HOKUGA 北海学園学術情報リポジトリ

学校法人北海学園
北海学園大学
北海商科大学

タイトル	相関係数の数学的性質にかんする一考察
著者	木村，和範；KIMURA，Kazunori
引用	季刊北海学園大学経済学会，65(3)：1-14
発行日	2017-12-30

Hokkai-Gakuen Organization of Knowledge Ubiquitous through Gaining Archives.

《論説》

相関係数の数学的性質にかんする一考察

木　村　和　範

〈要旨〉

コーシー＝シュワルツの不等式を援用して，２変量にかんする相関係数 r の数学的性質

$$-1 \leqq r \leqq 1$$

を証明するとともに，上式における等号の成立条件を解明した。

〈Summary〉

The author uses Cauchy-Schwarz's inequality to demonstrate the mathematical characteristic $-1 \leqq r \leqq 1$ of the correlation coefficient r for bivariate data and reveals the necessary and sufficient condition of $r = \pm 1$.

はじめに
1．予備的考察 ── コーシー＝シュワルツの不等式とその数学的性質 ──
(1)　コーシー＝シュワルツの不等式の誘導
(2)　コーシー＝シュワルツの不等式における等号の成立条件
2．相関係数 r の数学的性質
(1)　コーシー＝シュワルツの不等式による $-1 \leqq r \leqq 1$ の証明
(2)　等号 ($r = \pm 1$) の成立条件
むすび

はじめに

２変量データの系列 $\{(x_1,\ y_1),\ (x_2,\ y_2),\ \cdots\cdots,\ (x_n,\ y_n)\}$ にかんする相関係数 r の導出方法には少なくとも３つがある[(1)]。第１は，x および y にかんする標準化した２つの平均偏差[(2)]

$$\frac{x_i - \overline{x}}{\sigma_x} \quad \text{および} \quad \frac{y_i - \overline{y}}{\sigma_y}$$

の積の相加平均として，r を定義する方法である。第２は，y にたいする x の直線回帰の回帰係数 b と x にたいする y の直線回帰の回帰係数 b' の相乗平均（$\sqrt{bb'}$）として，r を定義する方法である。第３は，系列に当てはめた回帰直線の回りの散布度を示す決定係数の平方根として，r を定義する方法である。

(1)　木村和範「相関係数について」『経済論集』（北海学園大学）第 31 巻第 2 号，1984 年 1 月。［以下，木村（1984）］

(2)　標準化した x の相加平均 (${}^{s}\overline{x}$)，分散 (${}^{s}_{x}\sigma^2$)，標準偏差 (${}^{s}_{x}\sigma$) は，標準化する前の x の標準偏差 (σ_x) が $\sigma_x \neq 0$ のとき，

$$
\begin{aligned}
{}^{s}\overline{x} &= 0 \\
{}^{s}_{x}\sigma^2 &= 1 \\
{}^{s}_{x}\sigma &= 1
\end{aligned}
$$

2変量データの組にかんする相関係数rは，数学的には，当然，いずれの方法によっても同一であり，rは以下のように導出される。

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

しかしながら，相関係数を導出するための上述の3つの方法のうち，2変量間の回帰を前提とする後2者の方法は，相関係数の実質的な意味に鑑みて，問題なしとしない。相関係数は，2変量間に独立従属の関係が存在することを必ずしも前提するものではないにもかかわらず，後2者の方法は，一方が独立変量であり，他方が従属変量である場合の回帰を前提しているからである。

以上のような考えから，旧著では標準化したxとyの平均偏差の積和をデータの組数nで除した

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right) \tag{2}$$

から相関係数の定義式［(1)式］を以下のように誘導した[(3)]。

$$r=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right) \tag{3}$$

$$=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (1)［再掲］$$

そして，(1)式を誘導した後に，相関係数

であり，標準化したyについても同様に$\sigma_y\neq0$のとき，相加平均${}^s\bar{y}$はゼロであり，分散${}^s_y\sigma^2$および標準偏差${}^s_y\sigma$はいずれも1である。以下では，xについてのみ，そのことを示す。ただし，xの相加平均，分散，標準偏差をそれぞれ，$\bar{x}$，$\sigma_x{}^2$，σ_xとする。

$$\begin{aligned}
{}^s\bar{x}&=\frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\bar{x}}{\sigma_x}\\
&=\frac{1}{n}\cdot\frac{1}{\sigma_x}\sum_{i=1}^{n}(x_i-\bar{x})\\
&=\frac{1}{n}\cdot\frac{1}{\sigma_x}\left(\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\bar{x}\right)\\
&=\frac{1}{n}\cdot\frac{1}{\sigma_x}\left(\sum_{i=1}^{n}x_i-n\bar{x}\right)\\
&=\frac{1}{n}\cdot\frac{1}{\sigma_x}\left(\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_i\right)\\
&\quad\left(\because\ \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ n\bar{x}=\sum_{i=1}^{n}x_i\right)\\
&=0
\end{aligned}$$

$$\begin{aligned}
{}^s_x\sigma^2&=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}-{}^s\bar{x}\right)^2\\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}-0\right)^2 \quad \text{（上式より）}\\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)^2\\
&=\frac{1}{n}\cdot\frac{1}{\sigma_x{}^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\\
&=\frac{1}{\sigma_x{}^2}\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\\
&=\frac{1}{\sigma_x{}^2}\cdot\sigma_x^2\ \left(\because\ \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\sigma_x{}^2\right)\\
&=1
\end{aligned}$$

以上から，

$$\begin{aligned}
{}^s_x\sigma&=\sqrt{{}^s_x\sigma^2}\\
&=1
\end{aligned}$$

(3)　ここでは，旧著と較べてより丁寧に(3)式から(1)式を誘導する。

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sqrt{\sigma_x{}^2}}\right)\left(\frac{y_i-\bar{y}}{\sqrt{\sigma_y{}^2}}\right)\\
&=\frac{1}{n}\cdot\frac{1}{\sqrt{\sigma_x{}^2}\cdot\sqrt{\sigma_y{}^2}}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\
&=\frac{1}{n}\cdot\frac{1}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}\cdot\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}}}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\
&=\frac{1}{n}\cdot\frac{1}{\frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\
&=\frac{1}{n}\cdot\frac{n}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})
\end{aligned}$$

r が取り得る範囲は

$$-1 \leqq r \leqq 1 \tag{4}$$

であることを証明した[(4)]。

相関係数 r が $+1$ のときは，正の完全相関であり，データの組は傾きが正の2元1次直線上にあり，他方で，-1 のときは負の完全相関であり，データの組は傾きが負の2元1次直線上にあることは，改めて指摘するまでもない，言わば自明の事柄である。

自明すぎるせいであろうか，$r=\pm 1$ のときに，データの組が2元1次直線上にあることの数学的証明について触れられることは，意外と多くない。そのようななかにあって，森田優三・久次智雄『新統計概論　改訂版』（日本評論社　1993年）は，例外的と言うべきであろう[(5)]。そこではコーシーの不等式を用いて，2変量データがすべて単一の2元1次直線上に落ちるときに限って，$r=\pm 1$ となることを証明している。しかし，表現が難解である。

彌永昌吉は，対談の形式をとった著作のなかで，相関係数 r について

$$-1 \leqq r \leqq 1 \qquad (4)［再掲］$$

となることの証明を取り上げたことにたいして，統計学者を登場させて，「証明などに立ち入らなくとも，信用して使われればよい」と発言させた。そして，それを受けて，彌永は，「数学の本を読むときは，本をそんなに‘信用’せずに，数の計算でも，式の計算でも，公式の証明でも，鉛筆と紙をもって，自分で当た」ることが重要であると述べ，さらに別の箇所でも数学の勉強にとっては「‘信用してしまう’こと」がよくないと述べている[(6)]。

我が身を省みると，

$$=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

（木村（1984），146頁以下，および木村和範『数量的経済分析の基礎理論』日本経済評論社　2003年，31頁以下参照。［以下，木村（2003）］）

(4)　いくつかの証明があるが，旧著で述べた方法は以下の通りである（木村（1984），160頁以下，木村（2003），35頁）。

$$\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\pm\frac{y_i-\bar{y}}{\sigma_y}\right)^2 \geqq 0$$

左辺を整理すると

$$\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)^2 \pm 2\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\cdot\frac{y_i-\bar{y}}{\sigma_y}\right)+\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{\sigma_y}\right)^2 \geqq 0$$

上式の左辺第1項と第3項に $\frac{n}{n}$ を乗じて整理すれば，次式を得る。

$$\frac{1}{\sigma_x{}^2}\cdot\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\cdot n \pm 2\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\cdot\frac{y_i-\bar{y}}{\sigma_y}\right)$$

$$+\frac{1}{\sigma_y{}^2}\cdot\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\cdot n \geqq 0 \qquad ①$$

ここで，以下を想起する。

$$\sigma_x{}^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} \qquad ②$$

$$nr=\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right) \qquad ③$$

$$\left(\because\ r=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right) \quad (3)［再掲］\right)$$

$$\sigma_y{}^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n} \qquad ④$$

②式，③式，④式を①式に代入すると，

$$n \pm 2nr + n \geqq 0$$
$$\therefore\ 2n(1\pm r) \geqq 0$$
$$1 \pm r \geqq 0 \qquad (\because\ n>0)$$

以上より，

$$-1 \leqq r \leqq 1$$

q. e. d.

(5)　森田優三・久次智雄『新統計概論　改訂版』日本評論社　1993年，98頁。なお，河田敬義・丸山文行『基礎課程　数理統計』裳華房　1951年，6頁以下では，データの組がすべて2元1次直線上にある場合に，$r=\pm 1$ となることが証明されているが，叙述が必ずしも平易ではない。

$$-1 \leqq r \leqq 1 \qquad (4)[再掲]$$

は証明しているものの，(4)式における等号の成立条件については，どの拙稿のなかでも一切言及はしていない。上記した彌永の指摘には重いものがある。相関係数の数学的性質については自明のことではあるが，本稿をもって，旧著の不備を補うことにした次第である。

以下では，森田・久次の著書（前出）から得た示唆を拠り所として，コーシー＝シュワルツの不等式を援用して，$r=\pm1$ のときに，データの組が２元一次直線上にあること（および，その逆）の数学的証明を試みる。換言すれば，(4)式における等号の成立条件を明らかにする。

１．予備的考察 ── コーシー=シュワルツの不等式とその数学的性質 ──

（１）コーシー＝シュワルツの不等式の誘導

コーシー＝シュワルツの不等式はコーシーの不等式ともシュワルツの不等式とも言われ，またブニヤコフスキーの不等式とも言われるが[7]，本稿では，コーシー＝シュワルツの不等式とする。以下の考察のために，あらかじめ不等式そのものを述べ，その不等式における等号の成立条件を明らかにする[8]。

任意の実数の２つの系列 $\{(a_1, a_2, \cdots\cdots, a_n)$ と $(b_1, b_2, \cdots\cdots, b_n)\}$ にかんして，

$$a_i \in \mathbb{R} \qquad (i=1, 2, \cdots, n)$$
$$b_i \in \mathbb{R} \qquad (i=1, 2, \cdots, n)$$

のとき，

$$(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)^2$$
$$\leqq (a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2) \qquad (5)$$

$$\left(\sum_{i=1}^{n} a_ib_i\right)^2 \leqq \left(\sum_{i=1}^{n} a_i{}^2\right)\left(\sum_{i=1}^{n} b_i{}^2\right) \qquad (5)'$$

が成立する。

これをコーシー＝シュワルツの不等式という。後の叙述のために，(5)′式を次のように変形しておく[9]。

$$\frac{\left(\sum_{i=1}^{n} a_ib_i\right)^2}{\left(\sum_{i=1}^{n} a_i{}^2\right)\left(\sum_{i=1}^{n} b_i{}^2\right)} \leqq 1 \qquad (6)$$

以下では，(5)式または(5)′式で表されるコーシー＝シュワルツの不等式を誘導する。

(6)　彌永昌吉『数学のまなび方』（ちくま学芸文庫）筑摩書房　2008 年，31 頁以下，46 頁。

(7)　彌永前掲書，36 頁以下。

(8)　以下の叙述は，「コーシー・シュワルツの不等式」（http://www.mathlion.jp/article/ar051.html, accessed on Nov. 6, 2017.）を参照した。なお，春日正文『公式集（３訂版）』（科学振興社モノグラフ 24）科学振興社　1983 年，50 頁以下も参照。

(9)　この変形は，相関係数 r が－１から１までの値となることを証明するときに使用する。厳密には，(6)式はコーシー＝シュワルツの不等式［(5)式，(5)′式］と同値ではない。なぜならば，$\sum_{i=1}^{n} a_i{}^2=0$ と $\sum_{i=1}^{n} b_i{}^2=0$（すなわち，$i=1, 2, \cdots, n$ のすべてにおいて $a_i=0$ と $b_i=0$）の両方またはいずれか一方が成立するときにも，コーシー＝シュワルツの不等式は成立するが，他方で，(6)式は，$\sum_{i=1}^{n} a_i{}^2=0$ および/または $\sum_{i=1}^{n} b_i{}^2=0$ のときには，その分母がゼロとなるために，(6)式の成立条件は，$\sum_{i=1}^{n} a_i{}^2 \neq 0$ および/または $\sum_{i=1}^{n} b_i{}^2 \neq 0$ である。換言すれば，$i=1, 2, \cdots, n$ のすべてにおいて，a_i が同時にゼロにはならないこと，および b_i が同時にゼロにはならないこと，この２つの条件の両方が，(6)式の成立条件である。敷衍して言えば，コーシー＝シュワルツの不等式は任意の実数において成立し，$i=1, 2, \cdots, n$ のすべてについて $a_i=0$ および/または $b_i=0$ を排除することがない。これにたいして，コーシー＝シュワルツの不等式を変形した(6)式は，分母のとりうる値に制約があるために，任意の実数において成立するとは言いがたい。このために，(6)式は，厳密にはコーシー＝シュワルツの不等式ではなく，特定の条件のもとでのみ成立する特殊なコーシー＝シュワルツの不等式である。

上と同じ2つの系列 $\{(a_1, a_2, \cdots\cdots, a_n)$ と $(b_1, b_2, \cdots\cdots, b_n)\}$ において，t を任意の実数とするとき，次式が成立する。コーシー＝シュワルツの不等式の誘導は，この式から始まる。

$$(a_1t-b_1)^2+(a_2t-b_2)^2+\cdots\cdots+(a_nt-b_n)^2 \geqq 0 \quad (7)$$

この式は

$$(a_1^2+a_2^2+\cdots\cdots+a_n^2)t^2 -2(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)t +(b_1^2+b_2^2+\cdots\cdots+b_n^2) \geqq 0 \quad (8)$$

と展開され，t にかんする2次方程式が誘導される。

2次方程式

$$y=ux^2+vx+w \quad (9)$$

ただし，$u>0$

において

$$y \geqq 0$$

となるときの解は，①不等号が成立する場合には異なる2つの虚数解（共役な複素数）となり，また②等号が成立する場合には重複解（1つの実数解）となる。このとき，(9)式の判別式 D_y の値は

$$D_y=v^2-4uw \leqq 0 \quad (10)$$

であるが，上記した①の場合には，

$$D_y=v^2-4uw<0 \quad (10)'$$

となり，また②の場合には，

$$D_y=v^2-4uw=0 \quad (10)''$$

となる。

以上に述べた2次方程式の数学的性質を(8)式に適用すれば，(8)式の判別式 D は非正（non-positive）となり，

$$D=\{2(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)\}^2 -4(a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2)$$
$$=4(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)^2 -4(a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2) \leqq 0 \quad (11)$$

となる。したがって，(11)式は

$$\frac{D}{4}=(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)^2 -(a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2) \leqq 0 \quad (12)$$

となる。

(12)式は，

$$(a_1t-b_1)^2+(a_2t-b_2)^2+\cdots\cdots+(a_nt-b_n)^2 \geqq 0 \quad (7)\text{［再掲］}$$

の判別式から誘導されたが，この(12)式は冒頭で掲げたコーシー＝シュワルツの不等式

$$(a_1b_1+a_2b_2+\cdots+a_nb_n)^2 \leqq (a_1^2+a_2^2+\cdots+a_n^2)(b_1^2+b_2^2+\cdots+b_n^2) \quad (5)\text{［再掲］}$$

と同値である[10]。以上によりコーシー＝シュワルツの不等式が得られた。

（2）コーシー＝シュワルツの不等式における等号の成立条件

コーシー＝シュワルツの不等式（(5)式）において等号が成立し，

$$(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)^2 =(a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2) \quad (5)''$$

すなわち，コーシー＝シュワルツの不等式は

$$\forall a \in \mathbb{R}, \quad \forall b \in \mathbb{R}$$

という条件を満たすが，コーシー＝シュワルツの不等式の特殊型は上の条件を満たさず，その成立条件は

$$\exists a \in \mathbb{R}, \quad \exists b \in \mathbb{R}$$

と表すに止まる。

(10)　Herman Amadeus Schwarz（1843-1921）が判別式を使って(5)式を誘導・証明したために，コーシー＝シュワルツの不等式はシュワルツの不等式とも言われる。シュワルツの証明については，彌永前掲書，39頁以下参照。また，コーシーによる証明については，同，43頁以下参照。なお，ここにコーシーとは Augustin-Louis Cauchy（1789-1857）を指すことは言うまでもない。

となるときの条件を考察する[11]。

そこで，繰り返しになるが，コーシー＝シュワルツの不等式は

$$(a_1t-b_1)^2+(a_2t-b_2)^2+\cdots\cdots+(a_nt-b_n)^2\geqq 0 \quad (7)[再掲]$$

を展開した

$$\begin{aligned}&(a_1^2+a_2^2+\cdots\cdots+a_n^2)t^2\\&\quad-2(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)t\\&\qquad+(b_1^2+b_2^2+\cdots\cdots+b_n^2)\geqq 0 \quad (8)[再掲]\end{aligned}$$

の判別式，

$$\begin{aligned}D=&4(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)^2\\&-4(a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2)\leqq 0 \quad (11)[再掲]\end{aligned}$$

したがって，

$$\begin{aligned}\frac{D}{4}=&(a_1b_1+a_2b_2+\cdots\cdots+a_nb_n)^2\\&-(a_1^2+a_2^2+\cdots\cdots+a_n^2)(b_1^2+b_2^2+\cdots\cdots+b_n^2)\leqq 0 \quad (12)[再掲]\end{aligned}$$

から誘導されたことを想起する。コーシー＝シュワルツの不等式において等号が成立し，その不等式が(5)″式で表現可能ということは，(7)式の判別式（(11)式）の値がゼロであること，すなわち，

$$D=0,\ したがって\quad \frac{D}{4}=0$$

であることを意味する。さらにこのことは，判別式の値がゼロのときに２次方程式は重複解（１つの実数解）をもつことを示すという判別式の数学的性質により，元の判別式をあたえる２次方程式

$$(a_1t-b_1)^2+(a_2t-b_2)^2+\cdots\cdots+(a_nt-b_n)^2\geqq 0 \quad (7)[再掲]$$

が

$$(a_1t-b_1)^2+(a_2t-b_2)^2+\cdots\cdots+(a_nt-b_n)^2=0 \quad (7)'$$

であることを意味する。(7)′ 式は

$$a_it-b_i=0 \quad (13)$$

と同値である。こうして，コーシー＝シュワルツの不等式における等号の成立条件は(13)式であることが明らかになった。換言すれば，変量の組 $(a_i,\ b_i)$ が

$$b_i=a_it \quad (13)'$$

という線形関係にある場合に，コーシー＝シュワルツの不等式において等号が成立し，

$$\begin{aligned}&(a_1b_1+a_2b_2+\cdots+a_nb_n)^2\\&=(a_1^2+a_2^2+\cdots+a_n^2)(b_1^2+b_2^2+\cdots+b_n^2) \quad (5)''[再掲]\end{aligned}$$

となる。

なお，(13)′ 式は

$$a_1:a_2:\cdots\cdots:a_n=b_1:b_2:\cdots\cdots:b_n \quad (14)$$

と表現することもできる。これもまた，コーシー＝シュワルツの不等式において等号が成立するための条件ということができる。

以上により,コーシー＝シュワルツの不等式において等号が成立するための条件は，任意の実数 a_i と b_i について次のとおりである[12]。

---

(11)「シュワルツの不等式のエレガントな証明」(https://www.mathtrain.jp/schwarz，accessed on Nov. 6, 2017.）を参照。

(12)　(13)′ 式を変形して得られる

$$\frac{b_i}{a_i}=t \quad ⑤$$

あるいは(14)式を変形して得られる

$$\frac{b_1}{a_1}:\frac{b_2}{a_2}:\cdots\cdots:\frac{b_i}{a_i}:\cdots\cdots:\frac{b_n}{a_n} \quad ⑥$$

がコーシー＝シュワルツの不等式における等号の成立条件といわれることもある。しかし，この言い方は厳密性に欠ける。なぜならば，⑤式および⑥式が成立するのは

$$a_i\neq 0$$

のときであるが，コーシー＝シュワルツの不等式を誘導した元の数式

$$(a_1t-b_1)^2+(a_2t-b_2)^2+\cdots+(a_nt-b_n)^2\geqq 0 \quad (7)[再掲]$$

において a_i は任意の実数であり，(7)式は a_i がゼロ

$$a_i t - b_i = 0 \qquad (13)［再掲］$$

または

$$a_i t = b_i \qquad (13)'［再掲］$$

すなわち

$$a_1 : a_2 : \cdots\cdots : a_n = b_1 : b_2 : \cdots\cdots : b_n \qquad (14)［再掲］$$

q. e. d.

2．相関係数 *r* の数学的性質

（1）コーシー＝シュワルツの不等式による －1≦*r*≦1 の証明

コーシー＝シュワルツの不等式の特殊型（脚注（9）参照）

$$\frac{\left(\sum_{i=1}^{n} a_i b_i\right)^2}{\left(\sum_{i=1}^{n} a_i{}^2\right)\left(\sum_{i=1}^{n} b_i{}^2\right)} \leqq 1 \qquad (6)［再掲］$$

ただし $\sum_{i=1}^{n} a_i{}^2 \neq 0$，$\sum_{i=1}^{n} b_i{}^2 \neq 0$

を再掲する。

ここで，実数の組の系列 $\{(x_1,\ y_1),\ (x_2,\ y_2),\ \cdots\cdots,\ (x_n,\ y_n),\}$ において，x_i の相加平均を $\bar{x}$ とし，他方で y_i の相加平均を $\bar{y}$ とする。そして，それぞれの個別値 $(x_i,\ y_i)$ から平均偏差を求めると，次のようになる。

x_i の系列における平均偏差：$x_i - \bar{x}$
y_i の系列における平均偏差：$y_i - \bar{y}$

平均偏差は任意の実数であるから，(6)式の a_i と b_i を次のようにおくことができる。

$$a_i = x_i - \bar{x} \qquad (15)$$

ただし，$\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0$　$(\because \sum_{i=1}^{n} a_i{}^2 \neq 0)$

$$b_i = y_i - \bar{y} \qquad (16)$$

ただし，$\sum_{i=1}^{n}(y_i - \bar{y})^2 \neq 0$　$(\because \sum_{i=1}^{n} b_i{}^2 \neq 0)$

(15)式と(16)式を(6)式に代入すると，次式を得る。

$$\frac{\left\{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right\}^2}{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2\right\}} \leqq 1 \qquad (17)$$

ゆえに

$$-1 \leqq \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \leqq 1 \qquad (18)$$

相関係数 *r* は

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (1)［再掲］$$

となることを排除せず，

$a_i = 0$

となるときにも(7)式が成立する。なお，(13)′式において，

$a_i = 0$，したがって　$a_i t = 0$　のとき，$b_i = 0$

となるので，コーシー＝シュワルツの不等式における等号の成立条件は，(13)′式において

$b_i = 0$

を排除するものではない（前出「コーシー・シュワルツの不等式」［脚注８］を参照）。

このことは，次のようにしても証明することができる。コーシー＝シュワルツの不等式

$$(a_1 b_1 + a_2 b_2 + \cdots + a_n b_n)^2 \leqq (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2) \qquad (5)［再掲］$$

に

$a_i = 0$

を代入すれば，(5)式の左辺はゼロとなり，したがって右辺もゼロとなる。このとき，(5)式において等号が成立する。また，

$b_i = 0$

のときも，同様に(5)式において等号が成立する。

以上により，コーシー＝シュワルツの不等式は

$a_i = 0$　または　$b_i = 0$

のときにも成立することが分る。

であるから，(1)式を(18)式に代入すると，

$$-1 \leqq r \leqq 1 \qquad (4)［再掲］$$

となる。以上，コーシー＝シュワルツの不等式の特殊型によっても，相関係数は，その下限を -1 とし，その上限を $+1$ とすることが証明された。

q. e. d.

（2）等号（$r=\pm 1$）の成立条件

① $r=\pm 1$ ならば，2変量データの組(x, y)がすべて2元1次直線上に落ちることの証明

$$-1 \leqq r \leqq 1 \qquad (4)［再掲］$$

における等号の成立条件は，上式が誘導された

$$\frac{\left\{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right\}^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}} \leqq 1 \qquad (17)［再掲］$$

において等号が成立する条件と同一である。この条件を考察するためには，(17)式のもとになったコーシー＝シュワルツの不等式の特殊型

$$\frac{\left(\sum_{i=1}^{n}a_ib_i\right)^2}{\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right)} \leqq 1 \qquad (6)［再掲］$$

における等号の成立条件が，コーシー＝シュワルツの不等式における等号の成立条件

$$a_it-b_i=0 \qquad (13)［再掲］$$

と同一であることを想起すればよい。

すでに述べたように，

$$a_i=x_i-\bar{x} \qquad (15)［再掲］$$
$$b_i=y_i-\bar{y} \qquad (16)［再掲］$$

であるから，コーシー＝シュワルツの不等式における等号の成立条件である

$$a_it-b_i=0 \qquad (13)［再掲］$$

に(15)式と(16)式を代入すれば，(17)式における等号の成立条件を示すことができる。すなわち，その条件は

$$(x_i-\bar{x})t-(y_i-\bar{y})=0 \qquad (19)$$

である。これを整理すれば，

$$y_i=(\bar{y}-t\bar{x})+tx_i \qquad (20)$$

となるが，事柄をより明確にするために，(20)式右辺第1項$(\bar{y}-t\bar{x})$を

$$\bar{y}-t\bar{x}=\mathscr{C} \qquad (21)$$

とおく。(21)式を(20)式に代入すると，

$$y_i=\mathscr{C}+tx_i \qquad (22)$$

となる。これはデータの組$(x_i,\ y_i)$がすべて，切片を $\mathscr{C}$ とし，傾きを t とする2元1次直線の上にあることを意味する。

すでに述べたように，(19)式は

$$-1 \leqq r \leqq 1 \qquad (4)［再掲］$$

の範囲に落ちる相関係数 r において，

$$r=\pm 1$$

が成立するための条件である。(22)式は(19)式と同値であるから，$r=\pm 1$ となるときのデータの組$(x_i,\ y_i)$はすべて一様に，(22)式で示される2元1次直線上に落ちる。

以上が，$r=\pm 1$ の成立条件であるが，その直線の数学的性質について付言しておく。(19)式，したがって(20)式および(22)式は

$$x_i=\bar{x}$$
$$y_i=\bar{y}$$

のときにも成立するので，$r=\pm 1$ となる直線

$$y_i=\mathscr{C}+tx_i \qquad (22)［再掲］$$

は，座標$(\bar{x},\ \bar{y})$（すなわち，それぞれの変量の相加平均）を通ることになる。

以上で，2変量データにかんする（-1 か

ら +1 までの値をとる）相関係数 r において，$r=\pm 1$ のときには，2変量データの組 $(x_i,\ y_i)$ は，すべて点 $(\bar{x},\ \bar{y})$ を通る2元1次直線上に落ちることを証明した。

② 2変量データの組 $(x_i,\ y_i)$ がすべて2元一次直線上に落ちるならば，$r=\pm 1$ となることの証明

以下では，①とは逆に，2変量データの組 $(x_i,\ y_i)$ が，すべて2元1次直線上に落ちる場合には，$r=\pm 1$ となることを証明する。

これによって，2変量データにかんする相関係数 r において，$r=\pm 1$ が成立するための必要かつ十分な条件を述べる。そして，

$$(r=\pm 1) \leftrightarrows (\text{2変量データの組}\ (x_i,\ y_i)\ \text{はすべて2元1次直線上に落ちる})$$

が成立することを確認する。

そのために，冒頭で掲げた相関係数 r の定義式

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (1)[再掲]$$

を再掲する。その平方 r^2 は

$$r^2=\frac{\left\{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right\}^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}} \qquad (23)$$

である。この(23)式を用いて，以下では，2変量データの組 $(x_i,\ y_i)$ がすべて，2元1次直線上に落ちる場合には，$r=\pm 1$ となることを証明する。

データの組がすべて，直線

$$y=A+Bx \qquad (24)$$

ただし $B\neq 0$

の上に落ちるとき，この直線は点 $(\bar{x},\ \bar{y})$ を通る[(13)]。したがって，(24)式においては

$$\bar{y}=A+B\bar{x} \qquad (25)$$

が成立する。

ここで，$B\neq 0$ の場合を（ｉ）$\bar{x}\neq 0$ のときと，（ⅱ）$\bar{x}=0$ のときに分けて考察する（$B=0$ の場合は後述する）。

（ｉ）$\bar{x}\neq 0$ のとき

---

(13)　$y=A+Bx$ 　(24)［再掲］

を x 方向に $\bar{x}$，y 方向に $\bar{y}$，それぞれ平行移動させる。このとき，(24)式は

$$y-\bar{y}=A+B(x-\bar{x}) \qquad ⑦$$

となる。

$$X=x-\bar{x}$$
$$Y=y-\bar{y}$$

と置くと，⑦式は

$$Y=A+BX \qquad ⑧$$

となる。

ここで，点 $(\bar{x},\ \bar{y})$ を原点とする新しい座標系を考える。この新しい原点 (0，0) における切片 A は，⑧式に

$X=0$
$Y=0$

を代入すれば，

$$A=0$$

と得ることができる。したがって，新しい座標系では，(1)式は，

$$Y=BX \qquad ⑨$$

となる。原点を通る1次直線の切片はゼロであるから，このこと自体，取り立てて言うようなことではない。ここで注目すべきは，⑨式の傾き B である。この傾きは，(1)式と同一である。また，原点を $(\bar{x},\ \bar{y})$ とする新しい座標系を作るとき，平行移動させた直線［(1)式］がその座標系の原点を通るということにも注目したい。

ここで，この原点を，元の座標系の座標で表せば，$(\bar{x},\ \bar{y})$ であることを想起する。以上により，

$$y=A+Bx \qquad (24)[再掲]$$

を平行移動させる前には，この1次直線は座標 $(\bar{x},\bar{y})$ を通っていることが分かる。

$$\bar{y}=A+B\bar{x} \qquad (25)［再掲］$$

を変形すれば,

$$B=\frac{\bar{y}-A}{\bar{x}} \qquad (25)'$$

を得る。(25)′ 式を

$$y=A+Bx \qquad (24)［再掲］$$

に代入すれば,

$$y=A+\frac{\bar{y}-A}{\bar{x}}x \qquad (24)'$$

となる。(24)′ 式より,

$$y_i=A+\frac{\bar{y}-A}{\bar{x}}x_i \qquad (24)''$$

$$\bar{y}=A+\frac{\bar{y}-A}{\bar{x}}\bar{x} \qquad (24)'''$$

$\because$ 右辺$=A+(\bar{y}-A)=\bar{y}=$左辺

を得る。(24)″ 式と (24)‴ 式を (23) 式に代入すると,

$$r^2=\frac{\left\{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right\}^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}} \qquad (23)［再掲］$$

$$=\frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})\left\{\left(A+\frac{\bar{y}-A}{\bar{x}}x_i\right)-\left(A+\frac{\bar{y}-A}{\bar{x}}\bar{x}\right)\right\}\right]^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left[\sum_{i=1}^{n}\left\{\left(A+\frac{\bar{y}-A}{\bar{x}}x_i\right)-\left(A+\frac{\bar{y}-A}{\bar{x}}\bar{x}\right)\right\}^2\right]}$$

$$=\frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})\left\{\left(\frac{\bar{y}-A}{\bar{x}}\right)(x_i-\bar{x})\right\}\right]^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left[\sum_{i=1}^{n}\left\{\left(\frac{\bar{y}-A}{\bar{x}}\right)(x_i-\bar{x})\right\}^2\right]}$$

$$=\frac{\left(\frac{\bar{y}-A}{\bar{x}}\right)^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})\right\}^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left(\frac{\bar{y}-A}{\bar{x}}\right)^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}}$$

$$=\frac{\left(\frac{\bar{y}-A}{\bar{x}}\right)^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left(\frac{\bar{y}-A}{\bar{x}}\right)^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}}$$

$$=\frac{\left(\frac{\bar{y}-A}{\bar{x}}\right)^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}{\left(\frac{\bar{y}-A}{\bar{x}}\right)^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}$$

$$=1 \qquad (26)$$

よって,

$$r^2=1 \qquad (27)$$

となり,

$$r=\pm 1 \qquad (28)$$

である。

(ⅱ) $\bar{x}=0$ のとき

説明のために,(24)式を再掲する。

$$y=A+Bx \qquad (24)［再掲］$$

ただし,$B\neq 0$

$\bar{x}=0$ であるから,(25)式は

$$\bar{y}=A+B\bar{x} \qquad (25)［再掲］$$

$$=A \qquad (25)'$$

となる。

ここで,

$$\bar{x}=0$$

$$y=A+Bx \qquad (24)［再掲］$$

$$\bar{y}=A \qquad (25)'［再掲］$$

を(23)式に代入すると,次式を得る。

$$r^2=\frac{\left\{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right\}^2}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}} \qquad (23)［再掲］$$

$$=\frac{\left[\sum_{i=1}^{n}(x_i-0)\{(A+Bx_i)-A\}\right]^2}{\left\{\sum_{i=1}^{n}(x_i-0)^2\right\}\left[\sum_{i=1}^{n}\{(A+Bx_i)-A\}^2\right]}$$

$$=\frac{\left(\sum_{i=1}^{n}x_i\cdot Bx_i\right)^2}{\left(\sum_{i=1}^{n}x_i{}^2\right)\left\{\sum_{i=1}^{n}(Bx_i)^2\right\}}$$

$$=\frac{B^2\left(\sum_{i=1}^{n}x_i{}^2\right)^2}{\left(\sum_{i=1}^{n}x_i{}^2\right)B^2\left(\sum_{i=1}^{n}x_i{}^2\right)}$$

$$=\frac{B^2\left(\sum_{i=1}^{n}x_i{}^2\right)^2}{B^2\left(\sum_{i=1}^{n}x_i{}^2\right)^2}$$

$$=1 \tag{29}$$

よって，(29)式より，

$$r=\pm1$$

となり，$\bar{x}=0$ のときも，$\bar{x}\neq0$ のときと同様の結果となる。

以上により，データの組がすべて，直線

$$y=A+Bx \qquad (24)［再掲］$$

ただし $B\neq0$

の上に落ちるとき，相関係数 r は ±1 となる。

q. e. d.

以上を踏まえて，以下ではさらに論を進め，$r=1$ となるときと $r=-1$ になるときの２元１次直線［(24)式］は，それぞれどのような性質をもっているかを検討する。そのために，相関係数の定義式

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (1)［再掲］$$

において，分母の２項

$$\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \text{および} \quad \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

がいずれも正の値をとることに着目する。このことから，相関係数 r の符号はもっぱら(1)式の分子

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \tag{30}$$

の符号に依存していることが分かる[(14)]。

題意より，データの組 $(x_i,\ y_i)$ は，直線

$$y=A+Bx \qquad (24)［再掲］$$

ただし，$B\neq0$

の上に落ちるので，(30)式は次のように変形される。

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \qquad (30)［再掲］$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})\{(A+Bx_i)-(A+B\bar{x})\}$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})\{B(x_i-\bar{x})\}$$
$$=B\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad (30)'$$

(30)′式において

$$\sum_{i=1}^{n}(x_i-\bar{x})^2>0$$

であるから

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \qquad (30)［再掲］$$

の符号（したがって r の符号）を決定するのは，もっぱら(30)′式の B の値であることが

(14)　この積和を２変量データの組の個数 n で除して得られる統計量が共分散 $\mathrm{Cov}(x,y)$ である。相関係数 r は，共分散と２つの変量の標準偏差を用いれば，次のように表記できる。

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (1)［再掲］$$
$$=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\left(\sqrt{\frac{1}{n}}\right)^2\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$
$$=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$
$$=\frac{\mathrm{Cov}\,(x,y)}{\sigma_x\sigma_y}$$

分かる。こうして，

$B>0$ のとき，

$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})>0$,

よって $r>0$

$B<0$ のとき，

$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})<0$,

よって $r<0$

となるが，題意により相関係数 r は $+1$ か -1 のいずれかであるから，

$r=1$ のとき，$B>0$，よってデータの組は傾きが正の1次直線上にあること，

$r=-1$ のとき，$B<0$，よってデータの組は傾きが負の1次直線上にあること

が証明された。

q. e. d.

以上，2変量データの組がすべて，1次直線

$$y=A+Bx \qquad (24)［再掲］$$

の上に落ちる場合の相関係数 r を取り上げてきた。そこでは，

$$B\neq 0$$

を前提とした。さらに念のために，以下では

$$B=0$$

となるとき，すなわちデータの組のすべてにおいて，x の値の如何にかかわらず，直線

$$y=A$$

の上に落ちる場合を検討する。この場合を換言すれば，データの組のすべてが，x 軸と平行な直線上に落ちるときと言うことができる。なお，ついでながら，データの組が，すべて y 軸と平行な直線上に落ちる場合，すなわち，データの組において

$$x=C$$

が成立する場合も併せて検討することにする。そのために，改めて相関係数 r の定義式

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}} \qquad (1)［再掲］$$

を掲げる。

(ⅰ)$y=A$ のとき

すべての y の値が A であるので，y の相加平均 $\overline{y}$ も同様に A である。このとき，相関係数をあたえる(1)式の分母は

$$\begin{aligned}&\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}\\&=\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(A-A)^2}\\&=\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\times\sqrt{0}\\&=0\end{aligned}$$

となる。しかし，これを(1)式に代入すると，ゼロ除算（division by zero）となり，2変量データの組が x 軸と平行な直線上に落ちる場合には，相関係数を算出することができない。

以下で見るように，同様のことは，2変量データの組が y 軸と平行な直線上に落ちる場合についても妥当する。

(ⅱ)$x=C$ のとき

すべての x の値が C であるので，x の相加平均 $\overline{x}$ も同様に C である。このとき，相関係数の定義式［(1)式］の分母は

$$\begin{aligned}&\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}\\&=\sqrt{\sum_{i=1}^{n}(C-C)^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}\end{aligned}$$

$$=\sqrt{0}\times\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}$$

$$=0$$

となる。これを(1)式に代入すると，（ⅰ）($y=A$) と同様にゼロ除算となり，2 変量データの組がすべて y 軸と平行な直線上に落ちる場合には，相関係数を算出することができない。

以上から，2 変量データの組にかんして相関係数が ±1 となるのは，データのすべての組が直線

$$y=A+Bx \qquad (24)［再掲］$$

ただし，$B\neq 0$

の上に落ちるときに限られ，すべてのデータの組が x 軸と平行な直線上に落ちる場合，およびすべてのデータの組が y 軸と平行な直線上に落ちる場合には，相関係数を計算できないことが分かる[15]。

(15)　コーシー＝シュワルツの不等式の特殊型である

$$\frac{\left(\sum_{i=1}^{n}a_ib_i\right)^2}{\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right)}\leqq 1 \qquad (6)［再掲］$$

を誘導したとき，ゼロ除算を回避する目的で，上式の成立条件を $\sum_{i=1}^{n}a_i^2\neq 0$，$\sum_{i=1}^{n}b_i^2\neq 0$ とした。そして，「2. 相関係数の数学的性質」（1）では，（6）式の a_i, b_i にたいして

$$a_i=x_i-\overline{x} \qquad (15)［再掲］$$

$$b_i=y_i-\overline{y} \qquad (16)［再掲］$$

を代入したときに，（15）式と（16）式の成立条件が，それぞれ

$$\sum_{i=1}^{n}(x_i-\overline{x})^2\neq 0 \qquad ⑩$$

$$\sum_{i=1}^{n}(y_i-\overline{y})^2\neq 0 \qquad ⑪$$

であることを記した。したがって，本文の（ⅰ）($y=A$ のとき）の結論は $\sum_{i=1}^{n}(y_i-\overline{y})^2\neq 0$ と整合し，また（ⅱ）($x=C$ のとき）の結論は $\sum_{i=1}^{n}(x_i-\overline{x})^2\neq 0$ と整合する。

むすび

2 変量データの組における相関係数 r は，それぞれの変量にかんする標準偏差の積で共分散を除して得られる統計量である。①r が -1 から $+1$ までの値をとり，r が $+1$ に近い値をとるほど，傾きが正の 2 元 1 次直線の近傍にデータの組がプロットされ，逆に，r が -1 に近いほど，データの組は，負の傾きの直線の近傍に散布すること，そして，②データの組がすべて，直線上に落ちるとき，相関係数 r は $+1$ か -1 となること，これらは，自明の理と考えられており，そのこともあってか，数理統計学の教科書では簡潔に述べられることが多い。叙述が比較的丁寧な教科書では

$$-1\leqq r\leqq +1$$

を証明しているものもあるが，

$$r=\pm 1$$

については，散布図を援用して，データの組がすべて単一の直線上にプロットされる正（または負）の完全相関のときには，相関係数の値が $+1$（または -1）になると指摘するに止まることが少なくない。

表計算ソフトを起動し，データセットと所定の関数さえ入力して，クリックすれば，相加平均，相乗平均，分散，標準偏差はもとより，中央値（メディアン），最頻値（モード），回帰係数，相関係数など，様々な統計量が手軽に求められるようになった。所要の統計量を算出するために，自らソフトウェアを組むことを強いられる事態は，以前の比ではなくなった。実際に，経済学部におけるプログラミング関連科目の開講数は縮減している。本学も例外ではない。

これは，コンピュータ科学の進歩発展の結果であり，そのこと自体を低く評価するものではない。しかしながら，クリックすれば統

計量が自動的に算出されると言うことは，計算過程をブラックボックス化しかねないのではないかとも危惧される。

そのようななかにあっては，相関係数の数学的性質を検討することも意味なしとしないと考え，旧聞に属すことではあるが，調べた事柄を本稿にまとめた。また，データの組が，すべて x 軸（あるいは y 軸）と平行な直線上に落ちる場合について相関係数という概念が成立するかどうかを検討した著述は，寡聞にして知らず，先行研究の検討は他日を期したいが，本稿では如上の場合における相関係数についても検討した。その結果，かかる場合の相関係数の計算はゼロ除算という禁則演算となることを述べた。

(2017 年 11 月 22 日提出)