Clase auxiliar 4

Aplicaciones de probabilidades y estadística en gestión

Departamento de Ingeniería Civil Industrial
Universidad de Chile

Ronald Leblebici Garo

9 de septiembre de 2017

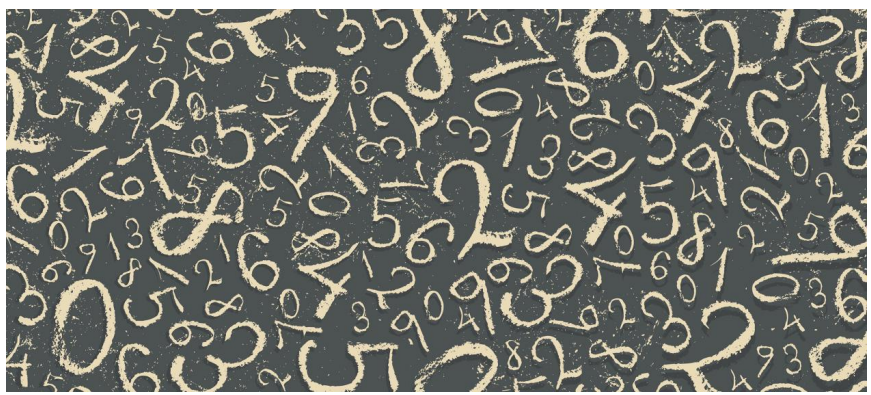
Antes de imprimir esta presentación, piense si es realmente necesario.

¿Qué veremos hoy?

- Breve repaso de teoría asintótica.
- Tests: significancia individual, global y restricciones lineales.
- Omisión de variables relevantes

Teoría asintótica

- Supongamos que tenemos un estimador $\hat{\theta}$ para un parámetro poblacional θ.
- Como buen estimador $\hat{\theta}$ es una función muestral. En particular, es función de la cantidad de muestras N.
- La teoría asintótica nos permite decir "en rigor, no tengo idea cómo se comporta $\hat{\theta}$, pero si N es suficientemente grande...".



- En palabras bonitas, la teoría asintótica nos permite conocer el comportamiento asintótico o distribución de probabilidad asintótica de nuestros estimadores de interés".

Teoría asintótica

La teoría asintótica es muy extensa y podríamos hacer un curso entero sobre ella. Por el momento, sólo necesitamos entender un par de conceptos y teoremas:

- Convergencia en probabilidad.
- Ley de los grandes números.
- Convergencia en distribución.
- Teorema de Slutsky.
- Teorema Central del Límite (TCL) univariado y multivariado.

Convergencia en probabilidad

Si la muestra es suficientemente grande, el estimador estará en una vecindad del verdadero parámetro.

$\hat{\theta}_N$ es un estimador de θ que converge en probabilidad a θ si:

$$\mathbb{P}(|\hat{\theta}_N - \theta| > \varepsilon) \to 0$$

$N = 2$

Ley débil de los grandes números

El promedio muestral converge en probabilidad a la esperanza.

Sea $\{X_i\}_{i=1}^N$ una muestra aleatoria i.i.d. con $\mathbb{E}(X_i) = \mu$ y $\mathbb{V}(X_i) = \sigma^2 < \infty$

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} \xrightarrow{p} \mu$$

Convergencia en distribución

Se dice que la función F_N converge en distribución a F si:

$$F_N(x) \to F(x), \forall x$$

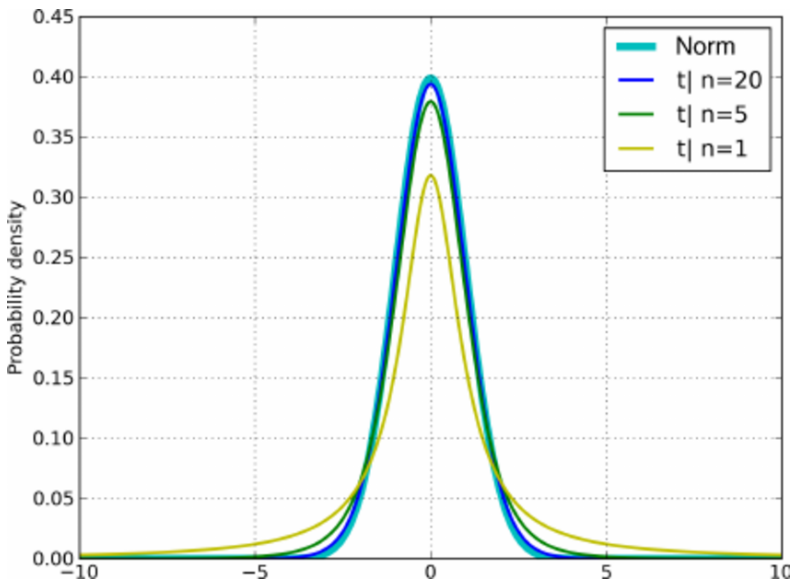


Figura: Una distribución T-Student con $N - K$ grados de libertad converge en distribución a una normal.

Teorema Central del Límite

El promedio variables aleatorias i.i.d. distribuye asintóticamente normal.

- Sea $\{X_i\}_{i=1}^N$ i.i.d. con $X_i \in \mathbb{R}$
- $\mathbb{E}(X_i) = \mu$.
- $\mathbb{V}(X_i) = \sigma^2 < \infty$.

$$\overline{X}_N \xrightarrow{d} \mathcal{N}(\mu, \sigma^2)$$

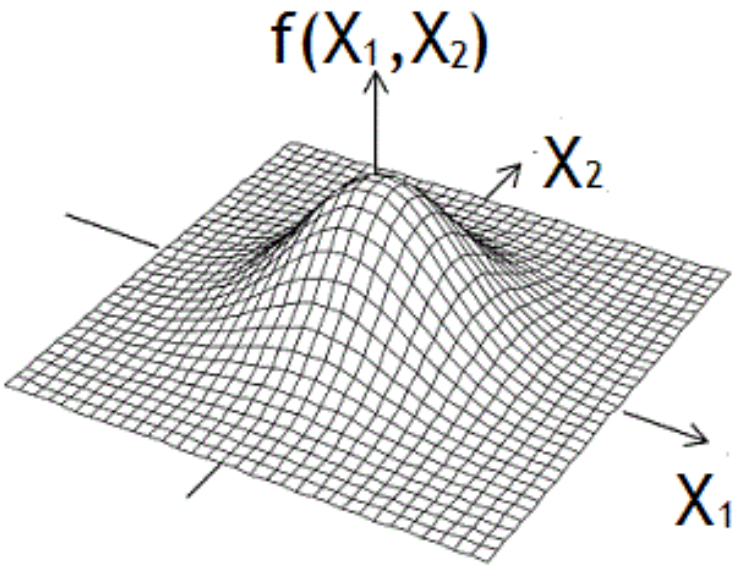
o equivalentemente...

$$\frac{\overline{X}_N - \mathbb{E}(\overline{X}_N)}{\sqrt{\mathbb{V}(\overline{X}_N)}} = \sqrt{N}\left(\frac{\overline{X}_N - \mu}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1)$$

Teorema Central del Límite Multivariado

- Sea $\{X_i\}_{i=1}^N$ i.i.d. con $X_i \in \mathbb{R}^K$
- $\mathbb{E}(X_i) = \vec{\mu}$.
- $\mathbb{V}(X_i) = \Sigma$.

$$\sqrt{N}(\overline{X}_N - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$



Teorema de Slutsky

Es como el álgebra de límites, pero para convergencias en probabilidad y distribución.

Si $X_n \xrightarrow{p} x$ e $Y_n \xrightarrow{d} Y$:

- $X_n Y_n \xrightarrow{d} xY$.
- $X_n + Y_n \xrightarrow{d} x + Y$

Teoremas de continuidad

Son como los límites de funciones para convergencias en probabilidad y distribución.

Teorema de continuidad: si $X_n \xrightarrow{p} x$ y $g(\cdot)$ es continua en x

$$g(X_n) \xrightarrow{p} g(x)$$

Teorema de mapeo continuo: si $X_n \xrightarrow{d} x$ y $g(\cdot)$ es continua en el soporte de x

$$g(X_n) \xrightarrow{p} g(x)$$

¿Qué resultados nos importan?

A partir de estos teoremas podemos demostrar (no lo haremos) lo siguiente:

- Si **conocemos** la varianza de los errores $\hat{\beta}$ distribuye Normal.
- Si **estimamos** la varianza de los errores $\hat{\beta}$ distribuye T-Student.
- El estadístico del test de **significancia global** distribuye F-Fischer.
- El estadístico del test de **restricciones múltiples** distribuye F-Fischer.

P1: Inferencia

Tenemos que saber interpretar la siguiente tabla. Iremos por partes...

```
. reg log_sal esc exp exp2
```

Source	SS	df	MS	Number of obs	=	97.926
				F(3, 97922)	=	11036,67
Model	15100,4243	3	5033,47478	Prob > F	=	0,0000
Residual	44659,1283	97.922	,456068384	R-squared	=	0,2527
				Adj R-squared	=	0,2527
Total	59759,5527	97.925	,610258388	Root MSE	=	,67533

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,114268	,0006729	169,82	0,000	,1129492	,1155869
exp	,0115993	,0004689	24,74	0,000	,0106803	,0125183
exp2	-,000063	8,17e-06	-7,71	0,000	-,000079	-,000047
_cons	7,38576	,0110628	667,62	0,000	7,364077	7,407443

P1.1: Significancia individual

Primero veamos qué distribución está ocupando *Stata* para los estimadores.

Primer intento

Si *Stata* conociera la matriz genérica de varianzas y covarianzas de los errores $\mathbb{V}(U)$...

$$\hat{\beta} \xrightarrow{d} \mathcal{N}(\beta, (X^T X)^{-1} X^T \mathbb{V}(U) X (X^T X)^{-1})$$

No conocemos $\mathbb{V}(U)$. Pero tenemos un supuesto que nos facilitará un poco la vida...

Segundo intento

El supuesto MCO de independencia y homocedasticidad nos dice que $\mathbb{V}(\sigma_U^2 I_{N\times N})$.

$$(X^T X)^{-1} X^T \mathbb{V}(U) X (X^T X)^{-1} = (X^T X)^{-1} X^T \sigma_U^2 I_{N\times N} X (X^T X)^{-1}$$

$$= \sigma_U^2 (X^T X)^{-1}$$

¡Excelente! Ahora...

$$\hat{\beta} \xrightarrow{d} \mathcal{N}(\beta, \sigma_U^2 (X^T X)^{-1})$$

¡Esperen un poco! σ_U^2 es un parámetro poblacional que nosotros y nuestro amigo *Stata* desconocemos.

P1.1: Significancia individual

Tercer intento

$$\hat{\beta} \xrightarrow{d} \mathcal{N}(\beta, \sigma_U^2 (X^T X)^{-1})$$

$$\Rightarrow \hat{\beta} - \beta \xrightarrow{d} \mathcal{N}(0, \sigma_U^2 (X^T X)^{-1})$$

Nos gustaría dividir por σ_U. Pero la gracia es que al lado izquierdo no tengamos datos desconocidos.

Por lo tanto, dividiremos por $\hat{\sigma}_U^2 = \frac{\hat{U}^T U}{N-K}$, que es un estimador consistente de σ_U^2. Los teoremas que vimos anteriormente nos permiten afirmar que:

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_U} \xrightarrow{d} \mathcal{T}_{N-K}$$

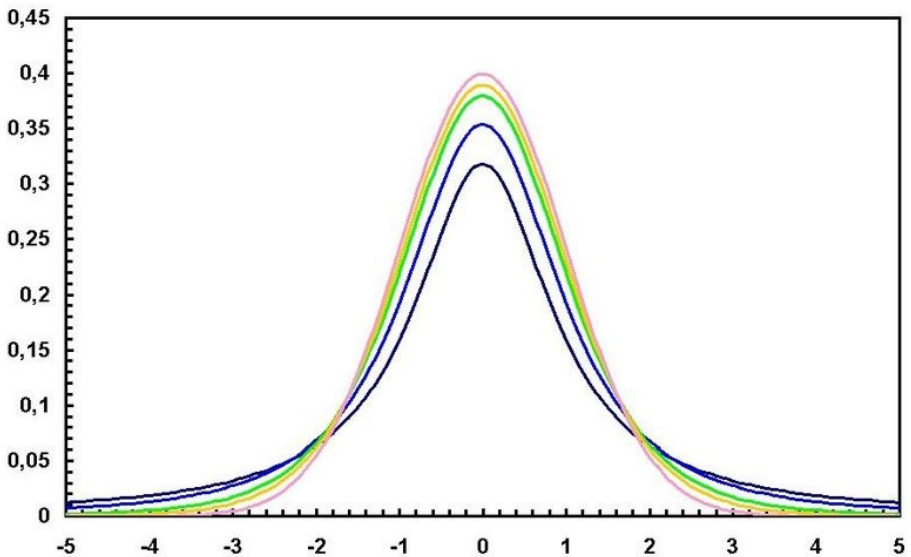
P1.1: Significancia individual

Tercer intento

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_U} \xrightarrow{d} \mathcal{T}_{N-K}$$

Bajo la hipótesis nula $H_0 : \beta_k = 0$ se tiene $\frac{\hat{\beta}_k}{\hat{\sigma}_U} \xrightarrow{d} \mathcal{T}_{N-K}$

La distribución T-Student es una campanita, que, cuando los $N - K$ grados de libertad que la caracteriza tienden a infinito, se parece mucho a la campanita de la distribución normal.



P1.2: Significancia individual

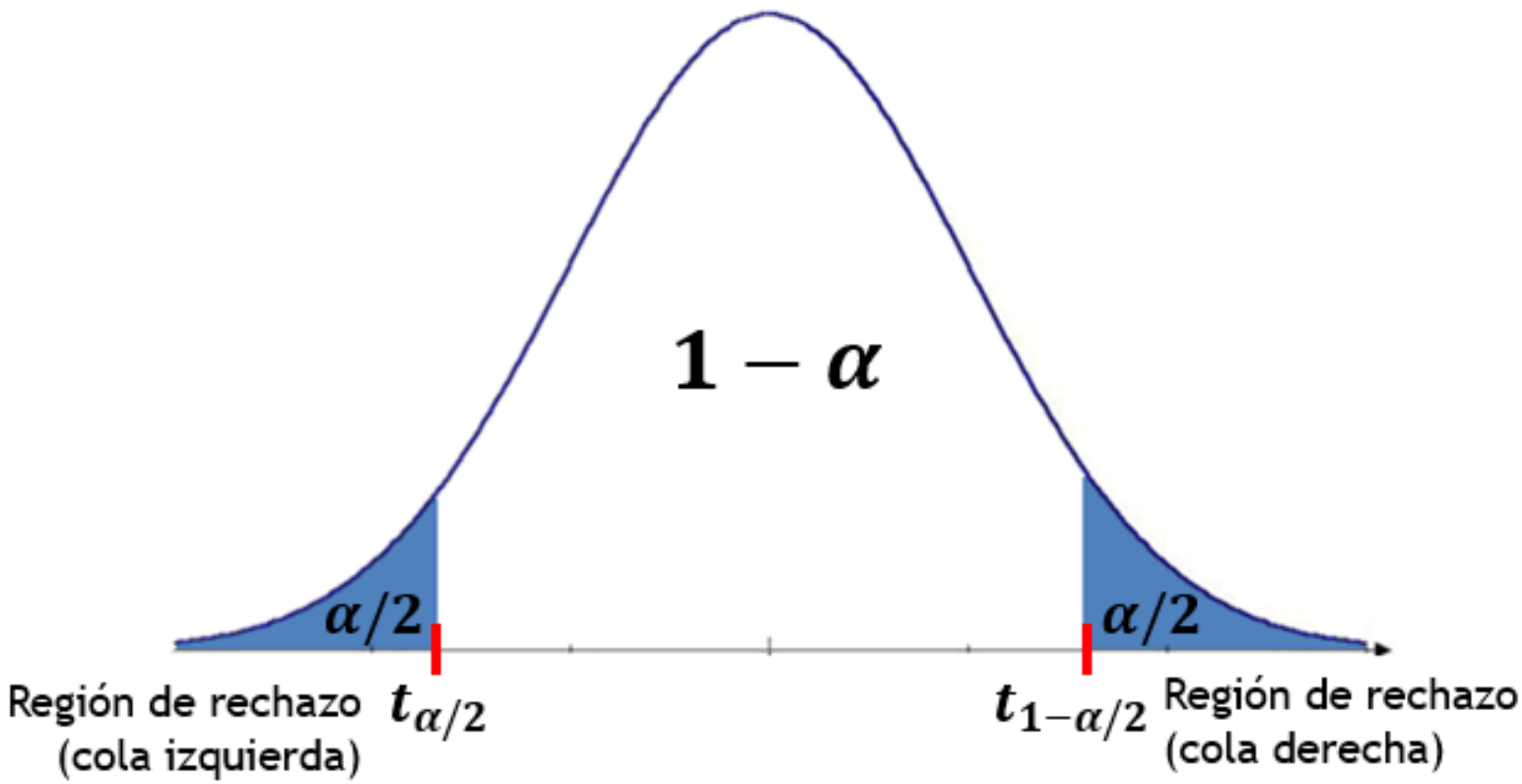
Antes de llevar a cabo el test debemos fijar un **nivel de significancia** (α).

¿Qué significa ésto?

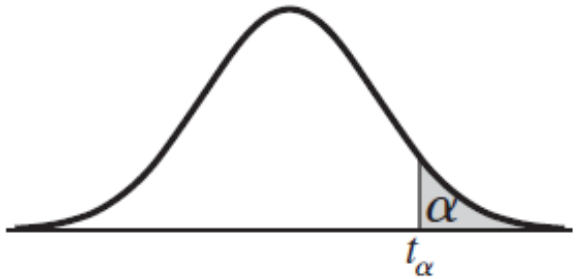
- Intentaremos rechazar una hipótesis nula, pero es imposible rechazarla con total certeza.
- Debemos estar dispuestos a aceptar una probabilidad de error tipo I (rechazar cuando no debimos haber rechazado).
- Esta probabilidad de error es el nivel de significancia (α).

P1.2: Significancia individual

Elegiremos el clásico $\alpha = 0{,}05$. Como haremos un test de dos colas, debemos obtener dos valores críticos, que definirán nuestra región de rechazo: $t_{\alpha/2}$ y $t_{1-\alpha/2}$.



P1.2: Significancia individual



$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	gl
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.315	1.706	2.056	2.479	2.779	26
1.314	1.703	2.052	2.473	2.771	27
1.313	1.701	2.048	2.467	2.763	28
1.311	1.699	2.045	2.462	2.756	29
1.282	1.645	1.960	2.326	2.576	inf.

P1.2: Significancia individual

De la tabla tenemos que $\hat{\beta}_{esc} = 0,114268$ y que $\hat{\sigma}_{\hat{\beta}_{esc}} = 0,0006729$.

$$\Rightarrow t = \frac{0,114268}{0,0006729} \approx 169,8 > 1,96$$

Como el estadístico t se encuentra en la región de rechazo, se rechaza la hipótesis nula.

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,114268	,0006729	169,82	0,000	,1129492	,1155869
exp	,0115993	,0004689	24,74	0,000	,0106803	,0125183
exp2	-,000063	8,17e-06	-7,71	0,000	-,000079	-,000047
_cons	7,38576	,0110628	667,62	0,000	7,364077	7,407443

Recordemos que $H_0 : \beta_{esc} = 0$.

Ésto quiere decir que existe evidencia estadística para rechazar la insignificancia individual de la variable escolaridad. O sea, la variable escolaridad es estadísticamente significativa.

P1.2: Significancia individual

Dijimos que la probabilidad de errar al momento de rechazar que estamos dispuestos a soportar es de $\alpha = 0{,}05$.

En este caso nuestra probabilidad de error tipo I, o p-valor es $0,000$.

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,114268	,0006729	169,82	0,000	,1129492	,1155869
exp	,0115993	,0004689	24,74	0,000	,0106803	,0125183
exp2	-,000063	8,17e-06	-7,71	0,000	-,000079	-,000047
_cons	7,38576	,0110628	667,62	0,000	7,364077	7,407443

Si esta probabilidad de errar al rechazar es menor, podemos rechazar con tranquilidad la hipótesis nula.

P1.2: Significancia individual

El intervalo de confianza que se ve en la tabla corresponde a $\hat{\beta}_k$, no a t.

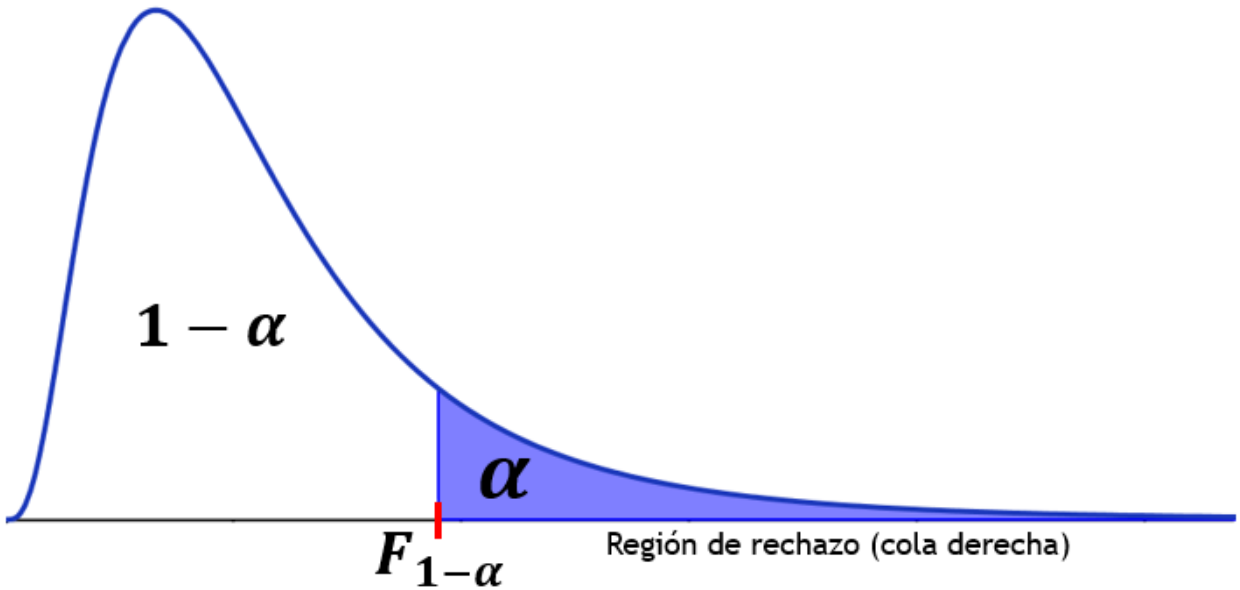
$$IC(\hat{\beta}_k) = [\hat{\beta}_k + z_{\alpha/2}\hat{\sigma}_{\hat{\beta}_k}, \hat{\beta}_k + z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_k}]$$

Dado que el cero no pertenece al intervalo de confianza, se rechaza la hipótesis nula de que $\beta_k = 0$.

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,114268	,0006729	169,82	0,000	,1129492	,1155869
exp	,0115993	,0004689	24,74	0,000	,0106803	,0125183
exp2	-,000063	8,17e-06	-7,71	0,000	-,000079	-,000047
_cons	7,38576	,0110628	667,62	0,000	7,364077	7,407443

P1.3: Significancia global

$$H_0 : \beta_1 = \beta_2 = ... = \beta_K = 0$$



$$\frac{R^2/(K-1)}{(1-R^2)/(N-K)} \sim \mathcal{F}(N-1, N-K)$$

P1.3: Significancia global

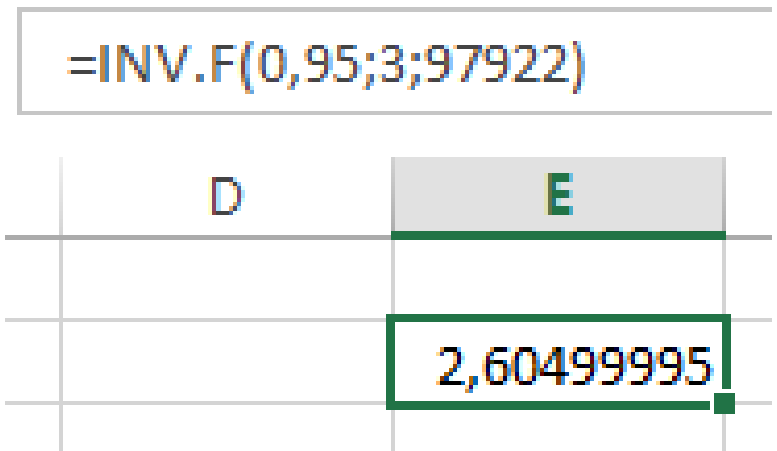
Number of obs	=	97.926
F(3, 97922)	=	11036,67
Prob > F	=	0,0000
R-squared	=	0,2527
Adj R-squared	=	0,2527
Root MSE	=	,67533

Reemplazando los valores de la tabla...

- $F = \frac{0{,}2527/(4-1)}{(1-0{,}2527)/(97{,}926-4)} = 11.037,4635$
- $\mathcal{F}(4-1, 97.926-4) = \mathcal{F}(3, 97.922)$
- Con $\alpha = 0{,}05$, se tiene $F_{1-\alpha}(3, 97.922) \approx 2,605$
- $F = 11.037,4635 > F_{1-\alpha}(3, 97.922) \approx 2,605$

P1.3: Significancia global

Entenderán que es difícil encontrar una tabla para la distribución F-Fischer con 3 grados de libertad en el numerador y 97.922 en el denominador. Sin embargo, siempre existen convenientes funciones en Excel, MATLAB, R, Stata, etc. para poder obtenerlos.



P1.4: Restricción lineal

¡Adivinen! Otro test T-Student. Si entendieron el de significancia individual, no deberían tener problemas con éste.

$$H_0 : \beta_{esc} + 2\beta_{exp} = 0,12$$

$$t = \frac{\hat{\beta}_{esc} + 2\hat{\beta}_{exp} - 0,12}{\widehat{DS}(\hat{\beta}_{esc} + 2\hat{\beta}_{exp})}$$

$$= \frac{1 \cdot 0,1185258 + 2 \cdot 0,012041 - 0,12}{\sqrt{(1)^2 \cdot (0,0011836)^2 + (2)^2 \cdot (0,0007351)^2 + 2 \cdot 1 \cdot 2 \cdot 0,0023}}$$

$$\approx 11,96 > 1,96$$

$\Rightarrow$ Se rechaza la hipótesis nula.

P1.5: Test de restricciones lineales múltiples

$H_0 : \beta_{esc} = 0 \wedge \beta_{exp} = 0$

- Modelo restringido: $X = [\vec{1}$ escolaridad$]$
- Modelo libre: $X = [\vec{1}$ escolaridad experiencia experiencia$^2]$

$$F = \frac{(R_L^2 - R_R^2)/r}{(1 - R_L^2)/(N - K)} \sim \mathcal{F}(r, N - K)$$

- $F = \frac{(0{,}2527 - 0{,}2337)/2}{(1 - 0{,}2527)/(97.926 - 4)} = 1.244,83$
- Con $\alpha = 0,05$, se tiene que $\mathcal{F}_{1-\alpha}(r, N - K) = \mathcal{F}_{1-\alpha}(2, 97.926 - 4) \approx 2,996.$
- $F = 1.244,83 > \mathcal{F}_{1-\alpha}(2, 97.926 - 4) \approx 2,996$

$\Rightarrow$ Se rechaza la hipótesis nula.

P2.1: Sesgo al omitir variables relevantes

PDQ: $\hat{\beta}_{MCO}$ es sesgado cuando se omiten variables relevantes.

Modelo estimado: $Y = X_1\beta_1 + U$

Modelo verdadero: $Y = X_1\beta_1 + X_2\beta_2 + U$

$$\begin{aligned}\hat{\beta}_1 &= (X_1^T X_1)^{-1} X_1^T Y \\ &= (X_1^T X_1)^{-1} X_1^T (X_1\beta_1 + X_2\beta_2 + U) \\ &= \underbrace{(X_1^T X_1)^{-1} X_1^T X_1}_{Id_{K_1 \times K_1}} \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2\beta_2 + (X_1^T X_1)^{-1} X_1^T U\end{aligned}$$

$$\Rightarrow \mathbb{E}(\hat{\beta}) = \underbrace{\mathbb{E}(\beta_1)}_{\beta_1} + \underbrace{\mathbb{E}[(X_1^T X_1)^{-1} X_1^T X_2\beta_2]}_{\neq 0 \text{ (sesgo)}} + \underbrace{\mathbb{E}[(X_1^T X_1)^{-1} X_1^T U]}_{=0} \neq \beta_1 \quad \square$$

Comentario

Es importante entender que bajo omisión de variables relevantes ocurren otros dos fenómenos:

- $\hat{\beta}$ es un estimador **inconsistente** de β, pues como el sesgo **no** converge en probabilidad a cero, $\hat{\beta}$ **no** converge en probabilidad a β.

- Las variables del modelo "absorben" el efecto de variables omitidas, por lo que al estimar omitiendo variables relevantes creemos tener mayor información de la que en realidad tenemos.

  $\therefore$ La varianza estimada $\hat{\mathbb{V}}(\hat{\beta}_1)$ es **erróneamente más eficiente** que la verdadera varianza $\mathbb{V}(\hat{\beta}_1)$.

P2.2: Efectos de omitir variables relevantes

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,0974079	,0005637	172,80	0,000	,0963031	,0985128
_cons	7,818565	,0068344	1144,01	0,000	7,80517	7,831961

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,114268	,0006729	169,82	0,000	,1129492	,1155869
exp	,0115993	,0004689	24,74	0,000	,0106803	,0125183
exp2	-,000063	8,17e-06	-7,71	0,000	-,000079	-,000047
_cons	7,38576	,0110628	667,62	0,000	7,364077	7,407443

- Como hemos explicado, al omitir variables $\hat{\beta}_{esc}$ se vuelve sesgado (e inconsistente, el sesgo no se va con muchas muestras), lo que explica la diferencia entre la estimación con y sin omisión de variables relevantes.
- Por otro lado, se observa que al omitir estas variables $\hat{\mathbb{V}}(\hat{\beta}_{esc})$ disminuye artificialmente porque $\hat{\beta}_{esc}$ está abarcando más información de la que le corresponde.