# 第9回　回帰分析（直線回帰）

高度な統計的分析の基本は、直線回帰分析である。ここでは、最も単純な単回帰分析について解説する。

単回帰分析は、ふたつの変量 $x$ と $y$ について、$x$ から $y$ を予測するのに、最も誤差の少ない直線を求めようとする。

$y' = a + bx$　　という式を考え、

誤差 $e_i = y_i - y'$ の2乗和を最小にするような$a$, $b$を求める。

$\sum e_i^2 = \sum (y_i - y_i')^2 = \sum \{y_i - (a + bx_i)\}^2$ が最小になればよい。

これは正規方程式

$$
\begin{cases}
na + b\sum x_i = \sum y_i & \cdots(A) \\
a\sum x_i + b\sum x_i^2 = \sum x_i y_i & \cdots(B)
\end{cases}
$$

の解である。

---

【証明】

$\sum e_i^2 = \sum (y_i - bx_i - a)^2$　　が最小になるように$a,b$の値を決定するには、

$$\frac{\delta}{\delta a}\sum e_i^2 = 0 \quad \text{かつ} \quad \frac{\delta}{\delta b}\sum e_i^2 = 0$$

となればよい。

$$
\begin{aligned}
\frac{\delta}{\delta a}\sum e_i^2 &= \sum \frac{\delta}{\delta a}(y_i - bx_i - a)^2 \\
&= \sum 2(y_i - bx_i - a)(-1) \\
&= \sum 2(a + bx_i - y_i) \\
&= 2\sum a + 2\sum bx_i - 2\sum y_i \\
&= 2\left(na + b\sum x_i - \sum y_i\right) = 0 \qquad \therefore na + b\sum x_i = \sum y_i \qquad \cdots(A)
\end{aligned}
$$

$$\frac{\delta}{\delta b}\sum e_i^2=\sum\frac{\delta}{\delta b}(y_i-bx_i-a)^2$$

$$=\sum 2(y_i-bx_i-a)(-x_i)$$

$$=2\sum(ax_i+bx_i^2-x_iy_i)$$

$$=2\left(a\sum x_i+b\sum x_i^2-\sum x_iy_i\right)=0 \qquad \therefore a\sum x_i+b\sum x_i^2=\sum x_iy_i \qquad \cdots\text{(B)}$$

---

この正規方程式(A)(B)の解は、

$b=\dfrac{S_{xy}}{S_x^2}$　　$a=\bar{y}-b\bar{x}$　　である。

---

【証明】この正規方程式を $b$ について解く。

(A)× $\sum x_i$

$$na\sum x_i+b\sum x_i\sum x_i=\sum y_i\sum x_i \qquad \cdots(1)$$

(B)× $n$

$$na\sum x_i+nb\sum x_i^2=n\sum x_iy_i \qquad \cdots(2)$$

(1)－(2)

$$b\sum x_i\sum x_i-nb\sum x_i^2=\sum y_i\sum x_i-n\sum x_iy_i$$

$$b\left(\sum x_i\sum x_i-n\sum x_i^2\right)=\sum y_i\sum x_i-n\sum x_iy_i$$

$$b=\frac{\sum y_i\sum x_i-n\sum x_iy_i}{\left(\sum x_i\right)^2-n\sum x_i^2}$$

$$=\frac{\frac{1}{n^2}\left(\sum x_i\sum y_i-n\sum x_iy_i\right)}{\frac{1}{n^2}\left(\sum x_i\right)^2-\frac{1}{n}\sum x_i^2}=\frac{S_{xy}}{S_x^2}$$

$\dfrac{1}{n^2}\left(\sum x_i\right)^2=\bar{x}^2$ であることに注意。

$a$ については、(A)を変形して、

$$na = \sum y_i - b \sum x_i$$

$$a = \frac{1}{n} \sum y_i - \frac{1}{n} b \sum x_i$$

$$= \bar{y} - b\bar{x}$$

---

よって、 $y' = a + bx$

$$= \bar{y} - b\bar{x} + bx$$

$$= \bar{y} + b(x - \bar{x})$$

$$= \bar{y} + \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

ところで、 $r_{xy} = \dfrac{S_{xy}}{S_x S_y}$

$$\therefore y' = \bar{y} + r_{xy}\frac{S_y}{S_x}(x - \bar{x})$$

つまり、回帰式は、平均値$\bar{x}$ $\bar{y}$をとおり、傾きが$r_{xy}\dfrac{S_y}{S_x}$の直線である。

また、回帰式の両辺を $S_y$ で割ると、

$$\frac{y' - \bar{y}}{S_y} = r_{xy}\frac{x - \bar{x}}{S_x}$$

$v' = r_{xy}u$　　　相関係数は、標準測度の回帰係数（標準化された回帰係数）である。