Econometría

Conceptos básicos (código de R de Edward Rubin)

Ana María Díaz

25 July 2023

Bienvenidos

Poblcaión *vs.* muestra

Modelos y notación

Escribimos el proceso genrador de datos (o modelo poblacional):

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

y nuestro modelo de regresión estimado basado en la muestra como

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

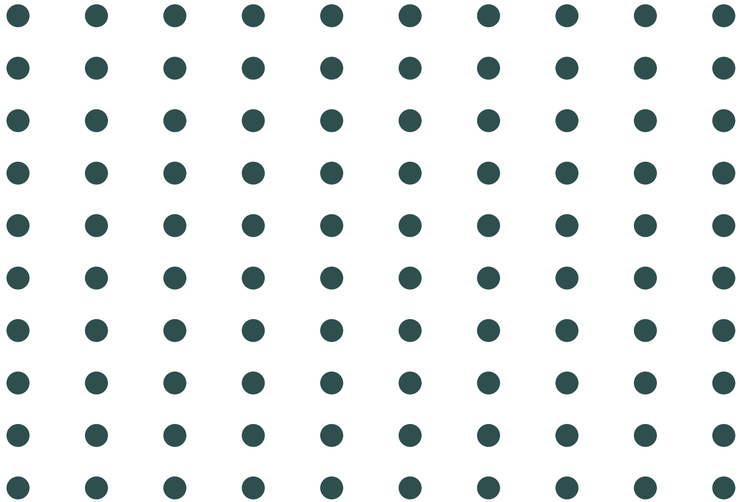
Un modelo de regresión estimado produce estimaciones para cada observación:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

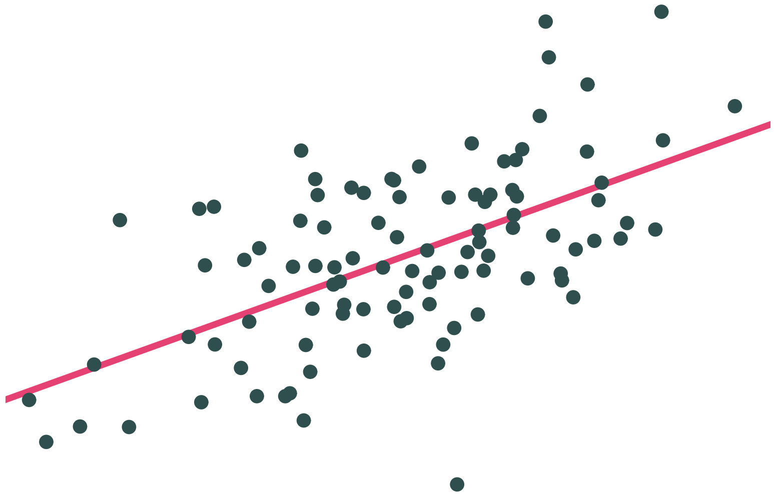
lo cual nos da la línea de *mejor ajuste* a través de nuestro conjunto de datos.

Población vs. muestra

Pregunta: ¿Por qué nos importa la *población vs. muestra*?



Población



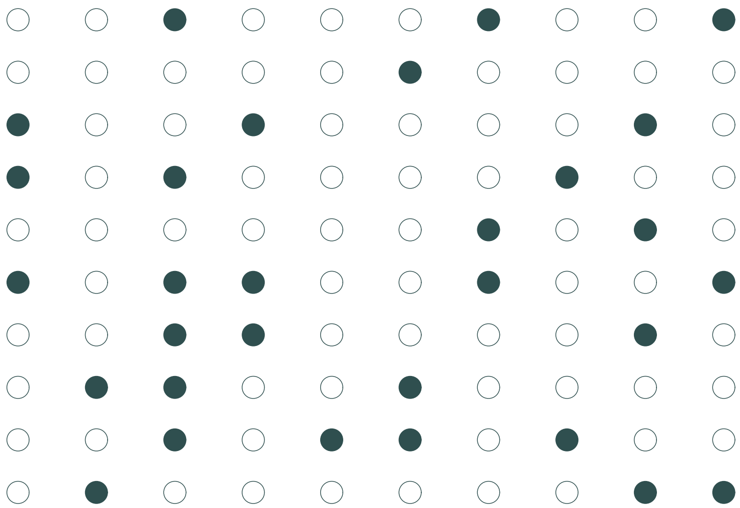
Relación Poblacional

$$y_i = 2.53 + 0.57x_i + u_i$$

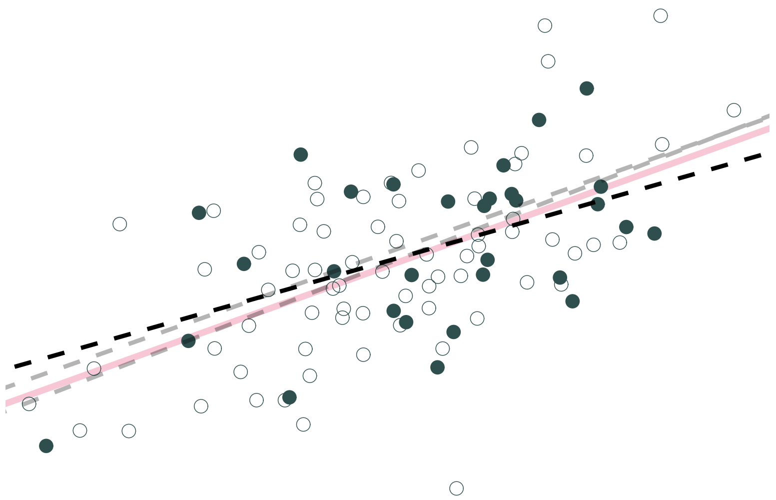
$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Población vs. muestra

Pregunta: ¿Por qué nos importa la *población vs. muestra*?



Sample 3: 30 individuos aleatorios



PGD Modelo Poblacional

$$y_i = 2.53 + 0.57 x_i + u_i$$

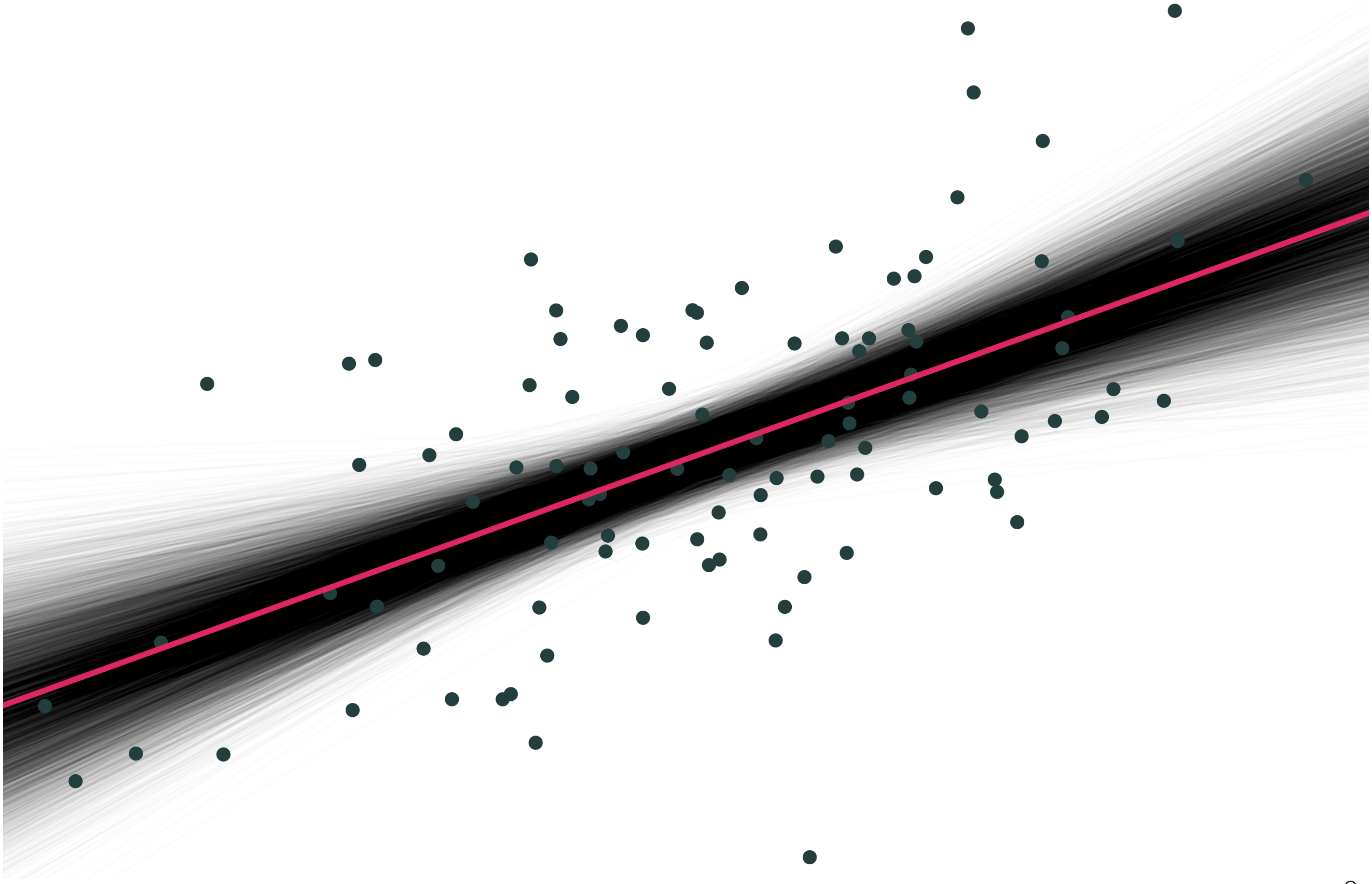
Modelo muestral

$$\hat{y}_i = 3.21 + 0.45 x_i$$

Ahora repitamos esto **10,000 veces**.

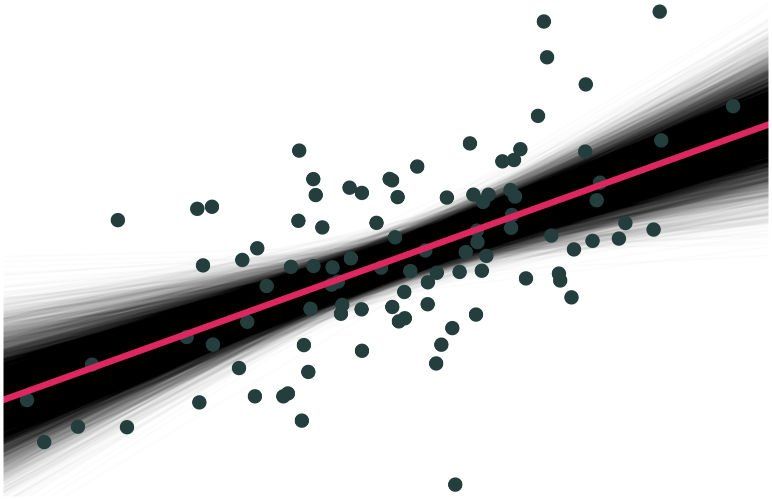
(Este ejercicio se llama ejercicio de Montecarlo)

Población *vs.* Muestra



Población vs. muestra

Pregunta: ¿Por qué nos importa la *población vs. muestra*?



- En **promedio**, nuestras líneas de regresión se ajustan muy bien a la línea de la población.
- Sin embargo, las **líneas individuales** (muestras) pueden desviarse significativamente.
- Las diferencias entre las muestras individuales y la población generan **incertidumbre** para el econometrista.

Población vs. muestra

Población vs. muestra

Pregunta: ¿Por qué nos importa la *población vs. muestra*?

Pregunta: ¿Por qué nos importa la *población vs. muestra*?

Población vs. muestra

Pregunta: ¿Por qué nos importa la *población vs. muestra*?

Respuesta: La incertidumbre es importante.

$\hat{\beta}$ en sí mismo es una variable aleatoria, dependiente de la muestra aleatoria. Cuando tomamos una muestra y realizamos una regresión, no sabemos si es una muestra 'buena' ($\hat{\beta}$ está cerca de β) o una muestra 'mala' (nuestra muestra difiere significativamente de la población).

Población *vs.* muestra

Incertidumbre

Mantener un registro de esta incertidumbre será un concepto clave a lo largo de nuestro curso.

- Estimación de errores estándar para nuestras estimaciones.
- Pruebas de hipótesis.
- Corrección para la heteroscedasticidad y autocorrelación.

Primero, repasemos cómo obtenemos estas estimaciones (inciertas) de regresión.

El estimador

Podemos estimar una línea de regresión en `Stata` (`reg y x`) y en `R` (`lm(y ~ x, my_data)`). Pero, ¿de dónde provienen estas estimaciones?

> $$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
>
> que nos da la línea de *mejor ajuste* a través de nuestro conjunto de datos.

¿qué queremos decir con "línea de mejor ajuste"?

Siendo el "mejor"

Pregunta: ¿Qué queremos decir con *línea de mejor ajuste*?

Respuestas:

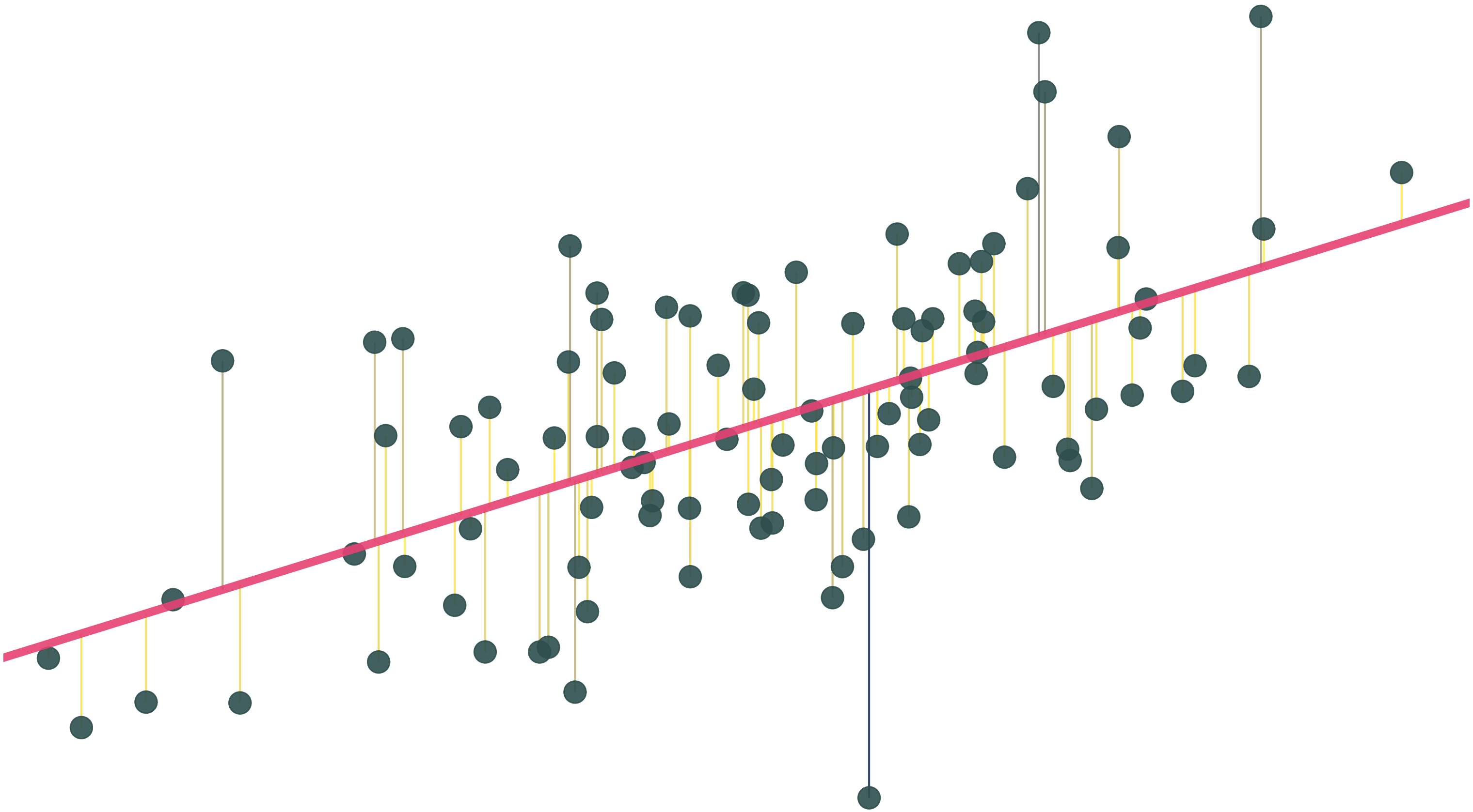
- En general (en econometría), *línea de mejor ajuste* significa la línea que minimiza la suma de residuos al cuadrado (SRC):

$$\text{SRC} = \sum_{i=1}^{n} \epsilon_i^2 \quad \text{donde} \quad \epsilon_i = y_i - \hat{y}_i$$

- Los **mínimos cuadrados ordinarios** (**MCO**) minimizan la suma de residuos al cuadrado.
- Basado en un conjunto de supuestos (GAUSS MARKOV), MCO es:
  - Es insesgado (y consistente)
  - Es el *mejor* (estimador insesgado lineal de mínima varianza *MELI*)

MCO *vs.* otras lineas / estimadores

La estimación de MCO es la combinación de $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimiza la SRC



MCO

Formalmente

En una regresión lineal simple, el estimador de MCO proviene de escoger $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimice la suma de residuos al cuadrado (SRC), *i.e.*,

$$\min_{\hat{\beta}_0,\, \hat{\beta}_1} \text{SRC}$$

pero nosotros sabemos que $\text{SRC} = \sum_i \tilde{\epsilon}_i^{\,2}$. Now use the definitions of $\tilde{\epsilon}_i$ and $\hat{y}$.

$$\begin{aligned} \tilde{\epsilon}_i^{\,2} &= \left(y_i - \hat{y}_i \right)^2 = \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2 \\ &= y_i^2 - 2 y_i \hat{\beta}_0 - 2 y_i \hat{\beta}_1 x_i + \hat{\beta}_0^2 + 2 \hat{\beta}_0 \hat{\beta}_1 x_i + \hat{\beta}_1^2 x_i^2 \end{aligned}$$

Recordatorio: Minimizar una función multivariada requiere (**1**) que las primeras derivadas sean iguales a cero (las *condiciones de primer orden*) y (**2**) las condiciones de segundo orden (concavidad).

Formalmente

Nos estamos acercando. Necesitamos **minimizar la SRC**.

$$\text{SRE} = \sum_i \tilde{e}_i{}^2 = \sum_i \left(y_i^2 - 2y_i\hat{\beta}_0 - 2y_i\hat{\beta}_1 x_i + \hat{\beta}_0^2 + 2\hat{\beta}_0\hat{\beta}_1 x_i + \hat{\beta}_1^2 x_i^2 \right)$$

For the first-order conditions of minimization, we now take the first derivates of SSE with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\begin{aligned} \frac{\partial \text{SRC}}{\partial \hat{\beta}_0} &= \sum_i \left(2\hat{\beta}_0 + 2\hat{\beta}_1 x_i - 2y_i \right) = 2n\hat{\beta}_0 + 2\hat{\beta}_1 \sum_i x_i - 2\sum_i y_i \\ &= 2n\hat{\beta}_0 + 2n\hat{\beta}_1 \overline{x} - 2n\overline{y} \end{aligned}$$

donde $\overline{x} = \frac{\sum x_i}{n}$ y $\overline{y} = \frac{\sum y_i}{n}$ son medias muestrales de x y y (de tamaño n).

Formalmente

Las condiciones de primer orden establecen que las derivadas deben ser iguales a cero:

$$\frac{\partial \mathrm{SSE}}{\partial \hat{\beta}_0} = 2n\hat{\beta}_0 + 2n\hat{\beta}_1\overline{x} - 2n\overline{y} = 0$$

Lo que implica

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

Ahora para $\hat{\beta}_1$.

Formalmente

Tomemos la derivada de la SRC con respecto a $\hat{\beta}_1$

$$\frac{\partial \text{SSE}}{\partial \hat{\beta}_1} = \sum_i \left(2\hat{\beta}_0 x_i + 2\hat{\beta}_1 x_i^2 - 2y_i x_i\right) = 2\hat{\beta}_0 \sum_i x_i + 2\hat{\beta}_1 \sum_i x_i^2 - 2\sum_i y_i x_i$$
$$= 2n\hat{\beta}_0 \overline{x} + 2\hat{\beta}_1 \sum_i x_i^2 - 2\sum_i y_i x_i$$

Igualarlo a cero

$$\frac{\partial \text{SSE}}{\partial \hat{\beta}_1} = 2n\hat{\beta}_0 \overline{x} + 2\hat{\beta}_1 \sum_i x_i^2 - 2\sum_i y_i x_i = 0$$

y reemplazarlo $\hat{\beta}_0$, *i.e.*, $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$. Thus,

$$2n\left(\overline{y} - \hat{\beta}_1 \overline{x}\right)\overline{x} + 2\hat{\beta}_1 \sum_i x_i^2 - 2\sum_i y_i x_i = 0$$

Formalmente

Continuando

$$2n\left(\overline{y}-\hat{\beta}_1\overline{x}\right)\overline{x}+2\hat{\beta}_1\sum_i x_i^2-2\sum_i y_i x_i=0$$

$$2n\overline{y}\,\overline{x}-2n\hat{\beta}_1\overline{x}^2+2\hat{\beta}_1\sum_i x_i^2-2\sum_i y_i x_i=0$$

$$\implies 2\hat{\beta}_1\left(\sum_i x_i^2-n\overline{x}^2\right)=2\sum_i y_i x_i-2n\overline{y}\,\overline{x}$$

$$\implies \hat{\beta}_1=\frac{\sum_i y_i x_i-2n\overline{y}\,\overline{x}}{\sum_i x_i^2-n\overline{x}^2}=\frac{\sum_i(x_i-\overline{x})(y_i-\overline{y})}{\sum_i(x_i-\overline{x})^2}$$

MCO

Formalmente

LISTOO!

Ahora tenemos nuestros lindos estimadores

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$

and the intercept

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Y ahora saben de dónde la Y ahora sabes de dónde proviene la parte de *mínimos cuadrados* en el término "mínimos cuadrados ordinarios". 🎊

Ahora pasamos a los supuestos y propiedades (implícitas) de los Mínimos Cuadrados Ordinarios (MCO / OLS).

MCO: Propiedades y supuestos

MCO: Propiedades y supuestos

Propiedades

Pregunta: ¿Qué propiedades podrían ser importantes para un estimador?

MCO: Propiedades y supuestos

Propiedades

Tangente: Primero revisemos las propiedades estadísticas.

MCO: Propiedades y supuestos

Propiedades

Repaso: Funciones de densidad

Recordemos que utilizamos las **funciones de densidad de probabilidad** (FDP- PDF) para describir la probabilidad de que una **variable aleatoria continua** tome valores en un rango dado. (El área total = 1).

Estas FDPs caracterizan distribuciones de probabilidad, y las distribuciones más comunes/famosas/populares reciben nombres (por ejemplo, normal, *t*, Gamma).

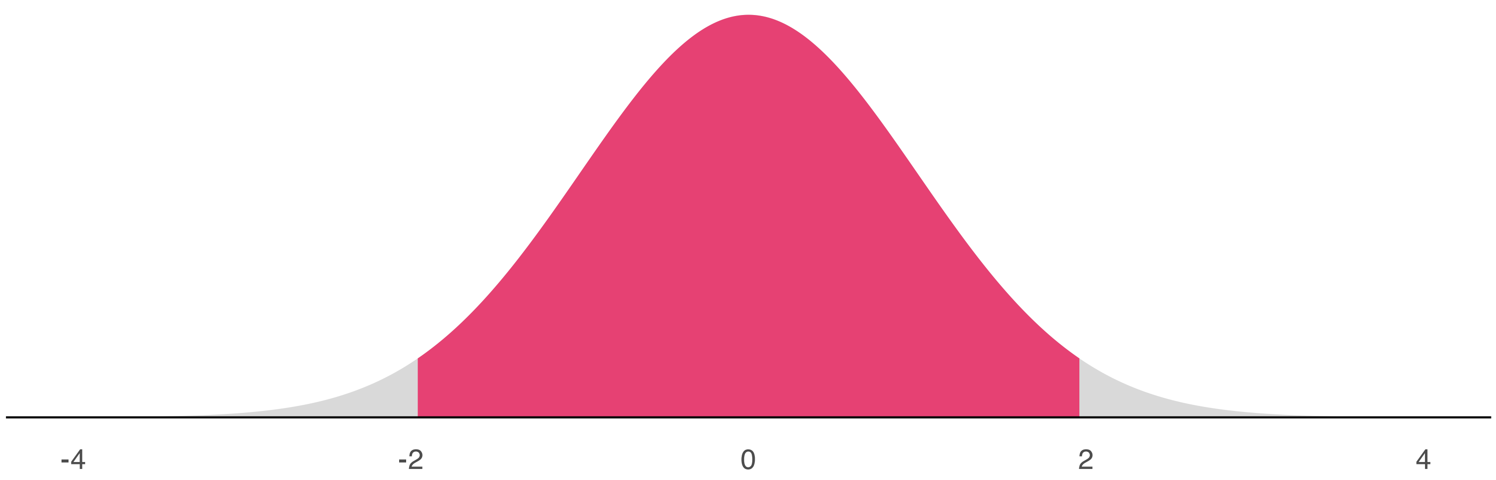
Repaso: Funciones de densidad

La probabilidad de que una variable aleatoria normal estándar tome un valor entre -2 y 0: $\mathrm{P}(-2 \leq X \leq 0) = 0.48$

Propiedades

Repaso: Funciones de densidad

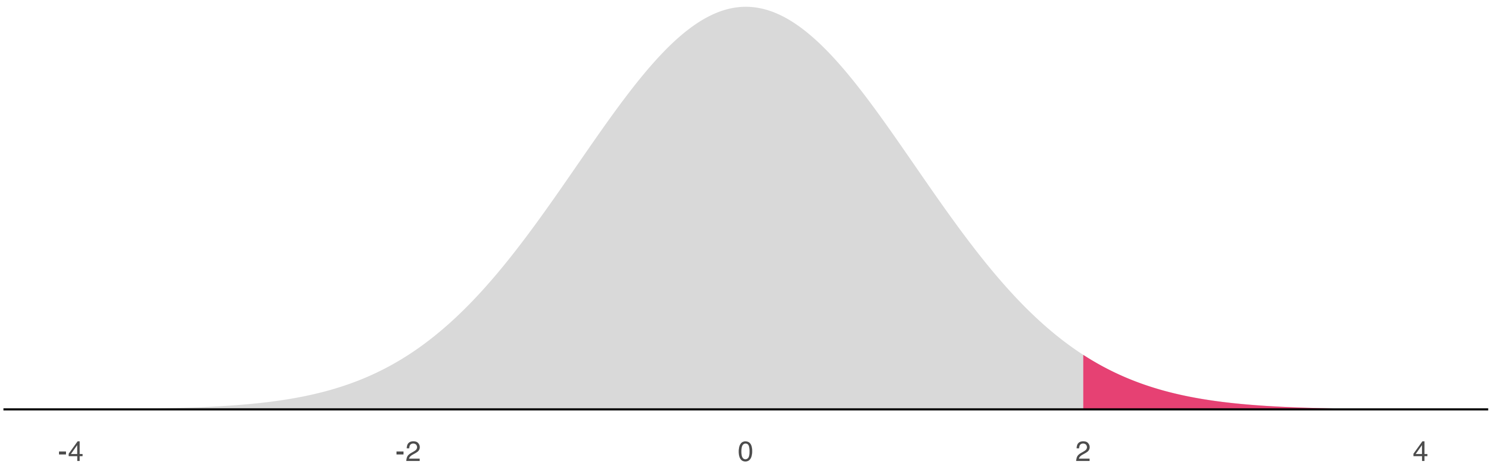
La probabilidad de que una variable aleatoria normal estándar tome un valor entre -1.96 y 1.96: $\mathbf{P(-1.96 \leq \mathit{X} \leq 1.96) = 0.95}$



Propiedades

Repaso Funciones de densidad

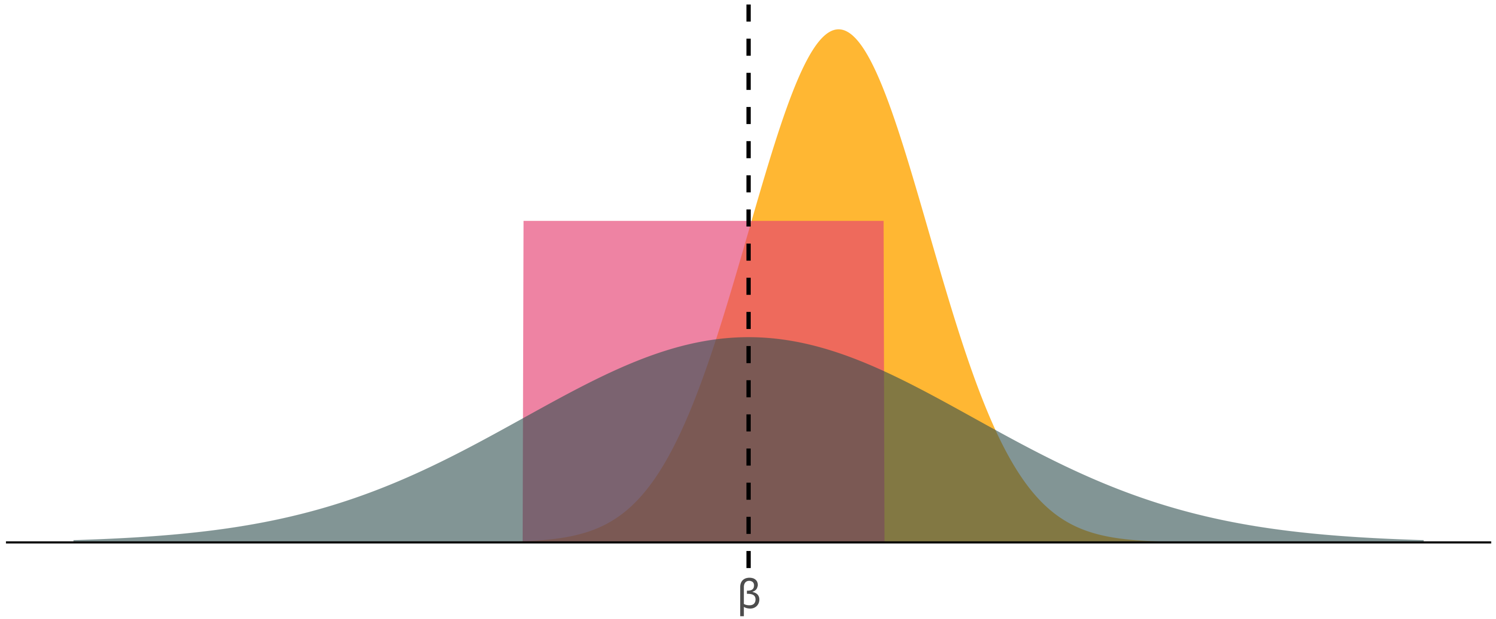
La probabilidad de que una variable aleatoria normal estándar tome un valor mayor a 2: $\mathrm{P}(X > 2) = \mathbf{0.023}$



MCO: Propiedades y supuestos

Propiedades

Imaginemos que estamos tratando de estimar un parámetro desconocido β, y conocemos las distribuciones de tres estimadores competitivos. ¿Cuál de ellos elegimos? ¿Cómo decidimos?



MCO: Propiedades y supuestos

Propiedades

Pregunta: ¿Qué propiedades podrían ser importantes para un estimador?

Respuesta uno: Sesgo (Bias).

En promedio (después de *muchas* repeticiones), ¿el estimador tiende hacia el valor correcto?

Más formalmente: ¿La media de la distribución del estimador es igual al parámetro que estima?

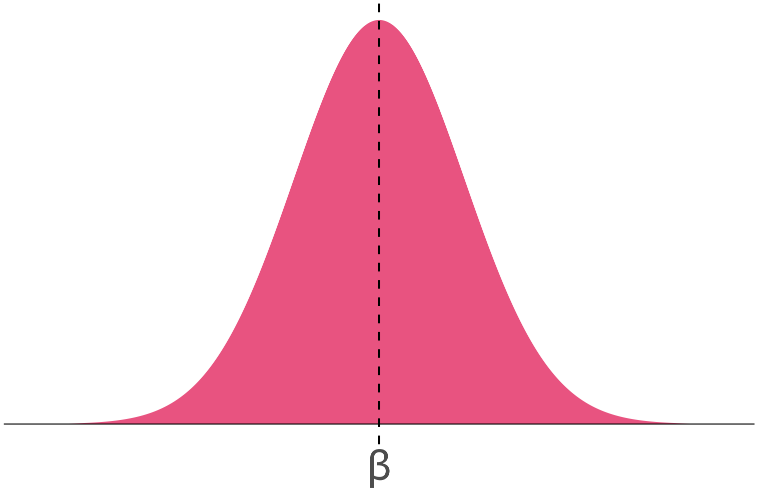
$$\underset{\beta}{\text{Sesgo}}\left(\hat{\beta}\right) = \boldsymbol{E}\left[\hat{\beta}\right] - \beta$$

MCO: Propiedades y supuestos

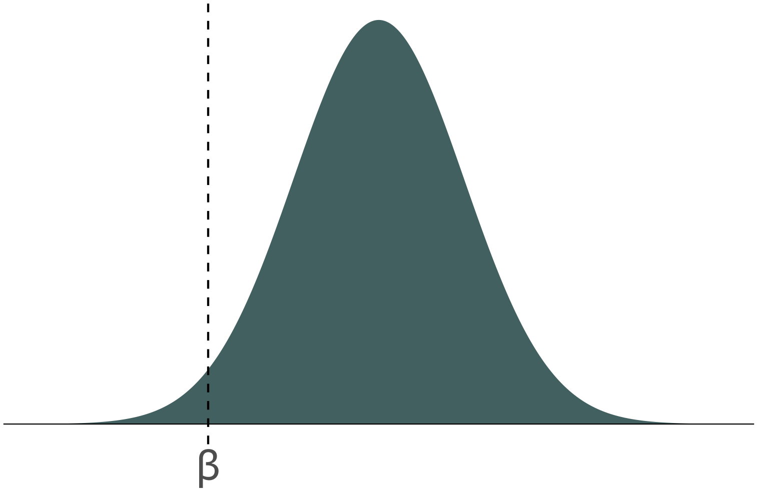
Propiedades

Respuesta uno: Sesgo (Bias).

Estimador Insesagado: $\boldsymbol{E}\left[\hat{\beta}\right] = \beta$

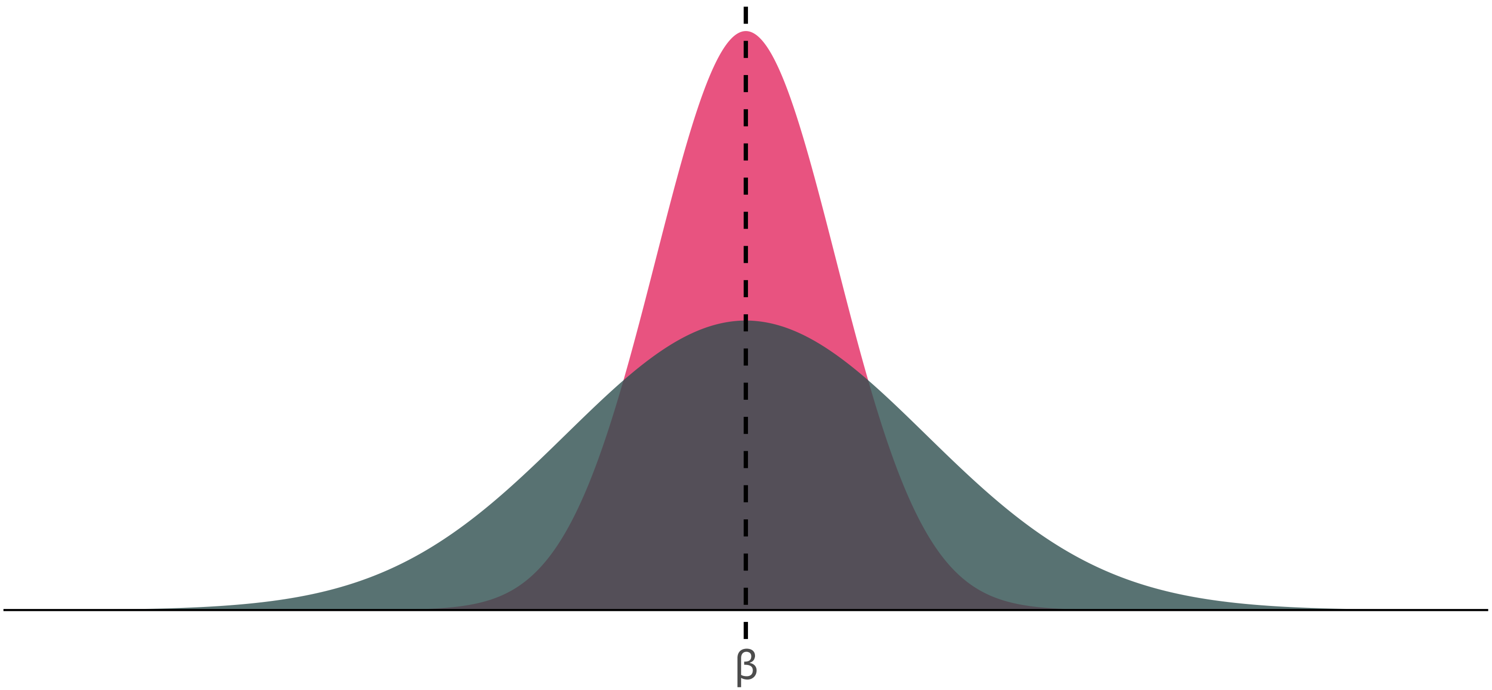


Estimador Sesagado: $\boldsymbol{E}\left[\hat{\beta}\right] \neq \beta$



Propiedades

Respuesta dos: varianza



MCO: Propiedades y supuestos

Propiedades

Respuesta uno: Sesgo

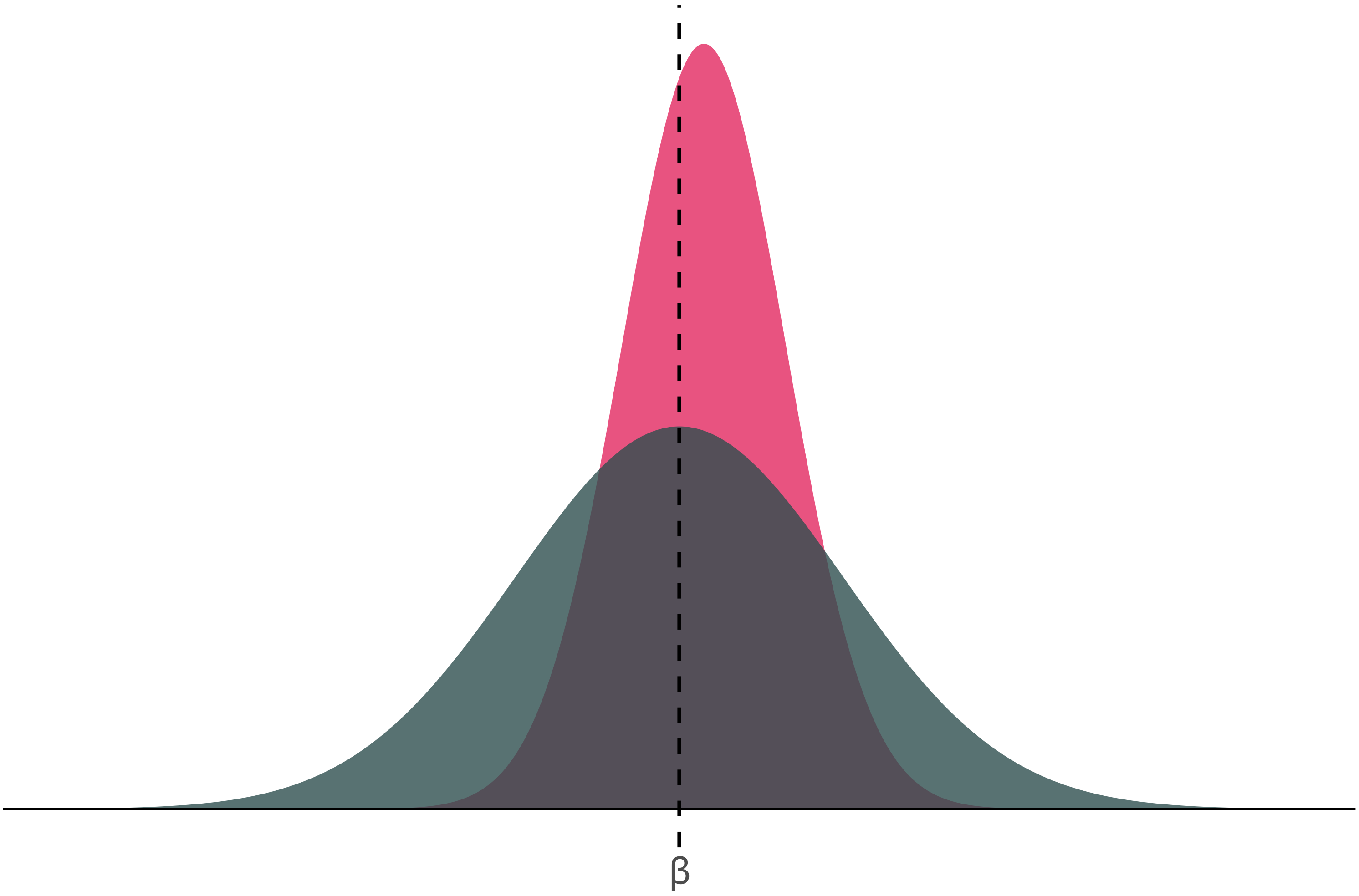
Respuesta dos: Varianza

El trade off: sesgo vs varianza .

¿Deberíamos estar dispuestos a aceptar un poco de sesgo para reducir la varianza?

En econometría, generalmente nos adherimos a estimadores insesgados (o consistentes). Pero en otras disciplinas (especialmente ciencias de la computación), se reflexiona un poco más sobre este compromiso.

El tradeoff.



MCO: Supuestos y propiedades

Propertiedades

- MCO is **insesgado**.
- MCO tiene la **menor varianza** de todos los estiamdores lineales e insesgados

MCO: Supuestos y propiedades

Supuestos

MCO: Supuestos y propiedades

Los supuestos del modelo clásico de regresión lineal están resumidos en la siguiente tabla:

Supuesto	Implicación
A1. Lineal	$y = X\beta + \epsilon$
A2. Exogeneidad Estricta	$E[\epsilon_i \mid X] = 0$
A3. Colinealidad Imperfecta	X es una matriz nxK con rango K
A4. Perturbaciones Esféricas	$Var[\epsilon_i \mid X] = \sigma^2$
	$Cov\left[\epsilon_i, \epsilon_j \mid X\right] = 0$
A5. Regresores no estocásticos	X es una matriz nxK no estocástica
A6. Normalidad	$\epsilon \mid X \sim N(0, \sigma^2 I)$
A.2, A.4-A.6	$\epsilon \mid X \sim i.i.d \quad N(0, \sigma^2 I)$

S1: Lineal en parámetros

El valor esperado de la distribución de y está relacionada con el valor de X_i de una manera lineal:

$$E[Y|X_i = x] = f(X_i) = X_i\beta$$

Por lo tanto el proceso generador de datos es igual a

$$Y_i = X\beta + \epsilon$$

S1: Lineal en parámetros

S1: Lineal en parámetros

S1: Lineal en parámetros

Modelos de Regresión

- **Lineal**: $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$
- **Log-log**: $ln(y_i) = \beta_1 + \beta_2 ln(x_i) + \epsilon_i$
- **Log-lineal**: $ln(y_i) = \beta_1 + \beta_2 x_i + \epsilon_i$
- **Lineal-log**: $y_i = \beta_1 + \beta_2 ln(x_i) + \epsilon_i$
- **Recíproco**: $y_i = \beta_1 + \beta_2 \frac{1}{x_i} + \epsilon_i$
- **Cuadrático**: $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \epsilon_i$
- **Interactuado**: $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \beta_4 (x_{i1} \times x_{i2}) + \epsilon_i$

En general, un modelo de regresión es no lineal cuando ni es lineal en su formación original, ni se puede convertir en un modelo lineal mediante

S2: Exogeneidad Estricta

Para muchas de las aplicaciones de la economía el supuesto más importante es la **EXOGENEIDAD**

$$E[\epsilon \mid X] = 0$$

Pero qué quiere decir?

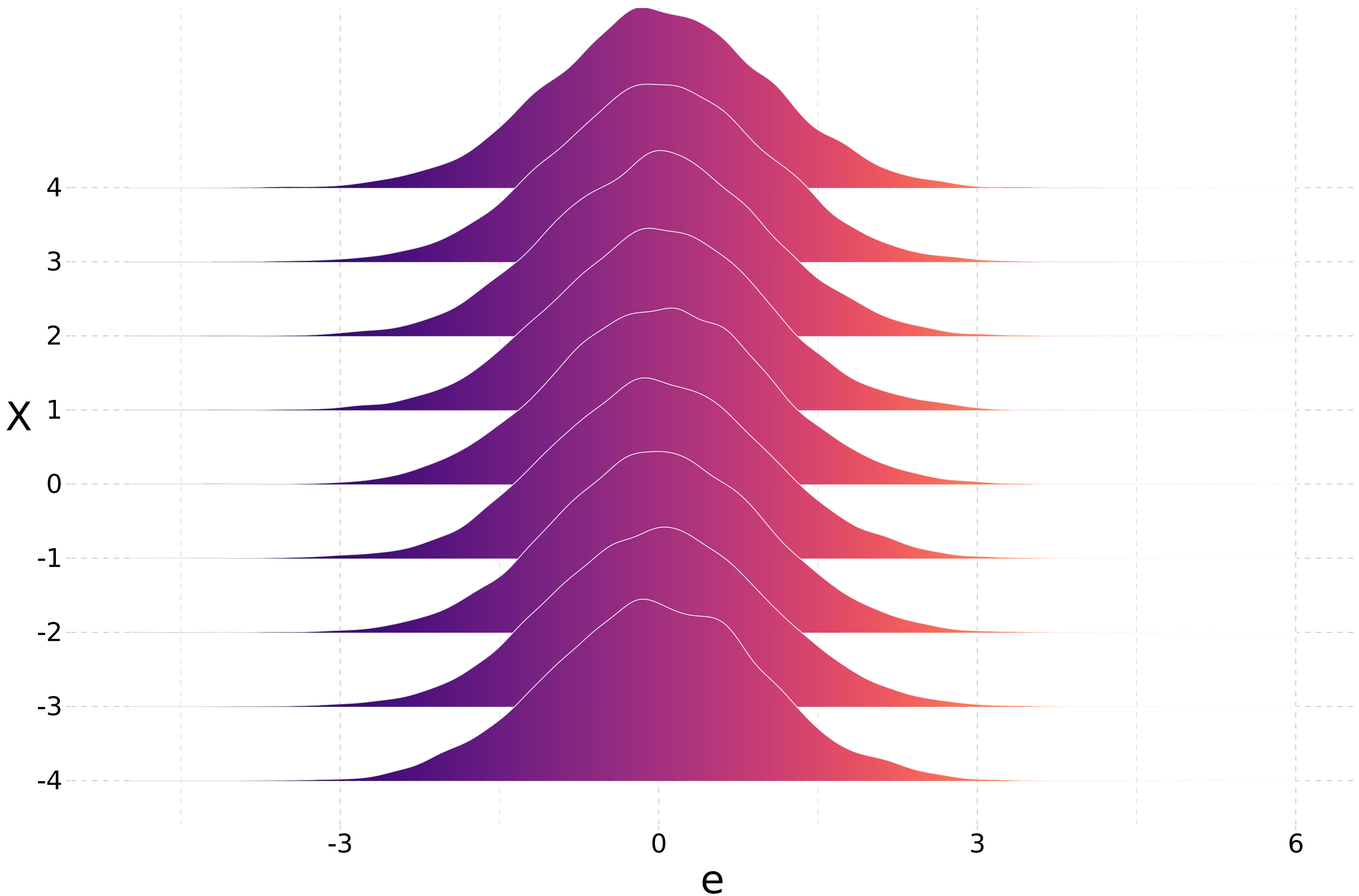
Una forma de pensar en esta definición es:

> Para *cualquier* valor de X, el valor esperado de los residuos debe ser igual a cero

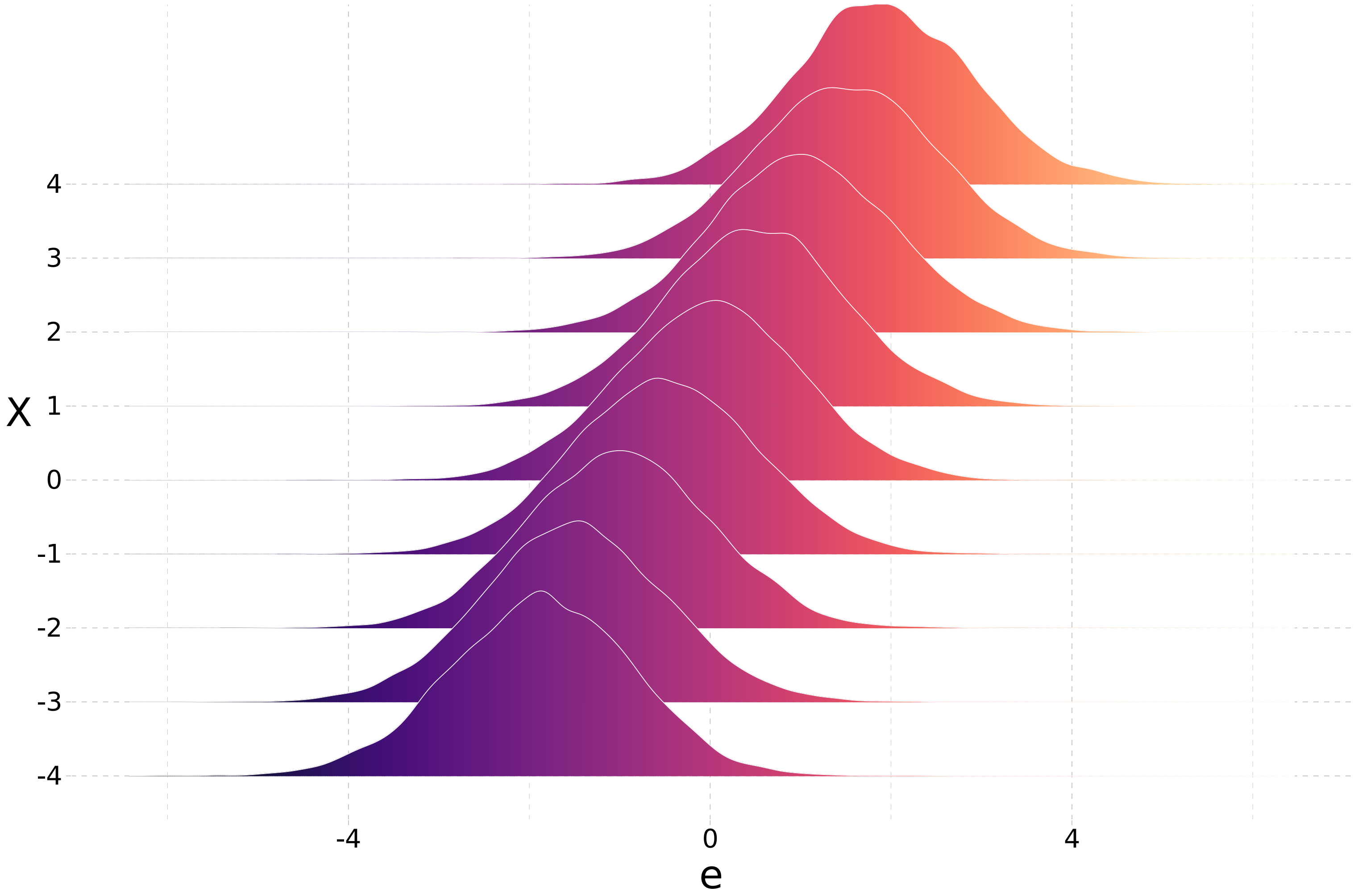
- *E.g.*, $E[u \mid X = 1] = 0$ *and* $E[u \mid X = 100] = 0$
- *E.g.*, $E[u \mid X_2 = \text{Mujer}] = 0$ *and* $E[u \mid X_2 = \text{Hombre}] = 0$
- Note: $E[u \mid X] = 0$ es más restrictivo que $E[u] = 0$

Graficamente...

Exogeneidad Estricta se cumple, $E[\epsilon \mid X] = 0$

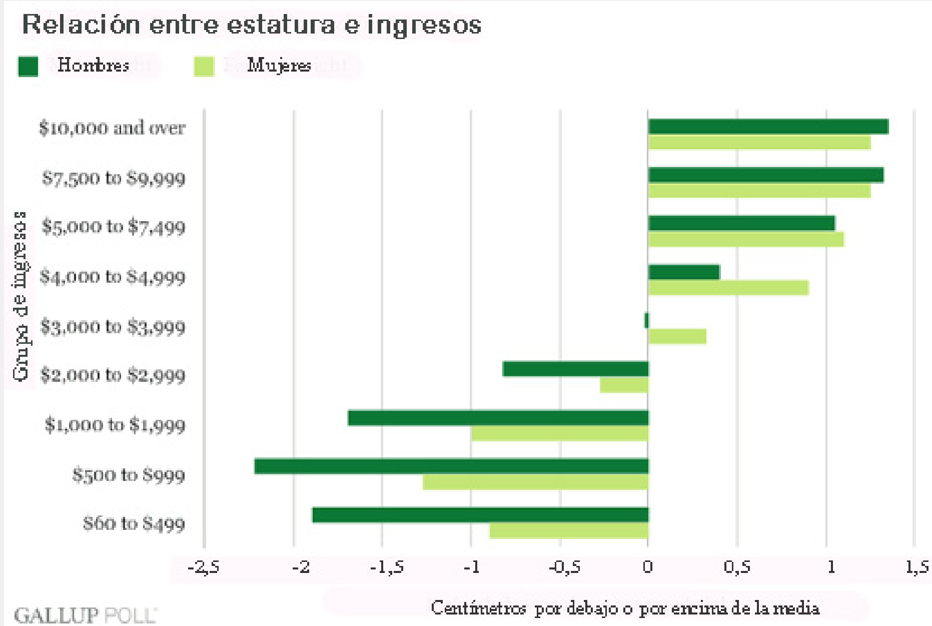


Exogeneidad Estricta se Incumple, *i.e.*, $E[\epsilon \mid X] \neq 0$



¿Por qué las personas más altas ganan más dinero?

Por Iñaki Berazaluce | 22.02.2012 | 14:10 h.



Hace años, un programa de la TV salió a las calles para preguntar a los transeúntes cuánto creían que ganaban dos hombres: uno de estatura media y otro alto. La totalidad de los encuestados asignaron al más alto un sueldo superior. Los participantes en aquel seudoestudio proyectaban algo que seguramente constataban en su vida cotidiana: en general, **las personas más altas ocupan puestos de más responsabilidad y ganan más** que las "personas verticalmente limitadas", por utilizar el argot políticamente correcto.

El efecto de la estatura en los salarios ha sido valorado con cierta precisión por diversos estudios. El que ilustra este artículo, por ejemplo, muestra una correlación entre ambas variables: **los hombres que ganan más de 10.000 dólares al mes miden 2 centímetros más que la media, mientras los "mildolaristas" miden unos 3 centímetros menos que la media**, según la encueta de Gallup. Entre las mujeres, las diferencias no son acentuadas pero también existen.

Más preciso aún es el estudio que realizó en 2006 un equipo inglés con datos de 4.630 finlandeses nacidos entre 1934 y 1940: **un centímetro más de estatura cuando el niño tiene 1 año de edad deviene en un aumento del 3,5% de ingresos en la edad adulta.** El estudio llevado a cabo por la Universidad de Southampton halló que este vínculo se da en la más temprana infancia y, por tanto, la explicación del fenómeno no tendría que ver con el estatus

Los niños golpeados, más propensos a la obesidad y los males cardiacos

Por primera vez un estudio explora el vínculo entre el "castigo físico severo" y las enfermedades en la edad adulta

Por **Antimio Cruz**
Jueves, 25 de julio de 2013 a las 09:54

Los golpes recibidos en la niñez podrían ser un factor para el desarrollo de enfermedades como la obesidad en la adultez (Getty Images).

Lo más importante

- Un niño que es golpeado es más propenso a desarrollar artritis, obesidad y males cardiovasculares, según un estudio
- El informe publicado en la revista 'Pediatrics' detalla que hay relación entre el

(CNNMéxico) — Empujar, sacudir o abofetear a un niño puede provocar que al crecer sea más propenso a desarrollar obesidad, males cardiovasculares y artritis, de acuerdo con una investigación pionera de la Universidad de Manitoba, Canadá.

El estudio, realizado a 34,226 adultos y publicado en la edición más reciente de la revista *Pediatrics*, explora por primera vez la relación entre el "castigo físico severo" (golpes, jalones y cachetadas) y las enfermedades de la edad adulta.

Vivir a menos de 300 metros de un parque reduce el riesgo de cáncer de mama

Un estudio con casi 3.000 mujeres de 10 provincias españolas respalda la creación de más zonas verdes

MANUEL ANSEDE

21 AGO 2018 - 18:22 CEST

Una mujer lee tumbada en el Parque del Retiro de Madrid, tras dos jóvenes. GORKA LEJARCEGI

Una de cada ocho mujeres tendrá cáncer de mama en algún momento de su vida. En muchos casos, el origen de este tumor es un misterio. Los factores de riesgo conocidos —como las mutaciones genéticas hereditarias, la obesidad y el consumo de alcohol— apenas explican la mitad de los casos detectados. Un nuevo estudio con casi 3.000 mujeres de 10 provincias españolas arroja un poco más de luz sobre este tumor. Vivir a menos de 300 metros de una zona verde urbana reduce el riesgo de sufrir cáncer de mama, según el trabajo, firmado por científicos de una quincena de instituciones punteras.

Los resultados del estudio muestran que vivir a menos de 300 metros de un parque o de un jardín disminuye el riesgo un 35%. Y la asociación es lineal. Vivir a menos de 100 metros supone una reducción del 44%. La epidemióloga Cristina O'Callaghan, coautora de la investigación, pide interpretar estos porcentajes "con todas las cautelas", dado el amplio margen de error estadístico, pero subraya que la asociación observada es "robusta".

Tomar más de tres tazas de café reduce tamaño de los senos

Científicos advierten que existe una relación inversamente proporcional entre el tamaño de los pechos y la cantidad de café ingerida.

Por: Elespectador.com

Científicos suecos descubrieron que los senos de las mujeres reducen su tamaño si toman tres o más tazas de café por día.

Según los científicos **existe una relación inversamente proporcional entre el tamaño de los pechos y la cantidad de café ingerida.**

Los resultados del estido advierten que **con cada taza de café adicional a la tercera el efecto se incrementa. Las mujeres que toman mucho café tienen los pechos pequeños.**

El estudio fue llevado a cabo con una muestra de 300 mujeres que fueron encuestadas sobre la medida de sus bustos y el consumo de café tomado por

Confirmada la mayor longevidad asociada al consumo de café

G+ Me gusta 2123 Tweet

Tras estudiar durante una década a 20.000 personas, investigadores españoles han confirmado que beber café se relaciona con una mayor longevidad. Los beneficios de la ingesta están especialmente presentes en las personas más mayores.

Más información sobre: café consumo longevidad

SINC | Seguir a @agencia_sinc | 30 noviembre 2018 13:19

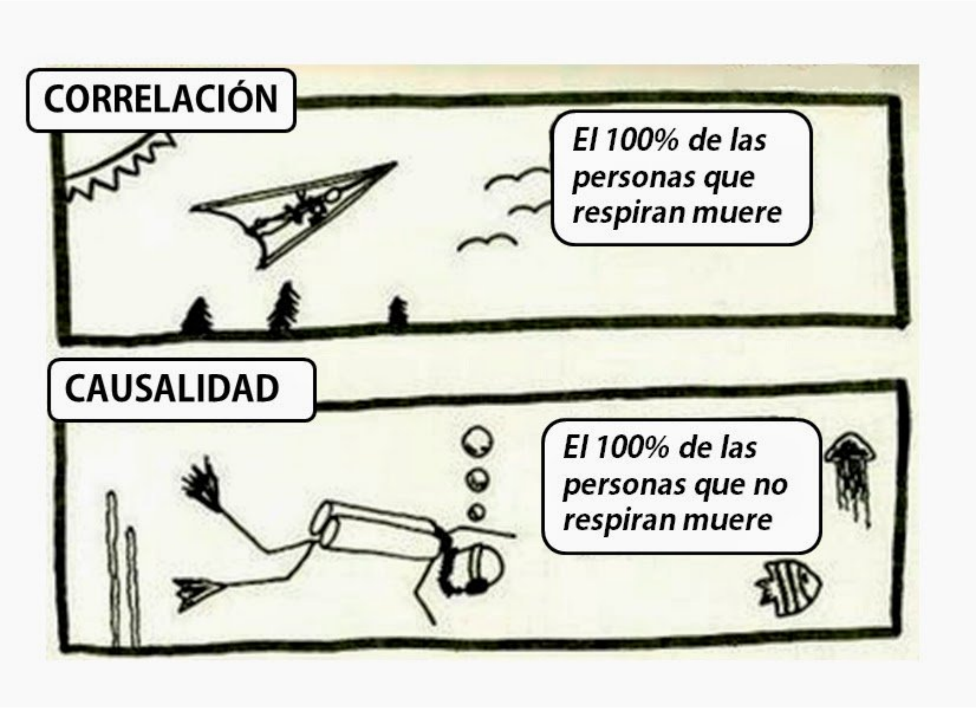
Estos investigadores encontraron beneficios claros de consumir café sobre el riesgo de mortalidad por todas las causas. / CIBEROBN

Científicos del Centro de Investigación Biomédica en Red de Obesidad y Nutrición (CIBEROBN) han encontrado mayor longevidad asociada al consumo de café tras estudiar a 20.000 voluntarios, graduados universitarios de toda España, durante una media de diez años.

La investigación constató que, con paridad en factores relevantes como la edad, consumir café habitualmente se asocia significativamente a una menor mortalidad. El efecto se observó tanto en el café con cafeína como con el descafeinado; tanto soluble como de máquina.

Según los datos del estudio, publicado en el *American Journal of Clinical Nutrition*, la protección era más fuerte en personas con 55 o más años y mostraba una clara tendencia dosis respuesta en el rango de consumo de entre 1 y 6 tazas al día.

Con paridad en factores relevantes como la edad,

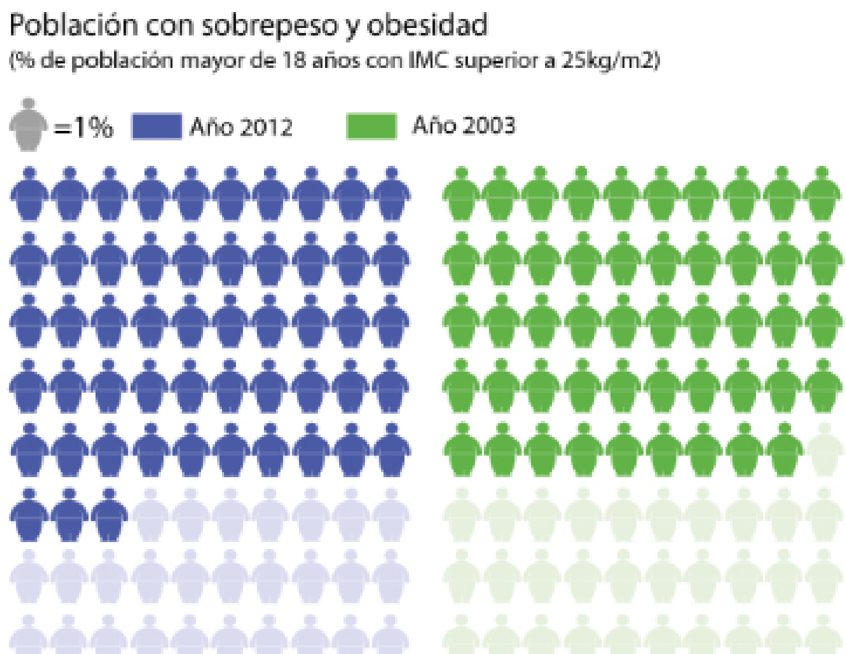

CORRELACIÓN
El 100% de las personas que respiran muere
CAUSALIDAD
El 100% de las personas que no respiran muere

Cuanto más ricos, más delgados

comentarios 12

tuenti

Muchas veces hemos oído eso de que "dejar de fumar engorda" y, más allá de lo simplista de la afirmación, podría decirse que algo así es lo que le ha pasado a España en la última década, según la Encuesta Nacional de Salud del INE [descargue aquí los datos]. **Un 53,6% de la población española tiene sobrepeso**, mientras que en 2003 la cifra no llegaba al 50%. Ahora en cambio, con respecto a 2003, fumamos un 4% menos.



Correlation vs causality. h/t Doug

S3: Colinealidad Imperfecta

$$X \text{ es una matriz } nxK \text{ con rango } K$$

Wooldridge (2003), este supuesto permite que las variables independientes estén correlacionadas, siempre y cuando no lo hagan de forma perfecta.

S3: Perturbaciones Esféricas

Homocedasticidad

$Var(\epsilon_i|X) = \sigma^2 \text{ para } i = 1, \ldots, n$

Homocedasticidad significa que la dispersión alrededor de la recta de regresión es igual para los diversos valores de X.

No Autocorrelación

$Cov(\epsilon_i, \epsilon_j|X) = 0 \text{ para } i \neq j$

La no autocorrelación significa que los errores no se encuentran relacionados entre sí. La autocorrelación generalmente aparece en datos de series de tiempo aunque también puede presentarse en el caso de una muestra de corte transversal (e.g., correlación espacial).

S3: Perturbaciones Esféricas

$$
\begin{aligned}
Var(\epsilon|X) &= E[\epsilon\epsilon'|X] - E[\epsilon|X]E[\epsilon'|X] \\
&= E[\epsilon\epsilon'|X] - \underbrace{E[\epsilon|X]E[\epsilon'|X]}_{0} \text{ por supuesto A2.} \\
&= E\left[\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_n \end{bmatrix} |X\right]
\end{aligned}
$$

S3: Perturbaciones Esféricas

$$Var(\epsilon|X) = \begin{bmatrix} E[\epsilon_1\epsilon_1|X] & E[\epsilon_1\epsilon_2|X] & \cdots & E[\epsilon_1\epsilon_n|X] \\ E[\epsilon_2\epsilon_1|X] & E[\epsilon_2\epsilon_2|X] & \cdots & E[\epsilon_2\epsilon_n|X] \\ \vdots & \vdots & \ddots & \vdots \\ E[\epsilon_n\epsilon_1|X] & E[\epsilon_n\epsilon_1|X] & \cdots & E[\epsilon_n\epsilon_n|X] \end{bmatrix}$$

$$= \begin{bmatrix} Var[\epsilon_1|X] & Cov[\epsilon_1\epsilon_2|X] & \cdots & Cov[\epsilon_1\epsilon_n|X] \\ Cov[\epsilon_2\epsilon_1|X] & Var[\epsilon_2|X] & \cdots & Cov[\epsilon_2\epsilon_n|X] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[\epsilon_n\epsilon_1|X] & Cov[\epsilon_n\epsilon_1|X] & \cdots & Var[\epsilon_n|X] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \text{ por supuesto A4.}$$

S3: Perturbaciones Esféricas

$$Var[\epsilon|X] = E[\epsilon\epsilon'|X] = \sigma^2 I$$