

変数ベクトルと相関

統計学と三角関数

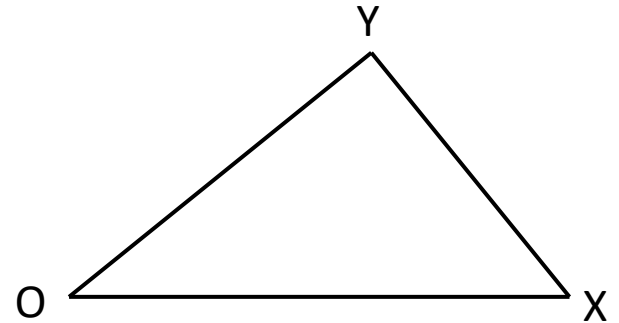
- 変数Xの中心化データ $(x_1, x_2, \dots, x_N)$ で表される点をX, 変数Yの中心化データ $(y_1, y_2, \dots, y_N)$ で表される点をYとする
- $\triangle OXY$ の各辺の長さ (の2乗)

$$OX^2 = \sum (x_i - 0)^2 = \sum x_i^2 = N S_X^2$$

$$OY^2 = \sum (y_i - 0)^2 = \sum y_i^2 = N S_Y^2$$

$$XY^2 = \sum (x_i - y_i)^2 = \sum (x_i^2 + y_i^2 - 2x_i y_i)$$

$$= \sum x_i^2 + \sum y_i^2 - 2 \sum x_i y_i = N S_X^2 + N S_Y^2 - 2 N S_{XY}$$



統計学と三角関数

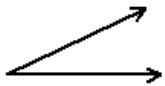
- $\angle XOY$ の大きさを θ とし, $\triangle OXY$ に余弦定理を適用すると,

$$\begin{aligned}\bullet \cos \theta &= \frac{OX^2 + OY^2 - XY^2}{2 OX OY} = \frac{NS_X^2 + NS_Y^2 - (Ns_x^2 + NS_Y^2 - 2NS_{XY})}{2NS_X S_Y} \\ &= \frac{2NS_{XY}}{2NS_X S_Y} = \frac{S_{XY}}{S_X S_Y} \\ &= r\end{aligned}$$

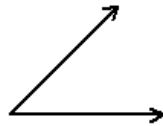
- 変数Xと変数Yを表す点のなす角は, 変数Xと変数Yの相関係数に対応する

変数を表す点の位置と相関係数の関係

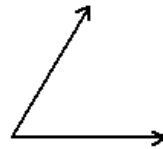
$r = 0.9$



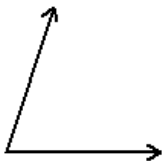
$r = 0.7$



$r = 0.5$



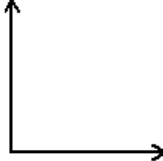
$r = 0.3$



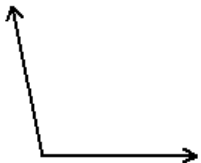
$r = 0.1$



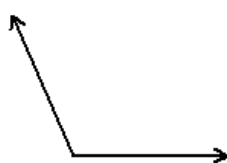
$r = 0$



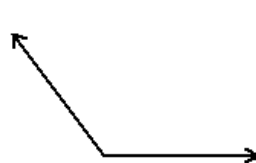
$r = -0.2$



$r = -0.4$



$r = -0.6$



- 無相関 ($r=0$) の場合,
- 変数Xの個人差と変数Yの個人差に,
- 一方が高ければ他方も高く、一方が低ければ他方も低いのような共変関係がないので,
- 2つの点の位置は直交する
- 相関係数0.3でも、個人差の方向性が似ているとは見なし難い

予測分散・残差分散

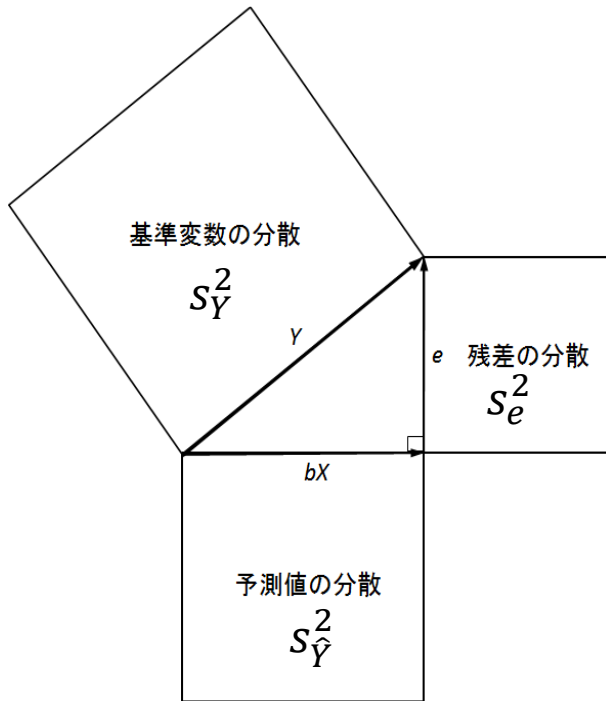


図17.8 予測分散・残差分散

- 説明変数 X （ひいては予測値 $\hat{Y} = bX$ ）と残差 e は無相関である（ X と e は直交する）

- 基準変数の分散は，予測値の分散と残差（誤差）分散に分割される

$$s_Y^2 = s_{\hat{Y}}^2 + s_e^2$$

- 予測の精度が高いほど，残差（誤差）は小さい

- 残差（誤差）が小さいほど，予測値の分散は，基準変数の分散に近くなる

決定係数（分散説明率）

- 基準値の分散に占める予測値の分散の割合

$$R^2 = \frac{\hat{Y} \text{の分散}}{Y \text{の分散}} = \frac{s_{\hat{Y}}^2}{s_Y^2} = 1 - \frac{s_e^2}{s_Y^2}$$

- $0 \leq R^2 \leq 1$
- R^2 が1に近いほど残差分散は小さい
- R^2 は予測の精度（良さ）を表す指標
- R^2 が1に近いほど，説明変数 X で，基準変数 Y の値を，よく予測できる

単回帰分析の決定係数

- $R^2 = \frac{\hat{Y} \text{の分散}}{Y \text{の分散}} = \frac{s_{\hat{Y}}^2}{s_Y^2} = 1 - \frac{s_e^2}{s_Y^2}$
- $\hat{Y} = a + bX$
$$b = r \frac{s_Y}{s_X}$$
- $s_{\hat{Y}}^2 = b^2 s_X^2 = r^2 \frac{s_Y^2}{s_X^2} s_X^2 = r^2 s_Y^2$
- $s_e^2 = s_Y^2 - r^2 s_Y^2 = (1 - r^2) s_Y^2$

- $R^2 = \frac{s_{\hat{Y}}^2}{s_Y^2} = \frac{r^2 s_Y^2}{s_Y^2} = r^2$

- 単回帰分析の決定係数は、相関係数の2乗に等しい