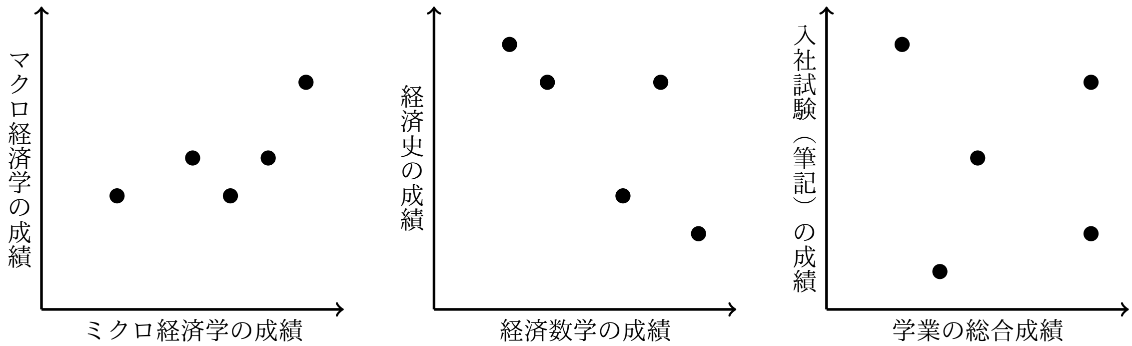
応用計量経済分析 TA セッション
第 2 回：共分散・相関係数

TA：北村友宏 *

2015 年 10 月 13 日

1　散布図

- 2 変数の関係を平面上に打点した図を**散布図**（scattergram）とよぶ。



正の相関　　負の相関　　ほぼ無相関

2　相関の尺度

2.1　共分散

- n 組の 2 変数データ $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ を観測したとき、**共分散**（covariance）は、

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

  - ⋆ 符号の向きが 2 変数の相関の方向を表す。
    $s_{xy} > 0 \Leftrightarrow$ 正の相関、$s_{xy} < 0 \Leftrightarrow$ 負の相関、$s_{xy} = 0 \Leftrightarrow$ 無相関。
  - ⋆ 相関の強さは表さない。
  - ⋆ $s_{xy} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ と定義する場合もある。
  - ⋆ x と y の単位に依存する。

* Email: kitamu.tom@gmail.com　URL: http://tomkitamura.html.xdomain.jp

例題 1.　共分散が $\dfrac{1}{n-1}\left(\displaystyle\sum_{i=1}^{n} x_i y_i - n\overline{x}\cdot\overline{y}\right)$ とも書けることを証明しなさい。

（証明）

$$
\begin{aligned}
s_{xy} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i y_i - x_i\overline{y} - \overline{x}y_i + \overline{x}\cdot\overline{y}) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\overline{y} - \sum_{i=1}^{n}\overline{x}y_i + \sum_{i=1}^{n}\overline{x}\cdot\overline{y}\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \overline{y}\sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} y_i + n\overline{x}\cdot\overline{y}\right).
\end{aligned}
$$

ここで、$\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$、$\overline{y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i$ なので、$n\overline{x} = \displaystyle\sum_{i=1}^{n} x_i$、$n\overline{y} = \displaystyle\sum_{i=1}^{n} y_i$ である。よって、

$$
\begin{aligned}
\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \overline{y}\sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} y_i + n\overline{x}\cdot\overline{y}\right) &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \overline{y}\cdot n\overline{x} - \overline{x}\cdot n\overline{y} + n\overline{x}\cdot\overline{y}\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\overline{x}\cdot\overline{y} - n\overline{x}\cdot\overline{y} + n\overline{x}\cdot\overline{y}\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\overline{x}\cdot\overline{y}\right).
\end{aligned}
$$

したがって、共分散は $\dfrac{1}{n-1}\left(\displaystyle\sum_{i=1}^{n} x_i y_i - n\overline{x}\cdot\overline{y}\right)$ とも書ける。（証明終）

2.2　相関係数

- n 組の 2 変数データ $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ を観測したとき、**相関係数**（correlation coefficient）は、

$$
r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}.
$$

  ⋆ $-1 \leq r_{xy} \leq 1$.

  ⋆ 符号の向きが 2 変数の相関の方向を表す。

  $r_{xy} > 0 \Leftrightarrow$ 正の相関、$r_{xy} < 0 \Leftrightarrow$ 負の相関、$r_{xy} = 0 \Leftrightarrow$ 無相関。

  ⋆ 1 や -1 に近いほど相関が強く、0 に近いほど相関が弱い。

  ⋆ 共分散とは異なり、x と y の単位に依存しない。