

応用システム工学 第五回 回帰分析

平成25年06月07日
重回帰分析

相関係数

$$s_e^2 = \frac{n}{n-2} s_{yy} \{1 - r_{xy}^2\}$$

– 相関係数の値の範囲

$$\frac{n}{n-2} s_{yy} \{1 - r_{xy}^2\} = s_e^2 \geq 0 \quad \text{より} \quad s_{yy} \geq 0 \quad \text{だから}$$

$$1 - r_{xy}^2 \geq 0$$

$$-1 \leq r_{xy} \leq 1$$

– 相関係数が $r_{xy} = \pm 1$ のとき, 直線上にある

$$s_e = 0$$

– 相関係数が0に近いほど, x,yの線形的な関係は小さくなる

決定係数

- 相関係数の方が予測誤差の評価に適當
 - 相関係数 r_{xy} は, 一次変換(単位変換等)
$$X = ax + b \quad Y = cy + d$$
によって変化しない
 - 予測誤差の標準偏差 s_e は, 一次変換により変化する
- 決定係数(寄与率) → 相関係数の2乗

$$r_{xy}^2 = \frac{s_{xy}^2}{s_{xx}s_{yy}} = 1 - \frac{\sum_{i=1}^n (y_i - Y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

決定係数

$$\begin{aligned}\sum_{i=1}^n (y_i - Y_i)^2 &= \sum_{i=1}^n \left\{ y_i - \left[\frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) + \bar{y} \right] \right\}^2 \\ &= \sum_{i=1}^n \left\{ y_i - \bar{y} - \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right\}^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2 \frac{s_{xy}}{s_{xx}} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) + \left(\frac{s_{xy}}{s_{xx}} \right)^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= ns_{yy} - 2 \frac{s_{xy}}{s_{xx}} ns_{xy} + \left(\frac{s_{xy}}{s_{xx}} \right)^2 ns_{xx} = n \left(s_{yy} - \frac{s_{xy}^2}{s_{xx}} \right)\end{aligned}$$

決定係数

$$\begin{aligned} 1 - \frac{\sum_{i=1}^n (y_i - Y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} &= 1 - \frac{n \left(s_{yy} - \frac{s_{xy}^2}{s_{xx}} \right)}{n s_{yy}} \\ &= 1 - \frac{s_{yy} - \frac{s_{xy}^2}{s_{xx}}}{s_{yy}} \\ &= 1 - \left(1 - \frac{s_{xy}^2}{s_{xx} s_{yy}} \right) = \frac{s_{xy}^2}{s_{xx} s_{yy}} = r_{xy}^2 \end{aligned}$$

変数変換による線形回帰

- 変量 x, y の関係は必ずしも線形ではない

- 線形回帰モデル

$$y_i = a_0 + a_1 x_i + e_i \quad (i = 1, 2, \dots, n)$$

- 変数変換により線形関係が得られることがある

- 変数変換した線形回帰モデル

$$x_i \Rightarrow X_i = \log x_i$$

$$y_i = a_0 + a_1 X_i + e_i \quad (i = 1, 2, \dots, n)$$

- 相関係数で評価

線形重回帰

- 1つの目的変数 y に対する複数(p 個)の説明変数 x_1, x_2, \dots, x_p

– 線形重回帰モデル

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi} + e_i \quad (i = 1, 2, \dots, n)$$

	目的変数:y	説明変数:x	予測誤差
1	y_1	$x_{11}, x_{21}, \dots, x_{p1}$	$e_1 = y_1 - (a_0 + a_1 x_{11} + a_2 x_{21} + \dots + a_p x_{p1})$
2	y_2	$x_{12}, x_{22}, \dots, x_{p2}$	$e_2 = y_2 - (a_0 + a_1 x_{12} + a_2 x_{22} + \dots + a_p x_{p2})$
...			
i	y_i	$x_{1i}, x_{2i}, \dots, x_{pi}$	$e_i = y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})$
...			
n	y_n	$x_{1n}, x_{2n}, \dots, x_{pn}$	$e_n = y_n - (a_0 + a_1 x_{1n} + a_2 x_{2n} + \dots + a_p x_{pn})$

線形重回帰

- 単回帰 (x,y)平面
- 重回帰 $(x_1, x_2, \dots, x_p, y)$ (p+1)次元空間
 - P次元超平面

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p$$

– 説明変数 $x_{1i}, x_{2i}, \dots, x_{pi}$ から目的変数 y_i を予測

- 予測誤差

$$e_i = y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \dots + a_px_{pi})$$

- 予測誤差の平方和

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left\{ y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \dots + a_px_{pi}) \right\}^2$$

線形重回帰

- 予測誤差の平方和の最小化

$$F(a_0, a_1, \dots, a_p) = \sum_{i=1}^n e_i^2$$
$$= \sum_{i=1}^n \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\}^2$$

$$\left\{ \begin{array}{l} \frac{\partial}{\partial a_0} F(a_0, a_1, \dots, a_p) = 0 \\ \frac{\partial}{\partial a_1} F(a_0, a_1, \dots, a_p) = 0 \\ \vdots \\ \frac{\partial}{\partial a_p} F(a_0, a_1, \dots, a_p) = 0 \end{array} \right.$$

線形重回帰

$$\frac{\partial}{\partial a_0} F(a_0, a_1, \dots, a_p) = \frac{\partial}{\partial a_0} \sum_{i=1}^n \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\}^2$$

$$= -2 \sum_{i=1}^n \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\} = 0$$

$$\frac{\partial}{\partial a_1} F(a_0, a_1, \dots, a_p) = \frac{\partial}{\partial a_1} \sum_{i=1}^n \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\}^2$$

$$= -2 \sum_{i=1}^n x_{1i} \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\} = 0$$

$$\vdots \quad \frac{\partial}{\partial a_p} F(a_0, a_1, \dots, a_p) = \frac{\partial}{\partial a_p} \sum_{i=1}^n \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\}^2$$

$$= -2 \sum_{i=1}^n x_{pi} \left\{ y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi}) \right\} = 0$$

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_p x_{pi})\} = 0 \\ \sum_{i=1}^n x_{1i} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_p x_{pi})\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_p x_{pi})\} = 0 \end{array} \right.$$

線形重回帰

- 正規方程式: 残差二乗和を最小にする推定値を与える方程式 $\rightarrow a_0, a_1, \dots, a_p$ の連立方程式

$$\left\{ \begin{array}{l} a_0 n + a_1 \sum_{i=1}^n x_{1i} + a_2 \sum_{i=1}^n x_{2i} + \dots + a_p \sum_{i=1}^n x_{pi} = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_{1i} + a_1 \sum_{i=1}^n x_{1i}^2 + a_2 \sum_{i=1}^n x_{1i} x_{2i} + \dots + a_p \sum_{i=1}^n x_{1i} x_{pi} = \sum_{i=1}^n x_{1i} y_i \\ \vdots \\ a_0 \sum_{i=1}^n x_{pi} + a_1 \sum_{i=1}^n x_{pi} x_{1i} + a_2 \sum_{i=1}^n x_{pi} x_{2i} + \dots + a_p \sum_{i=1}^n x_{pi} x_{pi} = \sum_{i=1}^n x_{pi} y_i \end{array} \right.$$

線形重回帰

$$\begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{pi} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & & \sum_{i=1}^n x_{1i}x_{pi} \\ \vdots & & & & \vdots \\ \sum_{i=1}^n x_{pi} & \sum_{i=1}^n x_{pi}x_{1i} & \sum_{i=1}^n x_{pi}x_{2i} & & \sum_{i=1}^n x_{pi}x_{pi} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i}y_i \\ \vdots \\ \sum_{i=1}^n x_{pi}y_i \end{bmatrix}$$

線形重回帰

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{pi} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & & \sum_{i=1}^n x_{1i}x_{pi} \\ \vdots & & & & \vdots \\ \sum_{i=1}^n x_{pi} & \sum_{i=1}^n x_{pi}x_{1i} & \sum_{i=1}^n x_{pi}x_{2i} & & \sum_{i=1}^n x_{pi}x_{pi} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i}y_i \\ \vdots \\ \sum_{i=1}^n x_{pi}y_i \end{bmatrix}$$

$$a_0, a_1, \dots, a_p \Rightarrow \hat{a}_0, \hat{a}_1, \dots, \hat{a}_p$$

分散共分散で表すことを考える

線形重回帰

- x_1, \dots, x_p の分散共分散行列

$$V = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1l} & \cdots & s_{1p} \\ s_{21} & s_{22} & & & & s_{2p} \\ \vdots & & & & & \vdots \\ s_{j1} & s_{j2} & \cdots & s_{jl} & \cdots & s_{jp} \\ \vdots & & & & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pl} & \cdots & s_{pp} \end{bmatrix}$$

$$s_{jl} = \frac{1}{n} \sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l)$$

$$j, l = 1, 2, \dots, p$$

- y と x_1, \dots, x_p の共分散

$$s_{yj} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_{ji} - \bar{x}_j) \quad j = 1, 2, \dots, p$$

線形重回帰

分散共分散行列であらわす

- 正規方程式の一行目

$$\hat{a}_0 n + \hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} + \cdots + \hat{a}_p \sum_{i=1}^n x_{pi} = \sum_{i=1}^n y_i$$

$$\begin{aligned} \hat{a}_0 &= \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} \left(\hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} + \cdots + \hat{a}_p \sum_{i=1}^n x_{pi} \right) \\ &= \bar{y} - \left(\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p \right) \end{aligned}$$

$$\hat{a}_0 = \bar{y} - \left(\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p \right)$$

– 正規方程式の a_0 に代入

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n \left\{ y_i - \left(\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p) \right) + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi} \right\} = 0 \\ \sum_{i=1}^n x_{1i} \left\{ y_i - \left(\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p) \right) + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi} \right\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{ji} \left\{ y_i - \left(\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p) \right) + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi} \right\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \left\{ y_i - \left(\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p) \right) + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi} \right\} = 0 \end{array} \right.$$

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n \left\{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \right\} = 0 \\ \sum_{i=1}^n x_{1i} \left\{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \right\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{ji} \left\{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \right\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \left\{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \right\} = 0 \end{array} \right.$$

各行j=1~pから, j=0行 $\times \bar{x}_j$ を引く

共通

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n (x_{1i} - \bar{x}_1) \{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \} = 0 \\ \vdots \\ \sum_{i=1}^n (x_{ji} - \bar{x}_j) \{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \} = 0 \\ \vdots \\ \sum_{i=1}^n (x_{pi} - \bar{x}_p) \{ (y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \cdots - \hat{a}_p(x_{pi} - \bar{x}_p) \} = 0 \end{array} \right.$$

各行の両辺をnで割ると

線形重回帰

$$\left\{ \begin{array}{l} s_{y1} - \hat{a}_1 s_{11} - \hat{a}_2 s_{12} - \cdots - \hat{a}_p s_{1p} = 0 \\ \vdots \\ s_{yj} - \hat{a}_1 s_{j1} - \hat{a}_2 s_{j2} - \cdots - \hat{a}_p s_{jp} = 0 \\ \vdots \\ s_{yp} - \hat{a}_1 s_{p1} - \hat{a}_2 s_{p2} - \cdots - \hat{a}_p s_{pp} = 0 \end{array} \right.$$

$$\left\{ \begin{array}{l} \hat{a}_1 s_{11} + \hat{a}_2 s_{12} + \cdots + \hat{a}_p s_{1p} = s_{y1} \\ \vdots \\ \hat{a}_1 s_{j1} + \hat{a}_2 s_{j2} + \cdots + \hat{a}_p s_{jp} = s_{yj} \\ \vdots \\ \hat{a}_1 s_{p1} + \hat{a}_2 s_{p2} + \cdots + \hat{a}_p s_{pp} = s_{yp} \end{array} \right.$$

線形重回帰

$$\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ \vdots & & & \\ s_{j1} & s_{j2} & & s_{jp} \\ \vdots & & & \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} s_{y1} \\ \vdots \\ s_{yj} \\ \vdots \\ s_{yp} \end{bmatrix}$$

分散共分散行列V

yのxに対する共分散

別途 $\hat{a}_0 = \bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)$ で \hat{a}_0 を求めればよい

目的変数yの, 説明変数 x_1, x_2, \dots, x_p に対する線形重回帰式

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \cdots + \hat{a}_p x_p$$

重回帰係数

$$\hat{a}_j \quad j = 1, 2, \dots, p$$