# El modelo lineal II: Inferencia en el modelo de regresión lineal múltiple

Econometría 2021-2022

Grado Ingenierías & ADE

## Contenidos

## Valor Esperado y Varianza de los EMCO

El estimador EMCO de $\vec{\beta}$ del modelo $\vec{y} = X\vec{\beta} + \vec{u}$ es:

$$\hat{\vec{\beta}} = \left(X^t X\right)^{-1}\left(X^t \vec{y}\right)$$

Sustituyendo $\vec{y}$ por $\vec{y} = X\vec{\beta} + \vec{u}$:

$$\begin{aligned}
\hat{\vec{\beta}} &= \left(X^t X\right)^{-1}\left(X^t \vec{y}\right) = \left(X^t X\right)^{-1} X^t\left(X\vec{\beta} + \vec{u}\right) \\
&= \vec{\beta} + \left(X^t X\right)^{-1} X^t \vec{u}.
\end{aligned}$$

Por tanto,

$$\hat{\vec{\beta}} = \vec{\beta} + \left(X^t X\right)^{-1} X^t \vec{u}.$$

- El valor esperado del EMCO es:

$$E\left[\hat{\vec{\beta}}\right] = E\left[\vec{\beta} + \left(X^t X\right)^{-1}\left(X^t \vec{u}\right)\right] = \vec{\beta} + \left(X^t X\right)^{-1} X^t E[\vec{u}] = \vec{\beta}$$

ya que $E[\vec{u}] = \vec{0} \Rightarrow \hat{\vec{\beta}}$ es un estimador insesgado de $\vec{\beta}$.

## Valor Esperado y Varianza de los EMCO

- La matriz de varianzas-covarianzas del EMCO es:

$$\begin{aligned}
\operatorname{var}\left(\hat{\vec{\beta}}\right) &= E\left[\left(\hat{\vec{\beta}} - E\left[\hat{\vec{\beta}}\right]\right)\cdot\left(\hat{\vec{\beta}} - E\left[\hat{\vec{\beta}}\right]\right)^t\right] = \\
&= E\left[\left(\hat{\vec{\beta}} - \vec{\beta}\right)\cdot\left(\hat{\vec{\beta}} - \vec{\beta}\right)^t\right] = \\
&= E\left[\left(X^t X\right)^{-1}\left(X^t \vec{u}\right)\cdot\left(\vec{u}^{\,t} X\right)\left(X^t X\right)^{-1}\right] = \\
&= \left(X^t X\right)^{-1} X^t \cdot E\left[\vec{u}\cdot\vec{u}^{\,t}\right]\cdot X\left(X^t X\right)^{-1} = \\
&= \sigma^2\cdot\left(X^t X\right)^{-1} X^t X\left(X^t X\right)^{-1} = \sigma^2\cdot\left(X^t X\right)^{-1},
\end{aligned}$$

## Teorema de Gauss-Markov

**Teorema (Gauss–Markov)**

*Los estimadores mínimos cuadrados ordinarios de $\vec{\beta}$ son BLUE (ELIO en español):*

- *óptimos, (Best)*
- *lineales, (Linear) y*
- *insesgados (Umbiased)*

Según el Teorema de Gauss-Markov podemos de entre todos los estimadores lineales e insesgados de $\vec{\beta}$, el EMCO es el presenta menor matriz de varianzas-covarianzas.

## Teorema de Gauss–Markov: Demostración (I)

- $\hat{\vec{\beta}}$ es lineal con respecto al vector $\vec{y}$.
  Definiendo la matriz de dimensión $k \times n$ como $W = \left(X^t X\right)^{-1} X^t$ es fácil comprobar que $\hat{\vec{\beta}}$ es un estimador lineal con respecto a las observaciones $\vec{y}$, ya que
  $$\hat{\vec{\beta}} = \left(X^t X\right)^{-1} \left(X^t \vec{y}\right) = W \vec{y}$$
- $\hat{\vec{\beta}}$ es un estimador insesgado de $\vec{\beta}$.
  Si $\hat{\vec{\beta}}$ es un estimador insesgado de $\vec{\beta}$ entonces es porque se verifica que $E\left[\hat{\vec{\beta}}\right] = \vec{\beta}$ (ya lo sabemos).

## Teorema de Gauss–Markov: Demostración (II)

- $\overrightarrow{\widehat{\beta}}$ es un estimador óptimo de $\overrightarrow{\beta}$.

  Consideremos otro estimador lineal e insesgado, alternativo a $\overrightarrow{\widehat{\beta}}$: $\overrightarrow{\widehat{\beta^*}}$.

  - Como $\overrightarrow{\widehat{\beta^*}}$ es lineal en $\overrightarrow{y}$ entonces se verifica que existe $C_{k\times n}$ tal que $\overrightarrow{\widehat{\beta^*}} = C\overrightarrow{y}$.
  - Al verificarse que $\overrightarrow{\widehat{\beta^*}}$ es un estimador insesgado de $\overrightarrow{\beta}$ entonces $E\left[\overrightarrow{\widehat{\beta^*}}\right] = \overrightarrow{\beta}$. Luego:

$$
\begin{aligned}
E\left[\overrightarrow{\widehat{\beta^*}}\right] &= E\left[C\overrightarrow{y}\right] = E\left[C\left(X\overrightarrow{\beta} + \overrightarrow{u}\right)\right] = E\left[CX\overrightarrow{\beta}\right] + \\
&+ E\left[C\overrightarrow{u}\right] = CE\left[X\overrightarrow{\beta}\right] + CE\left[\overrightarrow{u}\right] = CX\overrightarrow{\beta}
\end{aligned}
$$

  Así que para que $\overrightarrow{\widehat{\beta^*}}$ sea insesgado se debe verificar: $CX = I_k$.

## Teorema de Gauss–Markov: Demostración (III)

$$
\begin{aligned}
\operatorname{var}\left(\overrightarrow{\widehat{\beta^*}}\right) &= E\left[\left(\overrightarrow{\widehat{\beta^*}} - E\left[\overrightarrow{\widehat{\beta^*}}\right]\right)\cdot\left(\overrightarrow{\widehat{\beta^*}} - E\left[\overrightarrow{\widehat{\beta^*}}\right]\right)^t\right] = \\
&= E\left[\left(\overrightarrow{\widehat{\beta^*}} - \overrightarrow{\beta}\right)\cdot\left(\overrightarrow{\widehat{\beta^*}} - \overrightarrow{\beta}\right)^t\right] = E\left[C\overrightarrow{u}\cdot\overrightarrow{u}^t C^t\right] = \\
&= CE\left[\overrightarrow{u}\cdot\overrightarrow{u}^t\right]C^t = C\sigma^2 I_n C^t = \sigma^2 CC^t
\end{aligned}
$$

Tenemos que $\overrightarrow{\widehat{\beta}} = W\overrightarrow{y}$ y que $\overrightarrow{\widehat{\beta^*}} = C\overrightarrow{y}$.
Si denotamos por $D = C - W$:

$$DX = (C - W)X = CX - WX = CX - \left(X^t X\right)^{-1} X^t X = I_k - I_k = 0_k$$

Luego:

$$
\begin{aligned}
\operatorname{var}\left(\overrightarrow{\widehat{\beta^*}}\right) &= \sigma^2 CC^t = \sigma^2 (W + D)(W + D)^t = \\
&= \sigma^2\left(WW^t + WD^t + DW^t + DD^t\right)
\end{aligned}
$$

## Teorema de Gauss–Markov: Demostración (IV)

$$\begin{aligned}
WW^t &= \left(\left(X^tX\right)^{-1}X^t\right)\left(\left(X^tX\right)^{-1}X^t\right)^t = \left(X^tX\right)^{-1}, \\
WD^t &= \left(\left(X^tX\right)^{-1}X^t\right)D^t = 0_{k\times k}, \\
DW^t &= D\left(\left(X^tX\right)^{-1}X^t\right)^t = 0_{k\times k},
\end{aligned}$$

luego:

$$\begin{aligned}
\text{var}\left(\overrightarrow{\widehat{\beta^*}}\right) &= \sigma^2\left(WW^t + WD^t + DW^t + DD^t\right) = \\
&= \sigma^2 WW^t + \sigma^2 DD^t = \sigma^2\left(X^tX\right)^{-1} + \sigma^2 DD^t = \\
&= \text{var}\left(\overrightarrow{\widehat{\beta}}\right) + \sigma^2 DD^t
\end{aligned}$$

$\text{var}\left(\overrightarrow{\widehat{\beta^*}}\right) - \text{var}\left(\overrightarrow{\widehat{\beta}}\right) = \sigma^2 DD^t \succ 0$, por tanto: $\overrightarrow{\widehat{\beta}}$ es el estimador con menor varianza entre todos los estimadores lineales e insesgados.

♯

## Estimación de $\sigma_u^2$

El vector de residuos es:

$$\begin{aligned}
\overrightarrow{e} &= \overrightarrow{y} - \overrightarrow{\widehat{y}} = \overrightarrow{y} - X\overrightarrow{\widehat{\beta}} = \overrightarrow{y} - X\left(\left(X^tX\right)^{-1}\left(X^t\overrightarrow{y}\right)\right) = \\
&= \left(I_n - X\left(X^tX\right)^{-1}X^t\right)\overrightarrow{y} = M_X\overrightarrow{y}
\end{aligned}$$

donde $M_X = I_n - X\left(X^tX\right)^{-1}X^t$ es la matriz complemento del proyector ortogonal.
La matriz $M_X$ verifica:

- $M_X^t = M_X$ (simetrica)
- $M_X^2 = M_X$ (idempotente)
- $traza(M_X) = n - k$.
- $M_X X = \left(I_n - X\left(X^tX\right)^{-1}X^t\right)X = X - X = 0_{n\times k}$.

# Estimación de $\sigma_u^2$

Como $\overrightarrow{y} = X\overrightarrow{\beta} + \overrightarrow{u}$:

$$\overrightarrow{e} = M_X\overrightarrow{y} = M_X\left(X\overrightarrow{\beta} + \overrightarrow{u}\right) = M_X X\overrightarrow{\beta} + M_X\overrightarrow{u} = M_X\overrightarrow{u}$$

Tomando valores esperados se tiene:

$$\begin{aligned}
E\left[\overrightarrow{e}^{\,t}\overrightarrow{e}\right] = E\left[\left(M_X\overrightarrow{u}\right)^t M_X\overrightarrow{u}\right] \qquad &= E\left[\overrightarrow{u}^{\,t}M_X\overrightarrow{u}\right] \\
= E\left[traza\left(\overrightarrow{u}^{\,t}M_X\overrightarrow{u}\right)\right] \qquad &= traza\left(M_X E\left[\overrightarrow{u}\,\overrightarrow{u}^{\,t}\right]\right) \\
= traza\left(M_X\sigma^2 I_n\right) \qquad &= \sigma^2 traza\left(M_X\right) = \sigma^2\left(n-k\right)
\end{aligned}$$

(Nota: $\overrightarrow{u}^{\,t}M_X\overrightarrow{u}$ es un escalar.)

Como $E\left[\dfrac{\overrightarrow{e}^{\,t}\overrightarrow{e}}{n-k}\right] = \sigma^2$, se concluye que un estimador insesgado de la varianza de las perturbaciones, $\sigma_u^2$, será:

$$\widehat{\sigma}_u^2 = \frac{\overrightarrow{e}^{\,t}\overrightarrow{e}}{n-k} = \frac{SCR}{n-k}$$

# Estimación de la varianza de los EMCO

La varianza del EMCO $\overrightarrow{\beta}$ era:

$$\mathrm{var}\left(\widehat{\overrightarrow{\beta}}\right) = \sigma^2\cdot\left(X^tX\right)^{-1},$$

Como dicha expresión depende de la varianza de la perturbación, no se puede determinar. Teniendo en cuenta que el estimador insesgado de la varianza del término de perturbación es

$$\widehat{\sigma}^2 = \frac{\overrightarrow{e}^{\,t}\overrightarrow{e}}{n-k} = \frac{SCR}{n-k},$$

se obtiene que la estimación de la matriz de varianzas-covarianzas de $\widehat{\overrightarrow{\beta}}$ es:

$$\widehat{\mathrm{var}}\left(\widehat{\overrightarrow{\beta}}\right) = \frac{SCR}{n-k}\cdot\left(X^tX\right)^{-1}.$$

## Supuesto de normalidad e inferencia

- Suponíamos que $\vec{u} \sim \mathcal{N}\left(\vec{0}, \sigma^2 I_n\right)$.
- Como $\vec{\hat{\beta}} = \vec{\beta} + \left(X^t X\right)^{-1} X^t \vec{u}$, y por T. Gauss-Markov:

$$\vec{\hat{\beta}} \sim \mathcal{N}\left(E\left[\vec{\hat{\beta}}\right], \mathrm{var}\left(\vec{\hat{\beta}}\right)\right) = \mathcal{N}\left(\vec{\beta}, \sigma^2 \cdot \left(X^t X\right)^{-1}\right)$$

o equivalentemente:

$$\left(\vec{\hat{\beta}} - \vec{\beta}\right) \sim \mathcal{N}\left(\vec{0}, \sigma^2 \cdot \left(X^t X\right)^{-1}\right)$$

- Como se verifica que:

$$\vec{e}^{\,t}\vec{e} = \vec{u}^{\,t} M_X \vec{u}$$
$$\vec{u} \sim \mathcal{N}\left(\vec{0}, \sigma^2 I_n\right)$$
$M_X$ es simétrica, idempotente y $\rho(M_X) = n - k$

Concluimos que $\frac{1}{\sigma^2}\vec{u}^{\,t} M_X \vec{u} \sim \chi^2_{n-k}$.
Por tanto

$$\frac{(n-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$$

## Intervalos de confianza

- IC para $\beta_i$:

$$\widehat{\beta}_i \pm t_{n-k,1-\frac{\alpha}{2}} \cdot \sqrt{\widehat{\mathrm{var}}[\widehat{\beta}_i]}, \quad i = 1, \ldots, k.$$

siendo $\widehat{\mathrm{var}}[\widehat{\beta}_i]$ el elemento $(i,i)$ de la matriz de varianzas-covarianzas estimada del estimador $\vec{\hat{\beta}}$, es decir, el elemento $(i,i)$ de $\widehat{\mathrm{var}}\left(\vec{\hat{\beta}}\right)$.

- IC para $\sigma^2$

$$\left[\frac{(n-k)\cdot\widehat{\sigma}^2}{\chi^2_{n-k,1-\frac{\alpha}{2}}}, \frac{(n-k)\cdot\widehat{\sigma}^2}{\chi^2_{n-k,\frac{\alpha}{2}}}\right],$$

donde $\chi^2_{n-k,1-\frac{\alpha}{2}}$ y $\chi^2_{n-k,\frac{\alpha}{2}}$ son los cuantiles de una distribución chi-cuadrado con $n-k$ grados de libertad tal que
$P[\chi \leq \chi^2_{n-k,1-\frac{\alpha}{2}}] = 1 - \frac{\alpha}{2}$ y $P[\chi \leq \chi^2_{n-k,\frac{\alpha}{2}}] = \frac{\alpha}{2}$.

# Contraste de hipótesis

Suponiendo $m$ restricciones lineales independientes entre sí:

$$\begin{array}{rcl} a_{11}\beta_1 + a_{12}\beta_2 + \cdots + a_{1k}\beta_k & = & r_1 \\ a_{21}\beta_1 + a_{22}\beta_2 + \cdots + a_{2k}\beta_k & = & r_2 \\ \vdots \qquad\qquad \vdots & = & \vdots \\ a_{m1}\beta_1 + a_{m2}\beta_2 + \cdots + a_{mk}\beta_k & = & r_m \end{array}$$

La hipótesis nula a contrastar será:

$$H_0 : R\overrightarrow{\beta} = \overrightarrow{r}$$

donde

$$R = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mk} \end{pmatrix}_{m\times k}, \qquad \overrightarrow{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix}_{m\times 1}.$$

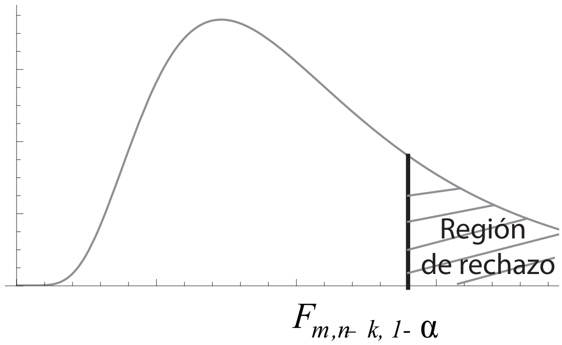
# Contraste de hipótesis

El estadístico de contraste asociado a la hipótesis $H_0 : R\overrightarrow{\beta} = \overrightarrow{r}$ es

$$\begin{aligned} F_{exp} &= \frac{\dfrac{\left(R\widehat{\overrightarrow{\beta}} - \overrightarrow{r}\right)^t \cdot \left[R\left(X^tX\right)^{-1}R^t\right]^{-1} \cdot \left(R\widehat{\overrightarrow{\beta}} - \overrightarrow{r}\right)}{m}}{\frac{\overrightarrow{e}^t\overrightarrow{e}}{n-k}} \\ &= \left(R\widehat{\overrightarrow{\beta}} - \overrightarrow{r}\right)^t \cdot \frac{\left[R\left(X^tX\right)^{-1}R^t\right]^{-1}}{m\widehat{\sigma}^2} \cdot \left(R\widehat{\overrightarrow{\beta}} - \overrightarrow{r}\right) \quad \sim F_{m,n-k}. \end{aligned}$$

# Contraste de hipótesis. Test general

La hipótesis nula se rechazará si el valor del estadístico cae en la región de rechazo.




Es decir, si $F_{exp} > F_{m,n-k,1-\alpha}$ entonces se rechaza la hipótesis $H_0 : R\overrightarrow{\beta} = \overrightarrow{r}$.

# Casos particulares (I): Significación individual

Si $m = 1$ y $r_i = 0\ \forall i$.

- Hipótesis:

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

- Estadístico de contraste:

$$t_{exp} = \left| \frac{\widehat{\beta}_i}{\sqrt{\widehat{\text{var}}[\widehat{\beta}_i]}} \right|$$

- Conclusión: Se rechaza $H_0$ si $t_{exp} > t_{n-k,1-\frac{\alpha}{2}}$.

## Casos particulares (II): Única restricción lineal

$m = 1$ y $r_i \neq 0\ \forall i$.

- Hipótesis: $\begin{cases} H_0 : \beta_i = r_i \\ H_1 : \beta_i \neq r_i \end{cases}$
- Estadístico de contraste:

$$t_{exp} = \left| \frac{\widehat{\beta}_i - r_i}{\sqrt{\widehat{\text{var}}[\widehat{\beta}_i]}} \right|$$

- Conclusión: Se rechaza $H_0$ si $t_{exp} > t_{n-k,1-\frac{\alpha}{2}}$.

## Casos Particulares (III): Significación global

Si $m = k - 1$ y $r_i = 0$ $i = 2, 3, \ldots, k$.

- Hipótesis: $\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \text{ con } i = 2, 3, \ldots, k \end{cases}$
- Estadístico de contraste:

$$F_{exp} = \left( R\,\overrightarrow{\widehat{\beta}} \right)^t \cdot \frac{\left[ R\left( X^t X \right)^{-1} R^t \right]^{-1}}{(k-1) \cdot \widehat{\sigma}^2} \cdot \left( R\,\overrightarrow{\widehat{\beta}} \right)$$

- Conclusión: Se rechaza $H_0$ si $F_{exp} > F_{k-1,n-k,1-\alpha}$.

## Contraste de Significación Global y la ANOVA

Teníamos:

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \text{ con } i = 2, 3, \ldots, k \end{cases}$$

Equivalentemente:

$$\begin{cases} H_0 : R\vec{\beta} = \vec{r} \\ H_1 : \text{No se verifica } H_0 \end{cases}$$

donde

$$R = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}_{(k-1)\times k} \qquad \vec{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{k\times 1} \qquad \vec{r} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{(k-1)\times 1}$$

## Contraste de Significación Global y la ANOVA

Realizando la partición de las matrices:

$$X^t X = \begin{bmatrix} 1^t \\ X_2^t \end{bmatrix} \begin{bmatrix} 1^t & X_2^t \end{bmatrix} = \begin{bmatrix} n & 1^t X_2 \\ X_2^t 1 & X_2^t X_2 \end{bmatrix} \qquad \vec{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \vec{\hat{\beta}}_2 \end{bmatrix}$$

Teniendo en cuenta que el número de restricciones es $k-1$ y las particiones realizadas, la expresión

$$\frac{\frac{\left(R\vec{\hat{\beta}} - \vec{r}\right)^t \cdot \left[R\left(X^t X\right)^{-1} R^t\right]^{-1} \cdot \left(R\vec{\hat{\beta}} - \vec{r}\right)}{m}}{\frac{\vec{e}^{\,t}\vec{e}}{n-k}}$$

queda reducida a

$$\frac{\frac{\vec{\hat{\beta}}{}_2^t \cdot \left[X_2^t X_2 - X_2^t 1 \frac{1}{n} 1^t X_2\right]^{-1} \cdot \vec{\hat{\beta}}_2}{k-1}}{\frac{\vec{e}^{\,t}\vec{e}}{n-k}} = \frac{\frac{\vec{\hat{\beta}}{}_2^t X_2^t A X_2 \vec{\hat{\beta}}_2}{k-1}}{\frac{\vec{e}^{\,t}\vec{e}}{n-k}} \sim F_{k-1,n-k}$$

donde $A = I_n - \frac{1}{n} 1 \cdot 1^t$.

## ANOVA

Considerando que se verifica

$$\begin{array}{ccccc} \overrightarrow{y}^t A \overrightarrow{y} & = & \overrightarrow{\widehat{\beta}}{}_2^t X_2^t A X_2 \overrightarrow{\widehat{\beta}}_2 & + & \overrightarrow{e}^t \overrightarrow{e} \\ (SCT) & = & (SCE) & + & (SCR) \end{array}$$

El estadístico experimental se puede escribir como:

$$F_{exp} = \frac{\frac{1}{k-1}}{\frac{1}{n-k}} \frac{SCE}{SCR}$$

## ANOVA

**Definición**

*ANOVA El* Análisis de la Varianza *es el contraste que estudia la significación global del modelo donde:*

- *Hipótesis* $\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_1 : \exists \beta_i \neq 0 \text{ con } i = 2, 3, \ldots, k \end{cases}$
- *Estadístico de contraste*

$$F_{exp} = \frac{\frac{1}{k-1}}{\frac{1}{n-k}} \frac{SCE}{SCR}$$

- *Conclusión* *Se rechaza* $H_0$ *si* $F_{exp} > F_{k-1,n-k,1-\alpha}$.

## Tabla ANOVA

| Fuente de variación | Suma de Cuadrados | Grados de Libertad | Medias |
|---|---|---|---|
| Explicada | $SCE = \overrightarrow{\widehat{\beta}}^t X^t \overrightarrow{y} - n\overline{Y}^2$ | $k-1$ | $\frac{SCE}{k-1}$ |
| Residuos | $SCR = \overrightarrow{y}^t \overrightarrow{y} - \overrightarrow{\widehat{\beta}}^t X^t \overrightarrow{y}$ | $n-k$ | $\frac{SCR}{n-k}$ |
| Total | $SCT = \overrightarrow{y}^t \overrightarrow{y} - n\overline{Y}^2$ | $n-1$ | |

Como

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

$$F_{exp} = \frac{\frac{1}{k-1}}{\frac{1}{n-k}} \frac{SCE}{SCT(1-R^2)} = \frac{\frac{1}{k-1}}{\frac{1}{n-k}} \frac{R^2}{(1-R^2)} = \frac{n-k}{k-1} \frac{R^2}{(1-R^2)}$$

## Cota $R^2$

Se rechaza $H_0$ cuando $F_{exp} > F_{k-1,n-k,1-\alpha}$, esto es si

$$\frac{n-k}{k-1} \frac{R^2}{(1-R^2)} > F_{k-1,n-k,1-\alpha} \Rightarrow R^2 > \frac{(k-1) \cdot F_{k-1,n-k,1-\alpha}}{(n-k) + (k-1) \cdot F_{k-1,n-k,1-\alpha}}.$$

## Predicción Puntual

Una vez estimado y validado el modelo $Y = \overrightarrow{\widehat{\beta}} X + u_t$:

- Si tenemos nuevos datos $\overrightarrow{x}_0^t = \begin{pmatrix} 1 & X_{20} & X_{30} & \dots & X_{k0} \end{pmatrix}$,
- Suponiendo que tenemos permanencia estructural en la especificación del modelo (el proceso de generación de los datos para la nueva observación $\overrightarrow{x}_0$ es el mismo que ha generado la información muestral)
- Podemos tomar como predictor para el valor medio y para el valor individual de la variable endógena

$$\overrightarrow{x}_0^{t} \overrightarrow{\widehat{\beta}} = \widehat{y}_0 = \widehat{\mathrm{E}[y_0 | \overrightarrow{x}_0]}$$

(el predictor mínimo cuadrático $\widehat{y}_0$ es lineal, insesgado y óptimo)

**Nota**

*Adviértase que, aunque se obtiene el mismo predictor para el valor individual $y_0$ como para $\widehat{E[y_0 | \overrightarrow{x}_0]}$, los errores de predicción obtenidos con ambas variables no coinciden entre sí, y por consiguiente las varianzas asociadas a dichos errores también serán distintas.*

## Predicción Puntual para el Valor Individual

Considerando el valor individual $y_0 = \overrightarrow{x}_0^t \overrightarrow{\beta} + u_0$ se tiene que el error de predicción corresponde:

$$e_0 = y_0 - \widehat{y}_0 = \overrightarrow{x}_0^t \overrightarrow{\beta} + u_0 - \overrightarrow{x}_0^t \overrightarrow{\widehat{\beta}} = u_0 - \overrightarrow{x}_0^t (\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta}) \sim \mathcal{N}(E[e_0], \mathrm{var}[e_0])$$

$$E[e_0] = E[u_0 - \overrightarrow{x}_0^t (\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})] = \overrightarrow{x}_0^t (E[\overrightarrow{\widehat{\beta}}] - \overrightarrow{\beta}) = 0.$$

## Predicción Puntual para el Valor Individual

$$
\begin{aligned}
\operatorname{var}[e_0] &= \operatorname{var}[u_0 - \overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})] = \operatorname{var}(u_0) + \operatorname{var}(\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})) - 2\operatorname{cov}(u_0\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta}) = \\
&= \operatorname{var}(u_0) + \operatorname{var}(\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})) - 2\operatorname{cov}(u_0\overrightarrow{x}_0^t(X^tX)^{-1}X^t\overrightarrow{u})) = {}^1 = \\
&= \operatorname{var}(u_0) + \operatorname{var}(\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})) = \sigma^2 + E[\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})^t\overrightarrow{x}_0] = \\
&= \sigma^2 + \overrightarrow{x}_0^t E[(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})^t]\overrightarrow{x}_0 = \sigma^2 + \overrightarrow{x}_0^t\sigma^2(X^tX)^{-1}\overrightarrow{x}_0 = \\
&=
\end{aligned}
$$

$$e_0 \sim N(0, \sigma^2(1 + \overrightarrow{x}_0^t(X^tX)^{-1}\overrightarrow{x}_0))$$

[1] $\operatorname{cov}(u_0\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})) = 0$ al no existir correlación entre $u_0$ y $\overrightarrow{u}$.

[1] $\operatorname{cov}(u_0\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})) = 0$ al no existir correlación entre $u_0$ y $\overrightarrow{u}$.

## Predicción Puntual para el Valor Esperado

$E[y_0|\overrightarrow{x}_0]$ tiene como residuo:

$$e_0^* = E[y_0|\overrightarrow{x}_0] - E\widehat{[y_0|\overrightarrow{x}_0]} = \overrightarrow{x}_0^t\overrightarrow{\beta} - \overrightarrow{x}_0^t\overrightarrow{\widehat{\beta}} = -\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta}) \sim \mathcal{N}(E[e_0^*], \operatorname{var}[e_0^*])$$

$$E[e_0^*] = E[-\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})] = \overrightarrow{x}_0^t(E[\overrightarrow{\widehat{\beta}}] - \overrightarrow{\beta}) = 0.$$

$$
\begin{aligned}
\operatorname{var}[e_0^*] &= \operatorname{var}[-\overrightarrow{x}_0^t(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})] = \overrightarrow{x}_0^t\operatorname{var}(\overrightarrow{\widehat{\beta}} - \overrightarrow{\beta})\overrightarrow{x}_0 = \\
&= \overrightarrow{x}_0^t\operatorname{var}(\overrightarrow{\widehat{\beta}})\overrightarrow{x}_0 = \overrightarrow{x}_0^t\sigma^2(X^tX)^{-1}\overrightarrow{x}_0 = \\
&= \sigma^2\overrightarrow{x}_0^t(X^tX)^{-1}\overrightarrow{x}_0.
\end{aligned}
$$

$$e_0^* = E[y_0|\overrightarrow{x}_0] - E\widehat{[y_0|\overrightarrow{x}_0]} \sim N(0, \sigma^2\overrightarrow{x}_0^t(X^tX)^{-1}\overrightarrow{x}_0)$$

## Predicción por intervalo

En las distirbuciones de $e_0$ y $e_0^*$, $\sigma^2$ es desconocida. Teniendo en cuenta que

$$\frac{(n-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k},$$

$$\frac{\frac{\widehat{y}_0 - \overrightarrow{x}_0^t \overrightarrow{\beta}}{\sqrt{\sigma^2(1+\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0)}}}{\sqrt{\frac{\frac{(n-k)\widehat{\sigma}^2}{\sigma^2}}{n-k}}} = \frac{\widehat{y}_0 - \overrightarrow{x}_0^t \overrightarrow{\beta}}{\widehat{\sigma}\sqrt{1+\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0}} \sim t_{n-k}$$

$$\frac{\frac{\widehat{E[y_0|\overrightarrow{x}_0]} - \overrightarrow{x}_0^t \overrightarrow{\beta}}{\sqrt{\sigma^2(\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0)}}}{\sqrt{\frac{\frac{(n-k)\widehat{\sigma}^2}{\sigma^2}}{n-k}}} = \frac{\widehat{E[y_0|\overrightarrow{x}_0]} - \overrightarrow{x}_0^t \overrightarrow{\beta}}{\widehat{\sigma}\sqrt{\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0}} \sim t_{n-k}$$

## Predicción por intervalo

- Intervalo de confianza para el valor individual $y_0$.

$$\begin{aligned} IC_{y_0} &= \widehat{y}_0 \pm t_{n-k,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}\sqrt{1+\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0} \\ &= \overrightarrow{x}_0^t \widehat{\overrightarrow{\beta}} \pm t_{n-k,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}\sqrt{1+\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0} \end{aligned}$$

- Intervalo de confianza para $E[y_0|\overrightarrow{x}_0]$.

$$\begin{aligned} IC_{E[y_0|\overrightarrow{x}_0]} &= \widehat{E[y_0|\overrightarrow{x}_0]} \pm t_{n-k,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}\sqrt{\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0} \\ &= \overrightarrow{x}_0^t \widehat{\overrightarrow{\beta}} \pm t_{n-k,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}\sqrt{\overrightarrow{x}_0^t (X^t X)^{-1} \overrightarrow{x}_0} \end{aligned}$$

# Distribuciones $\chi^2$ y $t$-Student

- La distribución $\chi^2$ de Pearson con $n$ grados de libertad, $\chi^2_n$, se construye como la suma de los cuadrados de $n$ variables aleatorias independientes e idénticamente distribuidas según una $\mathcal{N}(0,1)$:

$$\chi^2_n = \sum_{i=1}^{n} X_i^2, \quad X_i \sim N(0,1),\ \forall i.$$

- La distribución $t$-Student con $n$ grados de libertad, $t_n$, se construye como el cociente entre una variable aleatoria $\mathcal{N}(0,1)$ y la raíz cuadrada de una $\chi^2$-cuadrado de $n$ grados de libertad dividida entre sus grados de libertad, siendo ambas distribuciones independientes:

$$t_n = \frac{X}{\sqrt{\frac{Y}{n}}}, \quad X \sim N(0,1),\ \ Y \sim \chi^2_n.$$

Propiedades:

- $\chi^2$ NO es simétrica, $t$-Student SÍ es simétrica.
- Para $n > 30$, la distribución $t$-Student se puede aproximar a una distribución Normal.

# Distribución $F$-Snedecor

- La distribución $F$ de Snedecor con $n$ y $m$ grados de libertad, que denotaremos por $F_{n,m}$, se construye a partir del cociente de dos variables aleatorias independientes y distribuidas según chi-cuadrados con $n$ y $m$ grados de libertad, respectivamente, divididas entre sus correspondientes grados de libertad. Por tanto, la distribución $F$ de Snedecor responde a la siguiente estructura:

$$F_{n,m} = \frac{\frac{X}{n}}{\frac{Y}{m}}, X \sim \chi^2_n,\ \ Y \sim \chi^2_m.$$

Propiedades:

- La distribución $F_{n,m}$ NO es simétrica.
- $F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}$.
- $t_n^2 = F_{1,n}$.

Se define la matriz $A$ como

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}
\end{aligned}
$$

Propiedades:

- La matriz $A$ es simétrica e idempotente.
- $A(1, \dots, 1)^t = \overrightarrow{0}$ y $A\overrightarrow{e} = \overrightarrow{e}$.
- $A\overrightarrow{y}$ es:

$$
A\overrightarrow{y} = \begin{pmatrix} Y_1 - \overline{Y} \\ Y_2 - \overline{Y} \\ \vdots \\ Y_n - \overline{Y} \end{pmatrix}
$$