TEMA 3. EL MODEL DE REGRESSIÓ MÚLTIPLE: ESTIMACIÓ

Joan Llull

Materials: http://pareto.uab.cat/jllull

Tutories: dijous de 11:00 a 13:00h
(concertar cita per email)
—Despatx B3-1132—

joan.llull [at] movebarcelona [dot] eu

Continguts

T3. El model de regressió múltiple: Estimació

Continguts

T3. El model de regressió múltiple: Estimació

1. El model de regressió amb K variables
2. L'estimador Mínim Quadrat Ordinari: MQO
3. Distribució de l'estimador MQO
4. L'estimador Mínim Quadrat Restringit: MQR
5. Aplicacions
6. Variables fictícies

Motivació

La **publicitat**, el nombre d'**establiments** es comercialitza, la quantitat de **calci** de la llet... poden ser altres determinants del nombre de ampolles de llet venudes (a més del preu)

El **capital humà**, el nivell d'**obertura internacional**, la mida del **sector públic**,... poden ser altres determinants del PIB per càpita d'un país (a més del capital per càpita)

Els anys d'**experiència** en el sector, l'**antiguitat**, la **mida de l'empresa**,... poden ser altres determinants del salari (a més dels anys d'educació)

$\Rightarrow$ Necessitarem ampliar el model a K **variables** (model de regressió múltiple)

El model de regressió múltiple

Ara volem tenir K **variables** explicatives en lloc de una:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{K-1} x_{iK-1} + u_i \qquad u_i \sim \mathcal{N}(0, \sigma^2)$$

Per tant, donades les nostres N **observacions**, tindrem el següent sistema d'equacions:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_{K-1} x_{1K-1} + u_1 & u_1 \sim \mathcal{N}(0, \sigma^2) \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_{K-1} x_{2K-1} + u_2 & u_2 \sim \mathcal{N}(0, \sigma^2) \\ \vdots & \\ y_N = \beta_0 + \beta_1 x_{N1} + \beta_2 x_{N2} + \cdots + \beta_{K-1} x_{NK-1} + u_N & u_N \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

Interpretació?

Això també ho podem transformar a **notació matricial** (piss.)

El model de regressió múltiple: notació matricial

El **model en notació matricial** quedaria expressat de la següent manera:

$$y = X\beta + u \qquad U \sim \mathcal{N}(0, \sigma^2 I_N)$$

on:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1K-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2K-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{NK-1} \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{K-1} \end{bmatrix} u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

Dimensions?

Supòsits

Novament treballarem amb x **fixes**. Els **supòsits** seran els mateixos que en el tema anterior, i la seva interpretació també.

Podem veure'ls implícitament a l'expressió $u \sim \mathcal{N}(0, \sigma^2 I_N)$:

- $\mathbb{E}[u] = \begin{bmatrix} \mathbb{E}[u_1] \\ \mathbb{E}[u_2] \\ \vdots \\ \mathbb{E}[u_N] \end{bmatrix} =$
- $\mathrm{Var}(u) = \begin{bmatrix} \mathrm{Var}(u_1) & \mathrm{Cov}(u_1,u_2) & \dots & \mathrm{Cov}(u_1,u_N) \\ \mathrm{Cov}(u_1,u_2) & \mathrm{Var}(u_2) & \dots & \mathrm{Cov}(u_2,u_N) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(u_1,u_2) & \mathrm{Cov}(u_2,u_N) & \dots & \mathrm{Var}(u_N) \end{bmatrix} =$
- $u \sim$

Continguts

T3. El model de regressió múltiple: Estimació

Estimació amb K variables

Com ja sabem, l'estimador MQO minimitza la suma dels quadrats dels residus:

$$\min_{\hat{\beta}} \hat{u}'\hat{u}$$

Com hem vist en el tema 2:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Dimensions?

Els residus

Els **residus** són l'equivalent mostral al terme de pertorbació:

$$\hat{u} = y - \hat{y} = y - X\hat{\beta} \quad \Leftrightarrow \quad \hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_{K-1} x_{iK-1}$$

Recordeu també la **propietat numèrica** que vam derivar en el tema 2:

$$X'\hat{u} = 0$$

Quines propietats inclou ara?

El coeficient de determinació i el coeficient de determinació ajustat

El **coeficient de determinació** ens indicava quina proporció de la variació de y ve explicada pel model:

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT} = 1 - \frac{\hat{u}'\hat{u}}{(y - \bar{y}\iota)'(y - \bar{y}\iota)} = 1 - \frac{\widehat{\mathrm{Var}}(\hat{u})}{\widehat{\mathrm{Var}}(y)}$$

Problema: sempre augmenta quan augmentem el nombre de variables.

Solució: coeficient de determinació ajustat (o corregit):

$$\bar{R}^2 = 1 - \frac{SQR/(N-K)}{SQT/(N-1)} = 1 - \frac{SQR}{SQT}\frac{N-1}{N-K} \leq R^2$$

Continguts

T3. El model de regressió múltiple: Estimació

L'esperança i la variança de l'estimador

Recordem que $\hat{\beta}$ és una **variable aleatòria**.

Igual com en el tema anterior, podem escriure:

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \;\Rightarrow \hat{\beta} \text{ és un } \textbf{estimador lineal} \text{ de } \beta.$$

$$\mathbb{E}[\hat{\beta}] = \begin{bmatrix} \mathbb{E}[\hat{\beta}_0] \\ \mathbb{E}[\hat{\beta}_1] \\ \vdots \\ \mathbb{E}[\hat{\beta}_{K-1}] \end{bmatrix} = \beta \;\Rightarrow \hat{\beta} \text{ és un estimador } \textbf{no esbiaixat} \text{ de } \beta.$$

$$\mathrm{Var}(\hat{\beta}) = \begin{bmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \dots & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_{K-1}) \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_1) & \dots & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_{K-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_{K-1}) & \mathrm{Var}(\hat{\beta}_{K-1}, \hat{\beta}_1) & \dots & \mathrm{Var}(\hat{\beta}_{K-1}) \end{bmatrix} = \sigma^2 (X'X)^{-1}.$$

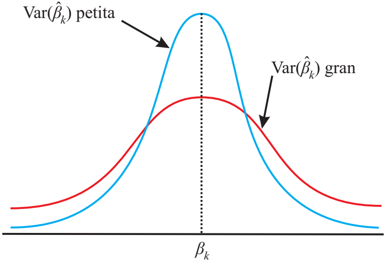
(demostracions pissarra)

Distribució de l'estimador MQO

Per tant, igual que en el tema anterior, $\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 (X'X)^{-1}\right)$.

Dimensions?

Per tant, $\hat{\beta}_{K-1} \sim \mathcal{N}\left(\beta_{K-1}, \sigma^2 (X'X)^{-1}_{kk}\right)$:



$\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ no és conegut. Per tant, ho estimarem:

$$\widehat{\text{Var}}(\hat{\beta}) =$$

El teorema de Gauss-Markov

Teorema de Gauss-Markov: "Sota els supòsits clàssics, MQO és l'estimador lineal no esbiaixat amb menor variança."

Aquest teorema:

- està comparant $\hat{\beta}$ amb els altres **estimadors lineals** (és a dir, els que es podem escriure com Ay o, el que és el mateix, com $Bu + c$).

- que són **no esbiaixats** (per exemple, un estimador que fos $\tilde{\beta} = 0$ tindria menor variança, però, en general, seria esbiaixat).

Continguts

T3. El model de regressió múltiple: Estimació

Motivació

Si tenim informació rellevant dels paràmetres és eficient usar-la en l'estimació. En aquest tema parlarem de **restriccions lineals d'igualtat**.

Per exemple, imaginem que volem estimar la funció de producció **Cobb-Douglas amb dades agregades**:

$$Y = AK^{\alpha}L^{1-\alpha}.$$

Si prenem logaritmes, el model serà:

$$\ln Y_i = \ln A + \alpha \ln K_i + (1-\alpha)\ln L_i.$$

Per tant, el nostre **model economètric** seria:

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + \beta_2 \ln L_i + u_i \quad s.a. \quad \beta_2 = 1 - \beta_1.$$

Les restriccions

Com hem vist abans, treballarem amb **restriccions lineals d'igualtat**.

algebra matricial?

Model 1: $y_i = \beta_0 + \beta_1 x_{i1} + u_i$
restricció: $\beta_0 = 2$ $\rightarrow$

Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 + \beta_2 = 1$ $\rightarrow$

Model 3: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 = 2\beta_2$ $\rightarrow$

Model 4: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + u_i$
restricció: $\beta_1 = 0 \quad \beta_2 + \beta_4 = 1$ $\rightarrow$

Recapitulació

Fixem-nos que qualsevol d'aquestes **restriccions** les podem escriure en **forma matricial** amb la següent estructura:

$$R\beta = r$$

Dimensions?

A continuació veurem dues formes d'**estimar** el model restringit:

- El **mètode de substitució**.
- L'estimador **mínim quadrat restringit** (MQR).

Mètode de substitució

El **mètode de substitució** consisteix en rescriure el model integrant la restricció i després estimar el "**model restringit**" per MQO:

model restringit?

Model 1: $y_i = \beta_0 + \beta_1 x_{i1} + u_i$
restricció: $\beta_0 = 2$ $\rightarrow$

Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 + \beta_2 = 1$ $\rightarrow$

Model 3: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 = 2\beta_2$ $\rightarrow$

Model 4: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + u_i$
restricció: $\beta_1 = 0 \quad \beta_2 + \beta_4 = 1$ $\rightarrow$

Mètode de substitució

El **mètode de substitució** consisteix en rescriure el model integrant la restricció i després estimar el "**model restringit**" per MQO:

model restringit?

Model 1: $y_i = \beta_0 + \beta_1 x_{i1} + u_i$
restricció: $\beta_0 = 2$ $\rightarrow$ $(y_i - 2) = \beta_1 x_{i1} + u_i$

Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 + \beta_2 = 1$ $\rightarrow$

Model 3: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 = 2\beta_2$ $\rightarrow$

Model 4: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + u_i$
restricció: $\beta_1 = 0 \quad \beta_2 + \beta_4 = 1$ $\rightarrow$

Mètode de substitució

El **mètode de substitució** consisteix en rescriure el model integrant la restricció i després estimar el "**model restringit**" per MQO:

model restringit?

Model 1: $y_i = \beta_0 + \beta_1 x_{i1} + u_i$
restricció: $\beta_0 = 2$ $\rightarrow$ $(y_i - 2) = \beta_1 x_{i1} + u_i$

Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 + \beta_2 = 1$ $\rightarrow$ $(y_i - x_{i2}) = \beta_0 + \beta_1 (x_{i1} - x_{i2}) + u_i$

Model 3: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$
restricció: $\beta_1 = 2\beta_2$ $\rightarrow$

Model 4: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + u_i$
restricció: $\beta_1 = 0 \quad \beta_2 + \beta_4 = 1$ $\rightarrow$

Mètode de substitució

El **mètode de substitució** consisteix en rescriure el model integrant la restricció i després estimar el "**model restringit**" per MQO:

		model restringit?
Model 1: $y_i = \beta_0 + \beta_1 x_{i1} + u_i$		
restricció: $\beta_0 = 2$	$\rightarrow$	$(y_i - 2) = \beta_1 x_{i1} + u_i$
Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$		$(y_i - x_{i2}) = \beta_0 +$
restricció: $\beta_1 + \beta_2 = 1$	$\rightarrow$	$\beta_1 (x_{i1} - x_{i2}) + u_i$
Model 3: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$		$y_i = \beta_0 +$
restricció: $\beta_1 = 2\beta_2$	$\rightarrow$	$+ \beta_1 (x_{i1} + 2x_{i2}) + u_i$
Model 4: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + u_i$		
restricció: $\beta_1 = 0 \quad \beta_2 + \beta_4 = 1$	$\rightarrow$	

Mètode de substitució

El **mètode de substitució** consisteix en rescriure el model integrant la restricció i després estimar el "**model restringit**" per MQO:

		model restringit?
Model 1: $y_i = \beta_0 + \beta_1 x_{i1} + u_i$ restricció: $\beta_0 = 2$	$\rightarrow$	$(y_i - 2) = \beta_1 x_{i1} + u_i$
Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ restricció: $\beta_1 + \beta_2 = 1$	$\rightarrow$	$(y_i - x_{i2}) = \beta_0 +$ $\beta_1 (x_{i1} - x_{i2}) + u_i$
Model 3: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ restricció: $\beta_1 = 2\beta_2$	$\rightarrow$	$y_i = \beta_0 +$ $+ \beta_1 (x_{i1} + 2x_{i2}) + u_i$
Model 4: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + u_i$ restricció: $\beta_1 = 0 \quad \beta_2 + \beta_4 = 1$	$\rightarrow$	$(y_i - x_{i4}) = \beta_0 +$ $\beta_2 (x_{i2} - x_{i4}) + u_i$

L'estimador mínim quadrat restringit (MQR)

L'estimador MQR és el que minimitza la suma de quadrats dels residus **subjecte a la restricció**:

$$\min_{\hat{\beta}} \hat{u}'\hat{u} = \min_{\hat{\beta}} (y - X\hat{\beta})'(y - X\hat{\beta})$$

$$s.a.$$

$$R\hat{\beta} = r.$$

I el resultat és:

$$\hat{\beta}_R = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(r - R\hat{\beta}\right).$$

Propietats de l'estimador MQR

Si les restriccions són certes ($R\beta = r$):

- $\mathbb{E}[\hat{\beta}_R] = \beta$.
- $\mathrm{Var}(\hat{\beta}_R) = \sigma^2 \left((X'X)^{-1} - (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1} \right).$

Continguts

T3. El model de regressió múltiple: Estimació

Introducció

A continuació ampliarem les nostres tres aplicacions de sempre per **il·lustrar** el que hem après en aquest tema.

Veurem com interpretar els resultats amb K **variables**.

A més, utilitzarem l'exemple de la funció de producció **Cobb-Douglas** per il·lustrar l'estimació amb **restriccions**.

I ampliarem l'exemple dels **salaris** per veure que passa si introduïm un **polinomi** de variables explicatives.

Noves variables per les nostres aplicacions

Model de **demanda de llet** (en milers d'ampolles):

- **Preu** unitari (en euros).
- Despesa en **publicitat** (en milers d'euros).
- Nombre d'**establiments** on es comercialitza (en milers).
- Quantitat de **calci** de la llet (en mg./100ml.).

Funció de producció **Cobb-Douglas** (versió agregada):

- **Capital** (en milers).
- **Treball** (en milers).

Salaris:

- Anys d'**educació**.
- Anys d'**experiència**.
- Anys d'**experiència al quadrat**.

Demanda de llet: el model complet

Estem interessats en introduir més informació en el nostre model de demanda de llet. Les variables que considerarem són:

- **Publicitat**: fa augmentar la demanda de llet. La mesurarem com despesa en publicitat (milers d'euros).

- **Establiments**: a més establiments, més proximitat del producte, i menor cost d'adquirir-lo per part dels consumidors, per tant, augmenta la demanda.

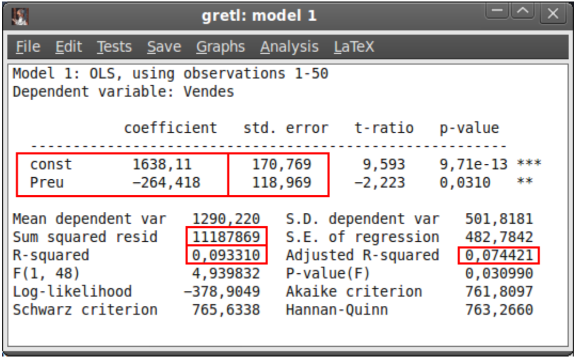
- **Calci**: la qualitat de la llet depèn del calci. Per tant, també esperem que a més calci, major demanda de llet.

El **model complet** quedaria així:

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 Pub_i + \beta_4 Estab_i + \beta_5 Calci_i + u_i$$

Demanda de llet: resultats (I)

Així era el model on només incloíem el preu:



Model 1: OLS, using observations 1-50
Dependent variable: Vendes

	coefficient	std. error	t-ratio	p-value	
const	1638,11	170,769	9,593	9,71e-13	***
Preu	−264,418	118,969	−2,223	0,0310	**

Mean dependent var	1290,220	S.D. dependent var	501,8181
Sum squared resid	11187869	S.E. of regression	482,7842
R-squared	0,093310	Adjusted R-squared	0,074421
F(1, 48)	4,939832	P-value(F)	0,030990
Log-likelihood	−378,9049	Akaike criterion	761,8097
Schwarz criterion	765,6338	Hannan-Quinn	763,2660

Interpretació?

Demanda de llet: resultats (II)

Anem a introduir les noves variables poc a poc. Primer **publicitat**:

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-50
Dependent variable: Vendes

               coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const          1615,28      144,319      11,19    7,46e-15 ***
  Preu           -364,626     102,914      -3,543   0,0009   ***
  Publicitat       52,2980     11,6106      4,504   4,39e-05 ***

Mean dependent var   1290,220   S.D. dependent var   501,8181
Sum squared resid     7814478   S.E. of regression   407,7567
R-squared            0,366697   Adjusted R-squared   0,339748
F(2, 47)             13,60706   P-value(F)           0,000022
Log-likelihood      -369,9336   Akaike criterion     745,8671
Schwarz criterion    751,6032   Hannan-Quinn         748,0515
```

Interpretació?

Què ha passat amb el coeficient del preu? Per què?

Com han canviat l'R^2 i l'$\bar{R}^2$?

Demanda de llet: resultats (III)

Després, **establiments**:

```
gretl: model 3
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 3: OLS, using observations 1-50
Dependent variable: Vendes

                 coefficient   std. error   t-ratio    p-value
  ------------------------------------------------------------
  const           1179,62      116,377       10,14     2,64e-13 ***
  Preu            -419,538      71,4707      -5,870    4,53e-07 ***
  Publicitat        32,9252      8,45135      3,896    0,0003   ***
  Establiments      10,7046      1,47655      7,250    3,85e-09 ***

Mean dependent var   1290,220   S.D. dependent var   501,8181
Sum squared resid     3647213   S.E. of regression   281,5799
R-squared            0,704422   Adjusted R-squared   0,685145
F(3, 46)             36,54240   P-value(F)           3,11e-12
Log-likelihood      -350,8832   Akaike criterion     709,7664
Schwarz criterion    717,4145   Hannan-Quinn         712,6788
```

Interpretació?

Què ha passat amb els coeficients del preu i publicitat? Per què?

Com han canviat l'R^2 i l'$\bar{R}^2$?

Demanda de llet: resultats (IV)

Finalment, **calci**:

gretl: model 4

File Edit Tests Save Graphs Analysis LaTeX

Model 4: OLS, using observations 1-50
Dependent variable: Vendes

	coefficient	std. error	t-ratio	p-value	
const	1186,59	130,374	9,101	9,12e-12	***
Preu	−418,495	72,7365	−5,754	7,24e-07	***
Publicitat	32,9186	8,54344	3,853	0,0004	***
Establiments	10,7193	1,49730	7,159	5,92e-09	***
Calci	−0,131206	1,05820	−0,1240	0,9019	

Mean dependent var	1290,220	S.D. dependent var	501,8181
Sum squared resid	3645967	S.E. of regression	284,6427
R-squared	0,704523	Adjusted R-squared	0,678258
F(4, 45)	26,82400	P-value(F)	2,06e-11
Log-likelihood	−350,8747	Akaike criterion	711,7493
Schwarz criterion	721,3094	Hannan-Quinn	715,3899

Excluding the constant, p-value was highest for variable 4 (Calci)

Interpretació?

Què ha passat ara amb els coeficients anteriors? Per què?

Com han canviat ara l'R^2 i l'$\bar{R}^2$?

Funció de producció Cobb-Douglas amb dades agregades

Ara volem estimar la funció de producció Cobb-Douglas amb **dades agregades**. Per tant,

$$Y = AK^{\alpha}L^{1-\alpha} \Rightarrow \ln Y_i = \beta_0 + \beta_1 \ln K_i + \beta_2 \ln L_i + u_i,$$

on ara les variables són **output** (Y_i, en milions de dòlars), **capital** (K_i, en milions de dòlars) i **treball** (L_i, en milions de treballadors).

El model, però, ens imposa una restricció:

$$\beta_2 = 1 - \beta_1 \Leftrightarrow \beta_1 + \beta_2 = 1$$

Cobb-Douglas amb dades agregades: estimació no restringida

```
gretl: model 3
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 3: OLS, using observations 1-188
Dependent variable: lnGDP

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       6,72771      1,56163       4,308    2,67e-05 ***
  lnK         0,372275     0,124220      2,997    0,0031   ***
  lnL         0,666764     0,113192      5,891    1,78e-08 ***

Mean dependent var   16,72470   S.D. dependent var   1,714682
Sum squared resid    444,7859   S.E. of regression   1,550564
R-squared            0,191012   Adjusted R-squared   0,182266
F(2, 185)            21,84040   P-value(F)           3,05e-09
Log-likelihood      -347,7086   Akaike criterion     701,4173
Schwarz criterion    711,1266   Hannan-Quinn         705,3512
```

Interpretació?

Quin és el valor de $\hat{\beta}_1 + \hat{\beta}_2$?

Cobb-Douglas amb dades agregades: estimació per MQR

Gretl permet l'estimació del model amb restriccions per MQR.

No ho veurem perquè sempre podem estimar el model pel mètode de substitució.

En el nostre cas, el resultat d'estimar el model per MQR és:

gretl: script output

```
Restriction:
 b[lnK] + b[lnL] = 1

Test statistic: F(1, 185) = 0,0539609, with p-value = 0,816566

Restricted estimates:

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       7,08690      0,218042     32,50     1,26e-78 ***
  lnK         0,350946     0,0834538     4,205    4,05e-05 ***
  lnL         0,649054     0,0834538     7,777    4,93e-13 ***

  Standard error of the regression = 1,54662
```

Com són els errors estàndard comparats amb els d'abans?

Cobb-Douglas amb dades agregades: mètode de substitució (I)

Quin seria el **model restringit**?

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + (1 - \beta_1) \ln L_i + u_i$$

Cobb-Douglas amb dades agregades: mètode de substitució (I)

Quin seria el **model restringit**?

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + (1 - \beta_1) \ln L_i + u_i$$

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + \ln L_i - \beta_1 \ln L_i + u_i$$

Cobb-Douglas amb dades agregades: mètode de substitució (I)

Quin seria el **model restringit**?

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + (1 - \beta_1) \ln L_i + u_i$$

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + \ln L_i - \beta_1 \ln L_i + u_i$$

$$(\ln Y_i - \ln L_i) = \beta_0 + \beta_1 (\ln K_i - \ln L_i) + u_i$$

Cobb-Douglas amb dades agregades: mètode de substitució (I)

Quin seria el **model restringit**?

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + (1-\beta_1) \ln L_i + u_i$$

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + \ln L_i - \beta_1 \ln L_i + u_i$$

$$(\ln Y_i - \ln L_i) = \beta_0 + \beta_1 (\ln K_i - \ln L_i) + u_i$$

$$\ln Y_i^* = \beta_0 + \beta_1 \ln K_i^* + u_i$$

Cobb-Douglas amb dades agregades: mètode de substitució (I)

Quin seria el **model restringit**?

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + (1 - \beta_1) \ln L_i + u_i$$

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + \ln L_i - \beta_1 \ln L_i + u_i$$

$$(\ln Y_i - \ln L_i) = \beta_0 + \beta_1 (\ln K_i - \ln L_i) + u_i$$

$$\ln Y_i^* = \beta_0 + \beta_1 \ln K_i^* + u_i$$

A què són iguals Y_i^* i K_i^*? Ens resulta familiar?

Aquest mètode ens donarà **exactament els mateixos resultats** que l'estimació per MQR.

Cobb-Douglas amb dades agregades: mètode de substitució (II)

Per estimar-lo:

- primer hem de **generar** les noves variables $\ln Y_i^*$ i $\ln K_i^*$.
- i després **estimar** la regressió com sempre hem fet.

```
gretl: model 1

File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 1: OLS, using observations 1-188
Dependent variable: lnGDPstar

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       7,08690      0,218042     32,50    1,26e-78 ***
  lnKstar     0,350946     0,0834538     4,205   4,05e-05 ***

Mean dependent var   6,302204   S.D. dependent var   1,614137
Sum squared resid    444,9157   S.E. of regression   1,546616
R-squared            0,086822   Adjusted R-squared   0,081913
F(1, 186)            17,68429   P-value(F)           0,000040
Log-likelihood      -347,7361   Akaike criterion     699,4721
Schwarz criterion    705,9450   Hannan-Quinn         702,0947
```

Com podem recuperar el valor de $\hat{\beta}_2$?

El model del salari ampliat: l'equació de Mincer

El **model de Mincer** ens diu que els salaris venen explicats per l'educació i l'experiència de la següent forma:

$$W_i = \omega_0 e^{\omega_1 Edu_i + \omega_2 Exp_i + \omega_3 Exp_i^2},$$

on W és el **salari**, Edu són els anys d'**educació** i Exp són els anys d'**experiència**.

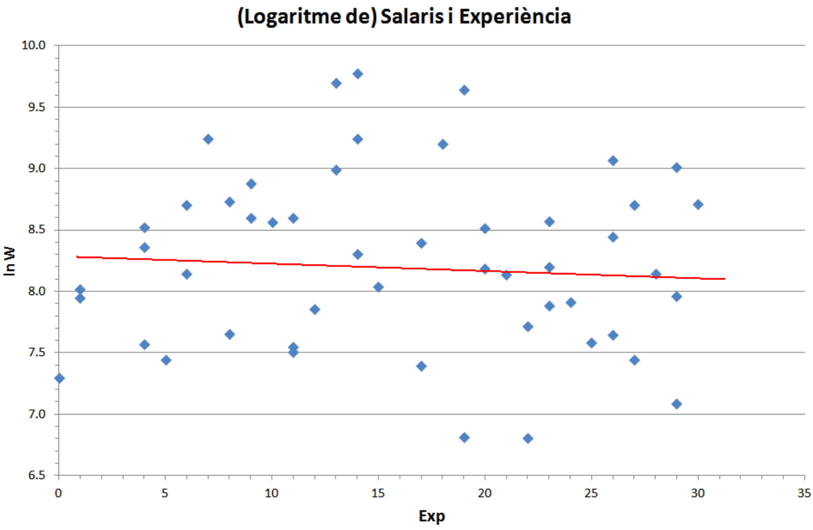
Per tant, el model que volem estimar serà:

$$\ln W_i = \beta_0 + \beta_1 Edu_i + \beta_2 Exp_i + \beta_4 Exp_i^2 + u_i$$

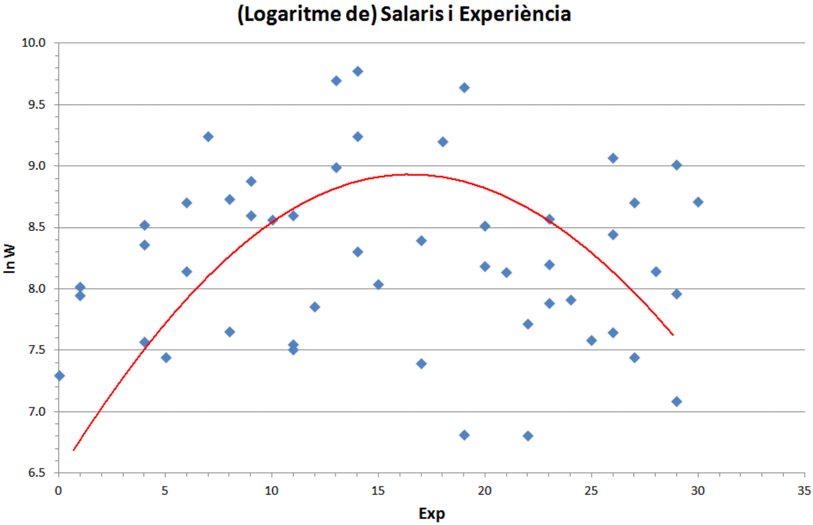
és un model de regressió lineal?

L'equació de Mincer: la relació entre salaris i experiència (I)

L'equació de Mincer: la relació entre salaris i experiència (II)



L'equació de Mincer: la relació entre salaris i experiència (III)



L'equació de Mincer: resultats (I)

Abans d'estimar el model, anem a veure més precisament quina és la relació entre experiència i el salari:

gretl: model 4

File Edit Tests Save Graphs Analysis LaTeX

```
Model 4: OLS, using observations 1-50
Dependent variable: LogSalari

                coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const          8,30157      0,214662     38,67     7,78e-38 ***
  Experiencia   −0,00332187   0,0117961   −0,2816    0,7795

Mean dependent var    8,248416   S.D. dependent var   0,716302
Sum squared resid     25,09990   S.E. of regression   0,723128
R-squared             0,001649   Adjusted R-squared  −0,019150
F(1, 48)              0,079303   P-value(F)           0,779455
Log-likelihood       −53,71795   Akaike criterion     111,4359
Schwarz criterion     115,2599   Hannan-Quinn         112,8921
```

Interpretació?

L'equació de Mincer: resultats (II)

Possible explicació d'aquest vincle tan dèbil, **dos efectes**:

- Més experiència $\Rightarrow$ més productivitat $\Rightarrow$ més salari
- Més experiència $\Rightarrow$ menys educació $\Rightarrow$ menys salari

gretl: model 5

File Edit Tests Save Graphs Analysis LaTeX

```
Model 5: OLS, using observations 1-50
Dependent variable: LogSalari

                 coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const          6,76018       0,156815     43,11     1,92e-39 ***
  Educacio       0,120703      0,00934623   12,91     4,52e-17 ***
  Experiencia    0,00641020    0,00563998    1,137    0,2615

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    5,518061   S.E. of regression   0,342645
R-squared            0,780519   Adjusted R-squared   0,771179
F(2, 47)            83,57063    P-value(F)           3,33e-16
Log-likelihood     -15,84701    Akaike criterion     37,69403
Schwarz criterion   43,43010    Hannan-Quinn         39,87836
```

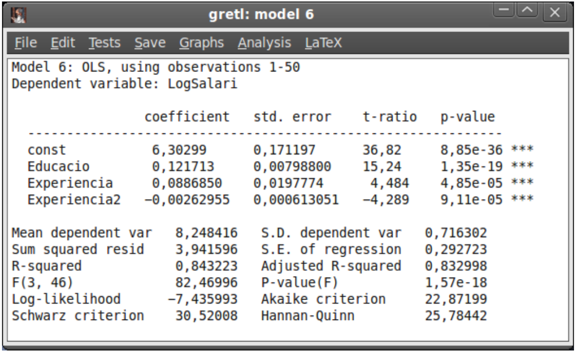
Interpretació?

Podem concloure que l'experiència té un efecte residual?

L'equació de Mincer: resultats (III)

Podem estimar l'equació de Mincer (completa):

$$\ln W_i = \beta_0 + \beta_1 Edu_i + \beta_2 Exp_i + \beta_4 Exp_i^2 + u_i$$



```
gretl: model 6
File  Edit  Tests  Save  Graphs  Analysis  LaTeX
Model 6: OLS, using observations 1-50
Dependent variable: LogSalari

                 coefficient   std. error    t-ratio   p-value
  ---------------------------------------------------------------
  const           6,30299      0,171197      36,82     8,85e-36 ***
  Educacio        0,121713     0,00798800    15,24     1,35e-19 ***
  Experiencia     0,0886850    0,0197774      4,484    4,85e-05 ***
  Experiencia2   -0,00262955   0,000613051   -4,289    9,11e-05 ***

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    3,941596   S.E. of regression   0,292723
R-squared            0,843223   Adjusted R-squared   0,832998
F(3, 46)             82,46996   P-value(F)           1,57e-18
Log-likelihood      -7,435993   Akaike criterion     22,87199
Schwarz criterion    30,52008   Hannan-Quinn         25,78442
```

Interpretació?

Continguts

T3. El model de regressió múltiple: Estimació

Motivació

No sempre volem considerar variables **quantitatives** (preu, capital, anys d'educació,...) en el nostre anàlisi.

A vegades ens pot interessar saber quin és l'efecte d'una variable **qualitativa** sobre la nostra variable depenent:

- Com afecta el **Tetra Brick** a les vendes de llet
- Com canvia l'output per capita en un país **democràtic** si el comparem amb un país autocràtic
- Quin és el diferencial salarial entre **homes** i **dones**

Variables fictícies

Per poder fer aquest anàlisi utilitzarem **variables fictícies**, d_i:

- Si la observació **compleix** la característica, $d_i = 1$:
  - L'empresa comercialitza la seva llet en Tetra Brick
  - El país gaudeix d'una democràcia
  - L'individu és una dona
- Si **no** la compleix, $d_i = 0$.

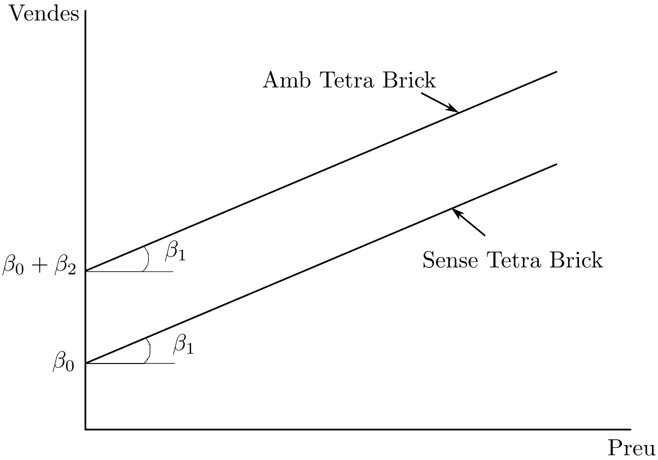
El Tetra-Brick i les vendes de llet (I)

Al model de demanda de llet simple (vendes-preu) li afegirem ara una variable TB_i:

$$Vendes_i = \beta_1 + \beta_2 Preu_i + \beta_3 TB_i + u_i$$

- Si l'empresa ven en Tetra Brick, $TB_i = 1$
- En cas contrari, $TB_i = 0$

Interpretació gràfica



El Tetra Brick i les vendes de llet (II)

En el model anterior estavem analitzant en quant augmenten les vendes (donat un preu).

Poster la llet en Tetra Brick és un "producte diferent" a la llet en ampolla, i per això tota **la demanda és diferent**:

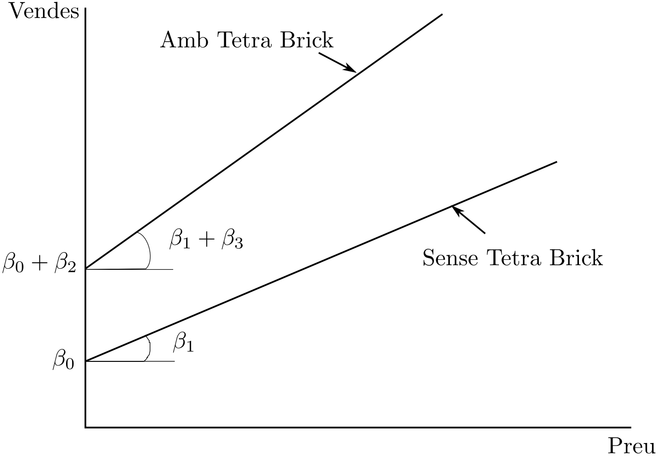
$$Vendes_i = \beta_1^{TB} + \beta_2^{TB} Preu_i + u_i \qquad \text{(TetraBrick)}$$

$$Vendes_i = \beta_1^{A} + \beta_2^{A} Preu_i + u_i \qquad \text{(Ampolla)}$$

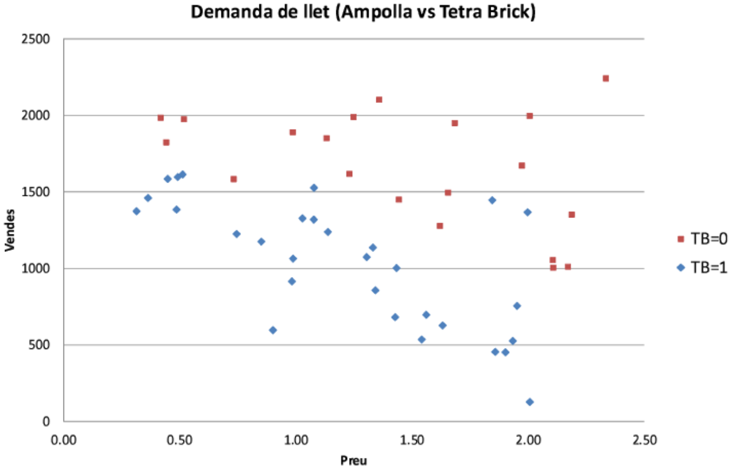
Això **es pot escriure** com:

$$Vendes_i = \beta_1 + \beta_2 Preu_i + \beta_3 TB_i + \beta_4 Preu_i TB_i + u_i$$

Interpretació gràfica



Aplicacions: Demanda de llet (I)



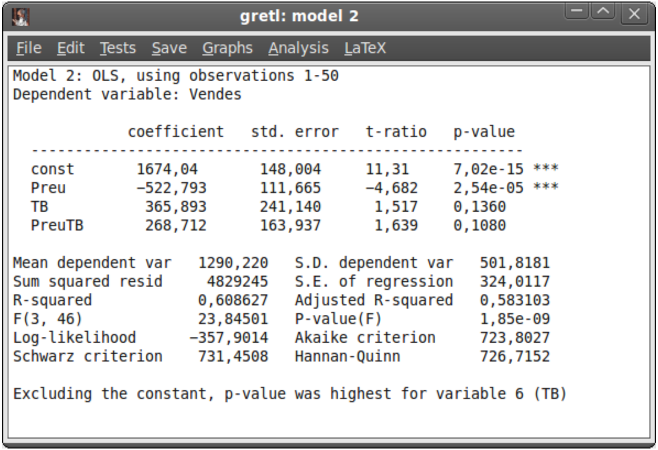
Aplicacions: Demanda de llet (II)

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-50
Dependent variable: Vendes

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       1522,57      117,666      12,94     4,20e-17 ***
  Preu        -398,122      83,2094     -4,785    1,74e-05 ***
  TB           728,638      97,4764      7,475    1,56e-09 ***

Mean dependent var   1290,220   S.D. dependent var   501,8181
Sum squared resid     5111303   S.E. of regression   329,7744
R-squared            0,585769   Adjusted R-squared   0,568142
F(2, 47)             33,23158   P-value(F)           1,01e-09
Log-likelihood      -359,3205   Akaike criterion     724,6410
Schwarz criterion    730,3770   Hannan-Quinn         726,8253
```

Aplicacions: Demanda de llet (III)

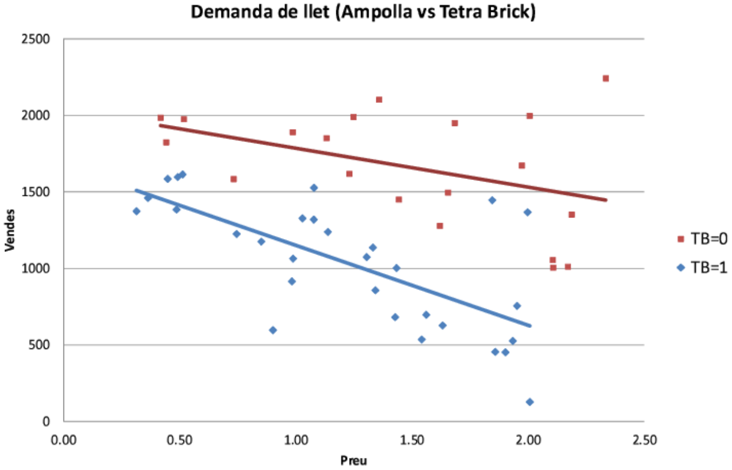


gretl: model 2

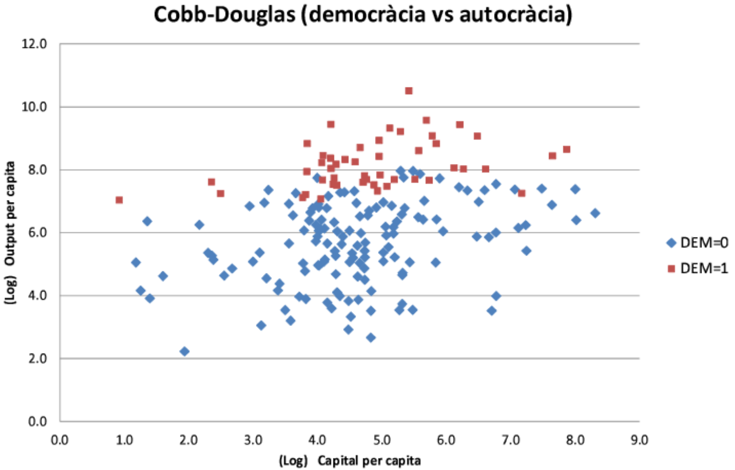
File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-50
Dependent variable: Vendes

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        1674,04       148,004     11,31    7,02e-15 ***
  Preu         −522,793      111,665     −4,682   2,54e-05 ***
  TB            365,893      241,140      1,517   0,1360
  PreuTB        268,712      163,937      1,639   0,1080

Mean dependent var   1290,220   S.D. dependent var   501,8181
Sum squared resid     4829245   S.E. of regression   324,0117
R-squared            0,608627   Adjusted R-squared   0,583103
F(3, 46)             23,84501   P-value(F)           1,85e-09
Log-likelihood      −357,9014   Akaike criterion     723,8027
Schwarz criterion    731,4508   Hannan-Quinn         726,7152

Excluding the constant, p-value was highest for variable 6 (TB)
```

Aplicacions: Demanda de llet (IV)



Aplicacions: Cobb-Douglas (I)



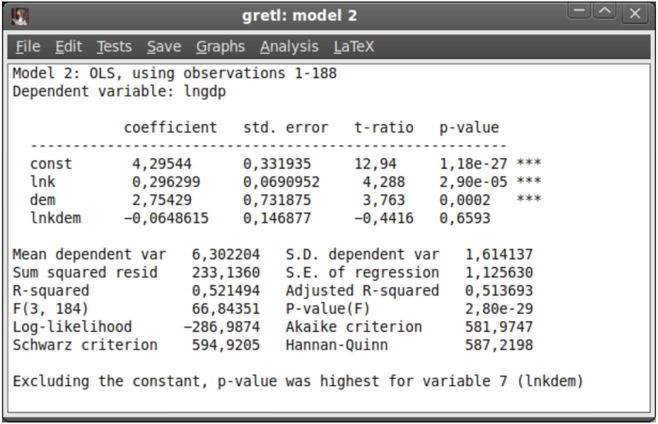
Aplicacions: Cobb-Douglas (II)

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-188
Dependent variable: lngdp

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       4,36151      0,295665     14,75     4,58e-33 ***
  lnk         0,281945     0,0608394     4,634    6,75e-06 ***
  dem         2,44205      0,188589     12,95     1,02e-27 ***

Mean dependent var   6,302204   S.D. dependent var   1,614137
Sum squared resid    233,3831   S.E. of regression   1,123179
R-squared            0,520987   Adjusted R-squared   0,515809
F(2, 185)            100,6055   P-value(F)           2,70e-30
Log-likelihood      -287,0869   Akaike criterion     580,1738
Schwarz criterion    589,8832   Hannan-Quinn         584,1077
```

Aplicacions: Cobb-Douglas (III)

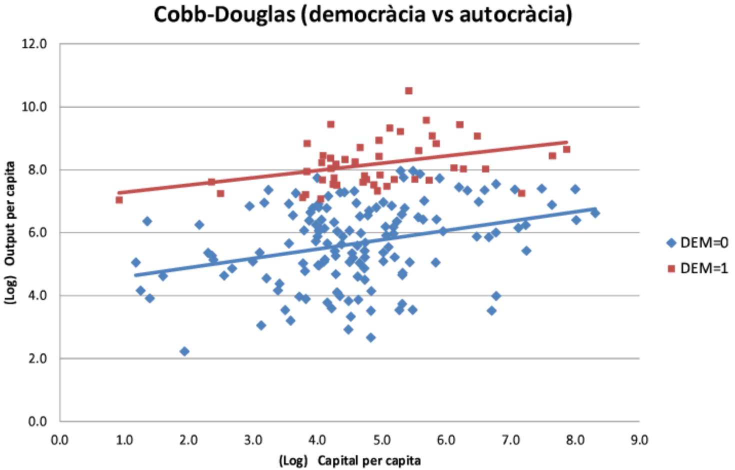


gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-188
Dependent variable: lngdp

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       4,29544      0,331935     12,94     1,18e-27 ***
  lnk         0,296299     0,0690952     4,288    2,90e-05 ***
  dem         2,75429      0,731875      3,763    0,0002   ***
  lnkdem     -0,0648615    0,146877     -0,4416   0,6593

Mean dependent var   6,302204   S.D. dependent var   1,614137
Sum squared resid    233,1360   S.E. of regression   1,125630
R-squared            0,521494   Adjusted R-squared   0,513693
F(3, 184)            66,84351   P-value(F)           2,80e-29
Log-likelihood      -286,9874   Akaike criterion     581,9747
Schwarz criterion    594,9205   Hannan-Quinn         587,2198

Excluding the constant, p-value was highest for variable 7 (lnkdem)
```

Aplicacions: Cobb-Douglas (IV)



Aplicacions: Salaris (I)

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-50
Dependent variable: LogSalari

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        8,67687      0,120289     72,13    1,33e-50 ***
  Dona        −0,823952     0,166811    −4,939    9,90e-06 ***

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    16,66873   S.E. of regression   0,589292
R-squared            0,337000   Adjusted R-squared   0,323187
F(1, 48)             24,39817   P-value(F)           9,90e-06
Log-likelihood      −43,48472   Akaike criterion     90,96944
Schwarz criterion    94,79349   Hannan-Quinn         92,42566
```

Aplicacions: Salaris (II)

gretl: model 4

File Edit Tests Save Graphs Analysis LaTeX

```
Model 4: OLS, using observations 1-50
Dependent variable: LogSalari

                 coefficient    std. error    t-ratio    p-value
  --------------------------------------------------------------
  const           6,50899      0,211170       30,82     6,73e-32 ***
  Educacio        0,112560     0,00968435     11,62     3,81e-15 ***
  Experiencia     0,0883958    0,0194414       4,547    4,09e-05 ***
  Experiencia2   -0,00266470   0,000603004    -4,419    6,18e-05 ***
  Dona           -0,162786     0,100808       -1,615    0,1133

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    3,725704   S.E. of regression   0,287738
R-squared            0,851810   Adjusted R-squared   0,838637
F(4, 45)             64,66598   P-value(F)           4,45e-18
Log-likelihood      -6,027744   Akaike criterion     22,05549
Schwarz criterion    31,61560   Hannan-Quinn         25,69603

Excluding the constant, p-value was highest for variable 5 (Dona)
```

Aplicacions: Salaris (III)

gretl: model 5

File Edit Tests Save Graphs Analysis LaTeX

```
Model 5: OLS, using observations 1-50
Dependent variable: LogSalari

                 coefficient   std. error    t-ratio   p-value
  ---------------------------------------------------------------
  const           6,47821      0,279139      23,21     2,58e-26 ***
  Educacio        0,114691     0,0158394      7,241    5,08e-09 ***
  Experiencia     0,0881946    0,0196897      4,479    5,27e-05 ***
  Experiencia2   -0,00265609   0,000611687   -4,342    8,16e-05 ***
  Dona           -0,120674     0,266301      -0,4531   0,6527
  EduDona        -0,00339592   0,0198398     -0,1712   0,8649

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    3,723224   S.E. of regression   0,290893
R-squared            0,851908   Adjusted R-squared   0,835080
F(5, 44)             50,62271   P-value(F)           3,79e-17
Log-likelihood      -6,011103   Akaike criterion     24,02221
Schwarz criterion    35,49434   Hannan-Quinn         28,39086

Excluding the constant, p-value was highest for variable 6 (EduDona)
```

Aplicacions: Salaris (IV)

gretl: model 7

File Edit Tests Save Graphs Analysis LaTeX

```
Model 7: OLS, using observations 1-50
Dependent variable: LogSalari

                 coefficient   std. error    t-ratio   p-value
  -----------------------------------------------------------
  const           6,44303      0,243116      26,50     1,10e-28 ***
  Educacio        0,112724     0,00976332    11,55     6,62e-15 ***
  Experiencia     0,0943952    0,0223194      4,229    0,0001   ***
  Experiencia2   -0,00276689   0,000634358   -4,362    7,67e-05 ***
  Dona           -0,0733810    0,188979      -0,3883   0,6997
  ExpDona        -0,00555626   0,00990335    -0,5610   0,5776

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    3,699239   S.E. of regression   0,289955
R-squared            0,852862   Adjusted R-squared   0,836142
F(5, 44)             51,00799   P-value(F)           3,29e-17
Log-likelihood      -5,849531   Akaike criterion     23,69906
Schwarz criterion    35,17120   Hannan-Quinn         28,06772

Excluding the constant, p-value was highest for variable 5 (Dona)
```

Aplicacions: Salaris (V)

gretl: model 8

File Edit Tests Save Graphs Analysis LaTeX

```
Model 8: OLS, using observations 1-50
Dependent variable: LogSalari

                  coefficient    std. error    t-ratio   p-value
  -----------------------------------------------------------------
  const            6,34726       0,355592      17,85     3,19e-21 ***
  Educacio         0,116481      0,0164187      7,094    1,07e-08 ***
  Experiencia      0,100385      0,0325941      3,080    0,0036   ***
  Experiencia2    -0,00293248    0,000950562   -3,085    0,0036   ***
  Dona             0,0658848     0,415271       0,1587   0,8747
  EduDona         -0,00603524    0,0206747     -0,2919   0,7718
  ExpDona         -0,0161229     0,0421078     -0,3829   0,7037
  Exp2Dona         0,000316211   0,00129975     0,2433   0,8090

Mean dependent var   8,248416   S.D. dependent var   0,716302
Sum squared resid    3,686622   S.E. of regression   0,296271
R-squared            0,853364   Adjusted R-squared   0,828925
F(7, 42)             34,91774   P-value(F)           1,58e-15
Log-likelihood      -5,764115   Akaike criterion     27,52823
Schwarz criterion    42,82441   Hannan-Quinn         33,35310

Excluding the constant, p-value was highest for variable 5 (Dona)
```