# ECONOMETRÍA

JOSÉ ALBERTO MAURICIO

3

## REGRESIÓN LINEAL MÚLTIPLE II

### 3.1 VARIABLES EXPLICATIVAS BINARIAS

Departamento de Análisis Económico y Economía Cuantitativa
Universidad Complutense de Madrid

EctrGr-JAM-3-1.pdf



ucm.randomshock.com/ectrgr

Versión 2.4 - Enero 2024

## REQUISITOS

En esta Sección 3.1 se suponen bien conocidos [1] los aspectos teóricos y prácticos del Análisis de Regresión Lineal cubiertos en los Temas 1 - 2, y [2] todos los procedimientos descritos en las Secciones 1 - 13 de la guía *Introducción al Uso de EViews 4.1*.

## BIBLIOGRAFÍA PARA LA SECCIÓN 3.1

Hill, Griffiths, Lim (2018): Capítulo 7.
Wooldridge (2020): Capítulo 7.

Sección 14 de la guía *Introducción al Uso de EViews 4.1*.
En adelante IEV41 ...

## PARTE 1 - INTRODUCCIÓN

Las variables explicativas binarias (cualitativas, categóricas, ficticias, de intervención, "dummy variables", "dummies", "indicator variables", ...) se utilizan para considerar posibles diferencias en los valores de los parámetros (uno, varios, o todos) de un modelo entre dos o más grupos de observaciones exhaustivos y excluyentes. Por ejemplo:

➪ Considerar la estabilidad de los parámetros de un modelo entre dos grupos (como dos periodos consecutivos) de observaciones de una serie temporal.

➪ Considerar la homogeneidad de los parámetros de un modelo entre dos grupos (como hombres y mujeres) de observaciones de una sección cruzada.

## PARTE 2 - EJEMPLO ST04-CHOW-1.WF1

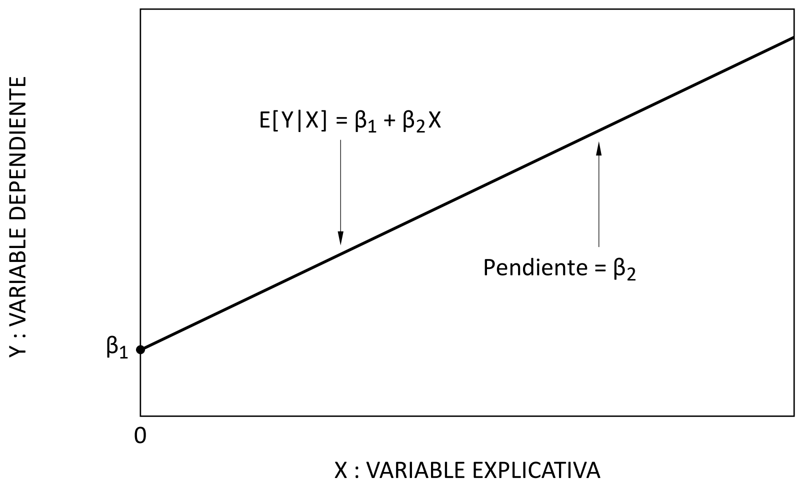
CONSUMO = CONSUMO AUTÓNOMO + PROPENSIÓN MARGINAL × RENTA DISPONIBLE + U

**Muestra Completa:** Desde 1985 hasta 1999 (15 observaciones). Dos grupos:

- **Grupo A:** Desde 1985 hasta 1989 (5 observaciones) : **DA = 1**, DB = 0.
- **Grupo B:** Desde 1990 hasta 1999 (10 observaciones) : DA = 0, **DB = 1**.

**Modelo Restringido:**

$$Y = \beta_1 + \beta_2 X + U. \tag{1}$$

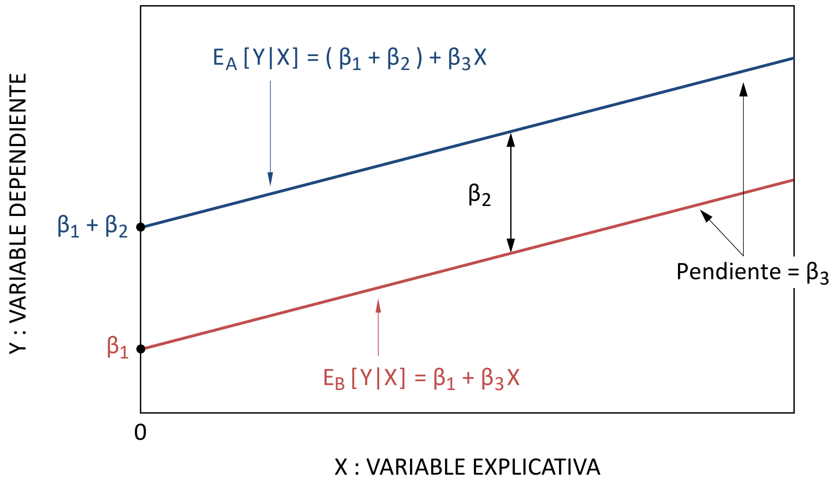



**CASO 1 - Modelo No Restringido - Versión 1:**

$$Y = \beta_1 + \beta_2 DA + \beta_3 X + U. \tag{2}$$

$$(2) \Rightarrow \begin{array}{l} \text{Grupo A (DA = 1)}: Y = (\beta_1 + \beta_2) + \beta_3 X + U. \\ \text{Grupo B (DA = 0)}: Y = \quad \beta_1 \quad + \beta_3 X + U. \end{array} \tag{3}$$

Por lo tanto, en (2): $\beta_1$ es el término constante del Grupo B (grupo "base"), $\beta_2$ es la diferencia entre los términos constantes de los grupos A y B (en ese orden), y $\beta_3$ es la pendiente común para ambos grupos. Por ejemplo, suponiendo que $\beta_2 > 0$ y $\beta_3 > 0$:




Los términos constantes de los grupos A y B son iguales cuando $\beta_2 = 0$ en (2). El contraste de $H_0 : \beta_2 = 0$ en (2) se puede llevar a cabo de varias formas (Tema 2) ...

**CASO 1 - Modelo No Restringido - Versión 2:**

$$Y = \beta_1 + \beta_2 DB + \beta_3 X + U. \tag{4}$$

**CASO 1 - Modelo No Restringido - Versión 3:**

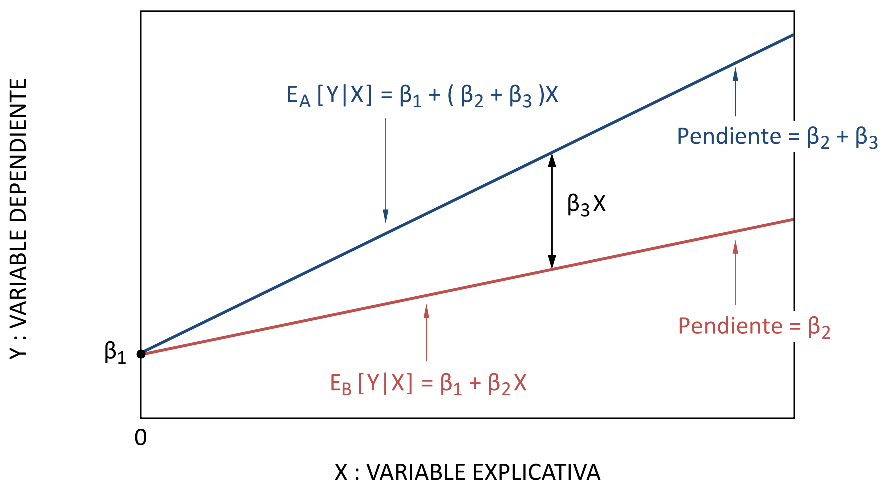
$$Y = \beta_1 DA + \beta_2 DB + \beta_3 X + U. \tag{5}$$

**CASO 2 - Modelo No Restringido – Versión 1:**

$$Y = \beta_1 + \beta_2 X + \beta_3 (DA \times X) + U. \tag{6}$$

$$(6) \Rightarrow \quad \begin{aligned} &\text{Grupo A (DA} = 1\text{)}: Y = \beta_1 + (\beta_2 + \beta_3) X + U. \\ &\text{Grupo B (DA} = 0\text{)}: Y = \beta_1 + \quad \beta_2 X \quad + U. \end{aligned} \tag{7}$$

Por lo tanto, en (6): $\beta_1$ es el término constante común para ambos grupos, $\beta_2$ es la pendiente del grupo B (grupo "base"), y $\beta_3$ es la diferencia entre las pendientes de los grupos A y B (en ese orden). Por ejemplo, suponiendo que $\beta_2 > 0$ y $\beta_3 > 0$:



Las pendientes de los grupos A y B son iguales cuando $\beta_3 = 0$ en (6). Como en el CASO 1, el contraste de $H_0: \beta_3 = 0$ en (6) se puede llevar a cabo de varias formas (Tema 2) ...

**CASO 2 - Modelo No Restringido - Versión 2:**

$$Y = \beta_1 + \beta_2 X + \beta_3 (DB \times X) + U. \tag{8}$$

**CASO 2 - Modelo No Restringido - Versión 3:**

$$Y = \beta_1 + \beta_2 (DA \times X) + \beta_3 (DB \times X) + U. \tag{9}$$

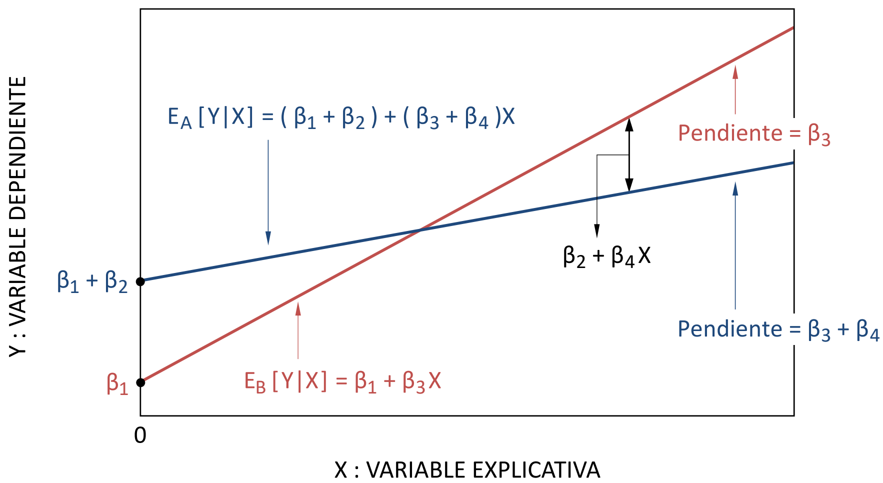
Los contrastes mencionados en los CASOS 1-2 sólo requieren el uso de estadísticos t. Sin embargo, alguno de los contrastes del CASO 3 requiere el uso de un estadístico F ...

**CASO 3 - Modelo No Restringido - Versión 1:**

$$Y = \beta_1 + \beta_2 DA + \beta_3 X + \beta_4 (DA \times X) + U. \tag{10}$$

$$(10) \Rightarrow \begin{array}{l} \text{Grupo A (DA} = 1\text{)}: Y = (\beta_1 + \beta_2) + (\beta_3 + \beta_4) X + U. \\ \text{Grupo B (DA} = 0\text{)}: Y = \quad \beta_1 \quad + \quad \beta_3 X \quad + U. \end{array} \tag{11}$$

Por lo tanto, en (10): $\beta_1$ es el término constante del Grupo B, $\beta_2$ es la diferencia entre los términos constantes de los grupos A y B, $\beta_3$ es la pendiente del Grupo B, y $\beta_4$ es la diferencia entre las pendientes de los grupos A y B. Por ejemplo, si $\beta_2 > 0$ y $\beta_4 < 0$:




Los términos constantes [resp. las pendientes] de los grupos A y B son iguales cuando $\beta_2 = 0$ [resp. $\beta_4 = 0$] en (10). En este CASO 3 hay varios contrastes interesantes ...

**CASO 3 - Modelo No Restringido - Versión 2:**

$$Y = \beta_1 + \beta_2 DB + \beta_3 X + \beta_4 (DB \times X) + U. \tag{12}$$

**CASO 3 - Modelo No Restringido - Versión 3:**

$$Y = \beta_1 DA + \beta_2 DB + \beta_3 (DA \times X) + \beta_4 (DB \times X) + U. \tag{13}$$

En la Sección 14 de IEV41 se puede encontrar el planteamiento, el desarrollo y el resultado comentados de todos los contrastes de hipótesis relevantes para los tres casos anteriores.

## PARTE 3 - EJEMPLO SC04-SALARIOS4.WF1

**Muestra Completa:** Sección cruzada de 528 personas trabajadoras clasificadas en diferentes grupos o categorías, según varios criterios o características.

**Modelo Restringido:**

$$\ln(SLRPH) = \beta_1 + \beta_2 EDUC + \beta_3 EXLP + U. \tag{14}$$

**Modelo con Clasificación por Género:**

$$\ln(SLRPH) = \beta_1 + \beta_2 MUJER + \beta_3 EDUC + \beta_4 EXLP + U. \tag{15}$$

En (15), $100 \times \beta_2$ representa la diferencia (logarítmica) porcentual entre los salarios de una mujer y de un hombre (en ese orden) con los mismos años de educación y de experiencia.

**Modelo con Clasificaciones por Varios Criterios:**

$$\begin{aligned} \ln(SLRPH) = \beta_1 &+ \beta_2 MUJER + \beta_3 ECIV + \beta_4 SIND \\ &+ \beta_5 ETN2 + \beta_6 ETN3 \\ &+ \beta_7 EDUC + \beta_8 EXLP + U. \end{aligned} \tag{16}$$

En (16), algunas hipótesis nulas interesantes serían las siguientes (frente a las alternativas que pudieran considerarse oportunas en cada caso):

- $H_0: \beta_2 = 0.$
- $H_0: \beta_3 = 0.$
- $H_0: \beta_4 = 0.$
- $H_0: \beta_5 = 0.$
- $H_0: \beta_6 = 0.$
- $H_0: \beta_5 - \beta_6 = 0.$
- $H_0: \beta_5 = \beta_6 = 0.$
- $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0.$

En los ejemplos de exámenes T2 (preguntas 11-20) y T4 (preguntas 1-12) disponibles en la página de la asignatura en Internet se consideran dos casos semejantes al de esta PARTE 3.

## PARTE 4 - RESUMEN

- Las variables binarias permiten clasificar de manera exhaustiva y excluyente todas las observaciones de una muestra (años, meses, personas, empresas, ...) en dos o más grupos o categorías en función de un criterio o de una característica determinados.
- Para cada grupo se define y se asocia una variable binaria (cualitativa) de manera que un 1 en dicha variable indica una observación dentro de ese grupo (una cualidad que está presente en esa observación) y un 0 indica una observación fuera de ese grupo (una cualidad que no está presente en esa observación).
- El carácter exhaustivo y excluyente de la clasificación se refleja en que la suma de las variables binarias asociadas con todos los grupos de dicha clasificación es exactamente igual a uno en todas las observaciones.
- En un modelo de regresión, las variables binarias se emplean para considerar distintos términos constantes y/o distintas pendientes en cada grupo de observaciones. Aunque hay muchas alternativas para especificar un modelo de regresión con variables binarias, una opción útil consiste en dejar fuera del modelo la variable binaria asociada con uno de los grupos. El grupo cuya variable binaria se deja fuera del modelo suele denominarse "grupo base" o "grupo de referencia".
- Los contrastes de igualdad, estabilidad, homogeneidad, o ausencia de cambio estructural entre grupos, pueden llevarse a cabo en muchos casos con estadísticos t, y en algunos otros empleando estadísticos F (que se pueden calcular a través de las sumas de cuadrados de residuos, o de los coeficientes de determinación, de los modelos restringido y no restringido correspondientes).
- En un modelo de regresión pueden incluirse al mismo tiempo varios conjuntos de variables binarias, para considerar simultáneamente varias clasificaciones en función de distintos criterios o características.

## APÉNDICE - EL TEST DE CHOW

En un modelo RLM, consideramos la posibilidad de que los parámetros tengan unos valores para cierto grupo de observaciones (grupo 1), y otros valores diferentes para el grupo de las observaciones restantes (grupo 2). Según esto, el modelo completo puede escribirse explícitamente en dos partes, una para cada grupo de observaciones, como

$$\underset{N_1\times 1}{\mathbf{Y}_1} = \underset{N_1\times K}{\mathbf{X}_1}\ \underset{K\times 1}{\boldsymbol{\beta}_1} + \underset{N_1\times 1}{\mathbf{U}_1}\ ,\quad \underset{N_2\times 1}{\mathbf{Y}_2} = \underset{N_2\times K}{\mathbf{X}_2}\ \underset{K\times 1}{\boldsymbol{\beta}_2} + \underset{N_2\times 1}{\mathbf{U}_2}\ ,$$

con $N_1 > K$ y $N_2 > K$ , o bien, de manera conjunta, como

$$\underbrace{\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}}_{\substack{\downarrow \\ \mathbf{Y}}} = \underbrace{\begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}}_{\substack{\downarrow \\ \mathbf{X}}} \underbrace{\begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}}_{\substack{\downarrow \\ \boldsymbol{\beta}}} + \underbrace{\begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix}}_{\substack{\downarrow \\ \mathbf{U}}}. \qquad \text{[A1]}$$

El llamado **Test de Chow** es el contraste de $H_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ frente a $H_1: \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ en un modelo de este tipo (contraste de **estabilidad** o **ausencia de cambio estructural** en los parámetros del modelo). El valor del estadístico F se puede calcular de acuerdo con [67] en la Sección 2.3 teniendo en cuenta que el modelo restringido por $H_0$ queda en este caso

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \boldsymbol{\beta}_2 + \begin{bmatrix} \mathbf{U}_1^* \\ \mathbf{U}_2^* \end{bmatrix}. \qquad [A2]$$

En el modelo original (no restringido) [A1], SCR $= \hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2$, donde $\hat{\mathbf{u}}_1$ y $\hat{\mathbf{u}}_2$ son los residuos MCO en las regresiones (estimadas por separado) de $\mathbf{y}_1$ sobre $\mathbf{X}_1$ y de $\mathbf{y}_2$ sobre $\mathbf{X}_2$, respectivamente:

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}}_{\downarrow \atop \mathbf{y}} = \underbrace{\begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}}_{\downarrow \atop \mathbf{X}} \underbrace{\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix}}_{\downarrow \atop \hat{\boldsymbol{\beta}}} + \underbrace{\begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix}}_{\downarrow \atop \hat{\mathbf{u}}},$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \end{bmatrix},$$

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= (\mathbf{y}_1'\mathbf{y}_1 + \mathbf{y}_2'\mathbf{y}_2) - [\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2'] \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \end{bmatrix} \\ &= \underbrace{(\mathbf{y}_1'\mathbf{y}_1 - \hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{y}_1)}_{\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1} + \underbrace{(\mathbf{y}_2'\mathbf{y}_2 - \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y}_2)}_{\hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2}. \end{aligned}$$

Por su parte, en el modelo restringido [A2], SCR$^* = \hat{\mathbf{u}}_*'\hat{\mathbf{u}}_*$, donde $\hat{\mathbf{u}}_*$ son los residuos MCO en la regresión de $[\mathbf{y}_1', \mathbf{y}_2']'$ sobre $[\mathbf{X}_1', \mathbf{X}_2']'$. En consecuencia, el valor calculado del estadístico F de acuerdo con [67] en la Sección 2.3 queda en este caso

$$F = \frac{N_1 + N_2 - 2K}{K} \times \frac{\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_* - (\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2)}{\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2}, \qquad [A3]$$

que bajo $H_0$ debe proceder de una distribución $F(K, N_1 + N_2 - 2K)$.

El uso del Test de Chow en la práctica presenta ciertas limitaciones y dificultades que se pueden solventar, en muchos casos, mediante el empleo de variables explicativas binarias (IEV4.1: Sección 14). Esta opción suele ser preferible al empleo del Test de Chow como instrumento para realizar contrastes de estabilidad, homogeneidad o ausencia de cambio estructural.