Clase auxiliar 12

Aplicaciones de probabilidades y estadística en gestión

Departamento de Ingeniería Civil Industrial
Universidad de Chile

Ronald Leblebici Garo

2 de diciembre de 2017

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

Econometría

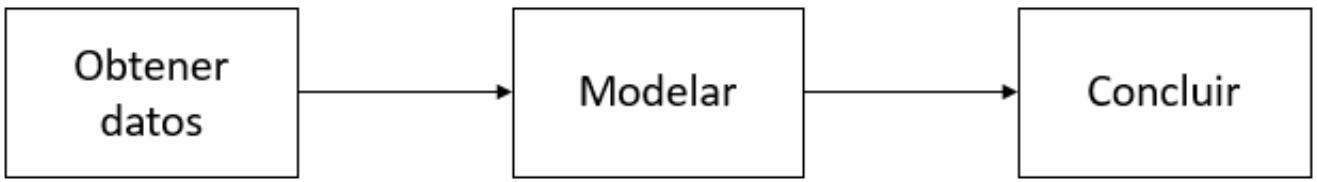
¿Qué es?

- Rama de la economía que brinda **contenido empírico** a relaciones teóricas.
- Permite analizar, interpretar y hacer predicciones sobre variables económicas a través de modelos que pueden ser testeados.
- Aplica matemáticas, estadística y computación.

Econometría

¿Por qué nos interesa estudiarla?

- Generar información consistente para la toma de decisiones.
- Proponer y generar modelos estadísticos que permitan hacer inferencias válidas y coherentes sobre los datos que se manejen en una organización.
- Obtener e interpretar inferencias causales.



¿Con qué problemas podemos encontrarnos en este proceso?

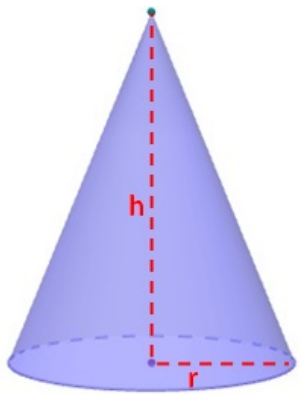
Modelo determinístico: una muestra de conos

Imaginemos un salón lleno de conos perfectos, de distintas alturas y distintos radios, pero todos conos al fin y al cabo.

Modelo determinístico: una muestra de conos

Imaginemos un salón lleno de conos perfectos, de distintas alturas y distintos radios, pero todos conos al fin y al cabo.

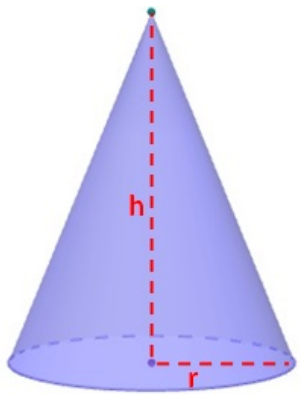
El volumen de un cono está dado por $V = \frac{\pi r^2 h}{3}$.



Modelo determinístico: una muestra de conos

Imaginemos un salón lleno de conos perfectos, de distintas alturas y distintos radios, pero todos conos al fin y al cabo.

El volumen de un cono está dado por $V = \frac{\pi r^2 h}{3}$.



Tenemos una fórmula completamente determinística. Teniendo las variables **independientes** –el radio (r) y la altura de un cono (h)– podemos obtener sin error alguno el valor de la variable **dependiente** (V).

Modelo aleatorio: demanda de camisetas de la selección

Supongamos que en una multitienda se tienen datos diarios relacionados a ventas de camisetas de la selección.

Modelo aleatorio: demanda de camisetas de la selección

Supongamos que en una multitienda se tienen datos diarios relacionados a ventas de camisetas de la selección.

Variable dependiente: camisetas vendidas por día (Y).

Modelo aleatorio: demanda de camisetas de la selección

Supongamos que en una multitienda se tienen datos diarios relacionados a ventas de camisetas de la selección.

Variable dependiente: camisetas vendidas por día (Y).

Variables independientes: días restantes para el próximo campeonato y goles que hizo Chile en el último partido (X_1, X_2).

Modelo aleatorio: demanda de camisetas de la selección

Supongamos que en una multitienda se tienen datos diarios relacionados a ventas de camisetas de la selección.

Variable dependiente: camisetas vendidas por día (Y).

Variables independientes: días restantes para el próximo campeonato y goles que hizo Chile en el último partido (X_1, X_2).

¿Si conozco X_1 y X_2, puedo conocer con total certeza Y?

Modelo aleatorio

En ciencias sociales, la no-existencia de componentes aleatorios ($Y_i = f(X_i)$) es un supuesto muy exigente. **Las relaciones nunca son perfectas**.

Modelo aleatorio

En ciencias sociales, la no-existencia de componentes aleatorios ($Y_i = f(X_i)$) es un supuesto muy exigente. **Las relaciones nunca son perfectas**.

$\therefore$ Suena razonable agregar una **variable aleatoria** al modelo. En consecuencia, trabajaremos con un modelo de la forma:

$$Y_i = f(X_i, U_i)$$

Modelo aleatorio

En ciencias sociales, la no-existencia de componentes aleatorios ($Y_i = f(X_i)$) es un supuesto muy exigente. **Las relaciones nunca son perfectas**.

$\therefore$ Suena razonable agregar una **variable aleatoria** al modelo. En consecuencia, trabajaremos con un modelo de la forma:

$$Y_i = f(X_i, U_i)$$

Luego, haremos un supuesto más fuerte: la linealidad de los parámetros:

Modelo aleatorio

En ciencias sociales, la no-existencia de componentes aleatorios ($Y_i = f(X_i)$) es un supuesto muy exigente. **Las relaciones nunca son perfectas**.

$\therefore$ Suena razonable agregar una **variable aleatoria** al modelo. En consecuencia, trabajaremos con un modelo de la forma:

$$Y_i = f(X_i, U_i)$$

Luego, haremos un supuesto más fuerte: la linealidad de los parámetros:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i$$

Linealidad de los parámetros

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i$$

Linealidad de los parámetros

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i$$

Notar que la linealidad es sobre los parámetros, no sobre las variables. Podemos tener perfectamente un modelo de la forma:

Linealidad de los parámetros

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i$$

Notar que la linealidad es sobre los parámetros, no sobre las variables. Podemos tener perfectamente un modelo de la forma:

$$\tanh(Y_i) = \beta_1 \ln(X_{i1}) + \beta_2 \exp(X_{i2}) + \beta_3 X_{i3}^{0,7} + \beta_4 \arctan(X_{i3}) + U_i$$

¿Y si quiero agregarle un intercepto a mi modelo?

Basta imponer que una de las variables dependientes sea 1 para todas las observaciones, ¿no?

Linealidad de los parámetros

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i$$

Notar que la linealidad es sobre los parámetros, no sobre las variables. Podemos tener perfectamente un modelo de la forma:

$$\tanh(Y_i) = \beta_1 \ln(X_{i1}) + \beta_2 \exp(X_{i2}) + \beta_3 X_{i3}^{0,7} + \beta_4 \arctan(X_{i3}) + U_i$$

¿Y si quiero agregarle un intercepto a mi modelo?

Basta imponer que una de las variables dependientes sea 1 para todas las observaciones, ¿no?

$$Y_i = \beta_1 \cdot 1 + \beta_2 X_{i2} + ... + \beta_K X_{iK} + U_i$$

Notación

- Número de observaciones: $N \in \mathbb{R}$.

Notación

- Número de observaciones: $N \in \mathbb{R}$.
- Número de variables explicativas : $K \in \mathbb{R}$.

Notación

- Número de observaciones: $N \in \mathbb{R}$.
- Número de variables explicativas : $K \in \mathbb{R}$.
- Vector de observaciones de la variable dependiente:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

Notación

- Número de observaciones: $N \in \mathbb{R}$.
- Número de variables explicativas : $K \in \mathbb{R}$.
- Vector de observaciones de la variable dependiente:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

- Matriz de observaciones de las variables independientes:

$$X = \begin{pmatrix} X_{1,1} & \cdots & \cdots & \cdots & X_{1,K} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ \vdots & \cdots & X_{i,k} & \cdots & \vdots \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ X_{N,1} & \cdots & \cdots & \cdots & X_{N,K} \end{pmatrix} \in \mathbb{R}^{N \times K}$$

Notación

- Vector de parámetros poblacionales:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \in \mathbb{R}^{K \times 1}$$

Notación

- Vector de parámetros poblacionales:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \in \mathbb{R}^{K \times 1}$$

- Vector de errores aleatorios:

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

Notación

- Vector de parámetros poblacionales:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \in \mathbb{R}^{K \times 1}$$

- Vector de errores aleatorios:

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

También se suele notar como ε.

Notación

- Vector de parámetros poblacionales:

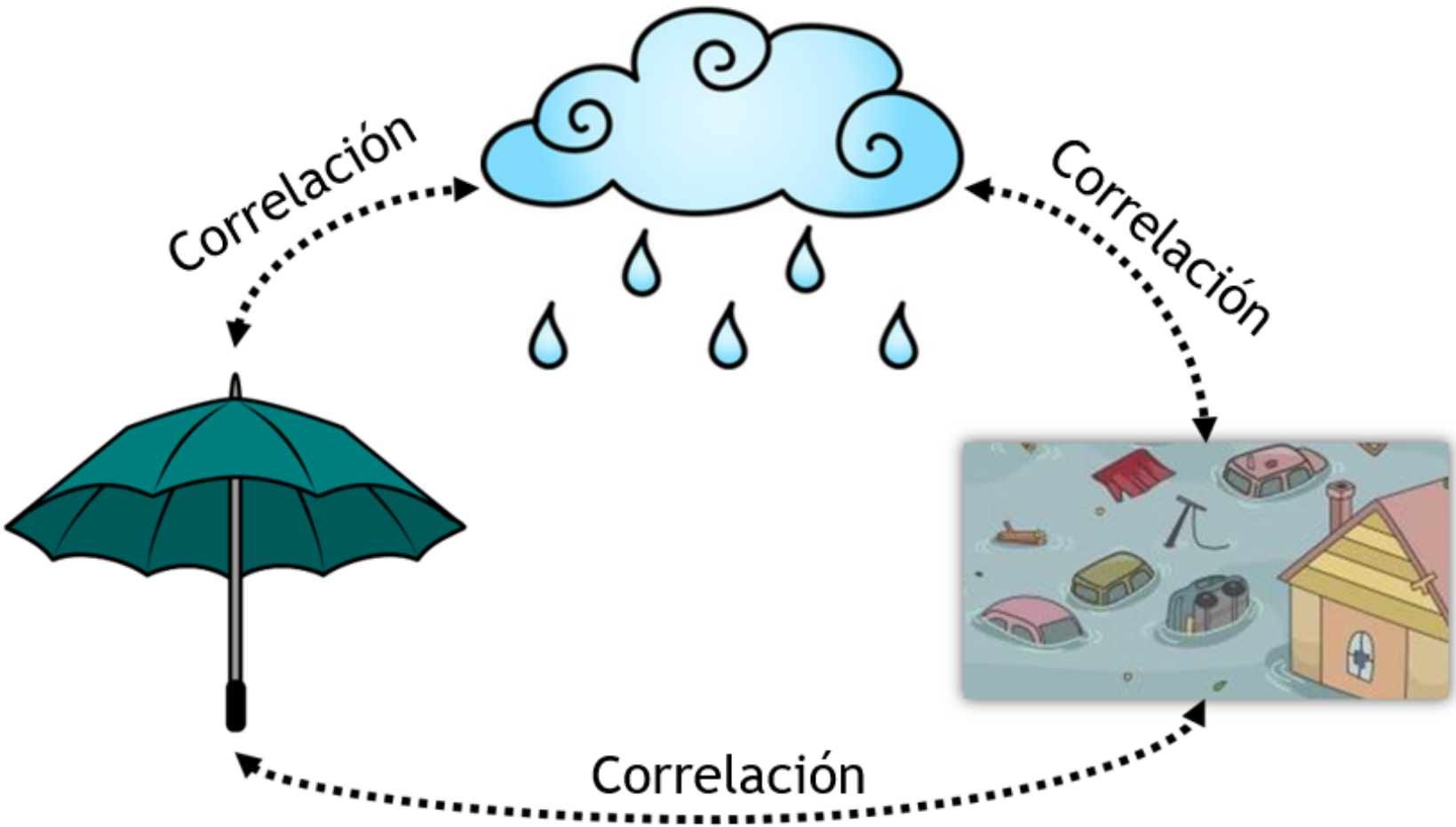
$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \in \mathbb{R}^{K \times 1}$$

- Vector de errores aleatorios:

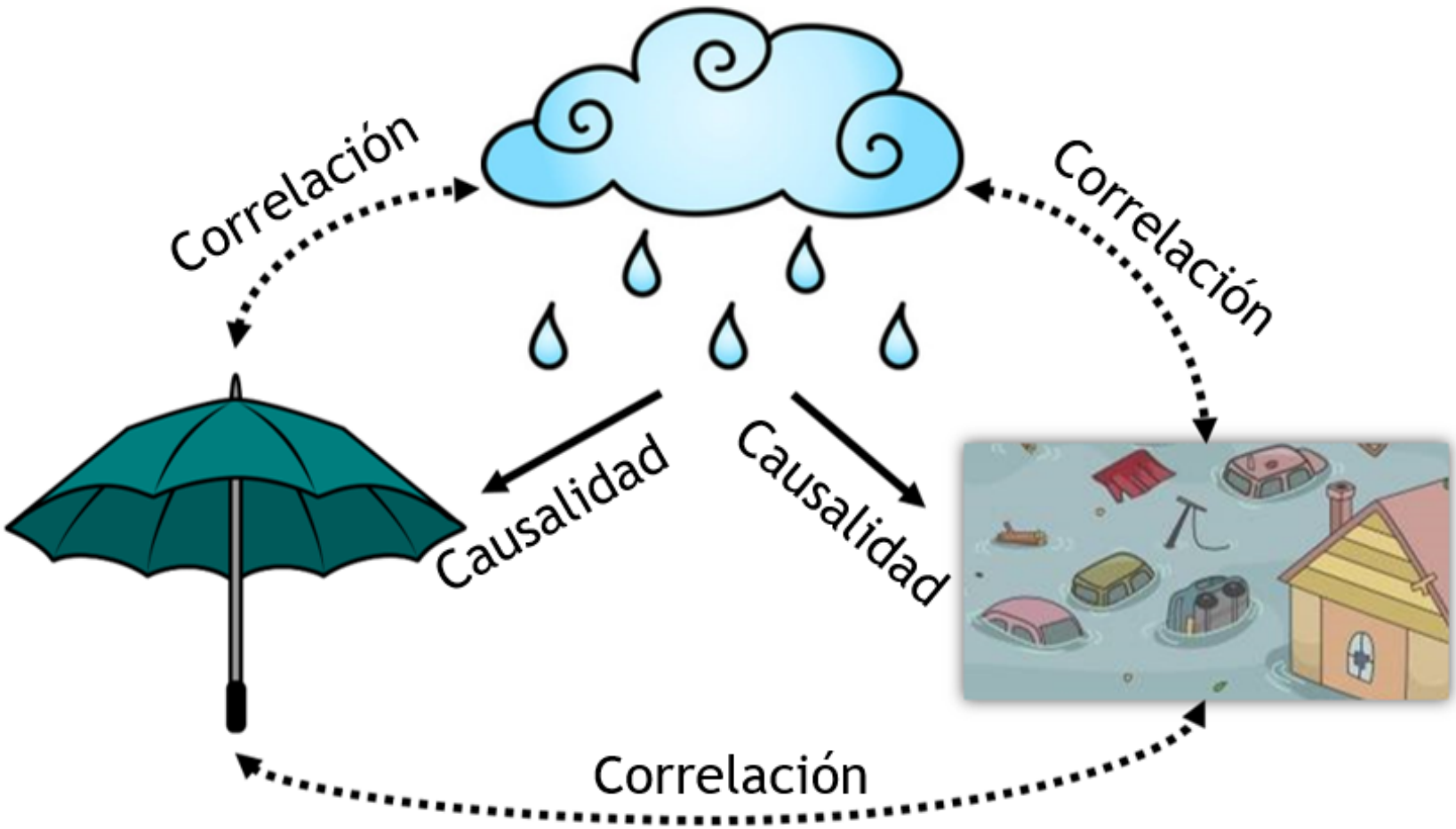
$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

También se suele notar como ε.

Causalidad y correlación: lluvia, paraguas y calles inundadas



Causalidad y correlación: lluvia, paraguas y calles inundadas



Betas

Ahora que sabemos que existe una **causalidad** entre los ml. de lluvia y los ml. de inundación en un sector, podría hacernos sentido plantear el siguiente modelo:

$$\text{Inundación}_i = \beta_0 + \beta_{Lluvia}\text{Lluvia}_i + U_i$$

Digamos que esta información corresponde a mediciones obtenidas a las afueras de Beauchef 850 y que cada observación i representa un día.

Donde el parámetro $\beta_{Lluvia} = \frac{\partial \text{Inundación}_i}{\partial \text{Lluvia}_i}$ representa el **incremento marginal** de la inundación en relación a la lluvia.

Lo triste es que **no podemos conocer este parámetro**.

¡Necesitaríamos datos de todos los días i habidos y por haber!

Estimación de Mínimos Cuadrados Ordinarios (MCO/OLS)

Pero no todo está perdido, tenemos esta maravilla:

	Inundacion	Lluvia
1	2,97	4,69
2	1,53	3,85
3	2,75	0,12
4	1,18	2,25
5	3,42	1,34
6	2,69	4,93
7	6,48	0,49
8	6,50	4,87
9	2,81	1,97

A pesar de que no conocemos β_{Lluvia}, podemos estimarlo.

Para lo anterior utilizaremos la famosa estimación de Mínimos Cuadrados Ordinarios (MCO u OLS, por sus siglas en inglés).

Supuestos de Mínimos Cuadrados Ordinarios (MCO)

1. **Linealidad de parámetros:** $Y = X\beta + U$.
   Una vez estimado el modelo, podemos realizar tests de hipótesis y obtener métricas de bondad de ajuste.

2. **Matriz $X^T X$ invertible:** $rg(X^T X) = K$. Se puede verificar.

3. **Errores nulos en esperanza:** $\mathbb{E}(U) = 0$. Supuesto aceptable.

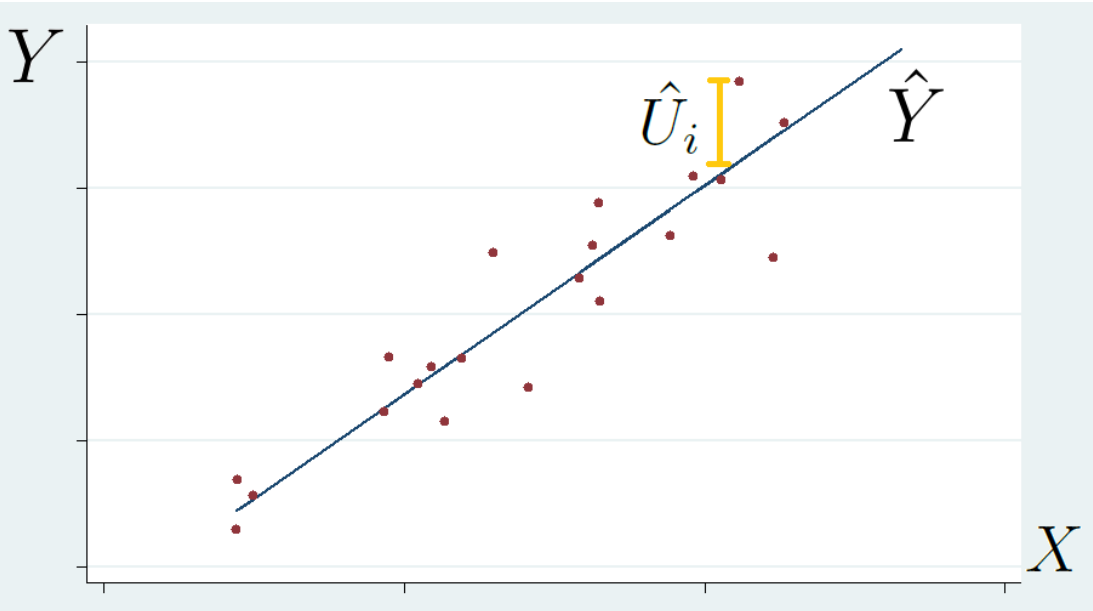
4. **Independencia y homocedasticidad:** $\mathbb{V}(U) = \sigma^2 I_N$.
   Se puede testear.

5. **Exogeneidad:** $\mathbb{E}(X^T U) = 0$. Supuesto fuerte y problemático.

A lo largo del curso veremos qué pasa si se violan algunos de estos supuestos y qué podemos hacer al respecto.

Matriz aniquiladora y Teorema de Frisch-Waugh

1. Modelo real: $Y = X\beta + U$.
2. Modelo estimado: $Y = X\hat{\beta} + \hat{U}$
3. Predicción: $\hat{Y} = X\hat{\beta}$



Si combinamos las ecuaciones 2 y 3, obtenemos lo siguiente:

$$Y = \hat{Y} + \hat{U}$$

Matriz aniquiladora y Teorema de Frisch-Waugh

$$Y = \hat{Y} + \hat{U}$$

Es decir, Y es la suma de dos componentes:

- La que se puede explicar con las variables X que observamos en nuestra base de datos ($\hat{Y}$).
- La que se puede explicar con aquellas variables que no observamos ($\hat{U}$).

Matriz aniquiladora y Teorema de Frisch-Waugh

- **Matriz de proyección:** ($P_X = X(X^TX)^{-1}X^T$) al pre-multiplicar un vector, lo estamos "proyectando" en el mundo de X.
  Lógicamente, la proyección de X en el mundo de las X es X...

$$P_X X = X(X^TX)^{-1}X^TX = X$$

  Como dijimos anteriormente, $\hat{Y}$ es la parte de Y que "vive" en el mundo de las X...

$$P_X Y = X(X^TX)^{-1}X^TY = \hat{Y}$$

- **Matriz aniquiladora:** ($M_X = I_{K\times K} - P_X$) al pre-multiplicar un vector, lo estamos "proyectando" dicho vector en el mundo de las variables no observadas.

Matriz aniquiladora y Teorema de Frisch-Waugh

Como bien dice su nombre, esta matriz "aniquila" a X...

$$M_X X = (I_{K\times K} - P_X)X = X - P_X X = X - X = 0_{N\times K}$$

Como dijimos anteriormente, $\hat{U}$ es la parte de Y que "vive" fuera del mundo de las X...

$$M_X Y = (I_{K\times K} - P_X)Y = Y - P_X Y = Y - \hat{Y} = \hat{U}$$

Donde $\hat{U}$ es el residuo de regresionar Y en función de X, por lo que lo notaremos como $\hat{U}_{Y,X}$. Recuerden esto último.

Matriz aniquiladora y Teorema de Frisch-Waugh

Consideremos el siguiente modelo y premultipliquémoslo por $M_{X_A} \in \mathbb{R}^{N \times N}$

$$\underbrace{Y}_{N\times 1} = \underbrace{\vec{\beta_0}}_{N\times 1} + \underbrace{X_A}_{N\times K_A}\underbrace{\beta_A}_{K_A\times 1} + \underbrace{X_B}_{N\times 1}\underbrace{\beta_B}_{1\times 1} + \underbrace{U}_{N\times 1}$$

$$\Rightarrow \underbrace{M_{X_A} Y}_{\hat{U}_{Y,X_A}} = \underbrace{M_{X_A}\vec{\beta_0}}_{\beta_0'} + \underbrace{M_{X_A}X_A}_{0_{N\times K}}\beta_A + \underbrace{M_{X_A}X_B}_{\hat{U}_{X_B,X_A}}\beta_B + \underbrace{M_{X_A}U}_{U'}$$

$$\Rightarrow \hat{U}_{Y,X_A} = \beta_0' + \hat{U}_{X_B,X_A}\beta_B + U'$$

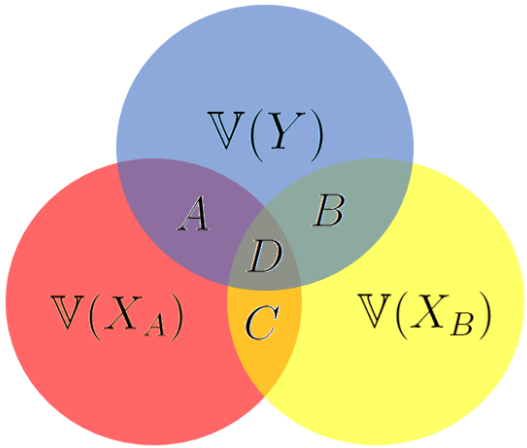
En pocas palabras, este teorema nos dice que la estimación por mínimos cuadrados ordinarios de β_B entrega el mismo valor para el modelo con y sin pre-multiplicar por M_{X_A}.

Matriz aniquiladora y Teorema de Frisch-Waugh

Lo anterior nos permite obtener $\hat{\beta}_B$ mediante el siguiente procedimiento.

1. Regresionar Y en función de X_A y guardar los residuos U_{Y,X_A}.
2. Regresionar X_B en función de X_A y guardar los residuos U_{X_B,X_A}.
3. Regresionar $\hat{U}_{Y,X_A}$ en función de $\hat{U}_{X_B,X_A}$. El coeficiente a estimar, asociado a $\hat{U}_{X_B,X_A}$ es nada más ni nada menos que $\hat{\beta}_B$.

Matriz aniquiladora y Teorema de Frisch-Waugh



Al regresionar Y en función de X_A y X_B, el efecto de la región A será capturado por $\hat{\beta}_A$ y el de la región B será capturado por $\hat{\beta}_B$. El área D es una región conflictiva (problema de identificación).

Si regresionamos Y en función de X_B se atribuirá a β_B el efecto de la región D, lo cual es engañoso porque parte de dicho efecto puede deberse a X_A. Ahora habría un problema de variable omitida.

Al pre-multiplicar por M_{X_A} podemos "limpiar" el efecto de X_A y no tenemos que preocuparnos de la región D, por lo que se puede estimar β_B sin problema.

Teoría asintótica

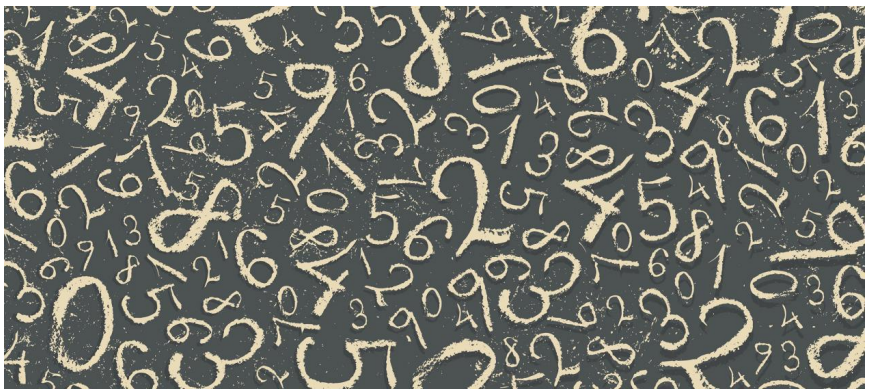
- Supongamos que tenemos un estimador $\hat{\theta}$ para un parámetro poblacional θ.

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.

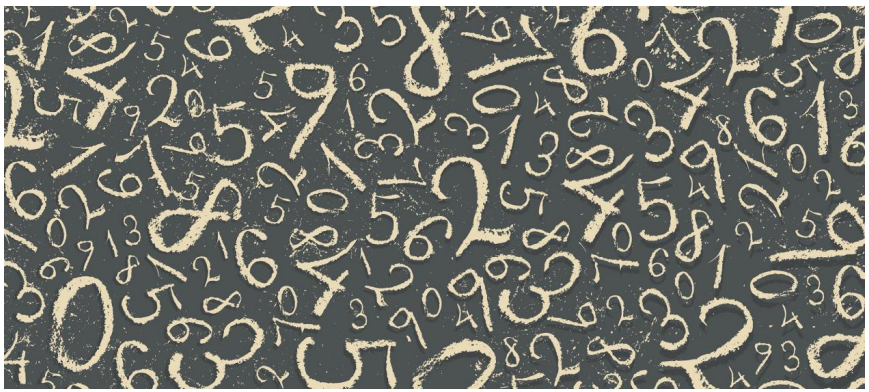
Teoría asintótica

- 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...".



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.

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:

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

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.

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 $\mathbb{V}(X_i) = \sigma^2 < \infty$

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

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:

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

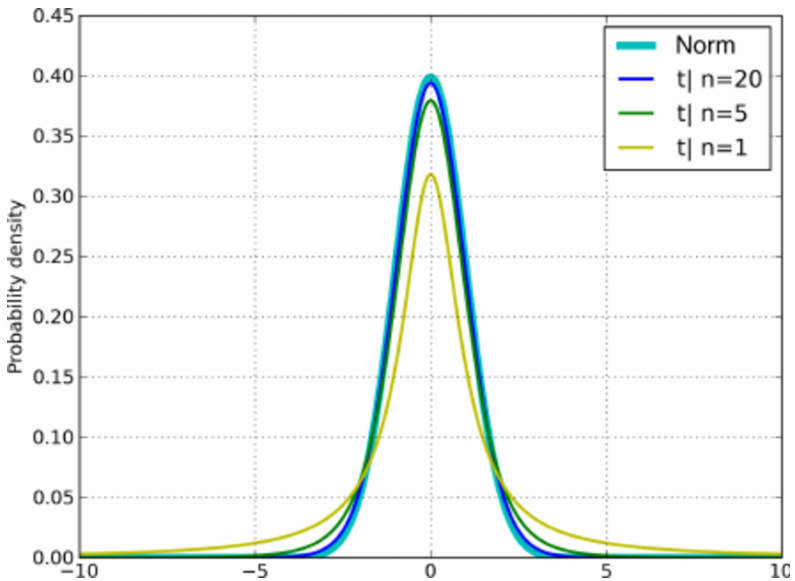


Figura: Una distribución T-Student con $N - K$ grados de libertad converge en distribución a una normal.

Teorema Central del Límite

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

Teorema Central del Límite

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

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

Teorema Central del Límite

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

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

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

Teorema Central del Límite

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

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

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

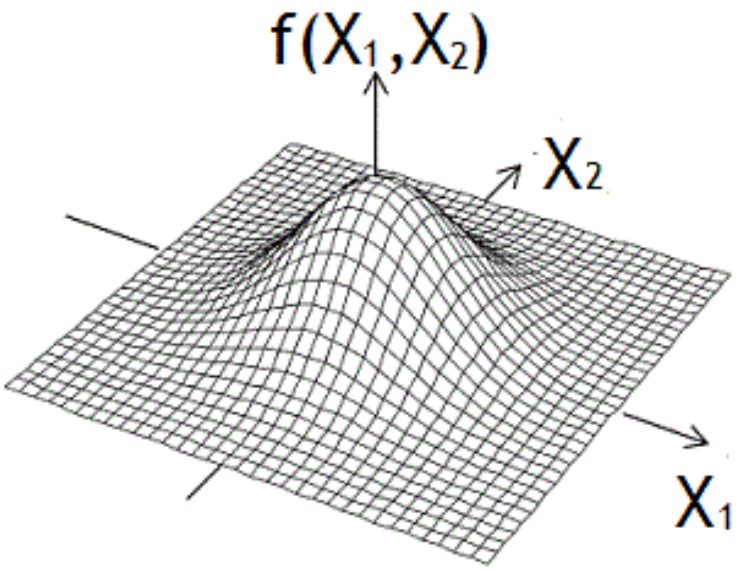
o equivalentemente...

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

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.

Teorema de Slutsky

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

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

- $X_n Y_n \xrightarrow{d} xY$.

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.

Inclusión de variables irrelevantes

El efecto de agregar una variable no es claro y dependería de su capacidad explicativa y de qué tan colineal es esta con otros regresores del modelo.

Inclusión de variables irrelevantes

El efecto de agregar una variable no es claro y dependería de su capacidad explicativa y de qué tan colineal es esta con otros regresores del modelo.

Si la variable es relevante y se omite, se tendrá sesgo e inconsistencia, aunque se los estimadores serán más eficientes ($\downarrow \widehat{\mathbb{V}}(\hat{\beta}_k)$).

Inclusión de variables irrelevantes

El efecto de agregar una variable no es claro y dependería de su capacidad explicativa y de qué tan colineal es esta con otros regresores del modelo.

Si la variable es relevante y se omite, se tendrá sesgo e inconsistencia, aunque se los estimadores serán más eficientes ($\downarrow \widehat{\mathbb{V}}(\hat{\beta}_k)$).

Si la variable es irrelevante y se incluye, $\hat{\beta}$ será insesgado, aunque se perderá eficiencia ($\uparrow \widehat{\mathbb{V}}(\hat{\beta}_k)$).

Inclusión de variables irrelevantes

El efecto de agregar una variable no es claro y dependería de su capacidad explicativa y de qué tan colineal es esta con otros regresores del modelo.

Si la variable es relevante y se omite, se tendrá sesgo e inconsistencia, aunque se los estimadores serán más eficientes ($\downarrow \widehat{\mathbb{V}}(\hat{\beta}_k)$).

Si la variable es irrelevante y se incluye, $\hat{\beta}$ será insesgado, aunque se perderá eficiencia ($\uparrow \widehat{\mathbb{V}}(\hat{\beta}_k)$).

Lo anterior se puede verificar a través de la siguiente fórmula:

$$\mathbb{V}(\hat{\beta}_k) = \frac{\mathbb{V}(Y)(1 - R^2)}{(N - K)\mathbb{V}(X_k)(1 - R^2_{k,-k})}$$

Inclusión de variables irrelevantes

En el modelo clásico de mínimos cuadrados ordinarios, incluir variables irrelevantes exógenas jamás causará sesgo en las estimaciones de β.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T ([X_1 X_2] \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + U)$$

Inclusión de variables irrelevantes

En el modelo clásico de mínimos cuadrados ordinarios, incluir variables irrelevantes exógenas jamás causará sesgo en las estimaciones de β.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T ([X_1 X_2] \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + U)$$

$$= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + (X^T X)^{-1} X^T U$$

Inclusión de variables irrelevantes

En el modelo clásico de mínimos cuadrados ordinarios, incluir variables irrelevantes exógenas jamás causará sesgo en las estimaciones de β.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T ([X_1 X_2] \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + U)$$

$$= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + (X^T X)^{-1} X^T U$$

$$\Rightarrow \mathbb{E}(\hat{\beta}|X) = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}$$

Inclusión de variables irrelevantes

En el modelo clásico de mínimos cuadrados ordinarios, incluir variables irrelevantes exógenas jamás causará sesgo en las estimaciones de β.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T ([X_1 X_2] \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + U)$$

$$= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + (X^T X)^{-1} X^T U$$

$$\Rightarrow \mathbb{E}(\hat{\beta}|X) = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}$$

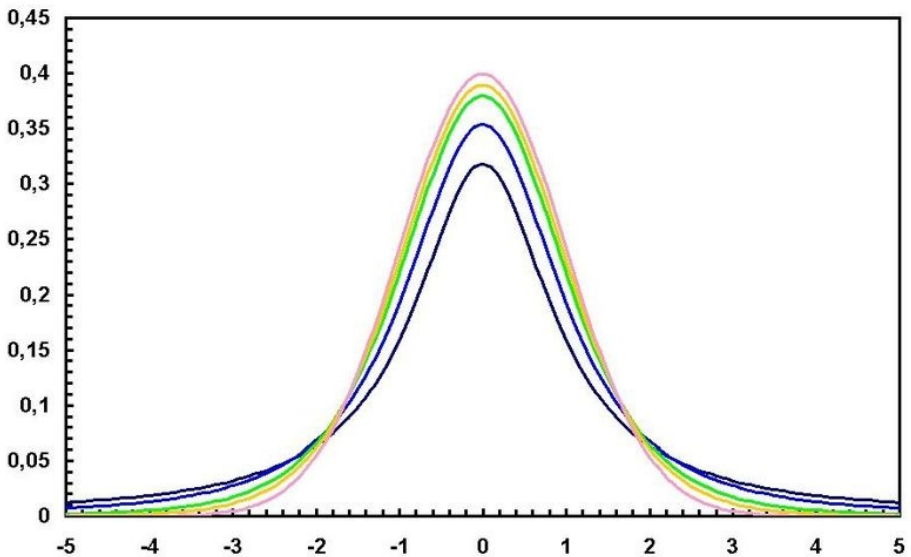
$$\Rightarrow \mathbb{E}(\hat{\beta}_1) = \beta_1$$

Significancia individual

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

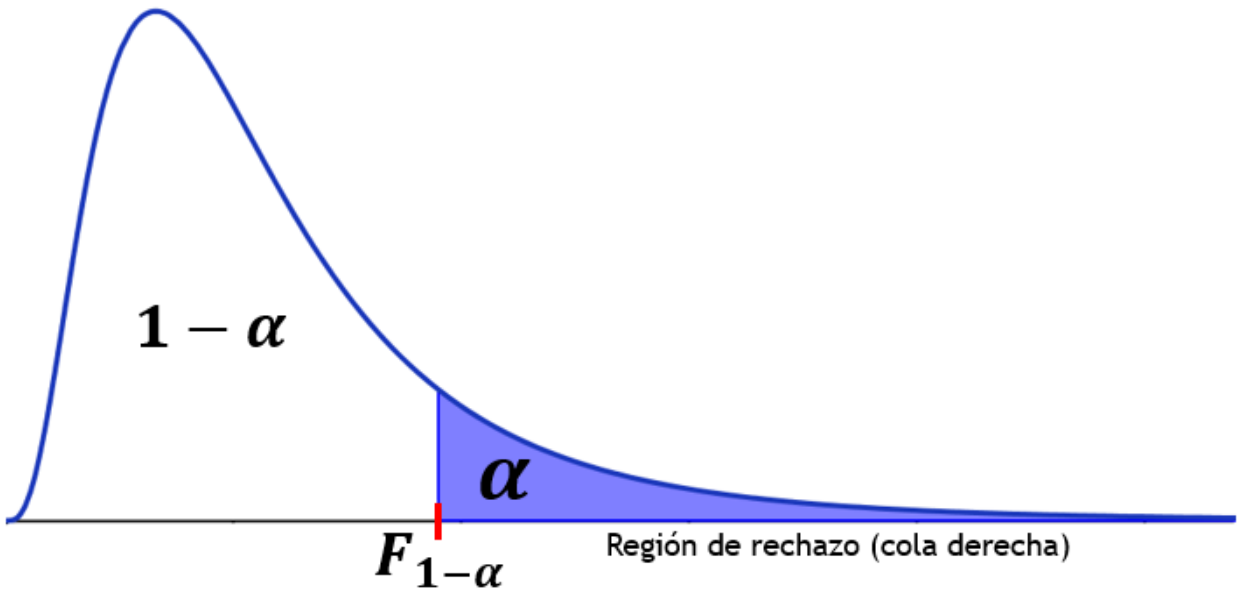
Bajo la hipótesis nula $H_0 : \beta_k = 0$ se tiene $\frac{\hat{\beta}_k}{\hat{\sigma}_U} \xrightarrow{d} \mathcal{T}_{N-K}$

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.



Significancia global

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



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

Test de restricción lineal

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

$$H_0 : \gamma_1\beta_1 + \gamma_2\beta_2 = \delta$$

Test de restricción lineal

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

$$H_0 : \gamma_1\beta_1 + \gamma_2\beta_2 = \delta$$

$$t = \frac{\gamma_1\hat{\beta}_1 + \gamma_2\hat{\beta}_2 - \delta}{\widehat{DS}(\gamma_1\hat{\beta}_1 + \gamma_2\hat{\beta}_2)}$$

Test de restricción lineal

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

$$H_0 : \gamma_1\beta_1 + \gamma_2\beta_2 = \delta$$

$$t = \frac{\gamma_1\hat{\beta}_1 + \gamma_2\hat{\beta}_2 - \delta}{\widehat{DS}(\gamma_1\hat{\beta}_1 + \gamma_2\hat{\beta}_2)}$$

Test de restricciones lineales múltiples

H_0 : alguna(s) variable(s) del modelo libre son cero.

- Modelo restringido: menos variables.
- Modelo libre: más variables.
- r es la cantidad de restricciones que diferencian al modelo libre del restringido.
- R_L^2 y R_R^2 son los coeficientes de ajuste de los modelos libre y restringido.

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