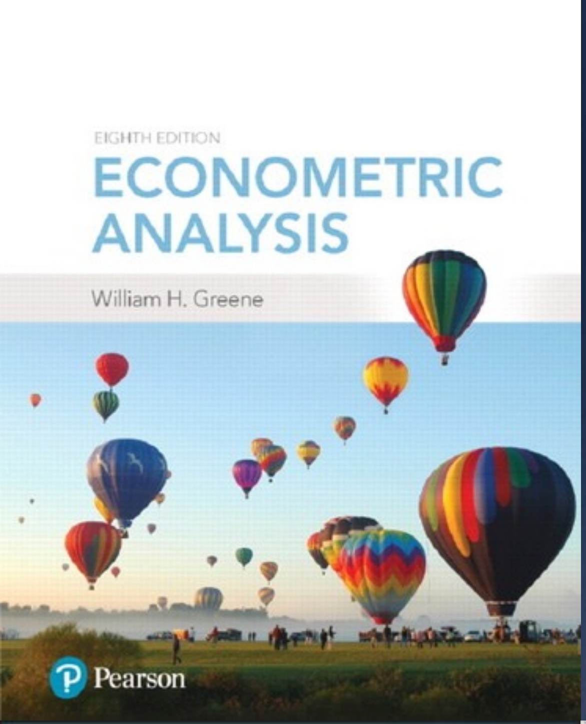
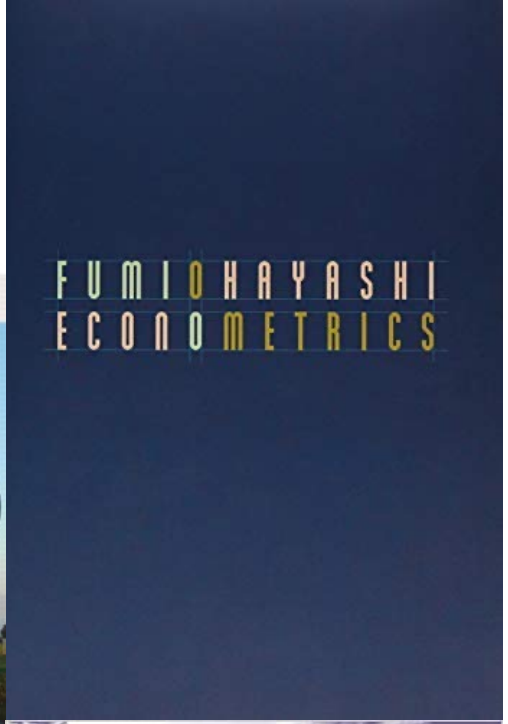
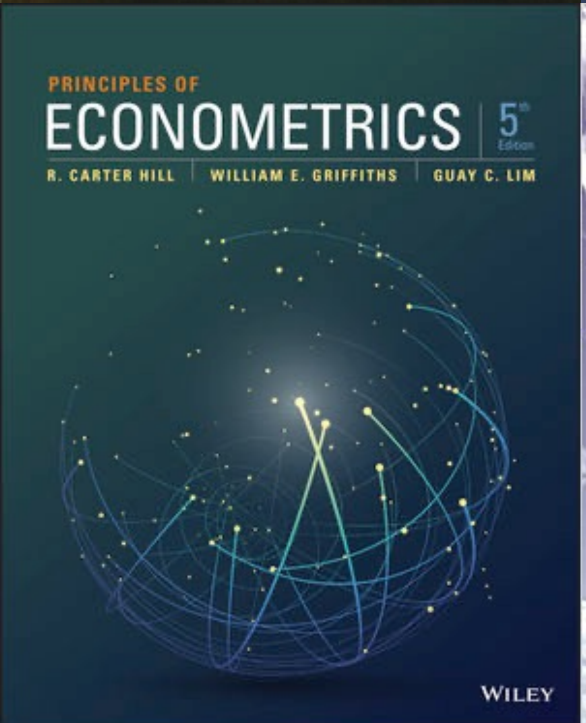
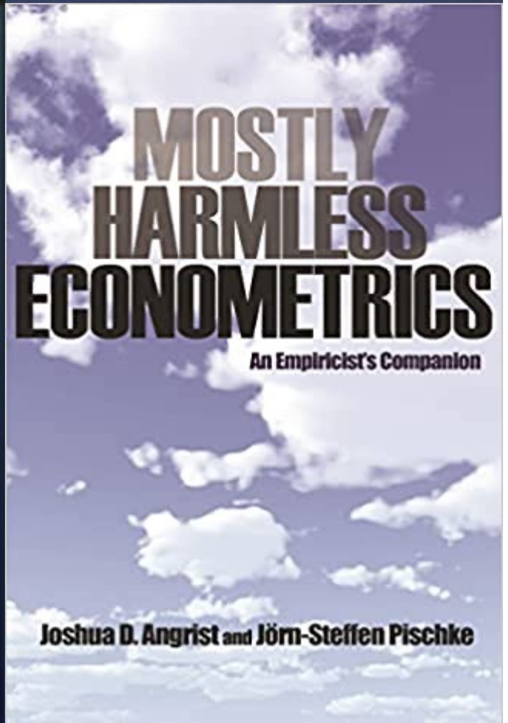
PUCP




**DEPARTAMENTO ECONOMÍA**
**FUNDAMENTOS DE ECONOMETRÍA**
**1ECO11 – HORARIO 0723**

# Sesión 2
# El Modelo de Regresión Lineal Bivariado

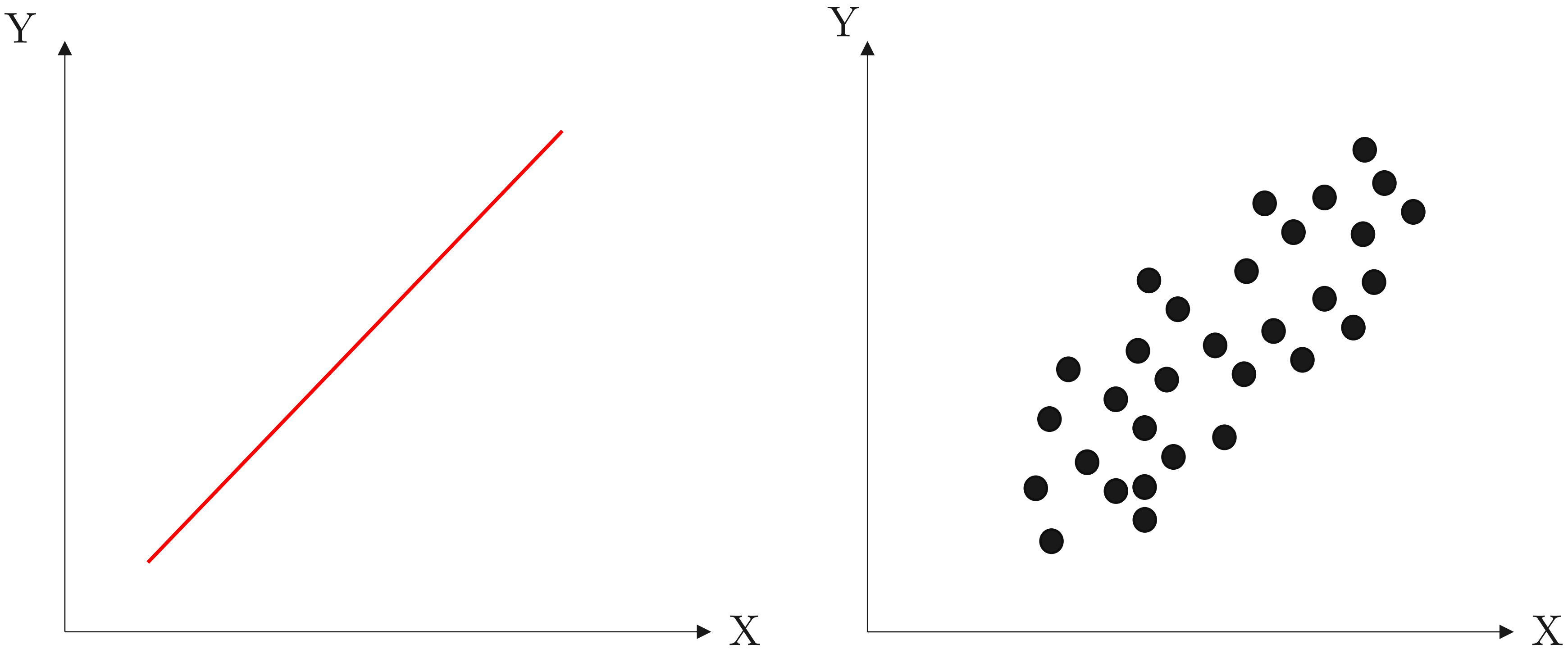
Docente: Juan Palomino

PUCP

# Índice

# 1. Análisis de Regresión

# Relaciones empíricas




a) Una relación determinística

b) Una relación estadística

Figura 1. Relaciones determinísticas y estadísticas

# Análisis de Regresión

**Objetivo**: explicar y pronosticar el comportamiento de la variable dependiente “$Y$” a través del comportamiento de la o las variables independientes “$X$”.

**Finalidad:**

- Analiza relaciones de causalidad entre las variables económicas
- Realizar pronósticos o proyecciones del comportamiento de una variable

# Función de Regresión Poblacional

Se le llama "regresión" o **Función de Regresión Poblacional (FRP)** a la expresión:

$$Y_i = \beta_1 + \beta_2 X_i$$

Esta expresión lineal resume la verdadera relación existente entre la variable $X$ e $Y$.

El modelo a estimar es:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

Donde:

- $Y$ es la variable dependiente o explicada
- $X$ es la variable independiente o explicativa
- $\varepsilon$ es la perturbación aleatoria

Esta expresión deja en claro que la relación $X$ e $Y$ no es exacta, debido a $\varepsilon$.

# 2. La Muestra de Observaciones

# La Muestra de Observaciones

1. Notación de los datos:

- **Datos de Corte Transversal:** $Y_i$, $X_i$ donde $i = 1,2,\ldots,n$
- **Datos de Series de Tiempo:** $Y_t$, $X_t$ donde $\mathrm{t} = 1,2,\ldots,T$
- **Datos de Panel:** $Y_{it}$, $X_{it}$ donde $i = 1,2,\ldots,n$ y $\mathrm{t} = 1,2,\ldots,T$

2. Importancia de la aleatoriedad de la muestra.

¿A qué problema puede conllevar una muestra no aleatoria?

# La Muestra de Observaciones

Figura 2. Problema de Muestreo

**Farid Matuk**
@matuk

#ESTADISTICA No hay evidencia que la cuarentena deba continuar en Mayo.
Con los datos de fallecidos al 1 de Abril, la tasa de crecimiento era 22% y esta tasa colapsaba el sistema de salud en la 3ra semana de Abril, al rebasar la capacidad de camas UCI con Ventiladores Mecánicos.

```
Poisson regression                              Number of obs     =         14
                                                LR chi2(1)        =      45.65
                                                Prob > chi2       =     0.0000
Log likelihood = -30.749794                     Pseudo R2         =     0.4260
```

| FALL | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| T | .2252912 | .0371667 | 6.06 | 0.000 | .1524458 | .2981365 |
| _cons | -3.406033 | .887811 | -3.84 | 0.000 | -5.14611 | -1.665955 |

761 11:21 - 12 abr. 2020

806 personas están hablando de esto

# 3. Supuestos del Modelo Clásico

# Supuestos del Modelo Clásico: Linealidad

### Supuesto 1: Linealidad

La relación entre la variable dependiente $Y$ y la variable independiente $X$ es lineal. El modelo a estimar es:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \qquad i = 1, \dots, n$$

Donde $\beta_1$ y $\beta_2$ son parámetros poblacionales desconocidos a ser estimados, y $\varepsilon_i$ son términos de errores observados.

# Supuestos del Modelo Clásico: Linealidad

Es importante entender que significa $\beta_1$ y $\beta_2$:

- $\beta_1$: intercepto, el cual multiplica al número 1 (la constante)
- $\beta_2$: es el efecto marginal o la pendiente o el impacto de $X$ sobre $Y$, $\frac{\partial Y}{\partial X}$

**Preguntas:**

Las siguientes especificaciones cumplen con la propiedad de Linealidad:

$Y_i = \beta_1 + \beta_2 X_i^2 + \varepsilon_i$ ➡ Es lineal

$Y_i = e^{\beta_1} X_i^{\beta_2} + \varepsilon_i$ ➡ No es lineal

**Nota**: La propiedad de linealidad es una propiedad de los parámetros, no de las variables.

# Supuestos del Modelo Clásico: Linealidad

**Ejemplo: Retornos a la Escolaridad**

Ecuación de Mincer: ¿Cuál es el impacto de los años de educación sobre los salarios?

$$\log(w_i) = \beta_1 + \beta_2 sch_i + \varepsilon_i$$

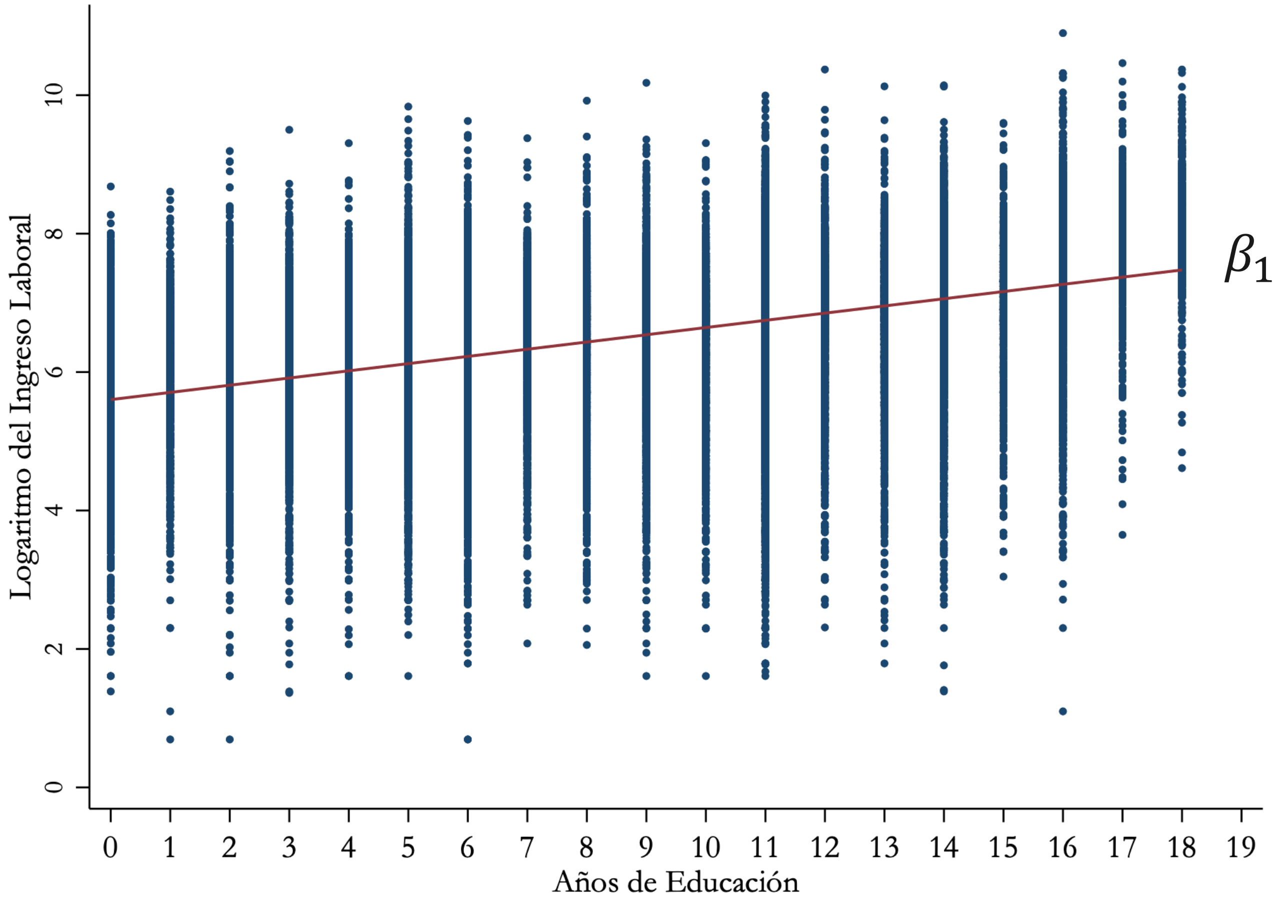
- $\beta_2$ mide el cambio proporcional en salarios asociado con el incremento de un año de escolaridad

$$\beta_2 = \frac{\partial log w_i}{\partial sch_i} = \frac{1}{w_i}\frac{\partial w_i}{\partial sch_i} = \frac{\frac{\partial w_i}{w_i}}{\partial sch_i} = \frac{cambio\ relativo\ en\ w_i}{cambio\ absoluto\ en\ sch_i}$$

- Notar que $\%\Delta w = 100 \cdot \beta_2 \cdot \Delta sch$, es decir, si aumenta un año de escolaridad, se espera que el salario cambie en $100 \cdot \beta_2$ por ciento.

# Supuestos del Modelo Clásico: Linealidad

Figura 3. Relación Lineal entre los años de educación y el salario




$\beta_1 + \beta_2 X_i$

# Supuestos del Modelo Clásico: Exogeneidad Estricta

**Supuesto 2: Exogeneidad Estricta**

El valor esperado de la perturbación aleatoria debe ser cero para cualquier observación:

$$E[\varepsilon_i|X_i] = 0 \qquad i = 1,2,\dots,n$$

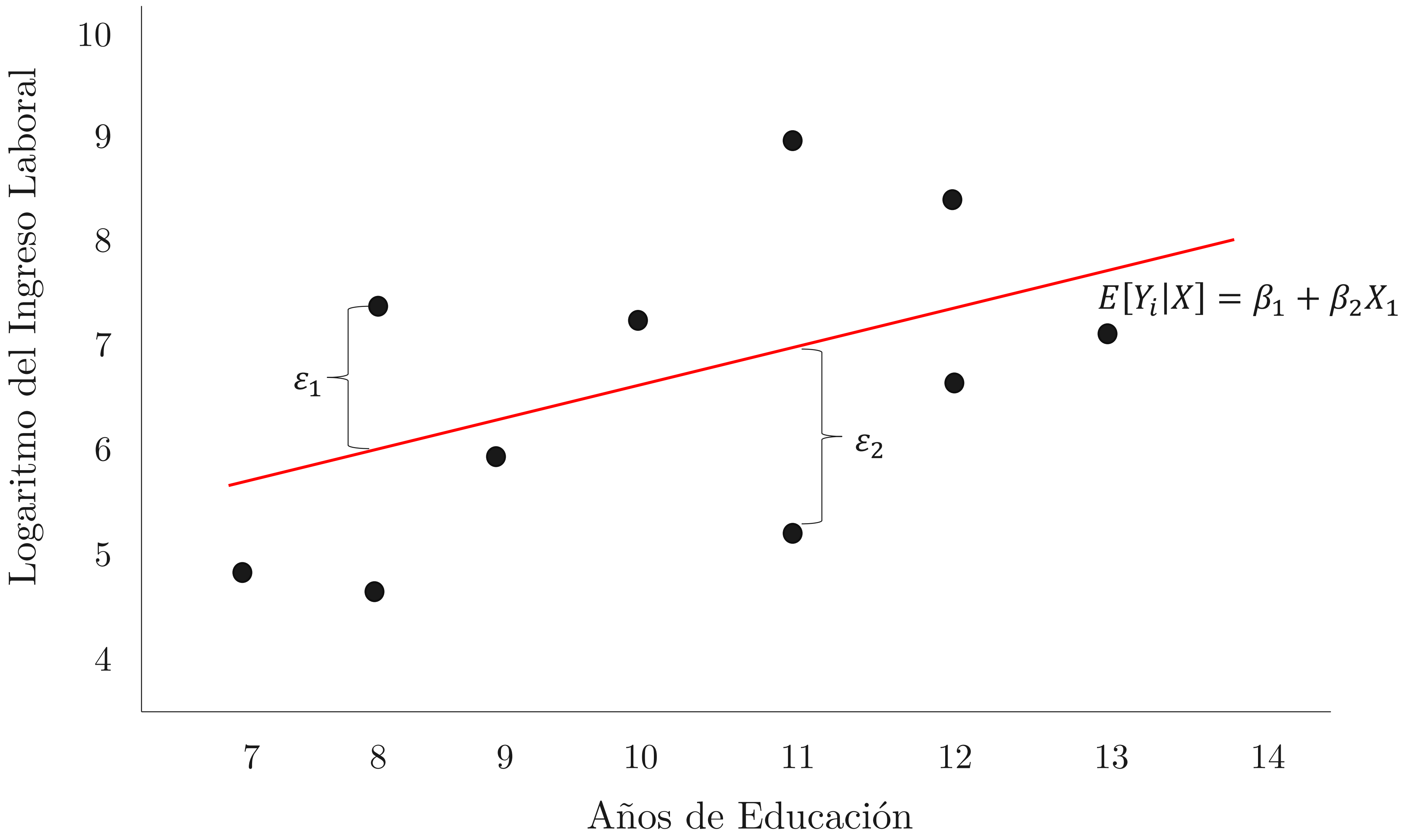
Dados diferentes valores de $X$ en la población, el promedio no observable es igual a cero.

Esto también se puede escribir como:

$$E[\varepsilon|X = X_i] = 0$$

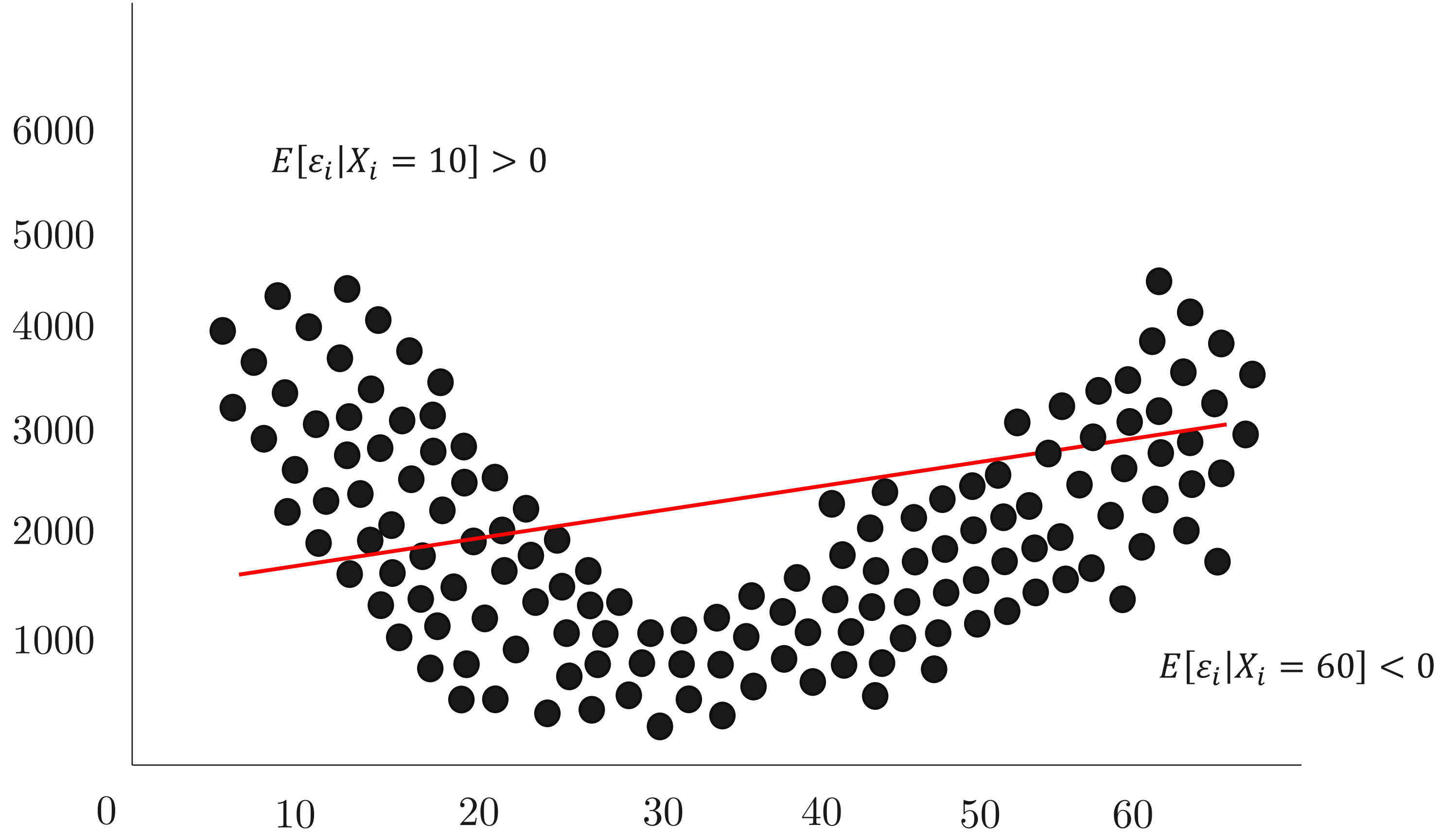
# Supuestos del Modelo Clásico: Exogeneidad Estricta

Figura 4. Recta de regresión poblacional y el término de perturbación



# Supuestos del Modelo Clásico: Exogeneidad Estricta

Figura 5. Caso donde no se cumple el supuesto



# Supuestos del Modelo Clásico: Exogeneidad Estricta

## Consecuencias de Exogeneidad Estricta:

**Definición: Ley de Esperanzas Iteradas**

Si $E|Y| < \infty$, es decir, la esperanza poblacional de $Y$ existe, entonces para cualquier vector $X$ aleatorio:

$$E_X[E(Y|X)] = E(Y)$$

**1. Media Incondicional**

La media incondicional del término de error es cero:

$$E(\varepsilon_i) = 0$$

**Demostración**. Por la Ley de Esperanzas Iteradas (LIE)

$$E(\varepsilon_i) = E[E(\varepsilon_i|X)] = 0$$

## Consecuencias de Exogeneidad Estricta:

**2. Regresores son ortogonales al término de error**

Si $E(xy)$ de dos variables aleatorias $x$ e $y$ es cero, entonces decimos que $x$ es ortogonal a $y$. Bajo exogeneidad estricta, el regresor $X$ es ortogonal al término de error $\varepsilon$ para todas las observaciones, es decir, no comparten información.

$$E(X_i \varepsilon_i) = 0 \qquad i = 1,2,\ldots,n$$

# Supuestos del Modelo Clásico: Exogeneidad Estricta

## Consecuencias de Exogeneidad Estricta:

### 3. Condición de cero correlación

El regresor no está correlacionado con el término de error:

$$\mathrm{Cov}(\varepsilon_i, X_i) = 0$$

**Demostración:**

$$E(\varepsilon_i X_i) - E(X_i)E(\varepsilon_i) = 0$$

Donde:

$$E(\varepsilon_i X_i) = 0$$

$$E(\varepsilon_i) = 0$$

### 4. Media condicional de $y_i$

La media condicional de la variable dependiente es una función lineal del regresor.

Bajo el supuesto de linealidad y exogeneidad estricta:

$$E(Y_i|X_i) = E(\beta_1 + \beta_2 X_i + \varepsilon_i|X_i) = \beta_1 + \beta_2 X_1 + E(\varepsilon_i|\mathrm{X_i})$$

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_i$$

## Supuesto 3: Perturbaciones Esféricas

### ❖ Homocedasticidad

Asume que todas las unidades tienen el mismo error de varianza

$$Var[\varepsilon_i|X_i] = \sigma^2 > 0 \qquad i = 1,2,\dots,n$$

### ❖ No Correlación

Asume también que no hay correlación entre las observaciones

$$\mathrm{Cov}\left[\varepsilon_i\varepsilon_j\middle|X_iX_j\right] = 0 \qquad i,j = 1,2,\dots,n \qquad i \neq j$$

# Supuestos del Modelo Clásico: Perturbaciones Esféricas

Este supuesto señala que la varianza de todos los términos de errores es constante. Para ver esto:

$$Var[\varepsilon_i|X_i] = E\left(\varepsilon_i^2\middle|X\right) - [E(\varepsilon_i|X)]^2 \quad (Por\ varianza\ condicional)$$

$$Var[\varepsilon_i|X_i] = E\left(\varepsilon_i^2\middle|X\right) \quad (Por\ exogeneidad\ estricta)$$

Similarmente, la covarianza de la distribución conjunta de $\varepsilon_i\varepsilon_j$ condicional a $X$ es cero.

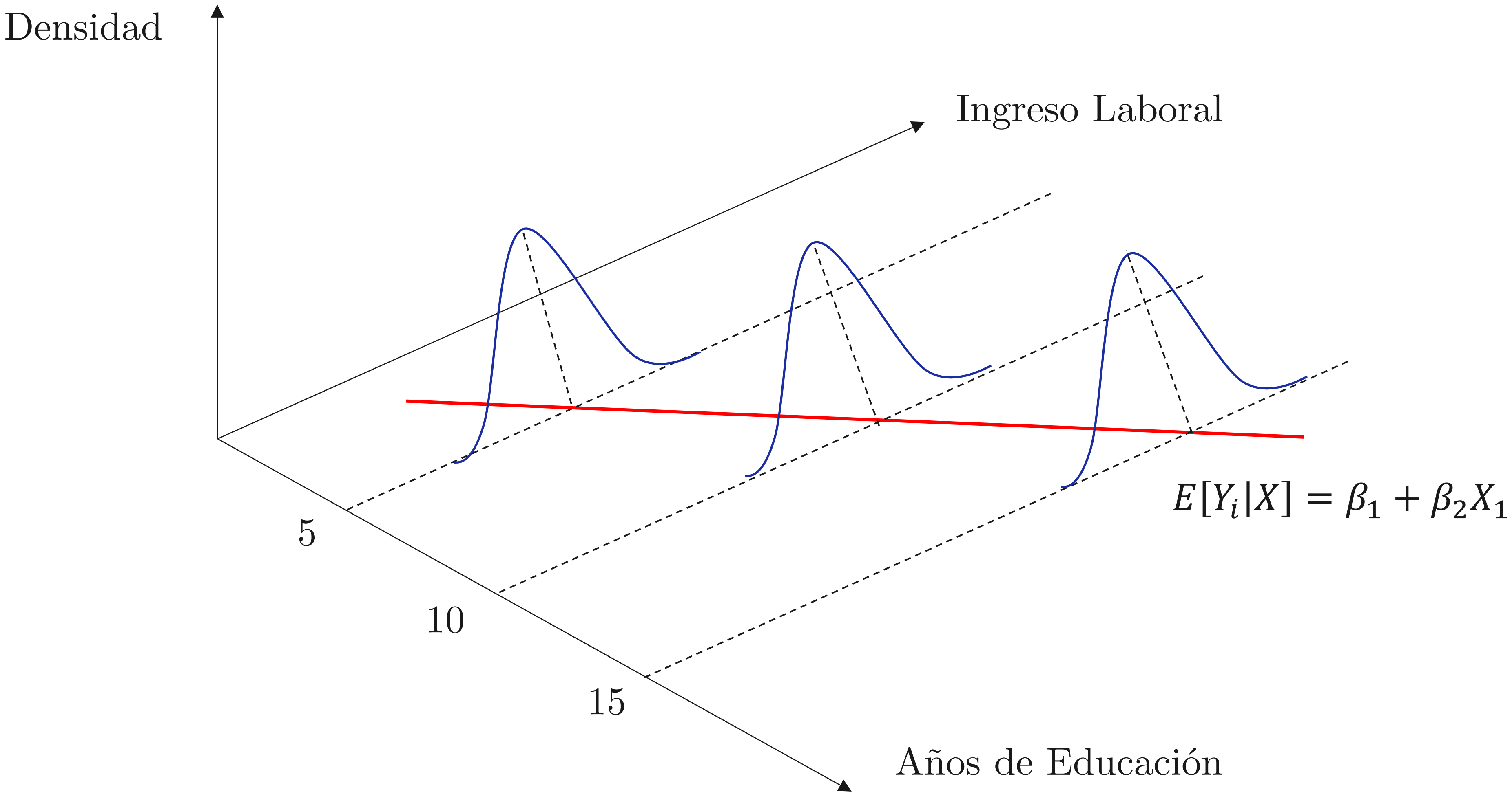
$$\mathrm{Cov}\left[\varepsilon_i\varepsilon_j\middle|X_iX_j\right] = 0 \qquad i,j = 1,2,\ldots,n \qquad i \neq j$$

$$\mathrm{E}\left[\varepsilon_i\varepsilon_j\middle|X_iX_j\right] - E(\varepsilon_i|X_i)E(\varepsilon_j|X_j) =$$

$$\mathrm{E}\left[\varepsilon_i\varepsilon_j\middle|X_iX_j\right] = 0$$

# Supuestos del Modelo Clásico: Perturbaciones Esféricas

Figura 6. Homocedasticidad



**Supuesto 4: Regresores no estocásticos**

Las observaciones de $X_i$ son fijos en muestras repetidas.

Asumir que los $X$ son fijos quiere decir que, en repetidas muestras de $X$, los valores obtenidos $X_1, X_2, \ldots, X$ van a ser siempre los mismos; es decir, dejan de ser aleatorios.

Bajo este supuesto, ya no es necesario hablar de esperanzas condicionales:

$$E(\varepsilon_i) = 0$$

$$\text{Var}(\varepsilon_i) = \sigma^2$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

# Supuestos del Modelo Clásico: Normalidad de los errores

**Supuesto 5: Normalidad de los errores**

$\varepsilon_i$ distribuye normal con media cero y varianza $\sigma^2$ condicional a $X$:

$$\varepsilon_i | X \sim N(0, \sigma^2), \qquad i = 1,2,\dots,n$$

# 4. La Función de Regresión Muestral

# La Función de Regresión Muestral

**Objetivo**: encontrar una aproximación a la Función de Regresión Poblacional (FRP) a partir de la muestra de observaciones.

La Función de Regresión Muestral (FRM) se define como:

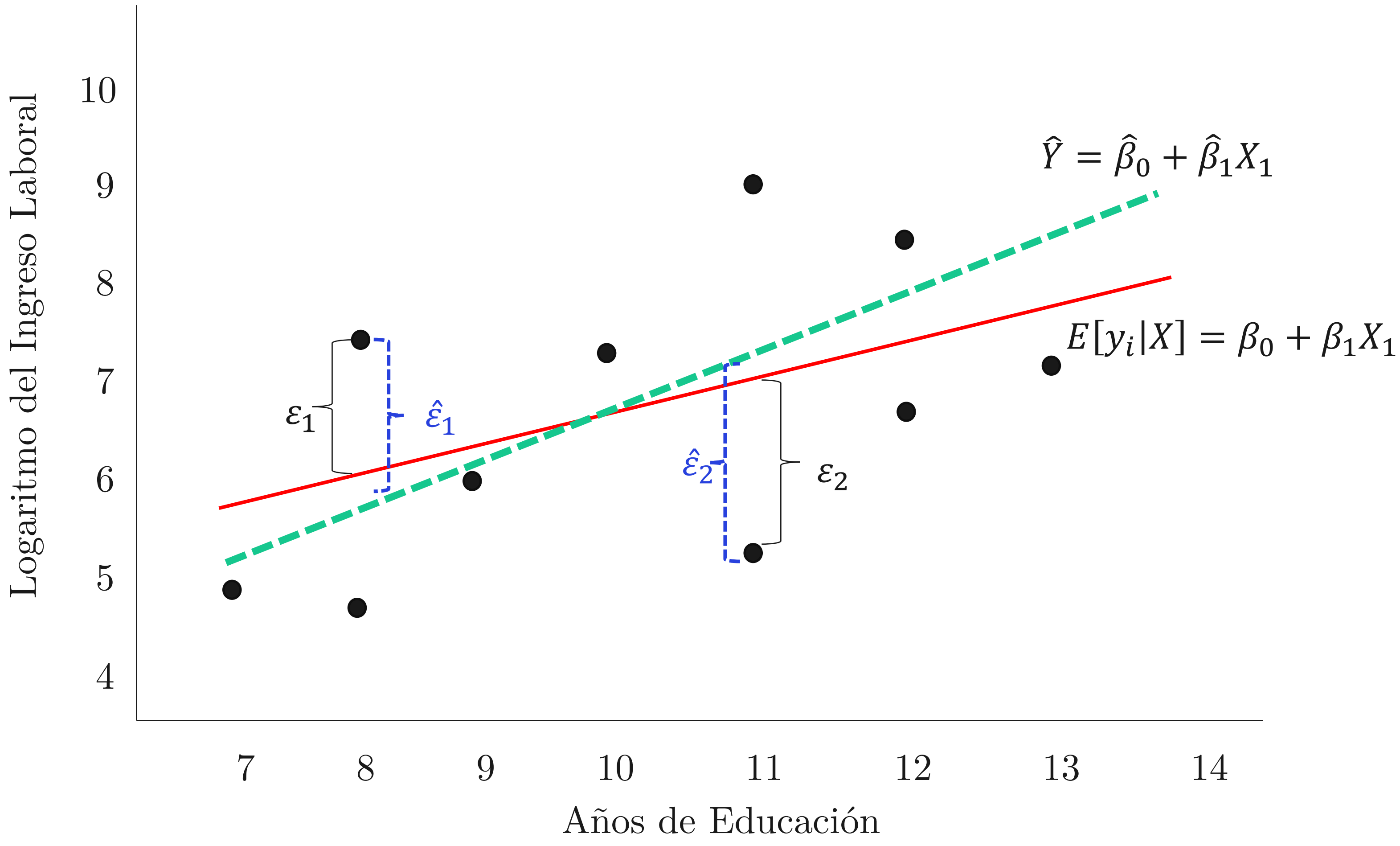
$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Donde:

- $\hat{Y}_i$ es un estimador de $E(Y_i|X_i)$
- $\hat{\beta}_1$ es un estimador del parámetro poblacional $\beta_1$
- $\hat{\beta}_2$ es un estimador del parámetro poblacional $\beta_2$

# La Función de Regresión Muestral

Figura 7. Rectas de Regresión Poblacional y Muestral



# La Función de Regresión Muestral

En esta Figura se muestran las distancias entre dos puntos de observaciones y lo estimado por la recta.

Esta distancia son los residuos definidos:

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 \mathrm{X_i}$$

Se puede escribir así también:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 \mathrm{X_i} + \hat{\varepsilon}_i$$

Este se denomina como **modelo estimado**.

# La Función de Regresión Muestral

Tenemos dos grupos de ecuaciones: las poblacionales y las muestrales.

- Modelo Econométrico:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

- Función de Regresión Poblacional

$$E(Y_i | X_i) = \beta_1 + \beta_2 X_i$$

- Modelo Estimado:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 \mathrm{X_i} + \hat{\varepsilon}_i$$

- Recta estimada o Función de Regresión Muestral:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

# 5. Estimación por Mínimos Cuadrados Ordinarios

# Estimación por MCO

Existen métodos para calcular la Función de Regresión Muestral, siendo el más usado el método de Mínimos Cuadrados Ordinarios (MCO)

La recta que mejor se ajuste a los datos será aquella que presente la menor suma de cuadrados de los residuos.

Para ello, se define a la sumatoria de cuadrados de los residuos (SCR) como:

$$SCR = \sum_{i}^{n} \hat{\varepsilon}_i^2 = \sum_{i}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$$

**Objetivo:** Hallar $\hat{\beta}_1$ y $\hat{\beta}_2$ de manera que minimice la $SCR$.

# Estimación por MCO

$$SCR = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$$

Entonces, obtenemos las denominadas “ecuaciones normales” de la estimación MCO:

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) \cdot X_i = 0$$

# Estimación por MCO

Entonces, se cumple que:

$$\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) = 0$$

Desarrollando el paréntesis:

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} X_i = 0$$

Dividiendo ambos lados por $n$, se obtiene:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

Despejamos el valor de $\hat{\beta}_0$ y obtenemos:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

# Estimación por MCO

De la ecuación de $\frac{\partial SSR}{\partial \hat{\beta}_2}$:

$$\sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) \cdot X_i = 0$$

Desarrollando el paréntesis:

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 = 0$$

# Estimación por MCO

Reemplazando $\hat{\beta}_1$:

$$\sum_{i=1}^{n} Y_i X_i = (\bar{Y} - \hat{\beta}_2 \bar{X}) \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i = \hat{\beta}_2 (\sum_{i=1}^{n} X_i^2 + \bar{X} \sum_{i=1}^{n} X_i)$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2 + \bar{X} \sum_{i=1}^{n} X_i}$$

Esta expresión es igual a:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

De aquí se obtiene $\hat{\beta}_1$

# Referencias


Capítulo 2 y 3 - Gujarati, D., & Porter, D. (2010). Econometría (Quinta edición ed.). *& P. Carril Villareal, Trad.) México: Mc Graw Hill educación*.

Capítulo 2 y 3 - Wooldridge, J. M. (2006). *Introducción a la econometría. Un enfoque moderno: un enfoque moderno*. Editorial Paraninfo.

PUCP