相関係数と偏相関係数の関係
— ベクトルによる幾何学的解釈の証明 —

数学科　三　橋　一　行

1．はじめに

確率変数 $X_1, X_2, \cdots X_n$，$Y_1, Y_2, \cdots Y_n$ から、それぞれの平均値 $\bar{X}$, $\bar{Y}$ をひいて並べたベクトル $\boldsymbol{x}$，$\boldsymbol{y}$ を、

$$\boldsymbol{x} = \begin{pmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{pmatrix}$$

とする。また、$\boldsymbol{x}$，$\boldsymbol{y}$ のなす角を $\theta (0 \le \theta \le \pi)$ とする。$X_i, Y_i \in \mathbb{R}$ $(i = 1,2,3,\cdots n)$．$(X_i, Y_i)(i = 1,2\cdots n)$ について２次元分布を仮定しているが，今回は深く言及しない。

このとき、$\boldsymbol{x}$，$\boldsymbol{y}$ の相関係数 ρ_{xy} は次の（１）式で与えられる。

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad \cdots\cdots \quad (1)$$

また、ベクトルの内積の定義より、（１）式は次のようにも表せる。

$$\rho_{xy} = \cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|} \quad \cdots\cdots \quad (2)$$

したがって、相関係数とは、ベクトルの内積の定義によって解釈すると、ベクトル $\boldsymbol{x}$，$\boldsymbol{y}$ のなす角 θ の余弦に他ならない。

一方、ベクトル $\boldsymbol{x}$，$\boldsymbol{y}$ のそれぞれから、ベクトル $\boldsymbol{z}$ の影響を除いた相関係数 $\rho_{xy\cdot z}$ すなわち偏相関係数 $\rho_{xy\cdot z}$ は、次の式で与えられる。

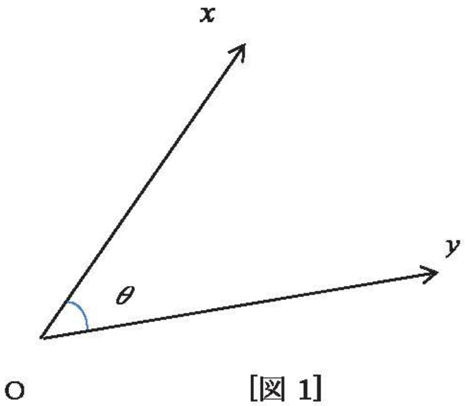
$$\rho_{xy\cdot z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1 - \rho^2_{xz}}\sqrt{1 - \rho^2_{yz}}} \quad \cdots\cdots \quad (3)$$

上の（３）式の右辺は、（１）式によって $\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{z}$ の内から、異なる２つを選んで求められる相関係数の組み合わせによって計算されていることがわかる。

（１）が（２）のようにベクトル表現できたのと同様に、（３）をベクトル表現するとどんなものになるのだろうか。（１）を（２）の様に表現すると幾何学的解釈が出来て、（２）の場合は、２つのベクトルがなす角 θ の余弦になることがわかった。こ

れは、２つのベクトルがどのくらい似ているか（同じ方向を向いているか）という解釈につながる［図 1］。

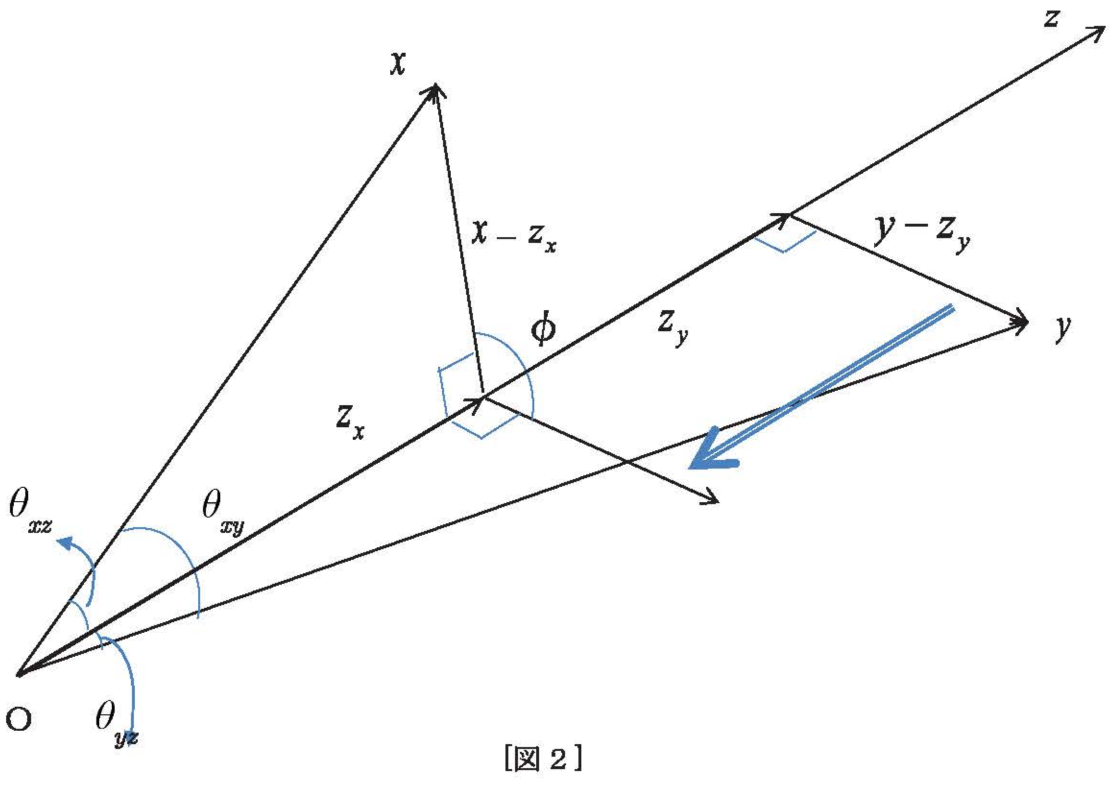
では、（３）式はどのような幾何学的意味を持っているのであろうか。相関係数を用いて偏相関係数を求めているので、（２）式の意味が反映されていると予想される。これらの疑問と予想に解決を試みたいと考えたのが、この研究の動機である。



[図 1]

２．偏相関係数の幾何学的解釈予想

偏相関係数の幾何学的表現はどのようになるか予想を立てて見た。それが以下の［図 2］である。なお、図にあるすべてのベクトルは零ベクトルではなく、任意の異なる２本のベクトルがなす角すべてが、0 [rad] 以上 π [rad] 以下の角としておく。

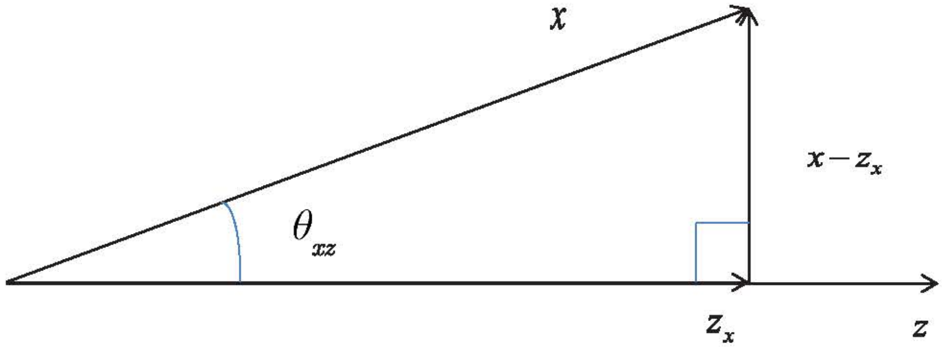


[図 2]

［図2］の説明をする。通常 $\boldsymbol{x}$，$\boldsymbol{y}$ の２つのベクトル（つまり２組のデータ列）で相関関係を調べる場合、相関係数 ρ_{xy} すなわち、$\cos\theta_{xy}$ を求めるのであるが、実際の分析の際には、往々にして、疑似相関という現象が生じる。これは、陰に隠れている「ある要因」が，得られた２組のデータに影響を落とし、本来の相関関係の強弱を狂わせてしまうものである。今、その「ある要因」のデータ列をベクトル $\boldsymbol{z}$ で表すことにする。ベクトル $\boldsymbol{z}$ は、ベクトル $\boldsymbol{x}$，$\boldsymbol{y}$ と同様に、次のように定義する。

$$z=\begin{pmatrix} Z_1-\bar{Z} \\ Z_2-\bar{Z} \\ \vdots \\ Z_n-\bar{Z} \end{pmatrix}$$

偏相関係数というのは、ベクトル $\boldsymbol{x}$ と $\boldsymbol{y}$ からベクトル $\boldsymbol{z}$ の影響を除いた相関係数の事を指すのであるから、まず、ベクトル $\boldsymbol{x}$ と $\boldsymbol{y}$ からベクトル $\boldsymbol{z}$ の影響を取り除くことを考える。まず、ベクトル $\boldsymbol{x}$ は、ベクトル $\boldsymbol{z}$ を用いて次のように分解される。



$$\boldsymbol{x}=\boldsymbol{z_x}+(\boldsymbol{x}-\boldsymbol{z_x})$$

ただし、ベクトル $\boldsymbol{z_x}$ は、ベクトル $\boldsymbol{x}$ のベクトル $\boldsymbol{z}$ への正射影ベクトルであるとする。

当然のことながら、

$$\boldsymbol{z}\perp(\boldsymbol{x}-\boldsymbol{z_x})$$

であるので、この２つのベクトル $\boldsymbol{z}$ と $(\boldsymbol{x}-\boldsymbol{z_x})$ の内積は０に等しく、相関係数も０となり互いに影響を及ぼさないことが分かる。(「2 つの確率変数ベクトルの内積が０⇔2 つの確率変数ベクトルが（統計的に）独立」は別に証明が必要であるが今回はそこには触れない。）したがって、ベクトル $\boldsymbol{x}$ から、ベクトル $\boldsymbol{z}$ の影響を取り除いたベクトルが $(\boldsymbol{x}-\boldsymbol{z_x})$ なのである。

同様に、ベクトル y から、ベクトル z の影響を取り除いたベクトルが $(y - z_y)$ である。したがって、図２の幾何学的解釈とあわせて、

> 「偏相関係数は、ベクトル $(x - z_x)$ とベクトル $(y - z_y)$ の相関係数に等しい、すなわち［図２］でいえば、 $\cos\phi$ に等しい。」　・・・・・(※)

以上が、私が立てた偏相関係数の幾何学的解釈予想である。

３．偏相関係数の幾何学的解釈予想の証明

偏相関係数の幾何学的解釈予想を証明する。すなわち、

$$\rho_{x\cdot y\ \cdot z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}} = \frac{(x-z_x)\cdot(x-z_y)}{|x-z_x||x-z_y|} = \cos\phi$$

証明すべきは、真ん中の等号であるので、以下の等式を証明すればよい。

$$\frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}} = \frac{(x-z_x)\cdot(x-z_y)}{|x-z_x||x-z_y|} \qquad \cdots\cdots(4)$$

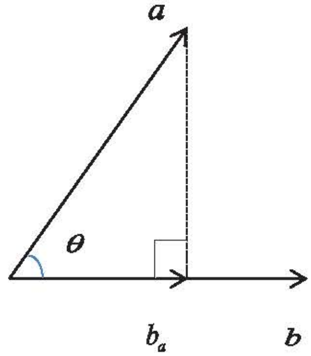
証明に入る前に補助定理を証明しておく。

＜補助定理＞

　ともに０ベクトルでない２つのベクトル a とベクトル b について、ベクトル a のベクトルへの正射影ベクトル b_a は次の式で与えられる。

$$b_a = \frac{a\cdot b}{|b|^2}\,b \qquad \cdots\cdots(5)$$

＜証明＞



　右図より b の単位ベクトル $\frac{b}{|b|}$ を考えると、

$b_a = \frac{|b_a|}{|b|}b$ と表せて、ベクトル a, b のなす角を

$\theta\ (0 \leq \theta \leq \pi)$ とすると、

$$b_a = \frac{|b_a|}{|b|}b = \frac{|a|\cos\theta}{|b|}b$$

$$= \frac{|a|}{|b|}\cdot(\cos\theta)\cdot b = \frac{|a|}{|b|}\,\frac{a\cdot b}{|a||b|}\,b = \frac{a\cdot b}{|b|^2}\,b$$

線形代数の射影行列を用いればさらに統一的に考えられるが、今回はこれで十分で

ある。また、高校の授業の延長として取り上げることも可能であるように配慮したつもりである。では、等式（４）の証明に入る。

まず、相関係数の幾何学的解釈より次の３本の等式が成立する。

$$\rho_{xy} = \cos\theta_{xy} = \frac{x \cdot y}{|x||y|}, \quad \rho_{xz} = \cos\theta_{xz} = \frac{x \cdot z}{|x||z|}, \quad \rho_{yz} = \cos\theta_{yz} = \frac{y \cdot z}{|y||z|}$$

これらを等式（４）の左辺に代入する、分母は *cos* の形で代入する。

$$\frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}} = \frac{\dfrac{x \cdot y}{|x||y|} - \dfrac{x \cdot z}{|x||z|} \times \dfrac{y \cdot z}{|y||z|}}{\sqrt{1-\cos^2\theta_{xz}}\sqrt{1-\cos^2\theta_{yz}}}$$

$$= \frac{\dfrac{x \cdot y}{|x||y|} - \dfrac{x \cdot z}{|x||z|} \times \dfrac{y \cdot z}{|y||z|}}{|\sin\ \theta_{xz}||\sin\ \theta_{yz}|} = \frac{x \cdot y - \dfrac{(x \cdot z)}{|z|} \times \dfrac{(y \cdot z)}{|z|}}{|x||y||\sin\ \theta_{xz}||\sin\ \theta_{yz}|}$$

$$= \frac{x \cdot y - \dfrac{(x \cdot z) \times (y \cdot z)}{|z|^2}}{|x||\sin\ \theta_{xz}||y||\sin\ \theta_{yz}|} = \frac{x \cdot y - \dfrac{(x \cdot z) \times (y \cdot z)}{|z|^2}}{|x - z_x||y - z_y|}$$

これで、分母は証明できたので、この続きは、分子のみを取り出して考える。

分子 $x \cdot y - \dfrac{(x \cdot z) \times (y \cdot z)}{|z|^2}$ が等式（４）の右辺分子に等しいことを示すために、次のように、打ち消し合って０になる項を挟んで変形する。

$$x \cdot y - \frac{(x \cdot z) \times (y \cdot z)}{|z|^2}$$

$$= x \cdot y - \frac{(x \cdot z) \times (y \cdot z)}{|z|^2} - \frac{(x \cdot z) \times (y \cdot z)}{|z|^2} + \frac{(x \cdot z) \times (y \cdot z)}{|z|^4}|z|^2$$

そして、上の式を次のように解釈すれば、

$$= x\cdot y-\frac{(x\cdot z)}{|z|^2}z\cdot y-\frac{(y\cdot z)}{|z|^2}z\cdot x+\frac{(x\cdot z)\times(y\cdot z)}{|z|^4}z\cdot z$$

$$=\left(x-\frac{(x\cdot z)}{|z|^2}z\right)\cdot\left(y-\frac{(y\cdot z)}{|z|^2}z\right)$$

と因数分解できて、補助定理の（５）式を用いると

$$\frac{(x\cdot z)}{|z|^2}z=z_x \quad , \quad \frac{(y\cdot z)}{|z|^2}z=z_y$$

であるので、

$$=(x-z_x)\cdot(y-z_y)$$

以上により、分母、分子ともに等しいことが示せたので、等式（４）が示せた。

しがたって、偏相関係数とは、[図２] の $\angle\phi$ の余弦のことを意味し、各ベクトル間の関係（構造）が [図２] のようであること、つまり偏相関係数の幾何学的解釈予想（※）が正しいことが証明された。

４．高等学校数学科における統計教育への示唆　～結びに代えて～

高等学校における新課程の統計教育がどのようなものになるか詳細はわからない。しかし、聴くところによると、統計的手法の数学的な部分を扱うのではなく、統計手法を用いた分析、つまりはデータ分析法に力が入れられるようである。極論だが、そのような内容であるなら、数学科で扱うものかどうか疑問である。使い方だけであるならば、コンピュータの普及状況から考えて他教科でも十分可能である。数学科において統計を扱うなら、指導内容にある他の単元の「活用」として関わりを持たせたい。統計学は純粋数学とは一線を画されてきたが，それは誤解ゆえの結果である。統計学は高等学校数学教育の延長、さらに大学教養課程の延長上に構築される。そこまで学んだとき、既習の数学力が発揮されて、数学の有用性と存在意義を実感できるのである。この域に達する前に数学や統計学を嫌いになってしまう人々がいるのは、残念でならない。今回紹介した内容は、高等学校数学の「活用」や「数学的活動」の一例として教材化できるだろう。だが、分析のみに偏った統計教育に加え、今回用いた「ベクトル」の単元が選択科目に移行させられてしまうのは、憂慮すべき事態である。