

Recap: We have discussed how to carry out hypothesis testing $H_0: \underline{\lambda}^T \underline{\beta} = m$ vs. $H_1: \underline{\lambda}^T \underline{\beta} \neq m$

We have discussed the usage of t -test and F -test in the last class.

Testing models:

start with a model $\underline{y} = \underline{X} \underline{\beta} + \underline{e}$, $\underline{e} \sim N_n(\underline{0}, \sigma^2 \underline{I})$

our wish is to reduce this model (i.e. use simpler model)

$$\underline{y} = \underline{X}_0 \underline{\gamma} + \underline{e}, \quad \underline{e} \sim N_n(\underline{0}, \sigma^2 \underline{I}), \quad C(\underline{X}_0) \subset C(\underline{X}).$$

Ex: One way ANOVA

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

$$\underline{X} = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}, \quad \text{let } \underline{X}_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\underline{Ex:} \quad y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \quad i=1,2,3$$

Define a reduced model that sets $\beta_2 = \beta_3$

\Rightarrow reduced model is

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

$$\underline{X}_0 = \begin{bmatrix} 1 & x_{11} & x_{21} + x_{31} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} + x_{3n} \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{bmatrix}$$

Consider the following hypothesis

$$H_0: E[\underline{y}] = \underline{x}_0 \beta \text{ for some } \beta \Leftrightarrow H_0: E[\underline{y}] \in C(\underline{x}_0)$$

vs.

$$H_1: E[\underline{y}] \in C(\underline{x}) \text{ and } E[\underline{y}] \notin C(\underline{x}_0)$$

Qn: How to build a test statistic?

Recall: Let \underline{P} and \underline{P}_0 be the perpendicular projection matrices onto $C(\underline{x})$ and $C(\underline{x}_0)$ respectively.

with $C(\underline{x}_0) \subset C(\underline{x})$, what is the projection matrix onto the orthogonal complement of $C(\underline{x}_0)$ w.r.t. $C(\underline{x})$?

$$\text{It is } \underline{P} - \underline{P}_0$$

If the null hypothesis is true, then

~~the~~ $\hat{\underline{y}}$ under the big model is $\hat{\underline{y}} = \underline{P}\underline{y}$ and $\hat{\underline{y}}$ under the reduced model is ~~the~~ $\underline{P}_0 \underline{y}$

Thus, if the reduced model is "close" to the full model $\underline{P}\underline{y} - \underline{P}_0 \underline{y}$ must be small.

$$\begin{aligned} \text{the length of the difference is} &= \|(\underline{P}\underline{y} - \underline{P}_0 \underline{y})\|^2 \\ &= (\underline{P}\underline{y} - \underline{P}_0 \underline{y})' (\underline{P}\underline{y} - \underline{P}_0 \underline{y}) = \underline{y}' (\underline{P} - \underline{P}_0)' (\underline{P} - \underline{P}_0) \underline{y} \\ &= \underline{y}' (\underline{P} - \underline{P}_0) \underline{y} \end{aligned}$$

$$\frac{\underline{y}'(\underline{P}-\underline{P}_0)\underline{y}}{\sigma^2} \sim \chi^2(\text{rank}(\underline{P}-\underline{P}_0), \frac{(\underline{X}\underline{\beta})'(\underline{P}-\underline{P}_0)(\underline{X}\underline{\beta})}{2\sigma^2})$$

$$M = \frac{\left\{ \frac{\underline{y}'(\underline{P}-\underline{P}_0)\underline{y}}{\sigma^2} \right\} / \text{rank}(\underline{P}-\underline{P}_0)}{\left\{ \frac{\underline{y}'(\underline{I}-\underline{P})\underline{y}}{\sigma^2} \right\} / (n - \text{rank}(\underline{X}))} \sim F(\text{rank}(\underline{P}-\underline{P}_0), n - \text{rank}(\underline{X}), \underbrace{\frac{(\underline{X}\underline{\beta})'(\underline{P}-\underline{P}_0)(\underline{X}\underline{\beta})}{2\sigma^2}}_{\phi})$$

M follows a non-central F with non-centrality parameter as ϕ .

when H_0 is true, $\phi = 0$

thus under H_0 , M follows a central F distribution with $\text{rank}(\underline{P}-\underline{P}_0)$, $n - \text{rank}(\underline{X})$ degrees of freedom.

Summarize:

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e}, \quad \underline{e} \sim N_n(\underline{0}, \sigma^2 \underline{I})$$

when the reduced model is $\underline{y} = \underline{X}_0 \underline{\gamma} + \underline{\tilde{e}}$,

$$\underline{\tilde{e}} \sim N_n(\underline{0}, \sigma^2 \underline{I}), \quad C(\underline{X}_0) \subset C(\underline{X})$$

\underline{P} , \underline{P}_0 are projection matrices onto $C(\underline{X})$ and $C(\underline{X}_0)$ respectively, then use the test statistic

M to test if the reduced model is legitimate.

Now, just recall the hypothesis testing

$$H_0: \underline{\Lambda}^T \underline{\beta} = \underline{m}, \quad \text{vs.} \quad H_1: \underline{\Lambda}^T \underline{\beta} \neq \underline{m}$$

Full model: $\underline{y} \sim N(\underline{X}\underline{\beta}, \sigma^2 \mathbf{I})$

with the parameter space $\Omega = \{(\underline{\beta}, \sigma^2) : \beta_j \in \mathbb{R}, \sigma^2 > 0\}$

the reduced model has the parameter space

$$\Omega_0 = \{(\underline{\beta}, \sigma^2) : \beta_j \in \mathbb{R}, \sigma^2 > 0, \underline{\Lambda}^T \underline{\beta} = \underline{m}\}$$

Likelihood: $L(\underline{\beta}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{\|\underline{y} - \underline{X}\underline{\beta}\|^2}{2\sigma^2}\right\}$

$$\phi(\underline{y}) = \frac{\max_{\Omega_0} L(\underline{\beta}, \sigma^2)}{\max_{\Omega} L(\underline{\beta}, \sigma^2)}$$

we reject if $\phi(\underline{y}) < c$ for some constant c depending on the choice of the level of this test.

$\max_{\Omega} L(\underline{\beta}, \sigma^2)$ can be done in two stages

when σ^2 is kept fixed, maximizing $L(\underline{\beta}, \sigma^2)$

over $\underline{\beta}$ is \Leftrightarrow minimizing $\|\underline{y} - \underline{X}\underline{\beta}\|^2$

\Rightarrow the minimizer will be one of the least square solutions $\hat{\underline{\beta}}$ and $\|\underline{y} - \underline{X}\hat{\underline{\beta}}\|^2 = \underline{y}^T (\mathbf{I} - \mathbf{P}) \underline{y}$

and you can check $\hat{\sigma}^2 = \frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{n}$

$$\begin{aligned} \max_{\underline{\sigma}^2} L(\underline{\beta}, \underline{\sigma}^2) &= \left(\frac{1}{\sqrt{2\pi \hat{\sigma}^2}} \right)^n \exp \left\{ - \frac{\| \underline{y} - \underline{X} \hat{\underline{\beta}} \|^2}{2 \hat{\sigma}^2} \right\} \\ &= \left(\frac{1}{\sqrt{2\pi \hat{\sigma}^2}} \right)^n \exp \left\{ - \frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{2 \frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{n}} \right\} \\ &= \left(\frac{1}{\sqrt{2\pi \hat{\sigma}^2}} \right)^n \exp \left\{ - \frac{n}{2} \right\} \end{aligned}$$

~~max~~ ~~max~~ Maximize ~~the~~ the likelihood w.r.t. $\underline{\sigma}^2$, we obtain (using similar calculations)

$$\max_{\underline{\sigma}^2} L(\underline{\beta}, \underline{\sigma}^2) = \left(\frac{1}{\sqrt{2\pi \hat{\sigma}^2}} \right)^n \exp \left\{ - \frac{n}{2} \right\}$$

where $\hat{\sigma}^2 = \frac{\underline{y}^T (\underline{I} - \underline{P}_0) \underline{y}}{n}$

Thus the likelihood ratio test statistic

$$LRT = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{n/2}$$

thus it is a function of $\frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{\underline{y}^T (\underline{I} - \underline{P}_0) \underline{y}}$

Note that, $\frac{\underline{y}^T (\underline{I} - \underline{P}_0) \underline{y}}{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}} - 1 = \frac{\underline{y}^T (\underline{P} - \underline{P}_0) \underline{y}}{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}$

thus the LRT test statistic is closely related to the already known ~~F~~-test statistic for testing the reduced models.

confidence interval:

Consider $\underline{\lambda}^T \underline{\beta}$ is estimable ~~is then~~.

Goal: Want $100(1-\alpha)\%$ confidence interval for $\underline{\lambda}^T \underline{\beta}$.

$$t = \frac{\underline{\lambda}^T \hat{\underline{\beta}} - \underline{\lambda}^T \underline{\beta}}{\sqrt{\frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{(n - \text{rank}(\underline{X}))} \underline{\lambda}^T (\underline{X}' \underline{X})^{-1} \underline{\lambda}}} \sim t(n - \text{rank}(\underline{X}))$$

$$P\left(|t| \leq t_{\alpha/2}(n - \text{rank}(\underline{X}))\right) = 1 - \alpha$$

$$\Rightarrow P\left(-\hat{\underline{\lambda}}^T \hat{\underline{\beta}} - \frac{t_{\alpha/2}(n - \text{rank}(\underline{X}))}{\sqrt{\frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{(n - \text{rank}(\underline{X}))} \underline{\lambda}^T (\underline{X}' \underline{X})^{-1} \underline{\lambda}}} \leq \underline{\lambda}^T \underline{\beta} \leq \hat{\underline{\lambda}}^T \hat{\underline{\beta}} + \frac{t_{\alpha/2}(n - \text{rank}(\underline{X}))}{\sqrt{\frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{(n - \text{rank}(\underline{X}))} \underline{\lambda}^T (\underline{X}' \underline{X})^{-1} \underline{\lambda}}}\right) = 1 - \alpha$$

$$\Rightarrow \hat{\underline{\lambda}}^T \hat{\underline{\beta}} \pm \frac{t_{\alpha/2}(n - \text{rank}(\underline{X}))}{\sqrt{\frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{(n - \text{rank}(\underline{X}))} \underline{\lambda}^T (\underline{X}' \underline{X})^{-1} \underline{\lambda}}}$$

is the $100(1-\alpha)\%$ CI for $\underline{\lambda}^T \underline{\beta}$.

Find $100(1-\alpha)\%$ confidence set for $\underline{\Lambda}^T \underline{\beta}$

We know,

$$\underline{\Lambda}^T \hat{\underline{\beta}} \sim N(\underline{\Lambda}^T \underline{\beta}, \underbrace{\sigma^2 \underline{\Lambda}^T (\underline{X}'\underline{X})^{-1} \underline{\Lambda}}_{\underline{H}}) = N(\underline{\Lambda}^T \underline{\beta}, \sigma^2 \underline{H})$$

~~100(1-\alpha)~~ $(\underline{\Lambda}^T \hat{\underline{\beta}} - \underline{\Lambda}^T \underline{\beta})$

$$\frac{(\underline{\Lambda}^T \hat{\underline{\beta}} - \underline{\Lambda}^T \underline{\beta})' \underline{H}^{-1} (\underline{\Lambda}^T \hat{\underline{\beta}} - \underline{\Lambda}^T \underline{\beta})}{\sigma^2 \text{rank}(\underline{\Lambda})}$$

$$\sim F(\text{rank}(\underline{\Lambda}), n - \text{rank}(\underline{X}))$$

$$\frac{\underline{y}'(\underline{I} - \underline{P})\underline{y}}{\sigma^2 (n - \text{rank}(\underline{X}))}$$

$100(1-\alpha)\%$ confidence set for $\underline{\Lambda}^T \underline{\beta}$ is

$$\left\{ \underline{d} : \frac{(\underline{\Lambda}^T \hat{\underline{\beta}} - \underline{d})' \underline{H}^{-1} (\underline{\Lambda}^T \hat{\underline{\beta}} - \underline{d})}{\text{rank}(\underline{\Lambda}) \frac{\underline{y}'(\underline{I} - \underline{P})\underline{y}}{n - \text{rank}(\underline{X})}} \leq F_{1-\alpha}(\text{rank}(\underline{\Lambda}), n - \text{rank}(\underline{X})) \right.$$

this is inside of an s -dimensional ellipse where $\underline{\Lambda}$ is of dimension $p \times s$.

Now if ~~you~~ the goal is to find $100(1-\alpha)\%$ CI for each $\lambda_j^T \underline{\beta}$, $j=1, \dots, s$ from here,

it can also be obtained ~~from~~ in the following way

$$\lambda_j^T \underline{\beta} \in \left(\lambda_j^T \hat{\underline{\beta}} \pm t_{\alpha/2} (n - \text{rank}(X)) \sqrt{\hat{\sigma}^2} \sqrt{H_{jj}} \right)$$

where H_{jj} is the ~~(i,j)~~ (j,j)th diagonal entry of \underline{H} .

This can be found by looking at different cross-sections.

$$\hat{\sigma}^2 = \frac{\underline{y}^T (\underline{I} - \underline{P}) \underline{y}}{n - \text{rank}(X)}$$