お茶の水女子大学附属高等学校　１年生

数学探究「相関係数の正体～ベクトルで解釈する相関係数の意味」

附属高等学校　三橋一行

１．学習のねらい

データの分析で相関係数について学ぶが，相関係数の値が-1 から 1 までの数しかとらないのはなぜか。そのことを，相関係数を求める公式から直感的に理解することは難しい。また，その公式も共分散をそれぞれの標準分散同士の積で割るという形である。これが，どんな意味を持っているのかということも分かりづらい。そこでその意味をベクトルの学習を通して，幾何学的に理解して見るという授業を考案した。相関係数の理解とデータの分析と数学のより深い関係を知ることで，上のような疑問に答えを見つけられることねらいとしている。

２．育てたい力（資質・能力）

- データの分析に使用する公式が，ベクトルを用いて幾何学的に解釈できる。
- データ分析の式や方法は，数学の活用・応用であり数学の便利さや良さが生かされていることを知ることができる。
- 既習事項を複数用いることで，数学の学習内容の復習と数学の応用力を高めることができる。

３．学習の展開

① 学習指導案

<table>
<tr><th>学習活動</th><th>指導の手立て留意点</th></tr>
<tr><td>
本時の流れを簡単に説明し，復習を兼ねて下の表を完成させ，相関係数を手計算で求めてもらう。電卓の使用も許可している。

●発問
「表埋めて相関係数をもとめよう。」

1　問題設定

下の表は、あるクラス（在籍2名）の英語と数学の期末試験の得点である。
<table>
<tr><th>生徒</th><th>x(英語の得点)</th><th>y(数学の得点)</th></tr>
<tr><td>A</td><td>40</td><td>90</td></tr>
<tr><td>B</td><td>80</td><td>70</td></tr>
<tr><td>計</td><td>120</td><td>160</td></tr>
</table>
（英語の平均点）＝60　　（数学の平均点）＝　80

下の表を埋めて、英語と数学の得点の相関係数 r を求めよ。
<table>
<tr><th>生徒</th><th>x</th><th>y</th><th>$x-\overline{x}$</th><th>$y-\overline{y}$</th><th>$(x-\overline{x})^2$</th><th>$(y-\overline{y})^2$</th><th>$(x-\overline{x})(y-\overline{y})$</th></tr>
<tr><td>A</td><td>40</td><td>90</td><td>－20</td><td>10</td><td>400</td><td>100</td><td>－200</td></tr>
<tr><td>B</td><td>80</td><td>70</td><td>10</td><td>－10</td><td>400</td><td>100</td><td>－200</td></tr>
<tr><td>計</td><td>120</td><td>160</td><td>/</td><td>/</td><td>800</td><td>200</td><td>－400</td></tr>
</table>
</td><td>
★通常「数学C」で学ぶベクトルに関しては，本校のSSH設定科目「数学探究」の中で，2次元ベクトルの内積の単元までは学習している。そのた上で，本時の授業が設定されている。

★復習なので，時間を区切り生徒自身に求めさせる。必要があれば，電卓を使用しても良いことにする。
</td></tr>
</table>

$$r=\frac{-400}{\sqrt{800}\sqrt{200}}=-1$$

以上のやり方を踏まえて，文字式で一般化を図る。

2　一般化

一般にそれぞれの得点を文字で以下の様に置くと、相関係数 r はどうなるか。

生徒	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
A	x_1	y_1	a_1	b_1	${a_1}^2$	${b_1}^2$	a_1b_1
B	x_2	y_2	a_2	b_2	${a_2}^2$	${b_2}^2$	a_2b_2
計	x_1+x_2	y_1+y_2	/	/	${a_1}^2+{a_2}^2$	${b_1}^2+{b_2}^2$	$a_1b_1+a_2b_2$

●発問
「この式は，どこかで見たことが無いか？」

●発問
「この式の r とは，何だろうか」

$$r=\frac{a_1b_1+a_2b_2}{\sqrt{{a_1}^2+{a_2}^2}\sqrt{{b_1}^2+{b_2}^2}}$$

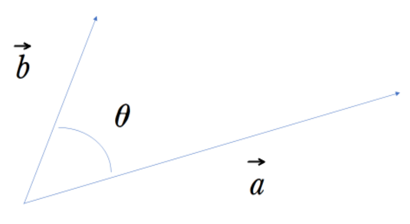
★　相関係数を求め終わったところを見計らって，指名するなどして，答えを発表してもらう。

★　表と相関係数の公式，求めた数値が正しいか答え合わせする。次のstepへ進めるよう，ここまでの間違いや誤解は修正しておく。

＜要注意＞
ベクトルで考える場合，データ数はベクトルの次元として現れる。今回は2次元べくとるなので，データは2件しかない設定である。この状態では，2点しかないので，直線が引けていまい，傾きが1もしくは−1の直線となってしまうそのため相関け数は1か−1になってしまう。

★　表中にあるように，ベクトルaとベクトルｂを設定する。
原点が，ｘの平均値とｙの平均値をｘ座標，ｙ座標にもつ点に移されるのであるが，深入りしない。相関係数が平均値からのズレを利用している量であることもこのことからわかる。

※この式では，まだひらめかない可能性があるので，次のようにベクトルとその成分を与える。

$\boxed{2}$の式を，$\vec{a}=(a_1, a_2)$，$\vec{b}=(b_1, b_2)$ として、rを$\vec{a}$と$\vec{b}$を用いて表せ。 $\displaystyle r=\frac{a_1b_1+a_2b_2}{\sqrt{{a_1}^2+{a_2}^2}\sqrt{{b_1}^2+{b_2}^2}} = \frac{\vec{a}\cdot\vec{b}}{\lvert\vec{a}\rvert\lvert\vec{b}\rvert}$	※　ここまで来ると，かなりの生徒がベクトルの内積の公式で，COS θを表すものであることに気づく。
●発問 「相関係数とは，何を表しているのか。」 相関係数 r の正体は何であると考えられるか ※　$-1 \leqq r \leqq 1$　となる理由が分かりましたか？　はい　・　いいえ $\displaystyle r=\frac{a_1b_1+a_2b_2}{\sqrt{{a_1}^2+{a_2}^2}\sqrt{{b_1}^2+{b_2}^2}} = \frac{\vec{a}\cdot\vec{b}}{\lvert\vec{a}\rvert\lvert\vec{b}\rvert} = \cos\theta$ **この式は，ベクトルの応用で学んだ。2本のベクトルのなす角を求める公式である。そして，これはなす角をθとするとCOS θ　に等しいものである。**	※　データを成分に持つベクトル同士がどのくらい同じ方向を向いているか，その尺度として，2本のベクトルのなす角をθとしたCOS θの値で2つのベクトルの類似度を表したものが相関係数であるという解釈ができる。 出来るだけ生徒に考えさせて，このような解釈を導き出してほしい。
相関係数の意味を幾何学的（図形的）に解釈すると。 	※　図に表すと左ようになる。 ※　相関係数を幾何学的に解釈すると，COS θなので，－１と１の間の数値しかとらないことがわかる。
以下，ワークシートの残りの課題に取り組む。	※　ワークシートの残りの部分に取り組んで学習事項を確認し，学習を振り返っての記述も行わせる。授業終了まで時間をとる。質問なども適宜受ける。

相関係数の正体とは？

(　　)組(　　)番　名前(　　　　　　　　　)

1　下の表は、あるクラス（在籍2名）の英語と数学の期末試験の得点である。

生徒	x(英語の得点)	y(数学の得点)
A	40	90
B	80	70
計		

（英語の平均点）＝　　　（数学の平均点）＝

下の表を埋めて、英語と数学の得点の相関係数 r を求めよ。

生徒	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
A	40	90					
B	80	70					
計			/	/			

2　一般にそれぞれの得点を文字で以下の様に置くと、相関係数 r はどうなるか。

生徒	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
A	x_1	y_1	a_1	b_1			
B	x_2	y_2	a_2	b_2			
計			/	/			

3　2の式を、$\vec{a}=(a_1, a_2)$、$\vec{b}=(b_1, b_2)$ として、r を $\vec{a}$ と $\vec{b}$ を用いて表せ。

4　相関係数 r の正体は何であると考えられるか

※　$-1 \leq r \leq 1$　となる理由が分かりましたか？　はい　・　いいえ

5　相関係数の意味を幾何学的（図形的）に解釈すると。

-1-

相関係数の正体とは？

(　　)組(　　)番　名前(　　　　　　　　　)

6　相関係数 r が次の値であるとき、$\vec{a}$，$\vec{b}$　の2本の矢印とそのなす角 θ を用いて図示せよ。また，θの値も答えよ。$|\vec{a}|>|\vec{b}|$とし、$0°\leq\theta\leq180°$ とする。

(1)　$r=1$　　(2)　$r=\frac{1}{2}$

(3)　$r=0$　　(4)　$r=-\frac{\sqrt{3}}{2}$

(5)　$r=-1$

7　相関係数 $r=$ 0.25と 相関係数 $r=$ 0.5を比較して、「相関係数が2倍になっているという表現は正しいか？　その正誤を根拠を持って答えよ。

8　【チャレンジ】英語と国語の成績に何らかの正の相関関係があるとき、英語得点データベクトル $\vec{a}$ から、国語得点データ ベクトル$\vec{c}$ の影響を取り除いた純粋な英語のデータデータベクトル $\vec{d}$ を図示せよ。ただし、垂直なデータベクトル同士は互いに影響しないことは既知とする。

9　相関係数とベクトルの関係を学んで感じたことや興味を持った点を書いてみよう。

-2-

② 授業活動の実際

実際の授業は，広い教室で 120 名ほどに対して一斉授業をおこなった。スライドを利用し板書の時間の短縮と大型教室ゆえの黒板の見にくさ解消をはかった。授業内容としては，薄目かと感じたが，1 コマ（45 分間）の授業の枠にほぼピタリと当て嵌まる感じであった。

比較的うまく行った授業と言えるだろう。

４．授業を振り返って

この授業のために，数学Ｃの内容であるベクトルのうち平面ベクトルの内積のところまでの学習を週に 1 回 45 分の授業で，4 週分ほどの短時間で生徒には先取り学習してもらっている。

ベクトルと相関係数という一見して全く別物であるとしか思えないような内容が実はつながっており，相関係数なるものはコサインの値と全く同じものであるという事実に驚いた生徒は多かった。しかし，ベクトルに十分に慣れておらず，その驚きが十分に得られなかった生徒もいる。指導時期としては，ベクトルに十分に慣れた数学Ｃのベクトルの学習が終わったあたりが適切であろう。今回 1 年生に持ってきたのは，①数学探究という授業で，数学の有用性と統合性について驚きをもってもらいたかったこと，②データの分析を行うにあたって，疑似相関関係を引き起こす要素を仮定してその影響を取り除くことによってより正確なデータ分析を行うことに役立てて貰うこと。の２つの目的のためであった。②については今回大幅な時間切れとなってしまって全く手つかずとなった。3 次元のベクトルまで，扱えるようになると相関係数の幾何学的解釈も多様性が生まれてくる。垂直をなす２つのデータベクトルは互いに影響しあわない（これについては大学レベルの証明が必要である）ことがわかっている。そこで，任意のデータベクトル **a** について，疑似相関を引き起こすと考えられるデータベクトル **c** への正射影ベクトルをつくり，この正射影ベクトル **a’** ともとのベクトル **a** の差をとると，疑似相関の要素を取り除いたデータベクトル **x** ができる。同様にデータベクトル **b** についても同様にベクトル **c** の影響を受けないデータベクトル y をつくる。すると， a と b の相関関係から c の要素を抜いた x と y の相関係数がえられることになる。これによって，疑似相関関係から真の相関関係が調べられることになる。相関係数は単にコサインと同じということだけではなく，ベクトルの利用によってより正確な相関関係の分析が行えるのである。このように，統計分析の手法と純粋数学は影響を与えつつ発展してきたし今後もそうあるべきである。

今回，後半の部分は実際には授業ができなかったが，機会を見て実施していきたい。また，今回行えた授業の部分においては，発問の仕方やワークシートの問いの設定など改善の余地がある。こちらの方も改善を考えて再度授業ができればよいと考えている。