Assignment 3. Problem 3. (c).

$AIC = n\log\left(\boxed{\frac{RSS}{n}}\right) + \boxed{2p}$. $BIC = n\log\left(\boxed{\frac{RSS}{n}}\right) + \boxed{2p\log n}$.

AIC / BIC + Forward / Backward may result in different

Occam's Razor

No more assumptions should be made than are necessary.

---

penalty: AIC: $2p$.   BIC: $2p\log n$.

greedy algorithm: choosing the best direction based on current state.

M_0: $y \sim \vec{x}$ $n \gg 3$.

$+X_k$

M_1: $y \sim \vec{x}_{new}\ (\vec{x} + X_k)$. $\log n > 1$.

$$\boxed{\Delta AIC} = n\log\left(\frac{RSS_{new}}{n}\right) + 2(p+1) - n\log\left(\frac{RSS_{old}}{n}\right) - 2p = \boxed{* + 2}$$

$$\boxed{\Delta BIC} = n\log\left(\frac{RSS_{new}}{n}\right) + 2(p+1)\log n - n\log\left(\frac{RSS_{old}}{n}\right) - 2p\log n = \boxed{* + 2\log n}$$

$\boxed{n>1}$ $\boxed{n>3}$. $2\log n > 2$. $\boxed{\Delta BIC > \Delta AIC}$. AIC reduces more.

(AIC) tends to accept more variables!

---

variables provide extra information →

Model	(x1 ~	x2)	(x3 ~	x4)	AIC	BIC
1					435.87	440.75
2	x				-2199.08	-2189.31
3		x			-2062.80	-2053.03
4			x		-2041.97	-2032.20
5				x	-2337.69	-2327.92
6	x	x			-2278.20	-2263.54
7	x		x		-3588.43	-3573.78
8	x			x	-3673.13	-3658.47
9		x	x		-3693.95	-3679.30
10		x		x	-3538.22	-3523.57
11			x	x	-2561.82	-2547.17
12	x	x	x		-3691.95	-3672.41
13	x	x		x	-3685.27	-3665.73
14	x		x	x	-3692.87	-3673.33
15		x	x	x	-3692.48	-3672.94
16	x	x	x	x	-3693.69	-3669.27

$y \sim X_3$.

$y \sim X_2$.
$y \sim X_1$
$y \sim (X_1, X_2)$.
adding X_1 to $y \sim X_2$ does not improve a lot!

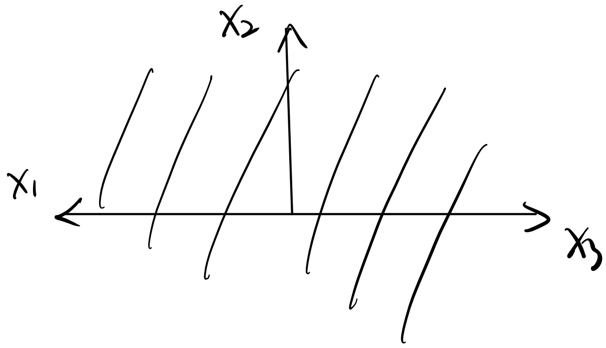
$y \sim X_2$
$y \sim X_3$.
$y \sim (X_2, X_3)$.
adding X_3 to $y \sim X_2$.

$X_2 \sim X_1 / X_3 / X_4$.   compare in pairs

improves a lot to the model.
X_4 is also good.
X_3, X_4 are NOT linearly related to X_2.

$\boxed{X_2 \text{ vs. } X_1}$.

cannot compare (X_1, X_3) vs (X_2, X_3).



$X_3 = -X_1$.

(X_1, X_2) is the same as (X_2, X_3).

X_1 v.s. X_2 NOT correlated.

$\boxed{X_1 \sim X_2}$ X_3, X_4, X.

Explanation for necessity?

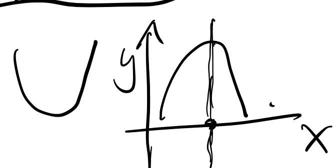
Problem 1.

① statistical evidence
- model diagnostic
- residual plots

② model interpretation

1. actual meaning of variables.
2. reasonable conclusions from model + purpose

1.(d). $\boxed{H_0: \beta_0 + \beta_1 X}$ $H_1: \boxed{\beta_0 + \beta_1 X + \beta_2 X^2}$.



true: $\beta_1 > 0$: $y \to \infty$ as $X \to \infty$

$\beta_1 < 0$: $y \nearrow$ as $X \to 0$.

give 100% of supplement.

give 0% of supplement

weights vs. supplement.

$\boxed{\text{best / optimal}}$ value of amount

Multicollinearity. X is not of full rank.

$(X'X)^{-1}$ NOT exist. $\boxed{r}$ linearly independent

$X = (X_1, X_2, \cdots, X_p)$. $\quad Span(X) = Span\boxed{\{X_1, \cdots, X_r\}} = Span\{X_1, \cdots, X_p\}$.

$X_i \in \mathbb{R}^n$. $\quad rank(X) = r < p$

$\hat{\beta} = \underline{(X^TX)^{-1}}X^TY$. $\quad \hat{Y} = E[Y|X]$.

$\hat{Y} = \boxed{Proj_{C(X)}(Y)}$ $\quad$ for linear space. $\boxed{\text{basis / base}}$.

unique $\hat{\beta}$ $\times$.

$X_0 = (\underline{X_1}, \cdots, \underline{X_r})$.

$Span(X_0) = Span(X)$.

X_0 is of full rank.

$(X_0^TX_0)^{-1}$ exists! $\quad C(X) = C(\underline{X_0})$.

$\hat{Y} = \hat{Y}_0 = Proj_{C(X)}(Y) = Proj_{C(X_0)}(Y)$.

$\boxed{\hat{Y}} = X_0(X_0^TX_0)^{-1}X_0^TY$.

$\hat{\beta}_0 = (X_0^TX_0)^{-1}X_0^TY \in \mathbb{R}^{r\times 1}$. $\quad \hat{\beta} = \boxed{\begin{pmatrix} \hat{\beta}_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}} \in \mathbb{R}^p$.

② generalized inverse. (may not be unique.)
for A is any $\boxed{G}$ such that $\boxed{AGA = A}$. $G = A^-$.

A invertible. $G = A^{-1}$. $AA^{-1}A = A$.

A is symmetric positive semidefinite. $\boxed{v^TAv \geq 0}$.

$$A = P\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}P^T.$$

P is orthogonal. $PP^T = I$
$P^TP = I$.

$$\nu_i = \begin{cases} \frac{1}{\lambda_i} & \lambda_i > 0 \\ 0 & \lambda_i = 0. \end{cases}$$

$$G = P\begin{pmatrix} \nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_n \end{pmatrix}P^T.$$

$$AGA = P\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \end{pmatrix}\overbrace{\cancel{P^T \cdot P}}^{I}\begin{pmatrix} \nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_n \end{pmatrix}\overbrace{\cancel{P^T \cdot P}}^{I}\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}P^T.$$

$$= P\begin{pmatrix} \cancel{\lambda_1\nu_1\lambda_1}\lambda_1 & & & \\ & \cancel{\lambda_2\nu_2\lambda_2}\lambda_2 & & \\ & & \ddots & \\ & & & \cancel{\lambda_n\nu_n\lambda_n}\lambda_n \end{pmatrix}P^T = A.$$

(X^TX) is not invertible. $v^TX^TXv = (Xv)^T \cdot Xv \geq 0$.

$$\hat{Y} = X\boxed{(X^TX)^-}X^TY = X_0(X_0^TX_0)^{-1}X_0^TY.$$

$$\mathrm{Proj}_{C(X)}(Y).$$