econometría

DARIEL AMADOR

Índice general

Índice de figuras

Índice de cuadros

Introducción a la probabilidad

Fundamentos de teoría estadística

Regresión simple y regresión múltiple

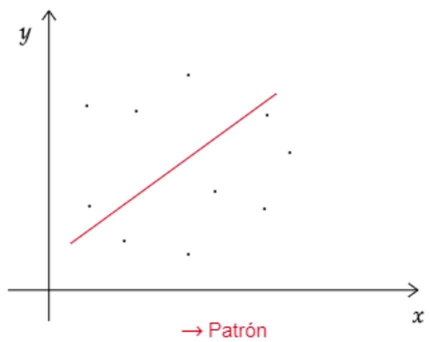
1. Qué es la econometría?

La econometría es un conjunto de herramientas <u>estadísticas</u> que permiten encontrar patrones en datos. **Es un medio y no un fin**.

- **NO PODEMOS VER** $\rightarrow$ Proceso generador de datos
- Datos

X	Y

- **SÍ PODEMOS VER**



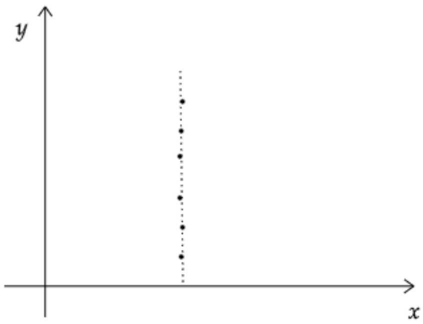
El método para encontrar ese patrón es $\rightarrow$ el <u>método de minimos cuadrados ordinarios</u>. **Los datos vinculan lo que no se puede ver con lo que sí se puede ver.**

1.1 Mínimos cuadrados ordinarios

■ **Ejemplo 1.1 — Horas de estudio y las notas de un curso. X: horas de estudio [0 , 15]**
Y: nota del curso
LO QUE NO PODEMOS VER
<u>Datos de proceso generador de datos</u>

Número de observación	Estudiante	X
1	Isela	8
2	Max	3
3	Valeria	12
4	Dariel	4
5	Kaleb	7

***Es importante saber qué es y qué no es aleatorio. En este caso X no sería aleatorio, es una característica ya dada por persona. No se podrían estimar relaciones estre X y Y si fuera un X igual para todos y hubiera notas diferentes**. Se vería algo así: ■



Entonces:

1. X no es aleatoria, "*es fija en muestreos aleatorios*".
2. La parte determinística entre X y Y:

$$Y_i = \beta_0 + \beta_1 X_i \tag{1.1}$$

Si Y_i fuera la nota, ¿qué serían los β's? $\rightarrow$ @ que la nota más baja posible, aún sin estudiar, fuera 20, por ende $\beta_0 = 20$. Y suponga la siguiente relación: el máximo disponible que se puede estudiar son 15 horas, y a esa cantidad de horas máximas posible, la nota más alta posible sería 90. Entonces se tiene la relación: $X = 15 \rightarrow Y = 90$. Esta relación no sabemos el por qué, es solo para poder obtener un β_1. Entonces:

$$90 = 20 + 15\beta_1$$

$$\Rightarrow 70 = 15\beta_1$$

$$\Rightarrow \frac{70}{15} = \beta_1$$

$$\Rightarrow 4{,}6 = \beta_1$$

$$Y_i \; = \; \underbrace{20 \; + \; 4{,}6X_i}_{No\ aleatorio}$$

3. Parte aleatoria: es un error $\rightarrow U_i$
   El modelo completo:

$$\boxed{Y_i = \underbrace{20 + 4,6X_i}_{\text{No aleatorio}} + \underbrace{U_i}_{\text{Aleatorio}}} \longrightarrow \textit{Modelo poblacional}$$

Es importante tener presente que se está asumiendo una relacional lineal con el error.

4. Modelo poblacional

$$Y_i = 20 + 4,6X_i + U_i \tag{1.2}$$

5. ¿De dónde "sale"U_i?

Dado	1	X2	3	4	5	6
U_i	-9	-6	-3	3	6	9
Probabilidad	1/6	1/6	1/6	1/6	1/6	1/6

Esto implica:

- $$E(U_i) = \frac{1}{6}\cdot -9 + \frac{1}{6}\cdot -6 + \frac{1}{6}\cdot -3 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 6 + \frac{1}{6}\cdot 9$$
  $$\Rightarrow E(U_i) = 0$$

  Datos del proceso generador de datos

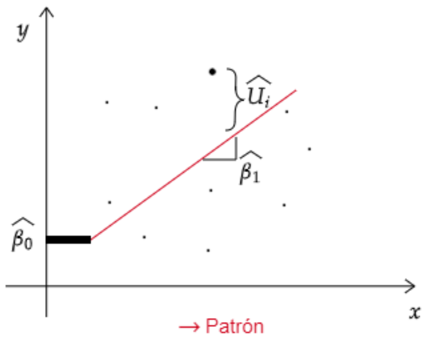
		X_i	U_i	Y_i
1	Isela	8	9	65.8
2	Max	3	3	36.8
3	Valeria	12	-9	66.2
4	Dariel	4	-9	29.4
5	Kaleb	7	-6	46.2

- $Var(U_i) = \sigma^2$
  - Cuando los valores de la varianza son igual para todos, hay: homocedasticidad.
  - Cuando los valores de la varianza no son igual para todos, hay: heterocedasticidad.
- No hay relación $\rightarrow$ porque fue con dado, pero podría darse de tal manera que U_i y U_j sí se relacionen. U_i es independiente de U_j.

LO QUE SÍ PODEMOS VER
Datos

	X_i	Y_i
1	8	65.8
2	3	36.8
3	12	66.2
4	4	29.4
5	7	46.2

Y el objetivo con estos datos, es encontrar la línea entre los puntos: El método de mínimos



cuadrados ordinarios es tomar los puntos y estimar la línea.

- Modelo muestral

$$Y_i \;=\; \underbrace{\widehat{\beta_0} + \widehat{\beta_1} X_i}_{Lnea} \;+\; \widehat{U_i}$$

$$\Rightarrow Y_i - \hat{\beta}_o - \hat{\beta}_1 X_i = \hat{U}_i \qquad\qquad i = 1, \ldots, N$$

<u>**Mínimos cuadrados ordinarios**</u>

$$\min_{\hat{\beta}_0, \hat{\beta}_0} \quad \sum_{i=1}^{N} \hat{U}_i^2 \tag{1.3}$$

Es decir, encontrar una fórmula para resolver el problema de minimización, que se puede reescribir de la siguiente forma:

$$\min_{\hat{\beta}_0, \hat{\beta}_0} \quad (\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i))^2 \tag{1.4}$$

Y se obtienen las siguientes condiciones de primer orden:

$$\frac{\partial \sum_{i=1}^{N} \hat{U}_i^2}{\partial \beta_0} : -2 \sum_{i=1}^{N} (Y_i - \hat{\beta}_o - \hat{\beta}_1 X_i)) = 0 \tag{1.5}$$

$$\sum_{i=1}^{N}(Y_i - \hat{\beta}_o - \hat{\beta}_1 X_i) = 0 \tag{1.6}$$

La ecuación (6) se conoce como *Ecuación normal 1*.

$$\sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} X_i = 0 \tag{1.7}$$

$$\sum_{i=1}^{N} Y_i - \hat{\beta}_1 \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} \hat{\beta}_0 \tag{1.8}$$

$$\sum_{i=1}^{N} Y_i - \hat{\beta}_1 \sum_{i=1}^{N} X_i = N\hat{\beta}_0 \tag{1.9}$$

$$\frac{\sum_{i=1}^{N} Y_i}{N} - \frac{\hat{\beta}_1 \sum_{i=1}^{N} X_i}{N} = \hat{\beta}_0 \tag{1.10}$$

Ahora, se hacen las siguientes definiciones:

$$\frac{\sum_{i=1}^{N} Y_i}{N} = \overline{Y} \qquad \frac{\sum_{i=1}^{N} X_i}{N} = \overline{X}$$

Y se puede reescribir:

$$\boxed{\overline{Y} - \hat{\beta}_1 \overline{X} = \hat{\beta}_0} \tag{1.11}$$

Ahora la segunda condición de primer orden:

$$\frac{\partial \sum_{i=1}^{N} \hat{U_i^2}}{\partial \hat{\beta}_1} : -2X_i \sum_{i=1}^{N}(Y_i - \hat{\beta}_o - \hat{\beta}_1 X_i) = 0 \tag{1.12}$$

$$X_i \sum_{i=1}^{N}(Y_i - \hat{\beta}_o - \hat{\beta}_1 X_i) = 0 \tag{1.13}$$

$$\sum_{i=1}^{N} Y_i X_i - \sum_{i=1}^{N} \hat{\beta}_o X_i - \sum_{i=1}^{N} \hat{\beta}_1 X_i^2 = 0 \tag{1.14}$$

$$\sum_{i=1}^{N} Y_i X_i - \hat{\beta}_o \sum_{i=1}^{N} X_i = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 \tag{1.15}$$

Evaluando el valor encontrado anteriormente de $\hat{\beta}_0$, se obtiene:

$$\sum_{i=1}^{N} Y_i X_i - (\overline{Y} - \hat{\beta}_1 \overline{X}) \sum_{i=1}^{N} X_i = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 \quad (1.16)$$

$$\sum_{i=1}^{N} Y_i X_i - \overline{Y} \sum_{i=1}^{N} X_i + \hat{\beta}_1 \overline{X} \sum_{i=1}^{N} X_i = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 \quad (1.17)$$

$$\sum_{i=1}^{N} Y_i X_i - \overline{Y} \sum_{i=1}^{N} X_i = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 - \hat{\beta}_1 \overline{X} \sum_{i=1}^{N} X_i \quad (1.18)$$

$$\sum_{i=1}^{N} Y_i X_i - \overline{Y} \sum_{i=1}^{N} X_i = \hat{\beta}_1 (\sum_{i=1}^{N} X_i^2 - \overline{X} \sum_{i=1}^{N} X_i) \quad (1.19)$$

$$\boxed{\frac{\sum_{i=1}^{N} Y_i X_i - \overline{Y} \sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} X_i^2 - \overline{X} \sum_{i=1}^{N} X_i} = \hat{\beta}_1 \quad (1.20)}$$

Los estimadores son funciones que se evalúan en los datos.

$\rightarrow$ Calculando

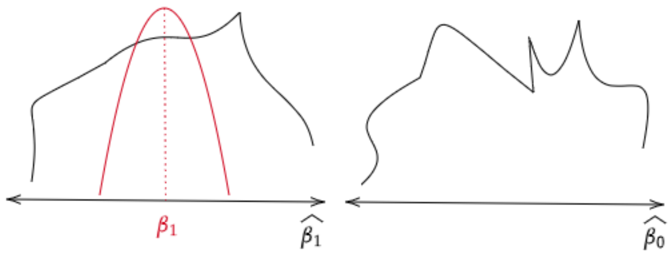
- $\overline{Y} = 48{,}88$
- $\overline{X} = 6{,}8$
- $\sum_{i=1}^{N} X_i = 34$
- $\sum_{i=1}^{N} Y_i X_i = 1872{,}2$
- $\sum_{i=1}^{N} X_i^2 = 282$

$\hat{\beta}_1 = \frac{1872{,}2 - 48{,}88 \cdot (34)}{282 - 6{,}8(34)}$

$\hat{\beta}_1 = 4{,}12 \rightarrow$ y el poblacional era 4.6

$\hat{\beta}_0 = 48{,}88 - 4{,}13 \cdot (6{,}8)$

$\hat{\beta}_0 = 20{,}79 \rightarrow$ y el poblacional era 20



Propiedades estadísticas:

1. Valor estimado
2. La varianza más pequena

*Las pruebas de hipótesis son fundamentales para hacer inferencias.

Haciendo otra muestra propia:

Número de observación	Nombre	X	U_i	Y_i
1	Isela	8	3	59.8
2	Max	3	-6	27.8
3	Valeria	12	-9	66.2
4	Dariel	4	-3	35.4
5	Kaleb	7	-9	43.2

$$Y_i = 20 + 4{,}6X_i + U_i$$

Datos

Número de observación	Nombre	X_i	Y_i
1	Isela	8	59.8
2	Max	3	27.8
3	Valeria	12	66.2
4	Dariel	4	35.4
5	Kaleb	7	43.2

$\rightarrow$ Calculando

- $\overline{Y} = 46{,}48$
- $\overline{X} = 6{,}8$
- $\sum_{i=1}^{N} X_i = 34$
- $\sum_{i=1}^{N} Y_i X_i = 1800{,}2$
- $\sum_{i=1}^{N} X_i^2 = 282$

$\hat{\beta}_1 = \frac{1800{,}2 - 46{,}48 \cdot (34)}{282 - 6{,}8(34)}$

$\hat{\beta}_1 = 4{,}3$

$\hat{\beta}_0 = 46{,}48 - 4{,}3 \cdot (6{,}8)$

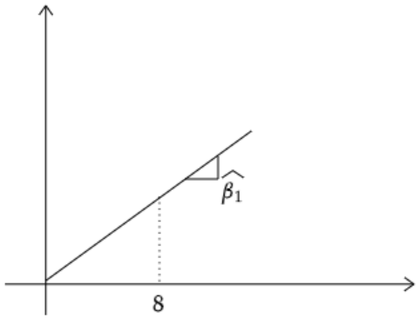
$\hat{\beta}_0 = 17{,}24$

Para efectos de las evaluaciones: **las demostraciones deben ir acompañadas de texto para facilitar la comunicación, no solo lenguaje matemático a pesar de que la matemática sea un lenguaje universal.** $\rightarrow$ **El orden es demasiado importante.** Todas las preguntas son válidas.

<u>Modelo Muestral</u>

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{U}_i \tag{1.21}$$

$$Y_i \;=\; \underbrace{\widehat{\beta}_0 + \widehat{\beta}_1 X_i \;+\; \widehat{U}_i}_{\hat{Y}}$$



$$\hat{Y} \;=\; \widehat{\beta}_0 + \widehat{\beta}_1 X_i$$

Y estimado

Entonces:

$$\hat{U}_i = Y_i - \hat{Y}_i \tag{1.22}$$

Recordando los estimadores MCO:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} Y_i X_i - \overline{Y} \sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} X_i^2 - \overline{X} \sum_{i=1}^{N} X_i = \hat{\beta}_1} \tag{1.23}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{1.24}$$

- Algunas propiedades matemáticas:

$$\text{EN1:} \qquad \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \beta_1 X_i) = 0 \tag{1.25}$$

Y entonces la propiedad matemática #1 dice que:

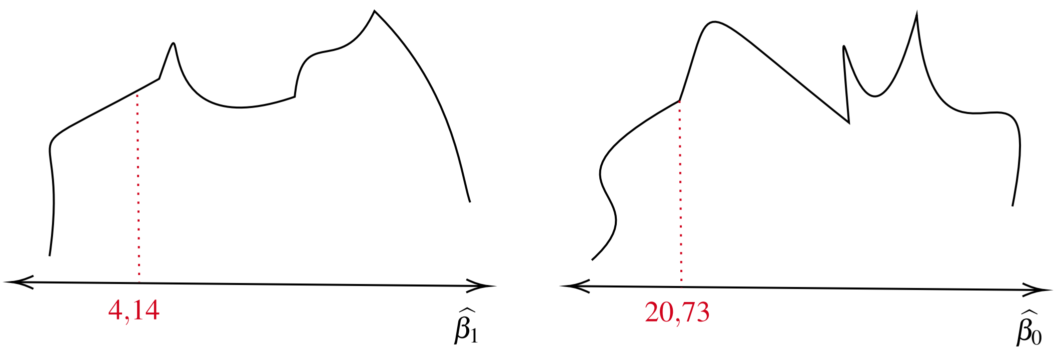
$$\sum_{i=1}^{N} \hat{U}_i = 0 \tag{1.26}$$

$$\text{EN2:} \qquad \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \beta_1 X_i) X_i = 0 \tag{1.27}$$

Mientras la propiedad matemática #2 dice que:

$$\sum_{i=1}^{N} \hat{U}_i X_i = 0 \tag{1.28}$$

Cambiando de tema, ahora se van a ver las distribuciones de probabilidad de $\hat{\beta}_0$ y $\hat{\beta}_1$:
$\hat{\beta}_0$ y $\hat{\beta}_1$ son funciones que se evalúan X_i y Y_i. Son aleatorias porque Y_i es aleatorio.



¿Qué podríamos cambiar?

- Aumentar N (se esperaría un promedio más cercano a 4.6)
- Cambiar distribución de U_i (por ejemplo de -6 a 6)
- Incorporar 12 posibles de los errores
- Establecer una distrución continúa del dado

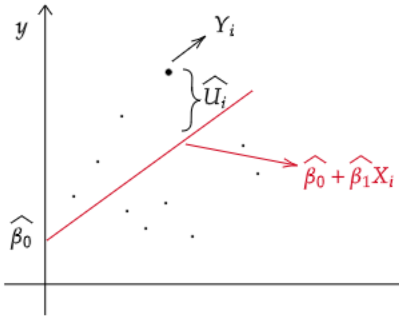
La clase pasada vimos una serie de definiciones:

- Modelo muestral

$$Y_i = \hat{\beta}_o + \hat{\beta}_1 X_i + \hat{U}_i \rightarrow \text{le llamamos modelo muestral (MM), el cual estimamos con MCO.} \tag{1.29}$$

- $\hat{Y}_i$ (Y_i estimado)

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{1.30}$$



- $\hat{U}_i$ (obtenemos del MM):

$$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \tag{1.31}$$

$$\hat{U}_i = Y_i - \hat{Y}_i \tag{1.32}$$

2. Propiedades matemáticas

Estas propiedades matematicas salen de las condiciones de primer orden y por ende no dependen del proceso generador de datos.

- Primera condición de primer orden: $\dfrac{\partial \sum_{i=1}^{N} \hat{U_i}^2}{\partial \hat{\beta}_0}$

$$\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{2.1}$$

$$\boxed{\sum_{i=1}^{N} \hat{U}_i = 0 \rightarrow \text{Propiedad matemática \#1}} \tag{2.2}$$

- Segunda condición de primer orden: $\dfrac{\partial \sum_{i=1}^{N} \hat{U_i}^2}{\partial \hat{\beta}_1}$

$$\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0 \tag{2.3}$$

$$\boxed{\sum_{i=1}^{N} \hat{U}_i X_i = 0 \rightarrow \text{Propiedad matemática \#2}} \tag{2.4}$$

***Nota: Cap.3 Gujarati y Porter**

- $X_i = X_i - \overline{X}$
- $Y_i = Y_i - \overline{Y}$

- Utilizando el modelo muestral (MM):

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{U}_i$$

Y se calcula el promedio del MM:

$$\frac{\sum_{i=1}^{N} Y_i}{N} = \frac{\sum_{i=1}^{N} \hat{\beta}_0}{N} + \frac{\sum_{i=1}^{N} \hat{\beta}_1 X_i}{N} + \frac{\sum_{i=1}^{N} \hat{U}_i}{N} \tag{2.5}$$

Usando la propiedad matemática #1: $\sum_{i=1}^{N} \hat{U}_i = 0$

$$\frac{\sum_{i=1}^{N} Y_i}{N} = \frac{N\hat{\beta}_0}{N} + \hat{\beta}_1 \frac{\sum_{i=1}^{N} X_i}{N} + \cancel{\frac{\sum_{i=1}^{N} \hat{U}_i}{N}} \tag{2.6}$$

$$\boxed{\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} \rightarrow 3.a} \tag{2.7}$$

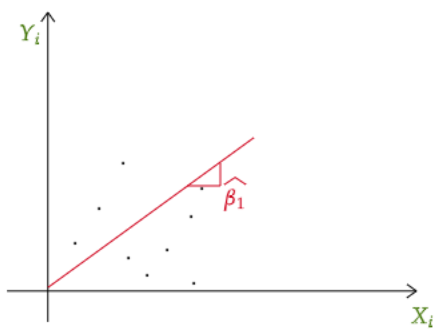
Usando el MM y 3.a:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{U}_i \rightarrow MM \tag{2.8}$$

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} \rightarrow 3.a \tag{2.9}$$

$$Y_i - \overline{Y} = \hat{\beta}_1 (X_i - \overline{X}) + \hat{U}_i \tag{2.10}$$

$$\boxed{Y_i = \hat{\beta}_1 X_i + \hat{U}_i \rightarrow \text{Propiedad matemática \#3}} \tag{2.11}$$



$$\boxed{\sum_{i=1}^{N} Y_i^2 = \hat{\beta}_1{}^2 \sum_{i=1}^{N} X_i{}^2 + \sum_{i=1}^{N} \hat{U}_i \rightarrow \text{Propiedad matemática \#4}} \tag{2.12}$$

Demuestre la propiedad matemática #4 a partir de la propiedad matemática #3:

Demostración. Tómese la propiedad matemática #3 y elévese ambos lados al cuadrado:

$$Y_i^2 = (\hat{\beta}_1 X_i + \hat{U}_i)^2 \tag{2.13}$$

Ahora aplíquese sumatoria a ambos lados de la igualdad

$$\Leftrightarrow \sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N} (\hat{\beta}_1 X_i + \hat{U}_i)^2 \tag{2.14}$$

Ahora desarróllese el producto notable del lado derecho

$$\Leftrightarrow \sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N} \hat{\beta}_1 X_i^2 + \sum_{i=1}^{N} 2\hat{\beta}_1 X_i \hat{U}_i + \sum_{i=1}^{N} \hat{U}_i^2 \tag{2.15}$$

$$\Leftrightarrow \sum_{i=1}^{N} Y_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + 2\hat{\beta}_1 \sum_{i=1}^{N} X_i \hat{U}_i + \sum_{i=1}^{N} \hat{U}_i^2 \tag{2.16}$$

Ahora se sustituye X_i en el término en el medio del lado derecho del igual:

$$\Leftrightarrow \sum_{i=1}^{N} Y_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + 2\hat{\beta}_1 \sum_{i=1}^{N} (X_i - \overline{X}) \hat{U}_i + \sum_{i=1}^{N} \hat{U}_i^2 \tag{2.17}$$

$$\Leftrightarrow \sum_{i=1}^{N} Y_i^2 = \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + 2\hat{\beta}_1 [\sum_{i=1}^{N} X_i \hat{U}_i - \overline{X} \sum_{i=1}^{N} \hat{U}_i] + \sum_{i=1}^{N} \hat{U}_i^2 \tag{2.18}$$

Ahora, obsérvese lo siguiente:

$$\textstyle\sum_{i=1}^{N} Y_i = \widehat{\beta_1^2} \sum_{i=1}^{N} X_i^2 + 2\widehat{\beta}_1 \left[\underbrace{\sum_{i=1}^{N} X_i \widehat{U}_i}_{\text{Propiedad matemtica \#2}} - \overline{X} \underbrace{\sum_{i=1}^{N} \widehat{U}_i}_{\text{Propiedad matemtica \#1}} \right] + \sum \widehat{U_i^2}$$

Por la propiedad matemática #2 : $\sum_{i=1}^{N} X_i \hat{U}_i = 0$

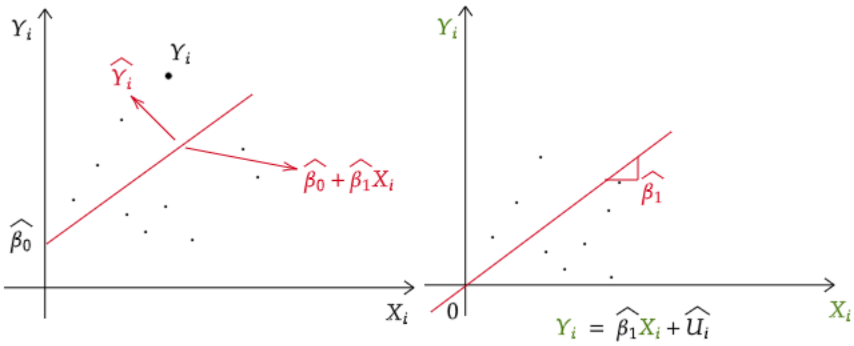
Por la propiedad matemática #1 : $\sum_{i=1}^{N} \hat{U}_i = 0$

Y, por tanto:

$$\boxed{\sum_{i=1}^{N} Y_i^2 = \hat{\beta}_1^{\,2} \sum_{i=1}^{N} X_i^2 + \sum_{i=1}^{N} \hat{U}_i^2} \tag{2.19}$$

■

$$X_i = X_i - \overline{X} \tag{2.20}$$



$$\sum_{i=1}^{N} X_i = \sum_{i=1}^{N}(X_i - \overline{X}) \tag{2.21}$$

$$\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N}\overline{X} \tag{2.22}$$

$$\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} X_i - \frac{\cancel{N}\sum_{i=1}^{N} X_i}{\cancel{N}} \tag{2.23}$$

$$\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} X_i \tag{2.24}$$

$$\sum_{i=1}^{N} X_i = 0 \tag{2.25}$$

$\hat{\beta}_1$: Cómo se relaciona la variabilidad de y con la variabilidad de x? $\rightarrow$ Cuánta nota más me daría una hora adicional de estudio?

Datos	Z	
		Varianza
	-	$\frac{\sum_{i=1}^{N}}{N}$
	-	
	-	
	-	

2.0.1 Indicador de bondad de ajuste

$$\sum_{i=1}^{N} Y_i^2 = \hat{\beta}_1{}^2 \sum_{i=1}^{N} X_i{}^2 + \sum_{i=1}^{N} \hat{U}_i{}^2 \tag{2.26}$$

- $\sum_{i=1}^{N} Y_i^2$: Suma Total de Cuadrados STC → Lo que se se quiere explicar
- $\hat{\beta}_1{}^2 \sum_{i=1}^{N} X_i{}^2$: Suma Estimada de Cuadrados SEC → Lo que el modelo explica
- $\sum_{i=1}^{N} \hat{U}_i{}^2$: Suma Residuos Cuadrados SRC → Lo que el modelo no explica

Indicador de bondad de ajuste:

$$R^2 = \frac{SEC}{STC} \tag{2.27}$$

Interpretación de R^2: $R^2 = \frac{SEC}{STC}$ Utilizando la propiedad matemática #4:

$$STC = SEC + SRC \tag{2.28}$$

Dividiendo a ambos lados:

$$\frac{STC}{STC} = \frac{SEC + SRC}{STC} \tag{2.29}$$

$$\frac{STC}{STC} = \frac{SEC}{STC} + \frac{SRC}{STC} \tag{2.30}$$

$$1 = R^2 + \frac{SRC}{STC} \tag{2.31}$$

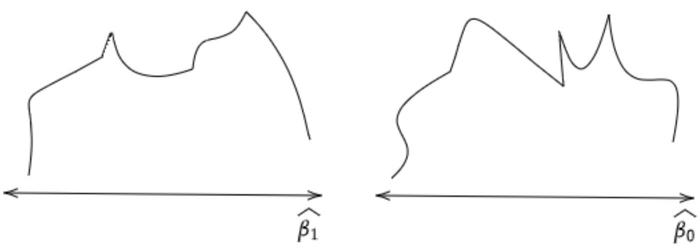
$$R^2 \in \quad]0,1[$$

Indica qué tanto explica el modelo. → Indica el menor SRC por cómo está disenado la función objetivo mín $\sum_{i=1}^{N} \hat{U}_i{}^2$

3. Propiedades estadísticas

Hasta ahora hemos visto propiedades matemáticas. Ahora vamos a ver las orpiedades esadísticas. Recordemos que:

$$\text{Estimadores de MCO} \rightarrow \begin{cases} \hat{\beta}_0 \\ \hat{\beta}_1 \end{cases}$$



Las propiedades estadísticas buscan decir algo sobre las distribuciones de $\hat{\beta}_0$ y $\hat{\beta}_1$.

$$\hat{\beta}_1 \rightarrow \begin{cases} E(\hat{\beta}_1) \rightarrow ideal : E(\hat{\beta}_1) = \hat{\beta}_1 \\ Var(\hat{\beta}_1) \rightarrow ideal : \text{que sea pequena} \end{cases}$$

$$\hat{\beta}_0 \rightarrow \begin{cases} E(\hat{\beta}_0) \rightarrow ideal : E(\hat{\beta}_0) = \hat{\beta}_0 \\ Var(\hat{\beta}_0) \rightarrow ideal : \text{que sea pequena} \end{cases}$$

¿Cómo determinar si una varianza es pequena $\rightarrow$ comparando métodos? Comparar la varianza de $\hat{\beta}_1$ bajo el método de MCO y la varianza de $\hat{\beta}_1$ bajo otro método.

***Leer Cap.3 Gujarati y Porter.** Ellos pasan el libro. El otro viernes hay quiz de lo que se ha visto hasta hoy.

Supuestos

Si el PGD se ve: - - - - (conjunto de supuestos)	¿Qué pasa con $E(\hat{\beta}_1)$ y la $Var(\hat{\beta}_1)$?	
Si el PGD se ve: - - - - (otro conjunto de supuestos)	¿Qué pasa con $E(\hat{\beta}_1)$ y la $Var(\hat{\beta}_1)$?	Otro método

1. $E(u_i) = 0$
2. $Var(u_i) = \sigma^2$
3. $Cov(u_i, u_j) = 0 \quad \forall i \neq j$
4. X no es aleatoria
5. Planteamiento Modelo Poblacional

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (M.P)$$

3.0.1 Prueba de insesgamiento

¿Qué es lo que queremos probar? $\rightarrow E(\hat{\beta}_1) = \beta_1$

$$\hat{\beta}_1 = \frac{\sum X_i Y_I - \overline{Y} \sum X_i}{\sum X_i^2 - \overline{X} \sum X_i}$$

Para simplificar $\hat{\beta}_1$ tomamos primero el numerador:

$$\begin{aligned}
\Rightarrow \sum X_i Y_i - \overline{Y} \sum X_i &= \\
\Rightarrow \sum X_i Y_i - \frac{\sum Y_i}{N} \sum X_i &= \\
\Rightarrow \sum X_i Y_i - \sum Y_i \cdot \frac{\sum X_i}{N} &= \\
\Rightarrow \sum X_i Y_i - \sum Y_i \cdot \overline{X} &= \\
\Rightarrow \sum X_i Y_i - \sum \overline{X} Y_i &= \\
\Rightarrow \sum (X_i Y_i - \overline{X} Y_i) &= \\
\Rightarrow \sum Y_i (X_i - \overline{X}) &= \\
\Rightarrow \sum Y_i X_i &=
\end{aligned} \tag{3.1}$$

Y luego tomamos el denominador:

$$
\begin{aligned}
&\Rightarrow \sum X_i^2 - \overline{X}\sum X_i = \\
&\Rightarrow \sum X_i^2 - \overline{X}\sum X_i - \overline{X}\sum X_i + \overline{X}\sum X_i = \\
&\Rightarrow \sum X_i^2 - 2\overline{X}\sum X_i + \overline{X}\sum X_i = \\
&\Rightarrow \sum X_i^2 - \sum 2\overline{X}X_i + \overline{X}\sum X_i = \\
&\Rightarrow \sum X_i^2 - \sum 2\overline{X}X_i + \frac{N\overline{X}\sum X_i}{N} = \\
&\Rightarrow \sum X_i^2 - \sum 2\overline{X}X_i + \overline{X}N\overline{X} = \\
&\Rightarrow \sum X_i^2 - \sum 2\overline{X}X_i + N\overline{X}^2 = \\
&\Rightarrow \sum X_i^2 - \sum 2\overline{X}X_i + \sum \overline{X}^2 = \\
&\Rightarrow \sum (X_i^2 - 2\overline{X}X_i + \overline{X}^2) = \\
&\Rightarrow \sum (X_i - \overline{X})^2 = \\
&\Rightarrow \sum {X_i}^2
\end{aligned}
\tag{3.2}
$$

Y entonces así:

$$\hat{\beta}_1 = \frac{\sum Y_i X_i}{\sum {X_i}^2} \tag{3.3}$$

$$\hat{\beta}_1 = \frac{}{\sum {X_i}^2} \cdot \sum Y_i X_i \tag{3.4}$$

Definimos:

$$\boxed{k_i = \frac{x_i}{\sum {x_i}^2}} \tag{3.5}$$

$$\boxed{\hat{\beta}_1 = \sum k_i Y_i} \tag{3.6}$$

Solo se emplean datos de $\{x_i\}$ y no se usan datos de $\{Y_i\}$.

4. Propiedades matemáticas de k_i

- $\sum k_i = 0$
- $\sum k_i X_i = 1$

$$\sum k_i = 0 = \frac{\sum x_i}{\sum {x_i}^2} = \frac{1}{\sum {x_i}^2} \cdot \sum x_i$$

sabiendo que $\sum x_i = 0$, entonces:

$$\sum k_i = 0$$

$$\begin{aligned} \sum k_i X_i &= \frac{\sum X_i}{\sum {X_i}^2} X_i \\ \Rightarrow \frac{1}{\sum {X_i}^2} \cdot \sum X_i X_i &= \frac{1}{\sum {X_i}^2} \cdot \sum (X_i - \overline{X}) X_i \\ \Rightarrow \frac{1}{\sum {X_i}^2} \cdot \sum (X_i^2 - \overline{X} X_i) &= \frac{\sum {X_i}^2}{\sum {X_i}^2} = 1 \end{aligned}$$

$$\hat{\beta}_1 = \sum k_i Y_i \tag{4.1}$$

Usando el supuesto 5, que dice que el modelo poblacional es:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{4.2}$$

Se reemplaza en la ecuación anterior:

$$\begin{aligned} \hat{\beta}_1 &= \sum k_i (\beta_0 + \beta_1 X_i + u_i) \\ \Rightarrow \hat{\beta}_1 &= \sum k_i \beta_0 + \sum k_i \beta_1 X_i + \sum k_i u_i) \\ \Rightarrow \hat{\beta}_1 &= \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i u_i) \end{aligned}$$

Usando las propiedades matemáticas de k_i :

$$\begin{aligned} \hat{\beta}_1 &= \beta_0 \cdot 0 + \beta_1 \cdot 1 + \sum k_i u_i \\ \Rightarrow \hat{\beta}_1 &= \beta_1 + \sum k_i u_i \end{aligned} \tag{4.3}$$

Calcular $E(\cdot)$ de $\hat{\beta}_1$:

$$E(\beta_1) = \beta_1 \hat{+} \sum E(k_i u_i)$$

Usando el supuesto de X_i no es aleatoria, lo que implica que k_i no es aleatorio:

$$E(\beta_1) = \beta_1 \hat{+} \sum k_i E(u_i)$$

Usando supuesto $E(u_i) - 0$,

$$E(\hat{\beta}_1) = \beta_1 \rightarrow Insesgamiento$$

5. Notación matricial

Modelo muestral:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_{k-1} X_{k-1} + \hat{u}_i \tag{5.1}$$

Definiciones:

- k = cantidad de coeficientes
- i = observación
- i = 1,..., N $\rightarrow$ N cantidad de observaciones

$$
\begin{array}{ccccc}
Y_{Nx1} & = & X_{NxK} & \hat{\beta}_{Kx1} & \hat{U}_{Nx1} \\
 & & \begin{array}{ccccc} & X_{1i} & X_{2i} & \cdots & Xk-1_i \end{array} & & \\
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} & = &
\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k-11} \\ 1 & X_{12} & X_{22} & \cdots & X_{k-12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & X_{2N} & \cdots & X_{k-1N} \end{bmatrix} &
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{k-1} \end{bmatrix} &
\begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \vdots \\ \hat{U}_N \end{bmatrix}
\end{array}
$$

Modelo muestral:

$$Y_{Nx1} = X_{XxK} \hat{\beta}_{Kx1} + \hat{u}_{iNx1} \tag{5.2}$$

Obtener estimador de MCO:

$$\hat{\beta} = f(X, Y)$$

Función objetivo:

$$\sum \hat{u}_i{}^2 = \hat{U}'\hat{U} \tag{5.3}$$

$$\begin{bmatrix} \hat{U}_1 & \hat{U}_2 & \cdots & \hat{U}_N \end{bmatrix} \cdot \begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \vdots \\ \hat{U}_N \end{bmatrix} = \widehat{U_1}^2 + \widehat{U_2}^2 + \ldots \widehat{U_N}^2$$

Usando el modelo muestral (MM):

$$\hat{U} = Y - X\hat{\beta} \tag{5.4}$$

Reemplazar en la función objetivo:

$$\begin{aligned} \hat{U}'\hat{U} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ \hat{U}'\hat{U} &= (Y' - \hat{\beta}'X')(Y - X\hat{\beta}) \\ \hat{U}'\hat{U} &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned} \tag{5.5}$$

Tarea: probar usando las matrices que:

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

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

5.1 Vectores

$$a_{mx1} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \qquad z_{mx1} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

$$a'_{1xm} \cdot z_{mx1} = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} = a_1z_1 + a_2z_2 + \ldots + a_mz_m = \sum_{i=1}^{m} a_iz_i$$

$$\frac{\partial \sum a_i z_i}{\partial z} = \begin{bmatrix} \frac{\partial \sum_{i=1}^{m} a_i z_i}{\partial z_1} \\ \frac{\partial \sum_{i=1}^{m} a_i z_i}{\partial z_1} \\ \vdots \\ \frac{\partial \sum_{i=1}^{m} a_i z_i}{\partial z_1} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = a$$

$$\frac{\partial a'z}{\partial z} = a \rightarrow \text{Regla \#1} \tag{5.6}$$

$$\frac{\partial z'Az}{\partial Z} = 2Az \rightarrow \text{Regla \#2} \tag{5.7}$$

CPO:

$$\frac{\partial [\cdot]}{\partial \hat{\beta}} : 0 - 2(Y'X)' + 2X'X\hat{\beta} = 0 \tag{5.8}$$

$$\begin{aligned} -(Y'X)' + X'X\hat{\beta} &= 0 \\ -X'Y + X'X\hat{\beta} &= 0 \\ X'X\hat{\beta} &= X'Y \\ (X'X)^{-1}(X'X)\hat{\beta} &= (X'X)^{-1}X'Y \\ I\hat{\beta} &= (X'X)^{-1}X'Y \end{aligned}$$

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y} \tag{5.9}$$

1. $Var[U]$
2. Prueba insesgamiento
3. $Var[\hat{\beta}]$

5.2 Varianza de los errores

Supuestos:

1. $E(U_i) = 0$
2. $Var(U_i) = \sigma^2$
3. $Cov(U_i, U_j) = 0$

Para cualquier variable:

$$Var(U_i) = E[U_i - E(U_i)^2]$$

$$Cov(U_i, U_j) = E[(U_i - E(U_i)^2)(U_j - E(U_j)^2)]$$

$$E(U_{Nx1}) = E\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix} = \begin{bmatrix} E(U_1) \\ E(U_2) \\ \vdots \\ E(U_N) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$E(U_{Nx1}) = 0_{Nx1}$

Definición de matriz de varianza de U:

$$Var[U] = E[(U - E(U))((U - E(U))']$$

$$U - E(U) = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix} - E\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix} = \begin{bmatrix} U_1 - E(U_1) \\ U_2 - E(U2) \\ \vdots \\ U_N - E(U_N) \end{bmatrix}$$

$$\begin{bmatrix} U_1 - E(U_1) \\ U_2 - E(U2) \\ \vdots \\ U_N - E(U_N) \end{bmatrix}_{Nx1} \cdot \begin{bmatrix} U_1 - E(U_1) & U_2 - E(U2) & \cdots & U_N - E(U_N) \end{bmatrix}_{1xN}$$

$$\begin{bmatrix} (U_1 - E(U_1))^2 & ((U_2 - E(U_2)(U_1 - E(U_1)) & \cdots & ((U_N - E(U_N))(U_1 - E(U_1)) \\ ((U_1 - E(U_1)(U_2 - E(U2)) & (U_2 - E(U_2))^2 & \cdots & ((U_N - E(U_N))(U_2 - E(U_2)) \\ \vdots & \vdots & \ddots & \vdots \\ ((U_1 - E(U_1))(U_N - E(U_N)) & ((U_2 - E(U_2))(U_N - E(U_N)) & \cdots & (U_N - E(U_N))^2 \end{bmatrix}$$

$$Var(U) = E\begin{bmatrix} (U_1-E(U_1))^2 & ((U_2-E(U_2)(U_1-E(U_1)) & \cdots & ((U_N-E(U_N))(U_1-E(U_1)) \\ ((U_1-E(U_1)(U_2-E(U2)) & (U_2-E(U_2))^2 & \cdots & ((U_N-E(U_N))(U_2-E(U_2)) \\ \vdots & \vdots & \ddots & \vdots \\ ((U_1-E(U_1))(U_N-E(U_N)) & ((U_2-E(U_2))(U_N-E(U_N)) & \cdots & (U_N-E(U_N))^2 \end{bmatrix}$$

1. Supuesto #1: $E(U_i) = 0$

$$Var(U) = E\begin{bmatrix} {U_1}^2 & U_2U_1 & \cdots & U_NU_1 \\ U_1U_2 & {U_2}^2 & \cdots & U_NU_2 \\ \vdots & \vdots & \ddots & \vdots \\ U_1U_N & U_2U_N & \cdots & U_N^2 \end{bmatrix}$$

$$Var(U) = \begin{bmatrix} E\left({U_1}^2\right) & E(U_2U_1) & \cdots & E(U_NU_1) \\ E(U_1U_2) & E\left({U_2}^2\right) & \cdots & E(U_NU_2) \\ \vdots & \vdots & \ddots & \vdots \\ E(U_1U_N) & E(U_2U_N) & \cdots & E\left(U_N^2\right) \end{bmatrix}$$

2. Aplicando el supuesto #2: $Var(U_i) = \sigma^2$

$$Var(U) = \begin{bmatrix} \sigma^2 & E(U_2U_1) & \cdots & E(U_NU_1) \\ E(U_1U_2) & \sigma^2 & \cdots & E(U_NU_2) \\ \vdots & \vdots & \ddots & \vdots \\ E(U_1U_N) & E(U_2U_N) & \cdots & \sigma^2 \end{bmatrix}$$

$$Cov(U_i,U_j) = E[(U_i - E(U_i)^2)(U_j - E(U_j)^2)]$$
$$Cov(U_i,U_j) = E(U_i,U_j) = 0$$

3. Aplicando el supuesto # 3: Cov (U$_i, U_j) = 0$

$$Var(U) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$Var(U) = \sigma^2 I_{NxN} \tag{5.10}$$

5.3 Prueba de insesgamiento

Los 3 supuestos anteriores eran sobre el error. Ahora hacemos supuestos sobre X_i.

- Supuesto #4: X no es aleatorio (es no estocástico)
- Supuesto #5: El modelo poblacional tiene lo misma forma que el modelo muestral. Ambos modelos tienen:
  - Una relación lineal entre X y Y, que incluye una constante (el intercepto β_0).
  - Un error (U) que se suma. Este supuesto resume las formas funcionales que hemos visto para el modelo poblacional:

$$Y = \beta X + U \tag{5.11}$$

Y el modelo muestral:

$$Y = \hat{\beta} X + \hat{U} \tag{5.12}$$

La prueba de insesgamiento inicia con los estimadores de MCO

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

Aplicando el supuesto #5: Modelo tiene lo mismo forma que el modelo muestral

$$Y = \beta X + U$$

Evaluando en $\hat{\beta}$:

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

Aplicar esperanza a ambos lados

$$E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'U)$$
$$E(\hat{\beta}) = E(\beta) + E((X'X)^{-1}X'U)$$
$$E(\hat{\beta}) = \beta + E((X'X)^{-1}X'U)$$

Aplicando el supuesto # 4. X está dado, no aleatorio

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(U)$$

Aplicando el supuesto 1. $E(U_i) = 0$

$$E(\hat{\beta}) = \beta$$

¿Por qué es importante que los estimadores sean insesgados?

1. Estimadores cuya distribución esté centrada en los valores poblaciones.
2. Si los estimadores están sesgados, se limita lo que podemos decir sobre los resultados de MCO.

5.4 Varianza de los estimadores

Definición: $Var(\hat{\beta})$

$$Var(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))']$$

Si $k = 2$:

$$Var(\hat{\beta}) = \boxed{\begin{matrix} (\hat{\beta}_0 - E(\hat{\beta}_0))^2 & ((\hat{\beta}_1 - E(\hat{\beta}_1)(\hat{\beta}_0 - E(\hat{\beta}_0)) \\ ((\hat{\beta}_0 - E(\hat{\beta}_0)(\hat{\beta}_1 - E(\hat{\beta}_1)) & (\hat{\beta}_1 - E(\hat{\beta}_1))^2 \end{matrix}}$$

Si $k = 3$:

$$Var(U) = \boxed{\begin{matrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_1, \hat{\beta}_0) & Cov(\hat{\beta}_2, \hat{\beta}_0) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) & Cov(\hat{\beta}_2, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_2) & Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) \end{matrix}}$$

Obtener $Var(\hat{\beta})$:

$$Var(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta})(\hat{\beta} - E(\hat{\beta})']$$

Usando la propiedad de insesgamiento: $E(\hat{\beta}) = \beta$

$$Var(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

Usando la expresión que encontramos en la prueba de insesgamiento

$$\hat{\beta} = \beta + (X'X)^{-1}X'U$$
$$\hat{\beta} - \beta = (X'X)^{-1}X'U$$
$$Var(\hat{\beta}) = E[(((X'X)^{-1}X'U)(((X'X)^{-1}X'U)']$$
$$Var(\hat{\beta}) = E[(((X'X)^{-1}X'U)(U'X((X'X)')^{-1})]$$
$$Var(\hat{\beta}) = E[(((X'X)^{-1}X'U)(U'X(X'X)^{-1})]$$
$$Var(\hat{\beta}) = E[(X'X)^{-1}X'UU'X(X'X)^{-1}]$$

Aplicamos el supuesto #4: X no es aleatorio

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

Como hicimos previamente, usando estos supuestos:
1. $E(U_i) = 0$
2. $Var(U_i) = \sigma^2$
3. $Cov(U_i, U_j) = 0$

Se puede obtener $Var(U) = E(UU') = \sigma^2 I_{NxN}$

$$
\begin{aligned}
Var(\hat{\beta}) &= (XX)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
Var(\hat{\beta}) &= (XX)^{-1}X'\sigma^2 X(X'X)^{-1} \\
Var(\hat{\beta}) &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
Var(\hat{\beta}) &= \sigma^2(X'X)^{-1}
\end{aligned}
$$

¿Por qué es importante esta matriz $Var(\hat{\beta})$?

1. Saber tanto como podamos sobre la distribución de los estimadores.
2. Obtener varianzas pequeñas, para que los estimadores estén concetados con mayor probabilidad alrededor de su esperanza, que es igual al valor poblacional, ya que son insesgadas.

Estábamos viendo el modelo matricial

$$Y = X\beta + U \qquad \text{Modelo poblacional}$$

Estábamos viendo las propiedades estadísticas que tienen que ver con los supuestos. *S̈i los supuestos son... entonces las propiedades estadśticas...¨.*

Supuestos:

- $E(U_i) = 0$
- $Var(U_i) = \sigma^2$
- $Cov(U_i, U_j) = 0 \qquad \forall i \neq j$
- X no es aleatorio
- Modelo muestral .es igual"(la forma funcional es igual) al modelo poblacional

$$Y = X\hat{\beta} + \hat{U} \rightarrow MM$$

$$Y = X\beta + U \rightarrow MP$$

MCO: $\hat{\beta} = (X'X)^{-1}X'Y$

- Modelo muestral:

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

$$MCO:\ \hat{\beta} = \underbrace{(X'X)^{-1}X'}_{KxN}\underbrace{Y}_{Nx1}$$

6. Propiedades estadísticas

- Insesgamiento
- Varianza mínima

Hay dos matrices que permiten distinguir modelos: dos matrices de varianzas y covarianzas de interés:

1. $Var(U)_{NxN}$
2. $Var(\hat{\beta})_{KxK}$

6.1 Matriz de varianza de los errores

No tiene techito. Da información sobre el proceso generador de datos (PGD) y lo vamos a ligar a los supuestos.

Definición:

$$Var(U) = \begin{bmatrix} Var(U_1) & Cov(U_1,U_2) & \cdots & Cov(U_1,U_N) \\ Cov(U_1,U_2) & Var(U_2) & \cdots & Cov(U_2,U_N) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(U_1,U_N) & Cov(U_2,U_N) & \cdots & Var(U_N) \end{bmatrix}$$

Usando los supuestos 1,2 y 3:

$$E(UU') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$E(UU') = \sigma^2 I$

$$Var(\hat{\beta}) = \boxed{\begin{matrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0,\hat{\beta}_1) & \cdots & Cov(\hat{\beta}_0,\hat{\beta}_{k-1}) \\ Cov(\hat{\beta}_0,\hat{\beta}_1) & Var(\hat{\beta}_1) & \cdots & Cov(\hat{\beta}_1,\hat{\beta}_{k-1}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_0,\hat{\beta}_{k-1}) & Cov(\hat{\beta}_1,\hat{\beta}_{k-1}) & \cdots & Var(\hat{\beta}_{k-1}) \end{matrix}}$$

¿Cómo se calcula esta matriz?

- $Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \rightarrow$ Matriz poblacional
- $\hat{Var}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} \rightarrow$ Matriz muestral (s^2)

6.2 Prueba de varianza mínima

$$Var(U) = E(UU') = \sigma^2 I$$
$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

$Var(\beta)$ no está deinida, no es aleatoria.

Este nombre es incorrecto, porque lo mínimo que podría ser es cero, y la varianza de MCO no va a ser cero.

Correcto: comparar las varianzas de dos estimadores

***MELI: Mejores Estimadores Lineales Insesgados**
***BLUE: Best Lineal Unbiased Estimater**

Comparar la varianza de dos estimadores lineales insesgados:

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

Y tomamos otro estimador:

$$\beta = CY$$

es lineal con una matriz de ponderadores C, calculada con datos de X, tal que

$$C \neq (X'X)^{-1}X'Y$$

¿Qué condiciones debe cumplir C para que β sea insesgado? $\rightarrow$ Para encontrar la condición, usamos el supuesto 5, que dice que:

$$Y = X\beta + U \qquad MP$$

reemplazamos en:

$$\beta = CY$$
$$\beta = C(X\beta + U)$$
$$\beta = CX\beta + CU$$

Entonces tenemos:

$$\beta = CX\beta + CU \qquad (**) \tag{6.1}$$

Calculando $E(\cdot)$:

$$E(\beta) = E(CX\beta) + E(CU)$$

Usando el supuesto 4, X no es aleatorio, y el hecho de que C se calculó solo con datos de X, entonces:

- CX no es aleatorio
- C no es aleatorio

Entonces

$$\begin{aligned} E(\beta) &= CXE(\beta) + CE(U) \\ \text{Como } \beta \text{ es poblacional (no aleatorio)} \\ E(\beta) &= CX\beta + CE(U) \\ \text{Usando supuesto 1: } E(U) = 0, \text{ por lo que } CE(U) &= 0 \end{aligned}$$

Para que se cumpla insesgamiento debe cumplirse que:

$$CX = \beta$$

Lo que se cumple si

$$CX = I$$

Entonces: $\beta = CY_{KxN}$ es lineal insesgado

- C se calcula con X
- $C = (X'X)^{-1}X'$
- C cumple que $CX = I$

Calculamos $Var(\beta)$:

$$Var\left(\tilde{\beta}\right) = E\left[\underbrace{\left(\underbrace{\tilde{\beta}}_{Kx1} - \underbrace{E\left(\tilde{\beta}\right)}_{Kx1}\right)\underbrace{\left(\underbrace{\tilde{\beta}}_{Kx1} - \underbrace{E\left(\tilde{\beta}\right)}_{Kx1}\right)'}_{1xK}}_{KxK}\right]$$

Usamos insesgamiento $E(\beta) = \beta$:

$$\begin{aligned} Var(\beta) &= E[(\beta - E(\beta))(\beta - E(\beta))'] \\ &\text{Usando (**) y CX = I implican:} \\ \beta &= \beta + CU \qquad (***) \\ \beta - \beta &= CU \\ &\text{Reemplazando en } Var(\beta) \\ Var(\beta) &= E[(\beta - \beta)(\beta - \beta)'] \\ Var(\beta) &= E[(CU)(CU)'] \\ Var(\beta) &= E[CUU'C'] \end{aligned}$$

Usando el supuesto 4, X no aleatorio y C se calcula con X, entonces:

$$Var(\beta) = CE[UU']C' \qquad \text{No son conmutativos} \rightarrow \text{respetar el orden!} \tag{6.2}$$

Usando el supuesto 1,2 y 3 se obtiene:

$$E(UU') = \sigma^2 I$$

Reemplazando:

$$Var(\beta) = C(\sigma^2 I)C'$$
$$Var(\beta) = \sigma^2 CC'$$

Definimos:

$$D = C - (X'X)^{-1}X' \tag{6.3}$$

Ejemplo: $k = 2$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum k_i^0 Y_i \\ \sum k_i^1 Y_i \end{bmatrix} = \begin{bmatrix} k_1^0 & k_2^0 & \cdots & k_N^0 \\ k_1^1 & k_2^1 & \cdots & k_N^1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$= \underset{KxN}{(X'X)^{-1}X'} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$D = C - (X'X)^{-1}X' = \begin{bmatrix} c_1^0 - k_1^0 & c_2^0 - k_2^0 & \cdots & c_N^0 - k_N^0 \\ c_1^1 - k_1^1 & c_2^1 - k_2^1 & \cdots & c_N^1 - k_N^1 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum k_i^0 Y_i \\ \sum k_i^1 Y_i \end{bmatrix} = \begin{bmatrix} k_1^0 & k_2^0 & \cdots & k_N^0 \\ k_1^1 & k_2^1 & \cdots & k_N^1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$= \underset{KxN}{(X'X)^{-1}X'} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$D = C - (X'X)^{-1}X' = \begin{bmatrix} c_1^0 - k_1^0 & c_2^0 - k_2^0 & \cdots & c_N^0 - k_N^0 \\ c_1^1 - k_1^1 & c_2^1 - k_2^1 & \cdots & c_N^1 - k_N^1 \end{bmatrix}$$

Usando $D = C - (X'X)^{-1}X'$, entonces:

$$C = D + (X'X)^{-1}X'$$

Reemplazamos en $Var(\beta)$:

$$Var(\beta) = \sigma^2(D + (X'X)^{-1}X')(D + (X'X)^{-1}X')'$$
$$Var(\beta) = \sigma^2(D + (X'X)^{-1}X')(D' + X'(X'X)^{-1})'$$
$$Var(\beta) = \sigma^2(DD' + DX(X'X)^{-1} + (X'X)^{-1} + (X'X)^{-1}X'D' + (X'X)^{-1}X'X(X'X)^{-1}$$

Para DX:

$$DX = (C - (X'X)^{-1}X')X$$
$$DX = CX - (X'X)^{-1}X'X$$
$$DX = I - I$$
$$DX = 0$$

Igualmente $X'D' = 0$

$$Var(\beta) = \sigma^2(DD' + (X'X)^{-1}X'X(X'X)^{-1})$$
$$Var(\beta) = \sigma^2(DD' + (X'X)^{-1})$$
$$Var(\beta) = \sigma^2DD' + \sigma^2(X'X)^{-1}$$
$$Var(\beta) = \sigma^2DD' + Var(\hat{\beta})$$

*Tarea: calcular DD' para ver si en la diagonal quedan valores positivos.

Se están comparando dos estimadores lineales insesgados

$$\hat{\beta} = \underbrace{(X'X)^{-1}X'}_{\text{matriz de ponderadores}}Y \qquad \neq \qquad \tilde{\beta} = \underbrace{C}_{\text{matriz de ponderadores}}Y$$

$$Var(\hat{\beta}) \;=\; \sigma^2(X'X)^{-1} \qquad \longleftrightarrow \qquad Var\left(\tilde{\beta}\right) = CC'$$

Matrices:

$$\hat{\beta} \qquad\qquad\qquad\qquad \tilde{\beta}$$

$$\boxed{\begin{array}{cccc} k_1^0 & k_2^0 & \cdots & k_N^0 \\ k_1^1 & k_2^1 & \cdots & k_N^1 \end{array}} \qquad \boxed{\begin{array}{cccc} c_1^0 & c_2^0 & \cdots & c_N^0 \\ c_1^1 & c_2^1 & \cdots & c_N^1 \end{array}}$$

donde:

$$Var(\beta) = (DD') + \sigma^2(X'X)^{-1}$$

donde

$$D \;=\; C - (X'X)^{-1}X' = \boxed{\begin{array}{cccc} c_1^0 - k_1^0 & c_2^0 - k_2^0 & \cdots & c_N^0 - k_N^0 \\ c_1^1 - k_1^1 & c_2^1 - k_2^1 & \cdots & c_N^1 - k_N^1 \end{array}}$$

$$DD' = \boxed{\begin{matrix} \Sigma\left(c_1^0 - k_1^0\right)^2 & \Sigma\left(c_1^0 - k_2^0\right)\left(c_1^1 - k_1^1\right) \\ \Sigma\left(c_1^1 - k_1^1\right)\left(c_1^0 - k_1^0\right) & \Sigma\left(c_1^1 - k_1^1\right)^2 \end{matrix}}_{2x2}$$

donde:

$$\begin{aligned} &\sum (c_1^0 - k_1^0)^2 \geq 0 \\ &\sum (c_1^1 - k_1^1)^2 \geq 0 \end{aligned} \tag{6.4}$$

$$Var\left(\tilde{\beta}\right)$$

$$\boxed{\begin{matrix} Var\left(\tilde{\beta}_0\right) & Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) \\ Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) & Var\left(\tilde{\beta}_1\right) \end{matrix}} = DD' + \boxed{\begin{matrix} Var\left(\tilde{\beta}_0\right) & Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) \\ Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) & Var\left(\tilde{\beta}_1\right) \end{matrix}}$$

$$\boxed{\begin{matrix} Var\left(\tilde{\beta}_0\right) & Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) \\ Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) & Var\left(\tilde{\beta}_1\right) \end{matrix}} = \boxed{\begin{matrix} \Sigma\left(c_i^0 - k_i^0\right)^2 & \Sigma\left(c_i^0 - k_i^0\right)\left(c_i^1 - k_i^1\right) \\ \Sigma\left(c_i^1 - k_i^1\right)\left(c_i^0 - k_i^0\right) & \Sigma\left(c_i^1 - k_i^1\right)^2 \end{matrix}}_{2x2} \boxed{\begin{matrix} Var\left(\tilde{\beta}_0\right) & Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) \\ Cov\left(\tilde{\beta}_0, \tilde{\beta}_1\right) & Var\left(\tilde{\beta}_1\right) \end{matrix}}$$

¿Bajo qué condiciones son iguales a 0?

$$\begin{aligned} &\sum (c_1^0 - k_1^0)^2 \quad \text{es igual a 0 si } \beta_0 = \hat{\beta}_0 \\ &\sum (c_1^0 - k_1^0)^2 \quad \text{es igual a 0 si } \beta_1 = \hat{\beta}_1 \end{aligned} \tag{6.5}$$

IV Inferencia y especificación

6.2.1 Inferencia (prueba de hipótesis)

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

- k =2

$$(X'X)^{-1} = \begin{bmatrix} \frac{\sum x_i^2}{N\sum x_i^2} & \frac{-\overline{X}}{\sum x_i^2} \\ \frac{-\overline{X}}{\sum x_i^2} & \frac{1}{\sum x_i^2} \end{bmatrix}$$

$$Var\left(\tilde{\beta}\right) = \begin{bmatrix} \frac{\sigma^2 \sum x_i^2}{N\sum x_i^2} & \frac{-\sigma^2\overline{X}}{\sum x_i^2} \\ \frac{-\sigma^2\overline{X}}{\sum x_i^2} & \frac{\sigma^2}{\sum x_i^2} \end{bmatrix}$$

***Las propiedades estadísticas nos dan información sobre la distribución**

$$\begin{aligned} Var(\hat{\beta}_0) &= \frac{\sigma^2 \sum x_i^2}{\sum x_i{}^2} \quad \text{más grande: } \uparrow \sigma^2 \downarrow N \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{\sum x_i{}^2} \quad \text{más grande: } \uparrow \sigma^2 \downarrow \sum x_i{}^2 \end{aligned} \tag{6.6}$$

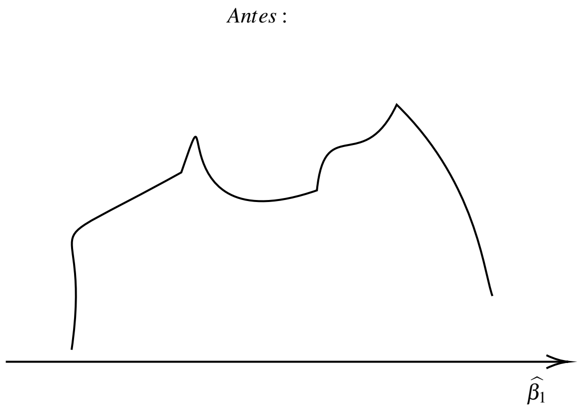
¿Como estimamos $Var(\hat{\beta})$? "Necesitamos.º "queremos":

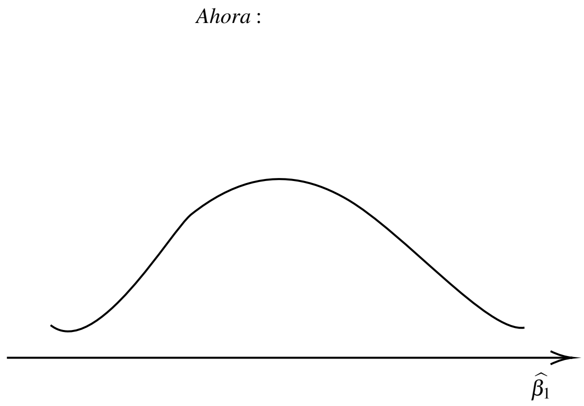
$$\hat{\sigma^2} = \frac{\sum \hat{U}_i}{N-K} = \frac{SRC}{N-K}$$

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

**Excel*

- Nuevo supuesto #6: $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$
  Esto nace más atrás de dos opciones:
  - $U_i \sim N(0, \sigma^2) \Rightarrow \hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$
  - Ley de los Grandes Números $U_i \rightarrow$
    - $E(U_i) = 0$
    - $Var(U_i) = \sigma^2$





Estimador de MCO tiene una particularidad: no se necesitan experimentos de Monte Carlo para saber su distribución

$$\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$$

Estandarizando:

$$z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}} \sim N(0,1) \tag{6.7}$$

*z no se puede ver en los datos.

Entonces:

$$z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{(\frac{\sigma^2}{\sum {x_i}^2})}} \sim N(0,1) \tag{6.8}$$

y

$$(N-K)\frac{\hat{\sigma^2}}{\sigma^2} \sim \chi^2_{N-K} \tag{6.9}$$

Definición T Student:

$$t = \frac{N(0,1)}{\sqrt{\frac{\chi^2_{N-K}}{N-K}}} \sim t_{N-K} \tag{6.10}$$

Así, reemplazando y evaluando:

$$t = \frac{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum x_i}}}}{\sqrt{\frac{(N-K)}{(N-K)}\frac{\hat{\sigma^2}}{\sigma^2}}} = \frac{\cancel{\sqrt{\sigma^2}}}{\sqrt{\hat{\sigma}^2}} \cdot \frac{\hat{\beta}_1 - \beta}{\frac{\cancel{\sqrt{\sigma^2}}}{\sum x_i}} = \frac{\hat{\beta}_1 - \beta}{\frac{\sqrt{\hat{\sigma}^2}}{\sum x_i}} \tag{6.11}$$

Así:

$$t = \frac{(\hat{\beta}_1 - \beta_1)}{\sqrt{\frac{\hat{\sigma^2}}{\sum {x_i}^2}}} \tag{6.12}$$

Los estadísticos son funciones que combinan resultados muestrales con parámetros poblacionales. Tienen una distribución de probabilidad.

6.2.2 Pruebas de hipótesis

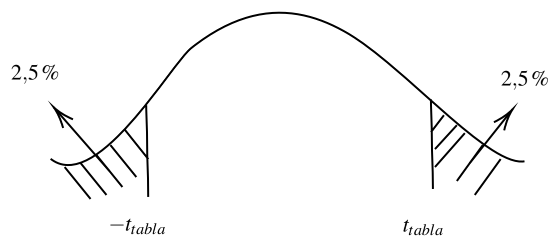
- Paso 0: saber o escoger previamente el estadístico a usar.
- Paso 1: plantear la hipótesis.
  - H_0: es un valor c a evaluar en el parámetro poblacional

$$\beta_1 = c$$

  - H_1: $\beta_1 \neq c$
- Paso 2: evaluar el estadístico en la hipótesis (obtener el t calculado)

$$t_{calculado} = \frac{\hat{\beta}_1 - c}{\sqrt{\hat{Var}(\hat{\beta}_1)}} \tag{6.13}$$

- Paso 3: comparar el t calculado con el t de la tabla. **Excel*



  - 3.1: ¿Cómo obtener el t tabla?
    - Seleccionar el nivel de significancia

$$\alpha = [10\,\%, 5\,\%, 1\,\%]$$

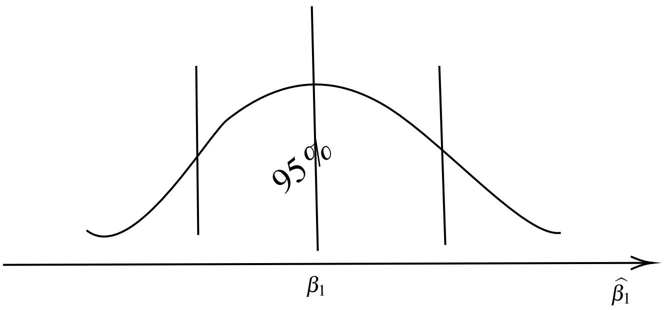
      Por ejemplo: $\alpha = 5\,\%$
    - Dividimos $\frac{\alpha}{2} = 2{,}5\,\%$ (porque son dos colas)
      Obtenemos 3.18
  - 3.2
    - Si $t_{cal} \in [-t_{tabla}, t_{tabla}]$ entonces NO se rechaza H_0.
    - Si $t_{cal} \notin [-t_{tabla}, t_{tabla}]$ entonces se rechaza H_0.
  - 3.3: Interpretación en palabras:
    - $H_0 : \beta_1 = 0$
      No rechazar implica que no hay evidencia de una relación estadística entre X y Y. Implica que no hay evidencia de relación entre X y Y (con un nivel de significancia de α).
    - Rechazar implica que sí hay evidencia de una relación estadística entre X y Y (con un nivel de significancia de α).

6.2.3 Intervalos de confianza ($1 - \alpha$ de confianza)

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{Var}(\hat{\beta}_1)}} \sim t_{N-K} \tag{6.14}$$

$$P[-t_{tabla} \leq t \leq t_{tabla}] = 95\,\%$$

$$
\begin{aligned}
-t_{tabla} \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} &\leq t_{tabla} \\
-t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} \leq \hat{\beta}_1 - \beta_1 &\leq t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} \\
-t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} \leq \hat{\beta}_1 - \beta_1 &\leq t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} \\
-t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} - \hat{\beta}_1 \leq -\beta_1 &\leq t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} - \hat{\beta}_1 \\
t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} + \hat{\beta}_1 \geq \beta_1 &\geq -t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} + \hat{\beta}_1 \\
\hat{\beta}_1 + t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)} \geq \beta_1 &\geq \hat{\beta}_1 - t_{tabla} \cdot \sqrt{V\hat{a}r(\hat{\beta}_1)}
\end{aligned}
\tag{6.15}
$$



De este lunes en 8: quiz de laboratorio en STATA (enfocado en manejo de datos), unas 2 horas, va a ser largo, como lo de la práctica.

6.3 Video Temático A: Multicolinealidad Perfecta

6.3.1 Existencia del estimador de MCO

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

Solo puede existir si $(X'X)^{-}1$, y esto se cumple si la matriz tiene rango completo.

Criterios de rango completo:

1. Para que una matriz tenga rango completo, sus columnas deben ser linealmente independientes.
2. Una matriz no tiene rango completo si al menos dos de sus columnas no son linealmente dependientes.

Concepto de multicolinealidad perfecta: se presenta multicolinealidad perfecta cuando la matriz X tiene al menos dos columnas que son linealmente dependientes. La multicolinealidad perfecta indica que el estimador de MCO no existe por que $(X'X)$ no se puede invertir.

Dos columnas son linealmente dependientes si una de ellas es una función lineal de la otra.

$$X_{1i} = cX_{2i}$$

En los datos, X_2 es el crédito . Esto podría ser en cualquier caso en el cual la misma variable se incluye dos veces, con diferentes unidades de medida.

$$X_{5i} = X_{1i} + X_{3i} + X_{4i}$$

Tampoco tiene sentido inlcuir el total y al mismo tiempo de los componentes de este total. POr lo que se debe eliminar una de las cuatro variables.

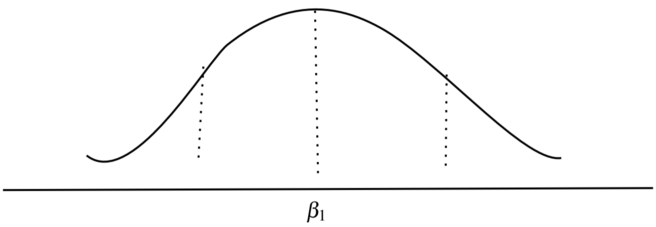
La multicolinealidad perfecta surge cuando se toman decisiones que no tienen sentido, porque se incluye la misma información dos o más veces dentro de una misma matriz X.

STATA sabe esto y la elimina automáticamente. Esto demuestra que el investigador no entiende el concepto de multicolinealidad perfecta. Además se quiere el poder de poder escoger la variable a eliminar.

6.3.2 Resumen de las pruebas posibles

	Hiptesis nula	*Hiptesis alternativa*	*Nivel de significancia*
Prueba de significancia (no correlacin entre variable X y Y)	$\beta_1 = 0$	*Dos colas* : $\beta_1 \neq 0$	$\alpha/2$
		Relacin positiva $\beta_1 > 0$	α
		Relacin negativa $\beta_1 < 0$	α
Prueba t general	$\beta_1 = c$	*Dos colas* : $\beta_1 \neq c$	$\alpha/2$
		Relacin positiva $\beta_1 > c$	α
		Relacin negativa $\beta_1 < c$	α

$$t \sim \frac{N(0,1)}{\chi^2_{N-K}} \tag{6.16}$$



$$\hat{\beta}_1 - \sqrt{v\hat{a}r}(\hat{\beta}_1) \cdot t_{\alpha/2} < \beta_1 < \hat{\beta}_1 + \sqrt{v\hat{a}r}(\hat{\beta}_1) \cdot t_{\alpha/2}$$

6.3.3 Criterio Intervalo de confianza

Una prueba de hipótesis se hace verificando si un valor específico de $\hat{\beta}_1$ está o no en el intervalo. Por ejemplo:

$$H_0 : \beta_1 = c$$

En una prueba de dos colas, se verifica si c está o no en el intervalo:

$$[\hat{\beta}_1 - \sqrt{v\hat{a}r}(\hat{\beta}_1) \cdot t_{\alpha/2}, \hat{\beta}_1 + \sqrt{v\hat{a}r}(\hat{\beta}_1) \cdot t_{\alpha/2}]$$

Reglas de decisión:

- Si c no pertenece al intervalo, se rechaza la hipótesis nula.
- Si c pertenece al intervalo, no se rechaza la hipótesis nula.

¿Qué nos dice esto sobre la distribución?

Conociendo la ecuación, podemos establecer un intervalo de confianza:

$$Prob(-t_{\alpha/2} < \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} < t_{\alpha/2}) = 0{,}95 \tag{6.17}$$

Despejando, se obtiene que β_1 pertenece, con 95 % de probabilidad, al siguiente intervalo:

$$\hat{\beta}_1 - \sqrt{v\hat{a}r}(\hat{\beta}_1) \cdot t_{\alpha/2} < \beta_1 < \hat{\beta}_1 + \sqrt{v\hat{a}r}(\hat{\beta}_1) \cdot t_{\alpha/2} \tag{6.18}$$

6.3.4 Otras formas de obtener conclusiones

Primero se debe calcular el estadstico :

$$t_c = \frac{\hat{\beta}_1 - c}{\sqrt{\widehat{var}\left(\hat{\beta}_1\right)}} \qquad t_c = \frac{\hat{\beta}_1 - 0}{\sqrt{\widehat{var}\left(\hat{\beta}_1\right)}}$$

Frmula general *Hiptesis de significancia*

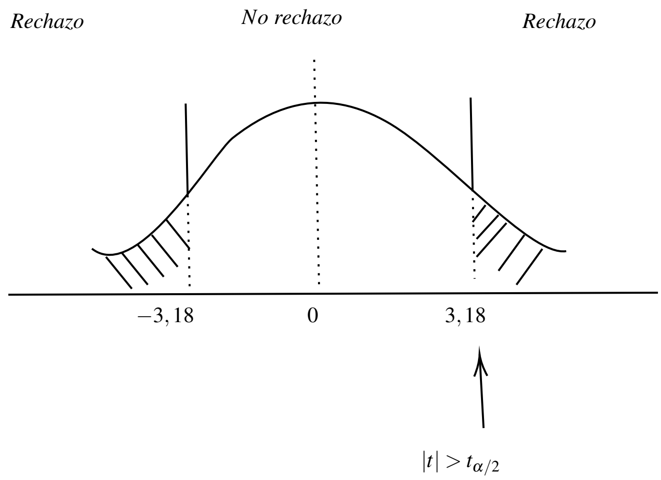
Adems del intervalo de confianza, hay dos formas de obtener las conclusiones en las cuales se usa el t calculado:

1. Usando la tabla t
2. Usando el valor P

6.3.5 Tabla t

- Escoger un nivel de significancia
- Para un nivel de significancia de 5 %, el valor t de tabla se busca para $(\alpha/2) = 0{,}025$
- En la tabla se ve que el t de tabla es 3.18
  Ahora vamos a obtener el t calculado

$$t_c = \frac{\hat{\beta}_1 - 0}{\sqrt{V\hat{a}r(\hat{\beta}_1)}}$$

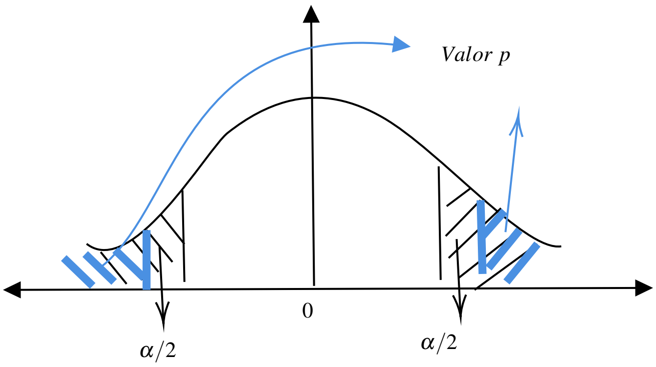


6.3.5.1 Regla de decisión:

- Si el t calculado (en valor absoluto) es mayor que el t de tabla, se rechaza la hipótesis nula. Esto se interpreta como que sí existe evidencia estadística de una relación entre X y Y. También se dice que X es significativo para explicar Y.
- Si el t calculado (en valor absoluto) es menor que el t de tabla, no se rechaza la hipótesis nula. Esto se interpreta como que no existe evidencia estadística de una relación entre X y Y. También se dice que X no es significativo para explicar Y.

6.3.6 Usando el valor P

El valor P es el área bajo la distribución *t-student* que se puede comparar directamente con la significancia



6.3.6.1 Regla de decisión:

- Si el valor P es **menor** a α **se rechaza** la hipótesis nula. Esto se interpreta como que **sí existe evidencia estadística** de una relación entre X y Y. También se dice que X es significativa para explicar Y.
- Si el valor P es **mayor** a α **no se rechaza** la hipótesis nula. Esto se interpreta como que **no existe evidencia estadística** de una relación entre X y Y. También se dice que X no es significativa para explicar Y.

6.3.7 Otras pruebas sobre los coeficientes

$$Y = X\beta + U$$

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

$$\hat{\beta}_{Kx1} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{k-1} \end{bmatrix} \qquad X_{NxK} = \begin{array}{c} \begin{array}{ccccc} & X_{1i} & X_{2i} & \cdots & Xk-1_i \end{array} \\ \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k-11} \\ 1 & X_{12} & X_{22} & \cdots & X_{k-12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & X_{2N} & \cdots & X_{k-1N} \end{bmatrix} \end{array}$$

$$Wage_i = \beta_0 + \beta_1 grade_i + U_i$$

$$Y_{i_{Nx1}} = X_{NxK-1}\beta_{Kx1} + U_{i_{Nx1}}$$

V Selección del modelo

6.3.8 Prueba de significancia conjunta

A veces se tiene más de una variable explicativa:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_i$$

- $H_0 : \hat{\beta}_1 = \hat{\beta}_2 = 0$ $\quad\rightarrow$ Hipótesis nula
- $H_1 :$ al menos uno distinto de 0 $\quad\rightarrow$ Hipótesis alternativa

Vamos a crear un estadístico:

$$Estadstico\ F \sim \underbrace{\frac{SEC/(k-1)}{SRC/(N-K)}}_{calculado} \sim F_{k-1,N-K}$$

6.3.8.1 Reglas de decisión

- Si el F calculado > F tabla → se rechaza H_0: hay evidencia estadística de que las variables explicativas en su conjunto sí son significativas.
- Si el F calculado < F tabla → no se rechaza H_0: evidencia estadística que las variables explicativas en su conjunto no son significativas.

6.3.9 Prueba de Wald

Permite introducir restricciones en el modelo: restricciones tipo lineales sobre los coeficientes.

→ **partiendo como base de un modelo poblacional** $Y = X\beta + U$

1. $\beta_3 = 0$
   $\beta_4 = 0$

$$\underset{R}{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}} \cdot \underset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}} = \underset{q}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

2. $3\beta_1 + \beta_2 = -4$
   $\beta_3 = 2$

$$\underset{R}{\begin{bmatrix} 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}} \cdot \underset{\beta}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}} = \underset{q}{\begin{bmatrix} -4 \\ 2 \end{bmatrix}}$$

3. $-\beta_1 = \beta_3$
   $\beta_0 = 5$
   $\beta_4 = 3$

$$\overset{R}{\begin{bmatrix}0&-1&0&-3&0\\1&0&0&0&0\\0&0&0&1&0\end{bmatrix}} \cdot \overset{\beta}{\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\\\beta_3\\\beta_4\end{bmatrix}} = \overset{q}{\begin{bmatrix}0\\5\\3\end{bmatrix}}$$

*Cada ecuación es una restricción. Las restricciones lineales se expresan en un sistema de m ecuaciones lineales.

Forma matricial que pueden tener las restricciones:

$$R_{mxk}\beta_{kx1} = q \qquad \rightarrow \text{R} \wedge \text{Q contienen parámetros no aleatorios} \tag{6.19}$$

En resumen:

$$H_0 : R\beta = q$$
$$H_1 : R\beta \neq q$$

- Cuando no se rechaza, la evidencia estadística no permite rechazar que las restricciones se cumplen.
- Cuando se rechaza, tenemos evidencia estadística de que al menos una restricción no se cumple.

Para verificar la hipótesis nula (H_0) se usa el estadístico de Wald:

$$W \sim F_{m,N-K} \tag{6.20}$$

Dos fórmulas:

1. Primera fórmula:

$$W = \frac{(R\beta - q)'[R\sigma^2(X'X)^{-1}R']^{-1}(R\beta - q)}{m} \sim F_{m,N-K} \tag{6.21}$$

   *Fórmula 1: la lógica es aplicar las restricciones lineales a los estimadores de MCO. Se define una diferencia S, brecha entre la restricción y los coeficientes estimados.

$$S \equiv R\hat{\beta} - q \tag{6.22}$$

2. Segunda fórmula: usa MCO restringidos para obtener lo siguiente

$$W = \frac{(SRC_R - SRC_{NR})/m}{SRC_{NR}/(n-k)} \tag{6.23}$$

6.3.9.1 Criterio de Wald

$$W = S'[Var(S)]^{-1}S \sim \chi^2_m$$
$$W = [R\hat{\beta} - q]'[Var(R\hat{\beta} - q)]^{-1}(R\hat{\beta} - q)$$
$$*Var(S) = Var(R\hat{\beta} - q)$$

Note que q contiene parámetros constantes

$$Var(q) = 0$$

$$Var(S) = Var(R\hat{\beta}) = E[(R\hat{\beta} - E(R\hat{\beta}))(R\hat{\beta} - E(R\hat{\beta}))']$$

*R no es aleatorio

$$Var(R\hat{\beta}) = E[(R\hat{\beta} - RE(\hat{\beta}))(R\hat{\beta} - RE(\hat{\beta}))']$$

$$Var(R\hat{\beta}) = E[(R(\hat{\beta} - E(\hat{\beta})))(R(\hat{\beta} - E(\hat{\beta})))']$$

$$Var(R\hat{\beta}) = E[(R(\hat{\beta} - E(\hat{\beta})))((\hat{\beta} - E(\hat{\beta})'R'))]$$

$$Var(R\hat{\beta}) = RE[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))']R'$$

$$Var(S) = RVar(\hat{\beta})R' \tag{6.24}$$

Evaluando:

$$W = (R\hat{\beta} - q)'[Var(S)]^{-1}(R\hat{\beta} - q)$$

$$W = (R\hat{\beta} - q)'[RVar(\hat{\beta})R']^{-1}(R\hat{\beta} - q)$$

$$\text{*Sabemos que } Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

$$W = (R\hat{\beta} - q)'[R\hat{\beta}) = \sigma^2(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)$$

$$\sigma^2 \text{ es un escalar}$$

$$W = \sigma^{-2}(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \sim \chi^2_m$$

$$\text{No conocemos } \sigma^2$$

Esta no es la expresión que estamos buscando. Por eso usaremos dos resultados:

$$\frac{\hat{\sigma}^2(N-K)}{\sigma^2} \sim \chi^2_{N-K} \sum^k x_i^2 = z \sim x_k \tag{6.25}$$

El segundo resultado:

$$\frac{\chi^2_n/n}{\chi^2_l/l} \tag{6.26}$$

Evaluando:

$$\frac{W/m}{\frac{\hat{\sigma}^2}{\sigma^2}/(N-K)} \sim F_{m,N-K}$$

$$W = \frac{\frac{\sigma^{-2}(R\hat{\beta}-q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-q)}{m}}{\frac{\hat{\sigma}^2\cancel{(N-K)}}{\sigma^2\cancel{(N-K)}}}$$

$$W = \frac{\cancel{\sigma^{-2}}(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)\cancel{\sigma^2}}{m\sigma^2}$$

$$W = \frac{\sigma^{-2}(R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)}{m}$$

$$W = \frac{(R\hat{\beta} - q)'[R\sigma^2(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)}{m}$$

*Fórmula 2: la lógica es estimar MCO restringido y construir un indicador de pérdida de bondad de ajuste. Es importante distinguir entre los estimadores restringidos y los no restringidos. Los estimadores que hemos visto hasta ahora en el curso son los no restringidos.

La función objetivo es:

$$min\{\hat{U}'_{NR}\hat{U}_{NR}\}$$

$\rightarrow \hat{NR} = (X'X)^{-1}X'Y$ Estadístico de Wald en este caso:

$$W = \frac{(R\hat{\beta}_{NR}-q)'[R\hat{\sigma}^2(X'X)^{-1}R'](R\hat{\beta}_{NR}-q)')}{m} \tag{6.27}$$

Estimador MCO restringido se obtiene resolviendo:

$$min\{\hat{U}'_{NR}\hat{U}_{NR}\} \qquad s.a \qquad R\hat{\beta}_R = q$$

$\hat{\beta}_R = \hat{\beta}_{NR} - (X'X)^{-1}R'[R((X'X)^{-1}R']^{-1}$
$\mathscr{L} : \hat{U}_R U_R + 2\lambda'[R\hat{\beta}_R - q]$
$*\hat{U}_R = Y - \hat{Y}_R \equiv Y - X\hat{\beta}_R$

Para encontrar el estadístico nos interesa definir la pérdida de ajuste.

Expresión que contenga errores estimados al cuadrado.

$$\begin{aligned} \hat{U}_R &= Y - X\hat{\beta}_R + O \\ \hat{U}_R &= Y - X\hat{\beta}_R + X\hat{\beta}_{NR} - X\hat{\beta}_{NR} \\ \hat{U}_R &= Y - X\hat{\beta}_{NR} - X\hat{\beta}_R + X\hat{\beta}_{NR} \\ \hat{U}_R &= \hat{U}_{NR} - X\hat{\beta}_R + X\hat{\beta}_{NR} \end{aligned}$$

$$\hat{U}_R = \hat{U}_{NR} - X(\hat{\beta}_R - \hat{\beta}_{NR}) \tag{6.28}$$

$SCR = \hat{U}'_R\hat{U}_R$ Evaluando en la expresión recién encontrada

$$\begin{aligned} \hat{U}'_R\hat{U}_R &= [(\hat{U}_{NR} - X(\hat{\beta}_R - \hat{\beta}_{NR}))'(\hat{U}_{NR} - X(\hat{\beta}_R - \hat{\beta}_{NR}))] \\ \hat{U}'_R\hat{U}_R &= [(\hat{U}'_{NR} - (\hat{\beta}_R - \hat{\beta}_{NR})'X')(\hat{U}_{NR} - X(\hat{\beta}_R - \hat{\beta}_{NR}))] \\ \hat{U}'_R\hat{U}_R &= \hat{U}'_{NR}\hat{U}_{NR} - \hat{U}'_{NR}X\cancel{(\hat{\beta}_R - \hat{\beta}_{NR})} - \cancel{(\hat{\beta}_R - \hat{\beta}_{NR})}X'\hat{U}_{NR} + (\hat{\beta}_R - \hat{\beta}_{NR})X'X(\hat{\beta}_R - \hat{\beta}_{NR}) \end{aligned}$$

Note que $X'\hat{U}_{NR} = 0$

$$\begin{aligned} X'\hat{U}_{NR} &= X'(Y - X\hat{\beta}_{NR}) \\ X'\hat{U}_{NR} &= X'Y - X'X(X'X)^{-1}X'Y \\ X'\hat{U}_{NR} &= X'Y - X'Y \\ X'\hat{U}_{NR} &= 0 \end{aligned}$$

Así

$$\begin{aligned} \hat{U}'_R\hat{U}_R &= \hat{U}'_{NR}\hat{U}_{NR} + (\hat{\beta}_R - \hat{\beta}_{NR})'X'X(\hat{\beta}_R - \hat{\beta}_{NR}) \\ \hat{U}'_R\hat{U}_R - \hat{U}'_{NR}\hat{U}_{NR} &= (\hat{\beta}_R - \hat{\beta}_{NR})'X'X(\hat{\beta}_R - \hat{\beta}_{NR}) \\ SCR_R - SCR_{NR} &= (\hat{\beta}_R - \hat{\beta}_{NR})'X'X(\hat{\beta}_R - \hat{\beta}_{NR}) \end{aligned}$$

Sabíamos que $\hat{\beta}_R = \hat{\beta}_{NR} - (X'X)^{-1}R'[R((X'X)^{-1}R']^{-1}(R\beta - q)$. Entonces:

$$SCR_R - SCR_{NR} = \{-(X'X)^{-1}R'[R((X'X)^{-1}R']^{-1}(R\beta - q)\}'X'X\{-(X'X)^{-1}R'[R((X'X)^{-1}R']^{-1}(R\beta - q)\}$$

Eliminando los negativos y transponiendo:

$$\begin{aligned} SCR_R - SCR_{NR} &= (R\beta - q)'[R((X'X)^{-1}R']^{-1}R\cancel{(X'X)}^{-1}\cancel{X'X}(X'X)^{-1}R'[R((X'X)^{-1}R']^{-1}(R\beta - q) \\ SCR_R - SCR_{NR} &= (R\beta - q)'[R((X'X)^{-1}R']^{-1}R\cancel{(X'X)}^{-1}R'[R(\cancel{(X'X)}^{-1}R']^{-1}(R\beta - q) \\ SCR_R - SCR_{NR} &= (R\beta - q)'[R((X'X)^{-1}R']^{-1}(R\beta - q) \end{aligned}$$

Retomando:

$$W=\frac{(R\beta-q)'[R\hat{\sigma}^2_{NR}((X'X)^{-1}R']^{-1}(R\beta-q)}{m}$$

$$W=\frac{(R\beta-q)'[R((X'X)^{-1}R']^{-1}(R\beta-q)}{m\hat{\sigma}^2_{NR}}$$

Se tiene:

$$W=\frac{SCR_R-SCR_{NR}}{m\hat{\sigma}^2_{NR}}$$

Y

$$\hat{\sigma}^2_{NR}=\frac{SCR}{(N-K)}=\frac{\hat{U}'_{NR}\hat{U}_{NR}}{(N-K)}$$

Entonces:

$$W=\frac{SCR_R-SCR_{NR}}{m\cdot\frac{SCR}{(N-K)}}=\frac{\frac{SCR_R-SCR_{NR}}{m}}{\frac{SCR_{NR}}{(N-K)}}$$

La prueba t es un caso de Wald

$$H_0:\beta_1=C$$
$$k=3$$

$$R=\begin{bmatrix}0 & 1 & 0\end{bmatrix}\cdot\quad q=c$$

$$R\beta=\begin{bmatrix}0 & 1 & 0\end{bmatrix}\cdot\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}$$

$$W=\frac{[R\hat{\beta}-q]'(R\hat{\sigma}^2(X'X)^{-1}R')[R\hat{\beta}-q]}{m}$$

$$W=\frac{(R\hat{\sigma}^2(X'X)^{-1}R')}{m}$$

$$=\begin{bmatrix}0 & 1 & 0\end{bmatrix}\begin{bmatrix}\widehat{Var}(\hat{\beta}_0) & \widehat{Cov}(\hat{\beta}_0,\hat{\beta}_2) & \widehat{Cov}(\hat{\beta}_2,\hat{\beta}_0)\\ \widehat{Cov}(\hat{\beta}_1,\hat{\beta}_0) & \widehat{Var}(\hat{\beta}_1) & \widehat{Cov}(\hat{\beta}_2,\hat{\beta}_1)\\ \widehat{Cov}(\hat{\beta}_0,\hat{\beta}_2) & \widehat{Cov}(\hat{\beta}_2,\hat{\beta}_1) & \widehat{Var}(\hat{\beta}2)\end{bmatrix}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}$$

$$=\begin{bmatrix}\widehat{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \widehat{Var}(\hat{\beta}_1) & \widehat{Cov}(\hat{\beta}_2,\hat{\beta}_1)\end{bmatrix}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}$$

$$=V\hat{a}r(\hat{\beta}_1)=RV\hat{a}r(\hat{\beta})R' \tag{6.29}$$

$$R\hat{\beta} - q =$$

$$R\hat{\beta} = \boxed{\begin{matrix} 0 & 1 & 0 \end{matrix}} \cdot \boxed{\begin{matrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{matrix}}$$

$$R\hat{\beta} = \hat{\beta}_1 \rightarrow 1x1$$
$$R\hat{\beta} - q_{\substack{mx1 \\ 1x1}} = \hat{\beta}_1 - C$$

Reemplazar información en W

$$W = \frac{[R\hat{\beta} - q]'[R\hat{Var}(\hat{\beta}_1)R']^{-1}(R\hat{\beta} - q)}{m}$$
$$W = \frac{(\hat{\beta}_1 - C)'[\hat{Var}(\hat{\beta}_1)]^{-1}(\hat{\beta}_1 - C)}{1}$$
$$W = \frac{(\hat{\beta}_1 - C)'(\hat{\beta}_1 - C)}{[\hat{Var}(\hat{\beta}_1)]}$$
$$W = \frac{(\beta_1 - C)^2}{[\hat{Var}(\hat{\beta}_1)}$$
$$W = t_C^2$$

6.3.10 Variables dicotómicas

¿Por qué hay diferencias salariales entre hombres y mujeres?

$\boxed{\text{Usos de las variables dicotómicas}}$

1. Diferencia en el intercepto
   Siguiente con el ejemplo de la diferencia salarial: hombres y mujeres
   <u>Variable dicotómica:</u> variable que solo puede tomar dos valores. Toma el valor de cero cuando la característica está ausente y toma el valor de 1 cuando la característica está presente.

$$D_i = \begin{cases} 0 & \text{si es mujer} \\ 1 & \text{si es hombre} \end{cases}$$ Con base en el modelo, queremos ver el salario esperado

Obs	Y_i	X_i	D_i
1	1003	8	0
2	578	7	1
3	311	11	1

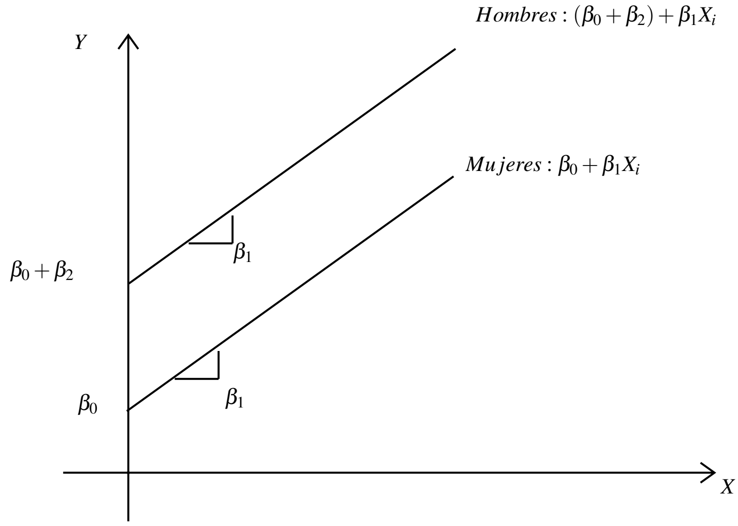
$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 D_i + U_i$$

Para los hombres, el valor esperado

$$\begin{aligned} E[Y_i | D_i = 1] &= E[\beta_0] + E[\beta_1 x_i] + E[\beta_2 D_i] + E[U] \\ &= \beta_0 + \beta_1 x_i + \beta_2 \cdot 1 + 0 \\ &= \beta_0 + \beta_1 x_i + \beta_2 \\ &= \underbrace{(\beta_0 + \beta_2)}_{intercepto} + \underbrace{\beta_1 x_i}_{pendiente} \end{aligned}$$

Para las mujeres

$$\begin{aligned} E[Y_i|\, D_i \;=\; 1] &= E[\beta_0]+E[\beta_1 x_i]+E[\beta_2 D_i]+E[U] \\ &= \beta_0+\beta_1 x_i+\beta_2 \cdot 0+0 \\ &= \beta_0+\beta_1 x_i \\ &= \underbrace{\beta_0}_{intercepto} + \underbrace{\beta_1 x_i}_{pendiente} \end{aligned}$$



2. Diferencia en el intercepto y pendiente

Obs	Y_i	X_i	D_i	$D_i x_i$
1	1003	8	0	0
2	578	7	1	7
3	311	11	1	11

Vamos a plantear el siguiente modelo:

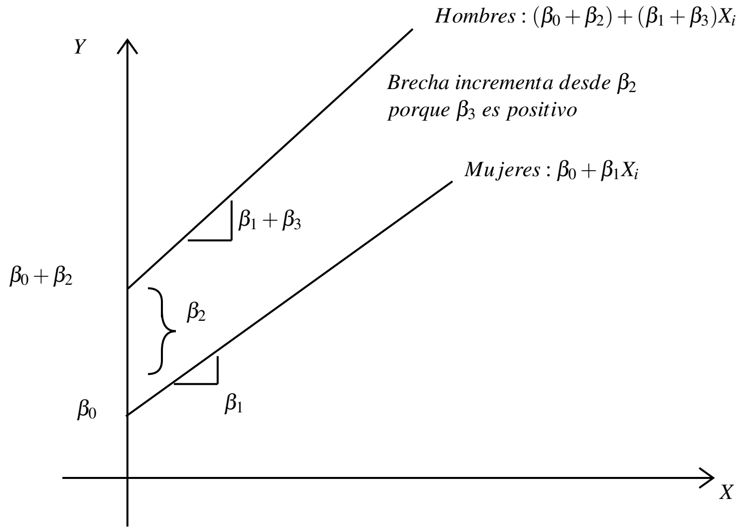
$$Y_i = \beta_0+\beta_1 X_i+\beta_2 D_i+\beta_3 D_i X_i+U_i \tag{6.30}$$

Obtener la esperanza para los hombres

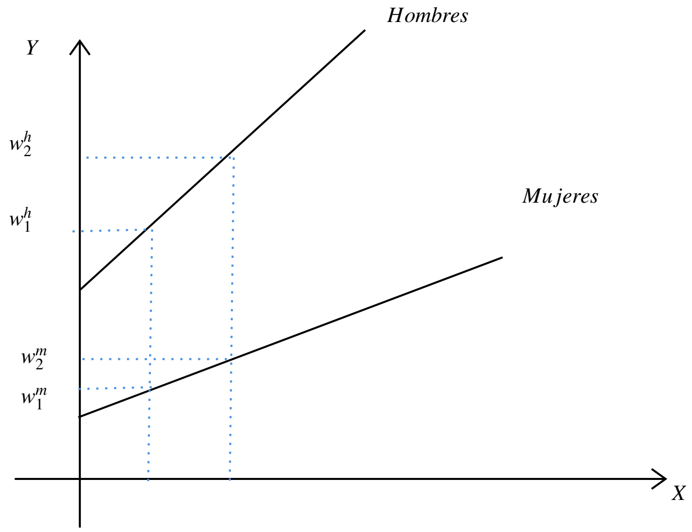
$$\begin{aligned} E[Y_i|D_i=1] &= E[\beta_0]+E[\beta_1 X_i]+E[\beta_2 D_i]+E[\beta_3 D_i X_i]+E[U_i] \\ E[Y_i|D_i=1] &= \beta_0+\beta_1 X_i+\beta_2+\beta_3 X_i \\ E[Y_i|D_i=1] &= (\beta_0+\beta_2)+(\beta_1+\beta_3)X_i \\ & \quad\; Intercepto \qquad Pendiente \end{aligned}$$

Para las mujeres

$$\begin{aligned} E[Y_i|D_i=0] &= E[\beta_0]+E[\beta_1 X_i]+E[\beta_2 D_i]+E[\beta_3 D_i X_i]+E[U_i] \\ E[Y_i|D_i=0] &= \beta_0+\beta_1 X_i \\ E[Y_i|D_i=0] &= \beta_0+\beta_1 X_i \\ & \quad\; Intercepto \qquad Pendiente \end{aligned}$$

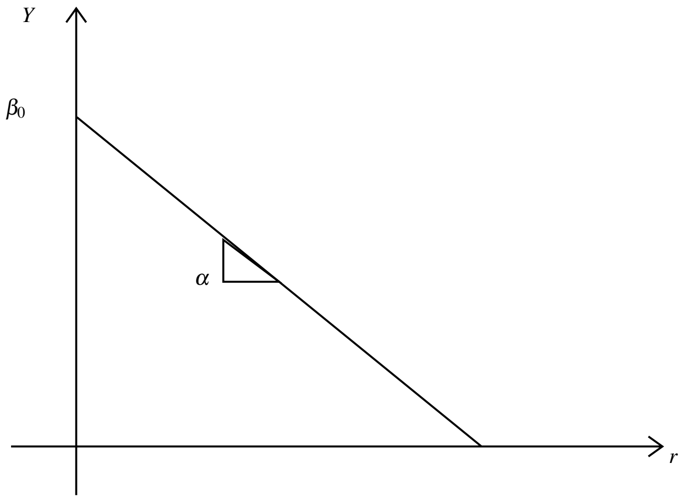


β_2 solo mide la diferencia salarial entre hombres y mujeres cuando la educación es cero.



3. Variables dicotómicas con base en variables categóricas
   En esta parte trabajaremos con un modelo en que la variable explicada es el crédito y los datos son trimestrales.
   Repaso: los tres tipos de datos
   a) Corte transversal: las observaciones son unidades. Por ejemplo: individuos o países.
   b) Series de tiempo: observaciones de momentos en el tiempo, como años o días.
   c) Datos de panel: observaciones son de unidades en diferentes momentos del tiempo. Es una mezcla.

Este es el modelo que vamos a trabajar. Relaciona crédito (c) y tasa de interés (r).



$c_t = \beta_0 + \alpha r_t + U_t$

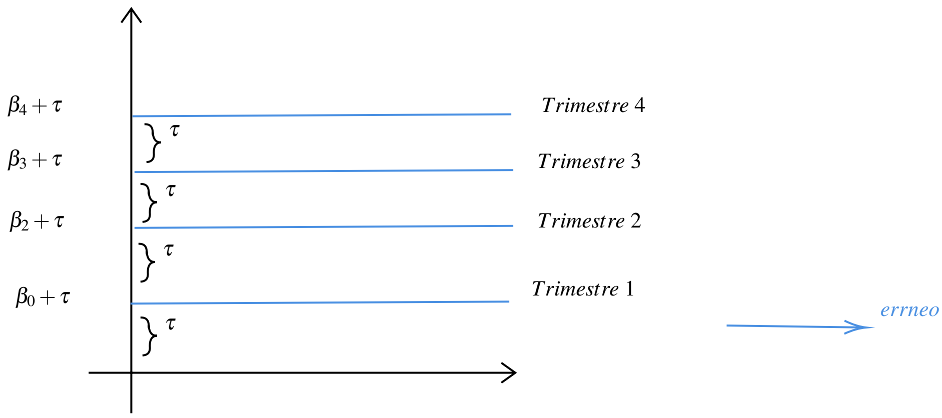
$\rightarrow$ lo estamos planteando como continuo: *spoiler*: esto no se debe hacer así, ya que son datos trimestrales

Modelo con variable categórica tratada como variable continua.

$$T \text{ es una variable categórica que se define como: } T_t = \begin{cases} 1: & \text{trimestre 1} \\ 2: & \text{trimestre 2} \\ 3: & \text{trimestre 3} \\ 4: & \text{trimestre 4} \end{cases}$$

t	C_t	T_t
12-1	281	1
12-2	197	2
12-3	201	3
12-4	210	4
13-1	303	1
13-2	295	2

Modelo relaciona crédito trimestral



$C_t = \beta_0 + \mathscr{T} T_t + U_t$

$E[C_t|T_t = 2] = \beta_0 + 2\mathscr{T}$

$E[C_t|T_t = 3] = \beta_0 + 3\mathscr{T}$

$E[C_t|T_t = 1] = \beta_0 + \mathscr{T}$

$\boxed{\text{Modelo que convierte la variable categórica en variables dicotómicas}}$
$\rightarrow$ Se crean variables dicotómicas con base en la variable $\mathscr{T}$:

2 $D_1 = \begin{cases} 1 & ;trimestre = 1 \\ 0 & ;trimestre \neq 1 \end{cases}$

$D_2 = \begin{cases} 1 & ;trimestre = 2 \\ 0 & ;trimestre \neq 2 \end{cases}$

$D_3 = \begin{cases} 1 & ;trimestre = 3 \\ 0 & ;trimestre \neq 3 \end{cases}$

$D_4 = \begin{cases} 1 & ;trimestre = 4 \\ 0 & ;trimestre \neq 4 \end{cases}$

t	C_t	T_t	D_1	D_2	D_3	D_4
12-1	241	1	1	0	0	0
12-2	301	2	0	1	0	0
12-3	259	3	0	0	1	0
12-4	310	4	0	0	0	1
13-1	297	1	1	0	0	0
13-2	300	2	0	1	0	0
13-3	258	3	0	0	1	0

Planteamiento del modelo:
$\boxed{\text{Modelo con variables dicotómicas e intercepto}}$

$$C_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + U_t$$

Se incluyen solamente 3 de las 4 variables dicotómicas. $\rightarrow$ Si se incluyeran las 4 variables dicotómicas: **se va a generar multicolinealidad**: no se podría obtener los estimadores de MCO Veamos la matriz X:

Unos	D_1	D_2	D_3	D_4	$\sum D_{it}$
1	1	0	0	0	1
1	0	1	0	0	1
1	0	0	1	0	1
1	0	0	0	1	1
1	1	0	0	0	1
1	0	1	0	0	1
1	0	0	1	0	1

¿Por qué las columnas son linealmente dependientes? $\rightarrow$ Sumas de las cuatro variables dicotómicas es igual a la columna de unos
Interpretación:

$$C_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + U_t$$

Calcular el valor esperado de cada trimestre:

- Primer trimestre

$$E[C_t|D_{1t}=1,D_{2t}=0,D_{3t}=0]=\beta_0+\beta_1\cdot 1+\beta_2\cdot 0+\beta_2\cdot 0$$
$$E[C_t|D_{1t}=1,D_{2t}=0,D_{3t}=0]=\beta_0+\beta_1$$

- Segundo trimestre

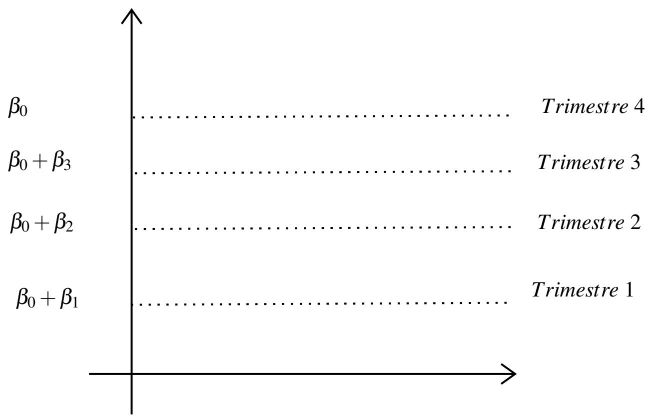
$$E[C_t|D_{1t}=0,D_{2t}=1,D_{3t}=0]=\beta_0+\beta_1\cdot 0+\beta_2\cdot 1+\beta_2\cdot 0$$
$$E[C_t|D_{1t}=1,D_{2t}=0,D_{3t}=0]=\beta_0+\beta_2$$

- Tercer trimestre

$$E[C_t|D_{1t}=0,D_{2t}=0,D_{3t}=1]=\beta_0+\beta_1\cdot 0+\beta_2\cdot 0+\beta_2\cdot 1$$
$$E[C_t|D_{1t}=1,D_{2t}=0,D_{3t}=0]=\beta_0+\beta_3$$

- Cuarto trimestre

$$E[C_t|D_{1t}=0,D_{2t}=0,D_{3t}=0]=\beta_0+\beta_1\cdot 0+\beta_2\cdot 0+\beta_2\cdot 0$$
$$E[C_t|D_{1t}=1,D_{2t}=0,D_{3t}=0]=\beta_0$$



*Regla de oro:
En el modelo en el cual se usa la variable categórica como continua, la diferente entre un trimestre y otro era $\mathscr{T}$.
En el modelo con variables dicotómicas, la diferencia es cualquiera que se calcula con los datos. Incluso sería posible que la línea del trimestre 3 está por debajo del trimestre 2.
<u>En conclusión:</u> variable categórica **no es continua**. Es incorrecto interpretar su coeficiente como una pendiente.

$$\text{Var. categrica} \xrightarrow[\text{se transforma}]{} \text{conjunto de variables dicotmicas}$$

6.3.11 Formas funcionales

¿Qué pasa si la relación entre las variables no es lineal?

$$Y=AL^{\beta}K^{\alpha}$$

$$\Leftrightarrow \ln(Y)=\ln(A)+\beta\ln(L)+\alpha\ln(K)$$

1. Modelo lineal

$$Y_t = \beta_0 + \beta_1 t + U_t$$
$$PIB_t = 4111,6 + 30674t$$

$\rightarrow$ Durante el tiempo de observación, el PIB incrementó en promedio alrededor de 31000 millones.

$$Y_i = \beta_o + \beta_1 x_i + U_i$$
$$\frac{Y_i}{dx} = \frac{\beta_o + \beta_1 X_i + U_i}{dx}$$
$$\frac{dy}{dx} = \beta_1$$

2. Modelo lineal-logarítmico

$$Y_i = \beta_0 + \beta_1 \ln(x_i) + U_i$$
$$\beta_1 = \frac{\Delta \text{ absoluto en Y}}{\Delta \text{ relativo en Y}} = \frac{\Delta \text{ absoluto en Y}}{\Delta \text{ en } \ln(x)}$$
$$\ln(x_i) + 1 = \ln(x_i) + \ln(e) = \ln(ex_i)$$

Otra forma de verlo:

$$\Delta \% x \rightarrow \ln\left[\frac{(100+p)}{100}\right] \cdot \beta_1$$

Por ejemplo: si $\%p = 10\,\%$

$$\text{Si p es muy pequeño } \approx 1\,\% \Rightarrow \ln\left[\frac{(100+p)}{100}\right] \cdot \beta_1 \approx \frac{p}{100} = \frac{1}{100}$$

***Un cambio de 1 % en X es aproximadamente un cambio en Y de** $\frac{1}{100} \cdot \beta_1$

$$\frac{Y_i}{dx} = \frac{\beta_0 + \beta_1 \ln(x_i) + U_i}{dx}$$
$$\frac{dY}{dx} = \frac{\beta_1}{x_i} \rightarrow pendiente$$

3. Modelo logarítmico-logarítmico

$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_i) + U_i$$
$$\mathbf{X}(k, P_x) = \frac{1}{2} \cdot \frac{k}{P_x} = \frac{k}{2} \cdot P_x^{-1}$$

$$\ln(x) = \ln(\frac{k}{2}) + \varepsilon \ln(P_x)$$
$$\ln(x) = \beta_0 + \beta_1 \ln(P_x)$$
$$\ln(x) = \beta_0 + 1,96 \ln(P_x)$$

$1\,\% en P_x \rightarrow$ incremento 1,96 % en X

$$\Delta \log(y) = \frac{y_1 - y_0}{y_0}$$

$$\frac{\ln(y)}{dx} = \frac{\beta_0 + \beta_1 \ln(x) + U}{dx} \quad \frac{1}{y} \cdot \frac{dy}{dx} = \frac{1}{x} \cdot \beta_1 \quad \frac{dy}{dx} = \frac{y}{x} \cdot \beta_1 \quad \varepsilon_{yx} = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{y\beta_1}{x} \cdot \frac{x}{y} = \beta_1$$

4. Modelo logarítmico-lineal

$\ln(x_i) = \beta_0 + \beta_1 x_i + U_i$ $\quad\rightarrow$ (se usa mucho en el PIB; se usa mucho en macroeconometría

$$y_t = (1+r)^t$$

$$ln(Y_t) = \underbrace{ln(y_0)}_{\beta_0} + \underbrace{t \cdot ln(1+r)}_{\beta_1}$$

"x"

$$\beta_1 = \frac{\Delta \text{ relativo en la variable dependiente o regresada}}{\Delta \text{ absoluto en la variable regresora o independiente}}$$

$$\ln(PIB) = 8{,}33 + 0{,}00705t$$
$$\beta_0 = \ln(y_0) = 8{,}33$$
$$\beta_0 = y_0 = e^{8{,}33}$$
$$\beta_1 = 0{,}00705$$

$\%\Delta$ PIB en promedio fue $\beta_1 \cdot 0{,}00705 = 0{,}00705 \cdot 100 = 0{,}705\,\%$

$$\frac{\ln(y)}{dx} = \frac{\beta_0 + \beta_1 x_i + U_i}{dx}$$
$$\frac{1}{y} \cdot \frac{dy}{dx} = \beta_1$$
$$\frac{dy}{dx} = \beta_1 \cdot y$$

*Hoy es una clase de transición de inferencia hacia levantamiento de los supuestos.

6.3.12 Descripción (formulación de procesos)

Proceso: preparar café. Hay que especificar un método → en este caso: método del *coffee maker*.

Suponga que previamente se tiene que el café y el *coffee maker* están disponibles. 2

1. Sacar el filtro viejo
2. Poner el aguar
3. Poner el filtro nuevo

En realidad, el proeso de debería ser más específico

Proceso: preparar dos tazas de café

Previamente: se tiene café y *coffee maker* disponible

2Sacar el filtro viejo Poner dos tazas de agua Poner el filtro nuevo Poner 4 cucharadas de café Cerrar el *coffee maker* Poner el pichel en el *coffee maker* Conectar el *coffee maker* Encender el *coffee maker* Buscar las tazas Cuando el café esté listo, servir las tazas Si se desea, agregar azúcar

***Indispensablemente, hay que conocer el punto de partida y el punto de llegada. Es buno saber qué cosas se pueden cambiar de orden sin alterar el resultado pero otros se tienen que respetar por son insumos necesarios para otros pasos.**

*Antes de hacer códigos, ella escribe un esquema general a seguir y sabiendo qué es insumo de otras cosas. → usar verbos en los pasos: para que siempre sean hacer cosas, pero no poner muchos por pasos (tal vez máximo 3). Sin embargo, uno no siempre puede visualizar cómo se van a ver las cosas. De ahora en adelante se va a pedir descripción de procesos.

Repaso de temas: Proceso generador de datos $\rightarrow$ $\boxed{\text{Datos}}$ $\rightarrow$

1. Estimación MCO
- Función objetivo
- Estimadores
  - Propiedades matemáticas
  - Propiedades estadísticas
    - Especificación
    - ¿Cómo se ve el modelo?
    - ¿Cuáles variables son estadísticamente significativas?
    - ¿Cuál es la magnitud de los coeficientes?
    - ¿Cuál es la forma funcional?
    - Familia pruebas Wald (t)

Y sobre el proceso generador de datos de hacen los siguientes supuestos vinculados a los procesos estadísticos:

- $U_i \sim N(0,\sigma^2)$
- $\hat{\beta}_1 \sim N(\beta, Var(\hat{\beta})$

$W = \dfrac{(R\hat{\beta}-q)'(R\hat{Var}(\hat{\beta}R'))(R\hat{\beta}-q)}{m}$ donde m es el número de restricciones

$W \sim F_{m,n-k}$

Matrices con coeficientes que se eligen de acuerdo a la hipótesis nula $\rightarrow \begin{cases} R \\ q \end{cases}$

Hipótesis conjunta $\begin{cases} H_0 : R\beta = q \\ H_1 : R\beta \neq q \end{cases}$ W siempre va a ser positivo o cero.

$Insumos : \hat{\beta}, \hat{Var}(\hat{\beta}), \{R,q\} \rightarrow H_0$

Modelo:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 Educ + \hat{\beta}_2 D_M + \hat{\beta}_3 D{*}_M \cdot Educ + \hat{U}_i$$

$D_M = \begin{cases} 1 & \text{si es mujer} \\ 0 & \text{si es hombre} \end{cases}$ Hombres y mujeres tienen la misma relación Educación-ingreso 2

$H_0 : \beta_2 = 0 \qquad \beta_3 = 0$

$H_1 : \beta_2 \neq 0 \qquad \beta_3 \neq 0$ *Cuando restricción lineal es Wald

*Estadístico t

$\dfrac{\hat{\beta}_1 - c}{\sqrt{\hat{Var}(\hat{\hat{\beta}}_1)}} \sim t_{N-K}$

Se puede hacer una prueba en cualquiera de las dos colas:

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 > c$

$W = 18,33$

*Cuando se rechaza, al menos una de las restricciones no se cumple, y entonces se podría hacerle t individualmente a cada una.

***Cuando $m = 1$, t es W^2.**

Para esta $H_0(*)$, obtenemos el siguiente modelo restringido:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 Educ + V_i \qquad \textbf{(R)}$$

El modelo restringido solo se puede saber de haber interpuesto una restricción $\boxed{\text{Prueba de significancia conjunta}}$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 w_i \hat{U}_i$$

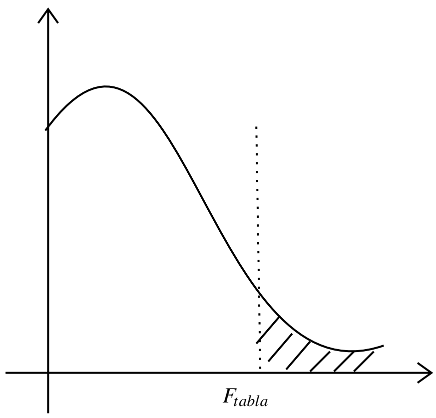
$H_0 : \beta_1 = \beta_2 = 0$ (los coeficientes asignados a las variables independientes)
H_1 : al menos un coeficiente es distinto de cero

$$F : \frac{SEC/(k-1)}{SCR/(N-K)} \sim F_{(K-1)(N-K)}$$

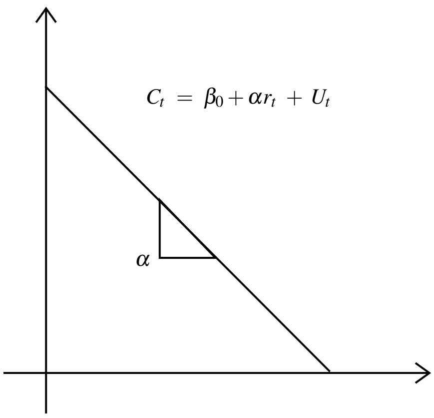
$$\begin{aligned} SCR_R - SCR_{NR} &= \\ \Leftrightarrow STC - SEC_R - (STC - SEC_{NR}) & \\ \Leftrightarrow \cancel{STC} - SEC_R - \cancel{STC} + SEC_{NR} & \\ \Leftrightarrow SEC_{NR} - SEC_R & \\ \Leftrightarrow SEC_{NR} & \end{aligned}$$

$\boxed{\text{Regla de decisión:}}$

- Si $F_{cal} > F_{tabla}$ se rechaza $H_0 \rightarrow$ Hay evidencia estadística de que las variables explicativas, **en su conjunto**, son significativas.
- Si $F_{cal} < F_{tabla}$ no se rechaza $H_0 \rightarrow$ Hay evidencia estadística de que las variables explicativas, **en su conjunto**, no son significativas.

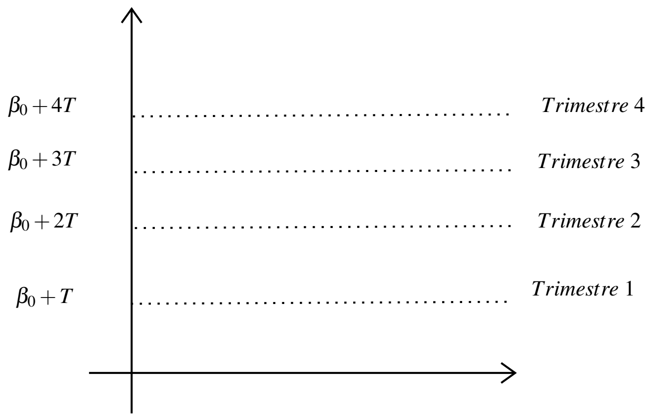


$\boxed{\text{Var. categóricas} \rightarrow \text{Var. dicotómicas}}$



Pensemos en una variable $T \rightarrow$ por trimestre

$$T_t = \begin{cases} 1 & \rightarrow trimestre\ 1 \\ 2 & \rightarrow trimestre\ 2 \\ 3 & \rightarrow trimestre\ 3 \\ 4 & \rightarrow trimestre\ 4 \end{cases}$$



t	C_t	T_t
1-12	250	1
2-12	300	2
3-12	310	3
4-12	270	4
1-12	290	1
2-12	280	2
3-12	300	3

*Pero esto estaría malo.

Se crean variables dicotómicas a partir de la variable T

$$D_1 = \begin{cases} 1 & \text{si trimestre} = 1 \\ 0 & \text{si trimestre} \neq 1 \end{cases}$$

$\vdots$

$$D_j = \begin{cases} 1 & \text{si trimestre} = \text{j} \\ 0 & \text{si trimestre} \neq \text{j} \end{cases}$$

t	C_t	T_t	D_1	D_2	D_3	D_4
1-12	250	1	1	0	0	0
2-12	300	2	0	1	0	0
3-12	310	3	0	0	1	0
4-12	270	4	0	0	0	1
1-13	310	1	1	0	0	0

Unos	D_1	D_2	D_3	D_4
1	1	0	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	0	1
1	1	0	0	0

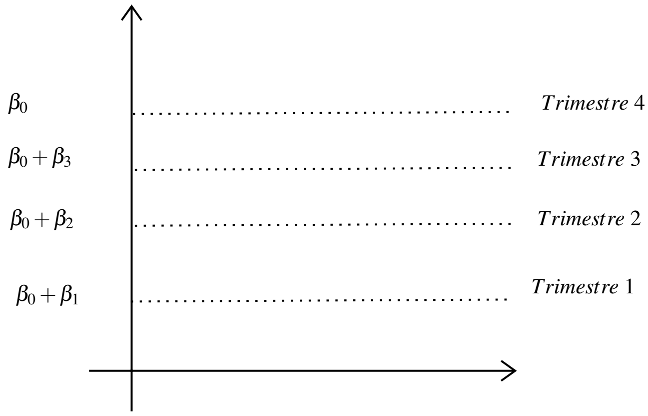
$$\sum D_i = 1$$

$\sum_1 = 1$
$\sum_2 = 1$
$\sum_3 = 1$
$\sum_4 = 1$

*Se incluyen solo 3 de las 4 variables, ya que de lo contrario, habría multicolinealidad perfecta. **Si se incluye el intercepto y las 4 variables dicotómicas → multicolinealidad.**

Interpretación:

$$C_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + U_t \quad \rightarrow Trimestre1$$
$$E[C_t|D_{1t} = 1, D_{2t} = 0, D_{3t=0}] = \beta_0 + \beta_1 \quad \rightarrow Trimestre2$$
$$E[C_t|D_{1t} = 0, D_{2t} = 1, D_{3t=0}] = \beta_0 + \beta_2 \quad \rightarrow Trimestre3$$
$$E[C_t|D_{1t} = 0, D_{2t} = 0, D_{3t=1}] = \beta_0 + \beta_3$$
$$E[C_t|D_{1t} = 0, D_{2t} = 0, D_{3t=0}] = \beta_0 \quad \rightarrow Trimestre4$$



$$*\left\{ B_j \forall j \in \{1,2,3,4\} \right.$$

*Clase importante: vamos a ir quitando uno a uno los supuestos.

Supuestos:

1. $E(U_i) = 0$
2. $Var(U_i) = 0$
3. $Cov(U_i, U_j) = 0 \quad \forall i \neq j$
4. X no es aleatoria
5. Modelo muestral es igual al Modelo poblacional
   $Y_i = \beta_0 + \beta_1 x_i + U_i$
   *No aleatorio
6. U_i es normal o $\hat{\beta}_\ell$ es normal, $\ell = 1, 2, ..., k-1$
   Hacemos supuestos sobre la distribución de Y_i
   $E(Y_i) = E(\beta_0 + \beta_1 x_i + U_i)$
   $E(Y_i) = \beta_0 + \beta_1 x_i + E(U_i)$
   Usando el supuesto 1: $E(U_i) = 0$
   $E(Y_i) = \beta_0 + \beta_1 x_i$

$E(Y_i) = \beta_0 + \beta_1 x_i$
$\hat{E}(Y_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$
$Var(Y_i) = E[(Y_i - E(Y_i)^2)]$
$Var(Y_i) = E[(\beta_0 + \beta_1 x_i + U_i - E(\beta_0 + \beta_1 x_i + U_i)^2)]$
$Var(Y_i) = E(U_i)^2$
Usando supuesto 2 y supuesto 1
$Var(Y_i) = \sigma^2$
$*Cov(U_i, U_j) = 0 \Rightarrow Cov(Y_i, Y_j) = 0$ (no se va a probar ahora)

VI

Evaluación de supuestos de MCO

7. Heterocedasticidad

Todos los otros supuestos no cambian, excepto 2: $Var(U_i) = \sigma^2$. Esto implica: $Var(Y_i) = \sigma^2$.

Comparemos la intuición con un par de ejemplos:

Y_i: nota de curso, X_i: horas de estudio 2

- Homocedasticidad
  - $Var(U_i) = \sigma^2$
  - $Var(Y_i) = \sigma^2$
- Heterocedasticidad
  - $Var(Y_i) = \sigma_i^2$

    *Es más realista

$Var(Y_i), Var(U_i) \rightarrow$ Determinantes: 2

- Conocimiento previo (-)
- Efectividad horas de estudio (-)
- Talento (-)
- Malas condiciones del ambiente de estudio (+)

Ejemplo económico:

Y_t : PIB	X_t : $tipo\ de\ cambio$
$Var(PIB_t)$	$E(PIB_t)$
$\left.\begin{array}{l} -Guerra \\ -Pandemia \\ -Crisis\ financiera\ internacional \\ -Clima \\ -Patrones\ de\ comercio\ internacional \\ *Tecnologa \end{array}\right\} U_t$	
$*Costos\ de\ transporte$	

*Homocedasticidad es un supuesto muy fuerte. Es más intuitivo pensar en modelos con heterocedasticidad. A partir de ahí se puede pensar en cosas que afectan la varianza de Y_t.

7.0.0.1 Propiedades estadísticas

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

- Insesgamiento usa supuestos 1,4 y 5, entonces sí se cumple insesgamiento
- Varianza mínima $\rightarrow$ revisar la prueba e identificar los posibles puntos (o líneas) donde habrían cambios.

Con homocedasticidad: $Var(U_i) = \sigma^2 I_N$.
Con heterocedasticidad: $Var(U_i) \neq \sigma^2 I_N$.

7.0.0.2 Otro método: mínimos cuadrados generalizados (MCG)

- Paso 0: establecer que se quiere trabajar con heterocedasticidad; específicamente el supuesto $Var(U_i) = \sigma_i^2$, el cual, junto con el supuesto 1, implica que $E(U_i) = \sigma_i^2$.
- Paso 1: transformar el modelo

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + U_i \\ \Leftrightarrow \frac{Y_i}{\sigma_i} &= \frac{\beta_0}{\sigma_i} + \frac{\beta_1 X_i}{\sigma_i} + \frac{U_i}{\sigma_i} \\ *Definimos : Y_i^* = \frac{Y_i}{\sigma_i}, X_{0i}^* = \frac{1}{\sigma_i}, X_i^* &= \frac{X_i}{\sigma_i}, U_i^* = \frac{U_i}{\sigma_i} \end{aligned}$$

  Entonces el modelo es:

$$Y_i^* = \beta_0^{MGC} X_{0i}^* + \beta_1^{MGC} X_i^* + U_i^* \qquad (MP) \tag{7.1}$$

- Paso 2: obtener los estimadores
  - $\hat{\beta}_0^{MCG}, \hat{\beta}_1^{MCG}$, estimado con MCO
- $Y_i^* = \hat{\beta}_0^{MGC} X_{0i}^* + \hat{\beta}_1^{MGC} X_i^* + \hat{U}_i^* \qquad (MP)$

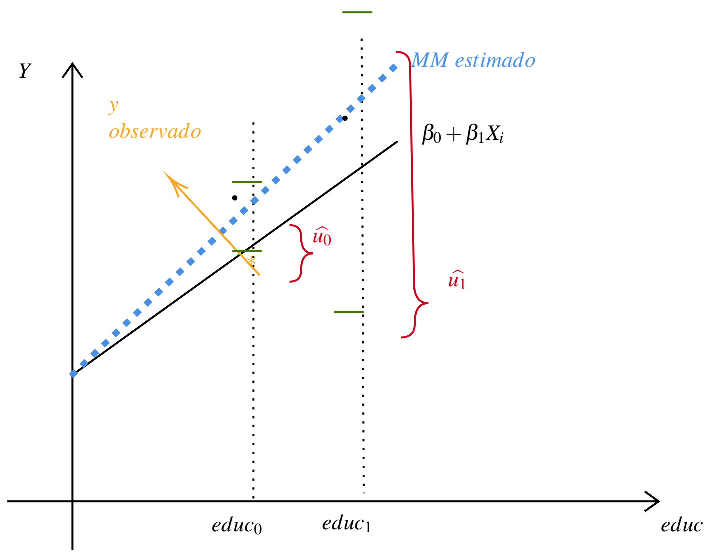
¿Por qué esto funciona? $\rightarrow$ Tómese el modelo poblacional

$$\begin{aligned} Var(U_i^*) &= Var\left(\frac{U_i}{\sigma_i}\right) \\ Var(U_i^*) &= \frac{1}{\sigma_i^2} Var(U_i) \\ Var(U_i^*) &= \frac{\sigma_i^2}{\sigma_i^2} \\ Var(U_i^*) &= 1 \end{aligned}$$

*A partir de aquí, se usa la presentación. Heterocedasticidad $\rightarrow$ Supuesto (2) modificado

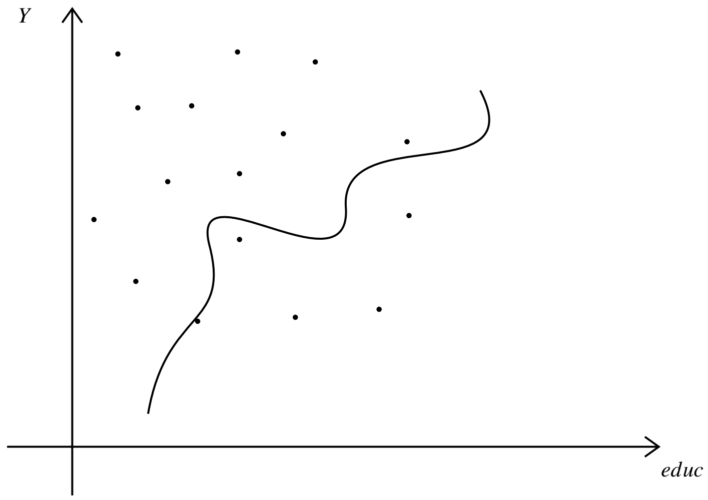
$Var(U_i) = \sigma_i^2 \neq \sigma^2 \rightarrow$ ¿Cómo hacer esto operativo? ¿Cómo manejarlo?

$\rightarrow \sigma_i^2$ relacionado con $X's$ (datos)



Posibles argumentos

1. + educación → más varianza → esperamos que $\hat{U}_0^2 < \hat{U}_1^2$
2. + educación → menos varianza



7.0.1 Paréntesis: prueba de varianza mínima

$Var(\beta) = \sigma^2 CC' \qquad Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

Para obtener estas varianzas, usamos supuestos 1,2 y 3 $\Rightarrow E(UU') = \sigma^2 I$ Con heterocedasticidad:

$$\begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} = \Omega \neq \sigma^2 I$$

$$Var(\beta) = CE(UU')C' = C\Omega C'$$

$$Var(\beta) = (X'X)^{-1} X' E(UU') X (X'X)^{-1} = (X'X)^{-1} X \Omega X (X'X)^{-1}$$

7.0.2 Mínimos Cuadrados Generales Factibles

- Paso 0: adivinar σ_i^2

- Paso 1: transformar los datos

$$Y_i^* = \frac{Y_i}{\sigma_i} \qquad X_{oi}^* = \frac{1}{\sigma_i} \qquad X_i^* = \frac{X_i}{\sigma_i}$$

- Paso 2: estimar con MCO, $\hat{\beta}_0^{MCG}$, $\hat{\beta}_1^{MCG}$, en la ecuación $Y_i^* = \hat{\beta}_0^{MCG} X_{0i}^* + \hat{\beta}_1^{MCG} X_i^* + \hat{U}_i^*$

$\boxed{\text{Alternativa 3:}}$

$$crdito_t = \beta_0 + \beta_1 IMAE_t + U_t$$

Medidos por mes:

$$Var(crdito_t) = \frac{\sum (crdito_{tj} - \overline{crdito_t})^2}{\text{días } mes_j} \qquad j = 1, 2, .., 12 meses$$

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{U}_i$$

Y_i: notas macroeconometría $\quad$ X_i: horas estudio macroeconometría
Otra base: notas de la persona i en todos los cursos que ha llevado

7.0.3 Prueba Breusch- Pagan-Godfrey

- Paso 1: estimar en MCO el modelo de interés

  $$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 W_i + \hat{U}_i$$

  y obtener $\hat{U}_i$.
- Paso 2: calcular una nueva variable:

  $$p_i = \frac{\hat{U}_i^2}{(\sum \hat{U}_i^2 / N)} \tag{7.2}$$

- Paso 3: estimar con MCO un modelo

  $$p_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_1 + \hat{\alpha}_2 Z_2 + \hat{\alpha}_3 Z_3 + \hat{V}_i \tag{7.3}$$

  y obtener SEC.
- Paso 4: Prueba de hipótesis
  - 4.1: plantear las hipótesis
    - $$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \rightarrow \text{implica homocedasticidad (no hay evidencia de patrón)}$$
    - H$_1$: al menos uno es distinto de cero
  - 4.2: calcular el estadístico

    $$\frac{1}{2} SEC \sim \chi^2_{m-1} \tag{7.4}$$

    donde m es la cantidad de α's.
  - 4.3: seleccionar α y encontrar χ^2 de tabla: $\chi^2_{m-1,\alpha,tabla}$.
  - 4.4: conclusiones
    - Si $\chi^2_{m-1} \leq \chi^2_{m-1,\alpha,tabla}$, no se rechaza H_0, no se rechaza homocedasticidad.
    - Si $\chi^2_{m-1} > \chi^2_{m-1,\alpha,tabla}$, se rechaza H_0, se rechaza que los errores sean homocedasticos.

7.0.4 Prueba de White

- Paso 1: estimar con MCO el modelo de interés

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 W_i + \hat{U}_i \qquad (MM)$$

  y obtener $\hat{U}_i$.
- Paso 2: calcular una nueva variable:

$$P_i = \frac{\hat{U}_i^2}{(\frac{\Sigma \hat{U}_i^2}{N})}$$

- Paso 3: estimar con MCO un modelo

$$\hat{U}_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 W_i + \hat{\alpha}_3 X_i^2 + \hat{\alpha}_4 W_i^2 + \hat{\alpha}_5 X_i W_i + \hat{V}_i$$

  obtener R^2.
- Paso 4:
  - 4.1: plantear las hipótesis.
    - $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_{m-1} = 0 \rightarrow$ implica homocedasticidad (no hay evidencia de patrón)
  - H_1 : al menos 1 es distinto de cero
  - 4.2: calcular el estadístico

$$NR^2 \sim \chi^2_{m-1}$$

    donde m es la cantidad de α's.
  - 4.3: seleccionar α y encontrar χ de tabla: $\chi^2_{m-1,\alpha,tabla}$
  - 4.4: conclusiones
    - Si $\chi^2_{m-1} \leq \chi^2_{m-1,\alpha,tabla}$, no se rechaza H_0, no se rechaza homocedasticidad.
    - Si $\chi^2_{m-1} > \chi^2_{m-1,\alpha,tabla}$, se rechaza H_0, se rechaza que los errores sean homocedasticos.

7.0.5 Supuestos: $Cov(U_i, U_j) \neq 0$

Este supuesto significa, en su versión original, sin modificar (o sea, igual a 0) "hay ausencia de correlación entre los errores.Reemplazar $Cov(U_i, U_j) = 0$

En las series de tiempo (autocorrelación):

$$Cov(U_t, U_{t-s}) = 0 \Rightarrow Cov(Y_t, Y_{t-s}) = 0$$

$$Cov(U_t, U_{t-s}) \neq 0 \Rightarrow Cov(Y_t, Y_{t-s}) \neq 0$$

Ejemplo de autocorrelación de orden q $AR(q)$

$$\begin{aligned} & & q = 1 \\ U_t &= \rho U_{t-1} + \mathscr{E}_t & (AR1) \\ & & q = 2 \\ U_t &= \rho U_{t-2} + \rho U_{t-1} + \mathscr{E}_t & (AR2) \end{aligned}$$

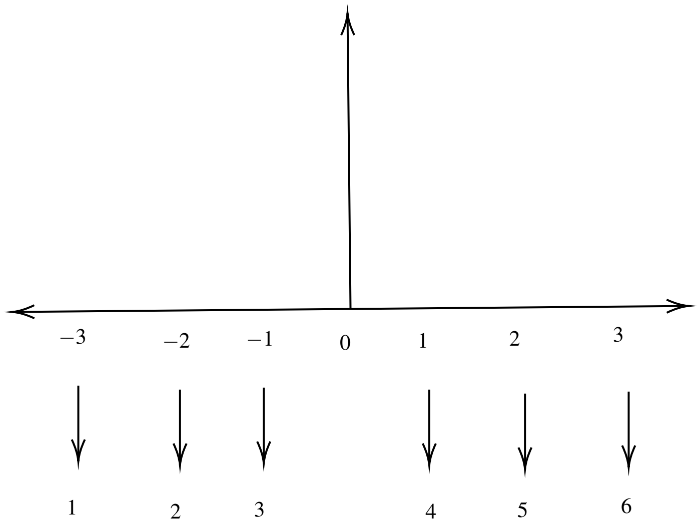
Estos dos anteriores son modelos poblacionales.

$\boxed{AR(1) \text{ en acción:}}$

$U_t = \rho U_{t-1} + \mathscr{E}_t$

Dos errores

$\mathscr{E}_t$: tiene o cumple todos los supuestos.



t	E_t	ρU_{t-1}	U_t
0	1	-	1
1	-1	0.7 $\cdot$1	-0.3
2	-2	0.7 $\cdot -0{,}3$	-2.21

Hoy se dan los parciales al final de la clase. Este curso no es de mucha práctica. Hay cosas que se pueden hacer de varias maneras y otras que solo de una forma. Nos va a dar unas prácticas.

Está pensando en 3 actividades para subir la nota: 2 prácticas y un quiz de 40 minutos. Así también se quitan cosas para el examen final. Lo primero es volver a resolver la parte 3 del examen.

7.0.6 Pruebas formales de heterocedasticidad

- Glaeser
- Park
- Goldfeld-Quant $\rightarrow$ Estudiar por cuenta propia; probablemente pregunte algo sobre esto
- Prueba de correlación de Spearman

8. Autocorrelación

$\boxed{\text{Ejemplo: AR(1)}}$

$$Y_t = \beta_0 + \beta_1 x_t \tag{8.1}$$

$$U_t = \rho U_{t-1} + \varepsilon_t \tag{8.2}$$

Solo esto refleja el pasado en el período t Antes:

1. $E(U_t) = 0$
2. $Var(U_t) = \sigma^2$
3. $Cov(U_t, U_{t-s}) = 0, \forall s$

$\boxed{\text{Supuestos con } AR(1)\text{:}}$ especíﬁcamente para AR(1)

- $E(u_t) = 0$
- $Var(u_t) = \sigma_u^2$
- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma_\varepsilon^2$
- $Cov(\varepsilon_t, \varepsilon_{t-s}) = 0, \quad \forall S$
- $Cov(\varepsilon_t, U_{t-s}) = 0, \quad \forall s$

¿Cómo es la matriz de Varianzas-Covarianzas de los errores?

$$Var(U_t) = E[(U_t - E(U_t))^2]$$
Usando el supuesto $\mathrm{E}(\mathrm{u}_t) = 0$
$$Var(U_t) = [E(U_t)]^2$$
Reemplazando y evaluando
$$Var(U_t) = E[(\rho U_{t-1} + \varepsilon_t)^2]$$
Calculando la expresión al cuadrado
$$Var(U_t) = E[\rho^2 U_{t-1}^2 + 2\rho U_{t-1}\varepsilon_t + \varepsilon_t^2]$$
$$Var(U_t) = \rho^2 E(U_{t-1}^2) + 2\rho E(U_{t-1}\varepsilon_t) + E(\varepsilon_t^2)$$
Para $E(\varepsilon_t^2)$, usando $Var(\varepsilon_t) = \sigma_\varepsilon^2$
Usando la definición de varianza
$$Var(\varepsilon_t) = [(\varepsilon_t - E(\varepsilon_t))(\varepsilon_t - E(\varepsilon_t))] = \sigma_\varepsilon^2$$
$E(\varepsilon_t) = 0$ y $Var(\varepsilon_t) = \sigma_\varepsilon^2$ implican que
$$\boxed{\mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^2}$$

Ahora vemos: $E(u_{t-1}, \varepsilon_t)$. Usando $Cov(\varepsilon_t, U_{t-s}) =$

$\mathrm{Cov}(\varepsilon_t, U_{t-s}) =$, $E(u_t) = 0$ y $E(\varepsilon_t) = 0$ implican
$\mathrm{E}(\varepsilon_t, u_{t-1}) = 0$

General:

$$\boxed{\mathrm{E}(\varepsilon_t, u_{t-s}) = 0, \forall s}$$
$Var(u_t) = \sigma_u^2$ y $E(u_t) = 0$ implican
$$\boxed{\mathrm{Var}(\mathrm{u}_t) = E(u_t^2) = \sigma_u^2}$$
$Var(\varepsilon_t) = \sigma_\varepsilon^2$ y $E(\varepsilon_t) = 0$ implican
$$\boxed{\mathrm{Var}(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2}$$
$Cov(\varepsilon_t, \varepsilon_{t-s}) = 0, \forall S$ y $E(\varepsilon_t) = 0$ implican
$$\boxed{\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-s}) = E(\varepsilon_t \varepsilon_{t-s}) = 0, \forall s}$$
$$Var(u_t) = \rho^2 E(u_{t-1}^2) + 2\rho E(u_{t-1}\varepsilon_t) + E(\varepsilon_t)^2$$

Usando $Var(u_t) = \sigma_u^2$ implica que $Var(u_t) = \sigma_u^2$

$$\sigma_u^2 = \rho^2 E(u_{t-1}^2) + 2\rho E(u_{t-1}\varepsilon_t) + E(\varepsilon_t)^2$$
Usando $Var(u_t) = E(u_t^2) = \sigma_u^2$
$$\sigma_u^2 = \rho^2 \sigma_u^2 + 2\rho E(u_{t-1}\varepsilon_t) + E(\varepsilon_t)^2$$
$$\sigma_u^2(1-\rho^2) = 2\rho E(u_{t-1}\varepsilon_t) + E(\varepsilon_t)^2$$
Usando $E(\varepsilon_t, u_{t-s}) = 0, \forall s$ y $Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2$
$$\sigma_u^2(1-\rho^2) = 0 + \sigma_\varepsilon^2$$
$$\boxed{\sigma_u^2 = \frac{\sigma_\varepsilon^2}{(1-\rho^2)}}$$

$$Cov(U_t, U_{t-1}) = E[u_t, u_{t-1}]$$

Reemplazando:

$$Cov(u_t, u_{t-1}) = E[(\rho u_{t-1} + \varepsilon_t)u_{t-1}]$$
$$Cov(u_t, u_{t-1}) = \rho E(u_{t-1}^2) + E(\varepsilon_t u_{t-1})$$

Usando $E(\varepsilon_t, u_{t-s}) = 0, \forall s$

$$Cov(u_t, u_{t-1}) = \rho E(u_{t-1}^2)$$

Usando $E(u_t, u_{t-1}^2) = Var(u_t) = \frac{\sigma_\varepsilon^2}{(1-\rho^2)}$

$$Cov(u_t, u_{t-1}) = \frac{\rho\sigma_\varepsilon^2}{(1-\rho^2)}$$

Se debe probar:

$$Cov(u_t, u_{t-s}) = \frac{\rho^s\sigma_\varepsilon^2}{(1-\rho^2)}$$

$$\begin{bmatrix}
\frac{\sigma_\varepsilon^2}{(1-\rho^2)} & \frac{\rho\sigma_\varepsilon^2}{(1-\rho^2)} & \frac{\rho^2\sigma_\varepsilon^2}{(1-\rho^2)} & \frac{\rho^3\sigma_\varepsilon^2}{(1-\rho^2)} & \frac{\rho^4\sigma_\varepsilon^2}{(1-\rho^2)} & \cdots & \frac{\rho^{N-1}\sigma_\varepsilon^2}{(1-\rho^2)} \\
\frac{\rho\sigma_\varepsilon^2}{(1-\rho^2)} & \frac{\sigma_\varepsilon^2}{(1-\rho^2)} & & & & & \\
\frac{\rho^2\sigma_\varepsilon^2}{(1-\rho^2)} & & \frac{\sigma_\varepsilon^2}{(1-\rho^2)} & & & & \\
\frac{\rho^3\sigma_\varepsilon^2}{(1-\rho^2)} & & & \frac{\sigma_\varepsilon^2}{(1-\rho^2)} & & & \\
\frac{\rho^4\sigma_\varepsilon^2}{(1-\rho^2)} & & & & \frac{\sigma_\varepsilon^2}{(1-\rho^2)} & & \\
\vdots & & & & & \ddots & \\
\frac{\rho^{N-1}\sigma_\varepsilon^2}{(1-\rho^2)} & & & & & & \frac{\sigma_\varepsilon^2}{(1-\rho^2)}
\end{bmatrix}$$

8.0.1 Mínimos Cuadrados Generalizados

$$Y_t = \beta_0 + \beta_1 x_t + u_t \tag{8.3}$$

$$U_t = \rho u_{t-1} + \varepsilon_t \tag{8.4}$$

Así, se tiene

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho u_{t-1}$$
$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - x_{t-1}) + u_t - \rho u_{t-1}$$

Reemplazar en la segunda ecuación, que implica: $\varepsilon_t = u_t - \rho u_{t-1}$

$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - x_{t-1}) + \varepsilon_t$$

ε_t cumple $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2$ y $Cov(\varepsilon_t, \varepsilon_{t-s}) = 0, \forall S =$ a MCO

1. Paso: transformar los datos

$$Y_t^* = Y_t - \rho Y_{t-1}$$
$$X_t^* = X_t - \rho X_{t-1}$$

2. Paso: estimar con MCO y $\lambda = \beta_0(1-\rho)$

$$Y_t^* = \hat{\lambda} + \hat{\beta}_1 x_t^* + \hat{\varepsilon}_t \tag{8.5}$$

Hoy vamos a ver tres cosas:
- Mínimos Cuadrados Generalizados factibles a mano
- Mínimos Cuadrados Generalizados con matrices
- Sesgo de variable omitida

Vamos a programar una iteración

MG AR(1)
- Paso 1: transformar los datos

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$x_t^* = x_t - \rho x_{t-1} \tag{8.6}$$

- Paso 2: estimar con MCO

$$Y_t^* = \hat{\lambda} + \hat{\beta}_1 x_t^* + \hat{\varepsilon}_t \tag{8.7}$$

$$\lambda = (1-\rho)\beta_0$$

$$Y_t - \rho Y_{t-1} = \underbrace{(1-\rho)\beta_0}_{Intercepto} + \beta_1(x_t - x_{t-1})$$

> <u>MCG Factibles PRO</u> (vamos a estimar ρ)
> - Paso 0.1: estimar con MCO
>
> $$Y_t = \hat{\beta}_0^0 + \hat{\beta}_1^0 x_t + \hat{u}_t^0$$
>
> obtener $\hat{u}_t^0$.
> - Paso 0.2: estimar con MCO
>
> $$\hat{u}_t = \hat{\rho}^0 u_{t-1}^0 + \varepsilon_t^0$$
>
> se estimar $\hat{\rho}^0$.

- Paso 1: transformar los datos

$$Y_t^* = Y_t - \hat{\rho}^0 Y_{t-1} \text{ y } x_t^* = x_t - \hat{\rho}^0 x_{t-1}$$

- Paso 2: estimar con MCO

$$Y_t^* = \hat{\lambda} + \hat{\beta}_1^1 x_t^* + \hat{\varepsilon}_t$$

- Paso 3: calcular

$$\hat{\beta}_0^1 = \frac{\hat{\lambda}}{(1-\hat{\rho}^0)}$$

- Paso 4: calcular $\hat{u}_t^1$

$$\hat{u}_t^1 = Y_t - \hat{\beta}_0^1 - \hat{\beta}_1^1 x_t$$

- Paso 5: estimar con MCO

$$\hat{u}_t^1 = \hat{\rho}^1 \hat{u}_{t-1}^1 + \hat{\varepsilon}_1^1$$

obtener $\hat{\rho}^1$.
- Paso 6: si $\hat{\rho}^0 = \hat{\rho}^1$ detener proceso y los estimadores de MCG son $\hat{\beta}_1^1$ y $\hat{\beta}_0^1$

8.0.2 MCG con matrices

Autocorrelación o heterocedasticidad:

$$Var(U) = E(UU') \neq \sigma^2 I$$
$$Var(U) = \Omega \neq \sigma^2 I$$

- MCO no son de "varianza mínima"(eficientes)
- MCG sí son de "varianza mínima"(eficientes)

****De aquí en adelante se usa la presentación**

8.0.3 Método MCO 2 Etapas

Se plantea una regresión de la variable endógena o instrumentalizada X, en términos de la variable omitida W y la exógena Z. Las dos cumplen el supuesto de exogeneidad. Es decir, X es una combinación lineal de variables endógenas, y así se va a obtener una mayor correlación entre las variables.

Entonces si se hace esta regresión, X_i sería ahora una variable instrumental para el modelo muestral, puesto que al ser una combinación lineal de variables exógenas, sería a su vez exógena.

8.0.4 X no aleatorio

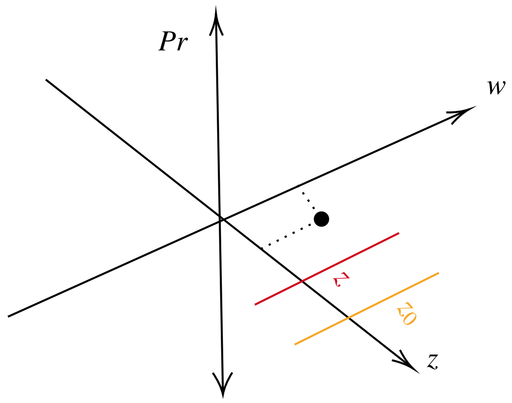
Hemos venido cambiando uno a uno los supuestos. Dentro de todos los supuestos, siempre hemos asumido que X no es una variable aleatoria. En el primer ejemplo de las notas del curso, la aleatoriedad de Y venía solo de u, pero tendría más sentido pensar en que X sí es aleatorio.
Entonces, cuando pasamos a suponer que X sí es aleatorio, dicha aleatoriedad podría provenirde varias causas distintas. En los experimentos sociales, en los experimentos casi no hay diseño posible que permita controlar las variables para que no sean aleatorias, a diferencia de los experimentos naturales.
X no aleatorias tiene más sentido en experimentos de laborario.

8.0.4.1 Paréntesis estadístico (Apéndice Greene)

Dos variables aleatorias W y Z con distribución conjunta:

$$Prob(W, Z) = f(W, Z) \tag{8.8}$$

Cuando pensamos en dos variables aleatorias, cualquier combinación de W y Z tiene asociada una probabilidad conjunto, generando así una superficie en el espacio.



A lo largo del curso, hemos estudiado dos características de la distribución de los beta's estimados. La pregunta ahora es cómo se calculan la varianza y la esperanza esta distribución con dos variables aleatorias. Para estos efectos ocupamos dos definiciones:

- Ley de expectativas iteradas
- Descomposición de la varianza

8.0.4.2 Ley de expectativas iteradas

Puedo obtener la esperanza de W en dos pasos.

1. Se fija z_0 y se integra sobre la línea sin cambiar z, que estaría fijo.
2. Luego integro hacia el otro lado, se forma un "piso con rayitas".

$$E(W) = E_Z(Var(W|Z)) \tag{8.9}$$

Descomposición de varianza

$$Var(W) = E_Z(Var(W|Z)) + Var_Z(E(W|Z)) \tag{8.10}$$

También vamos a ver algo sobre consistencia, y también definir algunos conceptos.

$\boxed{\text{Convergencia en probabilidad}}$

W_N una variable aleatoria. Se dice W_N converge en probabilidad a c si:

$$\lim_{n\to\infty} Pr(|W_n - c| > \varepsilon) = 0 \qquad \forall\varepsilon > 0$$

Se usa la siguiente notación:

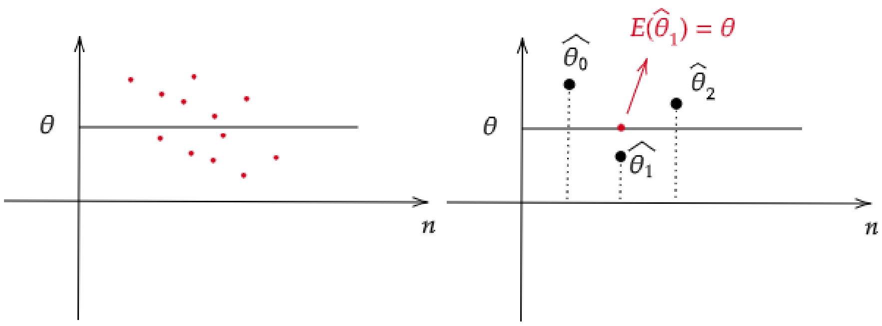
$$p\lim W_n = c$$

8.0.4.3 Consistencia

Un estimador $\hat{\theta}_n$ del parámetro poblacional θ es consistente si

$$\lim_{n\to\infty} Pr(|\hat{\theta}_n - \theta| > \varepsilon) = 0 \qquad \forall\varepsilon \tag{8.11}$$

$p\lim \hat{\theta}_n = \theta$



Cuando el n es pequeño, puede ser que $\hat{\theta}_n$ sea diferente a θ, pero cuando n tiende a infinita, el estimador converge al parámetro poblacional. Puede que no lo haga necesariamente en muestras pequeñas. (Aquí recordar lo que había dicho Sebastián de que muchos de los supuestos aplicaban a muestras grandes).

$$\text{insesgamiento} \Rightarrow consistencia$$
$$\text{insesgamiento} \not\Leftarrow \text{consistencia}$$

Ahora vamos a ver la prueba de consistencia.

8.0.5 Consistencia: Greene 5.2.1

No vamos a hacer propiamente la prueba de consistencia porque es una prueba que puede tomar hasta dos clases. Lo que vamos a hacer es hablar de la intución.

Cuando hacemos insesgamiento, llegamos a la siguiente expresión, y le calculamos el límite en probabilidad. Lo que se busca, es lo mismo que en la prueba de insesgamiento, es que el límite en probabilidad de la segunda parte, sea 0, pero hay que involucrar N en la expresión.

$$\begin{aligned}
\hat{\beta} &= \beta + (X'X)^{-1}X'U \\
p\lim\hat{\beta} &= p\lim(\beta + (X'X)^{-1}X'U) \\
p\lim\hat{\beta} &= p\lim(\beta) + p\lim[(X'X)^{-1}X'U)] \\
p\lim\hat{\beta} &= \beta + p\lim[(X'X)^{-1}X'U)] \\
p\lim\hat{\beta} &= \beta + p\lim\left[\left(\frac{X'X}{N}\right)^{-1}\left(\frac{X'U}{N}\right)\right]
\end{aligned}$$

$\boxed{\text{Intuición}}$

*Independiente X del error u (si X es aleatorio)

¿Converge en probabilidad?

$\frac{X'X}{N} \longmapsto Q$ invertible, finita positiva definida

$\frac{X'U}{N} \longmapsto 0$

Entonces la consistencia requiere de dos cosas:

- $\frac{X'X}{N} \longmapsto Q$
- $\frac{X'U}{N} \longmapsto 0$

8.0.6 X aleatoria

- Supuesto escenario 1
  1. $E(U_i) = 0$
  2. $Var(U_i) = \sigma^2$
  3. $Cov(U_i, U_j) = 0 \quad \forall i \neq j$
  4. MM = MP
  5. X no aleatoria
- Supuesto X aleatoria
  1. $E(U_i|X_i) = 0$
  2. $Var(U_i|X_i) = \sigma^2$
  3. $Cov(U_i, U_j|X_i, X_j) = 0 \quad \forall i \neq j$
  4. MM = MP
  5. X_i aleatoria independiente de U_i

$$\begin{aligned}
Y &= \beta_0 + \beta_1 x_i + u_i \\
&\quad\quad f(x_i, u_i)
\end{aligned}$$

1. X_i independiente de $U_i \Rightarrow Cov(x_i, u_i) = 0$
2. $Cov(X_i, U_i) \neq 0 \rightarrow X_i)$ es endógena

8.0.6.1 Insesgamiento (con X aleatoria e independiente de U_i

$$\hat{\beta} = \beta + (X'X)^{-1}X'U$$
$$E_U(\hat{\beta}|X) = E(\beta|X) + E[(X'X)^{-1}X'U|X]$$
β es un parámetro poblacional
$$E_U(\hat{\beta}|X) = \beta + E[(X'X)^{-1}X'U|X]$$

Debido a la independencia $(X'X)^{-1}X'$ (fija en muestreos repetidos) se puede extraer de la esperanza

$$E_U(\hat{\beta}|X) = \beta + (X'X)^{-1}X'E[U|X]$$

Supuesto 1: $E(U_i|X_i) = 0$

$$E_U(\hat{\beta}|X) = \beta \tag{8.12}$$

***Ley de expectativas iteradas**

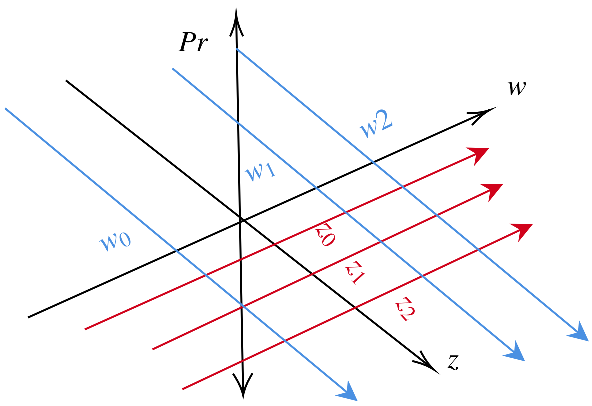
$$\boxed{\text{E(W) = E}_W(Var(W|Z))}$$

$$E(\hat{\beta}) = E_X[E(\beta|X)]$$

Reemplazando la ecuación anterior

$$E(\hat{\beta}) = E_X[\beta]$$
$$E(\hat{\beta}) = \beta$$

Hemos probado insesgamiento



Bibliografía