10/3(一) 出遊
12/13(一) 出遊(二)
10/14(五) 3,4 補回
1/6(五) 3,4 補課

期中考 11/7(一) (0補課) 期末考 1/9(一) (0補課) 自主學習週 11/21-25.

回歸分析 · Montgomery Peck & Vining

Introduction to linear regression analysis 5th ed

· 作業、期中考、期末考
30%, 30%, 40%
(資料分析 etc)
、R
、SAS

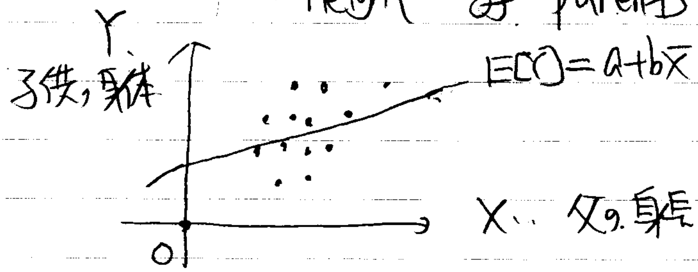
、線形代數
、微積分
、確率統計

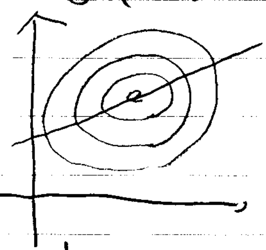
Regression. finding the relationship between variables discovering how variable(s) affect other variables. (response variables or dependent, or dependent variables.)

(covariates, independent variables. predictor variables or regressors)

· Francis Galton. フランシス・ゴルトン.

… height of parents / height of offsprings.





regression line. regression towards mediocrity

No.
Date : :

Model (モデル) $Y = f(X) + \varepsilon$

- X … independent variable
- Y … dependent variable
- ε … random error, $E[\varepsilon|X] = 0$

$$E[Y|X] = E[f(X) + \varepsilon|X] = f(X)$$

(regression function)

Y = response variable (反応変数)

- quantitative … 量的 …
  - discrete (離散)
  - continuous (連続)
- qualitative … 質的 …
  - nominal … (名義)
  - oridinal (順序)

model…

1. f … unknown
   - smooth or piecewise smooth
   - non-parametric model
2. $f = g(x|\theta)$ (g: known, θ: unknown)
3. $f(X) = \underline{\beta_0} + \underline{\beta_1} X$ simple linear regression
   未知
   (単回帰 model)

data: (X_i, Y_i) 數據

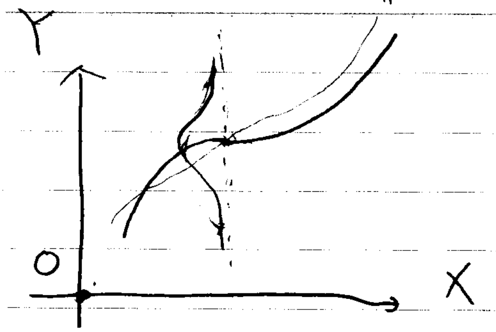
$(X_1, Y_1), (X_2, Y_2), \dots (X_n, Y_n)$

9/12
[8]

Cheryv culture

No.
Date 9 12 ①

$Y = f(X) + \varepsilon$. true.



Data. $(X_1, Y_1) \cdots (X_n, Y_n)$

- interpolation
- extrapolation.

Yの条件付分布 $(Y|X)$

- Model specification
- Model fitting
- Model checking
- Model validation.

• Data Collection (デタの収集)

1. retrospective study (回顧的) → 過去のデタに基づいてということだろうか?
2. observational study
3. designed experiment (デザインされた実験)

- multiple covariates

$Y = f(X_1, X_2 \ldots X_k) + \varepsilon$

$\alpha + f_1(X_1) + \cdots f_k(X_k)$

$\beta_0 + \beta_1 X_1 + \cdots \beta_k X_k$

linear regression

↑ パラメータの線形

linear in parameters

X_1, X_2

(- $y = \beta_0 + \beta_1 x_1^2 + \varepsilon$ 〃
- $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \cdots \varepsilon$

これらも線形回帰モデル

Chryo culture

No.

Date : :

Simple linear regression. (線形単回帰)

Data... $(X_1,Y_1),\dots,(X_n,Y_n)$ are coming from $Y=\beta_0+\beta_1X+\varepsilon$

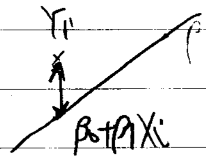
$E[\varepsilon]=0$

つまり $Y_i=\beta_0+\beta_1X_i+\underbrace{\varepsilon_i}_{i.i.d}$ $(i=1,2,3\dots)$ という事である。

$E[\varepsilon_i]=0$

Least Squares Fitting.. (最小二乗法)

$$\arg\min_{\beta_0,\beta_1}\underbrace{\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2}_{F(\beta_0,\beta_1)}$$



(これの解は $\hat\beta_0,\hat\beta_1$ と書く)

$$\underbrace{\frac{\partial F(\beta_0,\beta_1)}{\partial\beta_0}=0,\quad\frac{\partial F(\beta_0,\beta_1)}{\partial\beta_1}=0}$$

これを計算すると
$$\begin{cases}-2\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)=0\\-2\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)(-X_i)=0\end{cases}$$

となり
$$\begin{pmatrix}\beta_0=\bar{Y}-\beta_1\bar{X}\\ \sum X_iY_i-\beta_0\sum_{i=1}^{n}X_i-\beta_1\sum_{i=1}^{n}X_i^2=0\end{pmatrix}$$

注 $S_{xx}=\sum(X_i-\bar X)^2$

$S_{xy}=\sum(X_i-\bar X)(Y_i-\bar Y)$

$\hat\beta_1=\frac{S_{xy}}{S_{xx}}$, $\hat\beta_0=\bar Y-\frac{S_{xy}}{S_{xx}}\bar X$

が得られる。

Chryv culture

$E[\varepsilon_i]=0$

$V[\varepsilon_i]=\sigma^2$

Gauss-Markov Conditions

$E[\varepsilon_i\varepsilon_j]=0 \quad (i\neq j)$

No.
Date 9 : 12 ⑩室

このとき $\hat\beta_0, \hat\beta_1$ を最小二乗推定量 (Least Square Estimator) と言う

② $E[\hat\beta_0] = E[\bar Y - \hat\beta_1\bar X] = \beta_0+\beta_1\bar X - \beta_1\bar x = \beta_0$

① $E[\hat\beta_1] = E\left[\frac{S_{xy}}{S_{xx}}\right] = \frac{\sum_{i=1}^n x_i(\beta_0+\beta_1 x_i) - \bar X\sum_{i=1}^n(\beta_0+\beta_1 x_i)}{S_{xx}} = \frac{\beta_1\left\{\sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}\right\}}{S_{xx}}$

$=\beta_1$

一方分散に関して、$V[\hat\beta_1] = V\left[\frac{S_{xy}}{S_{xx}}\right] = V\left[\sum_{i=1}^n \frac{(x_i-\bar x)Y_i}{S_{xx}}\right]$

いま $Y_1\sim Y_n$ は互いに独立であるので $\sum_{i=1}^n V\left[\frac{(x_i-\bar x)Y_i}{S_{xx}}\right]$

$(\because \varepsilon_i : \text{iid})$

$\sum_{i=1}^n \frac{(x_i-\bar x)^2}{(S_{xx})^2}\underbrace{V[Y_i]}_{\sigma^2} = \frac{1}{(S_{xx})^2}\sum_{i=1}^n \sigma^2(x_i-\bar x)^2$

$=\frac{\sigma^2}{S_{xx}}$ となる

$V[\hat\beta_0] = Var[\bar Y - \hat\beta_1\bar X] = \underbrace{Var[\bar Y]}_{\frac{\sigma^2}{n}} + \underbrace{\bar X^2 Var[\hat\beta_1]}_{\frac{\sigma^2}{S_{xx}}\bar X^2} - \underbrace{2\bar X\cdot cov[\bar Y,\hat\beta_1]}_{0}$

$\left(\frac{1}{n}+\frac{\bar X^2}{S_{xx}}\right)\sigma^2$

以下、個人的なメモ

$$\underbrace{\begin{pmatrix} Y_1\\ \vdots \\ Y_n\end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} 1 & x_1\\ \vdots & \\ 1 & x_n\end{pmatrix}}_{X}\underbrace{\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}}_{\beta} + \begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{pmatrix}$$

$L=(Y-X\beta)^t(Y-X\beta) = (Y^t-\beta^tX^t)(Y-X\beta) = Y^tY - \beta^tX^tY - Y^tX\beta + \beta^tX^tX\beta$

Chryu culture

No.

Date 9 12 (10)

residuals … $e_i = Y_i - \hat{Y}_i \quad (i = 1 \sim n)$

残差

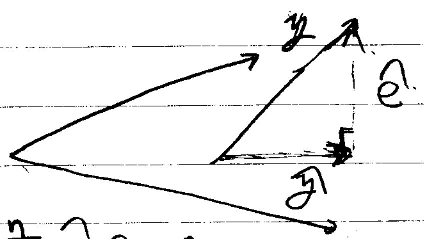
- 真のmodel …… $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- 仮定のmodel … $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

Properties (残差の性質)

1. $\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(\underbrace{Y_i - \bar{Y}}_{\to 0} - \hat{\beta}_1 \underbrace{(X_i - \bar{X})}_{\to 0} \right)$
2. $\sum_{i=1}^{n} \hat{Y}_i = \sum_{i=1}^{n} Y_i$ (ここでの Y_i は観測地データと助かり)
3. LS regression line passes $(\bar{X}, \bar{Y})$
4. $\sum_{i=1}^{n} X_i e_i = 0$

ベクトル空間で考えると、正射影として

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$



5. $\sum_{i=1}^{n} \hat{y}_i e_i = 0$

この辺りも復習しておく。

Chuyu culture

9/19.~

- Simple Linear Regression.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (i = 1 \sim n)$$

Gauss Markov Conditions
$$\begin{cases} E[\varepsilon_i] = 0 & (i = 1 \sim n) \\ V[\varepsilon_i] = \sigma^2 \\ E[\varepsilon_i \varepsilon_j] = 0 & (i \neq j) \end{cases}$$

- Least Square Estimators. (LSE)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$E[\hat{\beta}_0] = \beta_0, \quad E[\hat{\beta}_1] = \beta_1$$

$$Var[\hat{\beta}_0] = \frac{\sigma^2}{S_{xx}}, \quad Var[\hat{\beta}_1] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

- Residuals (残差)

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad (i = 1 \sim n)$$

いわゆ先回,内容

- Estimation of σ^2. (σ^2の推定)

$\sigma^2 = E[\varepsilon_i^2]$ であるが ε_iは直接に観測されない。

$$SS_{res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
(または SS_{error})

No.
Date 9/19

$E[SS_{Res}] = (n-2)\sigma^2$ である。

$\hat{\sigma}^2 = \frac{1}{n-2} SS_{Res} = MS_{Res}$ (Residual Mean Square)

$$SS_{Res} = \sum_{i=1}^{n} (Y_i - \bar{Y} + \hat{\beta}_1 \bar{X} - \hat{\beta}_1 X_i)^2$$

$$= \sum_{i=1}^{n} \{(Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X})\}^2$$

$$= \sum_{i=1}^{n} \{(Y_i - \bar{Y})^2 - 2\hat{\beta}_1 (X_i - \bar{X})(Y_i - \bar{Y}) + \hat{\beta}_1^2 (X_i - \bar{X})^2\}$$

$$= S_{yy} - \frac{2 S_{xy}^2}{S_{xx}} + \left(\frac{S_{xy}}{S_{xx}}\right)^2 \cdot S_{xx}$$

$$= \underbrace{\underset{\substack{\shortparallel \\ SST}}{S_{yy}} - \frac{S_{xy}^2}{S_{xx}}}$$ となる。

・Analysis of Variance … (分散分析)

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} ((Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}))^2$$

$$= \sum_{i=1}^{n} \{\underbrace{(Y_i - \hat{Y}_i)^2}_{SS_{Res}} + \underbrace{2(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})}_{e_i \cdot (\hat{Y}_i - \bar{Y}) = 0} + \underbrace{(\hat{Y}_i - \bar{Y})^2}_{SSR}\}$$

ここは先週に登場した数式で説明がつく。

$$= \underbrace{SS_{Res}}_{\text{モデルによって説明できないところ}} + \underbrace{SSR}_{\text{モデルが説明できるところ}}$$

No.
Date 9/19.

全分散の公式

- $V[Y] = E[V[Y|X]] + V[E[X|Y]]$

先ほどの式に戻って $SS_{RES} = SS_T - \hat{\beta}_1 \cdot S_{xy} = SS_T - \underbrace{\frac{S_{xy}^2}{S_{xx}}}_{SS_R}$

これが回帰モデルによる変動に一致することはすぐに分かる

- Degrees of Freedom (自由度) → 先ほど $E[SS_{RES}] = \sigma^2(n-2)$ から

$$\underbrace{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}_{\text{自由度}(n-1)} = \underbrace{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}_{\text{自由度}(n-2)} + \underbrace{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}_{\text{自由度}1}$$

- Mean Squares.

$$\begin{cases} MS_R = \dfrac{SS_R}{1} \\ MS_{RES} = \dfrac{SS_{RES}}{n-2} \end{cases}$$

- ANOVA Table....

	SS	df	MS	統計量 F
Source of Variation	Sum of Square	Degree of Freedom	Mean Square	
Regression (回帰)	SS_R	1	MS_R	$\dfrac{MS_R}{MS_{RES}}$
Residual (残差)	SS_{RES}	$n-2$	MS_{RES}	
計	SS_T	$n-1$		

No.
Date 9:19

• H_0 (帰無仮説) vs H_1 (対立仮説) に対して

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

$$E[MS_{Res}] = \sigma^2, \quad E[MS_R] = \begin{cases} \sigma^2 & \text{if } (\beta_1 = 0) \\ \sigma^2 + \beta_1^2 S_{xx} & \text{if } (\beta_1 \neq 0) \end{cases}$$

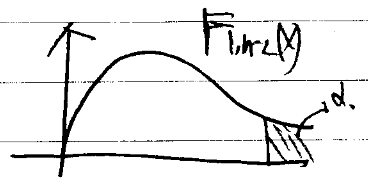
$$F_0 = \frac{MS_R}{MS_{Res}}$$ (もし $\beta_1=0$であれば F_0は1になるが $\beta_1 \neq 0$であれば F_0はそれより大きくなる。と考えられる。)

$$= t_0^2$$

$$t_0 = \frac{\hat{\beta}_1 - 0}{\sqrt{\left(\frac{\hat{\sigma}^2}{S_{xx}}\right)}} \qquad \left(\hat{\sigma}^2 = \frac{SS_{Res}}{n-2}\right)$$

• Normal Assumption + (Gauss-Markov Condition)

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \sim N(0, \sigma^2)$



$F_0 \sim F_{1,n-2}$ under H_0

Reject $H_0 : \beta_0 = 0$ when $F_0 > F_{1,n-2}(1-\alpha)$

• Confidence Interval (信頼区間)

$\hat{\beta}_1$ …… $\dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}$ に従う

Chryv culture

~~β_1の100(1-α)%~~

$\hat{\beta}_1 \pm t_{n-2}\left(\frac{\alpha}{2}\right) \cdot \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$ を信頼区間と呼ぶ。

次に σ^2 の信頼区間 $\hat{\sigma}^2 = MS_{RES} \sim \sigma^2 \cdot \chi_{n-2} / (n-2)$

$$1-\alpha = Pr\left(\chi^2_{n-2, \frac{\alpha}{2}} < \frac{\hat{\sigma}^2(n-2)}{\sigma^2} < \chi_{n-2, 1-\frac{\alpha}{2}}\right)$$

$$= Pr\left(\frac{\hat{\sigma}^2(n-2)}{\chi^2_{n-2, 1-\frac{\alpha}{2}}} < \sigma^2 < \frac{\hat{\sigma}^2(n-2)}{\chi^2_{n-2, \frac{\alpha}{2}}}\right)$$

これが信頼区間となる。

- Coefficient of Determination : R^2 (決定係数)

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR+SS_{RES}} \quad (0\sim1\text{の値をとる。})$$

$$= 1 - \frac{SS_{RES}}{SSR + SS_{RES}}$$

- 次に $\beta_0 + \beta_1 x_0$ の信頼区間について

$$\frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}} \sim t_{n-2}$$

$$Var(\hat{\beta}_0 + \hat{\beta}_1 x_0) = Var[\bar{Y} + \hat{\beta}_1(x_0 - \bar{x})]$$

$$= Var[\bar{Y}] + (x_0 - \bar{x})^2 Var(\hat{\beta}_1) + 0$$

No.
Date :

予測区間…

- Prediction Interval for a new observation

$$Y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0 \quad \varepsilon_0 \sim N(0, \sigma^2)$$

$$\frac{\hat{\beta}_0 + \hat{\beta}_1 x_0 - Y_0}{\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}} \sim t_{n-2} \text{ となる.}$$

$$\mathrm{Var}[\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0 - \varepsilon_0]$$
$$= \mathrm{Var}[\bar{Y}] + (x_0 - \bar{x})^2 \mathrm{Var}(\hat{\beta}_1) + \underbrace{\mathrm{Var}(\varepsilon_0)}_{\text{この部分は増加する}}$$

- Regression through the origin (原点を通る回帰モデル)
(つまり $\beta_0 = 0$ のモデル)

「$Y = \beta_1 X + \varepsilon_i$」

LSE for β_1 … $\displaystyle\min_{\beta_1} \sum_{i=1}^{n} (Y_i - \beta_1 x_i)^2$
(最小二乗推定量)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} \qquad E[\hat{\beta}_1] = \beta_1 \qquad \mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} \leq \frac{\sigma^2}{S_{xx}}$$

$$\hat{\sigma}^2 = MS_{Res} = \frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 x_i)^2}{n-1}$$ → 10/14, X-上 記載 [illegible]

$$\hat{Y}_i = \hat{\beta}_1 x_i$$

Intercept β_0 ないの model

No.
Date

$$R^2 = \frac{\sum \hat{Y_i}^2}{\sum Y_i^2}$$

- Random Design ... (X is random)

「$Y = \beta_0 + \beta_1 X + \varepsilon$」

$E[Y|X] = \beta_0 + \beta_1 x$, $E[\varepsilon|X] = 0$

$(X, Y)^t \sim$ 2次元正規分布... $N\left(\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$

$Y|X=x \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right)$

(注)

$E[Y|X] = E(Y) + \mathrm{cov}[Y, X]\, \Sigma_{X}^{-1} (x - E(X)) =$

$V(Y|X) = V(Y) - \mathrm{cov}(Y,X) \cdot \Sigma_X^{-1} \mathrm{cov}(Y,X)$

MLE = LSE.

Sample correlation coefficient.

$$r = \frac{S_{xy}}{(S_{xx} \cdot S_{yy})^{\frac{1}{2}}} = \frac{S_{xy}}{(S_{xx} SS_T)^{\frac{1}{2}}}$$

$H_0: \rho = 0$. $H_1: \rho \neq 0$

$$\hat{\beta_1} = \left(\frac{SS_T}{S_{xx}}\right)^{\frac{1}{2}} \cdot r$$

Chryu culture

No. 9
Date : :

Ⓐ
Chapter 2. Homework

…… ~~4~~, ~~7~~, 10, 12, 17, 25, 26, 27, 32, 33.

10/14までに提出。

Chapter 3. Multiple Linear Regression (9/26~

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \varepsilon$

$E[\varepsilon] = 0$

- $X_1, X_2, \ldots X_k$: Covariates
- Y : Response Variable
- $\beta_0, \beta_1 \ldots \beta_k$: Unknown Parameters

β_1… $X_2 \sim X_k$ を固定した状況下で X_1 の値が1増加したときに変動する量を表している。

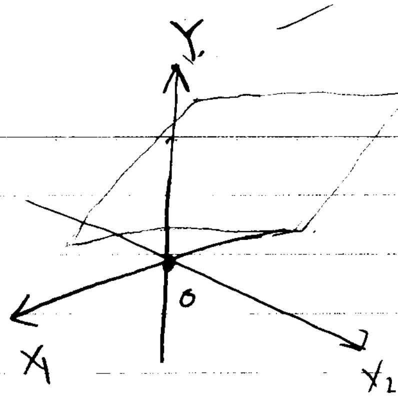
例、多項式回帰（これも重回帰と言える。）

X…

$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots \beta_k X^k + \varepsilon$

No
Date 9/26

Interactions.



(平面)

- 交互作用のない2変量

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- 交互作用を考慮したmodel

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \varepsilon$$

- データ (Data)

$$(Y_i, X_{i1}, X_{i2}, \ldots X_{ik}) \quad i = 1 \sim n$$

model… $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots \beta_k X_{ik} + \varepsilon_i \quad (i = 1 \sim n)$

Gauss-Markov-Conditions $E[\varepsilon_i] = 0, \quad V[\varepsilon_i] = \sigma^2 \quad (i = 1 \sim n)$

$$\mathrm{Cov}[\varepsilon_i, \varepsilon_j] = 0 \quad (i \neq j)$$

- 最小二乗法による推定

$$S(\beta_0, \beta_1, \ldots \beta_k) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2$$

$\underset{\begin{pmatrix}\beta_0 \\ \vdots \\ \beta_k\end{pmatrix}}{\arg\min}\ S(\beta_0, \beta_1, \ldots \beta_k)$ を求める

No.
Date 9/26

$$\begin{pmatrix} \frac{\partial}{\partial \beta_0} S(\beta_0, \dots, \beta_k) \\ \frac{\partial}{\partial \beta_1} S(\beta_0, \dots, \beta_k) \\ \vdots \\ \frac{\partial}{\partial \beta_k} S(\beta_0, \dots, \beta_k) \end{pmatrix} = \frac{\partial S}{\partial \boldsymbol{\beta}} = 0$$ と計算しても同じ

$$\frac{\partial S}{\partial \beta_j} = \sum_{i=1}^{n} 2(Y_i - \beta_0 - \beta_1 X_{i1} \cdots - \beta_k X_{ik}) \cdot (-X_{ij}) = 0 \quad (j \geq 1)$$

$(j = 1 \sim k)$

しかしこのまま計算すると少しめんどうなので工夫する

モデルの行列表示を考える

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & & & \\ 1 & X_{n1} & \cdots & X_{nk} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}}_{\boldsymbol{\beta}} + \varepsilon$$

これを利用して $(Y - X\beta)^t (Y - X\beta) = S(\beta)$

となるので $\frac{\partial S}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} (Y^t Y - \beta^t X^t Y - Y^t X\beta + \beta^t X^t X\beta)$

$$\frac{\partial}{\partial \beta}\left({}^t\beta\, {}^tX\, Y\right) = \begin{pmatrix} \frac{\partial}{\partial \beta_1} {}^t\beta\, {}^tX\, Y \\ \vdots \\ \frac{\partial}{\partial \beta} {}^t\beta\, {}^tX\, Y \end{pmatrix} = \begin{pmatrix} {}^tXY\text{の1列目} \\ \vdots \\ {}^tXY\text{のn列目} \end{pmatrix} = {}^tXY$$

$$\frac{\partial}{\partial \beta}\left({}^t\beta A \beta\right) = \frac{\partial}{\partial \beta} \sum_{i=1}^{k}\sum_{j=1}^{k} \beta_i \beta_j \cdot a_{ij}$$ であるから

$=$

$$\frac{\partial S}{\partial \beta} = -2\,{}^tXY + 2\,{}^tXX\beta = 0$$

$$\therefore \beta = ({}^tXX)^{-1}\,{}^tXY \Rightarrow \text{これは}\hat{\beta}\text{とする}$$

残差に関して $e = Y - X\hat{\beta} = Y - X({}^tXX)^{-1}\,{}^tXY = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$

回帰に関して $\hat{y} = X\hat{\beta} = \underbrace{X({}^tXX)^{-1}\,{}^tX}_{\text{これはHと書かれる。}}Y$

No. 9/26
Date

$$X = (1\ x_1\ \cdots\ x_k)$$

$$1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \uparrow n \downarrow \qquad x_j = \begin{pmatrix} x_{1,j} \\ \vdots \\ x_{n,j} \end{pmatrix}$$ $\boxed{\text{各データ1〜n の } x_j \text{ の値}}$ と見なせる

$$X\beta = (1_n\ x_1\ \cdots\ x_k)\begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$$

$= \beta_0 1 + \beta_1 x_1 + \cdots \beta_k x_k$ である。

$\hat{y} = X\hat{\beta} = H\underline{y}$ 観測値

$$H = X(X^t X)^{-1} X^t$$

$$HH = X(X^t X)^{-1} \underbrace{X^t X (X^t X)^{-1}}_{I} X^t = X(X^t X)^{-1} X^t = H$$

$\therefore H^2 = H$ である。

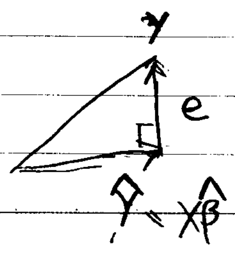
$$e = y - \hat{y} = (I - H)y$$

・ $X^t e = X^t(I - X(X^t X)^{-1} X^t)y$

$= (X^t - X^t)y = 0y = 0$ となる

(図) 正射影と思う

$X^t e = 0$



Chuyu culture

No.
Date 9/26

$$\hat{\beta} = (X^tX)^{-1}X^tY$$

$$E[\hat{\beta}] = \begin{pmatrix} E[\hat{\beta}_0] \\ \vdots \\ E[\hat{\beta}_k] \end{pmatrix}$$ の意味 $= E[(X^tX)^{-1}X^tY]$

$$= (X^tX)^{-1}X^tE[Y] = (X^tX)^{-1}X^tE[X\beta+\varepsilon] = (X^tX)^{-1}X^tX\beta = \beta$$ となる

$$V[\hat{\beta}] = E[(\hat{\beta}-\underset{\substack{\parallel \\ E[\hat{\beta}]}}{\beta})(\hat{\beta}-\beta)^t] =$$

($Cov[\hat{\beta}]$)
分散、共分散
行列

$\underline{k+1} \times \underline{k+1}$ の行列

$$\hat{\beta}-\beta = (X^tX)^{-1}X^tY - \beta = (X^tX)^{-1}X^t(Y-X\beta)$$

$$(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t = (X^tX)^{-1}X^t(Y-X\beta)(Y-X\beta)^t((X^tX)^{-1}X^t)^t$$

$$\therefore E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t] = (X^tX)^{-1}X^t\underbrace{E[(Y-X\beta)(Y-X\beta)^t]}_{\substack{\parallel \\ \sigma^2 I}}((X^tX)^{-1}X^t)^t$$

$$= \sigma^2 \cdot (X^tX)^{-1}X^t((X^tX)^{-1}X^t)^t$$

$$= \sigma^2 \cdot (X^tX)^{-1}X^tX((X^tX)^{-1})^t$$

$$= \sigma^2 \cdot ((X^tX)^{-1})^t = \sigma^2 \cdot (X^tX)^{-1}$$

culture

No.
Date 9.26.

~~SS_{Error}~~ (3節)

$\|$

- $SS_{Residual}$ ..... $\sum_{i=1}^{n} e_i^2 = e^t e$ である。

$$\left((I-H)y\right)^t (I-H)y = y^t (I-H)^t (I-H) y$$

$$= y^t (I - H^t I - IH + H^t H) \qquad H^t = H,\ H^2 = H$$

$$= y^t (I-H) y$$

$$= (X\beta + \epsilon)^t (I-H)(X\beta + \epsilon) \qquad (X\beta \text{の部分に注目すると消える})$$

$$= \varepsilon^t (I-H) \varepsilon$$

$$E[\varepsilon^t \underbrace{(I-H)}_{M} \varepsilon] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}\, \varepsilon_i \varepsilon_j\right] = \sigma^2 \cdot tr(M)$$

$$= \sigma^2 tr(I-H)$$

$$= \sigma^2 (\underbrace{tr I}_{n} - tr H)$$

$n \times (k+1)$ $\quad (k+1) \times n$

$$trH = tr[\underbrace{X}_{A} \underbrace{(X^t X)^{-1} X^t}_{B}] = tr(AB) = tr(BA)$$

$$= tr(I_{k+1}) = k+1.$$

$$= \sigma^2 (n - \underbrace{p}_{k+1})$$

- SS_{Total} .. $\sum_{i=1}^{n} (y_i - \bar{y})^2 = \left(y - \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\bar{y}\right)^t \left(y - \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\bar{y}\right)$

スカラー値 ←

$$= (y - \bar{y}\mathbf{1})^t (y - \mathbf{1}\bar{y})$$

$$= (\hat{y} + e - \bar{y}\mathbf{1})^t (\hat{y} + e - \bar{y}\mathbf{1})$$

ベクトルで考えると直交することから

$$= \underbrace{(\hat{y} - \bar{y}\mathbf{1})^t (\hat{y} - \bar{y}\mathbf{1})}_{SS_{Reg}} + \underbrace{(\hat{y} - \bar{y}\mathbf{1})^t e + e^t (\hat{y} - \bar{y}\mathbf{1})}_{0} + \underbrace{e^t e}_{SS_E}$$

・R^2 … $\frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$ $(0 \leqq R^2 \leqq 1)$」

・次に $\begin{pmatrix} Y \\ X_1 \\ \vdots \\ X_k \end{pmatrix}$ の分布（※今回 $X_1 \sim X_k$ は確率変数と見なす）

$$\sim N\left(\underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}}_{\mu}, \underbrace{\begin{pmatrix} \sigma_1^2 & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}}_{\Sigma} \right)$$

同時確率密度関数

$$\frac{1}{\sqrt{2\pi}^{k+1} \cdot |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left((y\ x_1 \cdots x_k) - \underbrace{(\mu_1\ \mu_2^t)}_{\mu^t} \right) \Sigma^{-1} \left\{ \begin{pmatrix} y \\ x_1 \\ \vdots \\ x_k \end{pmatrix} - \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}}_{\mu} \right\} \right)$$

・次に $Y \mid X_1, X_2 \cdots X_k$ の条件付分布について考える

・平均 : $\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X - \mu_2)$

・分散 : $\sigma_1^2 - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$

の正規分布 $N\left(\underbrace{\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X - \mu_2)}_{= X\beta}, \underbrace{\sigma_1^2 - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}_{= \sigma^2} \right)$

・MLEを求めると＝LSEに一致

（要証明）

No.
Date 10/14 (金) (前までの復習を兼ねて)

- 重回帰分析 $\underbrace{Y}_{n\times 1} = \underbrace{X}_{n\times p}\underbrace{\beta}_{p\times 1} + \varepsilon.$

- LSE : $\hat{\beta} = (X^tX)^{-1}X^tY$

$$E[\hat{\beta}] = \beta \qquad V[\hat{\beta}] = (X^tX)^{-1}\sigma^2$$

- $\hat{Y} = X\hat{\beta} = \underbrace{X(X^tX)^{-1}X^t}_{H}Y$

$$E[\hat{Y}] = X\beta = E[Y]$$

$$V[\hat{Y}] = X\,\mathrm{Cov}[\hat{\beta}]\,X^t = \sigma^2 X(X^tX)^{-1}X^t = \sigma^2 H$$

- $e = Y - \hat{Y} = (I-H)Y$

$V[e] = \sigma^2(I-H)$

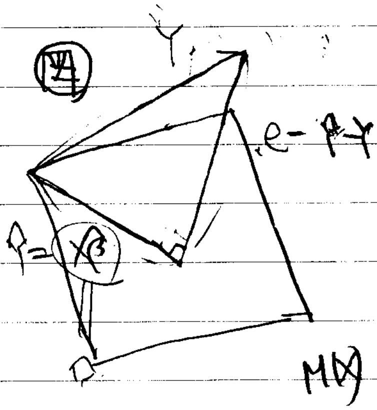
$\left(V[e_i] = \sigma^2(1-H_{ii})\right)$

$SS_{Res} = e^te$ (残差平方和) $\left(\hat{\sigma}^2 = \frac{1}{n-p}e^te\right)$

$$\underbrace{\sum_{i=1}^{n} Y_i^2}_{\text{Total}} = \underbrace{\sum_{i=1}^{n}\hat{Y}_i^2}_{\text{predicted}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\text{residual}}$$

$$Y^tY = \hat{Y}^t\hat{Y} + e^te$$

図



Chuyu culture

No. Date 10/14

Xの計画行列に $\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$ の vector 項が入っている

- Intercept β_0 in the model. (β_0がモデルに入っているとき)

$\sum_{i=1}^{n} e_i = 0$ つまり $\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^t \begin{pmatrix}e_1\\ \vdots \\ e_n\end{pmatrix} = e_1 + \cdots + e_n = 0$

$$\underbrace{\sum_{i=1}^{n} e_i^2}_{SS_E} = \sum_{i=1}^{n} Y_i^2 - \sum_{i=1}^{n} \hat{Y}_i^2 = \underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{SS_T} - \underbrace{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}_{SS_{reg}}$$

$$R^2(\text{決定係数}) = \frac{SS_{Residual}}{SS_T} = 1 - \frac{SS_{reg}}{SS_T}$$

また

$$\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n}(\underbrace{Y_i - \hat{Y}_i}_{e_i} + \hat{Y}_i - \bar{Y})(\hat{Y}_i - \bar{Y}) =$$

$$= \underbrace{\sum_{i=1}^{n} e_i \hat{Y}_i}_{\substack{\parallel \\ 0 \\ (e^t \cdot \hat{Y})}} - \bar{Y}\underbrace{\sum_{i=1}^{n} e_i}_{\substack{\parallel \\ 0 \\ (1_n^t e = 0)}} + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \text{ である}$$

$= SS_{reg}$

またこのことから $R^2 = \dfrac{\left(\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}$ と書けるので、

$= SS_{reg}$

∴ R = Yと$\hat{Y}$の標本相関係数と言える。

(つまり Rが1に近いほど、$\hat{Y}$がYに近いので良いモデルと考えられる)

Chuo culture

No.
Date 10/14

ANOVA Table :

$Y^t(H-\frac{J}{n})Y$

Source	自由度	SS
回帰	$p-1$ (K)	$\hat{\beta}^t X^t Y - n\bar{Y}^2$
残差	$n-p$	$Y^tY - \hat{\beta}^t X^t Y$
Total	$n-1$	$Y^tY - n\bar{Y}^2$

$= SS_{reg} = Y^t H Y - n\bar{Y}^2$

- Intercept β_0 not in the model.

(β_0がモデルに入っていないとき)

ANOVA Table :

why

Source	自由度	SS
回帰	k	$\hat{Y}^t\hat{Y} = \hat{\beta}^t X^t Y$
残差	$n-k$	e^te

SS_T
$=$
Y^tY とかになる
$=$
ヨとOにおきかわっている

- Best Linear Unbiased Estimator (BLUE)

$\underset{q\times n\ n\times 1}{AY}$ が $\underset{q\times p\ p\times 1}{L\beta}$ の BLUE (最良線形不偏推定量) で

あるとは $E[AY] = L\beta$ (for all β) が成り立ち

分散共分散行列の差 $V[CY] - V[AY]$ が

任意の $L\beta$ の不偏推定量 CY に対し

positive semi-definite になることをいう
(半正定値行列)

Chuyu culture

ガウス・マルコフ・モデル

$Y = X\beta + \varepsilon \quad (E[\varepsilon] = 0)$

定義 estimable
$l^t\beta$ が推定可能であるとは $l \in \mathrm{span}[x_1, \cdots, x_n]$ と同値である
$(X = [x_1 \cdots x_n])$

※

Gauss-Markov Theorem

ガウスマルコフの条件下で $l^t\hat{\beta}$ は推定可能な関数 $l^t\beta$ の BLUE になる。

証明

$l^t\beta$ 推定可能 $\Leftrightarrow \overset{n \times 1}{\exists c}, \; l^t\beta = E[c^tY] = c^tX\beta$ (for all β)

($l^t \to L$ にしても同じ) $\Leftrightarrow \exists c, \; l^t = c^tX$

ここで c^tY を任意の線形不偏推定量 ($l^t\beta$ の) とする。

$c^tX = l^t$ から $V[c^tY] - V[l^t\hat{\beta}] = \sigma^2 c^t I c - \sigma^2 l^t (X^tX)^{-1} l$

$= \sigma^2 c^t (I - X(X^tX)^{-1}X^t) c$

$= \sigma^2 c^t (I - H) c$

$= \mathrm{Var}[c^te] \geq 0$

Chryo culture

No.
Date 10/14

- Centered Model: $\underline{k+1=p}$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & \cdots & X_{1,k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n,1} & \cdots & X_{n,k} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} + \varepsilon$$

$$= \beta_0 1_{n\times 1} + \begin{pmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & & \vdots \\ x_{1,n} & & x_{n,k} \end{pmatrix} \beta_{(0)} + \varepsilon \qquad \beta_{(0)} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$= \beta_0 1_{n\times 1} + \underbrace{\begin{pmatrix} \sum_{j=1}^{k} \beta_j \bar{x}_j \\ \vdots \\ \sum_{j=1}^{k} \beta_j \bar{x}_j \end{pmatrix}}_{\sum_{j=1}^{k} \beta_j \bar{x}_j \cdot 1_{n\times 1}} + \underbrace{\begin{pmatrix} x_{1,1}-\bar{x}_1 & \cdots & x_{1,k}-\bar{x}_k \\ \vdots & & \vdots \\ x_{n,1}-\bar{x}_1 & & x_{n,k}-\bar{x}_k \end{pmatrix}}_{= Z} \beta_{(0)} + \varepsilon$$

$\gamma_0 1_{n\times 1}$

$= \gamma_0 1_{n\times 1} + Z\beta_{(0)} + \varepsilon$ と書ける

$= \underbrace{(1 \ Z)}_{\text{新しい計画行列}} \begin{pmatrix} \gamma_0 \\ \beta_{(0)} \end{pmatrix} + \varepsilon$ と書ける

最小二乗推定量 $\begin{pmatrix} \hat{\gamma}_0 \\ \hat{\beta}_{(0)} \end{pmatrix} = \left\{ (1 \ Z)^t (1 \ Z) \right\}^{-1} (1 \ Z)^t Y$

$$= \left\{ \begin{pmatrix} 1^t \\ Z^t \end{pmatrix} (1 \ Z) \right\}^{-1} (1 \ Z)^t Y$$

Chuyu culture

No.
Date 10/14

$$= \begin{pmatrix} 1^t 1 & 1^t Z \\ Z^t 1 & Z^t Z \end{pmatrix}^{-1} (1\ Z)^t Y \qquad 注\ 1^t Z = 0 \ (Z^t 1)$$

$$= \begin{pmatrix} n & 0 \\ 0 & Z^t Z \end{pmatrix}^{-1} \begin{pmatrix} 1^t \\ Z^t \end{pmatrix} Y = \begin{pmatrix} n & 0 \\ 0 & Z^t Z \end{pmatrix}^{-1} \begin{pmatrix} n\bar{Y} \\ Z^t Y \end{pmatrix}$$

$$= \begin{pmatrix} n^{-1} & 0 \\ 0 & (Z^t Z)^{-1} \end{pmatrix} \begin{pmatrix} n\bar{Y} \\ Z^t Y \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ (Z^t Z)^{-1} Z^t Y \end{pmatrix} \text{ と書ける}$$

$$\hat{Y} = (1\ Z) \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_{(0)} \end{pmatrix} = X\hat{\beta} \text{ になるはずである}$$

（元々は同じもの）

$$e = Y - \hat{Y}$$

$$e^t e = Y^t Y - n\bar{Y}^2 - Y^t Z (Z^t Z)^{-1} Z^t Y \text{ である}$$

$$R^2 = \frac{Y^t Z (Z^t Z)^{-1} Z^t Y}{Y^t Y - n\bar{Y}^2} = \frac{(Y - \bar{Y}1) Z (Z^t Z)^{-1} Z^t Y}{(Y - \bar{Y}1)^t (Y - \bar{Y}1)}$$

= sample multiple correlation between Y and $X_1 \sim X_k$

$$V\begin{bmatrix} Y \\ X_1 \\ X_k \end{bmatrix} = \begin{pmatrix} \sigma_Y^2 & \sigma_{YX} \\ \sigma_{YX} & \Sigma_{XX} \end{pmatrix}$$

multiple-correlation between Y and $\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ is

$$\frac{(\sigma_{YX}^t \Sigma_{XX}^{-1} \sigma_{YX})^{\frac{1}{2}}}{\sigma_Y}$$

10/14

（最後は高速に進んだので要確認）

No.
Date 10/17

Centered and Rescaled Model

$$Z_{(S)} = Z \cdot \underbrace{\begin{pmatrix} S_{11}^{-\frac{1}{2}} & & O \\ & S_{22}^{-\frac{1}{2}} & \\ & & \ddots \\ O & & & S_{kk}^{-\frac{1}{2}} \end{pmatrix}}_{D_{(0S)}^{-1}} = \left(\frac{X_{ij} - \bar{X}_j}{\sqrt{S_{jj}}}\right)_{ij}$$

ただし $S_{jj} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2$ である

$$\delta = \underbrace{\mathrm{diag}(S_{11}^{\frac{1}{2}}, \ldots, S_{kk}^{\frac{1}{2}})}_{D_{(0S)}} \beta_{(0)}$$

$$Y = \gamma_0 \cdot \mathbf{1} + \underbrace{Z D_{(0S)}^{-1}}_{Z_{(S)}} \underbrace{D_{(0S)} \beta_{(0)}}_{\delta} + \varepsilon$$

$$= (\mathbf{1} \; Z_{(S)}) \begin{pmatrix} \gamma_0 \\ \delta \end{pmatrix} + \varepsilon$$

と書き換えられ、δの最小2乗推定量 LSE $\hat{\delta}$

$$\hat{\delta} = (Z_{(S)}^t Z_{(S)})^{-1} Z_{(S)}^t Y \text{ である。}$$

$$\mathrm{cov}(\hat{\delta}) = \sigma^2 (Z_{(S)}^t Z_{(S)})^{-1}$$

$Z_{(S)}^t Z_{(S)}$: Sample Correlation Matrix of $X_1 \sim X_k$

⊗ Constrained Least Squares.

$\underset{m\times p}{C}\underset{p\times 1}{\beta}=\underset{m\times 1}{d}$ という制約付のパラメータ

$(m<p)$ $\quad \mathrm{rank}\, C = m$

問 $\displaystyle \min_{\substack{\beta \\ C\beta=d}} (Y-X\beta)^t(Y-X\beta)$

制約条件付の最小二乗推定量ということだろうか？
(そうであればラグランジュの未定乗数法を用いて解く)

ラグランジュの未定乗数法の $\lambda = \begin{pmatrix}\lambda_1\\ \vdots \\ \lambda_m\end{pmatrix}$ とし

$$\frac{\partial}{\partial\beta}\left\{(Y-X\beta)^t(Y-X\beta)+\lambda^t(d-C\beta)\right\}=0$$

$\rightarrow\ 2(X^tX)\beta-2X^tY-C^t\lambda=0$ を得る。

また $d-C\beta=0$ (制約条件)

(板書では ~~Xx~~ λ と書かれていたが $d-C\beta=0$ のみ使っていいだろう。)

$$C\tilde{\beta}=d$$

$$\tilde{\beta}=\underbrace{(X^tX)^{-1}X^tY}_{\hat{\beta}}+\frac{1}{2}(X^tX)^{-1}C^t\tilde{\lambda}$$

No.
Date 10/17

$$d = C\tilde{\beta} = C\hat{\beta} + \frac{1}{2} C (X^t X)^{-1} C^t \tilde{\lambda}$$

$$\tilde{\beta} = \hat{\beta} + (X^t X)^{-1} C^t \left(C (X^t X)^{-1} X^t \right)^{-1} (d - C\hat{\beta})$$

Note.
- $\tilde{\beta}$ は $C\tilde{\beta} = d$ を満足する
- もし $C\beta = d$ であれば $E[\tilde{\beta}] = \beta + 0 = \beta$
- $V[\tilde{\beta}] = V\left[\left(I - (X^t X)^{-1} C^t (C (X^t X)^{-1} C^t)^{-1} C \right) \hat{\beta} \right]$
  $= \sigma^2 (X^t X)^{-1} - \sigma^2 (X^t X)^{-1} C^t \left[C (X^t X)^{-1} C^t \right]^{-1} C (X^t X)^{-1}$
- $V[\hat{\beta}] - V[\tilde{\beta}] = \sigma^2 (X^t X)^{-1} C^t \left[C (X^t X)^{-1} C^t \right]^{-1} C (X^t X)^{-1} \geqq 0$

である

Hypothesis Testing

(I)
- $H_0 : \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = 0$
- $H_1 : {}^{\exists}\beta_j \neq 0 \quad (j = 1, \dots, k)$

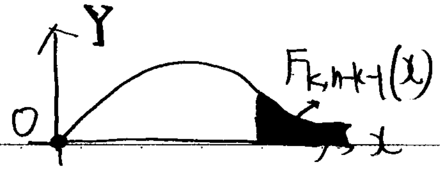
(II)
- $H_0 : C\beta = d$
- $H_1 : C\beta \neq d$

$(k+1 = p)$

ANOVA:

Source	df	SS	MS	F
Regression	k	SS_{Reg}	SS_{Reg}/k	$\dfrac{(SS_R/k)}{SS_{residual}/(n-k-1)}$
Residual	$n-k-1$	$SS_{Residual}$	$SS_{Residual}/(n-k-1)$	
Total	$n-1$	SS_T		

No.
Date 10/17.



もし $F > F_{k,n-k-1}(\alpha)$ であるとき H_0 を棄却する

- $SS_{Reg} = (\hat{Y} - 1_n\bar{Y})^t(\hat{Y} - 1_n\bar{Y}) = Y^t(H - \frac{1}{n}J)Y$
- $SS_{Residual} = e^t e = Y^t(I-H)Y$

$n\bar{Y}^2 = \frac{1}{n}(1Y)^t(1Y) = \frac{1}{n}Y^t \underbrace{1^t 1}_{J} Y$ $\quad J = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$ ($n \times n$)

$\underline{(H-\frac{1}{n}J)^t(H-\frac{1}{n}J) = H-\frac{1}{n}J}$ $\quad$ (∵ $HX = X(X^tX)^{-1}X^tX = X$ であるから..
$\therefore H\begin{bmatrix} 1 & x_{1,1} \\ \vdots & \vdots \\ 1 & x_{n,1} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$)

$\therefore HJ = J$, $J^t = J^t H^t = JH$

この前提

$Y = X\beta + \varepsilon \quad \varepsilon \sim \sigma^2 I_n \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \varepsilon_i \sim N(0, \sigma^2)$

$Y \sim N(X\beta, \sigma^2 I_n)$ $\quad$ 注 $H = X(X^tX)^{-1}X^t$

いま $T = (T_1, T_2, \dots T_k)^t \sim N(\mu, I_k)$ とする

$Q = \underset{1\times k}{T^t}\underset{k\times k}{A}\underset{k\times 1}{T}$ とおくとき

(べき等行列)

$\boxed{\text{定理}}$ $Q = T^tAT$ は χ^2 分布をもつ ⇔ Aは idempotent である。

このとき χ^2 分布の自由度は rank(A) = tr(A) であり、

非心度は $\mu^t A \mu$ である。である。
non-centrality parameter

べき等行列 なので

No
Date 10/17 $E((T-\mu)^t A^{-}(T-\mu))$

数理統計学の基礎 (P151 P152)

定理 もし $Q = Q_1 + Q_2$ where $Q \sim \chi^2_a$ $Q_1 \sim \chi^2_b$

さらに $Q_2 \geq 0$ であるとすると $Q_2 \sim \chi^2_{a-b}$ である。

定理 もし ${}^tTA_1T$ 及び ${}^tTA_2T$ が どちらも χ^2分布に従い、
それらが独立である $\Leftrightarrow$ $A_1A_2 = 0$ である。

(野田、宮岡、数理統計学の基礎 p152〜153)
(または教科書の最後)

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots \beta_k X_{i,k} + \varepsilon_i \quad (\text{パラメータ数} = p = k+1)$$

$$(\varepsilon_1 \cdots \varepsilon_n \sim N(0,1) \ \text{i.i.d})$$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = X\beta + \varepsilon \sim N(X\beta, \sigma^2 I) \quad \operatorname{rank} X = k+1 \ (= p)$$

定理: $R_0^2 \overset{\text{def}}{=} \min\limits_{\beta} (Y - \underset{X\hat\beta}{X\beta})^t (Y - \underset{X\hat\beta}{X\beta})$ とするとき

$$R_0^2 \sim \sigma^2 \chi^2_{n-\underset{p}{\underset{\parallel}{(k+1)}}}$$

定理: $R_1^2 \overset{\text{def}}{=} \min\limits_{\substack{\beta \\ C\beta = d}} (Y - \underset{X\hat\beta}{X\beta})^t (Y - \underset{X\hat\beta}{X\beta})$ (つまり制約付)

注
- C は $m \times (k+1)$ の行列
- $\operatorname{rank}(C) = m \quad (< \underset{p}{\underset{\parallel}{k+1}})$
- $d = m \times 1$

この辺りは 数理統計学の基礎 P251 とも関連あり　　No.　Date 10/17

・このとき $R_0^2,\ R_1^2-R_0^2$ は独立である …①

$$\underset{e^te}{R_1^2}-\underset{e^te}{R_0^2}$$

次に $R_0^2\sim\sigma^2\chi^2_{n-(k+1)}$, $R_1^2-R_0^2\sim$ 自由度mの非心χ^2分布 …②

もし $C\beta=d$ が成り立てば $R_1^2-R_2^2\sim\sigma^2\chi^2_m$

かつ $\dfrac{\left(\dfrac{R_1^2-R_0^2}{m}\right)}{\left(\dfrac{R_0^2}{n-(k+1)}\right)}\sim F_{m,\,n-(k+1)}$ …③

Full-model, Reduced-model について教科書を参照する

$$Y=X\beta+\varepsilon$$

$$\begin{array}{c} n\updownarrow \end{array}\left(\underset{p-m}{X_1}\ \underset{m}{X_2}\right)_{k+1}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}\begin{array}{l}\updownarrow p-m\\ \updownarrow m\end{array}$$

$H_0:\beta_2=0$ という仮説検定に関して

$$\underbrace{\begin{pmatrix} O & \\ & I \end{pmatrix}}_{C}\underbrace{\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}}_{\beta}\begin{array}{l}p-m\\ m\end{array}=\begin{pmatrix}0\\ d\end{pmatrix}$$

$\mathrm{rank}\,C=m$

という制約と考えれば ↑のパターンを用いて行うことが出来る

・chapter3の作業 (~10/31)

3.5, 3.6, 3.9, 3.14, 3.17, 3.21, 3.25, 3.26, 3.35, 3.36, 3.39

No
Date 「10.24〜

(ここでは尤度比に基づく.)

前回からの続きで制約付きの回帰モデルの検定について

$$\underset{n\times1}{Y} = \underset{n\times p}{X}\underset{p\times1}{\beta} + \underset{n\times1}{\varepsilon} \qquad H_0: \underset{m\times p}{C}\beta = r \quad \text{vs} \quad H_1: C\beta \neq d$$

$m\times p\ (p>m)$

$\mathrm{rank}\, C = m$ (制約の数)

制約付きの最小二乗解 $\tilde{\beta}$

$\varepsilon \sim N(0, \sigma^2 I) \quad Y \sim N(X\beta, \sigma^2 I)$

・ここで尤度比検定について考える。↓ 野田・宮岡、数理統計学の基礎にも同じようなのが載ってる

尤度関数 $L(\beta, \sigma^2 | Y, X) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(Y-X\beta)^t(Y-X\beta)\right)$

βとσ^2の最大推定量は $\hat{\beta}_{MLE} = (X^tX)^{-1}X^tY$

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}e^te = \frac{1}{n}(Y-X\hat{\beta})^t(Y-X\hat{\beta})$$

$\tilde{\beta}_{LSE}$と同じように最大推定法とラグランジュの未定乗数法を用いて

βの制限下での最大推定量 $\tilde{\beta}_{MLE}$を求める

$$\tilde{\beta}_{MLE} = \hat{\beta}_{MLE} - (X^tX)^{-1}C^t[C(X^tX)^{-1}C^t]^{-1}(C\hat{\beta}_{MLE} - r)$$

同様に制限下でのσ^2の最大推定量

$\tilde{\sigma}^2_{MLE}$を求める。 $\tilde{\sigma}^2_{MLE} = \frac{1}{n}(Y-X\tilde{\beta})^t(Y-X\tilde{\beta})$

No
Date 10.24.

$$= \frac{1}{n}\left[(Y-X\hat{\beta})+X(\hat{\beta}-\tilde{\beta})\right]^t\left[(Y-X\hat{\beta})+X(\hat{\beta}-\tilde{\beta})\right]$$

$$= \frac{1}{n}(Y-X\hat{\beta})^t(Y-X\hat{\beta}) + \frac{1}{n}(C\hat{\beta}-r)^t\left[C({}^tXX)^{-1}{}^tC\right]^{-1}(C\hat{\beta}-r)$$

次に尤度比検定について考える

$$\Lambda = \frac{\max_{(\beta,\sigma^2)} L(\beta,\sigma^2|Y,X)}{\max_{\{(\beta,\sigma^2) \mid C\beta=r\}} L(\beta,\sigma^2|Y,X)} = \frac{L(\hat{\beta}_{MLE},\hat{\sigma}^2_{MLE}|Y,X)}{L(\tilde{\beta}_{MLE},\tilde{\sigma}_{MLE}|Y,X)}$$ を計算

$\left(= \frac{L(H_1)}{L(H_0)}\right.$ とも書ける。H_1が真ならこの値は大きく(なり)棄却される$\left.\right)$

($\Lambda > c \Rightarrow$ 棄却するという形に考える.)

$$= \frac{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}}\exp\left(-\frac{n}{2}\right)}{(2\pi\tilde{\sigma}^2)^{-\frac{n}{2}}\exp\left(-\frac{n}{2}\right)} = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{\frac{n}{2}} > c \Rightarrow \text{棄却}$$

つまり $\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right) > c \Rightarrow$ 棄却

$$\Leftrightarrow \frac{\left\{\frac{(C\hat{\beta}-r)^t\left[C({}^tXX)^{-1}{}^tC\right]^{-1}(C\hat{\beta}-r)}{m}\right\}}{\left\{\frac{{}^tY\left[I-X({}^tXX)^{-1}{}^tX\right]Y}{n-p}\right\}} > c$$

↓

$Y|\beta+?$

これは非心F分布に従う

$F_{m,n-p}$

Chryv culture

No.
Date 10 24

・統計量の分布に関し　$H_0: C\beta = r$ の制限下に於いて……

$$[C(X^tX)^{-1}C^t]^{-\frac{1}{2}}(C\hat{\beta}-r) \sim N(0, \sigma^2 I_m) が成立する$$

・$C\hat{\beta} - r = C(\hat{\beta}-\beta) = C(X^tX)^{-1}[X^tY - X^tX\beta]$
$(\because C\beta = r)$

$= C(X^tX)^{-1}X^t\varepsilon$ と書ける

・$(C\hat{\beta}-r)^t[C(X^tX)^{-1}C^t]^{-1}C(C\hat{\beta}-r) = \varepsilon^t \underbrace{X(X^tX)^{-1}C^t(C(X^tX)^{-1}C^t)^{-1}C(X^tX)^{-1}X^t}_{P\ とする}\varepsilon$

（実はPもベキ等行列である。）
つまり Idempotent。

$tr(P) = tr(I_m) = m$ となる。

・$e^te = \varepsilon^t(I-H)\varepsilon = \varepsilon^t(I - X(X^tX)^{-1}X^t)\varepsilon \sim \sigma^2\chi_{n-p}$

ここで ⊛ $P(I-H) = 0$ になることから独立性が保証され
（Fの分布と同る）

$F \sim F_{m,n-p}$（H_0が真のとき）

もし H_1 下では、$(C\beta \neq r)$ は非心F分布（$F_{m,n-p}$）

に従うことである。

非心度 $(C\beta - r)^t [C(X^tX)^{-1}C^t]^{-1} C(\beta - r)/\sigma^2$ である

次に $C\beta$ の $100(1-\alpha)\%$ 信頼区間に関して考えると

$\{r \mid C(\hat{\beta} - r)^t [C(X^tX)^{-1}C^t]^{-1} C(\hat{\beta} - r) \leq m\hat{\sigma}^2 F_{m,n-p}(1-\alpha)\}$

となる。次に特殊な仮定について考える。

(I) $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$ (j は $0 \sim k$ のある数) の場合。

$C = (0, \cdots 1, \cdots 0)$ $(\mathrm{rank}\, C = 1 \therefore m = 1)$

$r = 0$

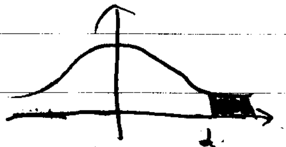
このとき 統計検定量は $F = T^2 = \dfrac{\hat{\beta}_j^2}{(X^tX)^{-1}_{jj}\hat{\sigma}^2} \sim F_{1,n-k-1}$

(Under H_0)

となるが T は t 分布に従うことに注意すると

$(1-\alpha)\cdot 100\%$ 信頼区間 (β_j) は

$\hat{b}_j \pm se(\hat{b}_j)\, t^{-1}_{n-k-1}\left(\frac{\alpha}{2}\right)$



(II) $H_0: \beta_1 = \beta_2 = \cdots = \beta_m = 0$

$Q = SS_{residual}(\text{reduced model}) - SS_{residual}(\text{full model})$
$= SS_{Reg}(\text{full model}) - SS_{Reg}(\text{reduced model})$

Chuyu culture

No.
Date 10.24.

(III) 2つの母集団について考える。

$$\begin{cases} \cdot\ Y_1 = X_1\beta_1 + \varepsilon_1 & \varepsilon_1 \sim N(0, \sigma^2 I_{n_1}) \\ \cdot\ Y_2 = X_2\beta_2 + \varepsilon_2 & \varepsilon_2 \sim N(0, \sigma^2 I_{n_2}) \end{cases}$$

$\varepsilon_1 \perp\!\!\!\perp \varepsilon_2$ (独立を表す記号)

$$\beta_1 = \begin{pmatrix} \beta^{(1)} \\ \beta_1^{(2)} \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} \beta^{(1)} \\ \beta_2^{(2)} \end{pmatrix} \begin{matrix} \updownarrow r \\ \updownarrow p-r \end{matrix}$$

ここで、$X_j = \begin{pmatrix} X_j^{(1)} & X_j^{(2)} \end{pmatrix} \quad j=1,2$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta^{(1)} \\ \beta_1^{(2)} \\ \beta_2^{(2)} \end{pmatrix}$$

$$X = \begin{pmatrix} X_1^{(1)} & X_1^{(2)} & 0 \\ X_2^{(1)} & 0 & X_2^{(2)} \end{pmatrix}$$ としてみると

$$X\beta + \varepsilon = \begin{pmatrix} X_1^{(1)} & X_1^{(2)} & 0 \\ X_2^{(1)} & 0 & X_2^{(2)} \end{pmatrix} \begin{pmatrix} \beta^{(1)} \\ \beta_1^{(2)} \\ \beta_2^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$= \begin{pmatrix} X_1^{(1)}\beta^{(1)} + X_1^{(2)}\beta_1^{(2)} \\ X_2^{(1)}\beta^{(1)} + X_2^{(2)}\beta_2^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

というように model がうまく書き換える事が出来るようになる。
(重要なモデルになるかもしれない)

↓ ②, ③は自分でも調べる，①②なども比較が考える

ちがうみたいの $(X^tX)\cdot(I-H)$ この… 10.24

これで検定、$H_0: \beta_1^{(1)}=\beta_2^{(2)}$ は、$H_0: C\beta=0$

ただし $C=(\underset{\leftrightarrow r}{O}\ \ I_{p-r}\ \ -I_{p-r})\ \updownarrow p-r$

の制約付きの回帰モデルの検定の手法が使えるようになる。

$(1-\alpha)\cdot100\%$ 信頼区間 (for β) は
(信頼域)

①
・$\{\beta \mid (\hat\beta-\beta)^t(X^tX)(\hat\beta-\beta)\le(k+1)\hat\sigma^2 F_{\underset{p}{k+1},\underset{n-p}{n-k-1}}(1-\alpha)\}$

回帰ごとに定める

② Bonferroni Confidence region
・ボンフェローニの信頼区間…

$$I_j=\hat\beta_j\pm se(\hat\beta_j)\,t_{n-k-1;\frac{\alpha}{2l}}\quad(j=1,2,\dots,l)$$

とおくと

$$1-P(\beta_j\in I_j,\ j=1\sim l)$$

$$=\Pr(\beta_j\notin I_j\ \text{for some}\ j=1\sim l)$$

$$\le\sum_{j=1}^{l}\Pr(\beta_j\notin I_j)=l\cdot\frac{\alpha}{l}=\alpha$$

ただし確率は α 以下におさえられる

∴ 信頼区間 for $\beta_1\sim\beta_l$ $\left\{\begin{pmatrix}\beta_1\\ \vdots\\ \beta_l\end{pmatrix}\middle|\ \beta_j\in I_j,\ j=1\sim l\right\}$

③ Scheffe'の信頼区間

おそらくt分布の外れ下方をつかうため

$$\{(\beta_1,\beta_2,\dots,\beta_l)^t \mid \beta_j\in\hat\beta_j\pm(k+1)^{\frac12}\cdot(F_{k+1,n-k-1}(1-\alpha))^{\frac12}\cdot se(\hat\beta_j),\ j=1\sim l\}$$

if l is such that $t_{n-k-1,\frac{\alpha}{2l}}\ge(k+1)^{\frac12}(F_{k+1,n-k-1}(1-\alpha))^{\frac12}$

Chuyu culture

No.
Date 10 24

Prediction region (Prediction Interval ではなくて 予測区間 ではなく、予測域と言うべきかって)
領域

- 次に予測区間(予測域)に関して考える.

$Y_0 = X_0^t\beta + \varepsilon_0$ (ε_0と$\varepsilon_1 \dots \varepsilon_n$は互いに独立)

Region
$(1-\alpha)\cdot 100\%$ 予測区間(予測域) for Y_0について考える.

$H = X(X^tX)^{-1}X^t \quad \hat{Y} = X\hat{\beta} = X(X^tX)^{-1}X^tY = HY$

$\because Var[\hat{Y}] = \sigma^2 H \quad (\sigma^2 HH^t = \sigma^2 H^2 = \sigma^2 H)$

$Var[Y_0 - \hat{Y}_0] = \sigma^2 + \sigma^2 X_0^t(X^tX)^{-1}X_0$

(なんか板書が適当だったので自分で調べる)

- VIF_jの話題(多重共線性)に関して
$= \frac{1}{1-R_j^2}$ (これも本のExerciseに出てきた内容を復習する)

- カテゴリカル、バイナリのデータ

$X_1 = \begin{cases} 0 & \text{storm window not present} \\ 1 & \text{o.w} \end{cases}$

or.

$X_1 = \begin{cases} 0 & \text{control} \\ 1 & \text{treatment} \end{cases}$

$\left.\begin{array}{l} Y_{1,1} \dots Y_{1,n_1} \sim N(\mu_1, \sigma^2) \\ Y_{2,1} \dots Y_{2,n_2} \sim N(\mu_2, \sigma^2) \end{array}\right\}$ i.i.d

$$\begin{pmatrix} \overbrace{1 \cdots 1}^{n_1} & \overbrace{0 \cdots 0}^{n_2} \\ 0 \cdots 0 & 1 \cdots 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} Y_{1,1} \\ \vdots \\ Y_{1,n_1} \\ Y_{2,1} \\ \vdots \\ Y_{2,n_2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_2 \end{pmatrix} + \varepsilon = \underbrace{\begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}}_{X} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \varepsilon$$

$$= \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \varepsilon \quad \begin{pmatrix} \beta_0 = \mu_1 \\ \beta_1 = \mu_2 - \mu_1 \end{pmatrix}$$

このあたりから Chapter 8 (?)

10/31. • One Indicator Variable ...

$$Y_i = \begin{cases} \mu_1 \\ \mu_2 \end{cases} + \varepsilon_i$$

• One Factor with k levels. (たぶん、分散分析，詳しく?)

実験計画が ↓

$$Y_i = \begin{cases} \mu_1 + \varepsilon_i & i = 1 \sim n_1 \\ \mu_2 + \varepsilon_i & i = n_1+1 \sim n_1+n_2 \\ \vdots & \\ \mu_k + \varepsilon_i & i = n_1 + \cdots + n_{k-1} + 1 \sim n_1 + \cdots + n_k. \end{cases}$$

$$Y_i = \sum_{j=1}^{k} \mu_j \cdot \underbrace{I(i,j)}_{} + \varepsilon_i$$

(もし i が j グループに属すれば 1)

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots \beta_k X_{i,k} + \varepsilon_i$$

(Centered Model を考える?)

$$\beta_0 = \overline{\mu} = \frac{1}{k}(\mu_1 + \cdots + \mu_k)$$

$$\beta_j = \mu_j - \overline{\mu} \qquad \sum_{j=1}^{k} \beta_j = 0$$

No. 10/31.

Date

これは覚えない
教わる

LSE…

- $\hat{\beta}_0 = \overline{Y}_{\cdot\cdot} = \frac{1}{n_1+\dots+n_k} \sum_{j=1}^{n_1+\dots+n_k} Y_j$
- $\hat{\beta}_j = \overline{Y}_{j\cdot} - \overline{Y}_{\cdot\cdot}$

(1要因)

one-way ANOVA…

㊟ $n_1+n_2+\dots+n_k = N_k$

Source	df	SS	
• treatment	$k-1$	$\sum_{j=1}^{k} n_j \overline{Y}_{j\cdot}^2 - n\overline{Y}_{\cdot\cdot}^2$	
• residual	$\sum_{j=1}^{k}(n_j-1)$	$\sum_{j=1}^{k}\sum_{i=N_{j-1}+1}^{N_j} (Y_{ij}-\overline{Y}_{j\cdot})^2$	
• total	$\sum_{j=1}^{k} n_j - 1$	$\sum_{j=1}^{k}\sum_{i=N_{j-1}+1}^{N_j} (Y_{ij}-\overline{Y}_{\cdot\cdot})^2$	

(2要因) 交互作用を考えない (各組み合わせに対して1つの値が観測される)

Block

	Factor A 1	2	…	I
Factor B 1				
2				
⋮				
J				

Block

randomized complete block design.

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

$$Y_{ij} = \beta_0 + \sum_{k=1}^{I} \alpha_k X_{ik} + \sum_{k=1}^{J} \beta_j X_{jik} + \varepsilon_{ij}$$

μ = grand mean

α_i = effect of ith block.

β_j = effect of jth treatment.

$$\sum_{i=1}^{I} \alpha_i = 0.$$

$$\sum_{j=1}^{J} \beta_j = 0$$

(LSE) $\hat{\mu} = \overline{Y}_{\cdot\cdot} = \frac{1}{IJ}\sum_i\sum_j Y_{ij}$

これは覚えておくと得になる。

No.
Date 10/31

$$\begin{cases} \hat{\alpha}_i = \overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot} \quad (i=1\sim I) & \overline{Y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij} \\ \hat{\beta}_j = \overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot} \quad (j=1\sim J) & \overline{Y}_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I} Y_{ij} \end{cases}$$

以下は Two-way ANOVA（乱塊法）

Source	df	SS
block	$I-1$	$J\cdot\sum_{i=1}^{I}(\overline{Y}_{i\cdot}-\overline{Y}_{\cdot\cdot})^2$
treatment	$J-1$	$I\sum_{j=1}^{J}(\overline{Y}_{\cdot j}-\overline{Y}_{\cdot\cdot})^2$
residual	$(I-1)(J-1)$	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y}_{\cdot\cdot})^2$
total	$IJ-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\overline{Y}_{\cdot\cdot})^2$

交互作用を考える場合

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad \begin{pmatrix} k=1\sim n_{ij} \\ i=1\sim I \\ j=1\sim J \end{pmatrix}$$

$$\sum_{i=1}^{I}\alpha_i = 0, \quad \sum_{j=1}^{J}\beta_j = 0$$

$$\sum_{i=1}^{I}\gamma_{ij} = 0 \quad (j=1\sim J)$$

$$\sum_{j=1}^{J}\gamma_{ij} = 0 \quad (i=1\sim I)$$

No. 2,3
Date 10/31.(月)　　宿題 8.7 / 8.10 / 8.11 / 8.12 / 8.13 / 8.16

mixed continuous and categorical variables.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \beta_2 x_{i2} + \varepsilon_i & \text{if } x_{i1} = 1 \\ \beta_0 + \beta_2 x_{i2} & \text{if } x_{i1} = 0 \end{cases}$$

(• $x_{i1} = 0$ or 1
• x_{i2} = continuous)

with interaction (x_{i2}の傾きも影響を受ける。)

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} \cdot x_{i2} + \varepsilon_i$$

宿) $\boxed{\text{Ch.8}}$ 8.7; 8.10; 8.11; 8.12; 8.13; 8.16 ← (Homework)

(第4章) Model-Checking (モデル検証)

GLM
(一般化線形モデル)

$$Y = X\beta + \varepsilon$$

$$\underset{[n\times 1]}{data} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \underset{[n\times p]}{X} = [X_{ij}]_{\substack{i=1\sim n \\ j=1\sim p}}$$

(我々が観測しているのはデータだけに過ぎず
モデルは我々が自分で決めたものに過ぎない)

• Gauss-Markov conditions ; Normality

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \sim N(0, \sigma^2 I)$$ … こういったデータに対する仮説は成立しているか?

ε_i, β … これらは観察出来ない

Chuyu culture

No

Date 10 31 (月)

残差… Residuals $e_i = Y_i - \hat{Y}_i$: $e = (I-H)Y$

$$E[e] = E[(I-H)Y] = (I-H)E[Y] = (I-H)X\beta = 0$$

$$(= (I - X(X^tX)^{-1}X^t)X\beta)$$

$$V[e] = V[(I-H)Y] = \sigma^2 \cdot (I-H)\, I \cdot (I-H)^t = \sigma^2 (I-H)$$

$\hat{\sigma}^2 = MS_{residual}$ $\qquad e \sim N(0, (I-H)\sigma^2)$

Standardized residuals (標準化された残差)

$$d_i = \frac{e_i}{\sqrt{MS_{residual}}} \qquad MS_{residual} = \frac{e^te}{n-p}$$

Studentised residuals (スチューデント化された残差)

$$r_i = \frac{e_i}{\sqrt{MS_{res} \cdot (1-h_{ii})}}$$

Press Residuals

$e_{(i)} = Y_i - \hat{Y}_{(i)}$ where $\hat{Y}_{(i)}$ is the predicted value of Y_i with all the data except X_i, Y_i

$$e_{(i)} = \frac{e_i}{1-h_{ii}}$$

(自分自身(X_i, Y_i)は除いた状況で用いた線形回帰モデルによって予測される)

Chryo culture

10.31

→この辺りは有名らしい 4章を読んだ方が良さそう

$$X = \begin{pmatrix} x_i \\ X_{(i)} \end{pmatrix}$$

(第3章の最後の方に出てきた ブロック分割の逆行列を用いる。 らしい…)

$$\hat{Y}_{(i)} = X_{(i)} \left(X_{(i)}^t X_{(i)}\right)^{-1} X_{(i)}^t Y_{(i)}$$

$$V[e_{(i)}] = \frac{\sigma^2}{1-h_{ii}}$$

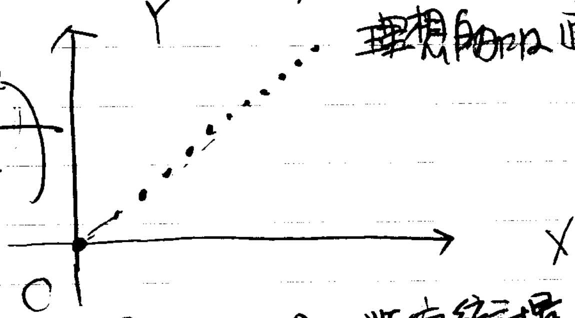
R-Student residual

$$MS_{Res} \rightarrow S_{(i)}^2$$

• Residual Plots.

Normal Probability Plot

$\Phi^{-1}\left(\frac{i-\frac{3}{8}}{n+\frac{1}{4}}\right)$

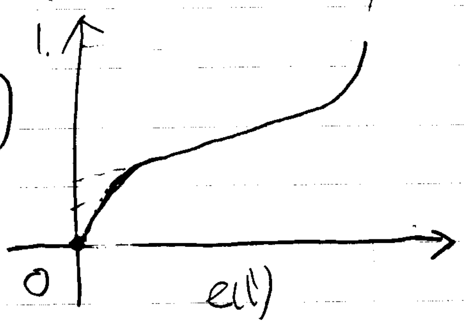


理想的には直線になるそう。

一般的に $+\frac{1}{4}$, $-\frac{3}{8}$ を加えるらしい

$e_{(i)}$ の順序統計量

$\Phi^{-1}\left(\frac{i}{n}\right)$



Heavy Tail.

(正規分布ではないときで)

p138に書いてあるかな。

Fig 4.7

No. 1031 (MON)
Date

Residuals against fitted values $\hat{Y}_i$.

check for unequal variance. non linearity.

Plot of residuals in time sequences. (p143)

Chryo culture

No

Date

Cheyu culture

12/16(金) 1/6(金)

・補課 10:20→12:10 10:20→12:10

Date 11 14 (月)

(residual plots)

Generalized Linear model

$Y = X\beta + \varepsilon$ (GM conditions) → $E[\varepsilon_i^2] = \sigma^2$ ((iに依らない) equal variance

(Normality) $E[\varepsilon_i] = 0$: partial regression

$E[\varepsilon_i \varepsilon_j] = 0$ $(i \neq j)$

Residuals : e_i vs $\hat{y}_i$ $V[\varepsilon] = \sigma^2 I_{n \times n}$

・F-test (F検定) : robust against nonnormality

(非正規性に対して頑健性をもつ)

しかし Confidence Interval は robust ではない.

(CI)

Tests of Normality (正規性の検定)

① Shapiro-Wilk Test $(n < 50)$

(シャピロ–ウィルク検定)

a_iの定義は wikipedia参照

・$u_1, u_2, \ldots u_n$ (データ)

・$u_{(1)}, u_{(2)} \ldots u_{(n)}$ ($u_1 \sim u_n$の順序統計量)

検定統計量… $W = \dfrac{\left(\sum_{i=1}^{n} a_i u_{(i)}\right)^2}{\sum_{i=1}^{n} (u_i - \bar{u})^2}$ or $\dfrac{\sum_{i=1}^{n}(a_i u_{(i)})}{\sqrt{\sum_{i=1}^{n}(u_i - \bar{u})^2}}$

a_i's depend on the mean of order statistics of $N(0,1)$

$W \leq W_\alpha \Rightarrow$ reject

Chryu culture

No.
Date 11 14 (月) 正規性の検定。

② Kolmogorov's Test (nが大きいとき)

$u_1, u_2 \cdots u_n$: iid とする (cdf が F の関数)

$H_0: F = \Phi$　　$H_1: F \neq \Phi$
(つまり正規分布)　　(正規分布では無い)

$$F_n(x) \overset{def}{=} \text{経験分布関数} = \frac{1}{n}\sum_{i=1}^{n} \underset{\sim\sim\sim}{I(u_i \leq x)}$$

注) $I(u_i \leq x) = \begin{cases} 1 & (u_i \leq x) \\ 0 & (u_i > x) \end{cases}$

$$\text{検定統計量} = D = \sup_{x \in R} \left| F_n(x) - \Phi(x) \right|$$

Reject Normality if D is large.
(Dが大きければ H_0 を棄却)

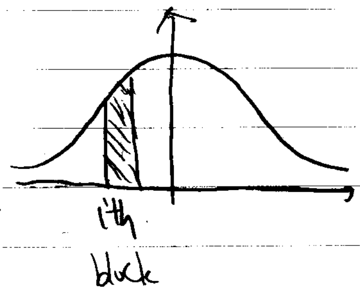
③ χ^2-test (カイ二乗検定)

$$\sum_{i=1}^{K} \frac{(E_i - O_i)^2}{E_i} \sim \chi_{K-1}$$ (これは昔から学んだカイ二乗検定)
(推定したパラメータがあれば自由度もその分減る)

(O_i … 観測値
E_i … 理論上の期待値)

これを正規性の検定にも応用する (明解演習にも触れたかな)

Date 11 14 (月)



K個の区間に区切って その中に含まれるサンプル数と[illegible]とみて χ^2-検定を行う.

E_i… expected number of samples in the ith block

Q_i… number of $u_1 \sim u_n$ in the ith block

- 正規性の仮定が無い状況における検定.

(Bootstrap法)

$H_0: C\beta = r$ vs $H_1: C\beta \neq r$

検定統計量 $T = \dfrac{(C\hat{\beta} - C\beta)^t (C(X^tX)^{-1}C^t)^{-1} (C\hat{\beta} - C\beta)}{\hat{\sigma}^2}$

$\left(= \varepsilon^t X (X^tX)^{-1} C^t [C(X^tX)^{-1}C^t]^{-1} C (X^tX)^{-1} X^t \varepsilon \right)$

$e_1, \dots, e_n$ $\quad T(e_1, e_2, \dots e_n)$

$e_1^*, \dots, e_n^*$ (random sample from $e_1, e_2, \dots e_n$) $\quad y^* = X\hat{\beta}^* + e_i^*$

(bootstrapped sample)

$T(e_1^*, \dots, e_n^*)$

repeat for B times ($B \geq 2000$) to obtain an approximation to the distribution of $T(e_1, e_2, \dots e_n)$ and its $(1-\alpha)$-th quantile T_α^*.

reject $H_0: C\beta = r$ (if: $T(e_1, e_2, \dots e_n) \geq T_\alpha^*$)

No.
Date 11/14.（月）17:00→

$Y = X\beta + \varepsilon$

Gauss Markov Conditions.

$E[\varepsilon_i] = 0 \ (i=1\sim n)$ （平均が0であること）

① residual plots.

② Partial Regression Plots.

$Y = X\beta + \varepsilon$

$Y = X_{(j)}\beta_{(j)} + X_j\beta_j + \varepsilon$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \boxed{\ } & X_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & & X_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \boxed{\ } \\ \beta_k \end{pmatrix} + \varepsilon$$ （j列目を抜き出す）

$[1 \ x_1 \cdots \underline{\underline{x_j}} \cdots x_n]$

$$= \underbrace{[1 \ x_1 \cdots x_{j-1} \ x_{j+1} \cdots x_k]}_{X_{(j)}} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} + X_j\beta_j + \varepsilon$$

$$\underbrace{(I - H_{(j)})Y}_{e(Y|X_{(j)})} = \underbrace{(I - H_{(j)})X_{(j)}}_{=0}\beta_{(j)} + (I - H_{(j)})X_j\beta_j + (I - H_{(j)})\varepsilon$$

$$e(Y|X_{(j)}) = \beta_j e(X_j|X_{(j)}) + \varepsilon^*$$

（教科書p145~146参照。）

Chryu culture

P134 あたりに $\hat{Y}_{(i)}$ の定義が述べられている

i番目のデータを除いてパラメタ推定したときの i th の推定値

③ Press Statistics

$$e_{(i)} = Y_i - \hat{Y}_{(i)}. \quad \text{PRESS} = \sum_{i=1}^{n} e_{(i)}^2 = \sum_{i=1}^{n} \left(\frac{e_i}{1-h_{ii}} \right)^2$$

※ 実はこう書けることに注意 (証明をしらべておく)

$$R^2_{prediction} = 1 - \frac{\text{PRESS}}{SS_T}$$

Lack-of-fit test (適合度検定)

- 繰り返し測定のあるモデル (Repeated Measurements) の場合

$$\begin{pmatrix} y_{1,1} \\ \vdots \\ y_{1,n_1} \\ \vdots \\ y_{m,1} \\ \vdots \\ y_{m,n_m} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} \\ \vdots & \vdots \\ 1 & x_{1,1} \\ \vdots & \vdots \\ \vdots & x_{m,1} \\ \vdots & \vdots \\ 1 & x_{m,1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \varepsilon$$

$\beta_0 = 2$

$$\underbrace{\sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_i)^2}_{SS_{Res}} = \underbrace{\sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}_{SS_{PE}} + \underbrace{\sum_{i=1}^{m} \sum_{j=1}^{n_i} n_i (\bar{Y}_i - \hat{Y}_i)^2}_{SS_{LOF}}$$

(m2)

SS_{Res} (df: $n-2$)

$= n_1 + n_2 + \dots n_m - 2$

SS_{PE} df: $n-m$

SS_{LOF} df: $m-2$

$$F_{LOF} : \frac{\{ SS_{LOF} / (m-2) \}}{\{ SS_{PE} / (n-m) \}}$$

Chrys culture

No.

HW、11/28(一)交.

P11/14. chapter 4の作業。 4.8, 4.14, 4.15, 4.17, 4.20, 4.21, 4.22, 4.24.

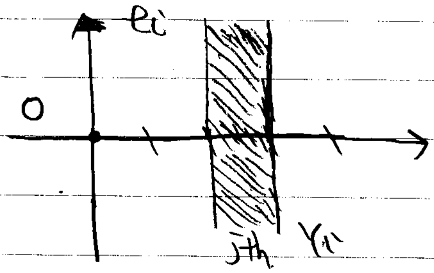
• No-repeated Measurements の場合

(near-neighborsの考え方を用いる)

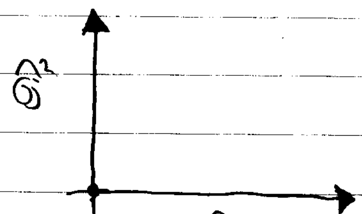
$$D_{ii'}^2 = \sum_{j=1}^{k} \frac{[\hat{\beta}_j (x_{ij} - x_{i'j})]^2}{MS_{Res}}$$ (教科書 P161に言及がある。)

次に、• $E[\varepsilon_i^2] = \sigma^2$ ($i=1 \sim n$) に関して。
(分散が等しいという仮定に関する検証)

① residual plots e_i's vs $\hat{y}_i$'s



$\hat{\sigma}_j^2$: sample variance of residual in jth interval



$\hat{y}_j$ … median of $\hat{y}_i$'s in jth interval
(中央値)

② White's Test

If equal variance assumption holds, and $h_{ii} \to 0$

then $S_1 = n^{-1}\hat{\sigma}^2 X^t X \approx S_2 = n^{-1} \sum_{i=1}^{n} e_i^2 x_i x_i^t$

Chuyu culture

③ Regress e_i's against X_{ij}'s

$nR^2 \overset{d}{\approx} \chi^2_{k-1}$ under equal variance assumption.

No. 「11/28→
Date

$$Y_i = \beta_0 + \sum_{j=1}^{k} X_{ij}\beta_j + \varepsilon_i \quad (i = 1 \sim n)$$

Gauss Markov Conditions:

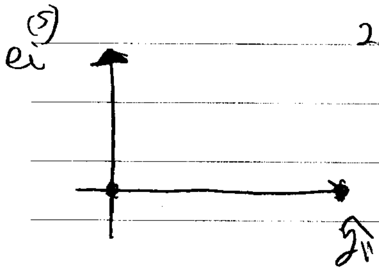
$$E[\varepsilon_i] = 0 \quad E[\varepsilon_i^2] = \sigma^2 \quad (i = 1 \sim n)$$

$$\mathrm{Cov}[\varepsilon_i, \varepsilon_j] = 0 \quad (i \neq j)$$

- Unequal Variances...

Heteroscedasticity... 等分散性の仮定に対する違反.

1. Residual plots $e_i^{(s)}$'s against $\hat{Y}_i$'s and each X-variable.



2. $$\hat{\sigma}_i^2 = \frac{\sum_{j \in i\text{th neighbor}} (e_j^{(s)} - \overline{e_i}^{(s)})^2}{\#\text{ pt in }i\text{th neighborhood}}$$

LSE

$\hat{\beta}_{OLS} = (X^tX)^{-1}X^tY$ ... unbiased even in presence of heterscedasticity, but not blue.
(非等分散性であるが、BLUEでは無い)

Chryu culture

No.
Date 11/28.

(I) Variance Stabilising Transformation.
(分散安定化変換)

$f(Y_i)$ がおおよそ等しい分散を持つような変換 f を求める。

・例えば $Y \sim Po(\eta)$, $E[Y] = \eta$; $V[Y] = \eta$ $(=\sigma^2)$

イータ

$Y \sim b(\eta)$, $E[Y] = \eta$; $V[Y] = \eta(1-\eta)$

$Y \sim Bin(n, p)$ $E[Y] = np$; $V[Y] = np(1-p)$

$= \eta$ とすると　　$= \eta(1-\frac{\eta}{n})$

$f(Y_i)$ を Taylor 展開すると

$$f(Y_i) = f(\eta_i) + f'(\eta_i)(Y_i - \eta_i) + f''(\theta_i)\frac{1}{2!}(Y_i-\eta_i)^2$$

こちらの項が比較的小さい

(ただし θ_i は η_i ~ Y_i の間のとある数。)

$$E[f(Y_i)] \fallingdotseq f(\eta_i)$$

$$V[f(Y_i)] \fallingdotseq f'(\eta_i)^2 \, Var[Y_i]$$

(これは δ-method と同じ)

以上の事から f の満たすべき要件は $(f'(\eta_i))^2 V[Y_i] =$ (何らかの)定数。

このような f を見つければ良い

No.
Date 11/28.

$f'(\eta) \propto \dfrac{1}{\sigma(\eta)}$ (その分布の標準偏差の逆数)

$\therefore f = \displaystyle\int \frac{1}{\sigma(\eta)} d\eta$ である

先ほどの例の場合だと

① $V[Y] = \eta$. $\sigma(\eta) = \sqrt{\eta}$

$\therefore f = \displaystyle\int \frac{1}{\sqrt{\eta}} d\eta = 2\sqrt{\eta}$ であるので $\sqrt{\eta}$ とする

② $V[Y] = \eta(1-\eta)$ $\sigma[\eta] = \sqrt{\eta(1-\eta)}$

$f = \displaystyle\int \frac{1}{\sqrt{\eta(1-\eta)}} d\eta$ $\eta = \sin^2\theta$ とおく $\dfrac{d\eta}{d\theta} = 2\sin\theta\cos\theta$

$= \displaystyle\int \frac{2\sin\theta\cos\theta}{\sin\theta\cos\theta} d\theta = 2\theta$ $\therefore$ $2\arcsin\sqrt{\eta}$ である。

③ 他にもし $V(Y) = \eta^2$ であると主体

$f(\eta) = \displaystyle\int \frac{1}{\eta} d\eta = \log\eta$ となる

注 $Y_i = \beta_0 + \displaystyle\sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i$ の場合

$f(Y_i) = f(\ \cdots\cdots\ )$

線形性が失われるかもしれない

Chuyu culture

④ $\sigma[\eta]=\eta^3$ $f(\eta)\propto \eta^{\frac{-1}{2}}$

- Box-Cox Transformation (Box-Cox 変換)

$$Y^{(\lambda)}=\begin{cases}\dfrac{Y^{\lambda}-1}{\lambda\cdot \dot{Y}^{\lambda-1}} & (\lambda\neq 0)\\ \dot{Y}\ln Y & (\lambda=0)\end{cases}$$

where.

$$\dot{Y}=\ln^{-1}\left(\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\ln Y_i}\right)$$ ($\ln^{-1}$ って何?)

$SS_{Res}(\lambda)$ (収入データの分析等で使用されるらしい)

$Y=\beta_0\exp(\beta_1 X)\cdot\varepsilon$ (真のモデル)

$\ln Y=\ln\beta_0+\beta_1 X+\ln\varepsilon$

(II) Weighting

$Y_i=\beta_0+\beta_1 X_{i1}+\cdots+\beta_k X_{ik}+\varepsilon_i$ ($i=1\sim n$)

$E[\varepsilon_i]=0$

$Var[\varepsilon_i]=\sigma_i^2=C_i^2\cdot\sigma^2$ (C_iは既知)

$E[\varepsilon_i\cdot\varepsilon_j]=0$ ($i\neq j$)

No. 3

Date 11/28

このとき $\frac{Y_i}{C_i} = \beta_0 \frac{1}{C_i} + \beta_1 \frac{X_{i1}}{C_i} + \cdots \beta_k \frac{X_{ik}}{C_i} + \frac{\varepsilon_i}{C_i}$ $(i=1 \ldots n)$

これはGauss Markov Conditionを満足する。

よって $\min_{\beta_0 \ldots \beta_k} \sum_{i=1}^{n} \left(\frac{Y_i}{C_i} - \frac{\beta_0}{C_i} \cdots - \beta_k \frac{X_{ik}}{C_i}\right)^2$ $\to (WY - WX\beta)$

として最小2乗推定量を求めれば良い。
(ordinary least square: OLS)

$V = \mathrm{diag}(C_1^2, \ldots C_n^2)$

$W = \mathrm{diag}(C_1^{-1} \ldots C_n^{-1})$

とおくと $\min_{\beta} (Y - X\beta)^t V^{-1} (Y - X\beta)$ と書き換える事が出来る。

$\hat{\beta}_{WLS} = (X^t V^{-1} X)^{-1} X^t V^{-1} Y$ である

(unbiased for β)

$V[\hat{\beta}_{WLS}] = (X^t V^{-1} X)^{-1} X^t V^{-1} (\sigma^2 V) ((X^t V^{-1} X)^{-1})^t$

$= \sigma^2 (X^t V^{-1} X)^{-1}$ である

$\leq V[\hat{\beta}_{OLS}] = V[(X^t X)^{-1} X^t Y]$

$= (X^t X)^{-1} X^t V ((X^t X)^{-1} X^t)^t$

$= (X^t X)^{-1} X^t V X (X^t X)^{-1}$ である

$$e_i^{(W)} = \frac{Y_i - \hat{Y}_{i,WLS}}{c_i}$$

$$\frac{1}{n-k-1}\sum_{i=1}^{n}\left(e_i^{(W)}\right)^2$$: unbiased for σ^2.

	unbiased	BLUE	MLE
OLS	✓	×	×
WLS	✓	✓	✓

以上、事に注意する ⇑

、Repeated Measurements.（繰り返しのある測定）

(Y_{il}, X_i) $l = 1 \sim n_i$

$$Y_{il} = X_i^t\beta + \varepsilon_{il}$$ （ε_{il} は G-M conditions を満足する。）

$$\sum_{i=1}^{m}\sum_{l=1}^{n_i}\left(Y_{il} - X_i^t\beta\right)^2 = \sum_{i=1}^{m}\sum_{l=1}^{n_i}\left(Y_{il} - \bar{Y}_i\right)^2 + \sum_{i=1}^{m} n_i\left(\bar{Y}_i - X_i^t\beta\right)^2$$

最小2乗推定量を考えるときこちらの項だけを考えれば良い。

$\bar{Y}_i = X_i^t\beta + \bar{\varepsilon}_i$

$Var[\bar{\varepsilon}_i] = \frac{\sigma^2}{n_i}$ となる

しかし n_i の項を考慮すると…?

すると 先ほど上で出てきた (II) Weighting の 分散の異なる モデルと見なしてやれば良い

Chryv culture

No.
Date 11 2日

$C_1 \cdots C_n$に関して Prior knowledge or information

estimate $C_1 \cdots C_n$ from OLS residuals.

WLS… iterated weighted least squares.

Correlated Errors : $V[\varepsilon] = \sigma^2 V$ (V: full rank)

- GLM : $Y = X\beta + \varepsilon$
- G.M-Conditions $E[\varepsilon] = 0$. $V[\varepsilon] = \sigma^2 I$

例 • income per capita over time
• clustered data

$$V = V^{\frac{1}{2}}(V^{\frac{1}{2}})^t = P^t D P = P^t D^{\frac{1}{2}} D^{\frac{1}{2}} P$$

$$V^{-\frac{1}{2}} Y = V^{-\frac{1}{2}} X\beta + V^{-\frac{1}{2}} \varepsilon$$

$$Y^{(V)} = X^{(V)}\beta + \varepsilon^{(V)}$$

このような変換を施すことで $V[\varepsilon^{(V)}] = V^{-\frac{1}{2}}(\sigma^2 V)V^{-\frac{1}{2}} = \sigma^2 I$

とすることが出来る。

(一般化された最小2乗推定量)

Generalized Least Squares ($Y^{(V)}$, $X^{(V)}$ に基づく最小2乗推定量)

$$\hat{\beta}_{GLS} = \left(X^{(V)t} X^{(V)}\right)^{-1} X^{(V)t} Y^{(V)}$$

$$e^{(V)}_{1} = \left({}^{t}X V^{-1} X\right)^{-1} {}^{t}X V^{-1} Y_{1} = Y^{(V)} - X^{(V)} \hat{\beta}_{GLS}$$

- Nested Errors:

$n = mM$ M sets of m observations
(m個の観測値を1つのセットと見なし、それがMセットある)

$$V = \begin{pmatrix} \Sigma & O & \cdots & O \\ O & \Sigma & & \\ \vdots & & \ddots & \\ O & & & \Sigma \end{pmatrix} \Big\updownarrow mM$$

(Σ: $m \times m$)

$$\underset{m\times m}{\Sigma} = \rho \mathbf{1}\mathbf{1}^{t} = \rho \begin{pmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{pmatrix}$$

Serial Correlation (時系列)

$$Y_t = X_t^{t} + \varepsilon_t \quad (t = 1, 2, \cdots n)$$

$\{\varepsilon_t\}$ … first order autoregressive process. AR(1)

$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$; ($u_1 \sim u_n$: iid, mean 0, Variance σ_u^2)

$|\rho| < 1$.

Chugoku culture

No.
Date 11 28

この辺り自分で計算して確認する

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t = \lim_{j \to \infty}\left\{\rho^{j+1}\varepsilon_{t-j-1} + \sum_{s=0}^{j}\rho^s u_{t-s}\right\}$$

$$= \sum_{s=0}^{\infty}\rho^s u_{t-s}$$

$E[\varepsilon_t] = 0 \quad V[\varepsilon_t] = \sigma_u^2 \sum_{s=0}^{\infty}\rho^{2s} = \dfrac{\sigma_u^2}{1-\rho^2}$ であ

$$\mathrm{Cov}[\varepsilon_t, \varepsilon_{t+h}] = \sum_{s=0}^{\infty}\rho^{2s+h} \cdot \sigma_u^2 = \frac{\rho^h}{1-\rho^2}\sigma_u^2$$

$$V[\varepsilon] = \frac{\sigma_u^2}{1-\rho^2} \cdot \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & \ddots & & & \rho^{n-2} \\ \vdots & & \ddots & & \vdots \\ \rho^{n-1} & \cdots & & \rho & 1 \end{pmatrix}$$

となる

$\hat{\rho} =$

OLS : $\hat{\varepsilon}_t = \hat{\rho}\varepsilon_{t-1} + u_t^*$

Chayu culture

「 12/5　chapter5の宿題: 5.2 / 5.8 / 5.10 / 5.12 / 5.14 / 5.16 / 5.17

No. 12.5
Date

Mixed Model … 混合 model　p194

$$Y_{ij} = \beta_0 + \beta_1 X_i + \underbrace{\delta_i}_{\text{random effect}} + \varepsilon_{ij} \qquad j = 1 \sim r_i,\ i = 1 \sim m.$$

fixed effect (unknown parameter)

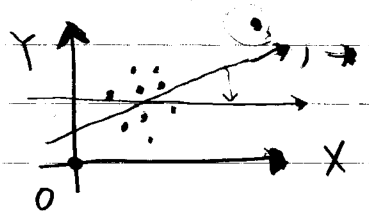
$\delta_i \sim N(0, \sigma_\delta^2)$　($i = 1 \sim m$). iid

$\varepsilon_{ij} \sim N(0, \sigma^2)$

· Residual maximum liklihood or called restricted maxmum liklihood (REML)　Jiang (1996) Ann. Statist. 24. 255~286

· outliers and influential observations.　chapter6. ↓
(外れ値と影響のある観測値)

influential points / un-usual x-values



→ これが回帰直線に大きな影響を与える.

外れ値の検出に関して… ↓ 次のページ以降述べる!
(決定)

No. 12/5
Date

chapter 6.

Leverage に関して…

$$X = \begin{pmatrix} \boxed{x_{1,1} \cdots x_{p,1}} \\ \vdots \\ x_{n,1} \cdots x_{p,n} \end{pmatrix}$$ (x_1^t)

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_1^t \\ \vdots \\ X_n^t \end{pmatrix} \beta + \varepsilon$$

X_1^t → ベクトル

計画行列

$$(X_i^t, Y_i) = (X_{1,i}, X_{2,i} \dots X_{p,i}, Y_i) \quad (i = 1 \sim n)$$

X_i の Leverage は「H_{ii}」として定義される

(つまり $H_{n \times n} = X(X^tX)^{-1}X^t = h_{ij} \quad (i,j) \in \{1 \dots n\}^2$)

- h_{ii} は X_i が すべての X_j's の 重心から どれくらい離れているかを 測る.
(centroid)
⇓ centered model を 思い出す

$$X^t = (x_1 \dots x_n)$$

$$Z^t = (X_1 - \bar{X}, \dots X_n - \bar{X}) \quad \text{where} \underset{p \times 1}{\bar{X}} = (\bar{X}_1, \dots \bar{X}_p)^t$$

$$\hat{H} = Z(Z^tZ)^{-1}Z^t = (\tilde{h}_{ij})_{(i,j) \in \{1,2,\dots n\}^2}$$

$$\hat{h}_{ii} = (X_i - \bar{X})^t (Z^tZ)^{-1} (X_i - \bar{X})$$

: standardized squared distance of X_i from $\bar{X}$

標準化された平方距離

3(?)

$$J = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

No. Date 12/5

chapter6

$$\left.\begin{aligned} \hat{Y} - 1\bar{Y} &= HY - \frac{1}{n}11^tY = HY - \frac{1}{n}JY \\ \hat{Y} &= Z(Z^tZ)^{-1}Z^t(Y-1\bar{Y}) + 1\bar{Y} = \tilde{H}Y + 1\bar{Y} \end{aligned}\right\}$$ これら2つを見比べる

Hence: $H - \frac{1}{n}J = \tilde{H}$ であるので

ゆえに $h_{ii} - \frac{1}{n} = \tilde{h}_{ii}$ である。

$\hat{Y} = HY$, $\hat{Y}_i = \sum_{j=1}^{n} h_{ij} Y_j$

$\sum_{i=1}^{n} h_{ii} = tr(H) = p = k+1$

rule of thumb … $h_{ii} > \frac{2}{n}(k+1)$

さて outliers (外れ値, 異常値) を検出するために、
(外部)スチューデント化残差を用いる。

$$e_i^* = \frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}}$$ (教科書で t_i と表記されたもの だろうか?)

$e_i = Y_i - \hat{Y}_i$ であり、$s_{(i)}$ は (x_i^t, Y_i) が除かれた

場合の最小2乗法を用いたときの s である。

(これは chapter4 で出現したはず。付録 C.6〜C.7 あたりにも載ってた。
(p590〜p594))

Chuyu culture

No. 12/5
Date

chapter 6

$\hat{\beta}_{(i)} = (X_{(i)}^t X_{(i)})^{-1} X_{(i)}^t Y_{(i)}$ where $X_{(i)}^t = (x_1, x_2 \cdots x_{i-1}, x_{i+1} \cdots x_n)$

$Y_{(i)} = (Y_1 \cdots Y_{i-1}, Y_{i+1}, \cdots Y_n)^t$

このとき $\hat{\beta}_{(i)} = \hat{\beta} + \frac{(X^tX)^{-1} x_i e_i}{1-h_{ii}}$ である

$(n-k-2) S_{(i)}^2 = (n-k-1) S^2 - (1-h_{ii})^{-1} e_i^2$ である

$= \sum_{j=1}^{n} (Y_j - x_j \hat{\beta}_{(i)})^2 - (Y_i - x_i \hat{\beta}_{(i)})^2$

$\hat{\beta}_{(i)} = \underbrace{(X^tX - x_i x_i^t)^{-1}}_{\shortparallel} (X^tY - x_i Y_i)$

問9.2 $(X^tX)^{-1} + \frac{(X^tX)^{-1} x_i x_i^t (X^tX)^{-1}}{1 - x_i^t (X^tX)^{-1} x_i}$

と書ける。

導出

$A \overset{\text{def}}{:=} ((X^tX) - (x_i x_i^t))^{-1}$ とおく

$A(X^tX - x_i x_i^t) = I$

$AX^tX - A x_i x_i^t = I$

$A - A x_i x_i^t (X^tX)^{-1} = (X^tX)^{-1}$

$A(I - x_i x_i^t (X^tX)^{-1}) = (X^tX)^{-1}$

右から x_i をかけると $A(x_i - x_i x_i^t (X^tX)^{-1} x_i) = (X^tX)^{-1} x_i$

$\therefore A x_i - A x_i x_i^t (X^tX)^{-1} x_i = (X^tX)^{-1} x_i$

chapter 6

ここで $A x_i x_i^t (X^tX)^{-1} x_i = (X^tX - x_i x_i^t) x_i x_i^t (X^tX)^{-1} x_i$

$\therefore A x_i = (X^tX - x_i x_i^t) x_i \underbrace{\boxed{x_i^t (X^tX)^{-1} x_i}}_{h_{ii}} + (X^tX)^{-1} x_i$

$= (X^tX)^{-1} x_i / (1 - h_{ii})$

$A = (X^tX)^{-1} + \dfrac{(X^tX)^{-1} x_i x_i^t (X^tX)^{-1}}{1 - h_{ii}}$ である

⋮

Plot residuals against $\hat{Y}_i$'s and any of the covariates

- 1. outliers ⇔ large residuals ✓
- 2. outliers and high leverage point ⇔ small residual
- 3. high leverage point ⇔ small residual

↑ Can be detected by deleted residuals.

$Y_i - \hat{Y}_{(i)}$

$= Y_i - X_i^t \hat{\beta}_{(i)}$

or externally studentized residuals (t_i)

No.
Date 12 5.

chapter6

Remedy

- outlierに出くわしたときの修正法...

① recording or measurement error → delete
(記録や測定上のエラー) (捨てる)

② rare event → retain
(非常に稀な事象) (残す)

③ coming from a different population → remove
(異なる母集団から来ている) (捨てる)

- 影響の測定 (Measurement of influence)

① leverage h_{ii}

「βの推定量の変化を比較する」

② compare change in estimation of β
with or without using (x_i^t, Y_i)

$$\hat{\beta}-\hat{\beta}_{(i)} = (X^tX)^{-1}x_i e_i / (1-h_{ii})$$

The jth component of $\hat{\beta}-\hat{\beta}_{(i)}$ is $DFBETA_{ij}$

$DFBETA_{ij}$ とは $\hat{\beta}-\hat{\beta}_{(i)}$ の第j成分
($\hat{\beta}_j-\hat{\beta}_{(i)j}$)

Standardized DFBETA (標準化された DFBETA)

$$\mathrm{DFBETAS}_{(ij)} = \frac{\hat{\beta}_j - \hat{\beta}_{(i)j}}{\sqrt{S_{(i)}^2 C_{jj}}} = \left(\frac{1}{(1-h_{ii})C_{jj}}\right)^{\frac{1}{2}} \cdot \left((X^tX)^{-1} x_i\right)_j e_i^*$$

$\left(C_{jj} \overset{\mathrm{def}}{:=} (X^tX)^{-1}_{(jj)} \text{ のこと} \right)$

$$\frac{\mathrm{DFBeta}(ij)}{\mathrm{s.e.}(\mathrm{DFBeta}\,(ij))} = e_i^*$$

③ Compare the fitted values: $\mathrm{DFFIT}_i = \hat{Y}_i - \hat{Y}_{(i)}$

$$= x_i^t\hat{\beta} - x_i^t\hat{\beta}_{(i)} = h_{ii} e_i / (1-h_{ii})$$

Standardized version : $\mathrm{DFFITS}_i = \dfrac{h_{ii}^{\frac{1}{2}} e_i}{S_{(i)} \cdot (1-h_{ii})} = \left(\dfrac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}} e_i^*$

• ratio of the determinants of the estimated covariance matrices of $\hat{\beta}$ and $\hat{\beta}_{(i)}$

$$\mathrm{COVRATIO}_i = \frac{\left| S_{(i)}^2 (X_{(i)}^t X_{(i)})^{-1} \right|}{\left| S^2 (X^tX)^{-1} \right|}$$

Deleted Residuals $Y_i - \hat{Y}_{(i)} = Y_i - x_i^t \hat{\beta}_{(i)} = \dfrac{e_i}{1-h_{ii}}$

Chrys culture

No. 12.5
Date

chapter 6. (其他影響力測度)

④ Cook's Distance: (p215 附近 參照)

$$D_i = \frac{(\beta - \hat{\beta}_{(i)})^t (X^t X)(\beta - \hat{\beta}_{(i)})}{(k+1)s^2}$$

$$= \frac{e_i^2 h_{ii}}{(k+1)s^2(1-h_{ii})^2}$$

- compute these measures.

- Histograms, box-plots.

- Critical level for h_{ii}'s $= \frac{2(k+1)}{n}$

  $\quad$ 〃 $\quad e_i^*$'s : 2

  $\quad$ 〃 $\quad$ DIFFITi's : $\frac{2(k+1)}{(n-k+1)}$

Polynominal Regression in One Variable.

12/16(金) 多項式回帰モデル

- $Y = \beta_0 + \beta_1 X + \varepsilon$ (単回帰モデル)
- $Y = \sum_{j=0}^{k} \beta_j \cdot x^j + \varepsilon$ (k次の多項式回帰)

(LSE) $\min_{\beta_0 \ldots \beta_p} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{k} \beta_j x_i^j \right)^2$

注1

- kが増大すれば、分散が増大してしまう。
- kが減少すれば、バイアスが増大してしまう。

} という問題が発生してしまう。

(次数)

kはなるべく小さな値に抑えておく必要がある。

注2.

- extrapolation

注3.

- Ill-conditioning $(X^tX)^{-1}$ 不安定

(P226も参照)

- Piecewise Polynominal Models.

Splines

piecewise polynominal of order k knots - points at which the segments are joined.

No.
Date 12 16

Cubic spline $k=3$の場合

$$Y = \sum_{j=0}^{3} \beta_{0j} x^j + \sum_{i=1}^{n} \left(\sum_{j=0}^{3} \right) \beta_{ij} (x-t_i)_+^j + \varepsilon$$

$$\sum_{i=1}^{n} \left(\beta_{i0} (x-t_i)_+^0 + \beta_{i1} (x-t_i)_+ + \beta_{i2} (x-t_i)_+^2 + \beta_{i3} (x-t_i)_+^3 \right)$$

$$t_1 < t_2 < \cdots < t_n$$

注 $(x-t_i)_+ = \begin{cases} (x-t_i) & \text{if } x-t_i > 0 \\ 0 & \text{else} \end{cases}$

Orthogonal Polynominals.

$$Y = \beta_0 + \beta_1 X + \cdots + \beta_k X^k + \varepsilon$$

$$Y = \alpha_0 P_0(x) + \cdots \alpha_k P_k(x) + \varepsilon$$

$P_j(x) \cdots$ jth order polynominal

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0 \quad (r \neq s) \quad (r,s) = (0,0) \cdots (k,k)$$

$$P_0(x) = 1$$

Nonparametric Regression.

$Y = m(X) + \varepsilon$

$E[\varepsilon] = 0 \quad V[\varepsilon] = \sigma^2$

m: regression function

1. Cubic Spline $h \to \infty$
2. Orthogonal basis $\quad Y = \sum_{j=0}^{\infty} \alpha_j P_j(X) + \varepsilon$

trigonometric regression wavelets

3. kernel regression.

$Y_i = m(X_i) + \varepsilon_i$

$$\hat{m}(X) = \frac{\sum_{i=1}^{n} K\left(\frac{X_i - X}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{X_i - X}{h}\right)}$$

where K is the kernel (pdf) and h is the bandwidth

local linear regression

$$\hat{m}(X) = \frac{S_2 T_0 - S_1 T_1}{S_0 S_2 - S_1 S_1} = \frac{\sum_{i=1}^{n} (S_2 - S_1 (X_i - X)) K\left(\frac{X_i - X}{h}\right) Y_i}{\sum_{i=1}^{n} (S_2 - S_1 (X_i - X)) K\left(\frac{X_i - X}{h}\right)} = \sum_{j=1}^{n} W_{jx} Y_j$$

where $S_l = \sum_{i=1}^{n} K\left(\frac{X_i - X}{h}\right)(X_i - X)^l \quad l = 0, 1, 2$

$T_l = \sum_{i=1}^{n} K\left(\frac{X_i - X}{h}\right)(X_i - X)^l Y_i \quad l = 0, 1$

Chryu culture

No
Date 12 16 /19 … 7.6, 7.7, 7.9, 7.13, 7.14, 7.18, 7.19, 7.20 (12/26)

chapter 7. 宿題…

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 (x_i - x))^2 K\left(\frac{x_i - x}{h}\right)$$

• chapter 9. Multicollinearity (多重共線性)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

X_1, X_2 are linearly dependent.
(X_1, X_2 … 線形従属であるとき)

$\exists t_1, t_2, c$ $t_1 X_1 + t_2 X_2 = C$ であるような状況を考える.
(0ではなくCであるようだ)

最小2乗推定量

LSE …
$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^t X)^{-1} X^t Y$$

$$= \begin{pmatrix} 1 & r_{12} \\ r_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix}$$

(centered and rescaled
中心化及び尺度を正規化)

$$= \frac{1}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{21} & 1 \end{pmatrix} \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix}$$

$$= \frac{1}{1 - r_{12}^2} \begin{pmatrix} r_{1y} - r_{12} r_{2y} \\ r_{2y} - r_{21} r_{1y} \end{pmatrix}$$

ch9 12 19

$$Var[\hat{\beta}_j] = \frac{\sigma^2}{1-r_{12}^2} \quad (j=1,2)$$

$|r_{12}| \to 1$ に近づくにつれて $Var[\hat{\beta}_j] \to \infty$ である。

$$V[\hat{\beta}] = \sigma^2 (X^t X)^{-1} = \sigma^2 C$$

$$V[\hat{\beta}_j] = \sigma^2 \cdot C_{jj} = \frac{\sigma^2}{1-R_j^2}$$

・ここで R_j^2 は説明変数 X_j に対して他の $p-1$ 個の変数により回帰を行ったときの決定係数である

・Source of Multicollinearity ...

Data collection method employed

Constraints（制約）

Model Specification 例 polynomial regression

Overdefined model

$$E[(\hat{\beta}-\beta)^t(\hat{\beta}-\beta)] = \sum_{j=1}^{p} V[\hat{\beta}_j]$$

$$= \sigma^2 \cdot \sum_{j=1}^{p} C_{jj} = \sigma^2 \cdot trC = \sigma^2 tr[(X^t X)^{-1}] = \sigma^2 \cdot \sum_{j=1}^{p} \frac{1}{\lambda_j}$$

（λ_j … 固有値）

メモ

Aが対角化できる逆行列とも $P^{-1}AP = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ $trP^{-1}AP = trAPP^{-1} = trA$

ch.9

No

Date 12 19

- Diagnostics

Variance Inflation Factors

$$VIF_j = C_{jj} = \frac{1}{1-R_j^2} \quad (j = 1 \ldots p)$$

(ex $VIF > 5 \rightarrow 1 - R_j^2 < \frac{1}{5} \;\therefore\; R_j^2 > 0.8$)

Eigensystem analysis of X^tX

the condition indices of X^tX:

$$K_j := \frac{\lambda_{max}}{\lambda_j} \quad (j = 1, 2 \ldots p)$$

λ_α ($\lambda_1, \ldots, \lambda_p$)

where $\lambda_{max} = \max\{\lambda_1, \lambda_2 \ldots, \lambda_p\}$.

Condition number

$$K = \frac{\lambda_{max}}{\lambda_{min}}$$

where $\lambda_{min} = \{\lambda_1, \ldots \lambda_p\}$

- $K < 100$ .... There is NO serious problem.
- $100 < K < 1000$ moderate to strong multicollinearity
- $K > 1000$ .. Strong multicollinearity

Chryr culture

1/6 (5) ch.9

No.
Date 12 19

第6章の宿題 6.8、6.10、6.11、6.14、6.16

第9章 9.2, 9.3、9.4、9.5、9.10、9.23

- Remedies（解決策）

① Collect more data.（さらにデータを集める）
② Respecify the model.（モデルを再検討する）
㊑ 変数の削除

Principal Component Regression

$Y_{n\times 1} = X\beta + \varepsilon = Z\alpha + \varepsilon$

$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}$

$Z = XT,\ \alpha = {}^tT\beta,\ {}^tZZ = {}^tT\,{}^tX XT = \Lambda$

（ここで $T\,{}^tT = I$ ということかな？つまり 実コンパクト行列）
= 直交行列）

→ おそらく ${}^tXX$…を対角化している
（対称行）

もし $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_r > \underbrace{\lambda_{r+1} = \cdots \lambda_p \fallingdotseq 0}$

$Z = [Z_{(r)}, Z_{(p-r)}]$ （左の方から何個か）

$\alpha = [{}^t\alpha_{(r)}, {}^t\alpha_{(p-r)}]^t = \begin{bmatrix} \alpha_{(r)} \\ \alpha_{(p-r)} \end{bmatrix}$

$T = [T_{(r)}, T_{(p-r)}]$

このとき $Y = \underbrace{U_{(r)}}_{(Z_{(r)}?)}\alpha_{(r)} + \varepsilon$

Chuyu culture

Ch.9

No. 12.19

Date

Biased Estimator.

Incomplete principle component regression.

$$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_r > \underbrace{\lambda_{r+1} \geqq \cdots \geqq \lambda_p}_{\fallingdotseq 0}$$

(ここは捨て去ってしまう.)

$$Y = Z_{(r)}\alpha_{(r)} + \tilde{\varepsilon}$$

$$\hat{\alpha}_{(r)} = \left(Z_{(r)}^t Z_{(r)}\right)^{-1} Z_{(r)}^t Y \quad (\text{biased for } \alpha_{(r)})$$

$$\mathrm{cov}[\hat{\alpha}_{(r)}] = \sigma^2 \cdot \left(Z_{(r)}^t Z_{(r)}\right)^{-1}$$ である. $(< \mathrm{cov}[\hat{\alpha}])$

Ridge Regression

$$h\beta^t I \beta$$

$$\shortparallel$$

L2 NORM

$$\min_{\beta} (Y - X\beta)^t (Y - X\beta) + \underbrace{h\beta^t\beta}$$

(ここはpenalty項として加える)

$$\frac{\partial L(\beta)}{\partial \beta} = -2X^tY + 2X^tX\beta + 2h\beta = 0$$

$$(X^tX + hI)\beta = X^tY$$

$$\therefore \hat{\beta}_R = (X^tX + hI)^{-1} X^tY$$

この解をridge estimatorと呼ぶ.

(なぜか突然変更でこれが成る) 12 19.

$$(X^tX+kI)^{-1}X^tY$$

続き…$= \underbrace{(X^tX+kI)^{-1}X^tX}_{Z_k}\hat{\beta} = Z_k\hat{\beta}$

(理由 $\hat{\beta} = (X^tX)^{-1}X^tY$ なら
$X^tX\hat{\beta} = X^tY$ となるため)

$\therefore E[\hat{\beta}_R] = E[Z_k\hat{\beta}] = Z_k\beta$

$V[\hat{\beta}_R] = \sigma^2 \cdot (X^tX+kI)^{-1}X^tX(X^tX+kI)^{-1}$ である.

$$MSE(\hat{\beta}_R) = cov(\hat{\beta}_R) + Bias(\hat{\beta}_R)^t Bias(\hat{\beta}_R)$$

$$TMSE(\hat{\beta}_R) = tr(MSE(\hat{\beta}_R)) = \sigma^2 \cdot \sum_{i=1}^{p} \frac{\lambda_j}{(\lambda_j+k)^2} + k^2\beta^t(X^tX+kI)^{-2} \cdot \beta$$

Ridge trace:

plot k against the co-efficient estimates.

There is a volume of k, say k_0 s.t $TMSE(\hat{\beta}_{k_0})$
$\leq TMSE(\hat{\beta}_k)$ for all $0<k<k_0$

12/19

ch9 HW… 9.2, 9.3, 9.4, 9.5, 9.10, 9.23

↓ chapter 10. 変数選択とモデル構築.

Variable Selection
many variables
exclude unimportant

真のモデル

$$X\beta = [X_1 \ X_2]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$$

X_1: p, X_2: $k+1-p = r$

True model.. $Y = X\beta + \varepsilon$

Effect of dropping variables

Suppose $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$ is the correct model.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

is considered $(p + r = k+1)$

・β_1 の最小2乗推定量… $(X_1^t X_1)^{-1} X_1^t Y = \hat{\beta}_1$

$$E[\hat{\beta}_1] = (X_1^t X_1)^{-1} X_1^t (X_1\beta_1 + X_2\beta_2)$$

$$= \beta_1 + \underbrace{(X_1^t X_1)^{-1} X_1^t X_2 \beta_2}_{\text{bias of } \hat{\beta}_1}$$

$$\mathrm{cov}[\hat{\beta}_1] = \sigma^2 (X_1^t X_1)^{-1}$$

$\mathrm{Bias}(\hat{\beta}_1) = 0$ if $X_1^t X_2 = 0$

If $X_1^t X_2 = 0$ then $\hat{\beta}$, LSE under the true model has

$$\mathrm{cov}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \sigma^2 (X^t X)^{-1} = \sigma^2 \begin{pmatrix}(X_1^t X_1)^{-1} & 0 \\ 0 & (X_2^t X_2)^{-1}\end{pmatrix}$$

((メモ) $V-W$: 半正定値 $\Rightarrow$ $W^{-1}-V^{-1}$ も半正定値

(証) C を正定値対称行列とし $C \geq I \Rightarrow I \geq C^{-1}$ を示す (固有値を考えよ)

$W^{-\frac{1}{2}}(V-W)W^{-\frac{1}{2}} \geq 0$ $\therefore W^{-\frac{1}{2}}VW^{-\frac{1}{2}} \geq I$

$\therefore I \geq W^{\frac{1}{2}}V^{-1}W^{\frac{1}{2}}$

$\therefore W^{-1} \geq V^{-1}$)

Otherwise. $\mathrm{COV}[\hat{\beta}_1] = \sigma^2\left[X_1^tX_1 - X_1^tX_2(X_2^tX_2)X_2^tX_1\right]^{-1}$

$= \sigma^2\left[X_1^t(I-H_2)X_1\right]^{-1}$ である

$\geq \sigma^2(X_1^tX_1)^{-1}$ (注) $\left([X_1 X_2]^t[X_1 X_2]\right)^{-1}$ を考えるとOK

$= \mathrm{COV}[\hat{\beta}_1]$ (教科書の Appendix C を参照)

$S^2 = \hat{\sigma}^2$ estimator of σ^2 under the true model. is

$$S^2 = \frac{1}{n-k-1}\underbrace{Y^t(I-H)Y}_{SS_{res}}$$

$\tilde{\sigma}^2 =$ estimator of σ^2 under $Y = X_1\beta_1 + \varepsilon$

is $\tilde{\sigma}^2 = \frac{1}{n-p}Y^t(I-H_1)Y$. $(SSE(p) = Y^t(I-H_1)Y.)$

$$(n-p)E[\tilde{\sigma}^2] = E\{\mathrm{tr}[Y^t(I-H_1)Y]\} = E\{\mathrm{tr}[(I-H_1)YY^t]\}$$

$$= \mathrm{tr}(I-H_1)E[YY^t] = \sigma^2(n-p) + \beta^tX^t(I-H_1)X\beta$$

$$\geq \sigma^2(n-p)$$

12/26 ~ (ch10)

No.

Date Variable Selection

$\hat{Y}_p = X_1 \tilde{\hat{\beta}} = H_1 Y$

$E[\hat{Y}_p] = H_1 X\beta = X\beta - \underbrace{(I-H_1)X\beta}$

biased unless $(I-H_1)X\beta = 0$

$\|Bias(\hat{Y}_p)\|^2 = (X\beta)^t (I-H_1)(X\beta)$ あれ?

$= E[SS_{res}^{(p)}] - (n-p)\sigma^2$

$MSE(\hat{Y}_p) = E[(\hat{Y}_p - X\beta)(\hat{Y}_p - X\beta)^t]$

$= V[\hat{Y}_p] + Bias(\hat{Y}_p) Bias(\hat{Y}_p)^t$

$= \sigma^2 H_1 + (I-H_1) X\beta\beta^t X^t (I-H_1)$

$TMSE(\hat{Y}_p) = p\sigma^2 + (X\beta)^t (I-H_1)(X\beta)$

$\dfrac{TMSE(\hat{Y}_p)}{\sigma^2} = p + \dfrac{(X\beta)^t (I-H_1)(X\beta)}{\sigma^2}$

$\dfrac{SS_{res}(p)}{\hat{\sigma}^2} - (n-2p)$ .. Mallow's C_p (p735)

↗ 先ほどとは逆の状況

・もし 真のmodelが $Y = X_1\beta_1 + \varepsilon$ であり $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ をモデルとしてあてはめたとする。

・β_1の最小2乗推定量 $\tilde{\beta}_1 = (X_1^t(I-H_2)X_1)^{-1}X_1^t(I-H_2)Y$

$E[\tilde{\beta}_1] = \beta_1$ (unbiased)

$V[\tilde{\beta}_1] = \sigma^2(X_1^t(I-H_2)X_1)^{-1} \geq cov[\hat{\beta}_1]$

βのLSE under $Y = X_1\beta_1 + \varepsilon$ は

$\hat{\beta}_1 = (X_1^tX_1)^{-1}X_1^tY$ $\quad E[\hat{\beta}_1] = \beta_1$

$V[\hat{\beta}_1] = \sigma^2(X_1^tX_1)^{-1}$

(注) $$\begin{bmatrix} X_1^tX_1 & X_1^tX_2 \\ X_2^tX_1 & X_2^tX_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1^tX_1)^{-1} + (X_1^tX_1)^{-1}X_1^tX_2G_{1\cdot2}X_2^tX_1(X_1^tX_1)^{-1} & -(X_1^tX_1)^{-1}X_1^tX_2G_{1\cdot2} \\ -G_{1\cdot2}X_2^tX_1(X_1^tX_1)^{-1} & G_{1\cdot2} \end{bmatrix}$$

ここで $G_{1\cdot2} = (X_2^t(I-H_1)X_2)^{-1}$

また $[X_1\ X_2]\begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} = [X_2\ X_1]\begin{bmatrix}\beta_2 \\ \beta_1\end{bmatrix}$ と見方を変えても同じことが起きる

(共分散の計算をすれば)

Chuyu culture

12/26
No
Date

- Subset Regression

k個の変数が存在し、2^kの可能なmodelが存在する

- モデル選択の基準

1. $R_p^2 = \frac{SS_{Reg}(p)}{SS_T} = 1 - \frac{SS_{Res}(p)}{SS_T}$

$R_{adj,p}^2 = 1 - \frac{\left(\frac{SS_{Res}(p)}{n-p}\right)}{\left(\frac{SS_T}{n-1}\right)}$ である

2. Residual Mean Square

$MS_{Residual} = \frac{SS_{Res}(p)}{n-p}$

3. Mallows Cp

$C_p = \frac{SS_{Res}(p)}{\hat{\sigma}^2} - n + 2p$

(small value of C_p)

4. $PRESS_p = \sum_{i=1}^{n} (Y_i - \hat{Y}_{(i),p})^2 = \sum_{i=1}^{n} \left(\frac{e_i}{1-h_{ii}}\right)^2$

No

Date

5、 Akaike Information Criterion (赤池の情報量規準)

$$AIC(p) = n \cdot \ln\left(\frac{SS_{res}(p)}{n}\right) + 2p$$

Baysian Information Criterion (ベイズ情報量基準)

$$BIC(p) = n \cdot \ln\left(\frac{SS_{res}(p)}{n}\right) + p \cdot \ln(n)$$

Stepwise Regression Methods

1、 Backward Elimination

- a、 Start with : Full model
- b、 partial F statistics for each variable
  - remove the variable with the smallest partial F statistics if it is not significant
  - or stop if all the partial F statistics are significant
- c、 repeat b

(Caution) Some variables are highly correlation they have low partial F-values and may be deleted early.

12.26

No

Date

2. Forward Selection

a. start with $Y_i = \beta_0 + \varepsilon_i$

b. partial F-statistics for every variables not yet in the model

- include the one with the largest partial F-value if it is significant.
- stop if none of the F-value is significant.

c. repeat b

3. Stepwise procedure

a. select the variable that has the highest correlation with Y.

b. After each new variable is entered, check every variable already in the model if it should be deleted.

期末テスト ~10章 1/9 ch.4~ch10

10.3 / 10.4 / 10.10 / 10.13

LASSO… $\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 x_{i1} \cdots \beta_K x_{iK})^2+\lambda\sum_{j=1}^{K}|\beta_j|$

(罰則項)

SACAD… $\sum(\qquad)^2+\lambda\sum_{j=1}^{K}a(|\beta_j|)$

その他キーワード　Screening　Fan & Lv 2008　JRSSB

Dimen reduction

Sparsity. Condition

Chuyu culture

No
Date 「1/6 (金)

- Non-parametric Regression
- parametric regression

- Non-linear Regression

$$y_i = f(x_i;\theta) + \varepsilon$$

$$E[\varepsilon] = 0$$

↓

このような場合も最小2乗法をあてはめる事が出来る

$$\cdots \underset{\theta}{\mathrm{argmin}} \sum_{i=1}^{n} (Y_i - f(x_i|\theta))^2$$

$$S(\vec{\theta}|x) \overset{\mathrm{def}}{=} \sum_{i=1}^{n} (Y_i - f(x_i|\theta))^2$$

$$\frac{\partial S}{\partial \theta} = 0 \quad \left(\Leftrightarrow \begin{pmatrix} \frac{\partial S}{\partial \theta_1} \\ \vdots \\ \frac{\partial S}{\partial \theta_p} \end{pmatrix} = \vec{0} \right)$$

- もし $f(x|\theta) = \theta_1 \exp({}^t x \theta_2)$ であるとき

$$\log f(x|\theta) = \log \theta_1 + {}^t x \theta_2 \qquad X = \begin{bmatrix} x_1^t \\ \vdots \\ x_n^t \end{bmatrix}$$

変数変換を行い $\ln Y$ と みればよい

- もし $f(x|\theta) = \dfrac{\theta_1 x}{x + \theta_2}$ と書けるとき

$$\frac{1}{y} = \frac{x + \theta_2}{\theta_1 x} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \cdot \frac{1}{x}$$

$$\therefore y^* = \underbrace{\frac{1}{\theta_1}}_{\theta_1^*} + \underbrace{\frac{\theta_2}{\theta_1}}_{\theta_2^*} \cdot \frac{1}{x} + \varepsilon^* \text{ と書ける}$$

(誤差項も逆数化されることに注意)

Chuyu culture

※ 2変数 $f(a+h, b+k) = \sum_{t=0}^{n} \frac{1}{t!} \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^t f(a,b)$

線形化: linearization

Taylor展開 (1次の項)

$$f(x|\theta) \doteqdot f(x|\theta_0) + \sum_{j=1}^{p} \left[\frac{\partial f(x;\theta)}{\partial \theta_j} \right]_{\theta=\theta_0} (\theta_j - \theta_{0j})$$

for some $\theta_0 = (\theta_{01} \dots \theta_{0p})$

- Binary Response Variables

$(X_i, Y_i) \quad i = 1 \sim n$

$$Y_i = \begin{cases} 0 & 1-\pi_i \\ 1 & \pi_i \end{cases}$$

$E[Y_i] = \pi_i$

$Var[Y_i] = \pi_i(1-\pi_i)$

- Logistic Regression

$$E[Y_i] = \pi_i = \frac{e^{x_i^t\beta}}{1+e^{x_i^t\beta}} \qquad \frac{\pi_i}{1-\pi_i} = e^{x_i^t\beta} \qquad x_i^t\beta = \log\frac{\pi_i}{1-\pi_i}$$

- likelihood function

$$L(\beta) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} = \prod_{i=1}^{n} e^{x_i^t\beta y_i}(1-\pi_i)$$

$$\ln L(\beta) = \sum_{i=1}^{n} x_i^t\beta Y_i - \sum_{i=1}^{n} \ln(1+e^{x_i^t\beta})$$

No.
Date

$$\frac{\partial}{\partial \beta}\ln L(\beta) = \sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} \pi_i X_i = X^t Y - X^t \mu$$

$$㊟\ \mu = (\pi_1 \cdots \pi_n)^t$$

- Liklihood Ratio Test

$$LR = 2\log\frac{L(\text{Full Model})}{L(\text{Redtced Model})} \xrightarrow{dist} \chi^2_k \quad k:\ \text{パラメタの自由度の差}$$

$$\left(= -2\ln\frac{L(\text{Reduced Model})}{L(\text{Full Model})}\right) \qquad \dim\Theta - \dim\Theta_0 \quad (df\Theta - df\Theta_0)$$

- Goodness-of-fit Test

・Deviance… (逸脱度)

$$D = 2\ln\frac{L(\text{Saturated Model})}{L(M)} \quad \left(= -2\ln\frac{L(M)}{L(\text{Saturated Model})}\right)$$

$$= 2\sum_{i=1}^{n}\left[Y_i \ln\frac{Y_i}{n_i\hat{\pi}_i} + (n_i - Y_i)\ n_i(1-\hat{\pi}_i)\right.$$

$$\xrightarrow{d} \chi^2_{n-p} \ (\text{under } M)$$

・ピアソンの χ^2 検定…

$$\sum_{i=1}^{n}\frac{(Y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i} + \frac{[(n_i - Y_i) - n_i(1-\hat{\pi}_i)]^2}{n_i(1-\hat{\pi}_i)} = \sum_{i=1}^{n}\frac{(Y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

$$\xrightarrow{d} \chi^2_{n-p}$$

ポアソン回帰　$Y_i \sim \text{Poisson}(\mu_i)$

$$g(\mu_i) = x_i^t\beta \qquad \mu_i = g^{-1}(x_i^t\beta) = e^{x_i^t\beta}$$

$$\underset{\log(\mu_i)}{\overset{\shortparallel}{g(\mu_i)}}$$

GLM (Generalized linear Model)

$$f(Y_i; \theta_i, \phi) = \exp\{[Y_i\theta_i - b(\theta_i)]/a(\phi) + h(Y_i, \phi)\}$$

$$\mu_i = g^{-1}(x_i^t\beta) \qquad x_i^t\beta = g(\mu_i)$$

Chuyu culture

No.

Date

Chryo culture