

MAT-266: Test de hipótesis y regiones de confianza

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Problema:

Desarrollar el test de razón de verosimilitudes para probar hipótesis lineales de la forma:

$$H_0 : \mathbf{G}\boldsymbol{\beta} = \mathbf{g}, \quad \text{versus} \quad H_1 : \mathbf{G}\boldsymbol{\beta} \neq \mathbf{g} \quad (1)$$

donde \mathbf{G} es una matriz de contrastes de orden $q \times p$ con $\text{rg}(\mathbf{G}) = q$ y $\mathbf{g} \in \mathbb{R}^q$.

Observación:

H_0 es expresada como un sistema de ecuaciones mientras que H_1 indica que al menos una ecuación no se satisface.



Ejemplo (Función de producción Cobb-Douglas):

Considere la función de producción:

$$Y = AK^{\alpha}L^{\beta},$$

donde Y es el nivel de producción, K es el stock de capital (maquinaria, equipamiento, construcciones), L es el trabajo (total de horas-hombre trabajadas en un año). Note que,

$$\log Y = \log A + \alpha \log K + \beta \log L,$$

donde A representa un factor de productividad (eficiencia), mientras que α y β son las elasticidades del capital y trabajo, respectivamente.

Por ejemplo, podemos estar interesados en,

$$H_0 : \alpha + \beta = 1, \quad \text{versus} \quad H_1 : \alpha + \beta \neq 1.$$



Hipótesis lineales

Para abordar hipótesis lineales, usaremos el principio de verosimilitud. Es decir, consideraremos el estadístico

$$\Lambda = \frac{\max_{G\beta=g} L(\beta, \sigma^2)}{\max_{\Theta} L(\beta, \sigma^2)} = \frac{L(\tilde{\beta}, \tilde{\sigma}^2)}{L(\hat{\beta}, \hat{\sigma}^2)}.$$

Asumiendo que $\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I})$ tenemos la función de verosimilitud

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) \right\}.$$

Sabemos que

$$\begin{aligned} \max_{\Theta} L(\beta, \sigma^2) &= L(\hat{\beta}, \hat{\sigma}^2) \\ &= (2\pi\hat{\sigma}^2)^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2 \right\} \\ &= \{2\pi Q(\hat{\beta})/n\}^{-n/2} \exp \left\{ -\frac{n}{2Q(\hat{\beta})} \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2 \right\} \\ &= \{2\pi Q(\hat{\beta})/n\}^{-n/2} \exp(-n/2). \end{aligned}$$



Mientras que bajo $H_0 : \mathbf{G}\boldsymbol{\beta} = \mathbf{g}$, tenemos¹

$$\begin{aligned}\max_{\mathbf{G}\boldsymbol{\beta}=\mathbf{g}} L(\boldsymbol{\beta}, \sigma^2) &= L(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2) = (2\pi\tilde{\sigma}^2)^{-n/2} \exp\left\{-\frac{1}{2\tilde{\sigma}^2} \|\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}\|^2\right\} \\ &= \{2\pi Q(\tilde{\boldsymbol{\beta}})/n\}^{-n/2} \exp(-n/2).\end{aligned}$$

De este modo, el estadístico de razón de verosimilitudes

$$\begin{aligned}\Lambda &= \frac{L(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)}{L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)} = \frac{\{2\pi Q(\tilde{\boldsymbol{\beta}})/n\}^{-n/2} \exp(-n/2)}{\{2\pi Q(\hat{\boldsymbol{\beta}})/n\}^{-n/2} \exp(-n/2)} \\ &= \left\{ \frac{Q(\tilde{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})} \right\}^{n/2},\end{aligned}$$

y de acuerdo con el principio de verosimilitud rechazamos $H_0 : \mathbf{G}\boldsymbol{\beta} = \mathbf{g}$ si Λ es pequeño.

¹ $\Theta_0 = \{\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top : \mathbf{G}\boldsymbol{\beta} = \mathbf{g}\}.$

Alternativamente, podemos considerar

$$\Lambda^{2/n} = \frac{Q(\hat{\beta})}{Q(\tilde{\beta})}.$$

Recuerde que $\tilde{\beta} = \hat{\beta} - B(G\hat{\beta} - g)$, así

$$Y - X\tilde{\beta} = Y - X(\hat{\beta} - B(G\hat{\beta} - g)) = Y - X\hat{\beta} + XB(G\hat{\beta} - g).$$

Además,

$$\begin{aligned} Q(\tilde{\beta}) &= \|Y - X\tilde{\beta}\|^2 = \|Y - \hat{\beta} - B(G\hat{\beta} - g)\|^2 \\ &= \|Y - X\hat{\beta}\|^2 + \|XB(G\hat{\beta} - g)\|^2 \\ &\quad + (Y - X\hat{\beta})^\top XB(G\hat{\beta} - g) + (G\hat{\beta} - g)^\top B^\top X^\top (Y - X\hat{\beta}). \end{aligned}$$

Sin embargo,

$$X^\top (Y - X\hat{\beta}) = X^\top (I - H)Y = 0$$



Luego,

$$\begin{aligned} Q(\tilde{\beta}) &= Q(\hat{\beta}) + (G\hat{\beta} - g)^\top B^\top X^\top X B (G\hat{\beta} - g) \\ &\geq Q(\hat{\beta}) \end{aligned}$$

Observación:

Es decir,

$$0 \leq \frac{Q(\hat{\beta})}{Q(\tilde{\beta})} \leq 1,$$

por tanto, $\Lambda^{2/n} \in [0, 1]$.



Además, recordando que

$$B = (X^T X)^{-1} G^T (G(X^T X)^{-1} G^T)^{-1},$$

obtenemos

$$B^T X^T X B = (G(X^T X)^{-1} G^T)^{-1}.$$

De este modo,

$$Q(\tilde{\beta}) - Q(\hat{\beta}) = (G\hat{\beta} - g)^T (G(X^T X)^{-1} G^T)^{-1} (G\hat{\beta} - g).$$

Así

$$\frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\hat{\beta})} = \Lambda^{-2/n} - 1,$$

es decir, valores pequeños de Λ (en cuyo caso rechazamos H_0) implican valores grandes de la razón anterior.



Ahora considere

$$\begin{aligned}E(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g}) &= \mathbf{G}E(\hat{\boldsymbol{\beta}}) - \mathbf{g} = \mathbf{G}\boldsymbol{\beta} - \mathbf{g} \\ \text{Cov}(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g}) &= \mathbf{G} \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{G}^\top = \sigma^2 \mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top.\end{aligned}$$

Note que

$$(\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} = \left\{ \frac{1}{\sigma^2} \text{Cov}(\mathbf{G}\hat{\boldsymbol{\beta}}) \right\}^{-1}.$$

De esta forma

$$\frac{(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})}{\sigma^2} \sim \chi^2(q; \delta),$$

pues $\sigma^{-2}(\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} \text{Cov}(\mathbf{G}\hat{\boldsymbol{\beta}}) = \mathbf{I}$, es matriz idempotente, y

$$\delta = \frac{1}{2\sigma^2} (\mathbf{G}\boldsymbol{\beta} - \mathbf{g})^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\mathbf{G}\boldsymbol{\beta} - \mathbf{g})$$

Por otro lado, sabemos que

$$\frac{Q(\hat{\boldsymbol{\beta}})}{\sigma^2} \sim \chi^2(n - p; 0)$$



Para notar la independencia, considere β^* cualquier vector que satisfice la condición $G\beta^* = g$. Entonces,

$$Y - X\hat{\beta} = (I - H)Y = (I - H)(Y - X\beta^*),$$

pues $(I - H)X = 0$ y

$$\begin{aligned} G\hat{\beta} - g &= G(X^\top X)^{-1}X^\top Y - G\beta^* = G\{(X^\top X)^{-1}X^\top Y - \beta^*\} \\ &= G\{(X^\top X)^{-1}X^\top Y - (X^\top X)^{-1}X^\top X\beta^*\} \\ &= G(X^\top X)^{-1}X^\top(Y - X\beta^*). \end{aligned}$$

En nuestro caso,

$$Y - X\beta^* \sim N_n(X\beta - X\beta^*, \sigma^2 I) \stackrel{d}{=} N_n(X(\beta - \beta^*), \sigma^2 I).$$



De este modo,

$$Q(\hat{\beta}) = \|Y - X\hat{\beta}\|^2 = (Y - X\beta^*)^\top (I - H)(Y - X\beta^*),$$

mientras que

$$\begin{aligned} Q(\tilde{\beta}) - Q(\hat{\beta}) &= (Y - X\beta^*)^\top X(X^\top X)^{-1}G^\top (G(X^\top X)^{-1}G^\top)^{-1} \\ &\quad \times G(X^\top X)^{-1}X^\top (Y - X\beta^*). \end{aligned}$$

Como

$$(I - H)X(X^\top X)^{-1}G^\top (G(X^\top X)^{-1}G^\top)^{-1}G(X^\top X)^{-1}X^\top = 0,$$

sigue la independencia y permite construir la estadística

$$F = \frac{\{Q(\tilde{\beta}) - Q(\hat{\beta})\}/q}{Q(\hat{\beta})/(n - p)} \sim F(q, n - p; \delta).$$

Esto lleva al siguiente resultado.



Resultado 1:

Para el modelo lineal $Y = X\beta + \epsilon$ con los supuestos A1 a A4*. Un test de tamaño α para probar

$$H_0 : G\beta = g, \quad \text{versus} \quad H_1 : G\beta \neq g,$$

es dado por, rechazar H_0 cuando

$$F = \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{qs^2} \geq F_{1-\alpha}(q, n - p; 0).$$

Demostración:

Bajo $H_0 : G\beta = g$, tenemos $\delta = 0$, de ahí que $F \sim F(q, n - p; 0)$, lo que lleva al resultado deseado.



Observación:

Note que podemos escribir el estadístico F de varias formas equivalentes, a saber:

$$\begin{aligned} F &= \left(\frac{n-p}{q} \right) \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\hat{\beta})} \\ &= \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{qs^2} \\ &= \frac{(G\hat{\beta} - g)^\top (G(X^\top X)^{-1}G^\top)^{-1} (G\hat{\beta} - g)}{qs^2} \\ &\sim F(q, n-p, \delta). \end{aligned}$$



Hemos notado la relación que existe entre el test de razón de verosimilitudes con el estadístico F para probar hipótesis lineales en el modelo de regresión lineal

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}).$$

A continuación exploramos la relación entre el estadístico F con los test de Wald, score y gradiente para hipótesis lineales, del tipo

$$H_0 : \mathbf{G}\boldsymbol{\beta} = \mathbf{g}, \quad \text{versus} \quad H_1 : \mathbf{G}\boldsymbol{\beta} \neq \mathbf{g}.$$

Primeramente, note que la matriz de información de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$, adopta la forma

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{pmatrix} \mathbf{X}^\top \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2\sigma^2} \end{pmatrix}.$$



El estadístico de Wald para hipótesis lineales de la forma $H_0 : G\beta = g$, es dado por

$$\begin{aligned} W &= n(\hat{G}\hat{\beta} - g)^\top (G\{\mathcal{F}(\hat{\beta})\}^{-1}G^\top)^{-1}(\hat{G}\hat{\beta} - g) \\ &= \frac{n(\hat{G}\hat{\beta} - g)^\top (G(X^\top X)^{-1}G^\top)^{-1}(\hat{G}\hat{\beta} - g)}{\hat{\sigma}^2}. \end{aligned}$$

Mientras que el test score es dado por

$$\begin{aligned} R &= \frac{1}{n}U^\top(\tilde{\beta})\{\mathcal{F}(\tilde{\beta})\}^{-1}U(\tilde{\beta}) \\ &= \frac{1}{n\tilde{\sigma}^2}(Y - X\tilde{\beta})^\top X(X^\top X)^{-1}X^\top(Y - X\tilde{\beta}). \end{aligned}$$



Como,

$$\begin{aligned} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\tilde{\beta}) &= \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}\mathbf{B}(\mathbf{G}\hat{\beta} - \mathbf{g})) \\ &= \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\hat{\beta}) + \mathbf{X}^\top \mathbf{X}\mathbf{B}(\mathbf{G}\hat{\beta} - \mathbf{g}) \\ &= \mathbf{X}^\top \mathbf{X}\mathbf{B}(\mathbf{G}\hat{\beta} - \mathbf{g}). \end{aligned}$$

Así

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\tilde{\beta})^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\tilde{\beta}) &= (\mathbf{G}\hat{\beta} - \mathbf{g})^\top \mathbf{B}^\top \mathbf{X}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\mathbf{B}(\mathbf{G}\hat{\beta} - \mathbf{g}) \\ &= (\mathbf{G}\hat{\beta} - \mathbf{g})^\top \mathbf{B}^\top \mathbf{X}^\top \mathbf{X}\mathbf{B}(\mathbf{G}\hat{\beta} - \mathbf{g}) \\ &= (\mathbf{G}\hat{\beta} - \mathbf{g})^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\mathbf{G}\hat{\beta} - \mathbf{g}). \end{aligned}$$

Finalmente,

$$R = \frac{(\mathbf{G}\hat{\beta} - \mathbf{g})^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\mathbf{G}\hat{\beta} - \mathbf{g})}{n\tilde{\sigma}^2}$$



Por otro lado, el estadístico gradiente es dado por

$$T = \mathbf{U}^\top (\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})^\top \mathbf{B}^\top \mathbf{X}^\top \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$$

Sabemos que $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathbf{B}(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})$, esto lleva a

$$\begin{aligned} T &= (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})^\top \mathbf{B}^\top \mathbf{X}^\top \mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} + \mathbf{B}(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})) \\ &= (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})^\top \mathbf{B}^\top \mathbf{X}^\top \mathbf{X} \mathbf{B}(\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g}) \\ &= (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g})^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\mathbf{G}\hat{\boldsymbol{\beta}} - \mathbf{g}) \end{aligned}$$



Observación:

Es decir, podemos escribir:

$$W = \frac{qn}{n-p} F$$

$$R = \left\{ 1 + \left(\frac{n-p}{q} \right) F^{-1} \right\}^{-1}$$

$$T = \frac{qs^2}{n-p} F$$

Lo que permite notar que basta usar el estadístico F para probar hipótesis lineales en el modelo de regresión lineal.



Regiones de confianza

Sea $\gamma = \mathbf{G}\beta$ un vector q -dimensional. Sabemos que

$$(\gamma - \hat{\gamma})^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\gamma - \hat{\gamma}) \leq qs^2 F_{1-\alpha}(q, n-p),$$

luego, puede ser usada como una región de confianza.

En particular, para $\mathbf{G} = \mathbf{I}$ tenemos

$$R(\beta) = \{\beta : (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta}) \leq ps^2 F_{1-\alpha}(p, n-p)\}.$$

Además, para el caso en que $\text{rg}(\mathbf{G}) = 1$, tenemos un intervalo de confianza para $\gamma = \mathbf{a}^\top \beta$, digamos

$$CI(\gamma) = [\mathbf{a}^\top \hat{\beta} \mp t_{1-\alpha/2}(n-p) s \sqrt{\mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{a}}],$$

o bien,

$$R(\mathbf{a}^\top \beta) = \{\gamma : (\gamma - \mathbf{a}^\top \hat{\beta})^2 \leq s^2 \mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{a} F_{1-\alpha}(1, n-p)\}.$$

