- Aguilar Hernández Erick Eduardo
- mat.ErickAguilar@ciencias.unam.mx

Modelo de Regresión Multiple

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$$

Y : Variable dependiente (Regresando)
X_i : Variables independientes (Predictora)
β_i : Regresores
ε : Ruido v.a. $\varepsilon \sim N(0, \sigma^2)$

Una muestra aleatoria del modelo se suele representar como una tabla

Y	X_1	$\cdots$	X_p
y_1	x_{11}	$\cdots$	x_{1p}
y_2	x_{21}	$\cdots$	x_{2p}
$\vdots$			
y_n	x_{n1}	$\cdots$	x_{np}

donde
$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ \vdots \\ y_p = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} + \varepsilon_n \end{cases}$$

Asumiremos que:
- $E(\varepsilon_i) = 0$
- $V(\varepsilon_i) = \sigma^2$
- $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{np} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{n\times(p+1)} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{p\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}$$

$$X_{n\times(p+1)} \cdot \beta_{(p+1)\times 1} + \varepsilon_{n\times 1} = Y_{n\times 1}$$

Obs.: • $E[\varepsilon_{n\times 1}] = 0_{n\times 1}$

• $Cov[\varepsilon_i, \varepsilon_j] = 0$ si $i \neq j$

$Cov(\varepsilon_i, \varepsilon_j) = Cov(\varepsilon_i, \varepsilon_i) = V[\varepsilon_i] = \sigma^2$

$$Cov(\varepsilon) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I_{n\times n}$$

$\Rightarrow E[Y_n] = X\beta$

$E[y_i] = E[\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i]$

$= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + E[\varepsilon_i]$

$E[y_i] = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$

$$\begin{cases} E[y_1] = \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} \\ E[y_2] = \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} \\ \vdots \\ E[y_n] = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} \end{cases}$$

$$E\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ & \vdots & & \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$E[Y_{n\times 1}] = X_{n\times(p+1)} \, \beta_{(p+1)\times 1}$$

• $Cov(y_i, y_j) = Cov(y_i, \beta_0 + \beta_1 x_{j1} + \cdots + \beta_p x_{jp} + \varepsilon_j)$

$= \cancel{Cov(y_i, \beta_0)} + \beta_1 \cancel{Cov(y_i, x_{j1})} + \cdots + \cancel{\beta_p Cov(y_i, x_{jp})} + Cov(y_i, \varepsilon_j)$

$$= Cov(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \varepsilon_j)$$

$$= Cov(\varepsilon_i, \varepsilon_j) = 0 \quad si\ i \neq j$$

$$= \sigma^2 \quad Si\ \ i = j$$

$$\Rightarrow Cov(Y_{n\times 1}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

Estimación de los parámetros

Mínimos cuadrados:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \quad (\text{Real})$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} \quad (\text{Estimado})$$

$$\mathcal{L}_2(\beta_0, \ldots, \beta_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (\text{Perdida cuadratica})$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_K x_{iK})^2$$

$$\varepsilon' \varepsilon = \sum_{i=1}^{n} (y_i - x_i' \hat{\beta})^2$$

$$\varepsilon' \varepsilon = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= (Y' - (X\hat{\beta})')(Y - X\hat{\beta})$$

$$= (Y' - \hat{\beta}' X')(Y - X\hat{\beta})$$

$$= Y'Y - Y'X\hat{\beta} - \hat{\beta}' X' Y + \hat{\beta}' X' X \hat{\beta}$$

$$\varepsilon' \varepsilon = Y'Y - 2X'Y\hat{\beta} + \hat{\beta}' X' X \hat{\beta}$$

$$\frac{\partial}{\partial \hat{\beta}} \varepsilon'\varepsilon = 0 - 2X'y + 2X'X\hat{\beta} = 0$$

$$-2X'y + 2X'X\hat{\beta} = 0$$

$$\cancel{2}X'X\hat{\beta} = \cancel{2}X'y$$

$$X'X\hat{\beta} = X'y$$

$$\cancel{(X'X)^{-1}(X'X)}\hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

Obs. $(X'X)^{-1}$ existe si y solo si X es rango $p+1 < n$. (Rango completo)

Mostrar que $\hat{\beta} = (X'X)^{-1}X'y$ mimiza $\varepsilon'\varepsilon$. Suponer que β^* es otro estimador lineal de β

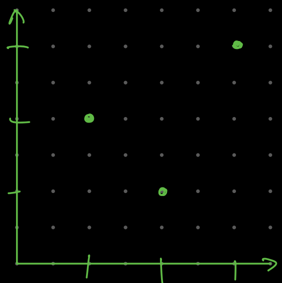
$$\begin{aligned}
\varepsilon'\varepsilon(\beta^*) &= (y - X\beta^*)'(y - X\beta^*) \\
&= (y - X\beta^* + X\hat{\beta} - X\hat{\beta})'(\underline{y - X\beta^*} + \underline{X\hat{\beta} - X\hat{\beta}}) \\
&= ((y - X\hat{\beta})' + (X\hat{\beta} - X\beta^*)') \\
&\quad ((y - X\beta) + (X\hat{\beta} - X\beta^*)) \\
&= (y - X\hat{\beta})'(y - X\hat{\beta}) + (y - X\hat{\beta})'(X\hat{\beta} - X\beta^*) + \\
&\quad (X\hat{\beta} - X\beta^*)'(y - X\hat{\beta}) + (y - X\hat{\beta})'(X\hat{\beta} - X\beta^*) \\
&= (y - X\hat{\beta})'(y - X\hat{\beta}) + \cancel{2(\hat{\beta} - \beta^*)(X'y - X'X\hat{\beta})}^{0} + \\
&\quad (\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*)
\end{aligned}$$

$$\varepsilon'\varepsilon(\beta^*, \hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \beta^*)'(X'X)(\hat{\beta} - \beta^*) \geq 0$$

Obs $\varepsilon'\varepsilon(\beta^*, \hat{\beta})$ es minima cuando $\hat{\beta} = \beta^*$ $\therefore$ $\hat{\beta}$ es el minimo

Ejercicio

X	Y
2	1
1	2
3	3



$$X = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

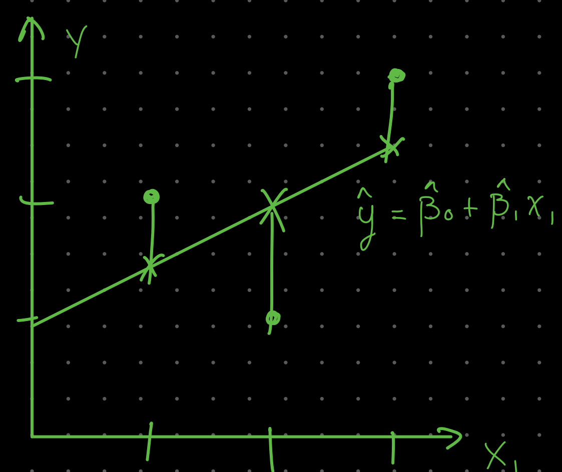
$$X'X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \quad X'y = \begin{bmatrix} 6 \\ 13 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 13 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} \begin{matrix} \leftarrow \hat{\beta}_0 \\ \leftarrow \hat{\beta}_1 \end{matrix}$$

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3/2 \\ 5/2 \end{bmatrix}$$

$$\hat{\varepsilon} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 3/2 \\ 5/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1/2 \\ 1/2 \end{bmatrix}$$



Propiedades de los estimadores $\hat{\beta}$

- $E[\hat{\beta}] = E[(X'X)^{-1}X'y] = (X'X)^{-1}X' E[y]$

$$= \cancel{(X'X)^{-1}X'X}\beta = I\beta = \beta$$

- $Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ $\qquad Cov(Ax) = A\, Cov(x)\, A'$

$$Cov(\hat{\beta}) = Cov((X'X)^{-1}X'y)$$

$$= ((X'X)^{-1}X')\, Cov(y)\, (((X'X)^{-1}X'))'$$

$$= (X'X)^{-1}X'\, Cov(y)\, X'' (X'X)^{-1'}$$

$$= (X'X)^{-1} X'(\sigma^2 I) X (X'X)^{-1}$$

$$= \sigma^2 \cancel{(X'X)^{-1}(X'X)} (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

$$\sigma^2 (X'X)^{-1} = \begin{bmatrix} V(\hat{\beta}_0) & \cdots & Cov(\hat{\beta}_0, \hat{\beta}_p) \\ \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_p, \hat{\beta}_0) & \cdots & V(\hat{\beta}_p) \end{bmatrix} \sigma^2$$

- Teorema de Gauss - Markov

Si $E[y] = X\beta$ y $Cov(y) = \sigma^2 I_n$ el estimador $\hat{\beta}$ por m.c. tiene la menor varianza de todos los estimadores insesgados lineales.

Sea $\tilde{\beta} = Ay$ otro estimador de β insesgado: $E[\tilde{\beta}] = \beta$

$E[Ay] = AE[y] = \underline{AX}\beta = \beta \Rightarrow \underline{AX = I}$ •

$$Cov(Ay) = A\Sigma_y A'$$

$Cov(Ay) = ACov(y)A' = A\sigma^2 I_n A'$

$Cov(\tilde{\beta}) = \sigma^2 AA' \Rightarrow V[\tilde{\beta}_i] = [\sigma^2 AA']_{ii} \leftarrow$

$$AA' = [\underline{A-(X'X)^{-1}X'} + (X'X)^{-1}X'][\underline{A-(X'X)^{-1}X'} + \underbrace{(X'X)^{-1}X'}]'$$

$$= [A-(X'X)^{-1}X'][A-(X'X)^{-1}X']' + [A-(X'X)^{-1}X'][(X'X)^{-1}X']' + [(X'X)^{-1}X'][A-(X'X)^{-1}X']'$$

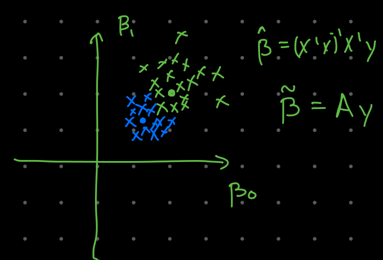
$$+ [(X'X)^{-1}X'][(X'X)^{-1}X']'$$

$$= [A-(X'X)^{-1}X'][A-(X'X)^{-1}X']' + AX(X'X)^{-1} - \cancel{(X'X)^{-1}X'X}^{I}(X'X)^{-1} + (X'X)^{-1}\cancel{X'A'}^{I}$$

$$- (X'X)^{-1}\cancel{X'X(X'X)^{-1}}^{I} + \cancel{(X'X)^{-1}X'X}^{I}(X'X)^{-1}$$

$$= [A-(X'X)^{-1}X'][A-(X'X)^{-1}X']' + I(X'X)^{-1} - I(X'X)^{-1} + (X'X)^{-1}I_n' - I_n(X'X)^{-1}$$

$$+ (X'X)^{-1}$$



$$AA' = [A-(X'X)^{-1}X'][A-(X'X)^{-1}X']' + (X'X)^{-1} \leftarrow$$

$$Cov(\tilde{\beta}) = \sigma^2[[A-(X'X)^{-1}X'][A-(X'X)^{-1}X']' + \underline{(X'X)^{-1}}]$$

$Cov(\tilde{\beta})$ alcanza su valor mínimo si $A = (X'X)^{-1}X'$

$$\therefore \tilde{\beta} = Ay = (X'X)^{-1}X'y = \hat{\beta}$$

$\therefore \hat{\beta}$ es el estimador con varianzas mínimas e insesgado

Estimar σ^2

Si x es un vector aleatorio y A una matriz cuadrada

i) $Cov(x) = E[xx'] - E[x']E[x] \leftarrow$

ii) $E[x'Ax] = tr[A\,Cov(x)] + tr[E[x']AE[x]]$

$$E[x'Ax] = E[tr(x'Ax)] = E[tr(Axx')]$$
$$= tr[E[Axx']] = tr[A\,E[xx']]$$
$$= tr[A[Cov(x) + E[x']E[x]]]$$
$$= tr[A\,Cov(x)] + tr(A\,E[x']E[x])$$
$$E[x'Ax] = tr[A\,Cov(x)] + tr(E[x']A\,E[x])$$

$E[y_i] = x_i'\beta$ para i-esimo renglón

$$V[y_i] = E[y_i - E[y_i]]^2 = E[y_i - x_i'\beta]^2$$
$$= E[\cancel{x_i'\beta} - \cancel{x_i'\beta} + \varepsilon]^2 = E[\varepsilon]^2 = \sigma^2$$

Consideremos:

Si $e = Y - X\hat{\beta}$ definimos al estadístico: la suma de cuadrados de los errores como $SSE = e'e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2$$
$$= (y - x\hat{\beta})'(y - x\hat{\beta})$$
$$= (y' - \hat{\beta}'x')(y - x\hat{\beta})$$
$$= y'y - y'x\hat{\beta} - \cancel{(x\hat{\beta})'y} + \cancel{\hat{\beta}'(x'x)\hat{\beta}}$$
$$= y'y - y'x\hat{\beta}$$
$$= y'y - y'x(x'x)^{-1}x'y$$
$$= y'(I - x(x'x)^{-1}x')y$$
$$= y'Ay$$
$$E[SSE] = tr(A\Sigma_y) + E[y]'A\,E[y]$$

$Cov(y) = \sigma^2 I \qquad E[y] = x\beta \qquad A = I - x(x'x)^{-1}x'$

$$E[SSE] = tr\left[(I - X(X'X)^{-1}X')\sigma^2\right] +$$
$$tr\left[(X\beta)'(I - X(X'X)^{-1}X')(X\beta)\right]$$
$$= \sigma^2 tr(I - X(X'X)^{-1}X') +$$
$$tr(\beta'X'X\beta - \beta'\cancel{X'X(X'X)^{-1}}X'X\beta)$$
$$= \sigma^2(tr(I_n) - tr(X(X'X)^{-1}X'))$$
$$= \sigma^2(n - tr(\cancel{(X'X)^{-1}X'X})I_{k+1})$$
$$E[SSE] = \sigma^2(n-(k+1))$$
$$E\left[SSE/n-k-1\right] = \sigma^2 \quad \therefore \frac{SSE}{n-k-1} \text{ estima a } \sigma^2$$

$$\hat{Cov}(\hat\beta) = \hat\sigma^2(X'X)^{-1} \text{ donde } \hat\sigma^2 = SSE/n-k-1$$

Modelo Normal: Se asume que $\varepsilon_i \sim N(0, \sigma^2)$

$$\varepsilon \sim N_n(0, \sigma^2 I_n), \quad y \sim N_n(X\beta, \sigma^2 I_n)$$

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

$$\ln\mathcal{L}(\beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

$$\frac{\partial}{\partial\beta}\ln\mathcal{L} = -\frac{1}{2\sigma^2}\frac{\partial}{\partial\beta}(y - X\beta)'(y - X\beta) = -\frac{1}{\cancel{2}\sigma^2}\cancel{2}X'(y - X\beta)$$

$$-\frac{1}{\sigma^2}X'(y - X\beta) = 0$$
$$X'(y - X\beta) = 0$$
$$X'y - X'X\beta = 0$$

$$X'Y = X'X\beta$$
$$(X'X)^{-1}X'Y = \hat{\beta}_{MLE} = \hat{\beta}_{MC}$$

$$\frac{\partial}{\partial\sigma^2}\ln\mathcal{L} = \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}(Y-X\beta)'(Y-X\beta)-n\right] = 0$$
$$0 = \frac{\cancel{1}}{\cancel{2}\sigma^2}\left[\frac{1}{\sigma^2}(Y-X\beta)'(Y-X\beta)-n\right]$$
$$0 = \frac{1}{\sigma^2}(Y-X\hat{\beta})'(Y-X\hat{\beta})-n$$
$$0 = \frac{1}{\sigma^2}SSE-n$$
$$n = \frac{1}{\sigma^2}SSE$$
$$\hat{\sigma}^2_{MLE} = \frac{1}{n}SSE \neq \hat{\sigma}^2$$

Obs $\hat{\beta} \sim N_{k+1}(\beta, \sigma^2(X'X)^{-1})$

Distribuciones No Centrales

Chi cuadrada no central:

Si $Z_1,\dots,Z_n$, $Z_i \sim N(0,1)$ entonces $\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$

Si $Y_1,\dots,Y_n$ $Y_i \sim N(\mu,1)$ entonces $\sum_{i=1}^{n} Y_i^2 \not\sim \chi^2(n)$

Si $Y = [Y_1,\dots,Y_n]'$ entonces $Y \sim N(\mu = [\mu_1,\dots,\mu_n], I_n)$

pero si $\sum_{i=1}^{n}(Y_i-\mu_i)^2 = (Y-\mu)'(Y-\mu) \sim \chi^2(n)$

pues $Y_i - \mu_i \sim N(0,1)$

A $y'y$ se le asocia una distribución chi cuadrada no central

$$y'y \sim \chi^2(n, \lambda)$$

λ: parametro de no centralidad:

$$\lambda = \frac{1}{2} \sum_{i=1}^{n} \mu_i = \frac{1}{2} \mu' \mu$$

Si $X \sim \chi^2(n, \lambda)$ su fdp:

$$f_X(x; n, \lambda) = \frac{1}{n} \exp\left(-\frac{1}{2}(x+\lambda)\right)\left(\frac{x}{\lambda}\right)^{\frac{n}{4}-\frac{1}{2}} I_{\frac{n}{2}-1}(\sqrt{\lambda x})$$

$I_v(x)$ es la función de Bessel de primer orden

Propiedades:

i) Si $\lambda = 0$ $\chi^2(n, \lambda) = \chi^2(n)$

ii) $E[x] = n + 2\lambda$

iii) $V[x] = 2n + 8\lambda$

iv) $M_x(t) = (1-2t)^{-n/2} \exp(-\lambda(1-(1-2t)^{-1}))$

vi) Si $X_1, \dots, X_k$ donde $X_i \sim \chi^2(n_i, \lambda_i)$

$$\sum_{i=1}^{k} X_i \sim \chi^2\left(\sum_{i=1}^{n} n_i, \sum_{i=1}^{n} \lambda_i\right)$$

Distribución F - No central

Si $\mathcal{U} \sim \chi^2(p, \lambda)$ y $\mathcal{V} \sim \chi^2(q, 0)$

$$F = \frac{U/p}{V/q} \sim F(p, q, \lambda)$$

Distribución T - no central

Si $X \sim N(\mu, 1)$ y $\mathcal{U} \sim \chi^2(p)$ $X \perp \mathcal{U}$

$$T = \frac{X}{\sqrt{U/p}} \sim t(p, \mu)$$

<u>Teorema 1</u>: Sea $y \sim N_p(\mu, \Sigma)$ y A una matriz simétrica de rango r, y $\lambda = \frac{1}{2}\mu'\mu$ se tiene que:

$$y'Ay \sim \chi^2(r, \lambda) \iff A\Sigma \text{ es indempotente}$$

<u>Corolario</u>: Si $y \sim N_p(\mu, \sigma^2 I)$ entonces

$$\frac{y'Ay}{\sigma^2} \sim \chi^2\left(\text{rango}(A), \lambda = \frac{1}{2\sigma^2}\mu'A\mu\right)$$

Aplicación al modelo normal

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = e'e = (y - \hat{y})'(y - \hat{y})$$

$$= (y - X\hat{\beta})'(y - X\hat{\beta})$$

$$= (y - X(X'X)^{-1}X'y)(y - X(X'X)^{-1}Xy)$$

$$= (y - Hat\, y)'(y - Hat\, y) \quad \text{Obs}$$

$$= (y' - y' Hat')(y - Hat\, y) \quad Hat\, y = \hat{y}$$

$$= y'y - y' Hat\, y - \cancel{y' Hat\, y + y' Hat^2 y}$$

$$SSE = y'y - y' Hat\, y$$

Obs: Hat es idempotente y simétrica

$$SEE = y'(I - Hat)y$$

i) $(I - Hat)' = I' - Hat' = I - Hat$

ii) $(I - Hat)^2 = (I - Hat)(I - Hat)$

$$= I^2 - I\, Hat + Hat^2 - Hat\, I_n$$

$$= I_n - 2\, Hat + Hat$$

$$= I_n - Hat$$

$A = I_n - Hat$: Es simétrica e idempotente

$$\frac{n}{\sigma^2} e'e \sim \chi^2\left(\text{rango}(A), \lambda = \frac{1}{2\sigma^2}\mu' A \mu\right)$$

$$\begin{aligned}
\text{rango}(I_n - Hat) &= tr(I_n - Hat) \\
&= tr(I_n) - tr(Hat) \\
&= n - tr(X(X'X)^{-1}X') \\
&= n - tr((X'X)^{-1}X'X) \\
&= n - tr(I_k) \\
&= n - k \\
&= n - K
\end{aligned}$$

$\mu = E[y] = X\beta$ y $A = I - Hat$

$$\lambda = \frac{\mu' A \mu}{2\sigma^2} = \frac{1}{2\sigma^2}(X\beta)'(I_n - Hat)(X\beta)$$

$$= \frac{1}{2\sigma^2}\left[\beta'X'X\beta - \beta'X'\,Hat\,X\beta\right]$$

$$= \frac{1}{2\sigma^2}\left[\beta'X'X\beta - \beta'\cancel{X'(X(X'X)^{-1}X')}X\beta\right]$$

$$= \frac{1}{2\sigma^2}\left[\cancel{\beta'X'X\beta - \beta'X'X\beta}\right]$$

$\lambda = 0$

$$\therefore \quad \frac{n}{\sigma^2} SSE \sim \chi^2(n-K, 0)$$

ii) $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad \bar{Y} = \begin{bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix}$

$$= (Y - \bar{Y})'(Y - \bar{Y})$$

$$= (Y - 1_{n\times 1}\bar{y})'(Y - 1_{n\times 1}\bar{y})$$

$$= (Y - 1_{n\times 1}(1/n\, 1_{1\times n} Y))' \cdot$$

$$(Y - 1_{n\times 1}(1/n\, 1_{1\times n} Y))$$

$$= (Y - 1/n\, 1_{n\times n} Y)'(Y - 1/n\, 1_{n\times n} Y)$$

$$= (y' - y'\, 1/n\, 1_{n\times n})(y - 1/n\, 1_{n\times n} y)$$

$$= (y'\underbrace{(I_n - 1/n\, 1_{n\times n})}_{})((I_n - 1/n\, 1_{n\times n})y)$$

$$SST = (y' H)(H y)$$

$H = I_n - \frac{1}{n} 1_{n\times n}$ "Matriz de centrado"

Obs: $X_{n\times p} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$ $HX = \begin{bmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1p} - \bar{x}_p \\ \vdots & \ddots & \vdots \\ x_{np} - \bar{x}_1 & & x_{np} - \bar{x}_p \end{bmatrix}$

$$HX = (I_n - \tfrac{1}{n} 1_{n\times n}) X = X - 1/n\, 1_{n\times n} X$$

Obs:

i) $H' = (I_n - 1/n\, 1_{n\times n})' = I_n' - 1/n\, 1_{n\times n}'$

$$= I_n - 1/n\, 1_{n\times n} = H$$

ii) H es indempotente

$$H^2 = (I_n - 1/n \, 1_{n\times n})(I_n - 1/n \, 1_{n\times n})$$
$$= I_n I_n - I_n \, 1/n \, 1_{n\times n} + 1/n^2 \, \underline{1_{n\times n}^2} - 1/n \, 1_{n\times n} I_n$$
$$= I_n - 2 \, 1/n \, 1_{n\times n} + 1/n^2 \, n_{n\times n}$$
$$= I_n - 2 \, 1/n \, 1_{n\times n} + 1/n \, 1_{0\times n}$$
$$= I_n - 1/n \, 1_{n\times n}$$
$$H^2 = H \quad \therefore \; H \text{ es indempotente}$$

$$\therefore \; SST = (y' H)(H y) = y' H^2 y = y' H y$$

$$\Rightarrow \quad \frac{n}{\sigma^2} SST \sim \chi^2\left(\text{Rango}(H), \lambda = \frac{\mu' H \mu}{2\sigma^2} \right)$$

$$\text{rango}(H) = tr(H) = tr(I_n - 1/n \, 1_{n\times n})$$
$$= tr(I_n) - 1/n \, tr(1_{n\times n})$$
$$= n - \cancel{1/n (n)}$$
$$= n - 1$$

obs $\mu = \overline{y}$ por $E[y] = \overline{y}$

$$\lambda = \frac{1}{2\sigma^2} \mu' H \mu = \frac{1}{2\sigma^2} \mu' (I_n - 1/n \, 1_{n\times n}) \mu$$

$$= \frac{1}{2\sigma^2} \left(\mu' \mu - \frac{1}{n} \mu' \, 1_{n\times n} \mu \right]$$

$$= \frac{1}{2\sigma^2}\left(\bar{y}'\bar{y} - \frac{1}{n}\bar{y}' \mathbb{1}_{n\times n}\bar{y}'\right)$$

$$= \frac{1}{2\sigma^2}\left(\bar{y}\,\mathbb{1}_{1\times n}\underline{\mathbb{1}_{n\times 1}\bar{y}} - \frac{1}{n}\bar{y}\,\mathbb{1}_{1\times n}\mathbb{1}_{n\times n}\mathbb{1}_{n\times 1}\bar{y}\right)$$

$$= \frac{1}{2\sigma^2}\left(\bar{y}^2 n - \frac{\bar{y}^2}{\cancel{n}}\, n^{\cancel{2}}\right)$$

$$\lambda = \cancel{\frac{1}{2\sigma^2}\left(\bar{y}^2 n - \bar{y}^2 n\right)}^{0}$$

$$\therefore \lambda = 0$$

$$\therefore SST \sim \chi^2(n-1, 0) \sim \chi^2(n-1)$$

iii) $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

- Si $X \sim N_p(\mu, \Sigma)$ y A matriz entonces: $y = AX \in \mathbb{R}^q$
  $y \sim N_q(A\mu, A\Sigma A')$

Recordando $Y \sim N_n(X\beta, \sigma^2 I_n)$

$$\hat{\beta} = (X'X)^{-1}X'y = [(X'X)^{-1}X']\,y = Ay$$

$$\hat{\beta} \sim N(A\,E[y], A\Sigma_y A')$$

$$E[y] = X\beta \quad \therefore \quad A\,E[y] = \cancel{(X'X)^{-1}X'X}\beta = \beta$$

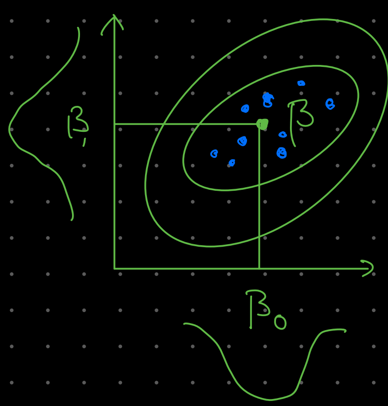
$$A\,\mathrm{Cov}(y)\,A' = (X'X)^{-1}X'(\sigma^2 I_n)\left((X'X)^{-1}X'\right)'$$

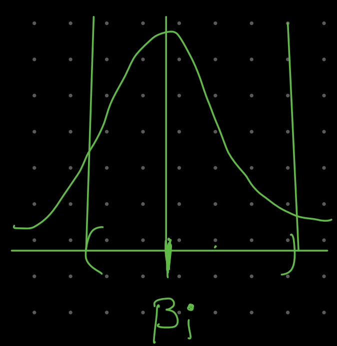
$$= \sigma^2 (X'X)^{-1}\cancel{X'X(X'X)^{-1}}$$

$$A\,Cov(y)\,A' = \sigma^2 (X'X)^{-1}$$

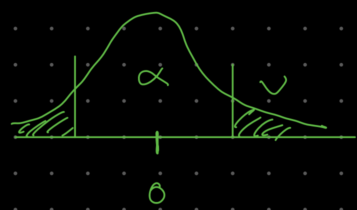
$$\Rightarrow \hat{\beta} \sim N_{k+1}\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

obs. $\hat{\beta}_i \sim N\left(\beta_i, \sigma^2 [(X'X)^{-1}]_{ii}\right)$ $\quad \sigma^2_{\hat{\beta}_i} = [\sigma^2 (X'X)^{-1}]_{ii}$





$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2_{\hat{\beta}_i}}} \sim N(0,1)$$



Coeficiente de determinación R^2

Definimos las fuentes de variación:

$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ Totales

$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ Residuales

$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ Regresión

$$R^2 = SSR/SST = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$SST = SSR + SSE$$

$$SST = [y - \bar{y}]'[y - \bar{y}]$$

$$= [(\hat{y} - \bar{y}) + (y - \hat{y})]'[(\hat{y} - \bar{y}) + (y - \hat{y})]$$

$$= [(\hat{y} - \bar{y})' + (y - \hat{y})'][(\hat{y} - \bar{y}) + (y - \hat{y})]$$

$$= (\hat{y} - \bar{y})'(\hat{y} - \bar{y}) + \cancel{(y - \hat{y})'(\hat{y} - \bar{y})} +$$

$$\cancel{(\hat{y} - \bar{y})'(y - \hat{y})} + (y - \hat{y})'(y - \hat{y})$$

$$SST = (\hat{y} - \bar{y})'(\hat{y} - \bar{y}) + (y - \hat{y})'(y - \hat{y})$$

$$SST = SSR + SSE$$

Teorema: $R^2 = \hat{\beta}'X'y \,/\, y'y$

$$(y - \bar{y})'(y - \bar{y}) = (y' - \bar{y}')(y - \bar{y})$$

$$= y'y - \bar{y}'y - y'\bar{y} + \bar{y}'\bar{y}$$

$$= y'y - 2n\bar{y}^2 + n\bar{y}^2$$

$$SST = y'y - n\bar{y}^2$$

$$SSR = (\hat{y} - \bar{y})'(\hat{y} - \bar{y}) = \hat{y}'\hat{y} \underbrace{- \bar{y}'\hat{y} - \hat{y}'\bar{y}} + \bar{y}'\bar{y}$$

$$= (X\hat{\beta})'(X\hat{\beta}) - 2\bar{y}'\hat{y} + n\bar{y}^2$$

$$= \hat{\beta}'(X'X)\hat{\beta} - 2\bar{y}'(X\hat{\beta}) + n\bar{y}^2$$

$$= \hat{\beta}' \cancel{(X'X)(X'X)^{-1}} X'y - 2n\bar{y}^2 + n\bar{y}^2$$

$$SSR = \hat{\beta} X'y - n\bar{y}^2$$

$$SSE = (y - \hat{y})'(y - \hat{y})$$

$$= (y' - \hat{\beta}'X')(y - X\hat{\beta})$$

$$= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'y$$

$$SSE = y'y - \hat{\beta}'X'y$$

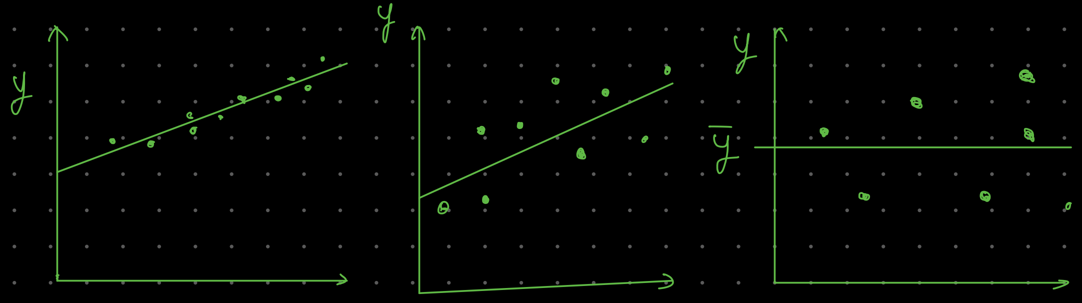
$$y'y - \cancel{n\bar{y}^2} = (\hat{\beta}'X'y - \cancel{n\bar{y}^2}) + (y'y - \hat{\beta}'X'y)$$

$$SST = SSR + SSE \Rightarrow$$

$$R^2 = \hat{\beta}'X'y / y'y = SSR/SST$$

Propiedades de R^2:

i) $0 \leq SSR \leq SST \Rightarrow 0 \leq R^2 \leq 1$



$R^2 \approx 1$ $\quad$ $R^2 \sim 0{,}5$ $\quad$ $R^2 \approx 0$

ii) Identidad

$$R^2 = 1 - \frac{(y-\hat{y})'(y-\hat{y})}{y'y}$$

iii) Sesgo: si $\beta = 0$ $(\beta_0 = 0, \beta_1 = 0, \ldots, \beta_K = 0)$

Entonces: $E[R^2] = K/n-1$!

Esto lo solucinamos ajustando el sesgo:

$$E[R_a^2] = 0 \text{ cuando } \beta = 0$$

$$R_a^2 = ((n-1)R^2 - K)/n-K-1$$

la R^2 cuadrada ajustada

Hay que recordar:

Como: $SST \sim \chi^2(n-1)$

$SSE \sim \chi^2(n-K)$

$SST = SSR + SSE$

$$\frac{1}{\sigma^2} SST = \frac{1}{\sigma^2} SSR + \frac{1}{\sigma^2} SSE$$

$$\underbrace{\chi^2(n-1)} = \underbrace{\chi^2(K-1)} + \underbrace{\chi^2(n-K)}$$

$$\therefore \frac{1}{\sigma^2} SSR \sim \chi^2(K-1)$$

Tabla ANOVA

$$F = \left[\frac{1}{\cancel{\sigma^2}(k-1)}\right] SSR \Big/ \left[\frac{1}{\cancel{\sigma^2}(n-k)}\right] SST$$

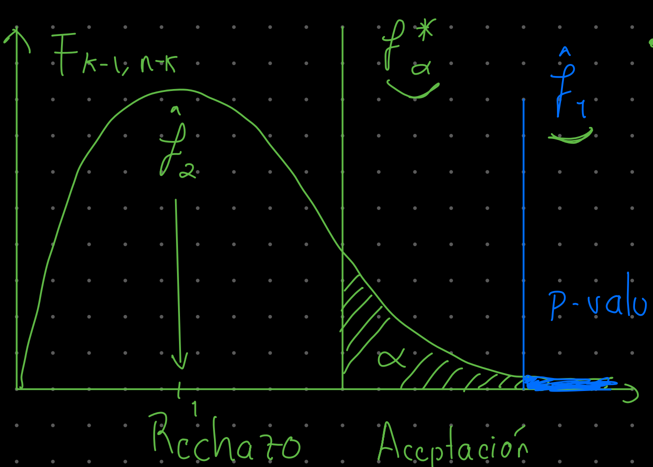
$$F = [1/(k-1)]\, SSR \,/\, [1/(n-k)]\, SST$$

$$F = [\chi^2(k-1)/k-1] \,/\, [\chi^2(n-k)/n-k]$$

$$F \sim Fisher(k-1, n-k)$$

Se define la tabla ANOVA

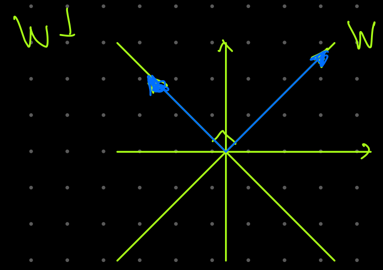
Fuente de variación	Suma de Cuadrados	Grados de libertad	Cuadrados medios	F
Regresión	$\hat{\beta}'X'y$	$k-1$	$SSR/k-1$	$\dfrac{SSR/k-1}{SSE/n-k}$
Residual	$(y-\hat{y})'(y-\hat{y})$	$n-k$	$SSE/n-k$	
Total	$y'y$	$n-1$		

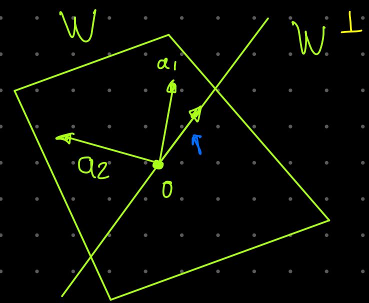


- Caso 1: Significativa
  $\hat{f}_1 > f^*_\alpha$
  con $1-F(\hat{f}_1) < \alpha$
- Caso 2: No significativa
  $\hat{f}_2 \leq f^*_\alpha$ con
  $1-F(\hat{f}_2) \geq \alpha$

Proyeciones:

Si $W \leq V$ $\quad W^{\perp} = \{ w \in V \mid \langle w, v \rangle = 0 \ v \in W \}$





Teorema:

Si $S = \{ a_1, \dots, a_n \}$ base de W y $A = [a_1 \cdots a_n]$

$$\boxed{W^{\perp} = \operatorname{Ker}(A')}$$

$$T[y] = \operatorname{Proy}_W y$$

$$\operatorname{Proy}_W y = \alpha_1 a_1 + \dots + \alpha_n a_n = \underline{A\alpha}$$

$$y - \operatorname{Proy}_W y = \operatorname{Proy}_{W^{\perp}} y \in W^{\perp} \in \operatorname{Ker}(A')$$

$$A' \operatorname{Proy}_{W^{\perp}} y = 0 \longleftarrow$$

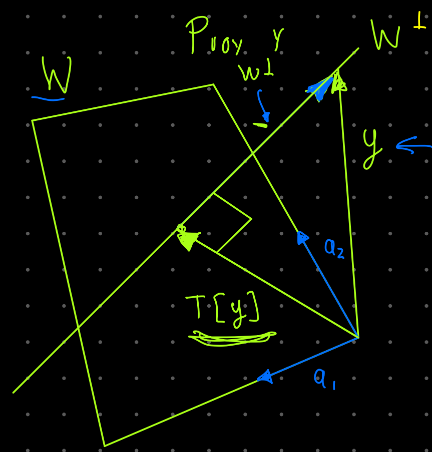
$$A'(y - \operatorname{Proy}_W y) = 0$$

$$A'y - A' \operatorname{Proy}_W y = 0 \longleftarrow$$

$$A'y = A' \underline{\operatorname{Proy}_W y}$$

$$A'y = A'A\alpha$$

$$(A'A)^{-1} A'y = \cancel{(A'A)^{-1}(A'A)}^{\,I} \alpha$$



$$A = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}$$

$$(A'A)^{-1}A'y = I\alpha$$

$$\underline{A(A'A)^{-1}A'}y = A\alpha$$

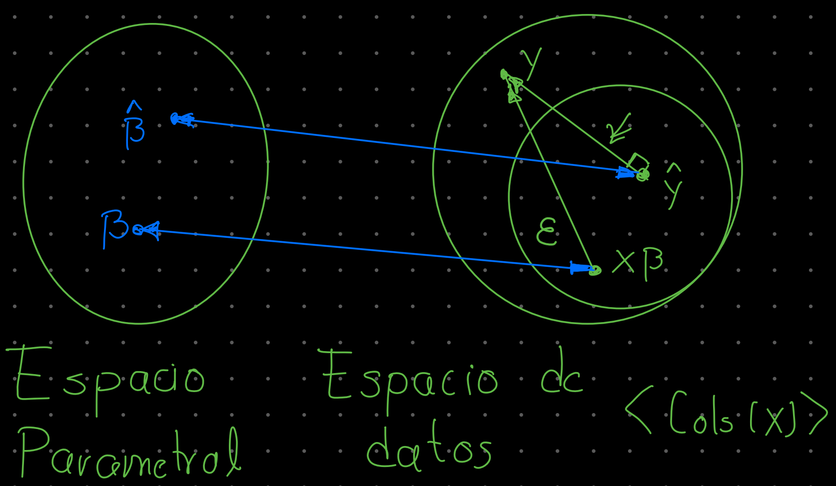
$$[P_A]y = Proy_W\, y$$

Obs: $X\hat{\beta} = \hat{y} \Rightarrow X(X'X)^{-1}X'y = Hat\ y = \hat{y}$

$$Hat = X(X'X)^{-1}X$$

$$Hat = Proy_{\langle Cols(X)\rangle}$$

$$\hat{y} = [Proy_{\langle Cols(X)\rangle}]\, y = Hat\ y$$



Hipotésis Lineal General

Supongamos que queremos contrastar la siguiente hipotesis nula:

$$H_0: C\beta = 0 \quad vs \quad H_a: C\beta \neq 0$$

Con $C \in M(\mathbb{R})\, q \times (K+1)$ conocida con rango $q_0 \leq K+1$

Teorema: Si $y \sim N_n(X\beta, \sigma^2 I)$ con $\text{rango}(C) = q_0 \leq K+1$

i) $C\hat{\beta} \sim N_q(C\beta, \sigma^2 C(X'X)^{-1}C')$

ii) $SSH = (C\hat{\beta})'[C(X'X)^{-1}C']C\hat{\beta}$

$\Rightarrow \dfrac{SSH}{\sigma^2} \sim \chi^2(q_0, \lambda)$; $\lambda = \dfrac{1}{2\sigma^2}(C\beta)'[C(X'X)^{-1}C']C\beta$

iii) $SSE = y'[I - X(X'X)^{-1}X']y$ $\leftarrow$

$\dfrac{SSE}{\sigma^2} \sim \chi^2(n-K-1, 0)$

iv) $SSE \perp SSH$

Definimos el contraste lineal general como

$$F_{H_0} = SSH/q_0 \,/\, SSE/(n-K-1)$$

- Si $H_0: C\beta = 0$ es falso:

$$F_{H_0} \sim F(q_0, n-K-1, \lambda)$$

• Si $H_0: C\beta = 0$ es verdadera

$$F_{H_0} \sim F(q_0, n-K-1)$$

Supongamos que queremos contrastar la siguiente hipotesis nula:

$$H_0: C\beta = t \quad vs \quad H_a: C\beta \neq t \quad con \ t \in \mathbb{R}^q$$

Con $C \in M(\mathbb{R})_{q \times (K+1)}$ conocida con rango $q_0 \leq K+1$

Teorema: Si $y \sim N_n(X\beta, \sigma^2 I)$ con $rango(C) = q_0 \leq K+1$

i) $C\hat{\beta} - t \sim N_q(C\beta - t, \sigma^2 C(X'X)^{-1}C')$

ii) $SSH = (C\hat{\beta} - t)'[C(X'X)^{-1}C'](C\hat{\beta} - t)$

$$\Rightarrow \frac{SSH}{\sigma^2} \sim \chi^2(q, \lambda) \ ; \ \lambda = \frac{1}{2\sigma^2}(C\beta - t)'[C(X'X)^{-1}C'](C\beta - t)$$

iii) $SSE = y'[I - X(X'X)^{-1}X']y$

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-K-1, 0)$$

iv) $SSE \perp SSH$

Definimos el estadistico contraste lineal general como

$$F_{H_0} = SSH/q_0 \Big/ SSE/(n-K-1)$$

• Si $H_0: C\beta = t$ es falsa:

$$F_{H_0} \sim F(q_0, n-K-1, \lambda)$$

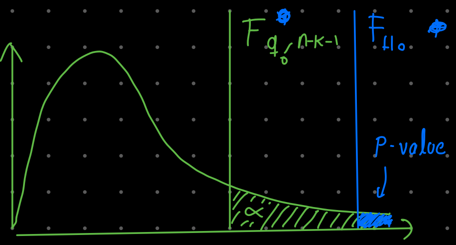
- Si $H_0: C\beta = 0$ es verdadera

$$F_{H_0} \sim F(q_0, n-k-1)$$

Regla de rechazo:

Rechazar H_0 si $F_{H_0} \geq F_{\alpha, q_0, n-k-1}$ donde $F_{\alpha, q_0, n-k-1} > \alpha$

es decir rechazar H_0 si $p \leq \alpha$ con $p =$ p-value F_{H_0} con $(q, n-k-1)$ g.l.



Ejemplos:

Contraste $H_0: C\beta = 0$ vs $H_a: C\beta \neq 0$ si $C = I_{k+1} \Rightarrow \text{rango}(I_{k+1}) = k+1$

$H_0: \beta = 0$ vs $H_a: \beta \neq 0$ (al menos una $\beta_i \neq 0$)

$$C\hat{\beta} \sim N_q(C\beta, \sigma^2 C(X'X)C') \quad q = k+1$$

$$\hat{\beta} \sim N_q(\beta, \sigma^2 (X'X))$$

$$\Rightarrow \frac{SSH}{\sigma^2} \sim \chi^2(q_0, \lambda) ; \quad \lambda = \frac{1}{2\sigma^2}(C\beta)'[C(X'X)'C']C\beta$$

$$\frac{SSH}{\sigma^2} \sim \chi^2(k+1, \lambda) ; \quad \lambda = \frac{1}{2\sigma^2}(\beta)'[(X'X)']\beta$$

$$\Rightarrow SSE = y'[I - X(X'X)^{-1}X']y$$

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$$

$$\Rightarrow SSH = (C\hat{\beta})'[C(X'X)^{-1}C']C\hat{\beta}$$

$$SSH = \hat{\beta}'[(X'X)^{-1}]\hat{\beta}$$

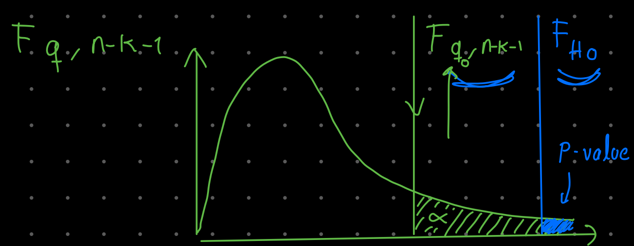
$$\Rightarrow \frac{SSH}{\sigma^2} \sim \chi^2(q, \lambda) \; ; \; \lambda = \frac{1}{2\sigma^2}(\beta)'[(X'X)^{-1}]\beta$$

$$SSE = y'[I - X(X'X)^{-1}X']y$$

$$F_{H_0} = \underline{(SSH/K+1)/SSE/(n-K-1)}$$

Rechazar H_0 si $F_{H_0} \geq F_{\alpha, q, n-K-1}$ donde $F_{\alpha, q, n-K-1} > \alpha$

es decir rechazar H_0 si $p \leq \alpha$ con $p = p\text{-value}$ F con $(q, n-K-1)$ g.l.



$q = K+1$

$\underline{H_0: \beta = 0}$

Ejemplo: Suponer el modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ queremos testear $\underline{\beta_1 = \beta_2 = \beta_0}$

$\underline{y = \beta_0 + \beta_0 x_1 + \beta_0 x_2}$ ie. $H_0: \beta_0 = \beta_1 = \beta_2$ vs $H_a: \beta_0 \neq \beta_1 \neq \beta_2$

$$H_0 \begin{cases} \beta_0 - \beta_1 = 0 \\ \underline{\beta_1 - \beta_2 = 0} \end{cases} \Rightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$C_{2\times3} \times \beta_{3\times1} = 0_{2\times1}$$

$$\text{Rango}(C_{2\times3}) = 2$$