TEMA 2: EL ESTIMADOR MCO

S. Álvarez, A. Beyaert, M. Camacho, M. González, A. Quesada
Departamento de Métodos Cuantitativos para la Economía y la Empresa
UNIVERSIDAD DE
MURCIA

Lo que estudiaremos en este tema:

UNIVERSIDAD DE **MURCIA**

1. El modelo de regresión múltiple

2. Funcionamiento e interpretación del estimador MCO

2.1 Introducción

2.2 Estimación MCO. Modelo de regresión simple

2.3 Estimación MCO. Modelo de regresión múltiple

2.4 Estimación MCO. Interpretación

2.5 Propiedades algebraicas del estimador MCO

2.6 Bondad del ajuste: R-cuadrado

2.7 Bondad del ajuste: R-cuadrado ajustado

3. Unidades de medida y forma funcional

3.1 Unidades de medida

3.2 Formas funcionales

Bibliografía básica: Wooldridge, 2008, cap. 2, 3 y 6

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$

- y: variable dependiente, variable explicada o regresando
- $x_1,\ldots,x_k$: variables independientes, variables explicativas o regresores
- $\beta_0,\beta_1,\ldots,\beta_k$: son los $K=(k+1)$ coeficientes del modelo
- ε: término de error, perturbación o shock

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon$$

- Los coeficientes son **efectos ceteris paribus**, si se cumple el supuesto de **media condicionada nula:**

$$E(\varepsilon / x_1, x_2, \ldots, x_k) = E(\varepsilon) = 0$$

Dos implicaciones:

  - $E(\varepsilon / x_1, x_2, \ldots, x_k) = E(\varepsilon)$

    Es un supuesto clave. Implica que el valor esperado de ε sea independiente de los valores de $x_1, \ldots, x_k$

  - $E(\varepsilon) = 0$

    No es restrictivo, cumpliéndose lo anterior, siempre que el modelo incluya término constante

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$

- Es un **modelo lineal en parámetros**
- La interpretación de β_1 es:

$$\beta_1 = \frac{\Delta E\left[y / x_1, \ldots, x_k\right]}{\Delta x_1} \quad \text{si } \Delta x_2 = \Delta x_3 = \ldots = \Delta x_k = 0$$

- Un caso particular: el **modelo de regresión simple**

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- El modelo de regresión múltiple es más útil que el simple porque:
  - permite controlar explícitamente los diversos factores que afectan a la variable dependiente
  - permite generalizar relaciones funcionales entre variables: por ejemplo, una función cuadrática

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon$$

- Objetivo: estimar $\beta_0, \beta_1, \ldots, \beta_k$ de **la función de regresión poblacional**

$$E\left(y/x_1, \ldots, x_k\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

- Necesitamos una **muestra** de la población

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i; \quad i = 1, \ldots, N$$

- $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ definen la **función de regresión muestral**

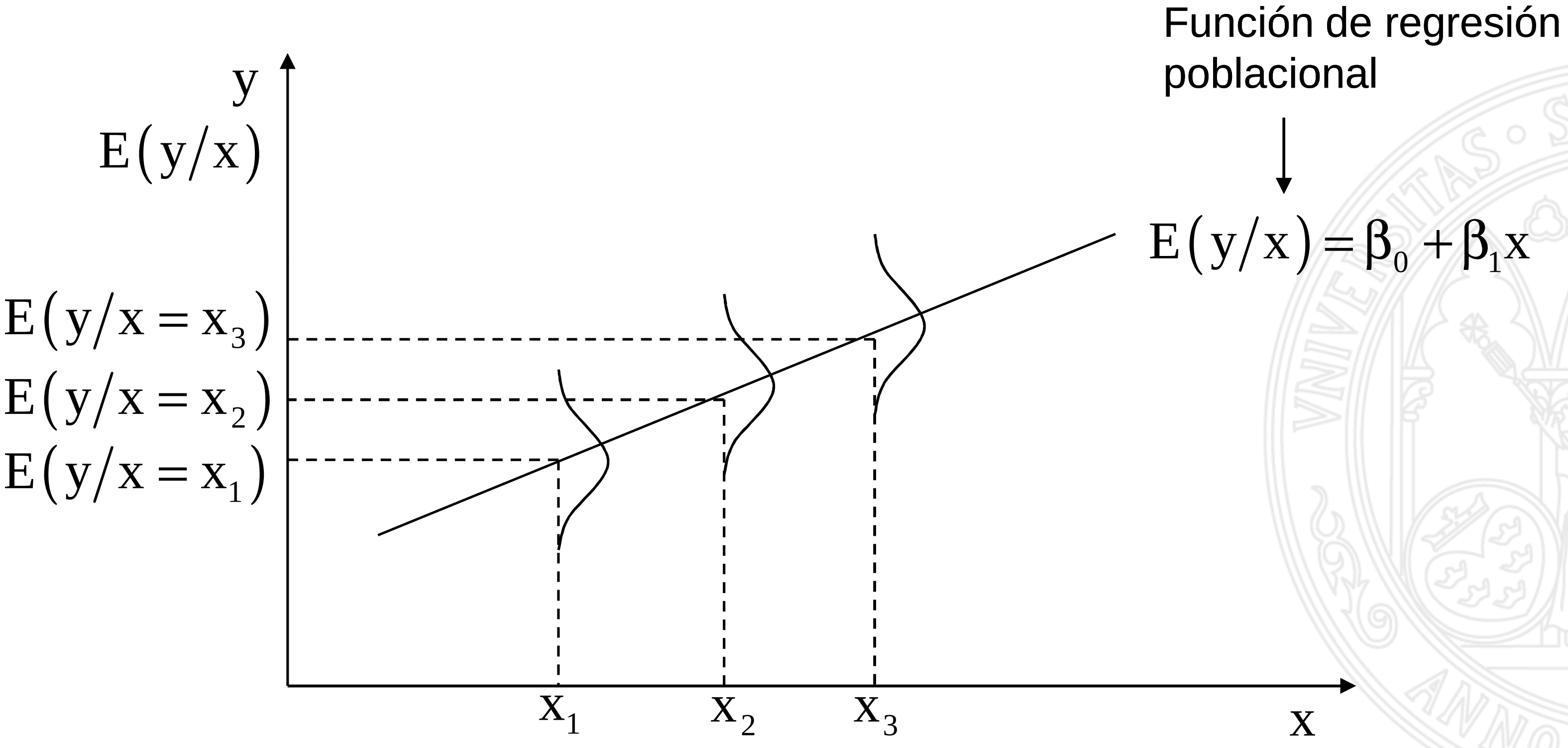
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_k x_{ki}$$

- La diferencia entre el valor verdadero y el ajustado es el **residuo**

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}$$

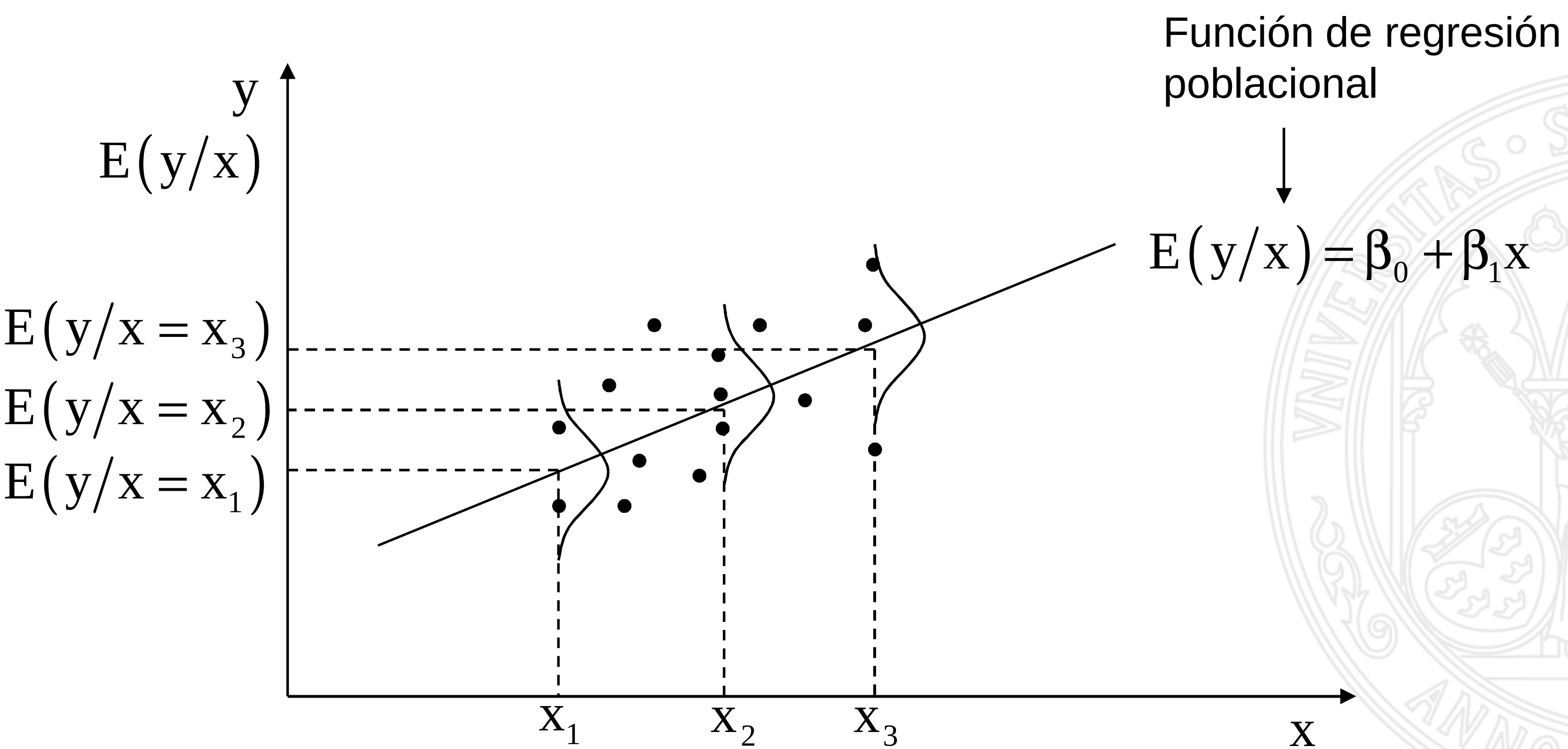
UNIVERSIDAD DE MURCIA

$$y = \beta_0 + \beta_1 x + \varepsilon$$



UNIVERSIDAD DE MURCIA

$$y = \beta_0 + \beta_1 x + \varepsilon$$

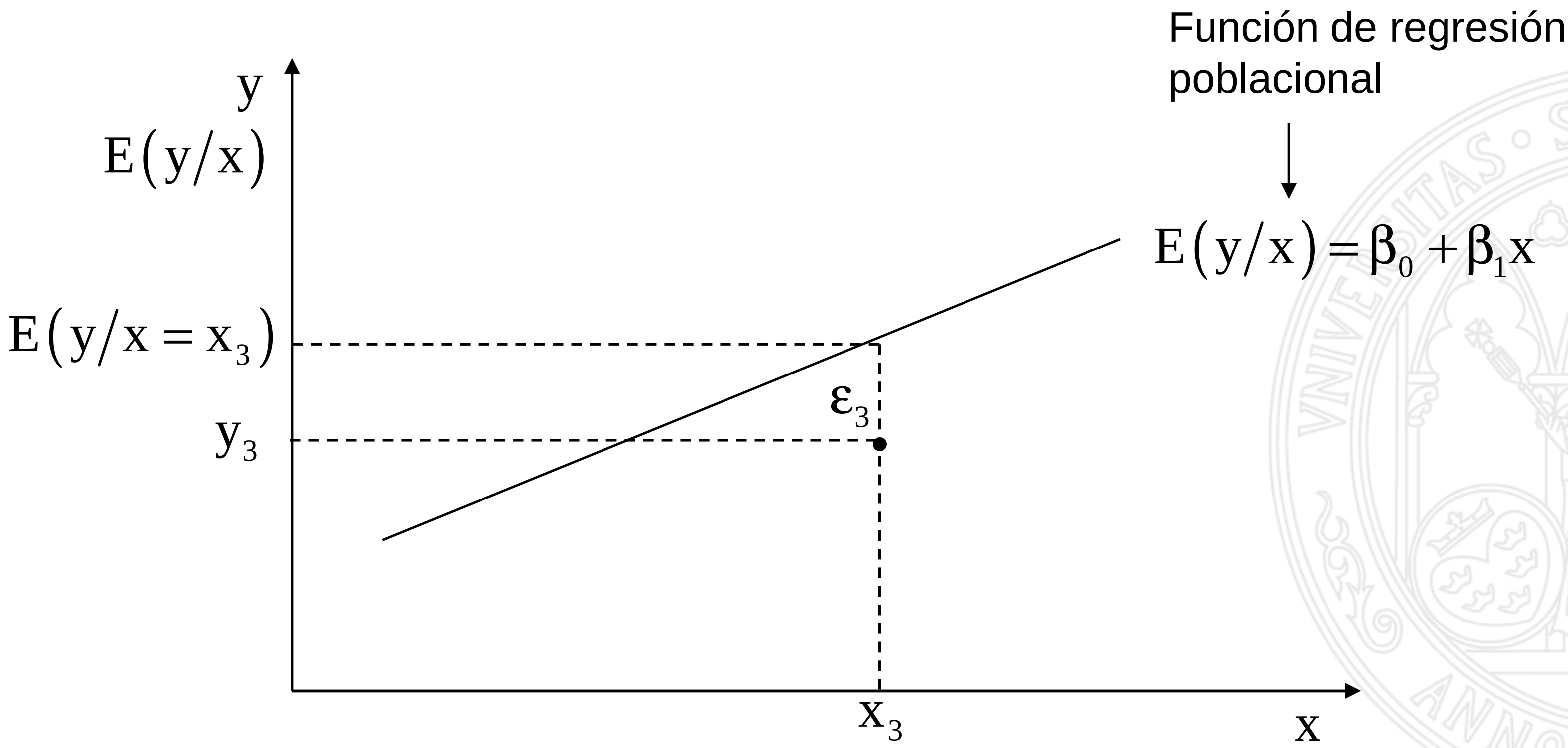


2. Funcionamiento e interpretación del estimador MCO

2.2 Estimación MCO. Modelo de regresión simple

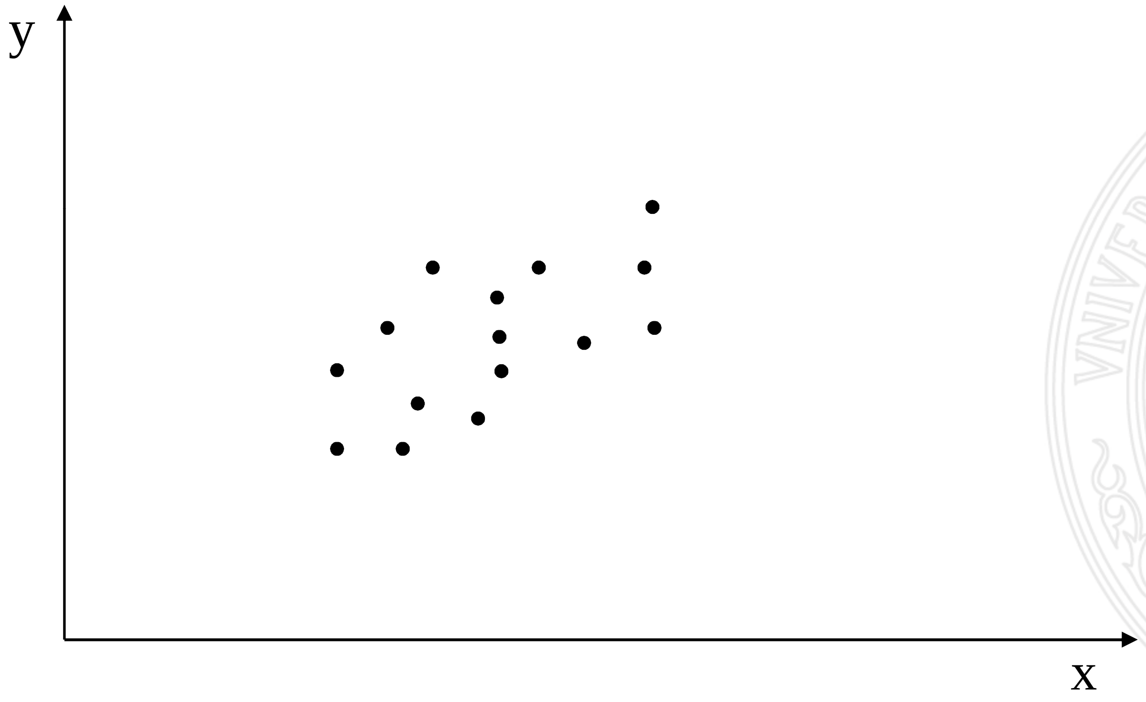
UNIVERSIDAD DE MURCIA

$$y = \beta_0 + \beta_1 x + \varepsilon$$

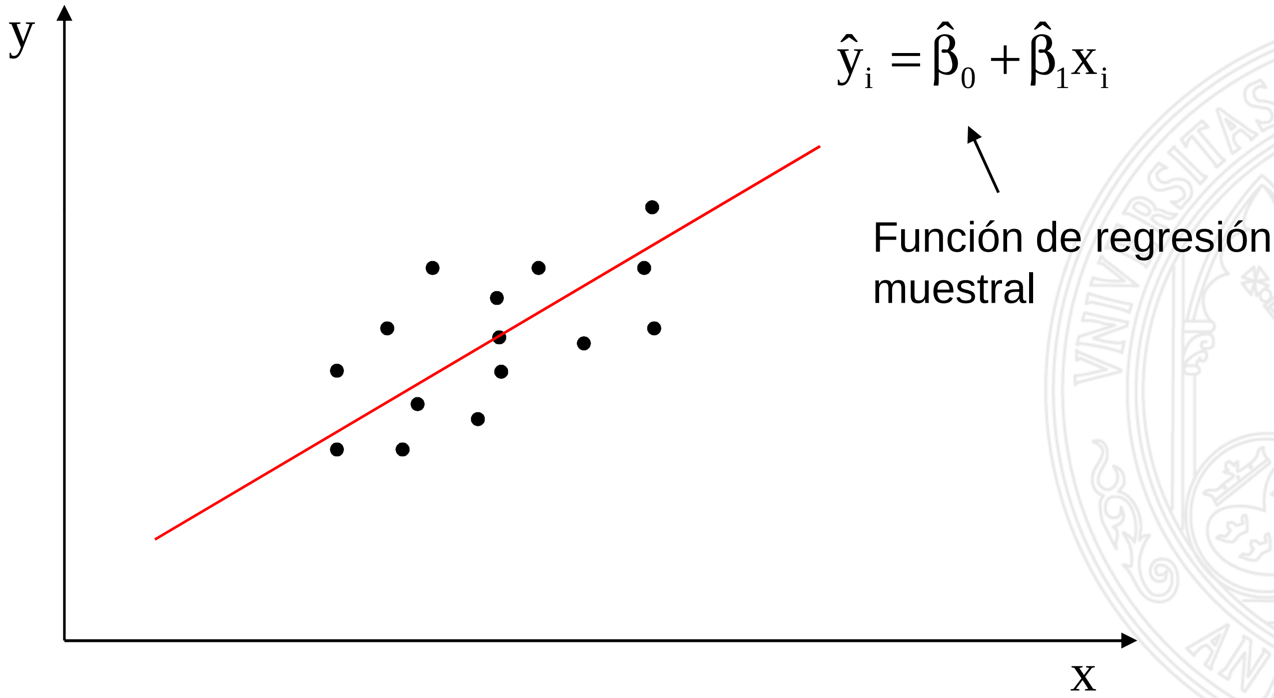


UNIVERSIDAD DE MURCIA

$$y = \beta_0 + \beta_1 x + \varepsilon$$



$$y = \beta_0 + \beta_1 x + \varepsilon$$

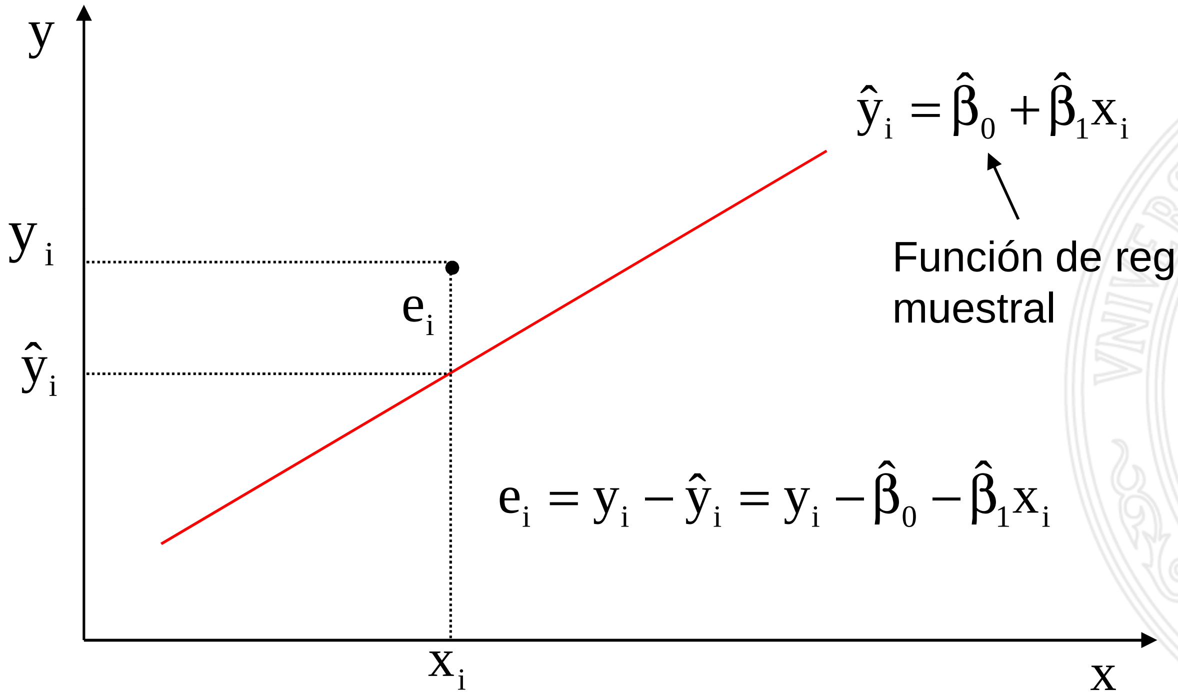


2. Funcionamiento e interpretación del estimador MCO

2.2 Estimación MCO. Modelo de regresión simple

UNIVERSIDAD DE MURCIA

$$y = \beta_0 + \beta_1 x + \varepsilon$$

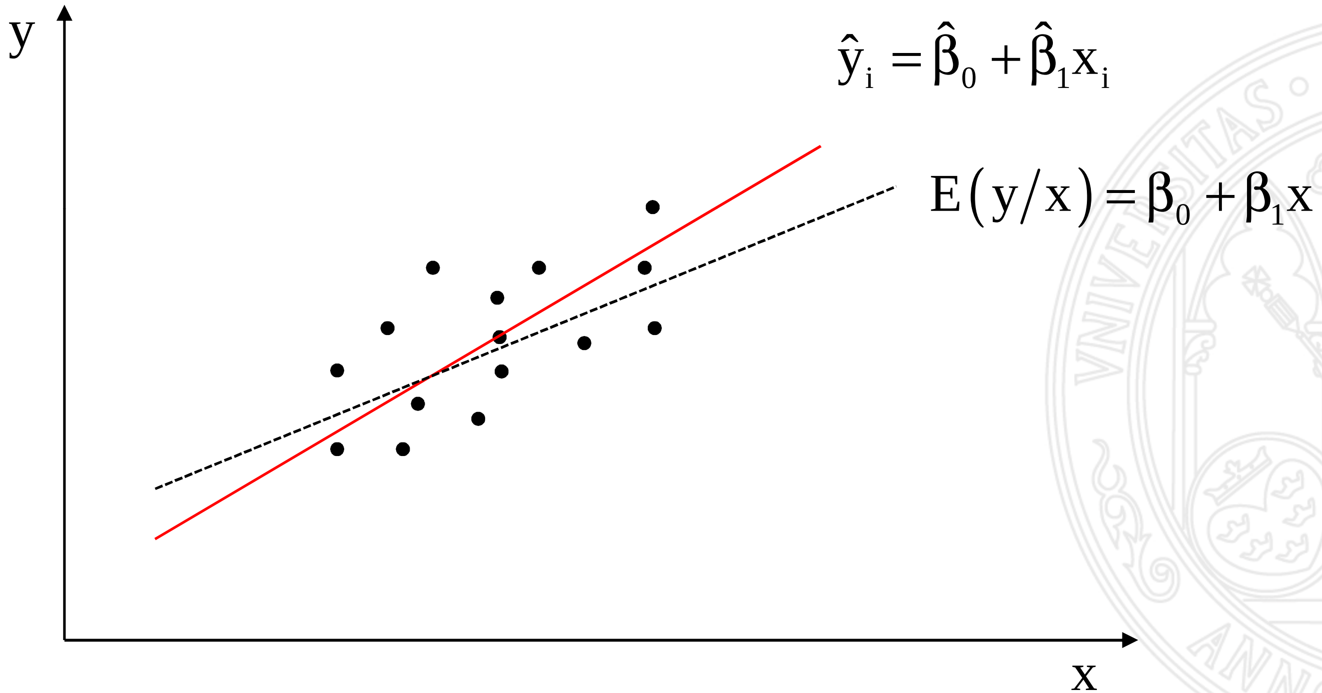


2. Funcionamiento e interpretación del estimador MCO

2.2 Estimación MCO. Modelo de regresión simple

UNIVERSIDAD DE MURCIA

$$y = \beta_0 + \beta_1 x + \varepsilon$$



UNIVERSIDAD DE MURCIA

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- MCO escoge $\hat{\beta}_0$ y $\hat{\beta}_1$ para minimizar la suma de los cuadrados de los residuos:

$$\underset{b_0, b_1}{\mathrm{Min}} \sum_{i=1}^{N} \left(y_i - b_0 - b_1 x_i \right)^2 \equiv \underset{b_0, b_1}{\mathrm{Min}}\, Q(b_0, b_1)$$

- Las condiciones de primer orden son:

$$\left. \frac{\partial Q(b_0, b_1)}{\partial b_0} \right|_{b_0 = \hat{\beta}_0, b_1 = \hat{\beta}_1} = 0$$

$$\left. \frac{\partial Q(b_0, b_1)}{\partial b_1} \right|_{b_0 = \hat{\beta}_0, b_1 = \hat{\beta}_1} = 0$$

UNIVERSIDAD DE MURCIA

- De las condiciones de primer orden obtenemos el **sistema de ecuaciones normales**:

$$\sum_{i=1}^{N}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)=0$$

$$\sum_{i=1}^{N}x_i\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)=0$$

- Resolviendo:

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

$$\hat{\beta}_1=\frac{\sum_{i=1}^{N}x_i\left(y_i-\overline{y}\right)}{\sum_{i=1}^{N}x_i\left(x_i-\overline{x}\right)}=\frac{\sum_{i=1}^{N}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sum_{i=1}^{N}\left(x_i-\overline{x}\right)^2}=\frac{S_{xy}}{S_x^2}$$

- El supuesto necesario para estimar es: $\sum_{i=1}^{N}\left(x_i-\overline{x}\right)^2>0$

UNIVERSIDAD DE MURCIA

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

- El problema de optimización es:

$$\underset{b_0, b_1, \ldots, b_k}{\text{Min}} \sum_{i=1}^{N} \left(y_i - b_0 - b_1 x_{1i} - \ldots - b_k x_{ki} \right)^2 \equiv \underset{b_0, b_1, \ldots, b_k}{\text{Min}} Q\left(b_0, b_1, \ldots b_k\right)$$

- Condiciones de primer orden:

$$\left. \frac{\partial Q\left(b_0, b_1, \ldots, b_k\right)}{\partial b_0} \right|_{b_0 = \hat{\beta}_0, b_1 = \hat{\beta}_1, \ldots, b_k = \hat{\beta}_k} = 0$$

$$\vdots$$

$$\left. \frac{\partial Q\left(b_0, b_1, \ldots, b_k\right)}{\partial b_k} \right|_{b_0 = \hat{\beta}_0, b_1 = \hat{\beta}_1, \ldots, b_k = \hat{\beta}_k} = 0$$

- El **sistema de ecuaciones normales** es:

$$\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}\right) = 0$$

$$\sum_{i=1}^{N} x_{1i}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}\right) = 0$$

$$\vdots$$

$$\sum_{i=1}^{N} x_{ki}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}\right) = 0$$

- Resolviendo obtenemos $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$

En **modelos sin termino constante**: $\hat{y} = \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$

- El sistema de ecuaciones normales es:

$$\sum_{i=1}^{N} x_{1i}\left(y_i - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}\right) = 0$$

$$\vdots$$

$$\sum_{i=1}^{N} x_{ki}\left(y_i - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}\right) = 0$$

- Resolviendo obtenemos $\hat{\beta}_1, \ldots, \hat{\beta}_k$

- Expresión algebraica del estimador MCO:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} r_{1i} y_i}{\sum_{i=1}^{N} r_{1i}^2}$$

r_1 es el residuo de la regresión de x_1 sobre $x_2,\ldots,x_k$, y una constante. Es la parte de x_1 que no está explicada con $x_2,\ldots,x_k$

$\hat{\beta}_1$ mide la relación muestral entre y y x_1 una vez tomado en cuenta el efecto de $x_2,\ldots, x_k$

$\Rightarrow$ estimación del efecto céteris paribus

UNIVERSIDAD DE MURCIA

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + ... + \hat{\beta}_k x_{ki} \quad \forall i=1,\ldots,N$$

- $\hat{\beta}_0$: Es el valor estimado de y cuando todas las variables explicativas valen cero
- $\hat{\beta}_1,\ldots,\hat{\beta}_k$: Efectos parciales o efectos céteris paribus

$$\hat{\beta}_1 = \frac{\hat{\Delta} y}{\Delta x_1}$$

Variación estimada en y ante cambios unitarios en x_1, cuando el resto de variables explicativas no cambian

- La regresión ofrece información céteris paribus aunque los datos no hayan sido recogidos de forma céteris paribus, si se cumple:

$$E\left(\varepsilon_i / x_{1i}, x_{2i}, \ldots, x_{ki}\right) = E\left(\varepsilon_i\right) = 0 \quad \forall i$$

UNIVERSIDAD DE MURCIA

Derivan del sistema de ecuaciones normales

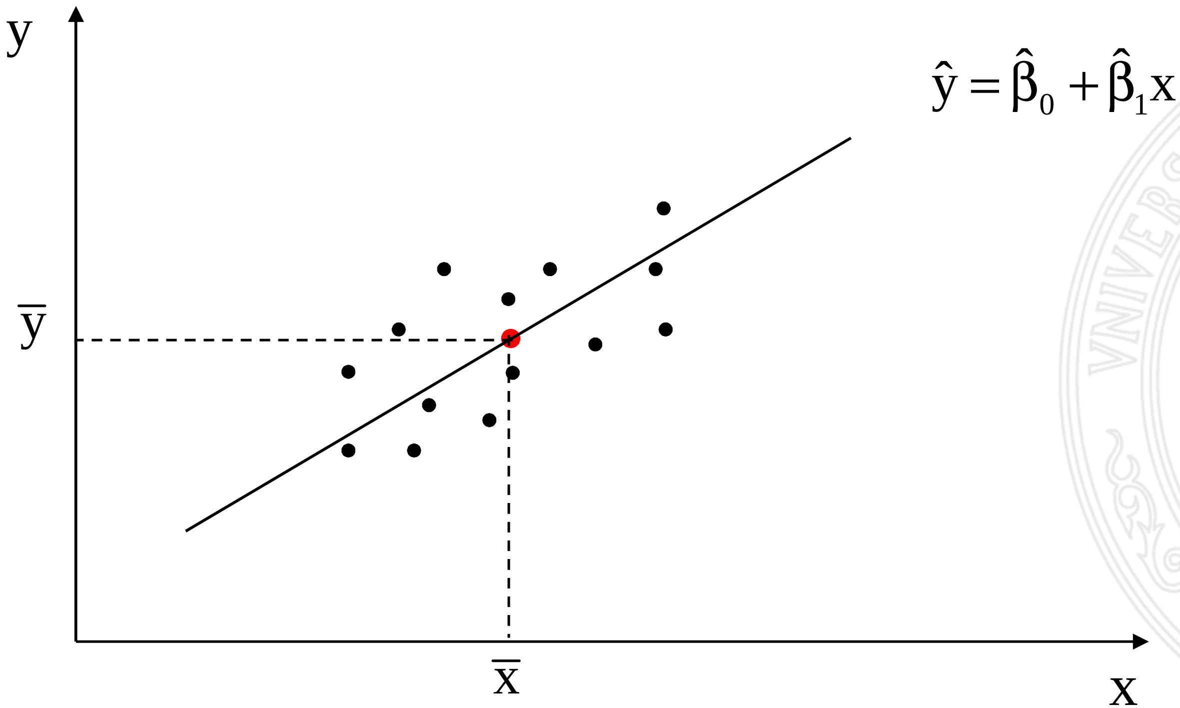
1. En **modelos con término constante**:

- La suma de los residuos MCO es nula: $\sum_{i=1}^{N} e_i = 0$
- La media de los residuos MCO es nula: $\bar{e} = 0$
- Las medias muestrales de y y de $\hat{y}$ son iguales: $\bar{y} = \bar{\hat{y}}$
- El punto $(\bar{y}, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ siempre está sobre la función de regresión MCO:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \cdots + \hat{\beta}_k \bar{x}_k$$

UNIVERSIDAD DE MURCIA

Gráficamente



2. **Con o sin término constante:**

- La covarianza muestral entre cada regresor y los residuos MCO es nula:

$$\sum_{i=1}^{N} x_{ji} e_i = 0; j = 1, \ldots, k$$

- La covarianza muestral entre $\hat{y}$ y e es nula:

$$\sum_{i=1}^{N} \hat{y}_i e_i = 0$$

Los valores ajustados y los residuos están incorrelacionados en la muestra

UNIVERSIDAD DE MURCIA

- El **coeficiente de determinación** es:

$$R^2 = 1 - \frac{SCE}{STC} \qquad R^2 \leq 1$$

donde:

$$STC = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

$$SCE = \sum_{i=1}^{N} e_i^2$$

en modelos con constante

$$\text{además}: SCR = \sum_{i=1}^{N} \left(\hat{y}_i - \bar{\hat{y}}\right)^2 = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$$

STC: Suma total de cuadrados

SCE: Suma de cuadrados de los residuos

SCR: Suma de cuadrados de la regresión

UNIVERSIDAD DE MURCIA

- En **modelos con término constante** STC= SCR+SCE

$$R^2 = 1 - \frac{SCE}{STC} = \frac{SCR}{STC} \qquad 0 \le R^2 \le 1$$

Interpretación: Es la fracción de la variación muestral de y explicada por la función de regresión muestral

- También se puede demostrar que

$$R^2 = \left(\frac{S_{y\hat{y}}}{S_y S_{\hat{y}}} \right)^2$$

- Si **no existe término constante** STC≠SCR+SCE
- El R^2 puede ser negativo sólo en modelos sin término constante

- Un R^2 bajo no implica que la regresión MCO no sea útil

- El R^2 nunca disminuye cuando se añade otro regresor

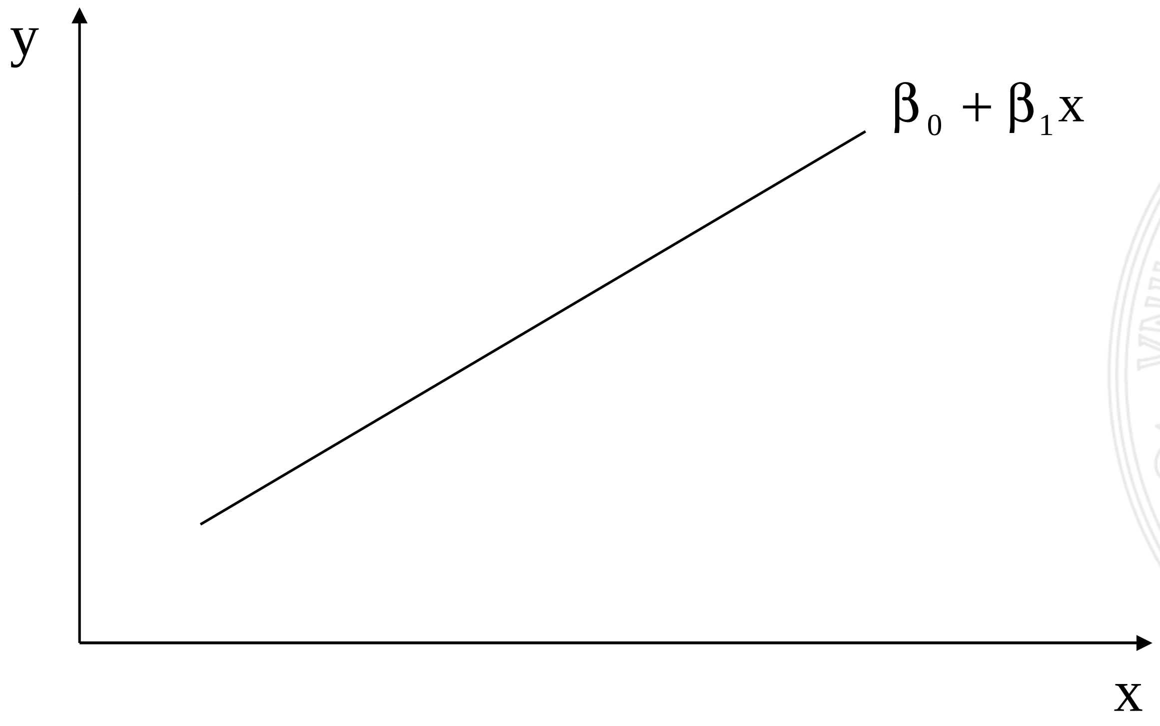
  $\Rightarrow$ no permite comparar modelos con distinto número de regresores

  Si $K = N \Rightarrow R^2 = 1$

UNIVERSIDAD DE MURCIA

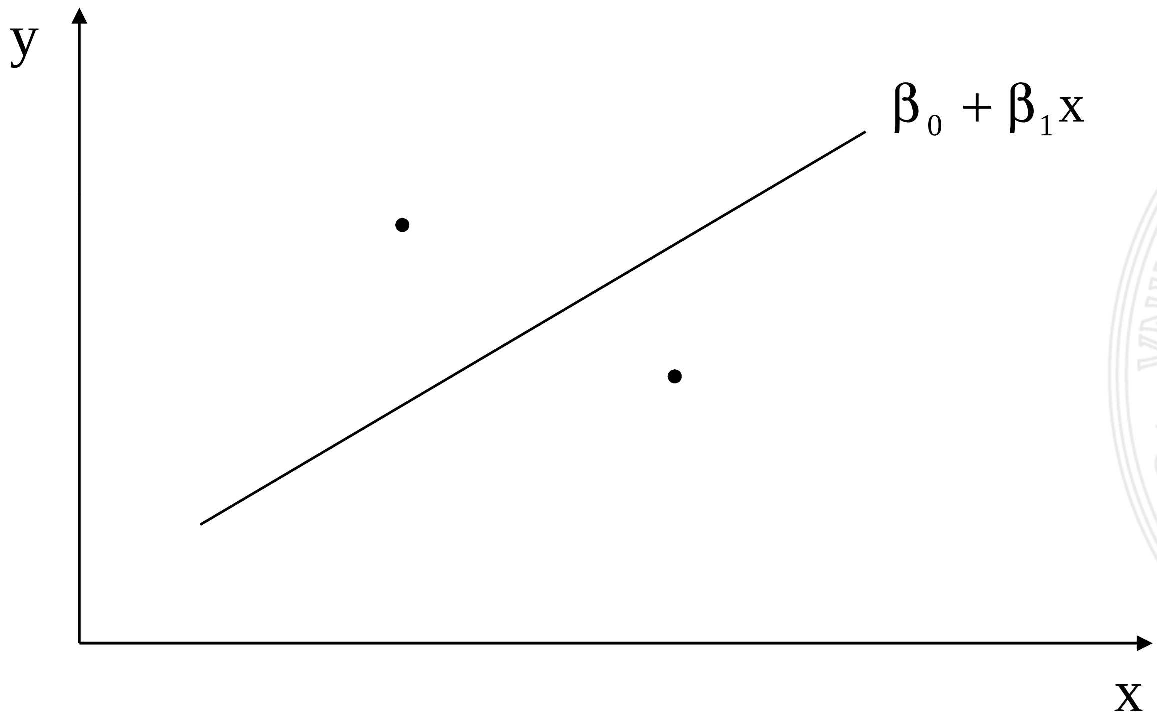
Gráficamente

Modelo de regresión simple, K=2

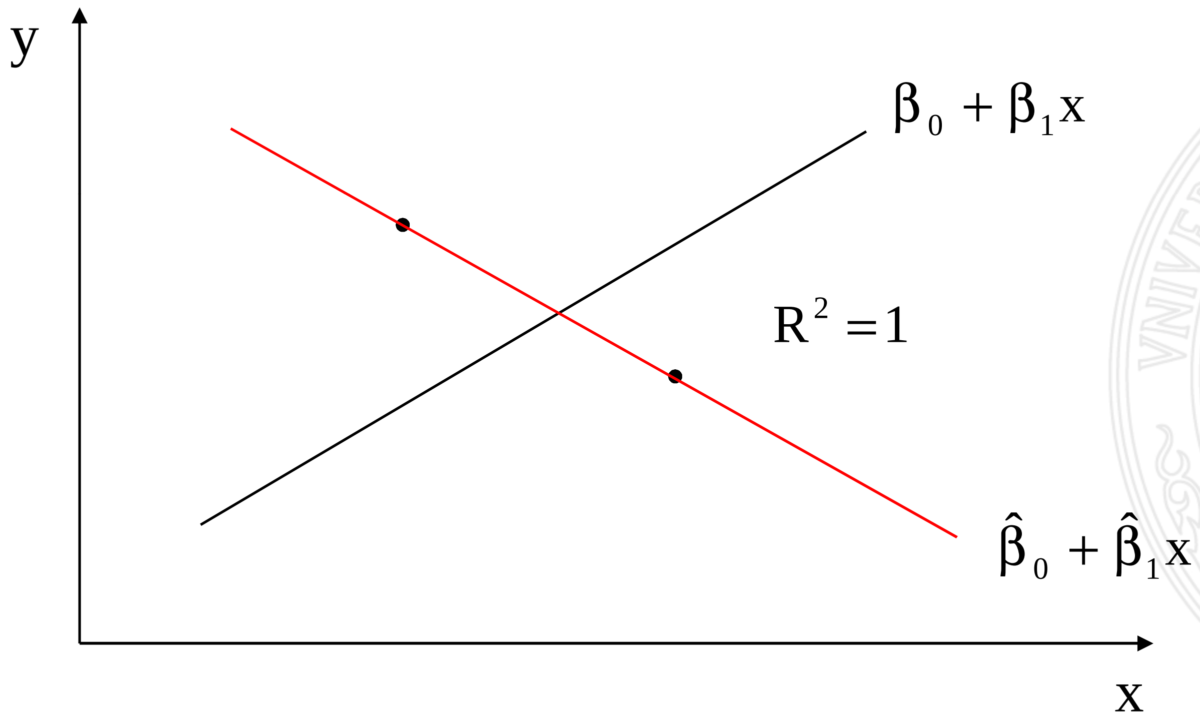


UNIVERSIDAD DE MURCIA

Si N = 2



Entonces, $K = N$ y $R^2 = 1$



UNIVERSIDAD DE MURCIA

- El coeficiente de determinación ajustado es:

$$R_a^2 = 1 - \frac{(N-1)\ SCE}{(N-K)\ STC} = 1 - \left(1 - R^2\right)\frac{N-1}{N-K} \qquad R_a^2 \leq 1$$

- Penaliza la inclusión de variables independientes

- Permite comparar algo mejor que el R^2 modelos con distinto número de regresores

- El R_a^2 puede ser negativo incluso en modelos con término constante

Modelo poblacional: $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon$

Modelo estimado: $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k + e$

- Los cambios de origen en las variables sólo afectan a β_0 y a $\hat{\beta}_0$
- Si y se multiplica por una constante c, entonces **todos** los coeficientes $(\beta_0, \beta_1, \ldots, \beta_k)$ y sus estimaciones MCO $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ se multiplican por c
- Si una x_j se multiplica (divide) por una constante c, entonces β_j y $\hat{\beta}_j$ se dividen (multiplican) por c
- Si las variables están en logaritmos los cambios de escala en las unidades de medida sólo afectan a β_0 y a $\hat{\beta}_0$
- El R-cuadrado es invariable a los cambios de unidades en las variables

Ejemplo: Cambio de origen en y

$$y^* = c + y$$

A. Efectos en el **modelo poblacional**: $y = \beta_0 + \beta_1 x + \varepsilon$

- El nuevo modelo es: $y^* = \beta_0^* + \beta_1^* x + \varepsilon^*$

- Deshaciendo el cambio de origen:

$$y + c = \beta_0^* + \beta_1^* x + \varepsilon^* \Rightarrow y = (\beta_0^* - c) + \beta_1^* x + \varepsilon$$

$$\beta_1^* = \beta_1 \qquad \beta_0^* = c + \beta_0$$

UNIVERSIDAD DE MURCIA

$$y^* = c + y$$

B. Efectos en el **modelo estimado**: $y = \hat{\beta}_0 + \hat{\beta}_1 x + e$

- El nuevo modelo estimado es: $y^* = \hat{\beta}_0^* + \hat{\beta}_1^* x + e^*$

$$\hat{\beta}_0^* = \overline{y}^* - \hat{\beta}_1^* \overline{x} \qquad \hat{\beta}_1^* = \frac{S_{xy^*}}{S_x^2}$$

- Deshaciendo el cambio de origen:

$$\hat{\beta}_1^* = \frac{S_{xy}}{S_x^2} = \hat{\beta}_1 \qquad \hat{\beta}_0^* = c + \overline{y} - \hat{\beta}_1 \overline{x} = c + \hat{\beta}_0$$

¿Qué pasa con el R-cuadrado?

¿Qué pasaría si el cambio de origen afectara a x?

Ejemplo: Cambio de escala en y

$$y^* = c \cdot y$$

A. Efectos en el **modelo poblacional**: $y = \beta_0 + \beta_1 x + \varepsilon$

- El nuevo modelo es: $y^* = \beta_0^* + \beta_1^* x + \varepsilon^*$

- Deshaciendo el cambio de escala:

$$c \cdot y = \beta_0^* + \beta_1^* x + \varepsilon^* \Rightarrow y = \frac{\beta_0^*}{c} + \frac{\beta_1^*}{c} x + \varepsilon$$

$$\beta_1^* = c \cdot \beta_1 \qquad \beta_0^* = c \cdot \beta_0$$

UNIVERSIDAD DE MURCIA

$$y^* = c \cdot y$$

B. Efectos en el **modelo estimado**: $y = \hat{\beta}_0 + \hat{\beta}_1 x + e$

- El nuevo modelo estimado es: $y^* = \hat{\beta}_0^* + \hat{\beta}_1^* x + e^*$

$$\hat{\beta}_0^* = \overline{y}^* - \hat{\beta}_1^* \overline{x} \qquad \hat{\beta}_1^* = \frac{S_{xy^*}}{S_x^2}$$

- Deshaciendo el cambio de escala:

$$\hat{\beta}_1^* = \frac{cS_{xy}}{S_x^2} = c\hat{\beta}_1 \qquad \hat{\beta}_0^* = c\overline{y} - c\hat{\beta}_1 \overline{x} = c\hat{\beta}_0$$

¿Qué pasa con el R-cuadrado?

¿Qué pasaría si el cambio de escala afectara a x?

- Modelo lineal en parámetros y variables: $y = \beta_0 + \beta_1 x + \varepsilon$

$$\beta_1 = \frac{\Delta E\left(y/x\right)}{\Delta x}$$ es el efecto marginal de x sobre $E(y/x)$

- Modelos lineales en parámetros, pero no en variables:
  - modelos que incluyen variables en logaritmos
  - funciones cuadráticas
  - otros

UNIVERSIDAD DE MURCIA

Modelos que incluyen variables en logaritmos:

- **Modelo log-log:** $\ln y = \beta_0 + \beta_1 \ln x + \varepsilon$

$$\beta_1 = \frac{\Delta E\left(\ln y / x\right)}{\Delta \ln x} \Rightarrow \beta_1 \approx \frac{\%\Delta E(y/x)}{\%\Delta x} = \text{elasticidad}$$

- **Modelo log-nivel:** $\ln y = \beta_0 + \beta_1 x + \varepsilon$

$$\beta_1 = \frac{\Delta E\left(\ln y / x\right)}{\Delta x} \Rightarrow 100 \cdot \beta_1 \approx \frac{\%\Delta E(y/x)}{\Delta x} = \text{semielasticidad}$$

- **Modelo nivel-log:** $y = \beta_0 + \beta_1 \ln x + \varepsilon$

$$\beta_1 = \frac{\Delta E\left(y / x\right)}{\Delta \ln x} \Rightarrow \frac{\beta_1}{100} \approx \frac{\Delta E(y/x)}{\%\Delta x} = \text{semielasticidad}$$

UNIVERSIDAD DE MURCIA

Ejemplo numérico:

- $\ln \hat{y} = 2 + 0,05x$

  Si $\Delta x = 1 \Rightarrow \Delta \hat{y}\% \approx (0.05 \cdot 100)\% = 5\%$

- $\hat{y} = 2 + 70 \ln x$

  Si $\Delta x\% = 1\% \Rightarrow \Delta \hat{y} \approx (70/100) = 0.7$ unidades

- $\ln \hat{y} = 2 - 2 \ln x$

  Si $\Delta x\% = 1\% \Rightarrow \Delta \hat{y}\% \approx -2\%$

UNIVERSIDAD DE MURCIA

Funciones cuadráticas:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

- Permiten captar efectos marginales crecientes o decrecientes

- El efecto marginal de x sobre y es:

$$\frac{\Delta E(y/x)}{\Delta x} \approx \beta_1 + 2\beta_2 x$$

Es creciente cuando $\beta_2 > 0$ y decreciente cuando $\beta_2 < 0$

Lo que hemos aprendido:

UNIVERSIDAD DE MURCIA

- Definir e interpretar el modelo de regresión múltiple
- El método de los Mínimos Cuadrados Ordinarios para estimar los coeficientes del modelo
- Interpretar las estimaciones
- Propiedades algebraicas de los estimadores MCO
- Utilizar el R-cuadrado y el R-cuadrado ajustado para medir la bondad del ajuste
- Cambio en las estimaciones MCO cuando cambian las unidades de medida de la variable dependiente o de las variables independientes
- Utilizar el modelo de regresión lineal para modelizar relaciones no lineales entre las variables:
    - el logaritmo neperiano permite trabajar con modelos de elasticidad constante y de semielasticidad constante
    - las funcionales cuadráticas permiten captar efectos marginales crecientes o decrecientes