシミュレーションを行なう．

たとえばサイコロを 60 回振れば，各目はおよそ 10 回ずつ出るはずである．しかし表計算のマクロを使ってそのシミュレーションをやってみると，5 割程度の誤差が出ることも多く，想像以上に誤差が大きいことがわかる（図 2）．しかし，回数を 600 回，6000 回，60000 回と増やしていけば，バラツキの相対的な値はどんどん小さくなる．これは期待値や独立試行だけでは簡単には説明できない．そこで，この値のバラツキと試行回数の関係を評価するためのシミュレーションを行なう．

2.1.3　円周率を求める

図 3 は，モンテカルロ法といえば真っ先に思い出されるシミュレーションである．ランダムに発生させた点が 1/4 円に入る確率は $\pi/4$ だから，当たりの数を総数で割ればおよそ $\pi/4 = 0.785\ldots$ になるので，逆にその割合を 4 倍することで円周率の近似値を求められるということだ．実行させてみると，かなり近い値を得る者もいれば，1 の位が 2 になってしまう者もいる．



(a) 60 回　(b) 600 回　(c) 6000 回　(d) 60000 回

図 2　サイコロを何回も投げる

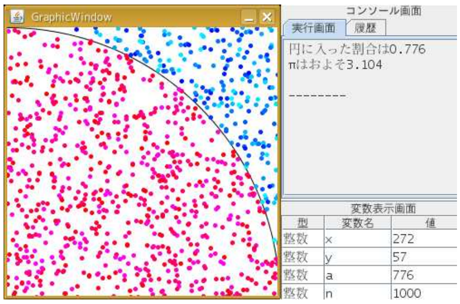


図 3　円周率を求める

点の個数を増やせば誤差を減らすことができることはわかっている．プログラム中の点の個数を変えて何度もやれば値のバラツキ具合を観察することができるのだが，それを手作業でやるのは面倒なので，自動的に反復実行するマクロを使った表計算のデータを配布して実行させる（図 4）．点の個数を 10,100,... と増やしていき，それぞれ 15 回[*2]ずつ前述のシミュレーションを行なって結果を表示させる．ここではその 15 回の最大値と最小値の差（これをバラツキと考える）に着目する．

回数が増えればバラツキは減るが，明らかに反比例ではない．よく見ると，回数が 100 倍になったところでバラツキが 1 桁小さくなっている．このことから，**バラツキは試行回数（標本の大きさ）のルー**

[*1] xDNCL を用いた初学者向けプログラミング教育環境．
`http://www.media.osaka-cu.ac.jp/PEN/`

[*2] 15 回というのは，本校の情報教室のディスプレイでスクロールせずに全体が見られるように決めた回数．

	A	B	C	D	E	F	G	H
1	回数	10	100	1000	10000	100000	1000000	10000000
4	1回目	0.8	0.88	0.796	0.7846	0.78624	0.785903	0.7853888
5	2回目	1.0	0.83	0.773	0.7902	0.78733	0.785915	0.7856327
6	3回目	0.9	0.77	0.799	0.7830	0.78512	0.785347	0.7854017
7	4回目	0.9	0.78	0.785	0.7829	0.78586	0.784810	0.7853023
8	5回目	0.9	0.79	0.790	0.7927	0.78709	0.785561	0.7854281
9	6回目	0.9	0.76	0.788	0.7791	0.78702	0.785116	0.7854764
10	7回目	1.0	0.80	0.780	0.7888	0.78468	0.785919	0.7854075
11	8回目	0.9	0.77	0.769	0.7856	0.78393	0.785014	0.7852441
12	9回目	0.9	0.83	0.791	0.7871	0.78421	0.784860	0.7853896
13	10回目	0.6	0.75	0.775	0.7884	0.78755	0.785443	0.7855042
14	11回目	0.8	0.78	0.796	0.7940	0.78431	0.785625	0.7853359
15	12回目	0.7	0.80	0.814	0.7884	0.78638	0.785581	0.7854387
16	13回目	0.8	0.75	0.787	0.7849	0.78753	0.785329	0.7854472
17	14回目	0.8	0.78	0.810	0.7920	0.78572	0.785185	0.7852999
18	15回目	0.7	0.80	0.796	0.7851	0.78669	0.785754	0.7853102
19	最大	1.0	0.88	0.814	0.7940	0.78755	0.785919	0.7856327
20	最小	0.6	0.75	0.769	0.7791	0.78393	0.784810	0.7852441
21	最大-最小	0.4	0.13	0.045	0.0149	0.00362	0.001109	0.0003886
22								

図 4　回数とバラツキの関係

トに反比例していると推測することができる．

2.2　正規分布

統計でよく用いられる値として，平均と標準偏差があることを紹介し，表計算ソフトで計算できることを確認させる．ところが，生徒にしてみれば平均は容易に想像できるが，標準偏差が何を意味しているのかはすぐにはわからない．

そこでいくつかの度数分布図（身体測定のデータや試験の得点分布を用いた）を見せて，これが一定の形—正規分布に近い形になることを示す．その上で，標準偏差はそのグラフの裾野の広さを表しており，$m \pm k\sigma$ の範囲に全体の何 % が含まれるかがわかっていることを説明し，それを検証できるようにマクロを組んだ表計算データを生徒に配布して確認させる（図 5）．

	A	B	C	D	E	F
1	身長			A2:A417に416人分のデータが保存されている		
2	164.3					
3	173.3	○		平均(m)	171.07	
4	171.7	○		標準偏差(σ)	5.52	
5	172.6	○				
6	165.5			k	1.00	
7	181.1					
8	170.8	○		m+1σ	176.6	
9	164.7			m-1σ	165.5	
10	170.3	○		上記範囲の人数	281	
11	170.7	○		その割合(%)	67.5	
12	163.0					
13	186.4			計算	クリア	
14	178.1					
15	171.6	○				

図 5　$m \pm k\sigma$ に含まれる割合

2.3　推定と検定

2.3.1　母平均の推定

本来なら数学的に公式を作るのだが，これまでに得られた結果を強引に組み合わせれば，次のように母平均の信頼区間を得ることができる．母平均を m，標本平均を $\overline{x}$ とするとき，

1. 1 つのデータが $m \pm 1.96\sigma$ に含まれる確率は 95%.
   $\therefore$ 1 つのデータと m の差は，95% の確率で 1.96σ 以下.
2. バラツキ（m との差）は標本の大きさのルートに反比例する．
   $\therefore$ $\overline{x}$ と m の差は，95% の確率で $1.96\sigma/\sqrt{n}$ 以下.
   $\therefore$ 95% の確率で m は $\overline{x} \pm 1.96\sigma/\sqrt{n}$ に含まれる．
3. m の 95% 信頼区間は $\overline{x} \pm 1.96\sigma/\sqrt{n}$ である．

公式ができたところで，いくつか計算練習をさせてみる．

2.3.2　母比率の推定

この公式を導くのは難しいので，これについては公式を与えて計算練習をさせる．出来合いの公式を与えてしまうのは面白くないが，計算結果の方で楽しむことを考えることにする．たとえば「1 の目が 100 回中 20 回出たサイコロの，1 の目のでる確率の 95% 信頼区間」と「. . . 10000 回中 2000 回. . .」を比較することで，2.1.2 のサイコロのシミュレーション結果の理由がわかってくる．

2.3.3　検定

サイコロが正常かどうかを論じるのは，検定を用いるのが筋ではある．そう考えて簡単な検定までを授業で扱ったこともあるのだが，生徒の反応と試験の結果はすこぶる悪かった．「. . . は棄却できないので，成績が等しくないとは断言できない（しかし等しいと断言するわけではない）」のような否定の連続は混乱しやすいし，公式に代入する値を勘違いしやすいし，かといって十分な時間もかけられないということで，2 年目からは推定だけを扱うことにした．

実は，この話を長く続けると生徒から「情報の授業でどうして数学を...」という不満が出て集中力が欠けてしまうということも，やめてしまった理由の一つである．

2.4 数学 B・C との対応

この内容は 1.1 で述べた数学の内容と重なる点が多い．しかし，本実践では標準偏差の定量的な意味づけを中心しているのに対し，数学 B「統計とコンピュータ」では表計算ソフトを用いて標準偏差や相関係数を求めさせているがその数値の意味についての説明はないし，数学 C「確率と確率分布」でも分散の性質について述べている他は数学 B と同様である．したがって本実践の内容と重なるのは数学 C「統計処理」であるが，結論に至るアプローチの方法が異なっている．数学 C では分散の性質を用いて数学的に導いているが，本実践ではシミュレーションの結果から推測する方法をとっている。前者の方が厳密であることは言うまでもないが，値を観察する後者の方法にも別の効果があると考えている。

また，2.2 で用いた身体測定のデータは数学 B で扱われている相関係数の題材としてとりあげることもできたのだが，授業時間数の都合で触れることができなかった．

3 まとめ

この実践にはいくつも問題があることはわかっている．たとえば，強引な導入を行なっている割には公式が不正確であるが，生徒たちがいずれ統計を用いるときになれば改めて勉強することになるのだから，正しいやり方はそのときに身につければいい．

また，そもそもこれは「情報」でなく数学で扱う内容だという意見もある．それは十分に承知しているのだが，しかし冒頭で述べたように，実際に数学の授業でこの内容が扱われることはほとんどない．だからこのような不完全なものでも，まったくやらないよりマシであると考えている．

参考文献

[1] 文部省. 高等学校学習指導要領 第 2 章第 4 節 数学, 3 1999.

[2] 文部省. 高等学校学習指導要領 第 2 章第 10 節 情報, 3 1999.

[3] 岡本敏雄・山極隆ほか 9 名. 最新情報 C. 実教出版.