偏相関係数に関するノート

2024 年 6 月 12 日

1　偏相関係数の導出

変数 x, y, z があったときに，x で統制した y と z の偏相関係数 $r_{yz|x}$ は，x で y を回帰した時の残差と，x で z を回帰した時の残差の相関として表現されます。数式では,

$$r_{yz|x} = \frac{r_{yz} - r_{xy}r_{xz}}{\sqrt{(1 - r_{xy}^2)}\sqrt{(1 - r_{xz}^2)}}$$

とされますが，どうしてこうなるのかを見ていきましょう。

まずは回帰式の確認から。

$$\hat{y} = ax + b, \hat{z} = cx + d$$

としします。ここで $\hat{y}, \hat{z}$ は x から説明される予測値です。回帰係数の導出はここではスキップしますが，以下のような関係があります。

$$a = \frac{s_{xy}}{s_x^2}, \quad b = \bar{y} - a\bar{x}$$
$$c = \frac{s_{xz}}{s_x^2}, \quad d = \bar{z} - c\bar{x}$$

ここで s_x^2 は x の分散，s_{xy} は x と y の共分散を表しています。

さて，x の影響を y から取り除くということは，$y - \hat{y}$ とすることでもありますから，これを y' とおきますと次のように表現できます。

$$\begin{aligned} y' &= y - \hat{y} \\ &= y - (ax + b) \\ &= y - (ax + \bar{y} - a\bar{x}) \\ &= y - (\bar{y} + a(x - \bar{x})) \\ &= (y - \bar{y}) - \frac{s_{xy}}{s_x^2}(x - \bar{x}) \end{aligned} \tag{1}$$

同様に

$$\begin{aligned} z' &= z - \hat{z} \\ &= z - (cx + \bar{z} - c\bar{x}) \\ &= z - (\bar{z} + c(x - \bar{x})) \\ &= (z - \bar{z}) - \frac{s_{xz}}{s_x^2}(x - \bar{x}) \end{aligned} \tag{2}$$

となります。

ここで，x の影響を取り除いた y' の平均は

$$\begin{aligned}
\bar{y'} &= \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \\
&= \frac{1}{N}\sum y_i - \frac{1}{N}\sum \hat{y}_i \\
&= \bar{y} - \frac{1}{N}\sum(ax_i + b) \\
&= \bar{y} - a\frac{1}{N}\sum x_i - b \\
&= \bar{y} - a\bar{x} - (\bar{y} - a\bar{x}) \\
&= 0
\end{aligned} \tag{3}$$

となります。同様に $\bar{z}' = 0$ です。

さて，相関係数は $r_{yz} = \frac{s_{yz}}{s_y s_z}$ であり，$s_{xy} = r_{xy}s_x s_y$ と表現できます。ここで s_x は x の標準偏差です。

したがって，

$$\begin{aligned}
a &= \frac{s_{xy}}{s_x^2} = r_{xy}\frac{s_y}{s_x} \\
c &= \frac{s_{xz}}{s_x^2} = r_{xz}\frac{s_z}{s_x}
\end{aligned}$$

です。

求めたい偏相関は $r_{y'z'}$ ですから，これは

$$r_{y'z'} = \frac{s_{y'z'}}{s_{y'}s_{z'}}$$

なので，これの各要素，$s_{y'}$ や $s_{y'z'}$ について考えます。

まず $s_{y'}$ を二乗した $s_{y'}^2$，すなわち y' の分散についてみていきましょう。

$$\begin{aligned}
s_{y'}^2 &= \frac{1}{N}\sum(y'_i - \bar{y}')^2 \\
&= \frac{1}{N}\sum y_i'^2 \\
&= \frac{1}{N}\sum((y_i - \bar{y}) - a(x_i - \bar{x}))^2 \\
&= \frac{1}{N}\sum(y_i - \bar{y})^2 - a\frac{2}{N}\sum(y_i - \bar{y})(x_i - \bar{x}) + a^2\frac{1}{N}\sum(x_i - \bar{x})^2 \\
&= s_y^2 - 2as_{xy} + a^2 s_x^2 \\
&= s_y^2 - 2\left(r_{xy}\frac{s_y}{s_x}\right)s_{xy} + \left(r_{xy}\frac{s_y}{s_x}\right)^2 s_x^2 \\
&= s_y^2 - 2\left(r_{xy}\frac{s_y}{s_x}\right)r_{xy}s_x s_y + \left(r_{xy}\frac{s_y}{s_x}\right)^2 s_x^2 \\
&= s_y^2 - 2r_{xy}^2 s_y^2 + r_{xy}^2 s_y^2 \\
&= s_y^2 - r_{xy}^2 s_y^2 \\
&= s_y^2\left(1 - r_{xy}^2\right)
\end{aligned} \tag{4}$$

同様に，$s_{z'}^2 = s_z^2\left(1 - r_{xz}^2\right)$ です。

次に共分散 $s_{y'z'}$ を見てみましょう。

$$
\begin{aligned}
s_{y'z'} &= \frac{1}{N}\sum(y'_i - \bar{y}')(z_i - \bar{z}') \\
&= \frac{1}{N}\sum y'_i z'_i \\
&= \frac{1}{N}\sum((y_i - \bar{y}) - a(x_i - \bar{x}))((z_i - \bar{z}) - c(x_i - \bar{x})) \\
&= \frac{1}{N}\sum\left\{(y_i - \bar{y})(z_i - \bar{z}) - a(x_i - \bar{x})(z_i - \bar{z}) - c(x_i - \bar{x})(y_i - \bar{y}) + ac(x_i - \bar{x})^2\right\} \\
&= s_{yz} - as_{xz} - cs_{xy} + acs_x^2 \\
&= s_{yz} - r_{xy}\frac{s_y}{s_x}\cdot s_{xz} - r_{xz}\frac{s_z}{s_x}\cdot s_{xy} + r_{xy}\frac{s_y}{s_x}\cdot r_{xz}\frac{s_z}{s_x}\cdot s_x^2 \\
&= r_{yz}s_y s_z - r_{xy}\frac{s_y}{s_x}r_{xz}s_x s_z - r_{xz}\frac{s_z}{s_x}r_{xy}s_x s_y + r_{xy}r_{xz}\frac{s_y s_z}{s_x^2}s_x^2 \\
&= r_{yz}s_y s_z - r_{xy}r_{xz}s_y s_z - r_{xy}r_{xz}s_y s_z + r_{xy}r_{xz}s_y s_z \\
&= r_{yz}s_y s_z - r_{xy}r_{xz}s_y s_z \\
&= s_y s_z(r_{yz} - r_{xy}r_{xz}) \qquad (5)
\end{aligned}
$$

従って，

$$
\begin{aligned}
r_{y'z'} &= \frac{s_{y'z'}}{s_{y'}s_{z'}} \\
&= \frac{s_y s_z(r_{yz} - r_{xy}r_{xz})}{\sqrt{s_y^2(1 - r_{xy}^2)}\sqrt{s_z^2(1 - r_{xz}^2)}} \\
&= \frac{s_y s_z(r_{yz} - r_{xy}r_{xz})}{s_y\sqrt{(1 - r_{xy}^2)}s_z\sqrt{(1 - r_{xz}^2)}} \\
&= \frac{r_{yz} - r_{xy}r_{xz}}{\sqrt{(1 - r_{xy}^2)}\sqrt{(1 - r_{xz}^2)}} \\
&= r_{yz|x} \qquad (6)
\end{aligned}
$$

となることが確認できました。

2　行列を使って偏相関行列を求める

分散共分散行列 $\mathbf{S}$ の逆行列 $\mathbf{S}^{-1}$ を使って偏相関係数を求めると，$\mathbf{S}^{-1}$ の要素 w_{ij} を使って

$$
r_{y'z'} = \frac{-w_{yz}}{\sqrt{w_{yy}\cdot w_{zz}}}
$$

となる，というのはよくある説明です。これも詳しくみてみましょう。

分散共分散行列，$\mathbf{S}$ を 3×3 の行列として、

$$
\mathbf{S} = \begin{pmatrix} s_x^2 & s_{xy} & s_{xz} \\ s_{xy} & s_y^2 & s_{yz} \\ s_{xz} & s_{yz} & s_z^2 \end{pmatrix}
$$

としましょう。これの逆行列を求めるときに，3×3 の行列ぐらいでしたら，頑張ってサラスの公式などを使って計算できるのですが，ここではさらなる一般化を見据えて余因子行列を使った逆行列の求め方を導入します。

$|A|$ は A の行列式であり，$|A| \neq 0$ とすると，逆行列は次のように表されます。

$$A^{-1} = \frac{1}{|A|}\tilde{A}$$

ここで $\tilde{A}$ は余因子行列と呼ばれ、正方行列 A の i 行 j 列目を除いた行列 A_{ij} を使って計算される要素,

$$\tilde{a} = (-1)^{i+j}|A_{ij}|$$

を転置したものです。行・列の番号から符号が決まり ($(-1)^{i+j}$ の箇所)，A_{ij} の行列式が余因子行列の要素となるうえ，それを転置したものを考えるのですから，大変面倒な式です。しかし大きなサイズの行列であっても，余因子に分解していけば最終的には 2×2 の行列式まで落としていくことができます。そうなると，$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ としたときの行列式 $|\mathbf{A}| = ad - bc$ が計算できますので，頑張って逆行列が計算できるということです。

どの要素を考えてるのか分かりやすくするために，余因子行列 $\tilde{A}$ の i, j 要素を $adj(A)_{ij}$ と書くことにしましょう。

$\mathbf{S}^{-1}$ の要素 w_{ij} は,

$$w_{11} = \frac{\mathrm{adj}(\mathbf{S})_{11}}{|\mathbf{S}|}, \quad w_{22} = \frac{\mathrm{adj}(\mathbf{S})_{22}}{|\mathbf{S}|}, \quad w_{12} = \frac{\mathrm{adj}(\mathbf{S})_{12}}{|\mathbf{S}|}$$

と書くことができます。ここから,

$$r_{y'z'} = \frac{\frac{-\mathrm{adj}(\mathbf{S})_{12}}{|\mathbf{S}|}}{\sqrt{\frac{\mathrm{adj}(\mathbf{S})_{11}}{|\mathbf{S}|} \cdot \frac{\mathrm{adj}(\mathbf{S})_{22}}{|\mathbf{S}|}}} \tag{7}$$

$$= \frac{\frac{-\mathrm{adj}(\mathbf{S})_{12}}{|\mathbf{S}|}}{\sqrt{\frac{\mathrm{adj}(\mathbf{S})_{11} \cdot \mathrm{adj}(\mathbf{S})_{22}}{|\mathbf{S}|}}} \tag{8}$$

$$= \frac{-\mathrm{adj}(\mathbf{S})_{12}}{\sqrt{\mathrm{adj}(\mathbf{S})_{11} \cdot \mathrm{adj}(\mathbf{S})_{22}}} \tag{9}$$

ここで，$s_{jk} = s_{kj}$ であることに注意して整理すると,

$$\begin{aligned} \mathrm{adj}(\mathbf{S})_{12} &= (-1)^{1+2} \begin{vmatrix} s_{xy} & s_{yz} \\ s_{xz} & s_z^2 \end{vmatrix} \\ &= -1(s_{xy}s_z^2 - s_{yz}s_{xz}) \\ &= (s_{yz}s_{xz} - s_{xy}s_z^2) \end{aligned} \tag{10}$$

同様に,

$$\begin{aligned} \mathrm{adj}(\mathbf{S})_{11} &= (-1)^{1+1} \begin{vmatrix} s_y^2 & s_{yz} \\ s_{yz} & s_z^2 \end{vmatrix} \\ &= (s_y^2 s_z^2 - s_{yz}^2) \end{aligned} \tag{11}$$

$$
\begin{aligned}
\mathrm{adj}(\mathbf{S})_{22} &= (-1)^{2+2}\begin{vmatrix} s_x^2 & s_{xz} \\ s_{xz} & s_z^2 \end{vmatrix} \\
&= (s_y^2 s_z^2 - s_{xz}^2)
\end{aligned} \tag{12}
$$

これを代入して,

$$
= \frac{-(s_{yz}s_{xz} - s_{xy}s_z^2)}{\sqrt{(s_y^2 s_z^2 - s_{yz}^2)(s_x^2 s_z^2 - s_{xz}^2)}} \tag{13}
$$

ここで $s_{xy} = r_{xy}s_x s_y$ より、

$$
\begin{aligned}
&= \frac{(s_{xy}s_z^2 - s_{yz}s_{xz})}{\sqrt{(s_y^2 s_z^2 - r_{yz}^2 s_y^2 s_z^2)(s_x^2 s_z^2 - r_{xz}^2 s_x^2 s_z^2)}} \\
&= \frac{(r_{xy}s_x s_y s_z^2 - r_{yz}s_y s_z r_{xz}s_x s_z)}{\sqrt{(s_y^2 s_z^2 - r_{yz}^2 s_y^2 s_z^2)(s_x^2 s_z^2 - r_{xz}^2 s_x^2 s_z^2)}} \\
&= \frac{s_x s_y s_z^2(r_{xy} - r_{yz}r_{xz})}{\sqrt{s_x^2 s_y^2 s_z^4 - r_{xz}^2 s_x^2 s_y^2 s_z^4 - r_{yz}^2 s_x^2 s_y^2 s_z^4 + r_{yz}r_{xz}s_x^2 s_y^2 s_z^4}} \\
&= \frac{s_x s_y s_z^2(r_{xy} - r_{yz}r_{xz})}{\sqrt{s_x^2 s_y^2 s_z^4(1 - r_{xz}^2 - r_{yz}^2 + r_{yz}r_{xz})}} \\
&= \frac{s_x s_y s_z^2(r_{xy} - r_{yz}r_{xz})}{s_x s_y s_z^2\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \\
&= \frac{r_{xy} - r_{yz}r_{xz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \\
&= r_{yz|x}
\end{aligned} \tag{14}
$$

となります。たしかに逆行列の要素で偏相関が算出できました。

今回は標本統計量 s_x^2, s_{xy}, r_{xy} をつかった表記をしましたが，一派には母数でのモデルなので，$\sigma_x^2, \sigma_{xy}, \rho_{xy}$ などで表記されています。私があんまり母数での – ギリシア文字での操作のイメージが得意じゃなく，あくまでも標本統計量，代数というより「文字と式」の感覚で表現したかったのでアルファベット表記になっていますが，母数の推定量として不偏分散で計算することなどに注意していただければ，基本的には同じことかと思います。