STAT 8004 Lecture 5

Thursday, February 19, 2015 5:25 PM

Review: normal theory based inference. $Y \sim N(X\beta, \sigma^2 I)$

- OLS of $c^T\beta$ (estimable) is the MLE

$$\boxed{\widehat{c^T\beta} = c^T(X^TX)^- X^TY}$$

- If $Y \sim N(\mu, \Sigma)$, $A_{n\times n}$, such that

$$\boxed{Y \quad E(Y) = X\beta \qquad a^T E(Y) = a^T X\beta = c^T\beta}$$

$A\Sigma$ is idempotent $((A\Sigma)(A\Sigma) = A\Sigma)$

then, $Y^TAY \sim \chi^2_{rank(A)}(\delta)$

$$\delta = \mu^T A \mu$$

$\Rightarrow$ distribution of $\frac{SSE}{\sigma^2} \sim \chi^2_{n-rank(X)}$

{ hypothesis testing for σ^2
{ confidence set for σ^2

— x —

- $\hat{Y} = PY$ & SSE are independent
  - inferences for $c^T\beta$, for
  - H_0: $\underline{\underline{C\beta = d}}$ for $C = \begin{pmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_l^T \end{pmatrix}$

$$\boxed{SS_{H_0}}: \frac{\cancel{\frac{1}{\sigma^2}} (\widehat{C\beta} - d)^T (C(X^TX)^- C^T)^{-1} (\widehat{C\beta} - d)/l \sim \chi^2_l}{\cancel{\frac{1}{\sigma^2}} SSE / n - rank(X)}$$

$$\sim F_{l,\, n-rank(X)}$$

— x —

Regression analysis:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \beta_r x_{ri} + \epsilon_i$$

This takes the form $Y = X\beta + \epsilon$ for $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{pmatrix}_{(r+1)}$

and $X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{r1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{rn} \end{pmatrix} = (1 | X_1 | X_2 | \cdots | X_r)$

and $X = \begin{pmatrix} 1 & \vdots & & \vdots \\ 1 & x_{1n} & & x_{rn} \end{pmatrix} = (1 | X_1 | X_2 | \cdots | X_r)$

conventionally ~~this~~ X is of full rank.

1) $\hat{Y} = PY$ $\quad P_{n\times n}$ projection matrix,

and it is also called "hat matrix" H

its diagonal terms h_{ii} are used as "measure of influence" ($h_{ii} \geq 0$

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

This says in average h_{ii} is about $\frac{r+1}{n}$

It is common as a "rule of thumb" to flag i as influential if $h_{ii} > 2\left(\frac{r+1}{n}\right)$.

(Cook's distance)

Recall $var(\hat{Y}) = \sigma^2 P = \sigma^2 H$

$\Rightarrow var(\hat{y}_i) = \underline{\underline{h_{ii}}}\,\sigma^2$

$\Rightarrow$ estimated std deviation for $\hat{y}_i$ is $\sqrt{h_{ii}}\sqrt{MSE}$

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

$var(e) = \sigma^2 (I - H)$

$var(e_i) = \sigma^2 (1 - h_{ii})$

So, it is common to use standardized residual.

$$e_i^* = \frac{e_i}{\sqrt{MSE}\sqrt{1-h_{ii}}}$$

3) $X_i = (1 | X_1 | X_2 | \cdots | X_i)$, $\quad E(Y) \in C(X_i)$

P_{X_i} : associated projection

$H_0: \beta_{p+1} = \beta_{p+2} = \cdots = \beta_r = 0$; $\quad (X, \ X_p)$

This hypothesis can be written as $H_0: \underline{\underline{C\beta = 0}}$

for $\underset{(r-p)\times(r+1)}{C} = \left(\underset{(r-p)\times(p+1)}{0} \,\middle|\, I_{(r-p)\times(r-p)} \right)$

$\underline{\underline{SS_{H_0}}} = \frac{1}{\sigma^2} (C\hat{\beta})^T \left(C (X^T X)^{-} C^T\right)^{-1} (C\hat{\beta})$.

use F test.

- Another approach is the Full/reduced model formulation

(ANOVA)

$$Y^T I Y = Y^T\left(P_1 + (P_{X_p} - P_1) + (P_X - P_{X_p}) + (I - P_X)\right) Y$$

$$= Y^T P_1 Y + Y^T(P_{X_p} - P_1)Y + Y^T(P_X - P_{X_p})Y + Y^T(I - P_X)Y$$

$$\underbrace{(Y^TY - Y^TP_1Y)}_{SS_{Tot}} = \underbrace{\underbrace{\boxed{Y^T(P_{X_p} - P_1)Y}}_{SSR\ reduced} + \underbrace{Y^T(P_X - P_{X_p})Y}_{SSH_0\ (Fact)}}_{SSR_{full}} + \underbrace{\boxed{Y^T(I - P_X)Y}}_{SSE_{full}}$$

It is common to organize this in an ANOVA table

Source	SS	df
Regression $(X_1, X_2 \ldots X_r)$	SSR_{full}	$rank(P_X - P_1) = r$
" " $(X_1, X_2 \ldots X_p)$	SSR reduced	$rank(P_{X_p} - P_1) = p$
$\boxed{(X_{p+1} \cdots X_r \mid X_1, X_2 \ldots X_p)}$	$\boxed{SSH_0}$	$rank(P_X - P_{X_p}) = r-p$
error	$\boxed{SSE_{full}}$	$\boxed{n-(r+1)}$
	SS_{tot}	$n-1$

Or, in terms of reduction in SS

$$Y^TY = \underset{R(\beta_0)}{Y^T\underline{P_1}Y} + \underset{R(\beta_1\beta_2, \ldots \beta_p \mid \beta_0)}{Y^T(P_{X_p} - P_1)Y} + \underset{R(\beta_{p+1}, \cdots \beta_r \mid \beta_0 \beta_1 \ldots \beta_p)}{Y^T(P_X - P_{X_p})Y} + Y^T(I - P_X)Y$$

P_1: projection using $(\mathbb{1})$

People also talk about "type I" or sequential SS.

$Y^TY =$	Y^TP_1Y	$R(\beta_0)$
	$+ Y^T(P_{X_1} - P_1)Y$	$R(\beta_1 \mid \beta_0)$
	$+ Y^T(P_{X_2} - P_{X_1})Y$	$R(\beta_2 \mid \beta_0 \beta_1)$
	$+ \cdots$	$\vdots$
	$+ Y^T(P_X - P_{X_{r-1}})Y$	$R(\beta_r \mid \beta_0 \cdots \beta_{r-1})$
	$+ Y^T(I - P_X)Y$	SSE

- In this case, for example $R(\beta_2 \mid \beta_0, \beta_1)$ is appropriate for testing $H_0: \beta_2 = 0$ in a model including X_1 & X_2

It is not correct for $\underline{\underline{H_0: \beta_2 = 0}}$ only, in the full model.
$(C^T\beta = 0)$

The proper way for the latter is using the "type IV" SS

let X_{-i} be X with ith column removed,

$$\underset{SSH_0}{\underbrace{Y^T(P_X - P_{X_{-i}})Y}} = R(\underline{\underline{\beta_i}} \mid \beta_0, \beta_1, \cdots \beta_{i-1}, \beta_{i+1} \cdots \beta_r)$$

— x —

- Remarks: - In the sequential / type I SS all pairs are independent.
  - This is actually the Cochran's theorem (Corollary 1 of Theorem 5.6c of R&S)

— x —

Linear model & 2-way factorial experiments

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

y_{ijk} is the kth observation at level i of A
level j of B

- Cell mean model. $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$
- let n_{ij} be # of observations at level i of A & level j of B
- provided $n_{ij} > 0$, every μ_{ij} is estimable.
- We assume $n_{ij} > 0$ for now.

All μ_{ij}s are estimable, then all linear combinations (lc) of them are estimable

- e.g. $\underline{\mu_{i\cdot}} = \frac{1}{J}\sum_{j=1}^{J} \mu_{ij}$ (row i average mean)

$\mu_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I} \mu_{ij}$ (column j " ")

$$\underline{\underline{\mu_{\cdot\cdot}}} = \frac{1}{IJ}\sum_{ij}\mu_{ij} \quad (\text{overall average mean})$$

Remark: - They are all estimable

- $\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}$ (main effect of A at level i)

- $\alpha_i - \alpha_{i'} = \mu_{i\cdot} - \mu_{i'\cdot}$

(difference in the main effect of A between the ith & jth level)

Similarly

$$\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$$

$$\beta_j - \beta_{j'} = \mu_{\cdot j} - \mu_{\cdot j'}$$

- $\underline{\underline{\delta_{ij}}} = (\mu_{ij} - (\mu_{\cdot\cdot} + \alpha_i + \beta_j))$

(interaction of A & B at level i & j)

$$= \underline{\underline{(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot})}}$$

(These definitions imply:

$$\sum_i \alpha_i = 0$$

$$\sum_j \beta_j = 0$$

$$\sum_i \delta_{ij} = 0 \quad \text{for any } j$$

$$\sum_j \delta_{ij} = 0 \quad \text{for any } i$$

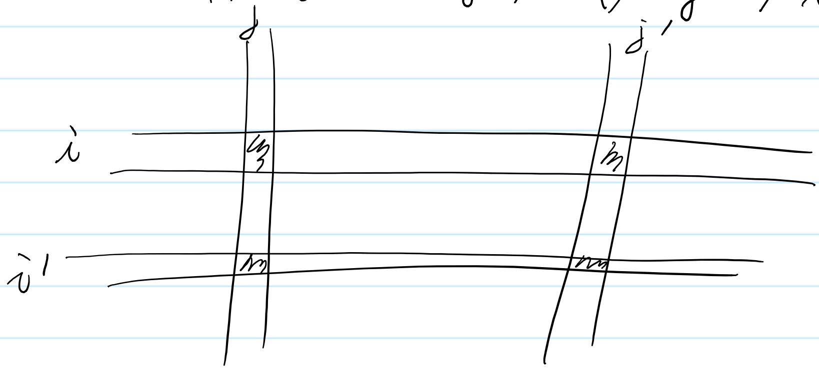
Question: whether or not $\underline{\delta_{ij} = 0}$.

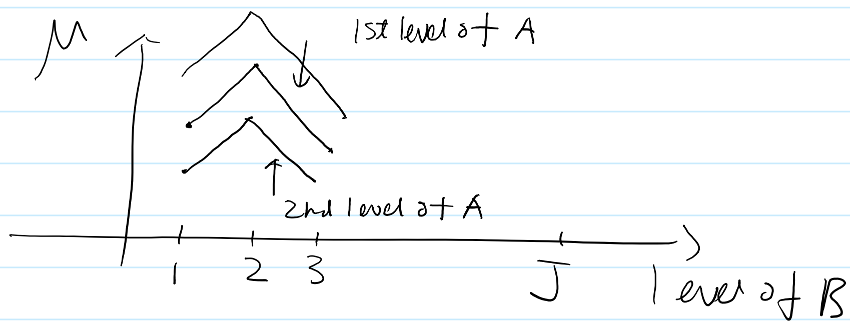
This $\Rightarrow$ $\left\{ \begin{array}{l} \mu_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j \\ \text{or} \\ \mu_{ij} = \mu_{i\cdot} + \mu_{\cdot j} - \mu_{\cdot\cdot} \end{array} \right]$ $\leftarrow$ what do they or

$$\mu_{ij} = \mu_{i\cdot} + \mu_{\cdot j} - \mu_{\cdot\cdot}$$ ← what do they mean?

This implies

$$(\mu_{ij} - \mu_{ij'}) - (\mu_{i'j} - \mu_{i'j'}) = 0$$





"all $\delta_{ij} = 0$" is saying that the "means" are parallel.

· Now. how to test $H_0: \delta_{ij} = 0$ for all i, j.

WAY 1). using the cell mean model

the statement $\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot} = 0$

$(I-1)(J-1)$ of them.

- This is actually to test

$$H_0: \underset{(I-1)(J-1)\times IJ}{C}\beta = 0$$

- It perhaps is more natural to view this problem from an "effect model" point view.

rather than $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, we could use $Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}$

- $k = 1 + I + J + IJ$: parameters,
- the design matrix will not be of full rank.
- To get simple calculations (in software), restrictions are necessary.

Sum restriction	Baseline restrictions	
$\sum \alpha_i = 0$	$\alpha_I = 0$	$\alpha_1 = 0$
$\sum \beta_j = 0$	$\beta_J = 0$	$\beta_1 = 0$
$\sum_i \delta_{ij} = \sum_j \delta_{ij}$	$\delta_{Ij} = 0$ for all j	$\delta_{1j} = 0$ for all j
	$\delta_{iJ} = 0$ for all i	$\delta_{i1} = 0$ for all i
	SAS	R

Example: 2×3 version with $n_{ij} = 1$

	1	2	3
1			
2			

Cell means

$$Y = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{pmatrix} = \underline{\underline{I}} \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{pmatrix} + \epsilon$$

for effect with sum constraints

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ \delta_{11} \\ \delta_{12} \end{pmatrix}$$

$E(y_{11}) = \mu + \alpha_1 + \beta_1 + \delta_{11}$

$\vdots$

$E(y_{23}) = \cdots$

baseline constraints (SAS)

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ \delta_{11} \\ \delta_{12} \end{pmatrix}$$

$E(y_{23}) = \delta_{12}$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ \vdots \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \\ \vdots \end{pmatrix}$$

$$\left(\mathbb{1}\;\begin{matrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{matrix}\right)\begin{pmatrix} \alpha_2 \\ \beta_2 \\ \beta_3 \\ \delta_{22} \\ \delta_{23} \end{pmatrix}$$

—— x ——

1) use l.c. of cell means

e.g. $\delta_{11} = \frac{1}{3}\mu_{11} - \frac{1}{6}\mu_{12} - \frac{1}{6}\mu_{13} - \frac{1}{3}\mu_{21} + \frac{1}{6}\mu_{22} + \frac{1}{6}\mu_{23}$

(convince yourself)

then use $H_0: C'\beta = 0$.

2) use regression analysis

$$X = \big(\underset{6\times1}{\mathbb{1}} \,\big|\, \underset{6\times1}{X_\alpha} \,\big|\, \underset{6\times2}{X_\beta} \,\big|\, \underset{6\times2}{X_\delta}\big)$$

and consider $H_0: E(Y) \in C(\mathbb{1}, X_\alpha, X_\beta)$

$$SS_{H_0} = Y'(P_X - P_{(\mathbb{1}, X_\alpha, X_\beta)})Y$$

with d.f. $(I-1)(J-1)$

Can use ANOVA to test for <u>interaction</u>.

- Remark: it is not clear what to expect in case, when $n_{ij} \neq m$ for all ij

$H_0: \alpha_i = 0$, may not give you what think you are testing.

example. 2×3 $n_{ij} = 1$

A	1	$\mu+\alpha_1+\beta_1+\delta_{11}$	$\mu_1+\alpha_1+\beta_2+\delta_{12}$	$\mu+\alpha_1$
	2	$\mu+\beta_1$	$\mu+\beta_2$	μ

$H_0: \alpha_1 = 0$, this is actually testing

$\mu_{13} = \mu_{23}$

Sum of squares & ANOVA

$$X = (\mathbb{1} \,|\, X_\alpha \,|\, X_\beta \,|\, X_\delta)$$

ANOVA table

Source	SS (type I)	SS (type II)	SS (type III)
A	$R(\alpha\|\mu)$	$R(\alpha\|\mu,\beta)$	SSH_0^{α}
B	$R(\beta\|\mu,\alpha)$	$R(\beta\|\mu,\alpha)$	SSH_0^{β}
A×B	$R(\delta\|\mu,\alpha,\beta)$	$R(\delta\|\mu,\alpha,\beta)$	SSH_0^{δ}
	SS_{tot}.		

1). what is it testing?

2) We need to be really careful for the practical interpretations of the tests ($C(X)$).

Remark: - in balanced case $n_{ij} = m$

all three types of SS are equivalent.

- But it is not ~~of~~ o.w. in un-balanced cases

For un-balanced data: $H_0: C\beta = 0$

- incompelte two-way factorial experiment

μ_{11}	μ_{12}	μ_{13}
μ_{21}	μ_{22}	μ_{23}

model of the means

—	—	empty
—	—	—

- With no (1,3) Combination, it is impossible to discuss inference for μ_{13}.

- For example: $\underline{\underline{\alpha_1 - \alpha_2}} = \mu_{1.} - \mu_{2.}$

$$= \frac{1}{3}(\mu_{11} + \mu_{12} + \boxed{\mu_{13}}) - \frac{1}{3}(\mu_{21} + \mu_{22} + \mu_{23}) \quad X$$

- Remark: is still about $C(X)$.

- But one could do inference for

$$\frac{1}{2}(\mu_{11}+\mu_{12}) - \frac{1}{2}(\mu_{21}+\mu_{22}) \checkmark$$

- It is about "what kind of problems" we can still handle.
  - $(\mu_{12}-\mu_{11}) - (\mu_{22}-\mu_{11})$ ✓

- How to handle interactions δ_{ij}?

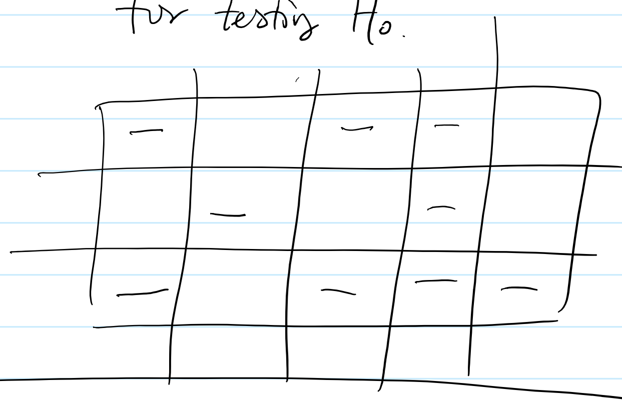
H_0: there is no estimable interactions.

- let X be the cell mean design matrix, (for k cells)

let $X^* = (\underset{n\times 1}{\mathbb{1}} \mid \underset{n\times(I-1)}{X_\alpha} \mid \underset{n\times(J-1)}{X_\beta})$ if X^* is of full rank from an effect model

$$F = \frac{Y'(P_X - P_{X^*})Y/(k-(I+J-1))}{Y'(I-P_X)Y/(n-k)}$$

for testing H_0.



this is still ok.

- H_0: $C\beta = d$ for testable C.

- Feb 26 - 5:30pm - 7:30pm A 745

  - Open note, Open book.
  - Calculator (non-programmable)

Coverage Lecture 1 - 5.