線形回帰モデル (P139)

$$\begin{cases} y_1 = \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\ \quad\vdots \\ y_N = \beta_1 x_{N1} + \beta_2 x_{N2} + \cdots + \beta_k x_{Nk} + \varepsilon_N \end{cases}$$

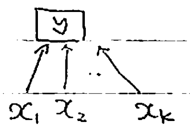
$\leftarrow y_i = \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$ $(i = 1, 2, \ldots, N)$.

y_i : 従属変数 or 被説明変数 (確率変数).

x_{ji} : 独立変数 or 説明変数 (定数) ← 与えられた変数とみなす

ε_i : 誤差項

y を $x_1, \ldots, x_k$ で説明する時, 説明できない部分が ε の中に含まれている.



$K=1$ の時, 単回帰モデル

$K \geq 2$ 〃 重回帰 〃

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}}_{= \mathbb{Y}} = \underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{pmatrix}}_{= \mathbb{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}}_{= \beta} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{= \mathbb{E}}.$$

$$\mathbb{Y} = \mathbb{X}\beta + \mathbb{E} \quad \leftrightarrow \quad \mathbb{Y} \sim N(\mathbb{X}\beta, \sigma^2 \mathbb{I}) \quad E(\mathbb{Y}) = \mathbb{X}\beta$$

$$V(\mathbb{Y}) = \sigma^2 \mathbb{I}.$$

P141

仮定

(i) $E(\mathbb{E}) = \mathbb{O}$

(ii) $\underset{\substack{\uparrow \\ \text{共分散行列}}}{V(\mathbb{E})} = E(\mathbb{E}\mathbb{E}') = \sigma^2 \mathbb{I} = \underbrace{\begin{pmatrix} \sigma^2 & & \mathbb{O} \\ & \sigma^2 & \\ \mathbb{O} & & \sigma^2 \end{pmatrix}}_{N} \Big\} N$.

$\varepsilon_1, \ldots, \varepsilon_N$ は互いに無相関

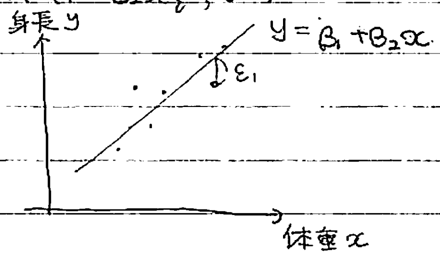
(iii) $(\mathbb{X}'\mathbb{X})^{-1}$ が存在 (or $\mathrm{rank}(\mathbb{X}) = K$).

No. Date . .

prob
ex1 $k=2$ $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$ $\mathbb{X}=\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$ $\beta=\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$

$y_i=\beta_1+\beta_2 x_i+\varepsilon_i \quad (i=1,2,\dots,N)$

$y_i \sim N(\beta_1+\beta_2 x_i, \sigma^2)$



最小2乗推定量 (Least Squares Estimator (LSE))

$\|\mathbb{y}-\mathbb{X}\beta\|^2=(\mathbb{y}-\mathbb{X}\beta)'(\mathbb{y}-\mathbb{X}\beta)$ を最小にする $\beta=\hat{\beta}$ を
最小2乗推定量と言う

$$0=\frac{\partial}{\partial\beta}\|\mathbb{y}-\mathbb{X}\beta\|^2=\frac{\partial}{\partial\beta}(\mathbb{y}'\mathbb{y}-\beta'\mathbb{X}'\mathbb{y}-\underline{\mathbb{y}'\mathbb{X}}\beta+\beta'\mathbb{X}'\mathbb{X}\beta)$$

$$=\frac{\partial}{\partial\beta}(\mathbb{y}'\mathbb{y}-2\beta'\mathbb{X}'\mathbb{y}+\beta'\mathbb{X}'\mathbb{X}\beta)$$

$$=-2\mathbb{X}'\mathbb{y}+2\mathbb{X}'\mathbb{X}\beta$$

$\Leftrightarrow \quad \mathbb{X}'\mathbb{X}\beta=\mathbb{X}'\mathbb{y}$

$\Leftrightarrow \quad \beta=(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{y}$

$$\underline{\hat{\beta}=(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{y}}$$

No. Date . .

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon)$$
$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon$$
$$= \beta + (X'X)^{-1}X'\varepsilon \qquad (\hat{\beta}\text{の確率的表現})$$

(1) 不偏性 (P144–145) $E(\hat{\beta}) = \beta$

PROOF

$$E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'\varepsilon)$$
$$= \beta + (X'X)^{-1}X'\underbrace{E(\varepsilon)}_{=0} = \beta$$

(2) 分散共分散行列 $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$

$$\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon$$

$$V(\hat{\beta}) = E((\hat{\beta}-\beta)(\hat{\beta}-\beta)')$$
$$= E\left((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right) \qquad X'X\text{は対称}$$
$$= (X'X)^{-1}X'\underbrace{E(\varepsilon\varepsilon')}_{=V(\varepsilon)=\sigma^2 I}X(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{=I_k}$$
$$= \sigma^2(X'X)^{-1}.$$

(3) $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

PROOF

Fact 23.2より

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_N)$$

$\hat{\beta}$は多変量正規分布に従う. $\hat{\beta}$の期待値と共分散行列は(1),(2)より

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

P150

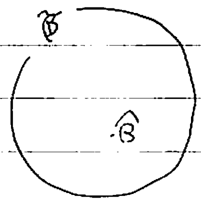
Gauss－Markovの定理

仮定 (i)－(iii).

$\hat{\beta}$ は BLUE (Best Linear Unbiased Estimator).

線形で不偏な推定量 ($\mathbb{C}y$) の集まりの中で

$\hat{\beta}$ は最も小さい分散を持つ



$E(\tilde{\beta}) = \beta$

i.e.,

$$\tilde{\beta} = \mathbb{C}y = \begin{pmatrix} c_{11} & \cdots & c_{1N} \\ \vdots & & \vdots \\ c_{K1} & \cdots & c_{KN} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 \\ \vdots \\ \tilde{\beta}_K \end{pmatrix}.$$

$$\hat{\beta} = \underbrace{(X'X)^{-1}X'}_{\mathbb{C}\text{の1つ}} y$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$$

$$V(\hat{\beta}_k) \leq V(\tilde{\beta}_k) \quad \text{for } \forall k$$

P151

補題1. $\mathbb{C}X = I_K$

PROOF: $\beta = E(\tilde{\beta})$ for $\forall \beta \neq 0$

$= E(\mathbb{C}y)$ $\quad y \sim N(X\beta, \sigma^2 I_N)$

$= \mathbb{C}X\beta$ $\quad \beta \neq 0$ より

$\Rightarrow \mathbb{C}X = I_K.$

補題2 $V(\tilde{\beta}) = \sigma^2 \mathbb{C}\mathbb{C}'$

PROOF $\tilde{\beta} = \mathbb{C}y = \mathbb{C}(X\beta + \varepsilon)$

$= \mathbb{C}X\beta + \mathbb{C}\varepsilon$ $\quad$ 補題1より

$= \beta + \mathbb{C}\varepsilon$

$\tilde{\beta} - \beta = \mathbb{C}\varepsilon$

No. Date . .

$$V(\hat{\beta}) = E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)']$$ $\quad \hat{\beta}-\beta = \mathbb{C}\varepsilon$
$$= E(\mathbb{C}\varepsilon(\mathbb{C}\varepsilon)')$$
$$= \mathbb{C}\,E[\varepsilon\varepsilon']\,\mathbb{C}'$$
$$= \mathbb{C}\,\sigma^2 \mathbb{I}_N\,\mathbb{C}'$$
$$= \sigma^2 \mathbb{C}\mathbb{C}'.$$

ここで $\mathbb{C} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' + D$ $\quad {}_{K}\{\underbrace{D}_{N}$: 任意の行列.

P153

不偏性の条件 $\quad \hat{\beta} = \mathbb{C}\mathbb{y} = \{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' + D\}\mathbb{y}$

$\boxed{\text{この時}\quad D\mathbb{X} = \underbrace{\mathbb{O}}_{K}\}{}^{K}}$

PROOF: $\quad \mathbb{C}\mathbb{X} = \mathbb{I}_K$

$$\Longrightarrow \{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' + D\}\mathbb{X} = \mathbb{I}_K$$
$$\Longrightarrow \mathbb{I}_K + D\mathbb{X} = \mathbb{I}_K$$
$$\Longrightarrow D\mathbb{X} = \mathbb{O}$$

任意の線形不偏推定量の分散の差

$$\boxed{\sigma^2\mathbb{C}\mathbb{C}' = \sigma^2(\mathbb{X}'\mathbb{X})^{-1} + \sigma^2 DD'}$$

PROOF. $\quad \sigma^2\mathbb{C}\mathbb{C}' = \sigma^2\{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' + D\}\{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' + D\}'$

$$= \sigma^2\left(((\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' + D)(\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1} + D')\right)$$
$$= \sigma^2\Big(\underbrace{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}}_{=\mathbb{I}}(\mathbb{X}'\mathbb{X})^{-1} + (\mathbb{X}'\mathbb{X})^{-1}\underbrace{\mathbb{X}'D'}_{=\mathbb{O}} + \underbrace{D\mathbb{X}}_{=\mathbb{O}}(\mathbb{X}'\mathbb{X})^{-1} + DD'\Big)$$
$$= \sigma^2\left((\mathbb{X}'\mathbb{X})^{-1} + DD'\right).$$

$$\longrightarrow \underline{V(\tilde{\beta}) = V(\hat{\beta}) + \sigma^2 DD'}$$

$$V(\tilde{\beta}_K) = V(\hat{\beta}_K) + \sigma^2 \cdot \underbrace{\text{対角成分}}_{\geq 0}$$

$$V(\tilde{\beta}_K) \geq V(\hat{\beta}_K)$$

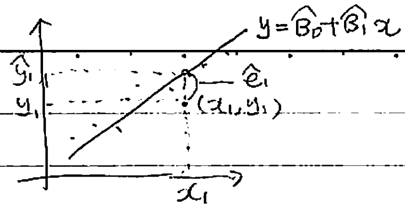
P155

$$D = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{pmatrix}$$

$$DD' = \begin{pmatrix} \sum d_{1i}^2 & \sum d_{1i}d_{2i} & \sum d_{1i}d_{3i} \\ - & \sum d_{2i}^2 & - \\ & \text{---} & \sum d_{3i}^2 \end{pmatrix}$$

対角成分は 0 以上

残差の分布について.



残差 $\hat{e} = y - X\hat{\beta}$ データと予測値のずれ

$$\begin{pmatrix}\hat{e}_1\\ \vdots \\ \hat{e}_N\end{pmatrix} \qquad \hat{e}'\hat{e} = (y-X\hat{\beta})'(y-X\hat{\beta}) = \sum_i \left(y_i - (\hat{\beta}_0+\hat{\beta}_1 x_{1i}+\cdots+\hat{\beta}_k x_{ki})\right)^2$$

表現 $\hat{e} = (I - X(X'X)^{-1}X')\varepsilon$. 確率ベクトル.

$$\hat{e} = y - X(X'X)^{-1}X'y$$
$$=(I - X(X'X)^{-1}X')y$$
$$=(I-X(X'X)^{-1}X')(X\beta+\varepsilon)$$
$$= X\beta - X\cancel{(X'X)^{-1}X'X}\beta + (I - X(X'X)^{-1}X')\varepsilon$$
$$= \cancel{X\beta} - \cancel{X\beta} + (I - X(XX)X')\varepsilon \quad \square$$

$M = I - X(X'X)^{-1}X'$ は対称なべき等行列.

・Mは対称、 $M'=M$ ← 演習問題

・Mはべき等行列 $M^2=M$.

残差平方和 $\hat{e}'\hat{e} = \varepsilon'(I - X(X'X)^{-1}X')\varepsilon$.

$$\hat{e}'\hat{e} = \varepsilon'(I-X(X'X)^{-1}X')(I-X(X'X)^{-1}X')\varepsilon$$
$$= \varepsilon'(I-X(X'X)^{-1}X')\varepsilon.$$

定理 (P255 . Fact 25.3)

A: 対称なべき等行列, $\mathrm{rank}(A)=R$

$\varepsilon \sim N(0, \sigma^2 I)$

$$\frac{1}{\sigma^2}\varepsilon' A \varepsilon \sim \chi^2(R)$$

No. Date . .

$\hat{\mathbb{e}}'\hat{\mathbb{e}} = \mathbb{\varepsilon}'\underbrace{(\mathbb{I}-\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')}_{M}\mathbb{\varepsilon}$

M：対称，べキ等行列

$\mathrm{rank}(M) = N-k$ (P235 Fact 22.37).

$\mathbb{\varepsilon} \sim N(\mathbb{0}, \sigma^2\mathbb{I})$

定理 (P255 Fact 25.3) より

$$\frac{1}{\sigma^2}\mathbb{\varepsilon}'(\mathbb{I}-\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')\mathbb{\varepsilon} \sim \chi^2(N-k)$$

$$\therefore E\left(\frac{1}{\sigma^2}\hat{\mathbb{e}}'\hat{\mathbb{e}}\right) = N-k$$

$$E\left(\frac{1}{N-k}\hat{\mathbb{e}}'\hat{\mathbb{e}}\right) = \sigma^2$$

◎ $\hat{\sigma}^2 = \frac{1}{N-k}\hat{\mathbb{e}}'\hat{\mathbb{e}} = \frac{1}{N-k}\sum_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}))^2.$

は σ^2 の不偏推定量となる

P119.

[決定係数]

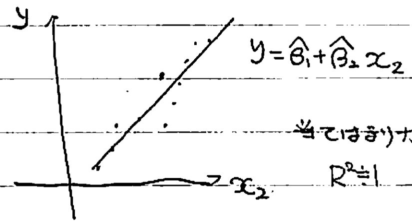
$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$$ ← 注意!

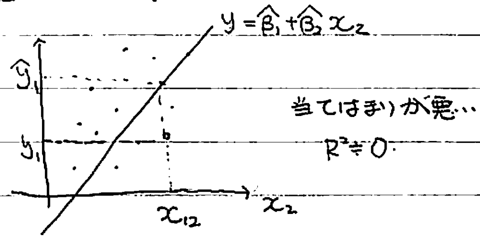
回帰式 $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_K x_{iK}$.

どのくらい回帰式の当てはまりが良いのか?

ex. (P109) K=2

$$y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i \qquad X = \begin{pmatrix} 1 & x_{12} \\ \vdots & \vdots \\ 1 & x_{N2} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$





$$R^2 = \frac{\overbrace{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}^{\text{説明された変動}}}{\underbrace{\sum_{i=1}^{N}(y_i - \bar{y})^2}_{\text{全変動}}} \qquad \bar{\hat{y}} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\text{実は} = \left\{ \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}} \right\}^2$$

yと$\hat{y}$の相関係数の2乗

$$\sum_{i=1}^{N}(y_i - \bar{y})^2 = \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2 + \sum_{i=1}^{N}\hat{e}_i^2 \quad (※) \qquad \leftarrow \hat{e} = y - X\hat{\beta} = \begin{pmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_N \end{pmatrix}$$

$$1 = \underbrace{\boxed{\frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}}}_{= R^2} + \frac{\sum_{i=1}^{N}\hat{e}_i^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

→ 回帰式$\hat{y}_i$でデータの何%を表わせているかを示している

No. Date . .

P121 Fact 13.2

$$\boxed{\mathbb{X}'\hat{e} = 0 \qquad \hat{e}'\mathbb{X} = 0'}$$

PROOF

$$\mathbb{X}'(\mathbb{y} - \mathbb{X}\hat{\beta}) = \mathbb{X}'\mathbb{y} - \mathbb{X}'\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{y} = 0$$

P122 Fact 13.3.

$$\bar{\hat{y}} = \bar{y}$$

PROOF. $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \} N$ とする. $\frac{1}{N}\mathbb{1}'\mathbb{y} = \frac{1}{N}(1 \cdots 1)\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \frac{1}{N}\sum_{i=1}^{N} y_i$

$$\frac{1}{N}\mathbb{1}'\hat{\mathbb{y}} = \frac{1}{N}\mathbb{1}'\mathbb{X}\hat{\beta} = \frac{1}{N}\mathbb{1}'(\mathbb{y} - \mathbb{X}\hat{\beta}) + \frac{1}{N}\mathbb{1}'\mathbb{y}$$

$$= \frac{1}{N}\mathbb{1}'\hat{e} + \bar{y}$$

$$= \underbrace{\frac{1}{N}\sum_{i=1}^{N}\hat{e}_i}_{\mathbb{X}'\hat{e}=0 \text{より}} + \bar{y} = \bar{y}$$

$$(\hat{e}_1 \cdots \hat{e}_N)\begin{pmatrix} 1 & x_{12} & \cdots & x_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N2} & \cdots & x_{NK} \end{pmatrix} = (0 \cdots 0)$$

補題

$$\mathbb{y}'\mathbb{y} = \hat{\mathbb{y}}'\hat{\mathbb{y}} + \hat{e}'\hat{e}. \qquad (※2).$$

PROOF. $\mathbb{y}'\mathbb{y} = (\mathbb{X}\hat{\beta} + \underbrace{\mathbb{y} - \mathbb{X}\hat{\beta}}_{=\hat{e}})'(\mathbb{X}\hat{\beta} + \underbrace{\mathbb{y} - \mathbb{X}\hat{\beta}}_{=\hat{e}})$

$$= (\hat{\beta}'\mathbb{X}' + \hat{e}')(\mathbb{X}\hat{\beta} + \hat{e})$$

$$= \hat{\beta}'\mathbb{X}'\mathbb{X}\hat{\beta} + \underbrace{\hat{e}'\mathbb{X}\hat{\beta}}_{=0} + \hat{\beta}'\underbrace{\mathbb{X}'\hat{e}}_{=0 \text{ Fact 13.2}} + \hat{e}'\hat{e}$$

$$= \hat{\mathbb{y}}'\hat{\mathbb{y}} + \hat{e}'\hat{e}.$$

(※2)は $\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{e}_i^2$ と表わせる

Fact 13.3.

$$\sum_{i=1}^{N} y_i^2 - N\bar{y}^2 = \sum_{i=1}^{N} \hat{y}_i^2 - N\bar{\hat{y}}^2 + \sum_{i=1}^{N} \hat{e}_i^2 \qquad \leftarrow \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = \frac{1}{N}\sum y_i^2 - (\bar{y})^2$$

$$\sum_{i=1}^{N}(y_i - \bar{y})^2 = \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2 + \sum_{i=1}^{N} \hat{e}_i^2$$

$\hat{\sigma}^2$ と $\hat{\beta}$ の関係

定理

$\boxed{\cdot\ \hat{\sigma}^2 \text{ と } \hat{\beta} \text{ は独立}}$

(review). ← 宮岡 数理統計学の基礎 P38−40.

確率ベクトル $\mathbb{X}=(X_1,\dots,X_n)'$ と $\mathbb{Y}=(Y_1,\dots,Y_m)'$ は独立　$\mathbb{X}\perp\!\!\!\perp\mathbb{Y}$ と書く

$$\overset{def}{\Leftrightarrow}\quad \underbrace{f_{\mathbb{X}\mathbb{Y}}(x,y)}_{\mathbb{X}\text{ と }\mathbb{Y}\text{ の同時pdf}} = f_{\mathbb{X}}(x)\, f_{\mathbb{Y}}(y)$$

or

$$P(\mathbb{X}\in\mathcal{A},\ \mathbb{Y}\in\mathcal{B}) = P(\mathbb{X}\in\mathcal{A})\, P(\mathbb{Y}\in\mathcal{B})$$

$$\mathbb{X}\perp\!\!\!\perp\mathbb{Y} \Rightarrow g(\mathbb{X})\perp\!\!\!\perp h(\mathbb{Y})$$

Fact 24.4. (P247)

$\varepsilon \sim N(0,\ \sigma^2 I_N)$

M: 対称なベキ等行列

$\varepsilon' M \varepsilon$ と $L\varepsilon$ が独立 $\Leftrightarrow$ $LM=0$

PROOF 省略

これを用いると　$\hat{\sigma}^2=\frac{1}{N-k}\varepsilon' M\varepsilon$　　$M=I-\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$

$$(N-k)\hat{\sigma}^2=\varepsilon' M\varepsilon.$$

$$\hat{\beta}-\beta=\underbrace{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'}_{=L}\varepsilon$$

$$LM=(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\left(I-\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\right)$$
$$=(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'-(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'=0$$

$\therefore\ \hat{\beta}-\beta$ と $(N-k)\hat{\sigma}^2$ は独立

$\Rightarrow$ $\hat{\beta}$ と $\hat{\sigma}^2$ は独立

No. Date . .

(review)

t-分布 (岩田 暁一 P137-141)

$X \sim N(0,1)$, $Y \sim \chi^2(\nu)$, $X \perp\!\!\!\perp Y$ とする.

$T = \dfrac{X}{\sqrt{Y/\nu}}$ は自由度 ν の t 分布に従う. ← $T \sim t(\nu)$ とかく

pdf $f(t) = \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1+\dfrac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ $-\infty < t < \infty$.

$E(T) = 0$ $V(T) = \dfrac{\nu}{\nu-2}$

定理 $(X'X)^{-1}$ の第 i 対角要素を a^{ii} とする.

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{a^{ii}}} \sim t(N-k).$$

Proof ・ $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$ ← $(X'X)^{-1} = \begin{pmatrix} a^{11} & a^{12} & \cdots \\ a^{21} & a^{22} & \\ \vdots & & \ddots & a^{kk} \end{pmatrix}$

$\Rightarrow \hat{\beta}_i \sim N(\beta_i, \sigma^2 a^{ii})$ $(i=1,2,\ldots,k)$.

$\Rightarrow \underbrace{\dfrac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{a^{ii}}}}_{= X} \sim N(0, 1^2).$

一方で $\dfrac{1}{\sigma^2}\hat{e}'\hat{e} \sim \chi^2(N-k)$

$= \dfrac{N-k}{\sigma^2}\hat{\sigma}^2 = Y$

ここで $\hat{\beta}_i \perp\!\!\!\perp \hat{\sigma}^2 \Rightarrow$ X と Y は独立

$T = \dfrac{X}{\sqrt{Y/\nu}} \sim t(N-k)$

$= \dfrac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{a^{ii}}} \Big/ \sqrt{\dfrac{(N-k)\hat{\sigma}^2}{\sigma^2 (N-k)}} = \dfrac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{a^{ii}}}$ □.

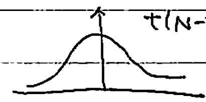
P159 検定

$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad (i=1,2,\dots,N)$

犬の体重　飼い主の体重　運動時間

帰無仮説　　　対立仮説

$H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$



① H_0が正しいと仮定

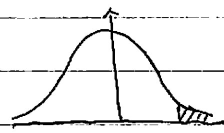
② ①の下で$\hat{\beta}_1$の理論的な確率分布を求める → $\dfrac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\sqrt{a''}} \sim t(N-2)$

③ 実際のデータから$\hat{\beta}_1$を計算する(これを$\hat{\beta}_1$の実現値と言い$\hat{\beta}_1^*$と書く).

そして$\hat{\beta}_1^*$とH_0のずれを測定

$$T = \frac{\hat{\beta}_1 - \overset{=0}{\beta_1}}{\hat{\sigma}\sqrt{a''}} \sim t(N-2)$$

$$t^* = \frac{\hat{\beta}_1^* - \overset{=0}{\beta_1}}{\hat{\sigma}\sqrt{a''}}$$



p-value と言う.

$2P(T > t^*) < 0.05 \Rightarrow H_0$を棄却

〃 $\geq 0.05 \Rightarrow H_a$を採択

No. Date . .

さいゆう
[最尤推定量] $\varepsilon \sim N(\mathbb{0}, \sigma^2 \mathbb{I}_N)$

(Maximum Likelihood Estimator)

MLE

$\mathbb{y} = \mathbb{X}\mathbb{B} + \varepsilon \qquad \mathbb{y} \sim N(\mathbb{X}\mathbb{B}, \sigma^2 \mathbb{I}_N)$

pdf $f(\mathbb{y}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B})\right\}$

対数尤度 $l(\mathbb{B}, \sigma^2) = \log f(\mathbb{y})$

$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B})$

$l(\mathbb{B}, \sigma^2)$を最大になるような $\mathbb{B}, \sigma^2$ を求める

$\mathbb{0} = \frac{\partial}{\partial \mathbb{B}} l(\mathbb{B}, \sigma^2) \qquad -\frac{1}{2\sigma^2}\frac{\partial}{\partial \mathbb{B}}(\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B}) = \mathbb{0}$

$\frac{\partial}{\partial \mathbb{B}}(\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B}) = \mathbb{0}$ i.e. $\hat{\mathbb{B}}_{MLE} = \hat{\mathbb{B}}$

$0 = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B})$

$\sigma^2 N = (\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B})$

$\hat{\sigma}^2_{MLE} = \frac{1}{N}(\mathbb{y}-\mathbb{X}\mathbb{B})'(\mathbb{y}-\mathbb{X}\mathbb{B}) \neq \hat{\sigma}^2$: 不偏推定量.

— 最尤推定量とは? <わしくは赤平「統計解析入門」P106, 統計学入門, 東京大学出版会 P217

表が出る確率が p の歪んだコインを N 回投げる.

i回目に投げたコインの結果を

Y_i	1(表)	0(裏)
確率	p	$1-p$

と表わす.

この時, Y_i の確率関数は $f_{Y_i}(y_i|p) = p^{y_i}(1-p)^{1-y_i} \quad (y_i = 0 \text{ or } 1)$

となる. よって $Y_1, \ldots, Y_N$ は iid より 同時pdfは

$f_{\mathbb{Y}}(\mathbb{y}|p) = \prod_{i=1}^{N}\{p^{y_i}(1-p)^{1-y_i}\} = p^{\sum y_i}(1-p)^{N-\sum y_i}$ となる

$L(p) = p^{\sum_{i=1}^{N} y_i}(1-p)^{N-\sum_{i=1}^{N} y_i}$ を尤度(ゆうど)と言う.

$l(p) = \log L(p)$ を対数尤度と言う. $l(p) = \sum_{i=1}^{N} y_i \log p + (N - \sum_{i=1}^{N} y_i)\log(1-p)$

$l(p)$を最大にする$\hat{p}$を p の最尤推定量と言う.

$\frac{d}{dp}l(p) = 0$ を解くと $\frac{d}{dp}l(p) = \left(\sum_{i=1}^{N} y_i\right)\frac{1}{p} - \left(N - \sum_{i=1}^{N} y_i\right)\frac{1}{1-p} = 0$. $\therefore \hat{p} = \frac{1}{N}\sum_{i=1}^{N} y_i$

モデル選択

Full model : $\mathbb{y} = \mathbb{X}\mathbb{B} + \mathbb{\varepsilon}$ $\quad \mathbb{X} = (\underbrace{\mathbb{X}_1}_{p}\ \underbrace{\mathbb{X}_2}_{k-p})$ $\quad \mathbb{B} = \begin{pmatrix} \mathbb{B}_1 \\ \mathbb{B}_2 \end{pmatrix}$

$M(p)$: $\mathbb{y} = \mathbb{X}_1\mathbb{B} + \mathbb{\varepsilon}$

↑
少ない説明変数

- 説明変数の数

多 ⟹ ・データのモデルへの当てはまりが良くなる

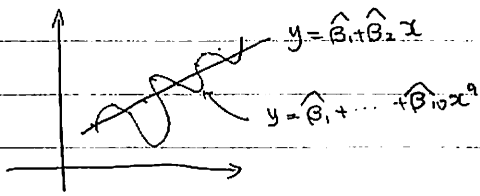
・$\mathbb{B}$の推定精度が落ちる

・モデルが不安定になり予測誤差が大きくなる

少 ⟹ ・$\mathbb{B}$の推定精度は高くなる

・データのモデルへの当てはまりが悪くなる

$y_i = \beta_1 + \beta_2 x_i + \cdots + \beta_k x_i^{k-1} + \varepsilon_i$



No. Date . .

- AIC (赤池情報量基準)

$Y_1, \dots, Y_N \overset{iid}{\sim} f(y|\theta)$

$M(p): \theta = (\theta_1, \dots, \theta_p)'$

$L(\theta) = \prod_{i=1}^{N} f(y_i|\theta)$: 尤度

$\hat{\theta}_{ML}$ $L(\hat{\theta}_{MLE}) = \sup_{\theta} L(\theta)$

$g(x)$: 真の確率分布

$\int \log \frac{g(y)}{f(y|\hat{\theta}_{ML})} g(y)dy$

$= \underbrace{\int \log g(y)\, g(y)dy}_{\text{ある数}} - \underbrace{\int \log f(y|\hat{\theta}_{ML})\, g(y)dy}_{\log L(\hat{\theta}_{ML}) - p}$

$M(p)$ に対する AIC は

$AIC(p) = \underbrace{-2\log L(\hat{\theta}_{ML})}_{\text{モデルの当てはまりの良さ}} + 2p$ ← パラメーターの数

$2p$: モデルの安定性

これを最小にするモデルを選ぶ

線形回帰モデルに当てはめてみる

$M(k+1): \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$

$$L(\theta) = \frac{1}{(\sqrt{2\pi})^N \sigma^N} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

MLE $\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

$\hat{\sigma}^2_{ML} = \frac{1}{N}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML})$

$$\frac{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML})}{\hat{\sigma}^2_{ML}} = N$$

$$L(\hat{\theta}_{ML}) = \frac{1}{(\sqrt{2\pi})^N \hat{\sigma}^N_{ML}} \exp\left\{-\frac{N}{2}\right\}$$

$$\log L(\hat{\theta}_{ML}) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\hat{\sigma}^2_{ML} - \frac{N}{2}$$

$$AIC(k+1) = N\log\hat{\sigma}^2_{ML} + N(\log 2\pi + 1) + 2(k+1)$$