STAT 8004 Lecture 2

Thursday, January 22, 2015 4:07 PM

- Gauss-Markov Linear Models

- $$Y_{n\times 1} = X_{n\times p} \beta_{p\times 1} + \epsilon$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ X_{12} & & \vdots \\ \vdots & & \vdots \\ X_{1n} & & X_{pn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$X = (\underset{n\times 1}{X_1}, X_2 \cdots X_p)_{n\times p}$$ predictor variable specific vectors

$$= \begin{pmatrix} \tilde{X}_1^T \\ \tilde{X}_2^T \\ \vdots \\ \tilde{X}_n^T \end{pmatrix}$$

- $Y = X\beta + \epsilon$

- usually, we assume $E(\epsilon) = 0$
  $\Rightarrow E(Y) = X\beta$
- assume $var(\epsilon) = \sigma^2 I$

$\Big\}$ Gauss-Markov model

- Additionally, assume $\epsilon \sim MVN(0, \sigma^2 I)$
  $(\epsilon_1, \epsilon_2 \cdots \epsilon_n \overset{iid}{\sim} N(0, 1))$

—— x ——

Examples

Yield (%) Y	Temperature (^{o}F) X_1	Time (hr) X_2
77	160	1
82	165	3
84	165	2
89	170	1
94	175	2

a) $n = 5$ $p = 2$

a) $n = 5 \quad p = 2$

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \epsilon_i$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_5 \end{pmatrix} \begin{pmatrix} 1 & X_{11} & X_{21} \\ \vdots & & \\ 1 & & \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_5 \end{pmatrix}$$

Thursday, January 22, 2015 5:49 PM

Example (b)

Diet 1	Diet 2	Diet 3
$Y_{11} = 62$	$Y_{21} = 71$	$Y_{31} = 72$
$Y_{12} = 60$		$Y_{32} = 68$
		$Y_{33} = 67$

b)

mean of the i th diet

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

"means" model.

i th Diet

j th rat

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ 1 & & \\ 0 & 1 & 0 \\ & & 1 \\ & 0 & 1 \\ & & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{33} \end{pmatrix}$$

effect

"effects" model

$$Y_{ij} = \underset{\text{baseline}}{\mu} + \alpha_i + \epsilon_{ij}$$

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix} = \begin{pmatrix} 1 & 1 & & \\ 1 & 1 & & \\ 1 & & 1 & \\ 1 & & & 1 \\ 1 & & & 1 \\ 1 & & & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{33} \end{pmatrix}$$

c) ANCOVA (one way)

- additional to diets

- age of the rat (X)

- "Means" $Y_{ij} = \mu_i + \underline{\underline{\gamma}} X_{ij} + \epsilon_{ij}$

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix} \quad \begin{pmatrix} & X_{11} \\ & X_{12} \\ & X_{21} \\ & X_{31} \\ & \vdots \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \gamma \end{pmatrix}$$

Thursday, January 22, 2015 5:59 PM

– (d) Example (d)

Percentage of apples with spots	Variety	Fungicide use
$Y_{111} = 4.6$	A	new
$Y_{112} = 7.4$	A	new
$Y_{121} = 18.3$	A	old
$Y_{122} = 15.7$	A	old
$Y_{211} = 9.8$	B	new
$Y_{212} = 14.2$	B	new
$Y_{211} = 21.1$	B	old
$Y_{222} = 18.9$	B	old

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}$$

2-way ANOVA with interaction

i: variety

j: fung.

k: index of the plot

$$\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{pmatrix}_{8\times 1} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & & & \\ & & & 0 & 1 & & 1 & & \\ & 0 & 1 & 1 & 0 & & & 1 & \\ & & & 0 & 1 & & & & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \delta_{11} \\ \delta_{12} \\ \delta_{21} \\ \delta_{22} \end{pmatrix}$$

$$\boxed{Y = X\beta + \epsilon}$$

——x——

- What is the objective of the linear models?
  - $E(Y) = X\beta$: how to estimate it?
    (mean of the response)
  - element(s) of β
    linear combinations of β $\left(\begin{matrix}\alpha_2-\alpha_1 \\ \mu_2-\mu_1\end{matrix}\right)$
    functions of β
    estimations
  - estimation of σ^2

- confidence intervals for σ^2, for β's l.c

- prediction Y with X_{new}

- prediction intervals

- testing

  $H_0: \beta_{k+1} = \beta_{k+2} = \cdots \beta_p = 0$

- Difficulty. 1) when p is large $p \gg n$

  (Methods IV 8102)

  2) when X is not of full rank.

$$X = (X_1, \cdots X_p) \qquad (n > p)$$

$$\text{rank}(X) < p$$

"ambiguity" "identifiability"

Example :

1) $$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix} = \underbrace{\boxed{\begin{pmatrix} 1 & & \\ 1 & & \\ & 1 & \\ & & 1 \\ & & 1 \\ & & 1 \end{pmatrix}}}_{= X} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \vdots \end{pmatrix}$$

2) $$\boxed{\begin{pmatrix} 1 & 1 & & \\ 1 & 1 & & \\ 1 & & 1 & \\ 1 & & & 1 \\ 1 & & & 1 \\ 1 & & & 1 \end{pmatrix}} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

$$\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} \qquad E\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 3 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} -5 \\ 6 \\ 7 \\ 8 \end{pmatrix} \qquad = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 3 \\ 3 \end{pmatrix}$$

- They have the same "column space"

$$C(X) = \{ z : z = \alpha_1 X_1 + \alpha_2 X_2 + \cdots \alpha_p X_p \}$$

"All linear combinations of the columns in X"

- Remark: $C(X)$ is the "fundamental" thing.

- Estimation of the response mean $E(Y)$ $\left(Y = X\beta + \epsilon \right)$

  - Fact from the linear model $E(Y) \in C(X)$

- $\hat{Y}$: ("best" one in $C(X)$)

- A plausible way is to minimize

$$(Y - \hat{Y})^T (Y - \hat{Y}) \quad \left(\|Y - \hat{Y}\|_2^2 \right)$$
$$= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

for $\hat{Y} \in C(X)$

- $$\boxed{\hat{Y} = \underset{a \in C(X)}{\arg\min} \|Y - a\|_2^2}$$

- example $\begin{pmatrix} Y_{11} \\ Y_{12} \\ \hline Y_{21} \\ \hline Y_{31} \\ Y_{32} \\ Y_{33} \end{pmatrix}$ $\qquad \boxed{\hat{Y} = \begin{pmatrix} a \\ a \\ b \\ c \\ c \\ c \end{pmatrix}}$

general form of an element in $C(X)$

$$(Y-\hat{Y})^T (Y-\hat{Y})$$
$$= (Y_{11}-a)^2 + (Y_{12}-a)^2 + (Y_{21}-b)^2 + (Y_{31}-c)^2 + (Y_{32}-c)^2 + (Y_{33}-c)^2$$

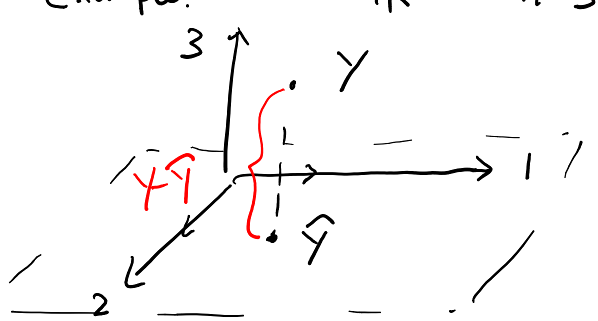
$$a = \frac{Y_{11}+Y_{12}}{2}, \quad b = Y_{21}, \quad c = \frac{Y_{31}+Y_{32}+Y_{33}}{3}$$

Remark: does not change for different X.

— x —

- Since $(Y-\hat{Y})^T(Y-\hat{Y})$ measures the distance between $\hat{Y}$ and Y, so finding $\hat{Y}$ is equivalent to find $\hat{Y} \in C(X)$ that is closest to Y.

- Example: $\mathbb{R}^3$ $\quad n=3$ $\quad Y = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$



$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

1) What is $C(X)$?

2) $\hat{Y} = ?$ $\qquad \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$

3) $Y - \hat{Y} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$

3) $Y - \hat{Y} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$.

- Definition: $\hat{Y}$ is the (perpendicular) projection of Y onto $C(X)$. $(E(Y))$.

- More about $\hat{Y}$.

  1) $\exists$ a unique $n \times n$ matrix P_X such that $\forall\, Y \in \mathbb{R}^n$, $\underline{P_X Y}$ is the projection of Y onto $C(X)$.

  2) $\boxed{P_X} = X \underbrace{(X^T X)^-}\, X^T$ $n \times n$

  $(X^T X)^-$ is $\underline{\text{any}}$ generalized inverse of $(X^T X)$

Remarks: if X is of full rank.

$$(X^T X)^- = (X^T X)^{-1}$$

for any $(X^T X)^-$, $\underline{P_X \text{ gives the same } \hat{Y}}$

- More about generalized inverse.

  1) ginv() of R (MASS)

  $\boxed{\text{2) Generalized inverse of matrix } A \text{ is any matrix } G, \text{ such that } AGA = A}$

  3) Generally, G is not unique.

  There are various algorithm for finding G.

  (R&S Chapter 2)

4). If A is square & non-singular, there exists a unique $G = A^{-1}$.

5) A symmetric & square A has at least one symmetric A^-.

— x —

- Properties of P_X $\left(= X(X^TX)^-X^T \right)$

1) $P_X^T = P_X$ (symmetry

2) $P_X P_X = P_X$ (idempotent)

$$P_X X = X$$

$$\boxed{P_X z = z \quad \text{if } z \in C(X)}$$

$$X^T P_X X = X^T X (X^T X)^- X^T X$$

$$= X^T X$$

$$\Rightarrow \underline{X^T (P_X X - X) = 0}.$$

3) $P_X (P_X Y) = P_X Y$ $(\hat{Y})$

- $X \to C(X) \to P_X$

— x —

- Now, let us consider $Y - \hat{Y}$ (residual)

$$= Y - P_X Y = \underbrace{(I - P_X)}Y$$

1) $(I - P_X)$ is symmetric

2) $(I - P_X)(I - P_X) = I - P_X$ idempotent.

3) $(I-P_X)$ projects Y onto $C(X)^{\perp}$

complementary space of $C(X)$.

- Properties:

1) $(I-P_X)Y$ & $\hat{Y}=P_XY$ are othogonal.

$$((I-P_X)Y)^T P_X Y = 0$$

$$\parallel$$

$$Y^T \underbrace{(I-P_X)P_X}\, Y$$

This means:

$$Y = \hat{Y} + (Y-\hat{Y})$$

is an othogonal decomposition of Y

- Fact: $\operatorname{rank}(X) = \operatorname{rank}(P_X) = tr(P_X)$

$$tr(I_n) = tr(P_X) + tr(I_n - P_X)$$

$$\operatorname{rank}(I_n) = \operatorname{rank}(P_X) + \operatorname{rank}(I_n - P_X)$$

$$\underline{\underline{n}} = \underbrace{\operatorname{rank}(X)} + \operatorname{rank}(I_n - P_X)$$

related to the d.f.

$$Y^T Y = ((\overbrace{P_X + I - P_X}^{I}) Y)^T Y$$

$$= Y^T P_X Y + Y^T (I-P_X) Y$$

$$= Y^T P_X P_X Y + Y^T (I-P_X)(I-P_X) Y$$

$$= Y^T P_X P_X Y + Y^T (I-P_X)(I-P_X) Y$$

$$= \hat{Y}^T \hat{Y} + e^T e$$

Decomposition of the total variation in Y

$$\left(\sum Y_i^2 = \sum \hat{Y}_i^2 + \sum_i (Y_i - \hat{Y}_i)^2 \right.$$

- summary: estimation of the mean response $\underline{\underline{E(Y)}}$

- Estimation of β. (?)

  - full rank case. $X = (X_1, \cdots X_p)_{n \times p}$

    $\text{rank}(X) = p$

    $\Longleftrightarrow$ $X_1, X_2 \cdots X_p$ are linearly independent

  in this case: $P_X = X \underbrace{(X^T X)^{-1}}_{\text{is unique}} X^T$

$$(X^T X)\hat{\beta} = P_X Y = (X^T X)\underline{\underline{(X^T X)^{-1}}} X^T Y$$

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

— ✓ —

If X is not of full rank.

$$X\beta = P_X Y = X \underbrace{(X^T X)^{-} X^T Y}$$

means there are multiple b, such that $Xb = P_X Y$. ($\underline{\underline{\hat{\beta}}}$ is not defined)

- what can be estimated for β? (estimability)

- what can be estimated for β?
(estimability)

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{31} \\ y_{32} \\ y_{33} \end{pmatrix} = \underline{\begin{pmatrix} 1 & & \\ 1 & & \\ & 1 & \\ & & 1 \\ & & 1 \\ & & 1 \end{pmatrix}} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \vdots \end{pmatrix}$$

$$\ldots \begin{pmatrix} 1 & 1 & & \\ 1 & 1 & & \\ 1 & & 1 & \\ 1 & & & 1 \\ 1 & & & 1 \\ 1 & & & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \;-\;-$$

$$(1, 1, 0, 0) \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \underline{\underline{\mu + \alpha_1}} \quad (\alpha_1 ?)$$

- $c^T \beta$

- for what kind of c, such that $c^T\beta$ is estimable.

- when is $c^T\beta$ estimable?

- Rationale:
  Key: $\boxed{P_X Y = \hat{Y}}$ does not have ambiguity.

$a^T \hat{Y}$: does not have ambiguity

$a^T \underbrace{X (X^TX)^- X^T Y}$: does not have ambiguity.

β

$\boxed{a^T X \beta}$

1) if $\overline{a^T X = c^T}$ then $c^T\beta$ is

↳ if $\boxed{a^T X = c^T}$, then $c^T\beta$ is <u>estimable</u>.
(does not have ambiguity)

- Theorem : - $c = X^T a$, then $c^T\beta$ is estimable.

- $c^T\beta$ is estimable if
  1). $\exists\, a$, $a^T X = c^T$
  2) $\underline{\underline{c}}$ is in the column space of X^T.
     $c \in C(X^T)$.

- example
$$\begin{pmatrix} 1 & 1 & & \\ 1 & 1 & & \\ 1 & & 1 & \\ 1 & & & 1 \\ 1 & & & 1 \\ 1 & & & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

Q : is $\mu + \alpha_1$ estimable? $(1,1,0,0)^T \in C(X^T)$
is α_1 .. ? $(0,1,0,0)^T \notin$
is $\alpha_3 - \alpha_2$ .. ? $\underline{(0,0,-1,1)^T} \in$

- Testability. $H_0: c^T\beta = d$
  for estimable $c^T\beta$.

Further $C = \begin{pmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_l^T \end{pmatrix}$

- $c_1^T\beta \cdots, c_l^T\beta$ are estimable,

$\boxed{H_0 : C\beta = d}$,

- $c_1\beta, \dots, c_\ell\beta$ are estimable,

$$\boxed{H_0: \; C\beta = \underset{\ell\times 1}{d}}$$

When is the hypothesis testable?

Condition 1. but it is not enough.

$$\begin{pmatrix} 1 & & \\ 1 & 1 & \\ & & 1 \\ & & 1 \\ & & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

$$H_0: \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Condition 2: $\underline{\text{rank}(C) = \ell}$

— x —

- $\boxed{Y = X\beta + \epsilon}$

1) mean of response (estimation) geometry

2) estimability

testability

-