第 1 章の補足

浅川 伸一

2006 年 11 月 21 日

1　補足

1.1　もう一つの共分散の定義

共分散の定義はもう一つあるが数学的には同じもの

$$\begin{aligned} s_{xy} &= \frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y}) && (1)\\ &= \frac{1}{n}\sum(x_iy_i-x_i\bar{y}-\bar{x}y_i+\bar{x}\bar{y}) && (2)\\ &= \frac{1}{n}\sum(x_iy_i)-\bar{x}\bar{y}-\bar{x}\bar{y}+\bar{x}\bar{y} && (3)\\ &= \frac{1}{n}\sum x_iy_i-\bar{x}\bar{y} && (4) \end{aligned}$$

すなわち共分散とは、積の平均から平均の積をひいたもの。分散が自乗の平均から平均の自乗を引いたものであることと対応する。

1.2　合成得点の分散

たとえば英語の得点を x、数学の得点を y とし、これらの合計点で入試が行われるとする。このとき合計点 $x+y$ の分散は元の分散 s_x^2, s_y^2 と比べると、

$$\begin{aligned} s_{x+y}^2 &= \frac{1}{n}\{x_i+y_i-(\bar{x}+\bar{y})\}^2 && (5)\\ &= \frac{1}{n}\{(x_i-\bar{x})+(y_i-\bar{y})\}^2 && (6)\\ &= \frac{1}{n}\left\{(x_i-\bar{x})^2+(y_i-\bar{y})^2+2(x_i-\bar{x})(y_i-\bar{y})\right\} && (7)\\ &= s_x^2+s_y^2+2s_{xy} && (8) \end{aligned}$$

となる。相関係数の定義は $r_{xy}=\dfrac{s_{xy}}{s_xs_y}$ であったから合計点の分散は

$$s_{x+y}^2=s_x^2+s_y^2+2s_xs_yr_{xy} \tag{9}$$

と書き表せる。ここで分散は 0 より大きい $s_x^2 \leq 0$, $s_y^2 \leq 0$ し、それを開平した正の数である標準偏差も正であるから、合計点の分散は x と y との相関が負 $r_{xy} < 0$ であればむしろ小さくなる。このことは多変数の場合にもあてはまる。

変数 x と y との和の分散をベクトル表現すると以下のようになる。

$$\begin{aligned}
s_{xy}^2 &= |\boldsymbol{x}+\boldsymbol{y}|^2 &(10)\\
&= (\boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}) &(11)\\
&= \frac{1}{n}(\boldsymbol{x},\boldsymbol{x}) + \frac{1}{n}(\boldsymbol{x},\boldsymbol{y}) + \frac{1}{n}(\boldsymbol{y},\boldsymbol{x}) + \frac{1}{n}(\boldsymbol{y},\boldsymbol{y}) &(12)\\
&= s_x^2 + 2s_{xy} + s_y^2 &(13)\\
&= \frac{1}{n}|x|^2 + \frac{1}{n}|y|^2 + 2|x||y|\cos\theta &(14)
\end{aligned}$$

前期の知識によれば、「独立した」二つの確率変数の和の分散はそれぞれの分散の和であった。上の式は、独立でない二つの確率変数の和の場合である。逆に言えば二つの確率変数が独立であるとは共分散がゼロ $s_{xy} = 0$ の場合である。幾何学的には二本のベクトルが直交することを独立であると言う。

1.3　変数変換

2 つの変数 x と y とがあったとき、これらの変数を変換することを考える。a, b, c, および d を定数とすると (ただし c と d とは 0 でないとする)、$v = ax + b$, $w = cy + d$ と変換したときの相関係数を考える。変換後の共分散 s_{vw} は

$$\begin{aligned}
s_{vw} &= \frac{1}{n}\sum(v_i - \bar{v})(w_i - \bar{w}) &(15)\\
&= \frac{1}{n}\sum\{ax_i + b - (a\bar{x} + b)\}\{cy_i + d - (c\bar{y} + d)\} &(16)\\
&= \frac{1}{n}\sum a(x_i - \bar{x})\,c(y_i - \bar{y}) &(17)\\
&= acs_{xy} &(18)
\end{aligned}$$

v と w のそれぞれの分散については、変数を定数倍すると分散は 2 乗に比例するので $s_v^2 = a^2 s_x^2$, $s_w^2 = c^2 s_y^2$ である。このことから

$$r_{vw} = \frac{abs_{xy}}{as_x bs_y} = r_{xy} \qquad (19)$$

となる。すなわち相関係数は単位によらず、変らない。

2　数値例

10 名の被験者について記憶力 x と 判断力 y のテストを行い次の結果を得た。このデータについて x と y との相関係数を求めよ (林, 1973 より)。

データ:

$$\boldsymbol{x} = (11, 10, 14, 18, 10, 5, 12, 7, 15, 16)' \tag{20}$$
$$\boldsymbol{y} = (6, 4, 6, 9, 3, 2, 8, 3, 9, 7)' \tag{21}$$

$$\sum_{i=1}^{10} x_i = 11 + 10 + \cdots + 16 = 118 \tag{22}$$
$$\sum_{i=1}^{10} y_i = 6 + 4 + \cdots + 7 = 57 \tag{23}$$

ゆえに $\bar{x} = 11.8$, $\bar{y} = 5.7$

$$\begin{aligned}
s_x^2 &= \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2 && (24)\\
&= \frac{1}{10}\left\{(11-11.8)^2 + (10-11.8)^2 + \cdots + (16-11.8)^2\right\} && (25)\\
&= 16.4 && (26)\\
s_y^2 &= \frac{1}{10}\sum_{i=1}^{10}(y_i - \bar{y})^2 && (27)\\
&= \frac{1}{10}\left\{(6-5.7)^2 + (4-5.7)^2 + \cdots + (7-5.7)^2\right\} && (28)\\
&= 6.666 && (29)
\end{aligned}$$

$$\begin{aligned}
s_{xy} &= \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) && (30)\\
&= \frac{1}{10}\{(11-11.8)(6-5.7) + (10-11.8)(4-5.7) + \cdots && (31)\\
&\qquad + (16-11.8)(7-5.7)\} && (32)\\
&= 8.34 && (33)
\end{aligned}$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{8.34}{\sqrt{16.4}\sqrt{6.666}} = 0.885 \tag{34}$$

3 相関係数の検定

母相関係数 ρ が 0 であるという帰無仮説 $H_0 : \rho = 0$ について

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \tag{35}$$

が自由度 $n-2$ の t 分布に従うことを利用する。ただし、この検定を証明するのは大変である。初等統計学のテキストには書かれていない。

4 参考文献

林周二 (1973) 統計学講義第 2 版, 丸善

南風原朝和 (2002) 心理統計学の基礎, 有斐閣アルマ