

# Tema 7: Variables Ficticias

Máximo Camacho

# Variables ficticias

- Bloque I: El modelo lineal clásico
  - × Tema 1: Introducción a la econometría
  - × Tema 2: El modelo de regresión lineal
  - × Tema 3: El método MCO
  - × Tema 4: Propiedades de la estimación MCO
  - × Tema 5: Inferencia y predicción
- Bloque II: Extensiones al modelo lineal clásico
  - × Tema 6: Multicolinealidad
  - × ***Tema 7: Variables ficticias***
  - × Tema 8: Heteroscedasticidad
  - × Tema 9: Endogeneidad

# Descripción de la clase

- Introducción
- Variables ficticias únicas con dos estados
- Variables ficticias en casos más generales
  - Más de dos estados
  - Más de una ficticia
- Siempre usaremos datos reales

# 1. Introducción

## 1.1. Ejemplo de clase

- Imaginemos que en una región (California) los responsables de educación quieren estudiar notas en 420 colegios en función estudiantes por profesor. Datos en 1998

- Notas  $Y_i$
- Ratio estudiantes por profesor  $X_i$  (REP)
- Aunque no la incluiremos como explicativa, tenemos información

Porcentaje de alumnos que no hablan bien el idioma (PNI)

- ¿Cómo estimamos esta relación?
- Modelo lineal clásico

# 1. Introducción

## 1.2. Supuestos del modelo lineal clásico

- Suponemos **relación lineal** entre las variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad Y = X\beta + \varepsilon \quad Y_i = \chi_i' \beta + \varepsilon_i$$

- Supuestos

- { Exogeneidad débil  $E(\varepsilon_i / \chi_i) = E(\varepsilon_i) = 0$

- } Muestras aleatorias  $E(\varepsilon_i / \chi_j) = E(\varepsilon_i) = 0 \quad E(\varepsilon_i \varepsilon_j) = E(\varepsilon_i)E(\varepsilon_j) = 0$

Momentos cuartos finitos  $0 < E(\varepsilon_i^4) < \infty, 0 < E(X_{1i}^4) < \infty, \dots, 0 < E(X_{ki}^4) < \infty$

No multicolinealidad exacta  $X_1, \dots, X_n$  no son linealmente dependientes

Normalidad  $\varepsilon / X \sim N$

- Homoscedasticidad  $\text{var}(\varepsilon_i / X) = \sigma^2 \mathbf{V}i$

## 2. Variables ficticias

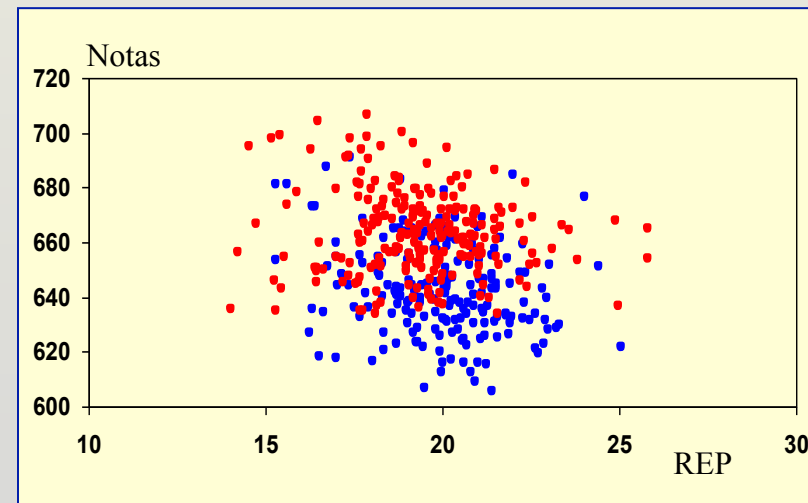
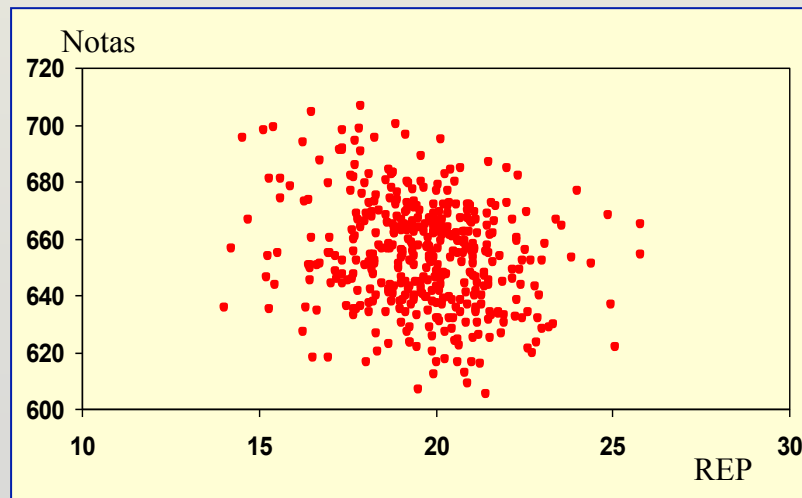
### ■ ¿Cómo tratar información cualitativa?

$$\text{Notas}_i = \beta_0 + \beta_1 \text{REP}_i + \varepsilon_i$$

$$\hat{\text{Notas}}_i = 698.93 - 2.27 \text{REP}_{1i}$$

(9.46)                      (0.48)

- Definimos “centros con problemas de aprendizaje” aquellos con PNI > 10%.
- ¿Se cumple la relación anterior para esos 192 centros (en azul)?



## 2. Variables ficticias

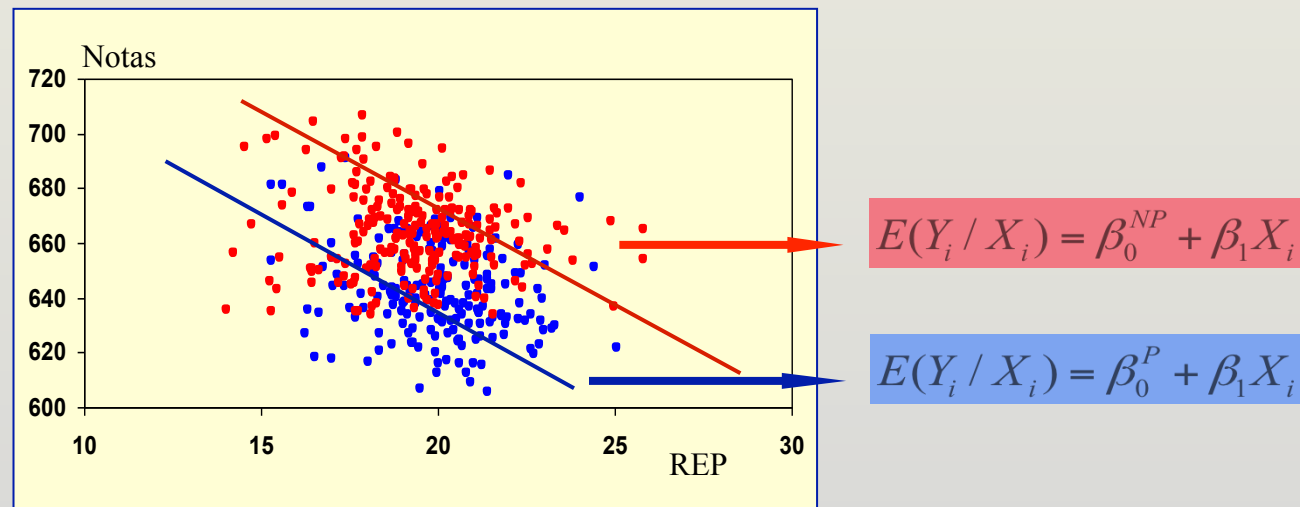
- Permiten tratar información cualitativa
- Ejemplos
  - Sexo en la determinación de salarios
  - Estación del año en el consumo de helados
  - Pertener a la UE en la determinación del crecimiento económico
  - Entrar en bancarrota en la predicción de beneficios
- Variables binarias cero-uno
- Se introducen en el modelo como explicativas adicionales

### 3. Variables ficticias únicas con dos estados

#### 3.1. Variables ficticias aditivas

- Supondremos que la “calidad” sólo afecta a la constante

- Si no tenemos en cuenta la calidad  $E(Y_i / X_i) = \beta_0 + \beta_1 X_i$
- 192 primeros tienen problemas de aprendizaje (  $PNI \geq 10\%$  )
- Esperamos que el “componente autónomo” de sus notas sea menor





## 3. Variables ficticias únicas con dos estados

### 3.1. Variables ficticias aditivas

#### ■ ¿Podemos estimar MCO por submuestras?

- Nada asegura que la pendiente sea la misma

× MCO en NP  $\hat{Notas}_i = 687.88 - 2.24REP_i$

× MCO en P  $\hat{Notas}_i = 682.24 - 0.96REP_i$

#### ■ Solución: variables ficticias aditivas

- Creamos  $FP_i$  valga 1 si el colegio tiene problemas de aprendizaje ( $PNI \geq 10\%$ )
- Creamos  $FNP_i$  valga 1 si el colegio no tiene problemas de aprendizaje ( $PNI < 10\%$ )

#### ■ En principio, vamos a añadir aditivamente las dos ficticias

$$Y_i = a_0 + b_0 FNP_i + c_0 FP_i + \beta_1 X_i + \varepsilon_i \quad \left\{ \begin{array}{l} i \in NP \Rightarrow E(Y_i / X_i) = a_0 + b_0 + \beta_1 X_i \\ i \in P \Rightarrow E(Y_i / X_i) = a_0 + c_0 + \beta_1 X_i \end{array} \right.$$

### 3. Variables ficticias únicas con dos estados

#### 3.1. Variables ficticias aditivas

- En principio la ficticia parece que no viola ningún supuesto clásico
- Problema: *trampa de las ficticias*

- El modelo propuesto  $Y_i = a_0 + b_0 FNP_i + c_0 FP_i + \beta_1 X_i + \varepsilon_i$

- En forma matricial  $Y = X\beta + \varepsilon$

- Donde  $X$  presenta multicolinealidad exacta (véase tema 6)

$$X = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} & X_1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{0} & \mathbf{1} & X_{192} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & X_{193} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & X_{420} \end{pmatrix}$$

- Por tanto no podemos encontrar de forma única  $\hat{\beta} = (X'X)^{-1} X'Y$

### 3. Variables ficticias únicas con dos estados

#### 3.1. Variables ficticias aditivas

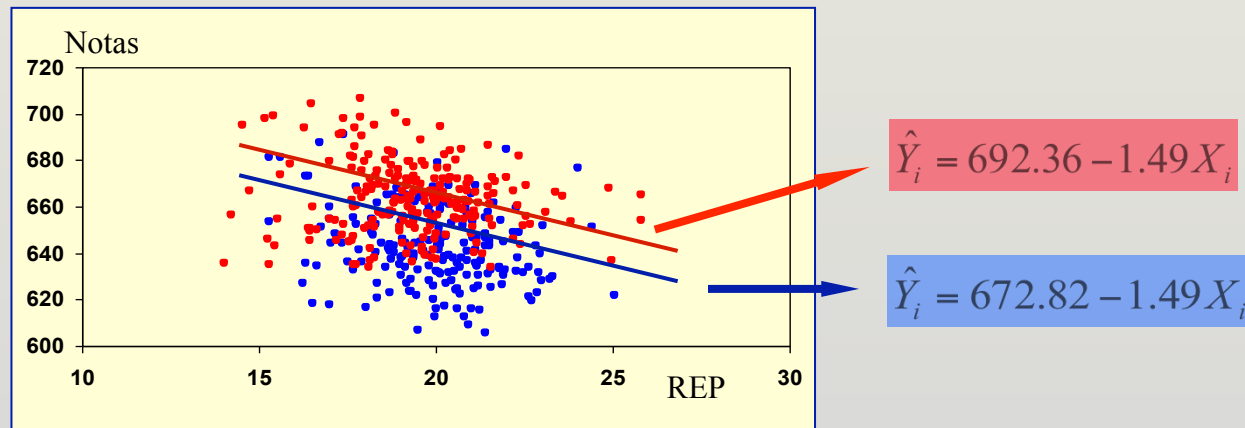
##### ■ Solución 1 a la *trampa de las ficticias*

- No incluimos la constante

× Partimos del modelo  $Y_i = a_0 + b_0 FNP_i + c_0 FP_i + \beta_1 X_i + \varepsilon_i$

× Usamos la relación  $1 = FNP_i + FP_i$   $Y_i = (a_0 + b_0)FNP_i + (a_0 + c_0)FP_i + \beta_1 X_i + \varepsilon_i$

× Estimamos  $\widehat{Notas}_i = 692.36 FNP_i + 672.82 FP_i - 1.49 REP_i$



### 3. Variables ficticias únicas con dos estados

#### 3.1. Variables ficticias aditivas

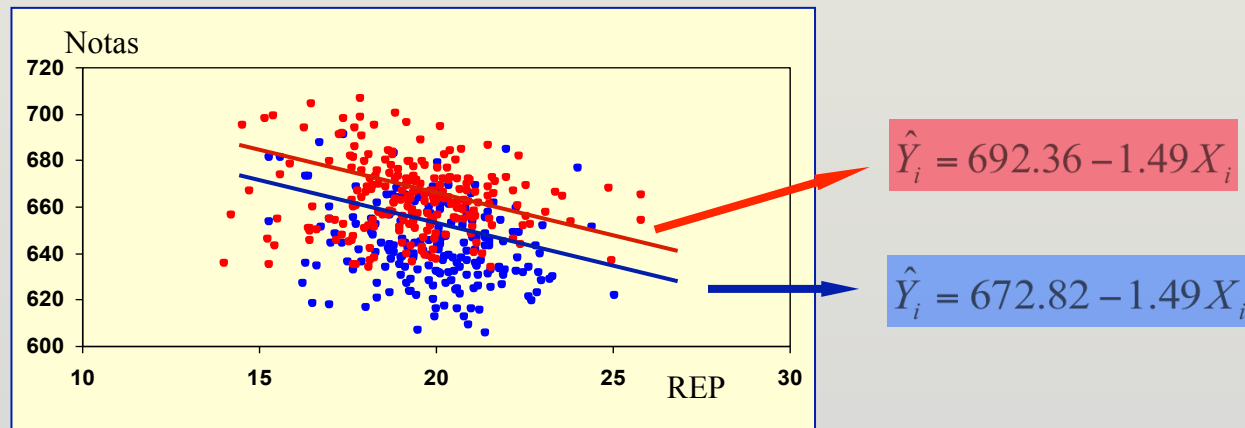
##### ■ Solución 2 a la *trampa de las ficticias*

- No incluimos alguna explicativa

- × Partimos del modelo  $Y_i = a_0 + b_0 FNP_i + c_0 FP_i + \beta_1 X_i + \varepsilon_i$

- × Usamos la relación  $FNP_i = 1 - FP_i$   $Y_i = (a_0 + b_0) + (c_0 - b_0)FP_i + \beta_1 X_i + \varepsilon_i$

- × Estimamos  $\widehat{Notas}_i = 692.36 - 19.53FP_i - 1.49REP_i$



### 3. Variables ficticias únicas con dos estados

#### 3.1. Variables ficticias aditivas

##### ■ Contrastes

- Tenemos que contrastar si efectivamente hay dos constantes distintas

$$Notas_i = \beta_0 + \lambda_0 FP_i + \beta_1 X_i + \varepsilon_i$$

$$H_0 : \lambda_0 = 0$$

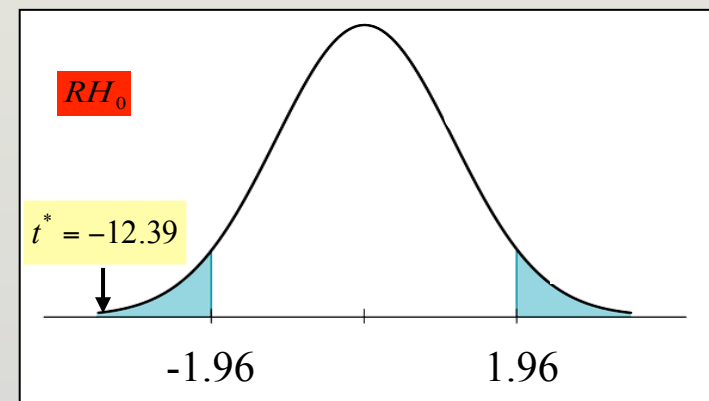
$$H_a : \lambda_0 \neq 0$$



$$No\hat{t}as_i = \underset{(8.12)}{692.36} - \underset{(1.57)}{19.53} FP_i - \underset{(0.41)}{1.49} X_i$$

$$t^* = \frac{\hat{\lambda}_0}{\sqrt{\hat{v}ar(\hat{\lambda}_0)}} = -12.39$$

$$t_{417,0.025} = 1.96$$

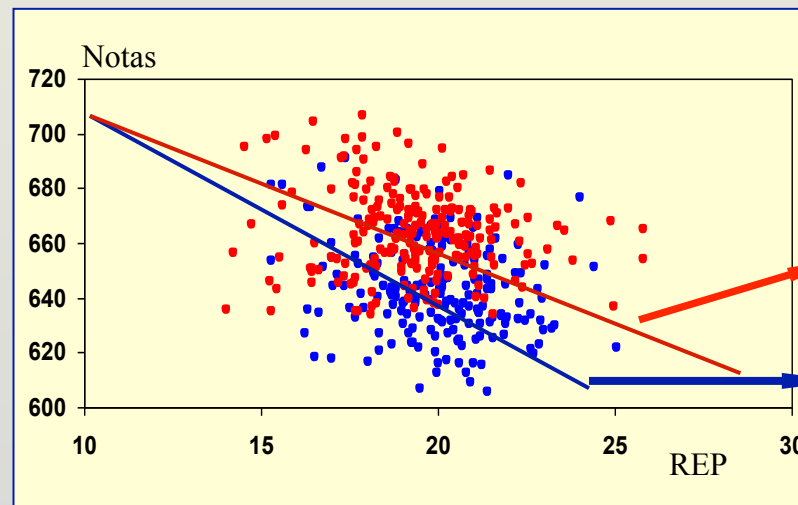


### 3. Variables ficticias únicas con dos estados

#### 3.2. Variables ficticias multiplicativas

■ Supondremos que la “calidad” sólo afecta a la pendiente

- Si no tenemos en cuenta la calidad  $E(Y_i / X_i) = \beta_0 + \beta_1 X_i$
- 192 primeros tienen problemas de aprendizaje (  $PNI \geq 10\%$  )
- Esperamos: reducir REP tenga más efecto en notas en los colegios  $P$  ( $PNI \geq 10\%$ )



$$E(Y_i / X_i) = \beta_0 + \beta_1^{NP} X_i$$

$$E(Y_i / X_i) = \beta_0 + \beta_1^P X_i$$

### 3. Variables ficticias únicas con dos estados

#### 3.2. Variables ficticias multiplicativas

- ¿Podemos estimar MCO por submuestras?

- Nada asegura que la constante estimada sea la misma

- × MCO en NP  $\hat{Notas}_i = 687.88 - 2.24REP_i$

- × MCO en P  $\hat{Notas}_i = 682.24 - 0.96REP_i$

- Solución: variables ficticias multiplicativas

- Creamos  $FP_i$  valga 1 si el colegio tiene problemas de aprendizaje ( $PNI \geq 10\%$ )
- Creamos  $FNP_i$  valga 1 si el colegio no tiene problemas de aprendizaje ( $PNI < 10\%$ )

- En principio, vamos a añadir multiplicativamente las dos ficticias

$$Y_i = a_0 + \beta_1 X_i + b_0 FNP_i X_i + c_0 FP_i X_i + \varepsilon_i \left\{ \begin{array}{l} i \in NP \Rightarrow E(Y_i / X_i) = a_0 + (\beta_1 + b_0) X_i \\ i \in P \Rightarrow E(Y_i / X_i) = a_0 + (\beta_1 + c_0) X_i \end{array} \right.$$

### 3. Variables ficticias únicas con dos estados

#### 3.2. Variables ficticias multiplicativas

- En principio la ficticia parece que no viola ningún supuesto clásico

- Problema: *trampa de las ficticias*

- El modelo propuesto  $Y_i = a_0 + \beta_1 X_i + b_0 FNP_i X_i + c_0 FP_i X_i + \varepsilon_i$

- En forma matricial  $Y = X\beta + \varepsilon$

- Donde  $X$  presenta multicolinealidad exacta (véase tema 6)

$$X = \begin{pmatrix} 1 & X_1 & 0 & X_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{192} & 0 & X_{192} \\ \hline 1 & X_{193} & X_{193} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{420} & X_{420} & 0 \end{pmatrix}$$

- Por tanto no podemos encontrar de forma única  $\hat{\beta} = (X'X)^{-1} X'Y$



### 3. Variables ficticias únicas con dos estados

#### 3.2. Variables ficticias multiplicativas

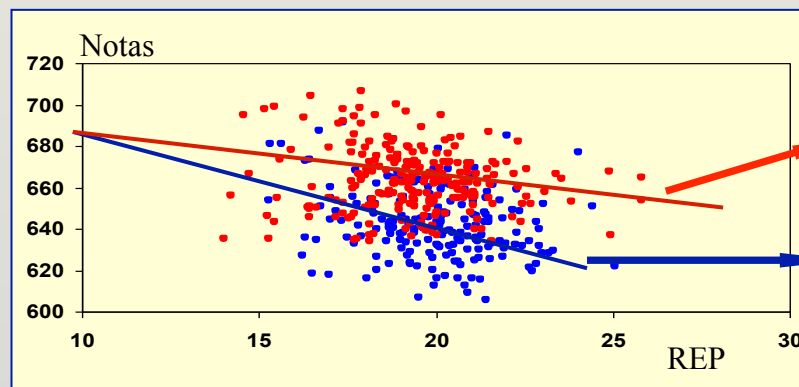
##### ■ Solución 1 a la *trampa de las ficticias*

- No incluimos la pendiente

× Partimos del modelo  $Y_i = a_0 + \beta_1 X_i + b_0 FNP_i X_i + c_0 FP_i X_i + \varepsilon_i$

× Usamos la relación  $1 = FNP_i + FP_i$   $Y_i = a_0 + (\beta_1 + b_0) FNP_i X_i + (\beta_1 + c_0) FP_i X_i + \varepsilon_i$

× Estimamos  $No\hat{t}as_i = 684.47 - 1.08 FNP_i X_i - 2.07 FP_i X_i$



$$No\hat{t}as_i = 684.47 - 1.08 X_i$$

$$No\hat{t}as_i = 684.47 - 2.07 X_i$$

### 3. Variables ficticias únicas con dos estados

#### 3.2. Variables ficticias multiplicativas

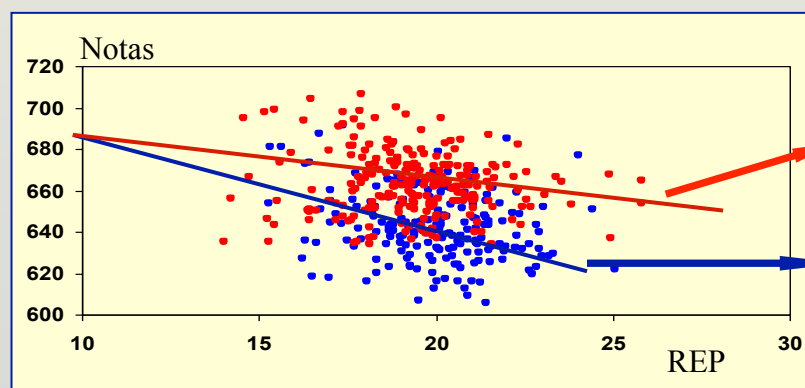
##### ■ Solución 2 a la *trampa de las ficticias*

- No incluimos una de las ficticias

× Partimos del modelo  $Y_i = a_0 + b_0 FNP_i X_i + c_0 FP_i X_i + \beta_1 X_i + \varepsilon_i$

× Usamos la relación  $FNP_i = 1 - FP_i$   $Y_i = a_0 + (\beta_1 + b_0) X_i + (c_0 - b_0) FP_i X_i + \varepsilon_i$

× Estimamos  $Notas_i = 684.47 - 1.08 X_i - 0.99 FP_i X_i$



$$Notas_i = 684.47 - 1.08 X_i$$

$$Notas_i = 684.47 - 2.07 X_i$$

### 3. Variables ficticias únicas con dos estados

#### 3.2. Variables ficticias multiplicativas

##### ■ Contrastes

- Tenemos que contrastar si efectivamente hay dos pendientes distintas

$$\text{Notas}_i = \beta_0 + \beta_1 X_i + \lambda_1 FP_i X_i + \varepsilon_i$$

$$H_0 : \lambda_1 = 0$$

$$H_a : \lambda_1 \neq 0$$

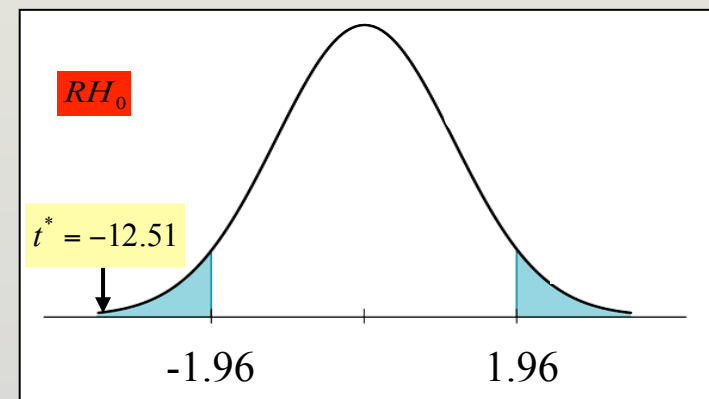


$$\widehat{\text{Notas}}_i = 684.47 - 1.08 X_i - 0.99 FP_i X_i$$

(8.16)            (0.42)            (0.07)

$$t^* = \frac{\hat{\lambda}_1}{\sqrt{\widehat{\text{var}}(\hat{\lambda}_1)}} = -12.51$$

$$t_{417,0.025} = 1.96$$

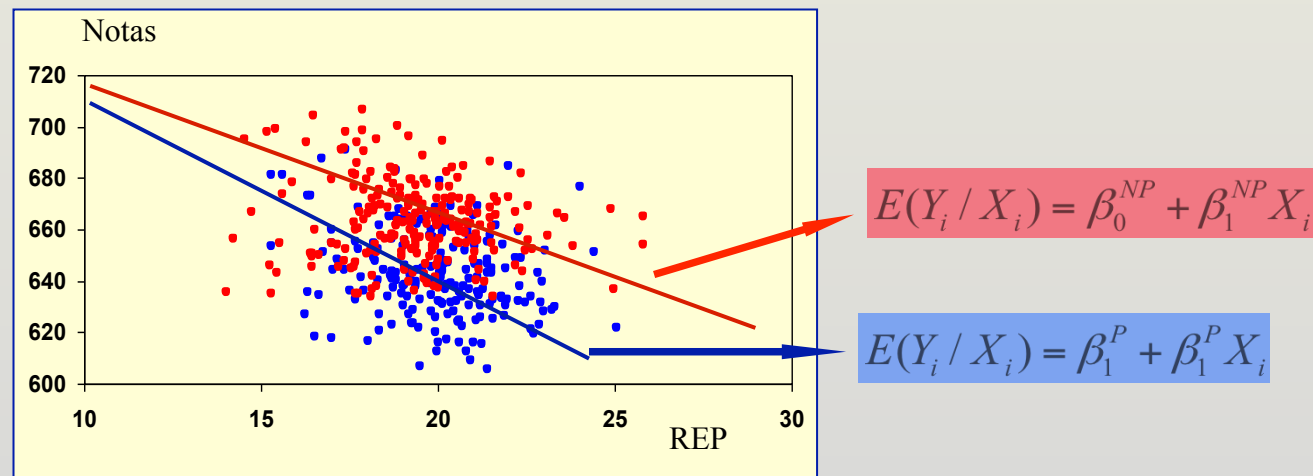


### 3. Variables ficticias únicas con dos estados

#### 3.3. Variables ficticias aditivas y multiplicativas

■ Supondremos que la “calidad” afecta la constante y pendiente

- Si no tenemos en cuenta la calidad  $E(Y_i / X_i) = \beta_0 + \beta_1 X_i$
- Si la tenemos en cuenta, esperamos
  - Colegios NP tengan sistemáticamente más notas
  - Reducir REP tenga más efecto en notas en los colegios P (PNI  $\geq 10\%$ )



### 3. Variables ficticias únicas con dos estados

#### 3.3. Variables ficticias aditivas y multiplicativas

- Según la discusión anterior, para evitar trampa de las ficticias

$$\text{Notas}_i = \beta_0 + \lambda_0 FP_i + \beta_1 X_i + \lambda_1 FP_i X_i + \varepsilon_i$$

- Contrastes: se conoce como **contraste de cambio estructural o de Chow**

$$H_0 : \lambda_0 = \lambda_1 = 0$$

$$H_a : \text{no } H_0$$



$$\widehat{\text{Notas}}_i = \underset{(10.51)}{682.24} + \underset{(16.71)}{5.63} FP_i - \underset{(0.53)}{0.96} X_i - \underset{(0.84)}{1.27} FP_i X_i \Rightarrow e'e = 104903$$

$$\widehat{\text{Notas}}_i = 698.93 - 2.28 X_i \Rightarrow e_R'e_R = 144315$$

$$F^* = \frac{(e_R'e_R - e'e)/q}{e'e/(n-K)} = \frac{(144315 - 104903)/2}{104903/(420 - 4)} = 78.14 > F_{2,\infty,0.05} = 3.00 \Rightarrow RH_0$$

- Cuidado con la multicolinealidad  $t_{\lambda_0}^* = 0.53$   $t_{\lambda_1}^* = -1.51$

- Ninguna de las ficticias son significativas individualmente por multicolinealidad
- No hacer contrastes individuales  $\Rightarrow$  mejor incluirlas una a una

## 4. Variables ficticias más generales

### 4.1. Más de dos estados

#### ■ Supongamos

- Mismo ejemplo notas y ratio estudiantes por profesor
- Calidad: colegios con problemas de aprendizaje en función  $PNI$
- Los problemas de aprendizaje afectan: ahora distinguimos tres estados
  - ▶ Colegios sin problemas:  $PNI_i \in [0-10) \Rightarrow FNP_i = 1$
  - ▶ Colegios con pocos problemas:  $PNI_i \in [10-50) \Rightarrow FPP_i = 1$
  - ▶ Colegios con muchos problemas:  $PNI_i \in [50-100) \Rightarrow FMP_i = 1$
- Sólo afectan al componente autónomo
- No podemos incluir las tres ficticias y constante (*trampa de las ficticias*)
- Solución
  - ▶ Incluir dos ficticias y término constante

## 4. Variables ficticias más generales

### 4.1. Más de dos estados

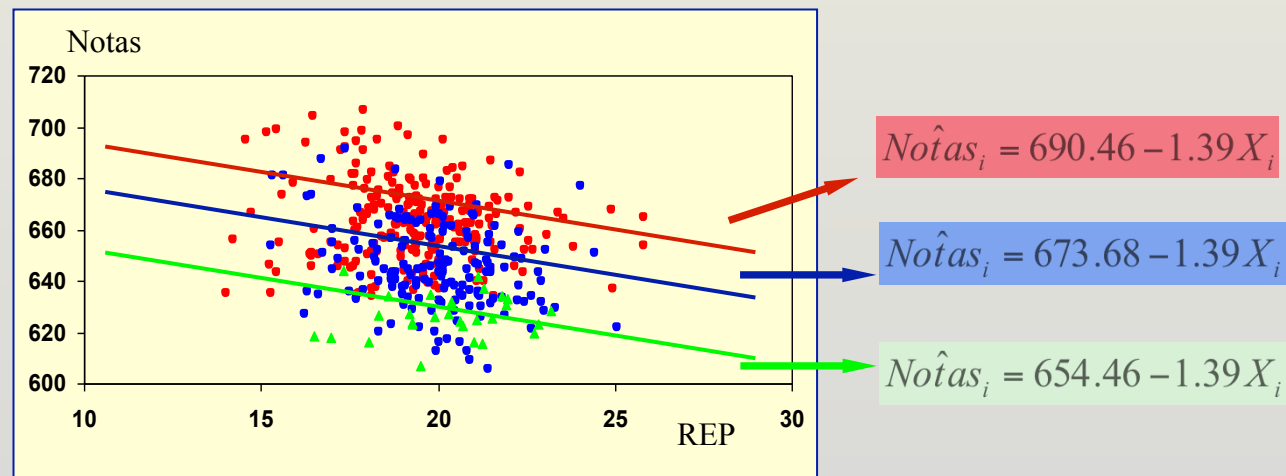
- Partimos del modelo

$$\hat{No\acute{t}a}_i = 690.46 - 16.78 FPP_i - 36.00 FMP_i - 1.39 X_i \Rightarrow e'e = 96662$$

(7.78)      (1.57)      (3.07)      (0.39)

- Contraste significatividad de las ficticias

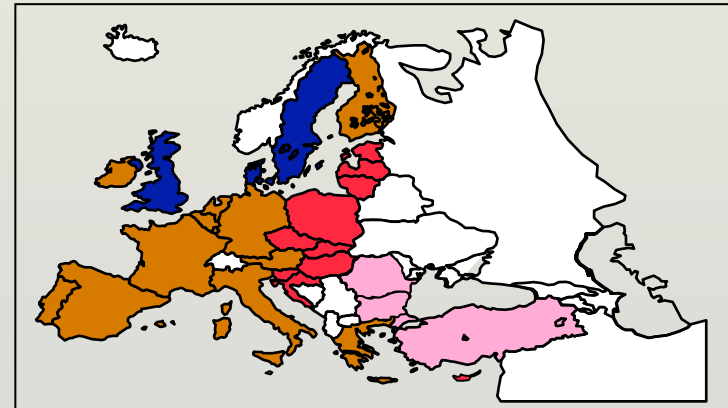
$$F^* = 102.53 > F_{2,\infty,0.05} = 3.00 \Rightarrow RH_0$$



## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

- Nuevo ejemplo: Grado de sincronización ciclo económico UE
  - ¿Hay más sincronización entre los que ya pertenecen?
  - ¿Afecta tener frontera?
- ¿Cómo medimos sincronización entre dos países?
  - Datos del IPI 1990.1-2004.3
  - “Correlación” entre ellos (dos a dos)
- ¿Cuántos países?
  - 15 Unión Europea
  - 12 Acceden excepto Malta y Bulgaria
  - 1 Negocia: Turquía
  - 4 países industrializados:  
( EE.UU, Japón, Canada, Noruega)
- Tamaño muestral: 435 datos de correlación





## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

#### ■ Supongamos

- La sincronización puede ser mayor
  - ▶ Ya pertenecían a UE antes de ampliación: creamos ficticia  $FUE_i = 1$  cuando la sincronización se mida entre dos de los 15
  - ▶ Comparten frontera: creamos ficticia  $FF_i = 1$  cuando la sincronización se mida entre dos que comparten frontera
- Sólo afectan al componente autónomo
- No podemos todas las ficticias y constante (*trampa de las ficticias*)
- Solución
  - ▶ Incluir dos ficticias  $FUE_i$  y  $FF_i$  y término constante
  - ▶ Posibilidad de *efecto interacción*

## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

- Partimos del modelo

$$y_i = a_0 + b_0 FUE_i + c_0 FF_i + d_0 FUE_i FF_i + \varepsilon_i$$

- Posibilidades

	$FUE_i = 0$	$FUE_i = 1$
$FF_i = 0$	$a_0$	$a_0 + b_0$
$FF_i = 1$	$a_0 + c_0$	$a_0 + c_0 + b_0 + d_0$

Efecto interacción

## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

) ¿Existe correlación positiva?

$$y_i = a + \varepsilon_i$$

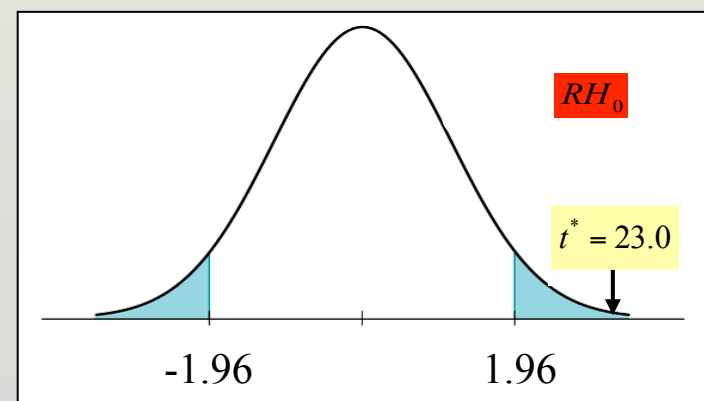
■ Estimación MCO

$$\hat{y}_i = \frac{0.23}{(0.01)}$$

■ ¿Es significativa?

$$\left. \begin{array}{l} H_0 : a = 0 \\ H_a : a \neq 0 \end{array} \right\}$$

$$t = \frac{0.23}{0.01} = 23.0$$



## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

( ¿Hay más sincronización entre los que ya pertenecen a la UE?

- Creamos  $FUE_i = 1$  si la sincronización se mide entre dos UE
- Proponemos el modelo

$$y_i = a + bFUE_i + \varepsilon_i$$

- Estimación MCO

$$\hat{y}_i = \underset{(0.01)}{0.20} + \underset{(0.02)}{0.08} FUE_i$$

- Contraste

$$\left. \begin{array}{l} H_0 : b = 0 \\ H_a : b \neq 0 \end{array} \right\}$$

$$t^* = \frac{0.08}{0.02} = 4.0$$

Sí afecta ser de UE

## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

) ¿Hay más sincronización entre los que tienen frontera?

- Creamos  $FF_i = 1$  si la sincronización se mide entre dos con frontera
- Proponemos el modelo

$$y_i = a + bFUE_i + cFF_i + \varepsilon_i$$

- Estimación MCO

$$\hat{y}_i = 0.20 + 0.08 FUE_i + 0.13 FF_i$$

(0.01)      (0.02)      (0.04)

- Contraste

$$\left. \begin{array}{l} H_0 : c = 0 \\ H_a : c \neq 0 \end{array} \right\}$$

$$t^* = \frac{0.13}{0.04} = 3.2$$

Sí afecta tener frontera

## 4. Variables ficticias más generales

### 4.2. Más de una variable ficticia

#### ¿Existe efecto interacción?

- Proponemos el modelo  $y_i = a + bFUE_i + cFF_i + dFUE_iFF_i + \varepsilon_i$

- Estimación MCO

$$\hat{y}_i = 0.20 + 0.06 FUE_i + 0.09 FF_i + 0.09 FUE_i FF_i$$

(0.01)      (0.02)                      (0.04)                      (0.07)

- Contraste

$$\left. \begin{array}{l} H_0 : d = 0 \\ H_a : d \neq 0 \end{array} \right\}$$

$$t^* = \frac{0.09}{0.07} = 1.3 < 1.96$$

¿Hay efecto interacción?

- A favor del efecto interacción
  - × Hay razones económicas para pensar que sí
  - × Si hay multico el  $|t^*| \ll 1$  para eliminarla y no es el caso
- En contra: la multico suele afectar a la significatividad de varias explicativas

## 5. ¿Qué hemos aprendido?

- Concepto de variables ficticias
  - Secuencias de 1 y 0
  - Nos sirven para introducir en el modelo aspectos cualitativos
- Ficticias aditivas y multiplicativas
  - Al principio, únicas y con dos estados
  - Cuidado con trampa de ficticias
  - Contraste de cambio estructural
- Más tarde:
  - Más de dos estados: introducir tantas ficticias como estados menos una
  - Más de una ficticia: posibilidad de efecto interacción