MCO: primeros pasos

Econometría I

Paula Pereda (ppereda@correo.um.edu.uy)
27 de agosto de 2021

Repaso

Población *vs.* muestra

Modelos y notación

Escribimos nuestro modelo poblacional (simple) así:

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

y nuestro modelo de regresión estimado basado en la muestra así:

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

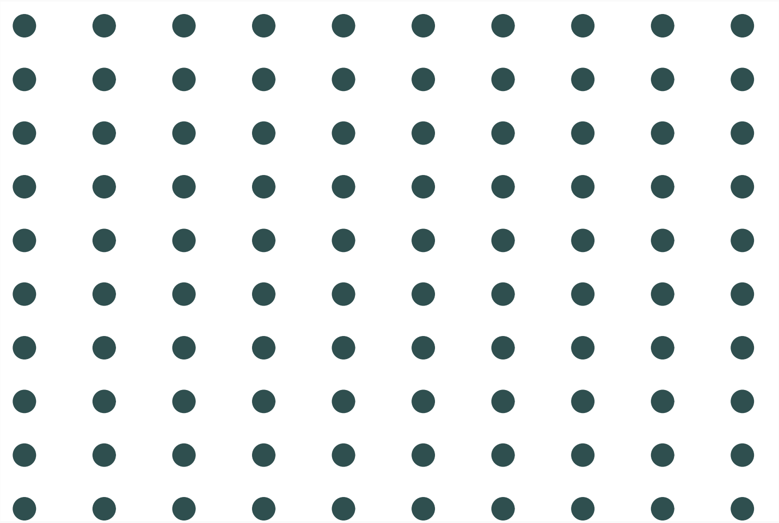
Un modelo estimado de regresión produce estimaciones para cada observación:

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

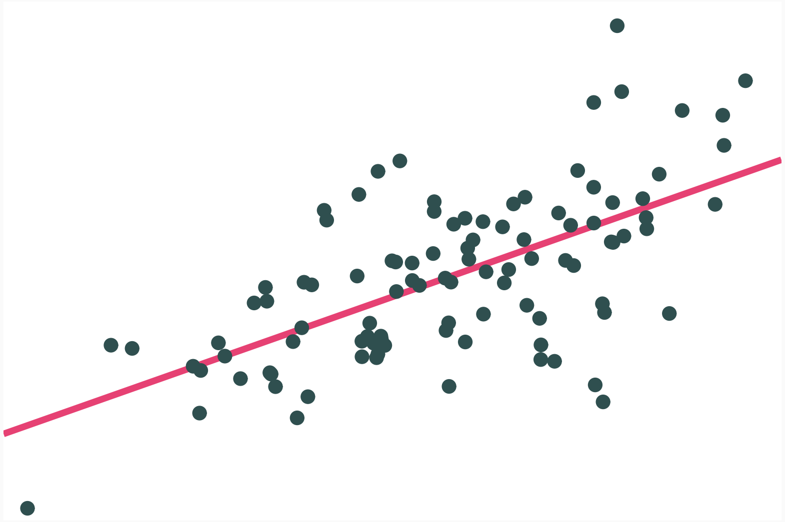
lo que nos da la línea que *mejor se ajusta* a nuestra muestra.

Población *vs.* muestra

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



Población



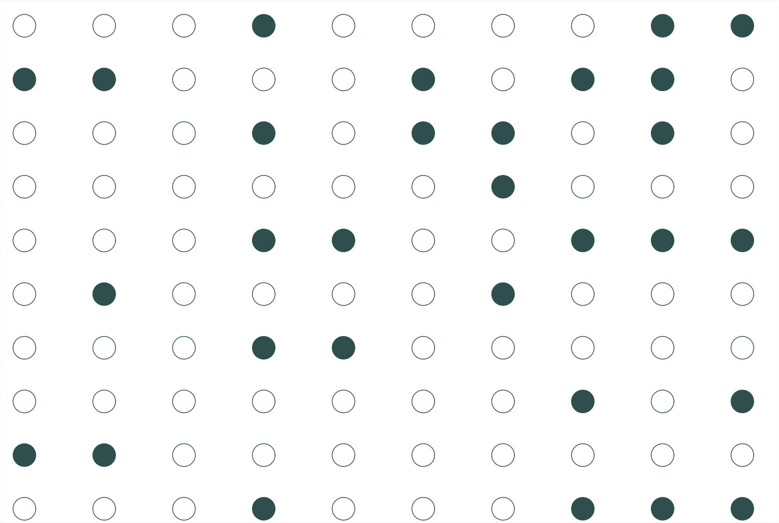
Relación poblacional

$$y_i = 2.95 + 0.5x_i + u_i$$

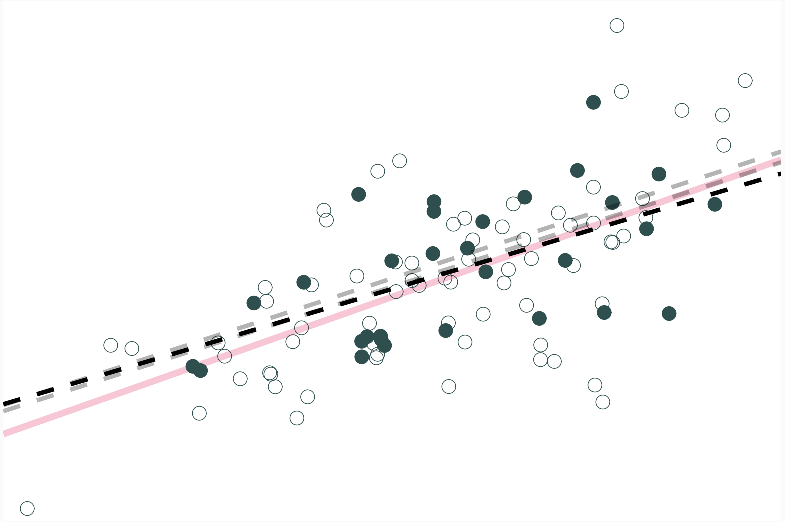
$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Población *vs.* muestra

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



Muestra 3: 30 individuos aleatorios



Relación poblacional

$$y_i = 2.95 + 0.5x_i + u_i$$

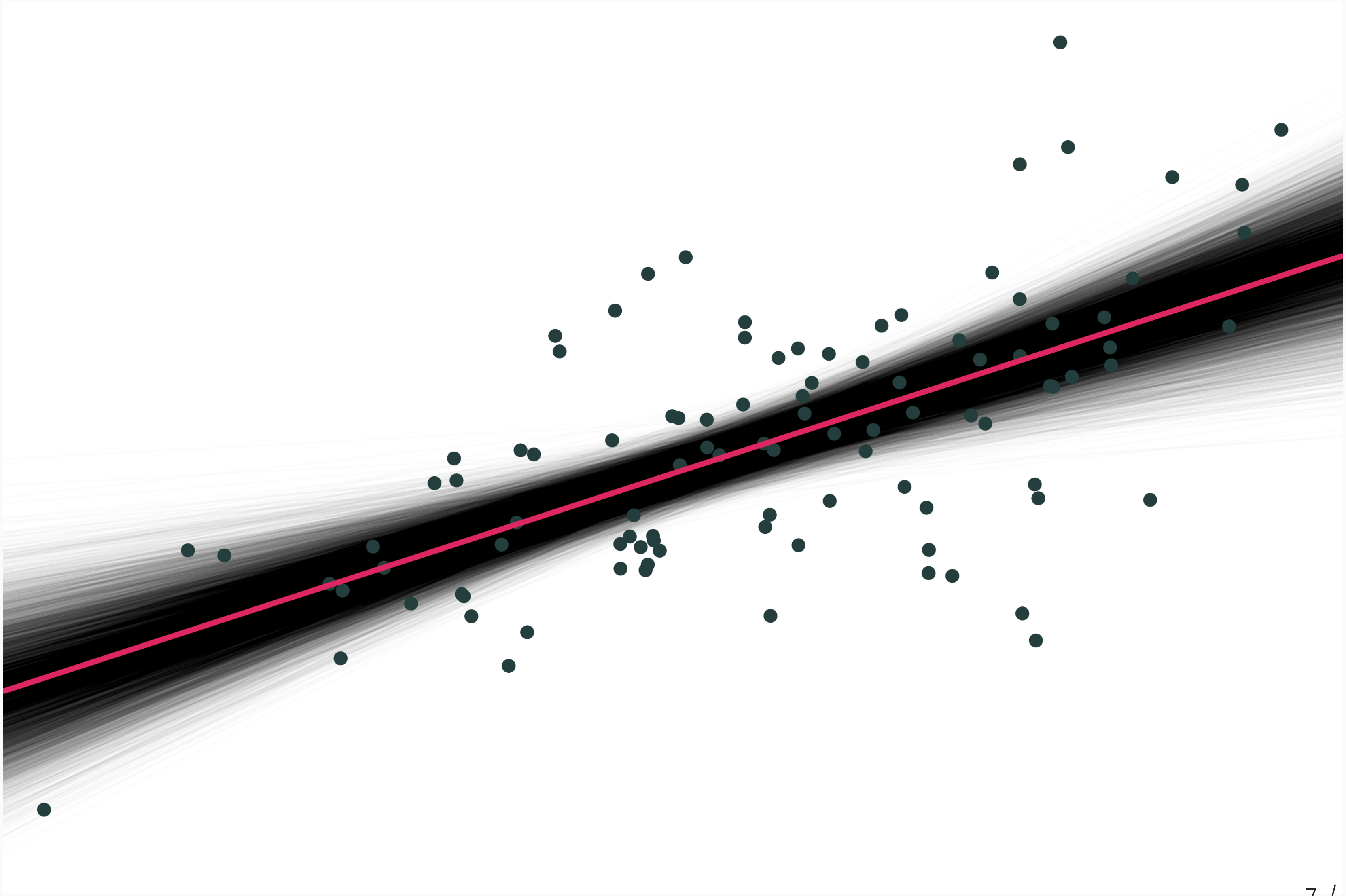
Relación muestral

$$\hat{y}_i = 3.41 + 0.42x_i$$

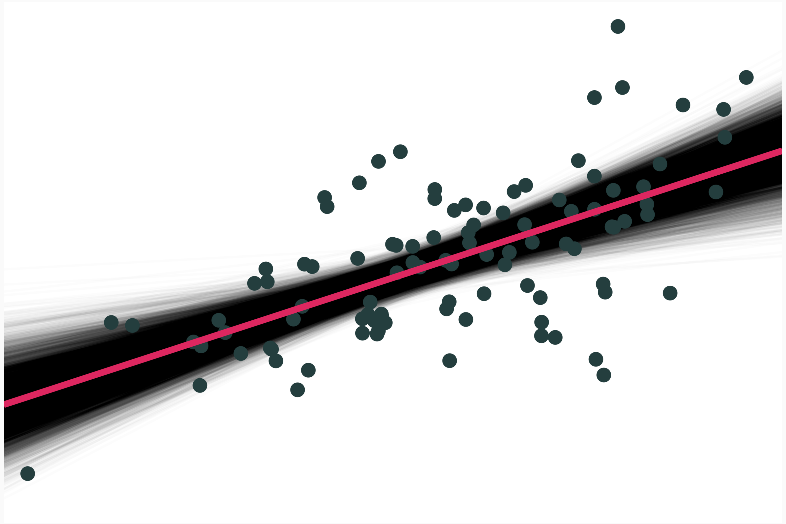
Vamos a repetir esto **10.000 veces**.

(Este ejercicio se llama simulación Monte Carlo.)

Población *vs.* muestra



Población *vs.* muestra



- En **promedio**, nuestras líneas de regresión coinciden muy bien con la línea poblacional.
- Sin embargo, **líneas individuales** (muestras) realmente pueden fallar.
- Las diferencias entre las muestras individuales y la población generan **incertidumbre** para el o la econometrista.

Población *vs.* muestra

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

Respuesta: La incertidumbre importa.

$\hat{\beta}$ en sí mismo es una variable aleatoria, que depende de la muestra aleatoria. Cuando tomamos una muestra y corremos una regresión, no sabemos si es una 'buena' muestra, $\hat{\beta}$ está cerca de β, o una 'mala muestra', nuestra muestra difiere mucho de la población.

Población *vs.* muestra

Incertidumbre

Hacer un seguimiento de esta incertidumbre será clave en el curso.

- Estimación de errores estándar para nuestras estimaciones.
- Prueba de hipótesis.
- Corrección de heterocedasticidad y autocorrelación.

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

El estimador

Podemos estimar la línea de regresión en R (`lm(y ~ x, my_data)`) pero... ¿de dónde vienen estas estimaciones?

Unas diapositivas atrás:

> $$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
>
> lo que nos da la línea que *mejor se ajusta* a nuestra muestra (set de datos).

¿Pero a qué nos referimos con la "línea que mejor se ajusta"?

Siendo el "mejor"

Pregunta: ¿A qué nos referimos con la *línea que mejor se ajusta*?

Respuestas:

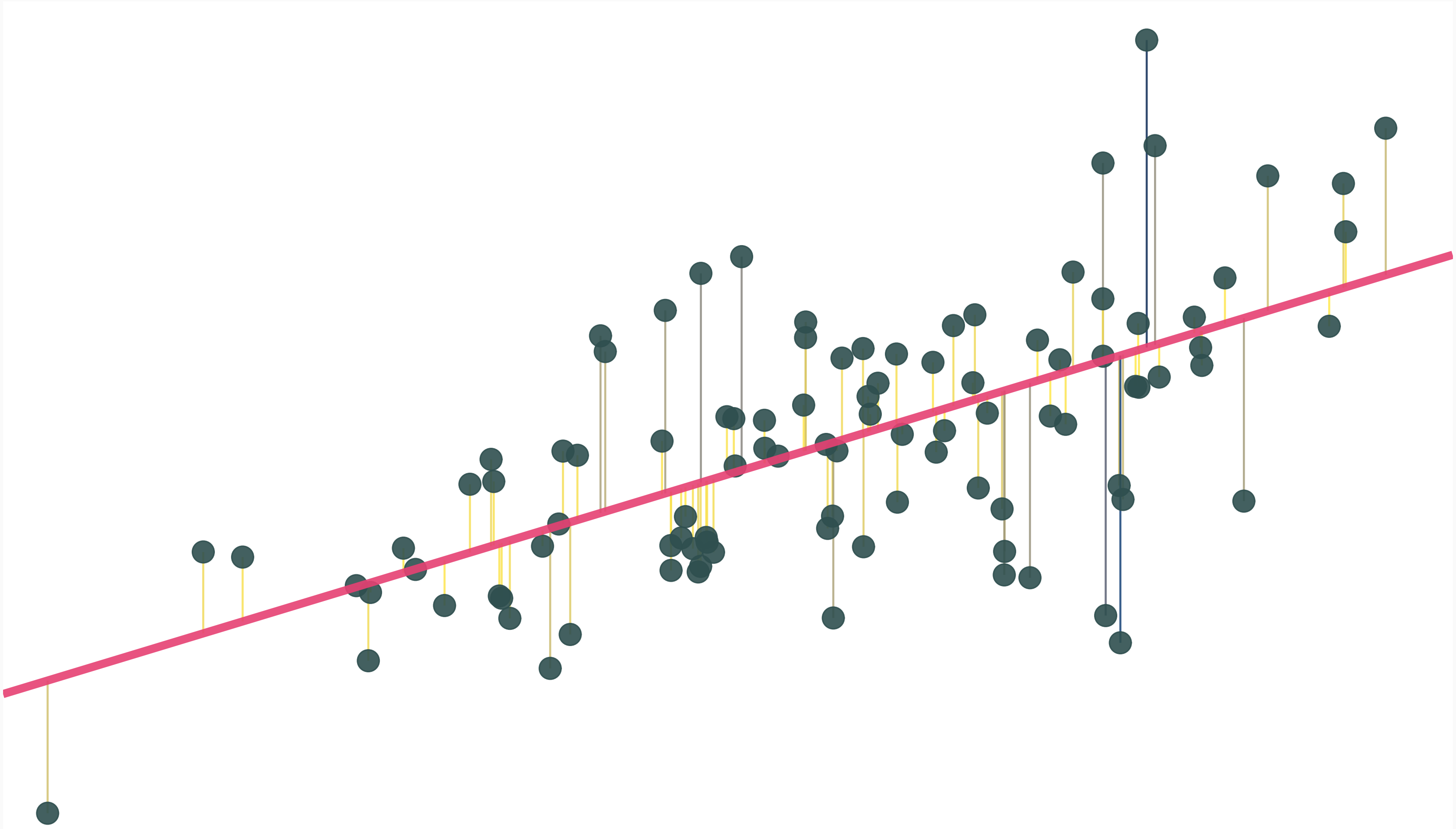
- En general (en econometría), la *línea que mejor se ajusta* significa la línea que minimiza la suma de minimizando la suma de errores o residuos al cuadrado (SCR):

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

- **Mínimos Cuadrados** Ordinarios (**MCO**) minimiza la suma de errores o residuos al cuadrado.
- Basado en un conjunto de supuestos (en su mayoría aceptables), MCO
    - Es insesgado (y consistente)
    - Es el *mejor* (mínima varianza) estimador lineal insesgado (MELI)

MCO *vs.* otras líneas/estimadores

El estimador de MCO es una combinación de $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimizan la SCR.



MCO

Formalmente

En una regresión lineal simple, el estimador de MCO proviene de escoger los $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimizan la suma de errores o residuos al cuadrado (SCR), *es decir*,

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \mathrm{SCR}$$

pero ya sabemos que $\mathrm{SCR} = \sum_i e_i^2$, ahora usamos las definiciones de e_i y $\hat{y}$.

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

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

MCO

Formalmente

Nos estamos acercando. Necesitamos **minimizar las SCR**. Hemos mostrado cómo SCR se relaciona con nuestra muestra (nuestros datos: $\boldsymbol{x}$ y $\boldsymbol{y}$) y nuestras estimaciones, *es decir*, $\hat{\beta}_0$ y $\hat{\beta}_1$.

$$\text{SCR} = \sum_i e_i^2 = \sum_i \left(y_i^2 - 2y_i\hat{\beta}_0 - 2y_i\hat{\beta}_1 x_i + \hat{\beta}_0^2 + 2\hat{\beta}_0\hat{\beta}_1 x_i + \hat{\beta}_1^2 x_i^2 \right)$$

Para las condiciones de primer orden de minimización, tomamos la primera derivada de la SCR con respecto $\hat{\beta}_0$ y $\hat{\beta}_1$.

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

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

MCO

Formalmente

Las condiciones de primer orden establecen que las derivadas son iguales a cero, entonces:

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

lo que implica que

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

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

Formalmente

Tomamos la derivada de SCR con respecto a $\hat{\beta}_1$

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

lo igualamos a cero (condiciones de primer orden, de nuevo 😅)

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

y sustituimos en nuestra relación por $\hat{\beta}_0$, *es decir*, $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$. Entonces,

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

Formalmente

Continuando de la diapositiva anterior

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

multiplicamos

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

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

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

Formalmente

¡Listo!

Ahora tenemos adorables estimadores de MCO para la pendiente

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

y el intercepto

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

Ahora sabemos de donde viene la parte de *mínimos cuadrados* de Mínimos Cuadrados Ordinarios 🤓🎊

Pasamos ahora a los supuestos y propiedades (implícitas) de MCO.

MCO: Supuestos y propiedades

Propiedades

Pregunta: ¿Qué propiedades nos podrían interesar para un estimador?

Tangente: Revisemos primero las propiedades estadísticas.

MCO: Supuestos y propiedades

Propiedades

Refrescando: Funciones de densidad

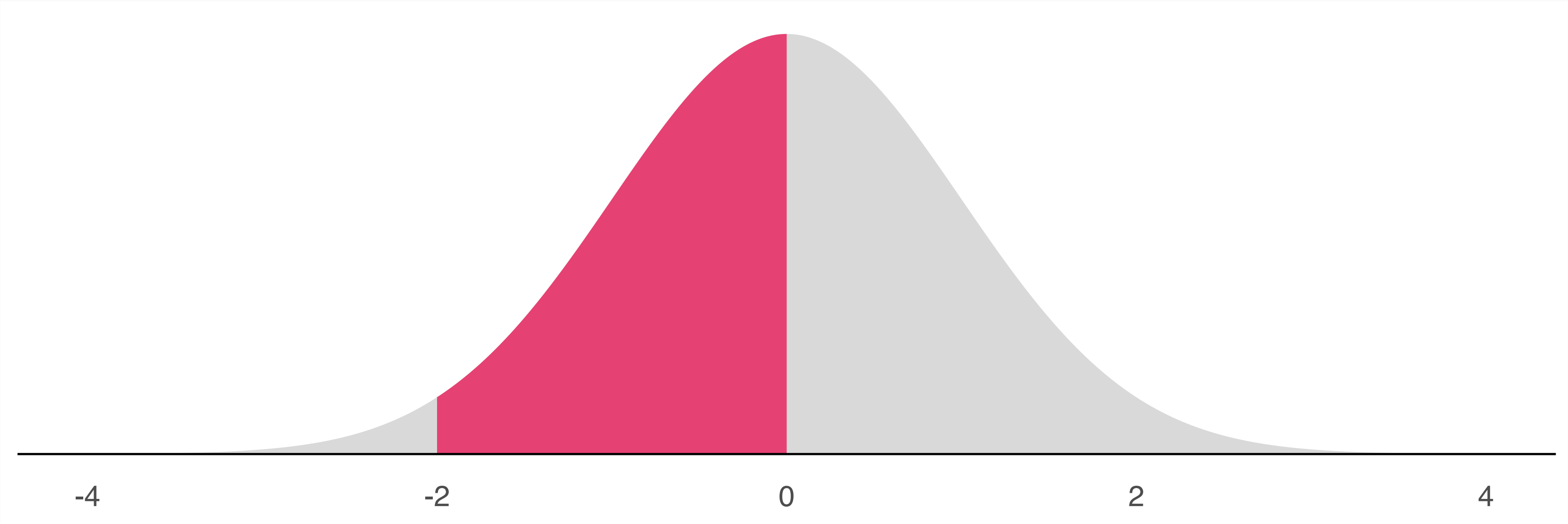
Recuerden que usamos **funciones de densidad de probabilidad** para describir la probabilidad que una **variable aleatoria continua** adopte un rango de valores. (El área total = 1)

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

Propiedades

Refrescando: Funciones de densidad

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

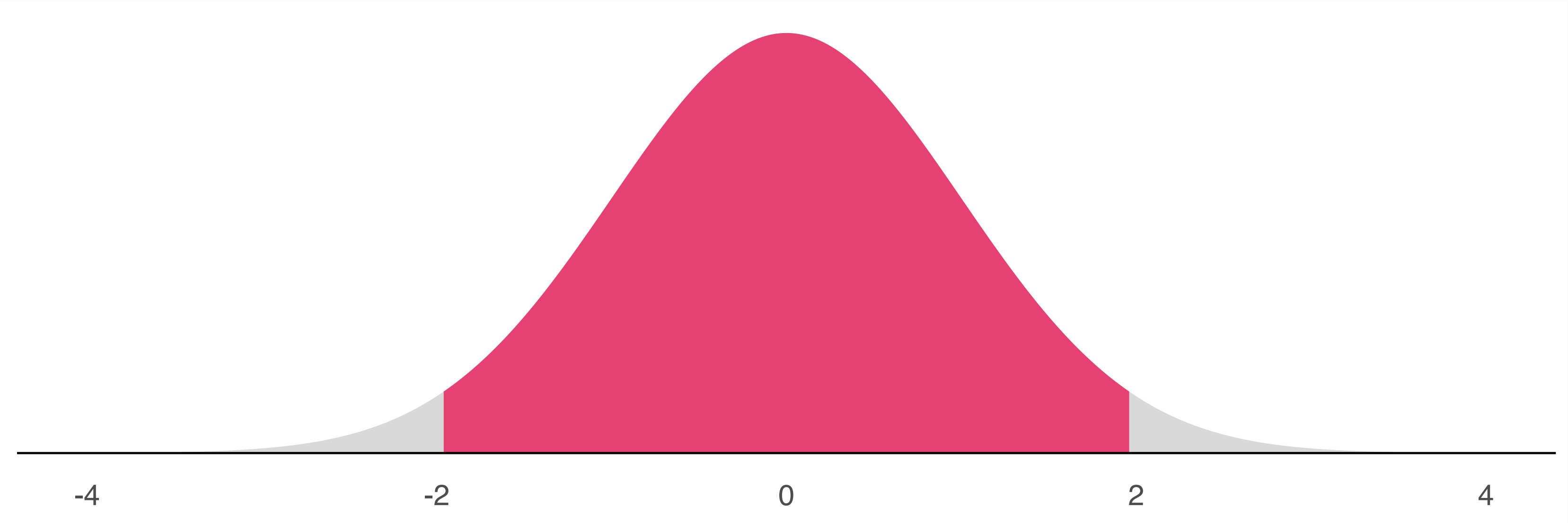


MCO: Supuestos y propiedades

Propiedades

Refrescando: Funciones de densidad

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



Propiedades

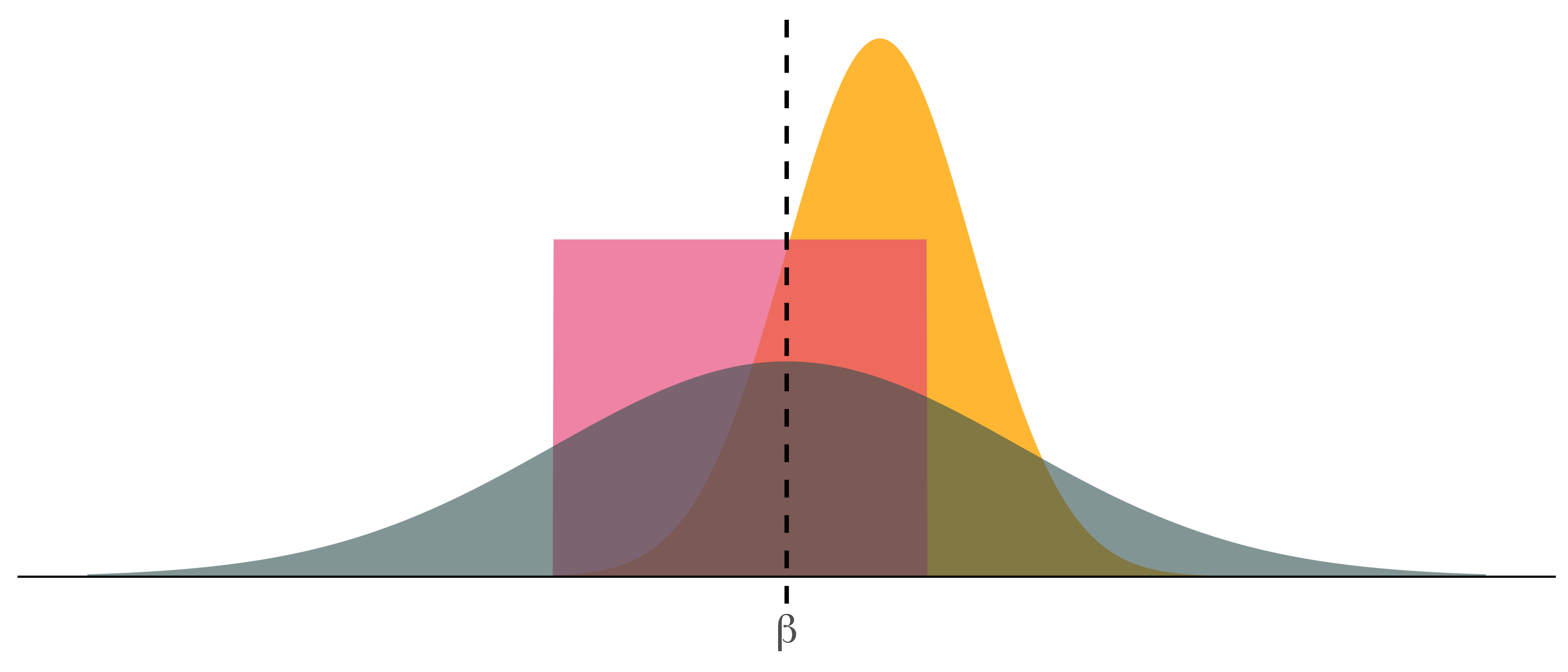
Refrescando: Funciones de densidad

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

MCO: Supuestos y propiedades

Propiedades

Imagine que estamos tratando de estimar un parámetro desconocido β, y conocemos las distribuciones de tres estimadores en competencia. ¿Cuál querríamos? ¿Cómo decidiríamos?



Propiedades

Pregunta: ¿Qué propiedades nos podrían interesar para un estimador?

Respuesta uno: Sesgo.

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

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

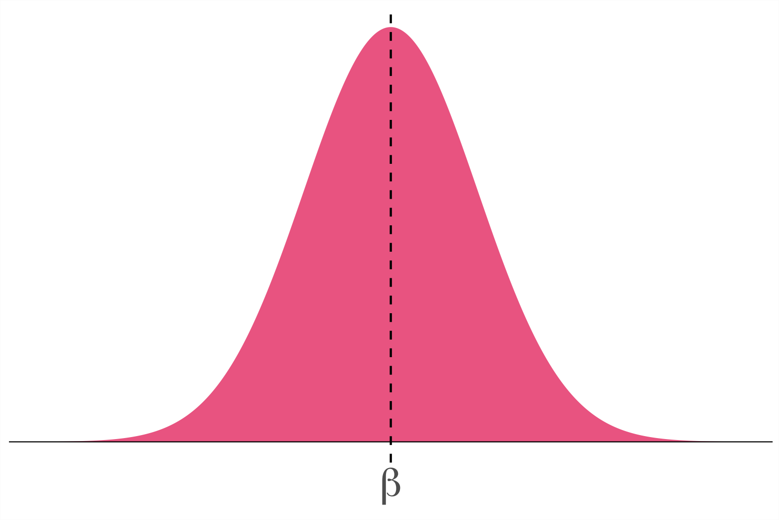
$$\underset{\beta}{\mathrm{Sesgo}}\left(\hat{\beta}\right) = \boldsymbol{E}\left[\hat{\beta}\right] - \beta$$

MCO: Supuestos y propiedades

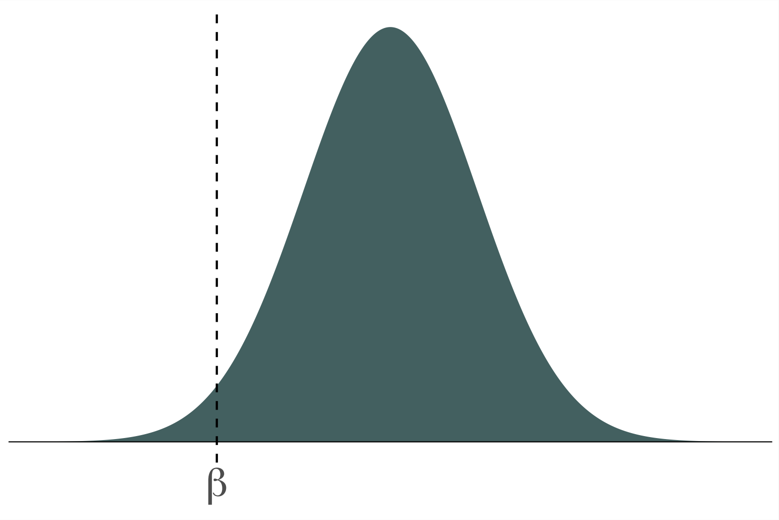
Propiedades

Respuesta uno: Sesgo.

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

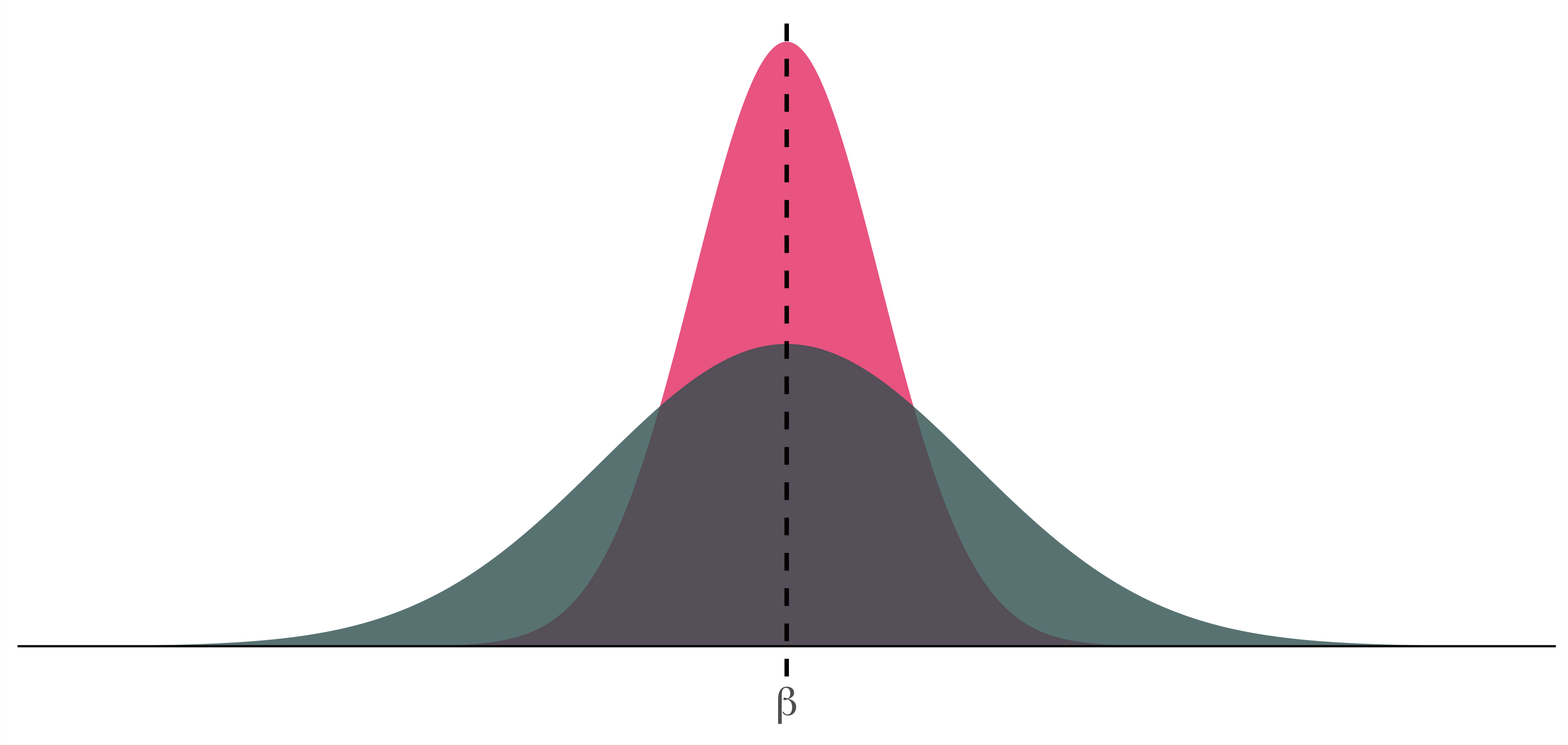


Estimador sesgado: $\boldsymbol{E}\left[\hat{\beta}\right] \neq \beta$



Propiedades

Respuesta dos: Varianza.



Propiedades

Respuesta uno: Sesgo.

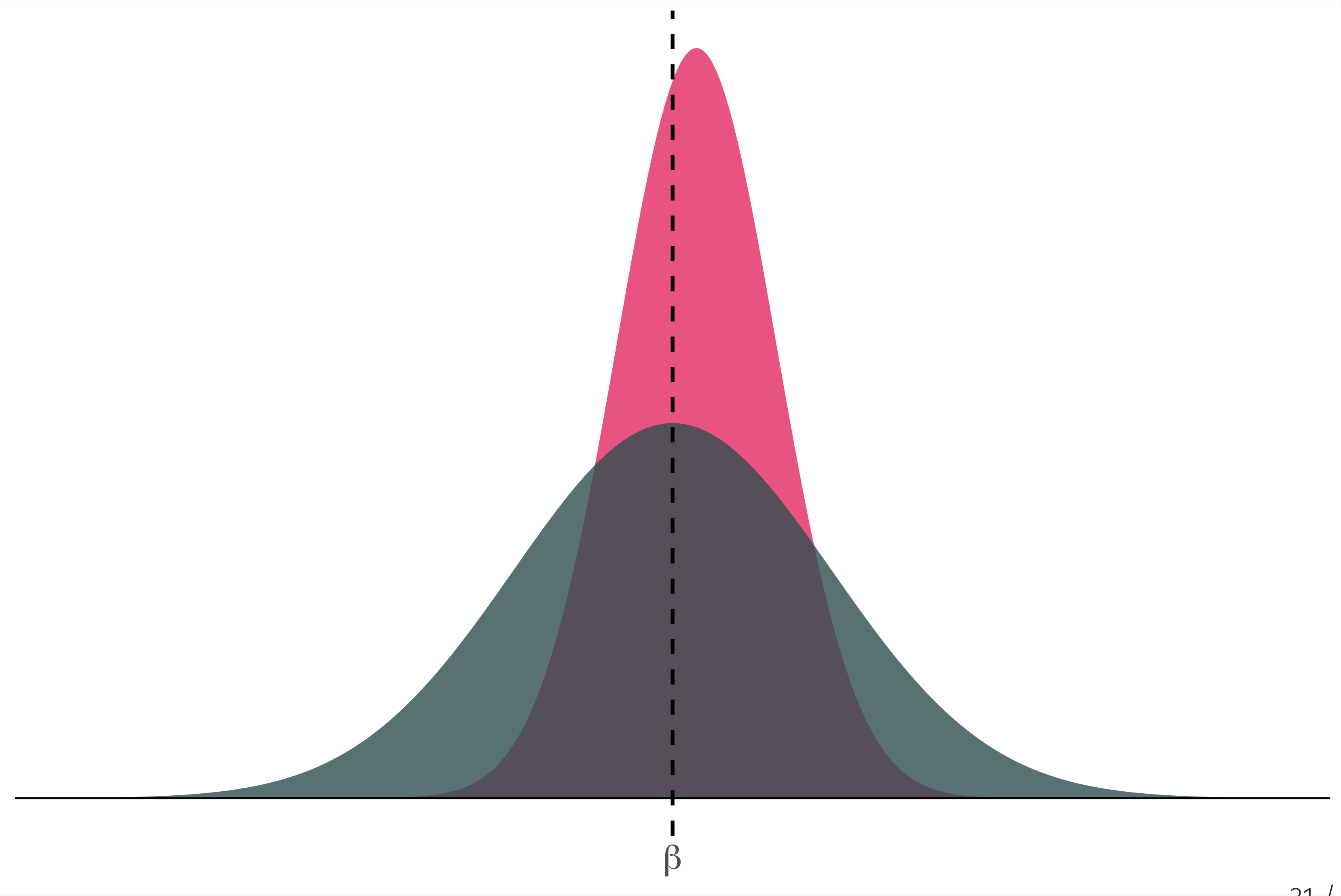
Respuesta dos: Varianza.

Sútilmente: El trade-off sesgo-varianza.

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

En econometría, generalmente nos apegamos a estimadores insesgados (o consistentes). Pero otras disciplinas (especialmente las ciencias de la computación) piensan un poco más en esta compensación.

El trade-off sesgo-varianza.



MCO: Supuestos y propiedades

Propiedades

Como ya habrán adivinado,

- MCO es **insesgado**.
- MCO tiene la **varianza mínima** de todos los estimadores lineales insesgados.

MCO: Supuestos y propiedades

Propiedades

Pero… estas (muy buenas) propiedades dependen de un conjunto de supuestos:

1. La relación de población es lineal en parámetros con una perturbación aditiva.
2. Nuestra variable $\boldsymbol{X}$ es **exógena**, *es decir*, $\boldsymbol{E}[u \mid X] = 0$.
3. La variable $\boldsymbol{X}$ tiene variación. Y si hay múltiples variables explicativas, no son perfectamente colineales.
4. Las perturbaciones de la población $\boldsymbol{u_i}$ se distribuyen de forma independiente e idéntica como variables aleatorias normales con una media de cero $\left(\mathop{\boldsymbol{E}} \left[u \right] = 0 \right) $ y varianza $\boldsymbol{\sigma^2}$ (*es decir*, $\boldsymbol{E}\left[u^2\right] = \sigma^2$). Independientemente distribuidos y significan cero conjuntamente implican $\boldsymbol{E}\left[u_i u_j\right] = 0$ para cualquier $\boldsymbol{i \neq j}$.

MCO: Supuestos y propiedades

Supuestos

Diferentes supuestos garantizan diferentes propiedades:

- Supuestos (1), (2), y (3) hace al insesgado MCO.
- Supuesto (4) nos da un estimador insesgado para la varianza de nuestro estimador MCO.

Durante nuestro curso, discutiremos las muchas formas en que la vida real puede **violar estas suposiciones**. Por ejemplo:

- Relaciones no lineales en nuestros parámetros/perturbaciones (o errores de especificación).
- Perturbaciones que no se distribuyen de forma idéntica y/o no son independientes.
- Violaciones de exogeneidad (especialmente sesgo de variable omitida).

Expectativa condicional

Para muchas aplicaciones, nuestro supuesto más importante es **exogeneidad**, *es decir*,

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

¿pero qué significa esto?

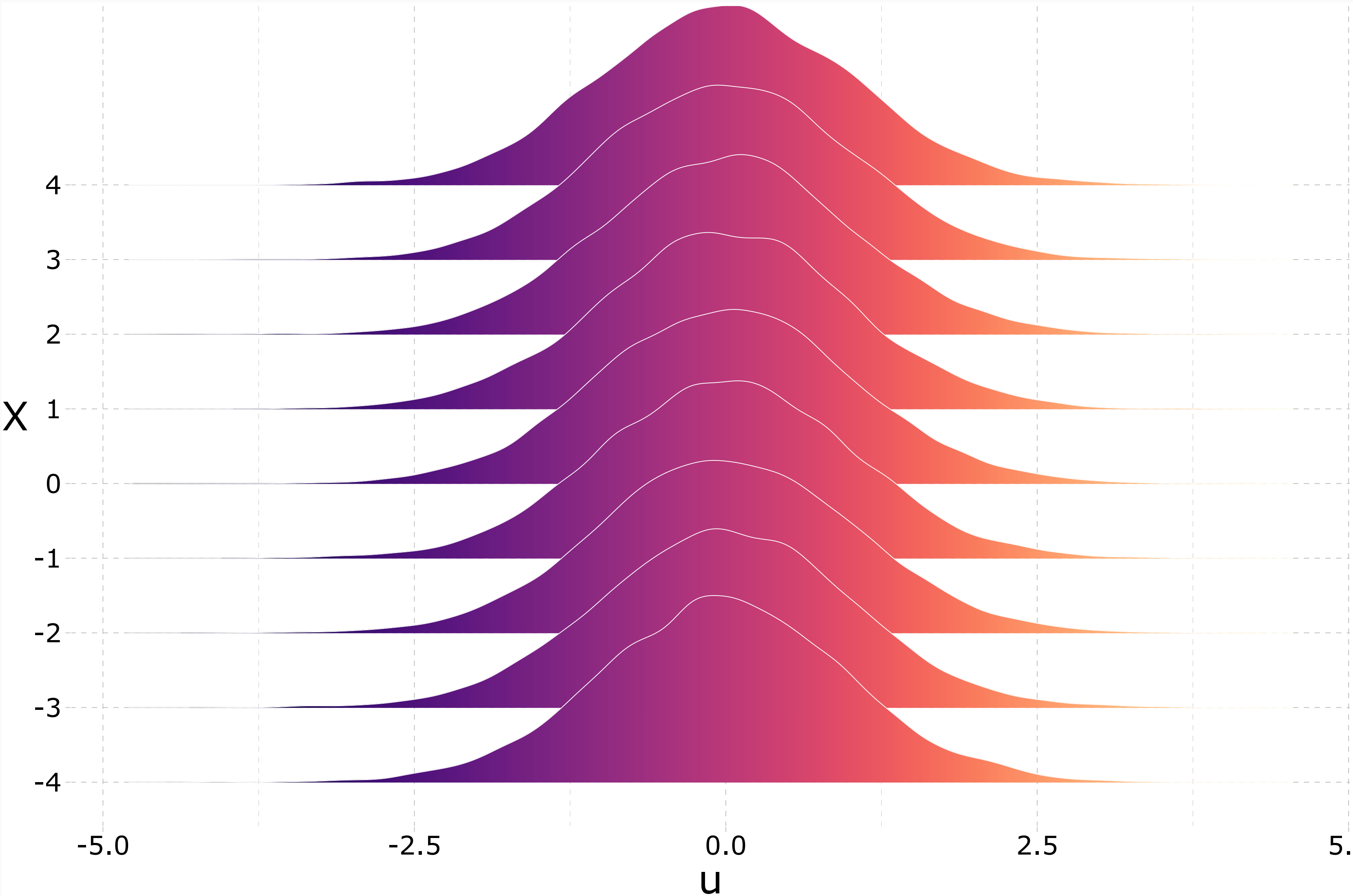
Una forma de pensar en esta definición:

> Para *cualquier* valor de X, la media de los residuos debe ser cero.

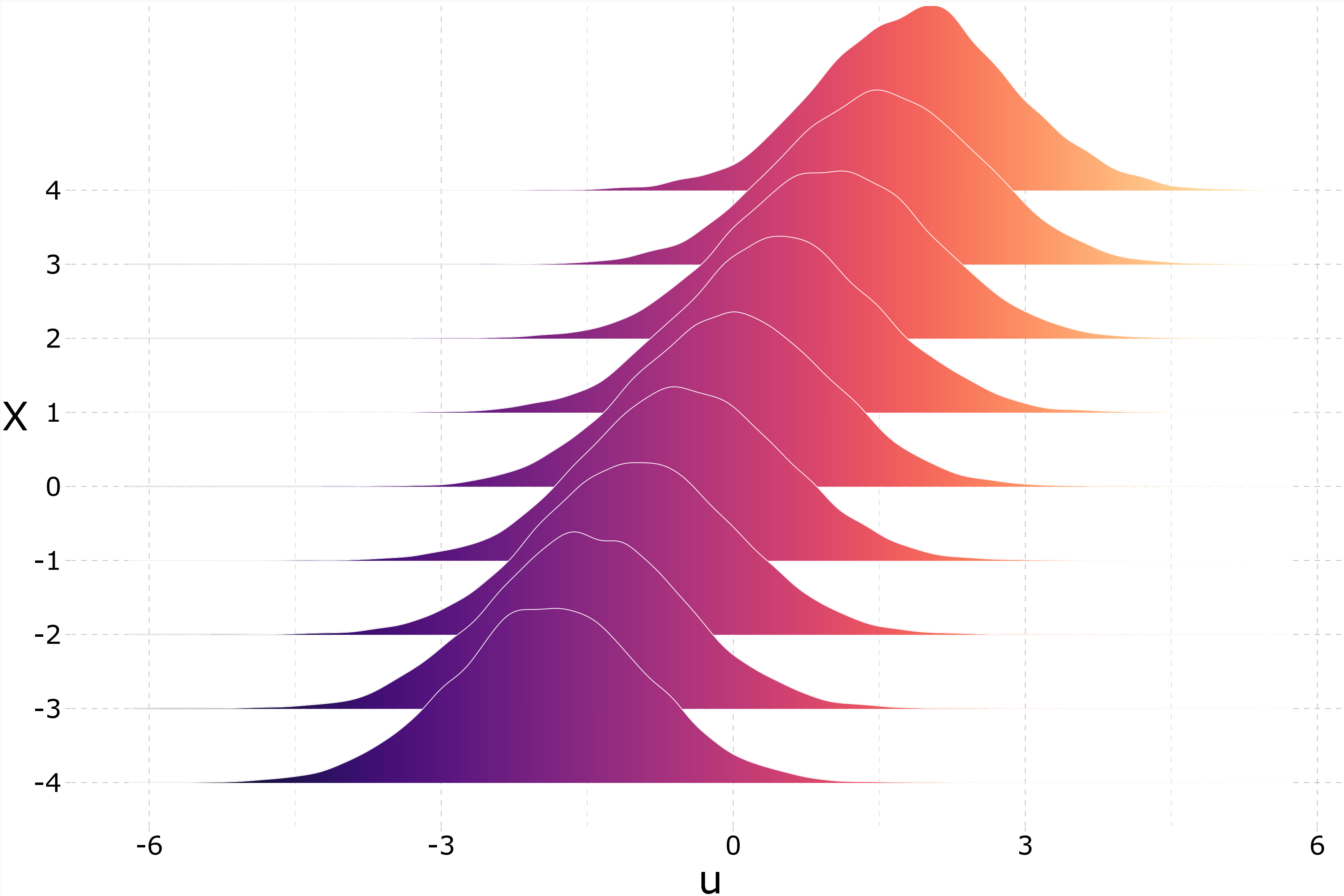
- *Por ejemplo*, $E[u \mid X = 1] = 0$ *y* $E[u \mid X = 100] = 0$
- *Por ejemplo*, $E[u \mid X_2 = \text{Mujer}] = 0$ *y* $E\left[u \mid X_2 = \text{Varón}\right] = 0$
- Aviso: $E[u \mid X] = 0$ es más restrictiva que $E[u] = 0$

Gráficamente...

La exogeneidad válida, *es decir*, $E[u \mid X] = 0$



La exogeneidad inválida, *es decir,* $\boldsymbol{E[u \mid X] \neq 0}$



Incertidumbre e inferencia

Incertidumbre e inferencia

¿Hay algo más?

Hasta este punto, sabemos que MCO tiene algunas propiedades interesantes, y sabemos cómo estimar un coeficiente de pendiente de intersección y a través de MCO.

Nuestro flujo de trabajo actual:

- Obtener datos (puntos con valores x y y)
- Regresar y en x
- Trace la línea MCO (*es decir*, $\hat {y} = \hat {\beta}_0 + \hat{\beta}_1 $)
- ¿Hecho?

Pero, ¿cómo podemos **aprender** algo de este ejercicio?

Hay más

Pero, ¿cómo podemos **aprender** algo de este ejercicio?

- Con base en nuestro valor de $\hat{\beta}_1$, ¿podemos descartar valores hipotetizados previamente?
- ¿Qué confianza debemos tener en la precisión de nuestras estimaciones?
- ¿Qué tan bien explica nuestro modelo la variación que observamos en y?

Necesitamos poder lidiar con la incertidumbre. Ingresa en escena: **La inferencia.**

Aprendiendo de nuestros errores

Como señaló nuestra simulación anterior, nuestro problema con la **incertidumbre** es que no sabemos si nuestra estimación muestral está *cerca* o *lejos* del parámetro de población desconocido. †

Sin embargo, no todo está perdido. Podemos usar los errores $\left(e_i = y_i - \hat{y}_i \right)$ para tener una idea de qué tan bien nuestro modelo explica la variación observada en y.

Cuando nuestro modelo parece estar haciendo un trabajo "agradable", es posible que tengamos un poco más de confianza al usarlo para conocer la relación entre y y x.

Ahora solo tenemos que formalizar lo que realmente significa un "buen trabajo".

.pink [†]: Excepto cuando corremos la simulación nosotros mismos, por eso nos gustan las simulaciones 😉

Aprendiendo de nuestros errores

En primer lugar, estimaremos la varianza de u_i (recuerde: $\mathbf{Var}(u_i) = \sigma^2$) usando nuestros errores al cuadrado, *es decir*,

$$s^2 = \frac{\sum_i e_i^2}{n-k}$$

donde k da el número de términos de pendiente y intersecciones que estimamos (*por ejemplo*, β_0 y β_1 daría $k = 2$).

s^2 es un estimador insesgado de $sigma^2$.

Aprendiendo de nuestros errores

Luego demostramos que la varianza de $\hat{\beta}_1$ (para regresión lineal simple) es

$$\operatorname{Var}\left(\hat{\beta}_1\right) = \frac{s^2}{\sum_i \left(x_i - \overline{x}\right)^2}$$

lo que muestra que la varianza de nuestro estimador de pendiente:

1. aumenta a medida que nuestras perturbaciones se vuelven más ruidosas
2. disminuye a medida que aumenta la varianza de x

Aprendiendo de nuestros errores

Más comúnmente: El **error estándar** de $\hat{\beta}_1$

$$\hat{\text{EE}}\left(\hat{\beta}_1\right) = \sqrt{\frac{s^2}{\sum_i \left(x_i - \overline{x}\right)^2}}$$

Recuerden: El error estándar de un estimador es la desviación estándar de la distribución del estimador.

Aprendiendo de nuestros errores

El error estándar en R's `lm`, se ve así:

```
tidy(lm(y ~ x, pop_df))
```

```
> # A tibble: 2 x 5
>   term        estimate std.error statistic  p.value
>   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
> 1 (Intercept)    2.95     0.314       9.40 2.37e-15
> 2 x              0.495    0.0615      8.06 1.88e-12
```

Interpretando coeficientes

Interpretando coeficientes

Variables continuas

Consideramos la siguiente relación:

$$\textbf{salario}_i = \beta_0 + \beta_1 \, \textbf{educación}_i + u_i$$

donde

- $\textbf{salario}_i$ es una variable continua que mide el salario de cada individuo
- $\textbf{educación}_i$ es una variable continua que mide los años de educación de cada persona

Interpretaciones

- β_0: es el intercepto de y, *es decir*, el $\textbf{salario}$ cuando $\textbf{educación} = 0$
- β_1: el aumento esperado en el $\textbf{salario}$ para un aumento unitario de $\textbf{educación}$

Interpretando coeficientes

Variables continuas

Considerando el siguiente modelo:

$$y = \beta_0 + \beta_1 \, x + u$$

Diferencio el modelo:

$$\frac{dy}{dx} = \beta_1$$

Es decir, la pendiente nos dice el aumento esperado en la variable dependiente para un aumento en una unidad de la variable explicativa, **manteniendo todas las demás variables constantes** (*ceteris paribus*).

Interpretando coeficientes

Variables binarias

Considero la relación:

$$\textbf{salario}_i = \beta_0 + \beta_1 \, \textbf{mujer}_i + u_i$$

donde

- $\textbf{salario}_i$ es una variable continua que mide el salario de cada individuo
- $\textbf{mujer}_i$ es un variable binaria que toma el valor 1 cuando i es mujer

Interpretaciones

- β_0: el $\textbf{salario}$ esperado para los hombres (cuando $\textbf{mujer} = 0$)
- β_1: la diferencia esperada en el $\textbf{salario}$ entre hombres y mujeres
- $\beta_0 + \beta_1$: el $\textbf{salario}$ esperado para mujeres

Interpretando coeficientes

Variables binarias

Derivación:

$$\begin{aligned}\boldsymbol{E}[\text{salario}|\text{hombre}] &= \boldsymbol{E}[\beta_0 + \beta_1 \times 0 + u_i] \\ &= \boldsymbol{E}[\beta_0 + 0 + u_i] \\ &= \beta_0\end{aligned}$$

$$\begin{aligned}\boldsymbol{E}[\text{salario}|\text{mujer}] &= \boldsymbol{E}[\beta_0 + \beta_1 \times 1 + u_i] \\ &= \boldsymbol{E}[\beta_0 + \beta_1 + u_i] \\ &= \beta_0 + \beta_1\end{aligned}$$

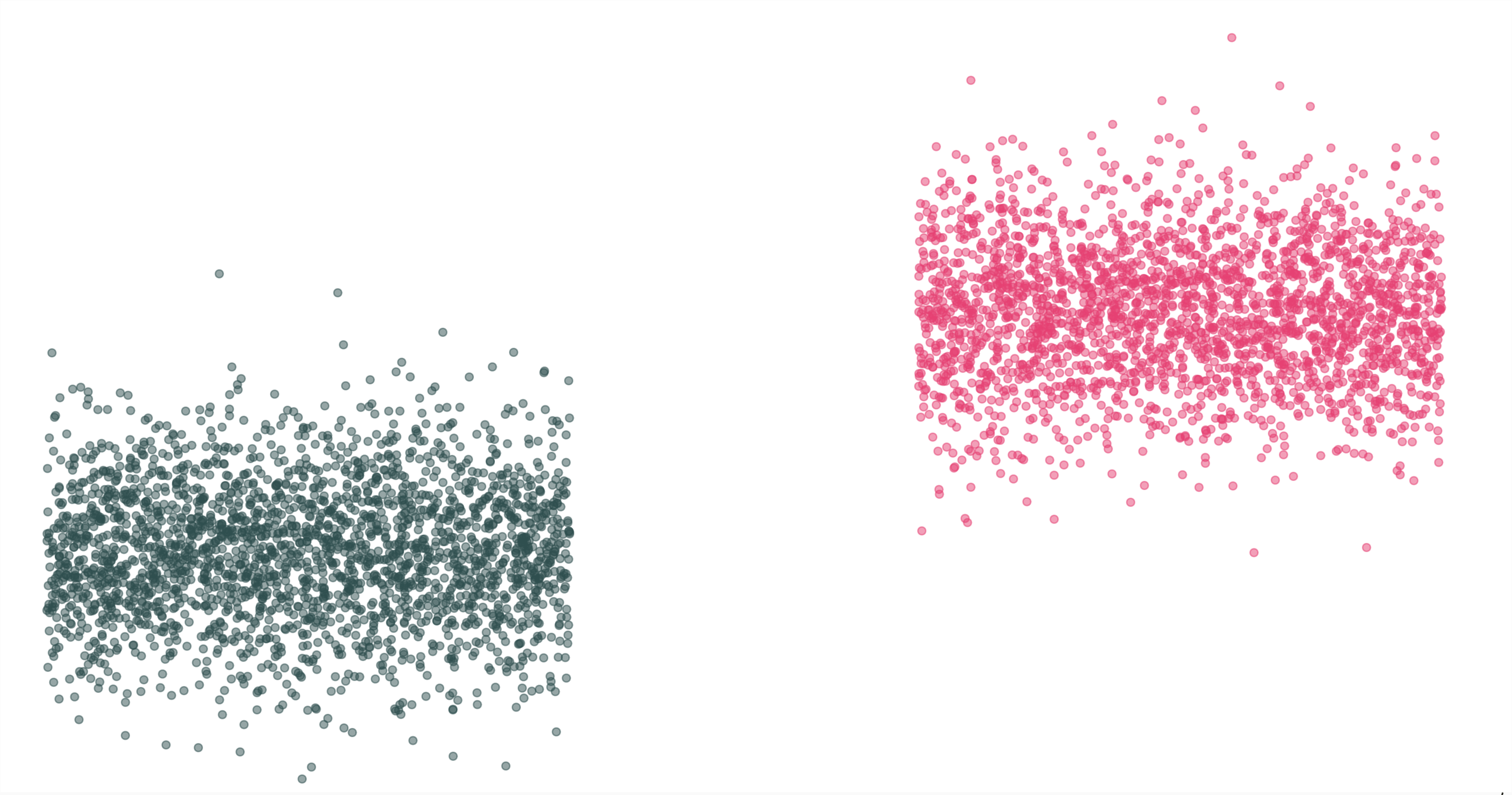
Nota: Si no hay variables adicionales de control, entonces $\hat{\beta}_1$ equivale a la diferencia entre las medias de los grupos, *en este caso*, $\overline{x}_{\text{mujer}} - \overline{x}_{\text{hombre}}$.

Nota$_2$: El *manteniendo todo lo demás constante* también aplica en regresiones que tienen variables binarias.

Interpretando coeficientes

Variables binarias

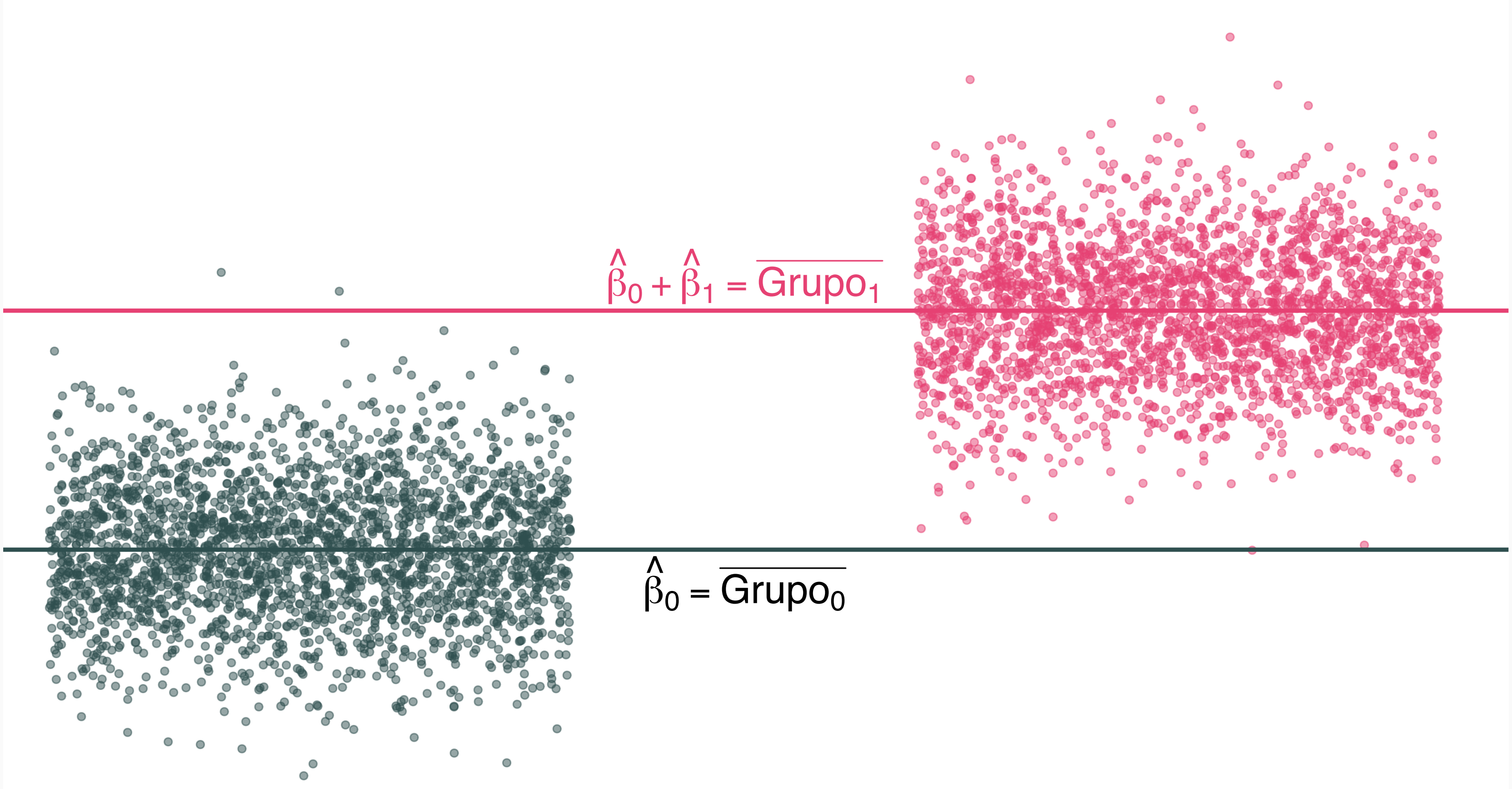
$y_i = \beta_0 + \beta_1 x_i + u_i$ para la variable binaria $x_i = \{0,\, 1\}$



Interpretando coeficientes

Variables binarias

$y_i = \beta_0 + \beta_1 x_i + u_i$ para la variable binarias $x_i = \{0, 1\}$



Interpretando coeficientes

Interacciones

Las interacciones permiten que el efecto de una variable cambie según el nivel de otra variable.

Ejemplos

1. ¿Cambia el efecto de la escolarización en el salario por sexo?
2. ¿El efecto del género en el salario cambia según la etnia?
3. ¿Cambia el efecto de la escolarización en el salario según la experiencia?

Interpretando coeficientes

Interacciones

Anteriormente, considerábamos un modelo que permitía a mujeres y hombres tener salarios diferentes, pero el modelo asumía que el efecto de la escuela sobre el salario era el mismo para todos:

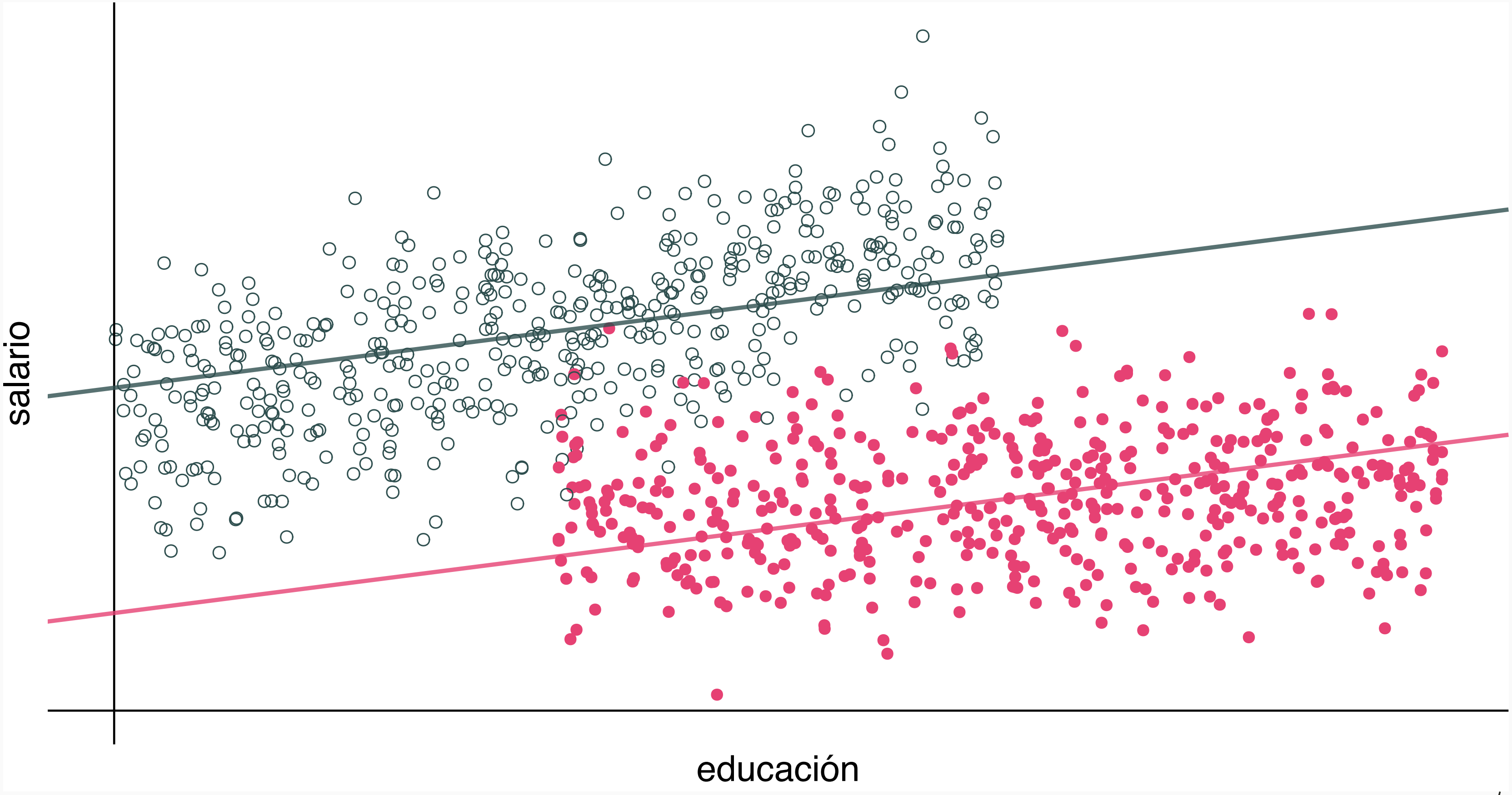
$$\textbf{salario}_i = \beta_0 + \beta_1\,\textbf{educación}_i + \beta_2\,\textbf{mujer}_i + u_i$$

pero también podemos permitir que el efecto de la escuela varíe según el sexo:

$$\textbf{salario}_i = \beta_0 + \beta_1\,\textbf{educación}_i + \beta_2\,\textbf{mujer}_i + \beta_3\,\textbf{educación}_i \times \textbf{mujer}_i + u_i$$

Interacciones

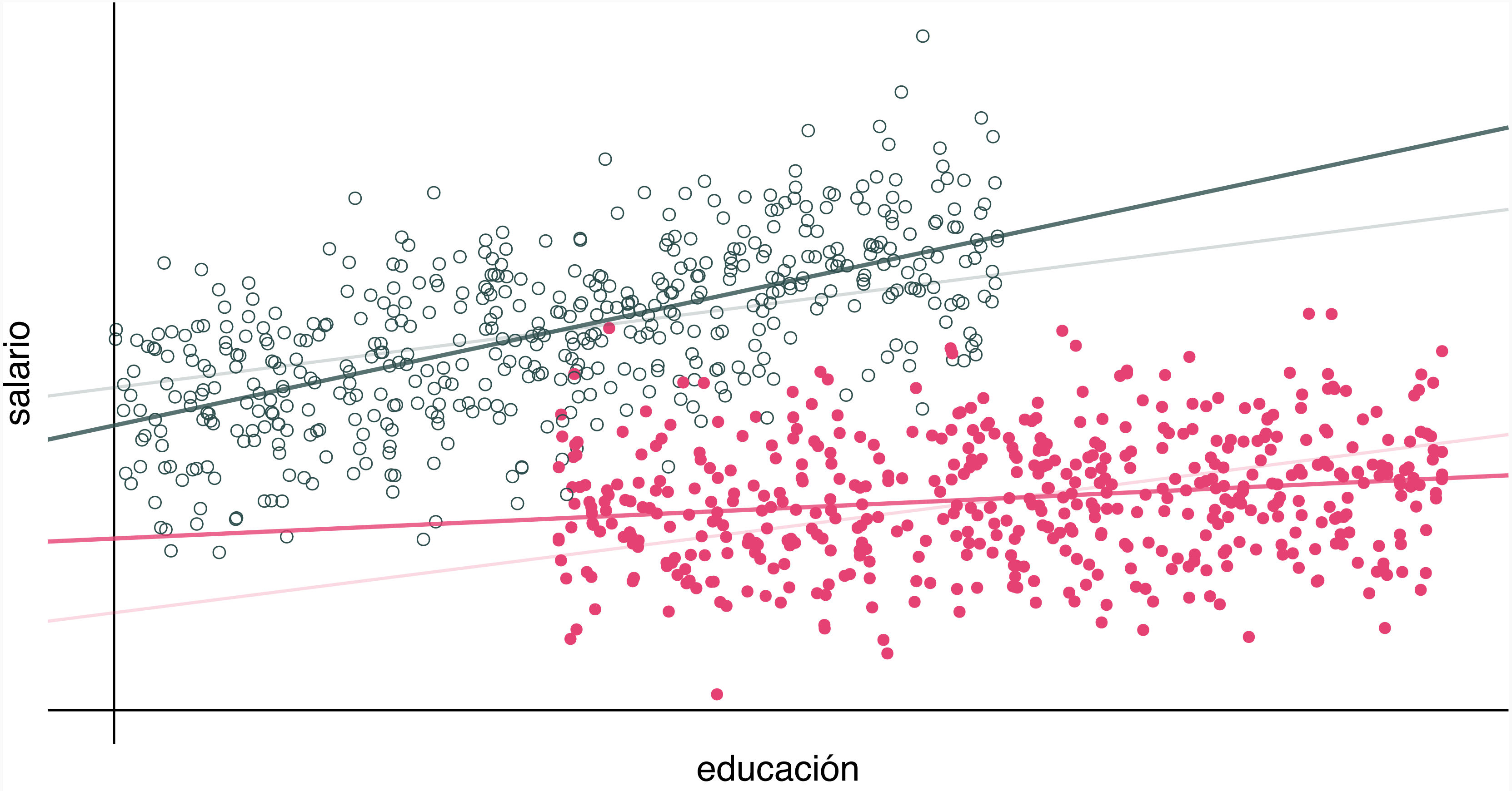
El modelo donde la escolarización tiene el mismo efecto para todos (**M** y **H**):



Interpretando coeficientes

Interacciones

El modelo donde el efecto de la educación puede variar por sexo (**M** y **H**):



Especificación log-lineal

En economía, con frecuencia se emplean variables explicativas en logaritmos, por ejemplo,

$$\log(\text{salario}_i) = \beta_0 + \beta_1\, \textbf{educación}_i + u_i$$

Esta especificación cambia nuestra interpretación de los coeficientes de pendiente.

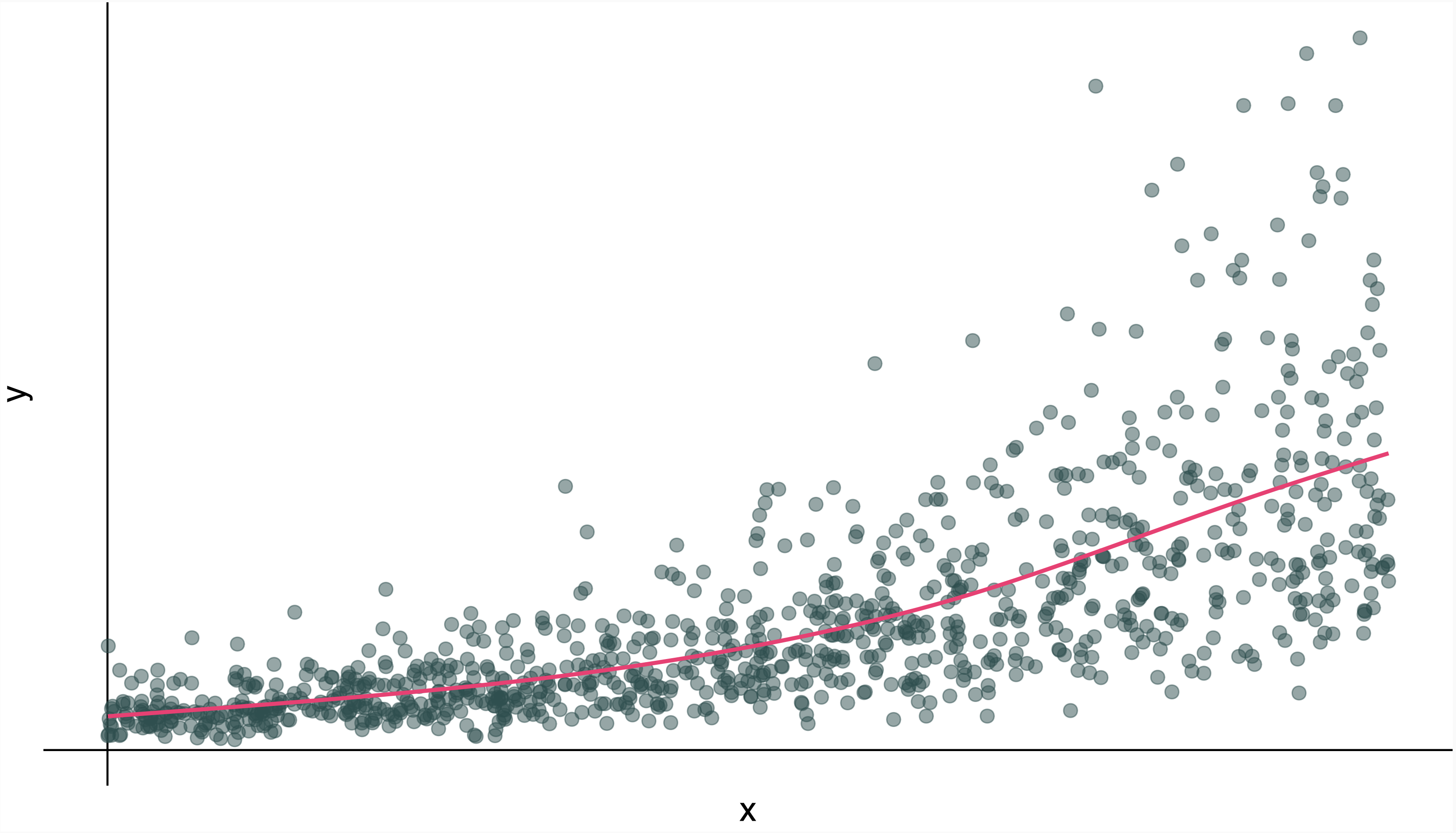
Interpretación

- Un aumento de una unidad en nuestra variable explicativa aumenta la variable de resultado en aproximadamente β_1 veces 100 por ciento.

- *Ejemplo:* Un año adicional de educación aumenta el salario en aproximadamente un 3 por ciento (para $\beta_1 = 0.03$).

Interpretando coeficientes

Especificación log-lineal



Interpretando coeficientes

Especificación log-log

De manera similar, los econometristas emplean con frecuencia modelos log-log, en los que y se está en logaritmos y al menos una variable explicativa, también:

$$\log(\text{salario}_i) = \beta_0 + \beta_1 \log(\text{educación}_i) + u_i$$

Interpretación:

- Un aumento del uno por ciento en x dará lugar a un cambio porcentual de β_1 en y.
- Se interpreta como una elasticidad.

Especificación log-log

Derivación

Consideramos el siguiente modelo log-log:

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

y diferenciamos

$$\frac{dy}{y} = \beta_1 \frac{dx}{x}$$

que dice que para un aumento del uno por ciento en x, veremos un aumento del β_1 por ciento en y. Como elasticidad:

$$\frac{dy}{dx}\frac{x}{y} = \beta_1$$

Interpretando coeficientes

Log-lineal con variable binaria

Nota: Si tenemos un modelo log-lineal con una variable binaria, la interpretación del coeficiente de esa variable cambia.

Consideramos:

$$\log(y_i) = \beta_0 + \beta_1 x_1 + u_i$$

siendo x_1 una variable binaria.

La interpretación de β_1 ahora es:

- Cuando x_1 cambia de 0 a 1, y cambiará en $100 \times \left(e^{\beta_1} - 1\right)$ por ciento.

Interpretando coeficientes

Resumen

Modelo	Interpretación
Nivel-nivel	Incremento de unidades en "y" cuando aumenta 1 unidad la "x" (ambas en sus unidades de medida originales)
Log-nivel	β*100 (incremento porcentual de "y" cuando aumenta una unidad la "x")
Nivel-log	β/100 (incremento en unidades de "y" cuando aumenta un 1% la "x")
Log-log	Incremento porcentual de "y" cuando aumenta un 1% la "x"