M 6

Ronald 3.1 - 3.5

* Now we will talk about **testing**.

likelihood ratio test

** Testing linear parametric functions (first principles test)

** Testing models

** Confidence intervals and multiple comparisons

- **Ex 1** (Interaction) Consider two-way crossed model w/ interactions

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad \begin{array}{l} i=1,\dots,a(=2) \\ j=1,\dots,b(=2) \\ k=1,\dots,n_{ij}(=2) \end{array}$$

$$X = \begin{array}{c} \begin{array}{ccccccccc} \mu & \alpha_1 & \alpha_2 & \beta_1 & \beta_2 & \gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{22} \end{array} \\ \left[\begin{array}{ccccccccc} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]_{8\times 9} \end{array}$$

X_0

Goal: test for no interactions

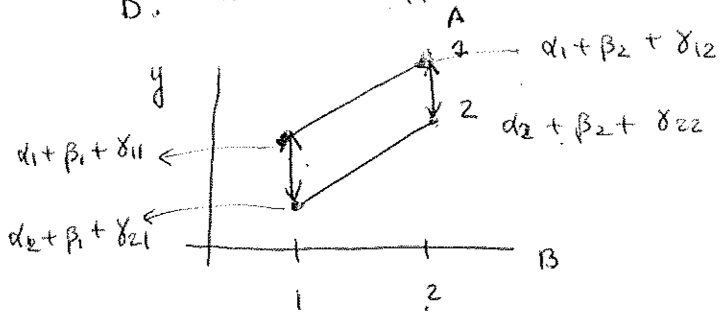
A. model comparison

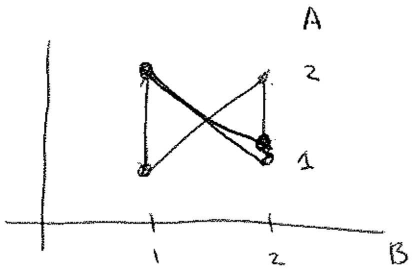
(reduced) $H_0: y = X_0\beta_0 + e$ vs (full model) $H_1: y = X\beta + e$

$$r(X_0) = 3 = a+b-1 = 2+2-1 = 3$$

$$r(X) = 4 = a\times b = 2\times 2 = 4$$

B. linear hypothesis





$$(\cancel{\alpha_1}+\cancel{\beta_1}+\gamma_{11}) - (\cancel{\alpha_2}+\cancel{\beta_1}+\gamma_{21})$$

$$= (\cancel{\alpha_1}+\cancel{\beta_2}+\gamma_{12}) - (\cancel{\alpha_2}+\cancel{\beta_2}+\gamma_{22})$$

$$\gamma_{11} - \gamma_{21} = \gamma_{12} - \gamma_{22}$$

$H_0: (\gamma_{11} - \gamma_{21}) - (\gamma_{12} - \gamma_{22}) = 0$ $\qquad H_a: (\gamma_{11} - \gamma_{21}) - (\gamma_{12} - \gamma_{22}) \neq 0$

$\Leftrightarrow \quad H_0: \lambda^T\beta = 0$ where $\lambda = [\,0\ 0\ 0\ 0\ 0\ 1\ {-1}\ {-1}\ 1\,]^T$

vs $H_a: \lambda^T\beta \neq 0$

Ex 3

One-way ANOVA.

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad i = 1, 2, 3, \quad j = 1, \dots, \underline{\underline{n_i}}_{=2}$$

(a) Want to test the hypothesis of a linear effect w/ the group covariate $x_i = i$ (linear trend)

A. model comparison

$H_0: y_{ij} = \mu + i\alpha + e_{ij}$ $\qquad H_a: y_{ij} = \mu + \alpha_i + e_{ij}$

$$X_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \qquad X = \begin{array}{c} \left[\begin{array}{c|ccc} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{array}\right] \\ \begin{array}{cccc} X_0 & & X_1 & \end{array} \end{array}$$

$\mathrm{rank}(X_0) = 2$ $\qquad \mathrm{rank}(X) = 3$

B. Linear Hypothesis

$H_0: \alpha_1 - \alpha_2 = \alpha_2 - \alpha_3$ $\qquad \underline{\underline{\lambda}}^T\beta = 0 \quad \lambda^T = [\,0\ 1\ {-2}\ 1\,]$

vs $H_1: \alpha_1 - \alpha_2 \neq \alpha_2 - \alpha_3$ $\qquad \lambda^T\beta \neq 0$

(b) Want to test that the effects α_i are all equal

A: model comparison $\qquad \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \begin{matrix} \beta_0 \\ \beta_1 \end{matrix}$

$H_0: y_{ij} = \mu + e_{ij}$ vs $H_a: y_{ij} = \mu + \alpha_i + e_{ij}$

$\mathrm{rank}(X_0) = 1$ $\qquad \mathrm{rank}(X) = 3$

B. linear hypothesis

$H_0: \alpha_1 = \alpha_2$ & $\alpha_1 = \alpha_3$ vs $H_a: \alpha_1 \neq \alpha_2$ $\underline{\text{or}}$ $\alpha_1 \neq \alpha_3$

$$\begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\Rightarrow H_0: \underset{\uparrow}{\Lambda}^T \beta = 0 \qquad H_a: \Lambda^T\beta \neq \underset{\sim}{0}$

$\Lambda: s \times p = 2 \times 4$

- $\underline{\text{Ex }2}$ (General subset)

$$X\beta = [X_0 \mid X_1] \begin{bmatrix} \beta_0 \\ \hline \beta_1 \end{bmatrix} \begin{matrix} p_0 \times 1 \\ p_1 \times 1 \end{matrix} \qquad p = p_0 + p_1$$

Test if $\beta_1 = 0$

A. Model Comparison

$H_0: y = X_0\beta_0 + e$ vs $H_1: y = X\beta + e$

$\underbrace{\text{rank}(X_0) = r_0 \qquad\qquad \text{rank}(X) = r}$

$\Rightarrow \underline{\underline{r - r_0 = s}}$

B. Linear Hypothesis

$H_0: \overbrace{(I - P_{X_0})X_1}^{\Lambda^T}\beta_1 = 0$ vs $H_1: (I - P_{X_0})X_1\beta_1 \neq 0$

why? ⊚ $\mathcal{C}(\underbrace{P_X - P_{X_0}}_{\uparrow}) = \mathcal{C}(X_0)^{\perp}_{\mathcal{C}(X)} = \mathcal{C}(\underbrace{(I - P_{X_0})}\underset{\cancel{X_0}\ X_1}{X}) = \underbrace{\mathcal{C}((I - P_{X_0})X_1)}_{\circledast \ \text{rank}(\circledast) = s}$

the perpendicular projection operator onto orthogonal complement of $\mathcal{C}(X_0)$ with respect to $\mathcal{C}(X)$

♣ Given the Gauss-Markov model with normal errors, $\mathbf{y} \sim N_n(\mathbf{X}\beta, \sigma^2 I)$, let's consider the **general *linear* hypothesis**.

$*$ Consider a system of linear equations:

$$H_0 : \Lambda^T \beta = \mathbf{m} \qquad \text{vs} \qquad H_1 : \Lambda^T \beta \neq \mathbf{m}$$

- Λ is $p \times s$ with full-column rank (to avoid redundancy in writing hypotheses).
- Each component of $\Lambda^T \beta$, $\boldsymbol{\lambda}^T \beta$ is estimable ($\boldsymbol{\lambda} \in C(\mathbf{X}^T)$)

♣ Let's do **testing models.**

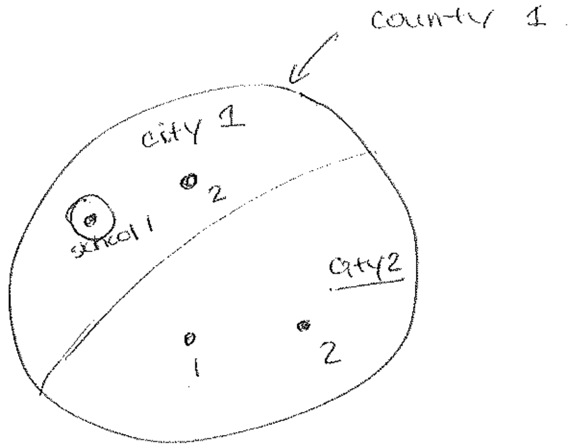
Start with a model that we assume valid

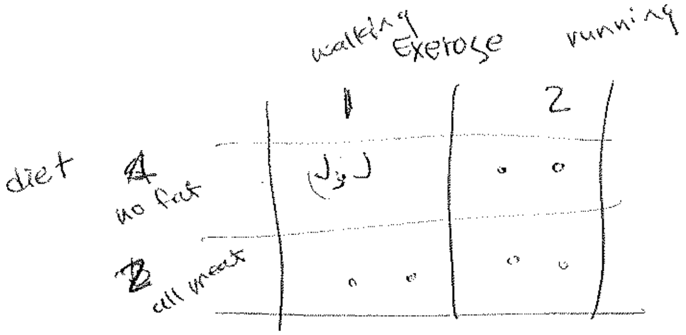
$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I). \tag{1}$$

Our wish is to reduce this model (i.e. simpler model, putting more constraints on the estimation space)

$$\mathbf{y} = \mathbf{X}_0\gamma + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I) \qquad C(\mathbf{X}_0) \in C(\mathbf{X}). \tag{2}$$

Note: if the reduced model is correct, the big model is also correct. **Our question is whether the reduced model is correct.**


county 1
city 1
2
school 1
city2
1
2


walking
Exercise
running
1
2
diet
1
no fat
2
all meat
(i,j)

♣ Given the Gauss-Markov model with normal errors, $\mathbf{y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$, let's consider the **general *linear* hypothesis**.

∗ Consider a system of linear equations:

$$\Lambda^T \beta = 0$$

$$\text{rank}(\Lambda) = s \quad H_0 : \Lambda^T \boldsymbol{\beta} = \mathbf{m} \qquad \text{vs} \qquad H_1 : \Lambda^T \boldsymbol{\beta} \neq \mathbf{m}$$

- $\underline{\Lambda}$ is $p \times s$ with full-column rank (to avoid redundancy in writing hypotheses). $\Lambda = [\lambda_1 \vdots \lambda_2 \cdots \vdots \lambda_s]$
- Each component of $\Lambda^T \boldsymbol{\beta}$, $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ is estimable ($\boldsymbol{\lambda} \in C(\mathbf{X}^T)$)

$\Lambda : p \times s \qquad r(\Lambda) = s \leq r$

† Def 6.1 The general linear hypothesis $H_0 : \Lambda^T \boldsymbol{\beta} = \mathbf{m}$ is **testable** iff Λ has full-column rank and each component of $\Lambda^T \boldsymbol{\beta}$ is estimable. If any of the components of $\Lambda^T \boldsymbol{\beta}$ are not estimable, then the hypothesis is considered **nontestable**.

$*$ Intuition?

Estimation space = $\mathcal{C}(X)$

$*$ Ex3 (contd): Consider the one-way ANOVA model;

$$y_{ij} = \mu + \alpha_i + e_{ij}, i = 1, 2, 3 \quad \text{and} \quad j = 1, 2.$$

$\lambda_1 \in \mathcal{C}(X^T)$ estimable $H_0: \lambda_1^T \beta = 0$

- $\underline{\boldsymbol{\lambda}_1^T = [0, 1, 0, -1]}$ $\alpha_1 - \alpha_2 = 0$ $\Rightarrow$ reduce $r(X)$

nonestimable

- $\underline{\boldsymbol{\lambda}_2^T = [0, 1, 1, 1]}$ $H_0: \lambda_2^T \beta = 0$ does not reduce $r(X)$

$\lambda_2 \notin \mathcal{C}(X^T)$

$\alpha_1 + \alpha_2 + \alpha_3 = 0$

$\boxed{\sum \alpha_i = 0}$

$\ast$ Recall: we found that the BLUE of $\Lambda^T\boldsymbol{\beta}$, $\Lambda^T\hat{\boldsymbol{\beta}}$ follows

$$\Lambda^T\hat{\boldsymbol{\beta}} \sim \underline{\mathrm{N}_s}(\underbrace{\Lambda^T\boldsymbol{\beta}}, \sigma^2\underbrace{\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda}_{\equiv H}).$$

$= \Lambda^T(X^TX)^- X^T y$

● Result 6.3 If $\Lambda^T\boldsymbol{\beta}$ is estimable, then $s \times s$ matrix $\mathbf{H} = \Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda$ is nonsingular. $\Rightarrow$ $\Lambda^T\hat{\boldsymbol{\beta}}$ is nonsingular!

$H_0: \Lambda^T\beta = m$

$*$ We have $\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m} \sim \mathrm{N}_s(\Lambda^T\boldsymbol{\beta} - \mathbf{m}, \sigma^2\mathbf{H})$.

$$\Rightarrow \quad \underline{(\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m})^T(\sigma^2\mathbf{H})^{-1}(\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m})} \sim \chi^2(\underset{\uparrow}{s}, \underset{\uparrow}{\phi})$$

where $\phi = \frac{1}{2}(\Lambda^T\boldsymbol{\beta} - \mathbf{m})^T(\sigma^2\mathbf{H})^{-1}(\Lambda^T\boldsymbol{\beta} - \mathbf{m})$

H: nonsingular $\Rightarrow$ H^{-1} : positive definit

$x^T(H^{-1})x > 0$ for all $x \neq 0$

$*$ $\phi > 0$ for $(\Lambda^T\boldsymbol{\beta} - \mathbf{m}) \neq \mathbf{0}$. Why?

$$\underset{\underset{H_0}{\uparrow}}{\Lambda^T\beta - m = 0} \longrightarrow \phi = 0$$

$$\frac{(\Lambda^T\hat\beta - m)^T H^{-1} (\Lambda^T\hat\beta - m)}{\sigma^2} \sim \chi^2(s, \phi)$$

indep

$SSE = Y^T(I-P)Y \qquad \frac{(X\beta)^T(I-P)(X\beta)}{2\sigma^2}$

$$\frac{Y^T(I-P)Y}{\sigma^2} \sim \chi^2(n-r, 0) \perp \Lambda^T\hat\beta$$

$$\frac{\frac{(\Lambda^T\hat\beta - m)^T H^{-1}(\Lambda^T\hat\beta - m)}{\cancel{\sigma^2}} / s}{\frac{Y^T(I-P)Y}{\cancel{\sigma^2}} / (n-r)} \sim F(s, n-r, \phi)$$

$$F = \frac{\dfrac{(\Lambda^T\hat{\beta} - m)^T H^{-1} (\Lambda^T\hat{\beta} - m)}{s}}{\underset{\substack{\| \\ MSE}}{\dfrac{SSE}{(n-r)}}} \sim F(s,\ n-r,\ \phi)$$

$\Rightarrow$ under $H_0: \Lambda^T\beta = m$, $\phi = 0$

$\Leftrightarrow$ under H_0; $F \sim F(s, n-r)$

reject H_0 if $F > F_\alpha(s, n-r)$

$*$ Consider

$$F = \frac{(\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m})^T(\sigma^2\mathbf{H})^{-1}(\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m})/s}{SSE/(n-r)} \quad \sim F(s, n-r, \phi)$$

Q: What is the distribution of F?

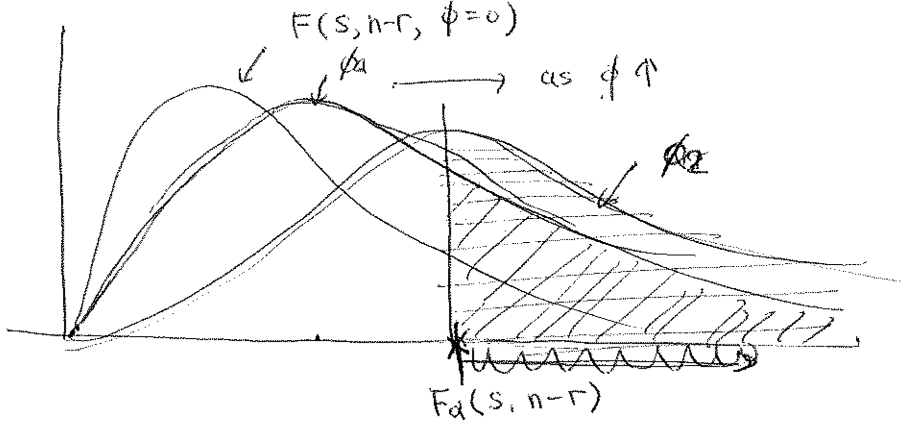
$*$ Resulting test procedure:
Reject H_0 if

$$F = \frac{(\Lambda^T \hat{\boldsymbol{\beta}} - \mathbf{m})^T (\sigma^2 \mathbf{H})^{-1} (\Lambda^T \hat{\boldsymbol{\beta}} - \mathbf{m})/s}{SSE/(n-r)} > F_\alpha(s, n-r)$$

$**$ test with level α

$**$ increasing power when the alternative is true

$$\phi \cancel{\propto} \frac{(\Lambda^T\beta - m)^T H^{-1} (\Lambda^T\beta - m)}{2\sigma^2}$$



power = P(reject Ho when Ho is not true)

♣ Let's consider $s = 1$, that is, test $H_0 : \boldsymbol{\lambda}^T\boldsymbol{\beta} = m$.

$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - m \sim \mathrm{N}(\boldsymbol{\lambda}^T\boldsymbol{\beta} - \mathbf{m}, \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^{-}\boldsymbol{\lambda}).$$

$$\Rightarrow \quad (\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - m)/\sqrt{\sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^{-}\boldsymbol{\lambda}} \sim \mathrm{N}(\boldsymbol{\lambda}^T\boldsymbol{\beta} - \mathbf{m}, 1)$$

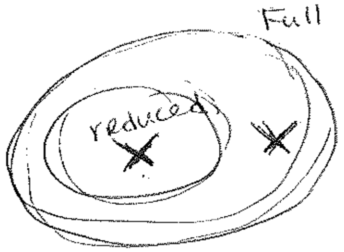
$*$ Consider

$$t = (\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - m)/\sqrt{\hat{\sigma}^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^{-}\boldsymbol{\lambda}}$$

Q: What is the distribution of t? $t\left(n-r, \phi = \frac{\lambda^T\beta - m}{\sqrt{\sigma^2\lambda^T(X^TX)^-\lambda}}\right)$

under H_0: $\lambda^T\beta = m$, $\phi = 0$

Roject H_0 if $|t| > t_{\alpha/2}(n-r)$



♣ Let's do **testing models**.

Start with a model that we assume valid

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I). \qquad \text{Full model} \tag{3}$$

Our wish is to reduce this model (i.e. simpler model, putting more constraints on the estimation space)

$$\mathbf{y} = \mathbf{X}_0\gamma + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I) \qquad C(\mathbf{X}_0) \overset{\subset}{\bullet} C(\mathbf{X}). \tag{4}$$

Note: if the reduced model is correct, the big model is also correct. **Our question is whether the reduced model is correct.**

$\ast$ Ex3 (contd): Consider the one-way ANOVA model;

$$y_{ij} = \mu + \alpha_i + e_{ij}, i = 1, 2, 3 \quad \text{and} \quad j = 1, 2.$$

- Define the reduced model to test for no treatment effects.

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \qquad X_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

- Define the reduced model to test $\alpha_1 - \alpha_3 = 0$. $\Leftrightarrow \alpha_1 = \alpha_3$

$$X_0 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

$*$ Ex4: Consider the full multiple regression the one-way ANOVA model;

$$y_i = \beta_0 + \beta_1 \cancel{x}_{1i} + \beta_2 x_{2i} + \cancel{\beta_3 x_3}_i + e_i, i = 1,2,3 \text{ and } j = 1,2.$$

- Define the reduced model to test whether $\underline{x_1}$ and $\underline{x_3}$ are adding significantly to the explanatory capability of the regression model.

$$y_i = \gamma_0 + \gamma_2 x_{2i} + e_i$$

- Define the reduced model to test $\beta_2 - \beta_3 = 0$. ($\beta_2 = \beta_3$)

$$y_i = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 (x_{2i} + x_{3i}) + e_i$$

$$\gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

$\ast$ Consider the following hypothesis.

$$H_0 : \mathrm{E}(\mathbf{y}) = \underline{\mathbf{X}_0}\gamma \quad \text{for some } \gamma \qquad \Leftrightarrow \qquad H_0 : \mathrm{E}(\mathbf{y}) \in C(\mathbf{X}_0)$$

versus

$$H_1 : \underset{= X\beta}{\mathrm{E}(\mathbf{y})} \in C(\mathbf{X}) \qquad \text{and} \qquad \mathrm{E}(\mathbf{y}) \notin C(\mathbf{X}_0)$$

$$E(y) = X\beta \quad \Rightarrow \quad \hat{y} = P_x y$$

Q: How to build a test statistic?

- Recall! Let $\mathbf{P}$ and $\mathbf{P}_0$ be the perpendicular projection operators onto $C(\mathbf{X})$ and $C(\mathbf{X}_0)$, respectively.

With $C(\mathbf{X}_0) \subset C(\mathbf{X})$, $\mathbf{P}-\mathbf{P}_0$ is the perpendicular projection operator onto the orthogonal complement of $C(\mathbf{X}_0)$ with respect to $C(\mathbf{X})$, that is, $C(\mathbf{P}-\mathbf{P}_0) = C(\mathbf{X}_0)^{\perp}_{C(\mathbf{X})}$.

- $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$ and $\hat{\mathbf{y}} = \mathbf{P}_0\mathbf{y}$ are estimates of $\mathrm{E}(\mathbf{y})$ under models (1) and (2), respectively.

$E(y) = X_0 \cdot \gamma$

$E(y) = X\beta$

reduced model

- **If model (2) is true,** $X_0\gamma = X\beta$

$*$ $\mathbf{P}\mathbf{y}$ and $\mathbf{P}_0\mathbf{y}$ are estimates of the same quantity

$*$ $\mathrm{E}(\mathbf{P} - \mathbf{P}_0)\mathbf{y} = \mathbf{0}$.

$$\Rightarrow \text{small } \mathbf{P}\mathbf{y} - \mathbf{P}_0\mathbf{y} = (\mathbf{P} - \mathbf{P}_0)\mathbf{y}$$

- **If $\overbrace{\textbf{model (2)}}^{\text{reduced model}}$ is NOT true,**

$*$ $\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta \in C(\mathbf{X}) = C(\mathbf{P})$

$*$ $\mathbf{P}_0\mathrm{E}(\mathbf{y}) = \mathbf{P}_0\mathbf{X}\beta \neq \mathrm{E}(\mathbf{y})$

$*$ $\mathbf{Py}$ and $\mathbf{P}_0\mathbf{y}$ estimate different things.

$\Rightarrow$ large $\mathbf{Py} - \mathbf{P}_0\mathbf{y} = (\mathbf{P} - \mathbf{P}_0)\mathbf{y}$

$$\overset{n\times 1}{\swarrow}$$

- The difference is $\mathbf{P}\mathbf{y} - \mathbf{P}_0\mathbf{y} = (\mathbf{P} - \mathbf{P}_0)\mathbf{y}$.

$*$ The length of the difference is $\|(P-P_0)y - 0\|^2 = \mathbf{y}^T(\mathbf{P} - \mathbf{P}_0)\mathbf{y}$

$$= \|(P-P_0)y\|^2$$

$$= (\underbrace{(P-P_0)}y)^T((P-P_0)y)$$

$$= y^T(P-P_0)y$$

$*$ The distribution of $\mathbf{y}^T(\mathbf{P} - \mathbf{P}_0)\mathbf{y}$?