Tema 2. El model de regressió simple

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

Tema 2. El model de regressió simple

1. El model.
2. L'estimador mínim quadrat ordinari (MQO).
3. Bondat de l'ajust.
4. Distribució de l'estimador MQO.
5. Experiments de Monte Carlo.
6. Aplicacions.

Continguts

Tema 2. El model de regressió simple

El model de regressió simple

Recordeu que el **model de regressió** simple tenia la següent estructura:

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad u_i \sim i.i.\mathcal{N}(0, \sigma^2).$$

Si tenim una mostra de N individus, tindrem les següents **observacions**:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + u_1 & u_1 &\sim \mathcal{N}(0, \sigma^2), \\ y_2 &= \beta_0 + \beta_1 x_2 + u_2 & u_2 &\sim \mathcal{N}(0, \sigma^2), \\ &\vdots \\ y_N &= \beta_0 + \beta_1 x_N + u_N & u_N &\sim \mathcal{N}(0, \sigma^2). \end{aligned}$$

Això ho podem transformar a **notació matricial** (pissarra).

Model de regressió simple: 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:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}.$$

Supòsits

Els **supòsits** que farem sobre $u_1, u_2, ..., u_N$ (coneguts com a supòsits "**clàssics**" del model de regressió amb x **fixes**) són:

1. Totes les pertorbacions tenen **esperança** igual a **zero**:

$$\mathbb{E}[u_i] = 0 \quad \forall i.$$

2. Totes les pertorbacions tenen la mateixa variança (***homosquedasticitat***):

$$\text{Var}(u_i) = \sigma^2 \quad \forall i.$$

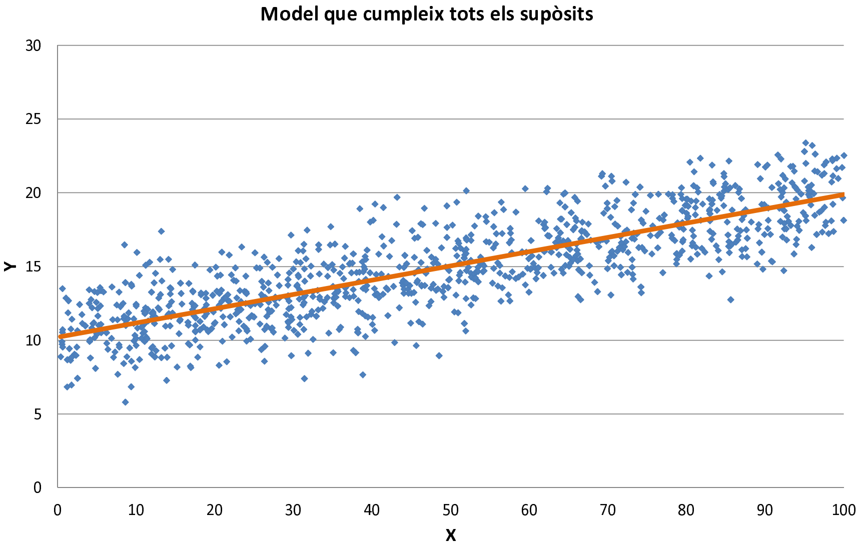
3. Les pertorbacions **no** estan **correlacionades** entre elles:

$$\text{Cov}(u_i, u_j) = 0 \quad \forall i \neq j.$$

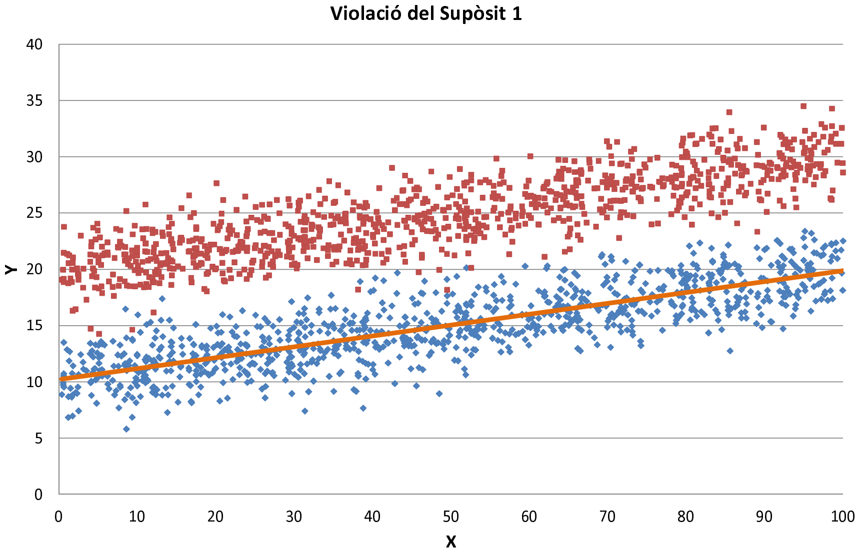
4. Totes les pertorbacions segueixen una distribució **normal**:

$$u_i \sim \mathcal{N}(\ ).$$

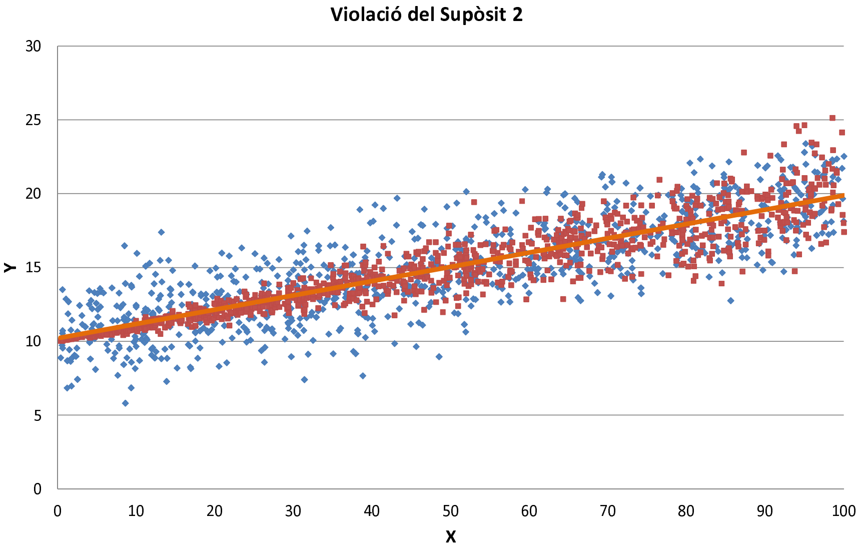
Tots els supòsits es compleixen



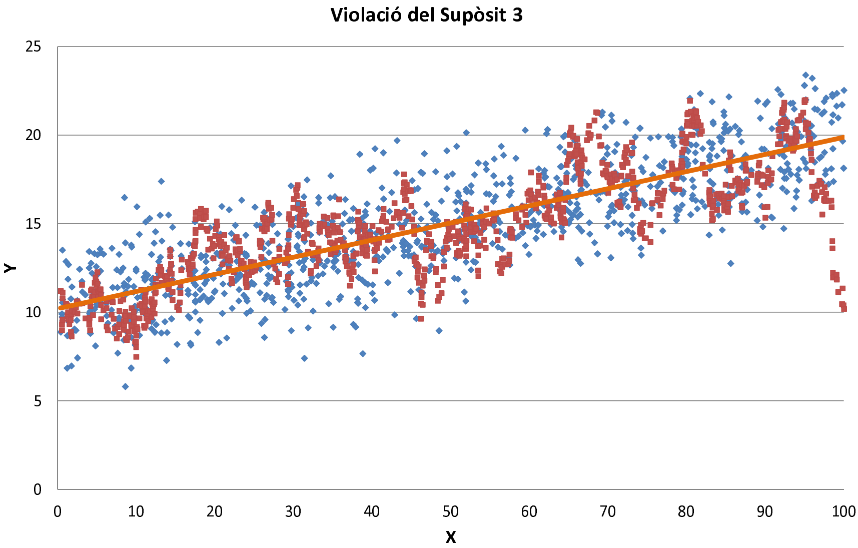
Desviacions (I): $\mathbb{E}[u_i] \neq 0$



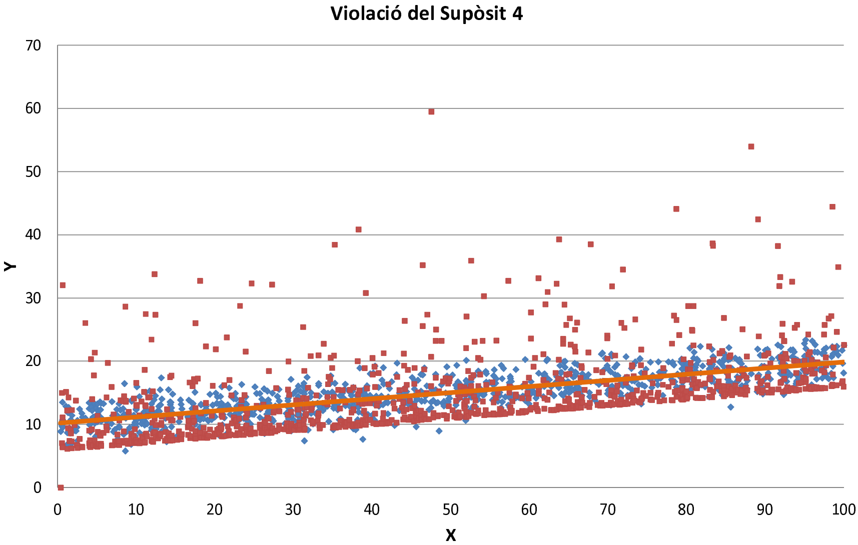
Desviacions (II): $\mathrm{Var}(u_i) = \sigma_i^2 \neq \sigma_j^2 = \mathrm{Var}(u_j)$



Desviacions (III): $\mathrm{Cov}(u_i, u_j) \neq 0$



Desviacions (IV): $u_i \not\sim \mathcal{N}(\ )$



Implicaciones sobre y_i *dels supòsits sobre* u_i

1. $\mathbb{E}[u_i] = 0 \quad \forall i \quad \rightarrow \quad \mathbb{E}[y_i] =$

2. $\mathrm{Var}(u_i) = \sigma^2 \quad \forall i \quad \rightarrow \quad \mathrm{Var}(y_i) =$

3. $\mathrm{Cov}(u_i, u_j) = 0 \quad \forall i \neq j \quad \rightarrow \quad \mathrm{Cov}(y_i, y_j) =$

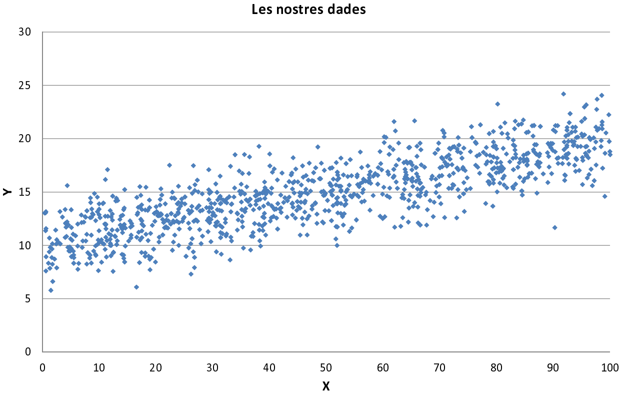
4. $u_i \sim \mathcal{N}(0, \sigma^2) \quad \forall i \quad \rightarrow \quad y_i \sim$

Continguts

Tema 2. El model de regressió simple

Objectiu del procés d'estimació

Quin procés pot haver generat les nostres dades?



No coneixem la recta de regressió poblacional!

Objectiu: aproximar β_0 i β_1 a partir de les nostres dades. És a dir, "estimar" β_0 i β_1 (recta de regressió mostral).

Escollir la recta de regressió mostral

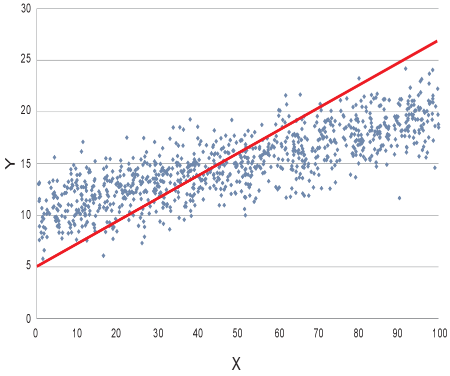
Com escollirem la recta de regressió mostral? Quina recta és el més probable mecanisme generador de les dades?

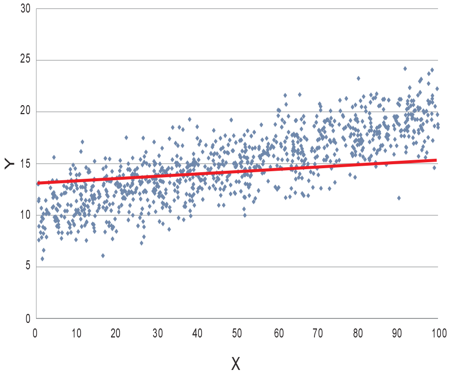
El més probable és que la nostra mostra contengui **moltes** observacions **a prop** de la RRP i **poques enfora**.

Cal triar aquella recta que "**millor s'ajusti**" a la mostra!

Solució: triar aquella recta que generi els residus més petits.

Dos exemples del que no volem



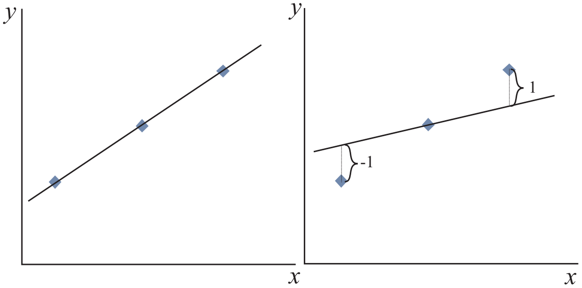


Els residus més petits (I)

Un podria pensar que la manera fer els residus el més petits possibles seria minimitzar la **suma dels residus**:

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{\mathrm{i=1}}^{\mathrm{N}} \hat{u}_i.$$

Problema: els valors positius i negatius es compensen!

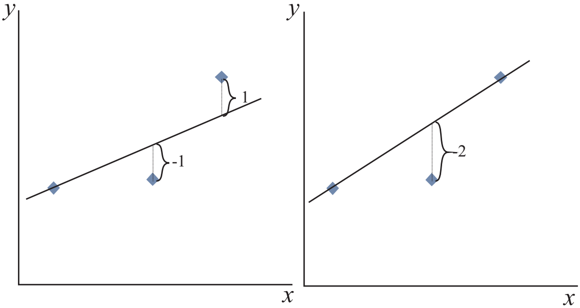


Els residus més petits (II)

Una solució al problema anterior podria ser minimitzar la suma dels **valors absoluts** dels residus:

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{\mathrm{i=1}}^{\mathrm{N}} |\hat{u}_i|.$$

Problema: aquest mètode no penalitza que les observacions estan molt allunyades!



Els residus més petits (III)

La millor alternativa serà, per tant, minimitzar la **suma dels quadrats dels errors**:

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{\mathrm{i=1}}^{\mathrm{N}} \hat{u}_i^2.$$

Aquest és l'estimador de **mínims quadrats ordinaris** (MQO).

L'estimador MQO:

- No compensa els valors **positius** amb **negatius**.
- Penalitza les (menys probables) observacions **allunyades**.

L'estimador MQO

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{\mathrm{i=1}}^{\mathrm{N}} \hat{u}_i^2 \quad \Leftrightarrow \quad \min_{\hat{\beta}} \hat{u}'\hat{u} \quad \Leftrightarrow \quad \min_{\hat{\beta}} (y - X\hat{\beta})'(y - X\hat{\beta}).$$

(pissarra)

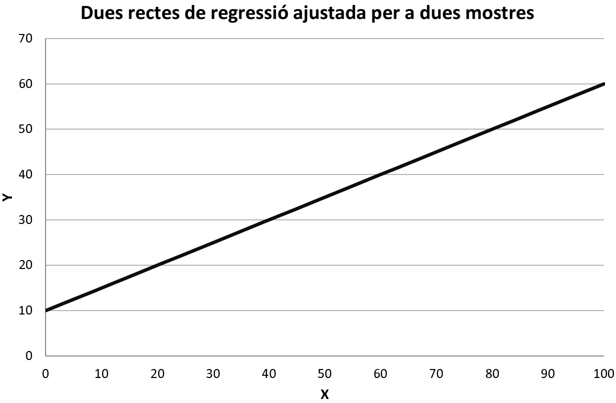
$$\hat{\beta} = (X'X)^{-1}X'Y.$$

(pissarra)

$$\hat{\beta}_1 = \frac{\sum_{\mathrm{i=1}}^{\mathrm{N}}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{\mathrm{i=1}}^{\mathrm{N}}(x_i - \bar{x})^2}; \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

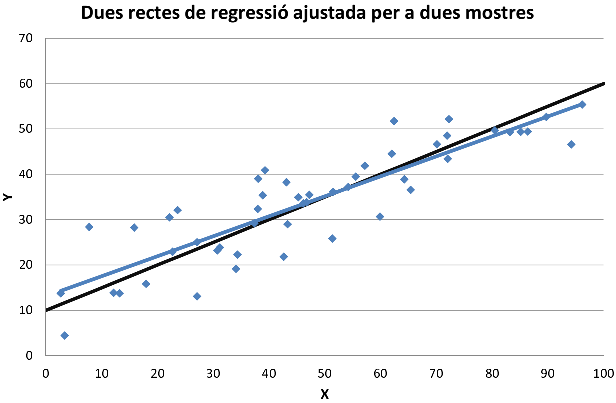
La recta de regressió ajustada

La **recta de regressió ajustada** ve donada per: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, on $\hat{y}_i$ és el **valor ajustat**:



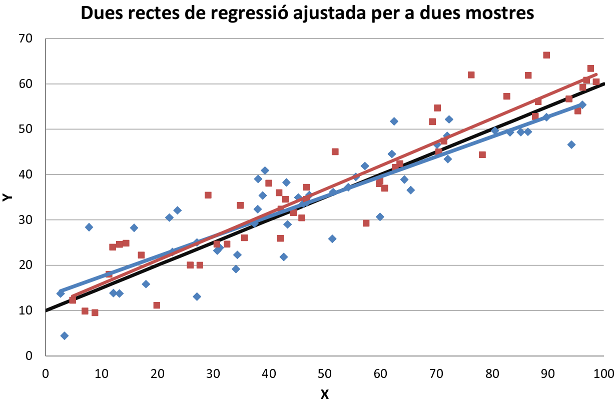
La recta de regressió ajustada

La **recta de regressió ajustada** ve donada per: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, on $\hat{y}_i$ és el **valor ajustat**:



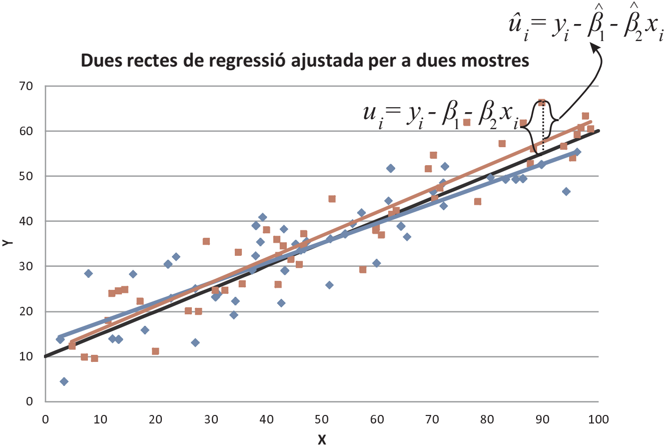
La recta de regressió ajustada

La **recta de regressió ajustada** ve donada per: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, on $\hat{y}_i$ és el **valor ajustat**:



La recta de regressió ajustada

La **recta de regressió ajustada** ve donada per: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, on $\hat{y}_i$ és el **valor ajustat**:



El **residu** és l'estimació de la pertorbació: $\hat{u}_i = y_i - \hat{y}_i$.

Propietats numèriques de l'estimador MQO

Hi ha dues **propietats numèriques** de l'estimador MQO que són força interessants:

$$\sum_{\mathrm{i=1}}^{\mathrm{N}} \hat{u}_i = 0.$$

$$\sum_{\mathrm{i=1}}^{\mathrm{N}} x_i \hat{u}_i = 0.$$

(demostracions a la pissarra)

Continguts

Tema 2. El model de regressió simple

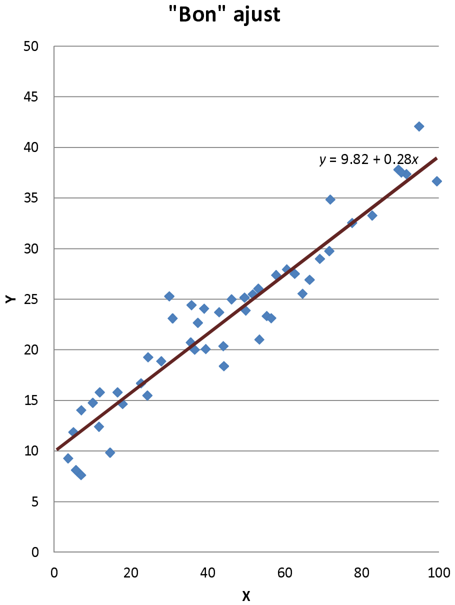
Què és la "Bondat de l'ajust"?

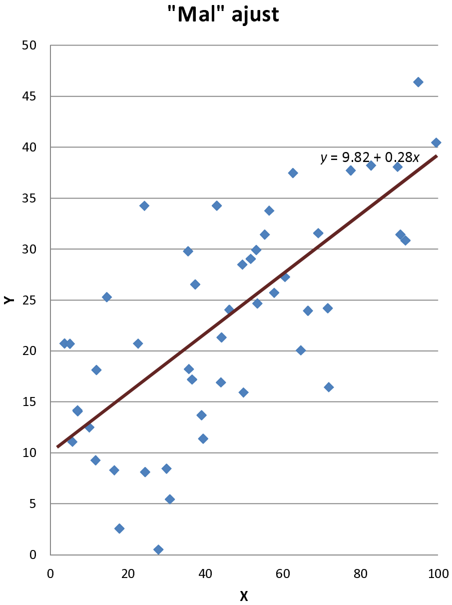
Una vegada tenim un estimador (MQO) pels nostres paràmetres ens podem demanar:

- És el nostre model una **bona representació** de les dades?
- Quina part de y ve **explicada pel model** i quina part per la **pertorbació**?
- Ens hem deixat moltes **variables importants** sense incloure?

La **bondat de l'ajust** és la resposta a aquestes preguntes: ens indica si el model és una bona aproximació de les nostres dades.

Bondat de l'ajust vista gràficament





Sumes de quadrats

Per poder establir una mesura de bondat de l'ajust hem de definir primer els següents conceptes:

- **Suma dels quadrats totals**:
$$SQT = (y - \bar{y}\iota)'(y - \bar{y}\iota) = \sum_{\mathrm{i=1}}^{\mathrm{N}} (y_i - \bar{y})^2 = N\widehat{\mathrm{Var}}(y_i).$$

- **Suma dels quadrats explicada**:
$$SQE = (\hat{y} - \bar{y}\iota)'(\hat{y} - \bar{y}\iota) = \sum_{\mathrm{i=1}}^{\mathrm{N}} (\hat{y}_i - \bar{y})^2 = N\widehat{\mathrm{Var}}(\hat{y}_i).$$

- **Suma dels quadrats dels residus**:
$$SQR = \hat{u}'\hat{u} = \sum_{\mathrm{i=1}}^{\mathrm{N}} \hat{u}_i^2 = N\widehat{\mathrm{Var}}(\hat{u}_i).$$

Es pot demostrar que $SQT = SQE + SQR$. (pissarra)

El coeficient de determinació (R^2)

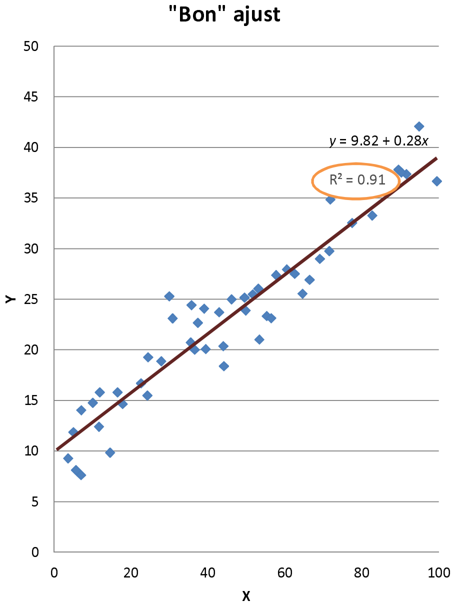
La nostra mesura ens ha de dir quina **proporció** de la **variació de** y està **explicada** pel nostre model de regressió.

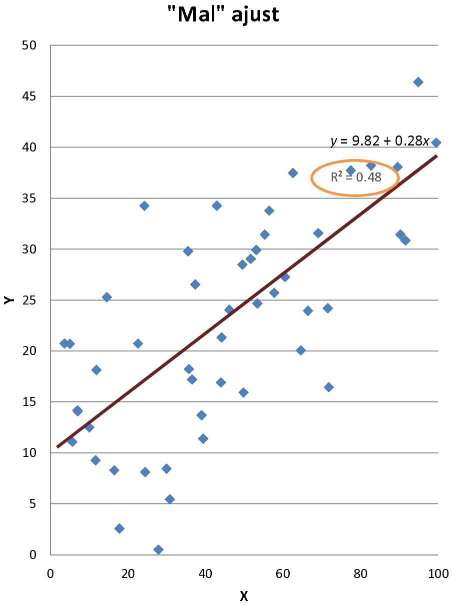
Aquesta és, de fet, la definició del **coeficient de determinació**:

$$R^2 = \frac{SQE}{SQT} = \frac{SQT - SQR}{SQT} = 1 - \frac{SQR}{SQT}.$$

Noteu que aquesta mesura estarà sempre **entre 0 i 1**. Quan més proper a 1, millor serà l'ajust.

El coeficient de determinació gràficament





Continguts

Tema 2. El model de regressió simple

1. El model.
2. L'estimador mínim quadrat ordinari (MQO).
3. Bondat de l'ajust.
4. Distribució de l'estimador MQO.
5. Experiments de Monte Carlo.
6. Aplicacions.

L'estimador MQO com a variable aleatòria

Fins i tot amb x fixes, l'estimador és una **variable aleatòria**: Y és una **variable aleatòria** (perquè u és una v.a.).

Obtindrem un **estimador diferent** per cada **mostra diferent**.

Funció de densitat mostral: distribució de l'estimador en mostres repetides de la mateixa mida i amb les mateixes x (x fixes).

L'esperança i la variança de l'estimador

Partirem de la següent **transformació** de l'estimador:

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

(demostració pissarra)

A partir d'aquí és fàcil demostrar que: (pissarra)

- $\mathbb{E}[\hat{\beta}] = \mathbb{E}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \beta.$

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

Exercici: trobar les expressions per $\mathrm{Var}(\hat{\beta}_0)$, $\mathrm{Var}(\hat{\beta}_1)$ i $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$.

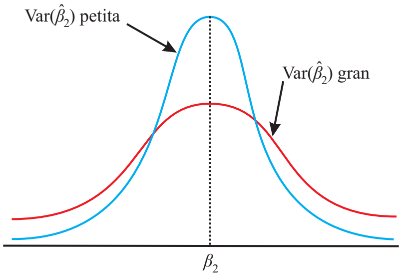
Distribució de l'estimador

Recordeu la **propietat** de la distribució **normal** que diu que si $Z \sim \mathcal{N}(\mu, \Sigma)$, aleshores $W = AZ + b \sim \mathcal{N}(A\mu + b, A\Sigma A')$.

Per tant, ja sabem quina serà la **distribució de** $\hat{\beta}$:

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 (X'X)^{-1}\right).$$

Gràficament, per cada paràmetre (p.ex. $\hat{\beta}_1$):



Estimador no esbiaixat i eficient

Com hem demostrat, l'estimador MQO sempre satisfà $\mathbb{E}[\hat{\beta}] = \beta$. Quan un estimador compleix aquesta propietat es diu que és un estimador **no esbiaixat**.

Es pot demostrar també que l'estimador MQO és el que té menys variança d'entre tots aquells estimadors lineals[1] no esbiaixats (Teorema Gauss-Markov).

Quan un estimador compleix aquesta propietat es diu que és un estimador **eficient**.

[1] Es diu que un estimador és un estimador lineal quan es pot escriure com una combinació lineal de les pertorbacions, és a dir, que té aquesta forma: $Au + b$

Estimació de la variança de l'estimador

La variança de l'estimador $\hat{\beta}$ **no és coneguda**. Per estimar-la necessitarem un **estimador de** σ^2.

Recordem que, per a tot i:

$$\mathrm{Var}(u_i) = \sigma^2 = \mathbb{E}[(u_i - \mathbb{E}[u_i])^2] = \mathbb{E}[u_i^2].$$

Per tant, un **estimador ideal** seria $\bar{\sigma}^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} u_i^2}{N}$.

No coneixem u, però podríem utilitzar $\hat{u}$. No obstant això, $\tilde{\sigma}^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \hat{u}_i^2}{N}$ és un estimador **esbiaixat** de σ^2: $\mathbb{E}[\tilde{\sigma}^2] \neq \sigma^2$.

L'estimador alternatiu **no esbiaixat** és:

$$\hat{\sigma}^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \hat{u}_i^2}{N-2} = \frac{\hat{u}'\hat{u}}{N-2} = \frac{SQR}{N-2}.$$

Continguts

Tema 2. El model de regressió simple

Motivació

En l'apartat anterior hem vist que $\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2(X'X)^{-1}\right)$.

Aquest resultat es basa en una **derivació "teòrica"**, a partir de certes propietats estadístiques i matemàtiques.

Com a alternativa, **aproximarem** la distribució de l'estimador a partir de la **simulació** (experiments de Monte Carlo).

Objectiu

La simulació serveix per entendre **quin paper** juga cada element del model (Tema 1).

En l'econometria, els experiments de Monte Carlo són una eina útil perquè **a vegades no podem derivar la distribució d'un estimador** de forma "teòrica".

Idea bàsica: si els vertaders paràmetres són $\beta = b$ i $\sigma^2 = s$, i tenim mostres de $N = n$ observacions, quina seria la distribució de probabilitat del nostre estimador MQO.

Experiments de Monte Carlo amb x fixes

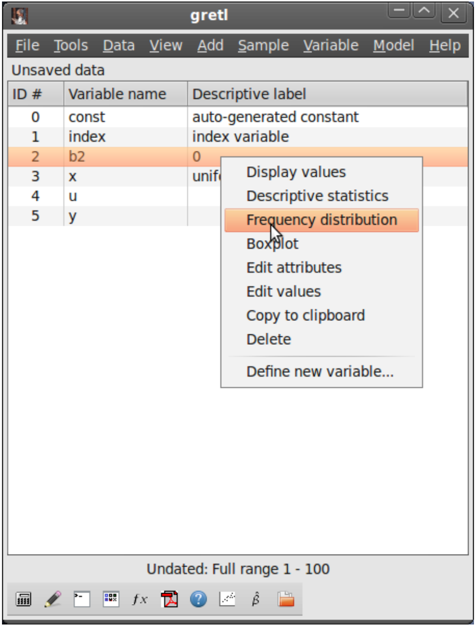
1. Triar els paràmetres del model:
   - Escollir els valors de tots els **paràmetres** del model: β i σ^2.
   - Triar la **mida mostral** N.
   - Generar una **mostra de** x de mida N (x fixes!).
2. Obtenir M estimadors $\hat{\beta}$ i $\hat{\sigma}^2$:
   - Generar una **mostra de** y de mida N (utilitzant β i σ^2).
   - **Estimar** $\hat{\beta}$ i $\hat{\sigma}^2$ per MQO amb la mostra generada.
   - **Guardar** els resultats.
   - Repetir el procés M **vegades** (bucle).
3. Presentar els resultats (histograma).

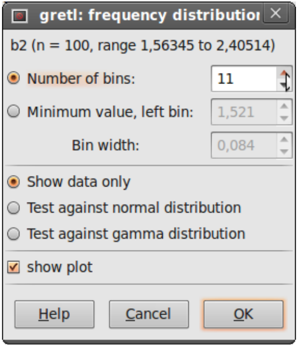
Experiments de Monte Carlo amb Gretl (I)

gretl: MonteCarlo_Tema2.inp

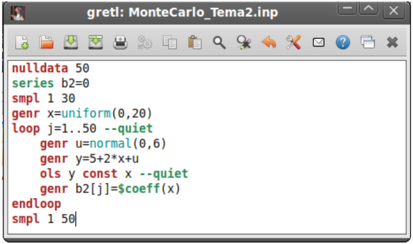
```
# Triem el nombre de mostres que volem generar (M):
nulldata 100

# Indiquem que es creï una variable, b2, per guardar posteriorment els resultats, i que,
# de moment, l'ompli de zeros
series b2=0

# Indiquem quina serà la mida mostral (N)
smpl 1 30

# Generem una mostra de x de mida N (que utilitzarem en totes les repeticions ja que hem
# assumit que les x són fixes!). En aquest cas les treurem d'una distribució uniforme
genr x=uniform(0,20)

# Aquesta línia indica que tot el que hi ha comprés entre aquest punt i l'ordre
# "endloop" es repetirà M=100 vegades (quiet indica que no mostri output)
loop j=1..100 --quiet

    # Generem les pertorbacions de la nostra mostra (assumim sigma^2=36)
    genr u=normal(0,6)
    # Generem la mostra de y (hem assumit beta1=5 i beta2=2)
    genr y=5+2*x+u
    # Indiquem que estimi el model per MQO (i que no mostri output)
    ols y const x --quiet
    # Guardem l'estimació del paràmetre associat a x en el lloc corresponent de b2
    genr b2[j]=$coeff(x)

endloop

# Indiquem que ens deixi generar totes les observacions generades
smpl 1 100
```

Experiments de Monte Carlo amb Gretl (II)

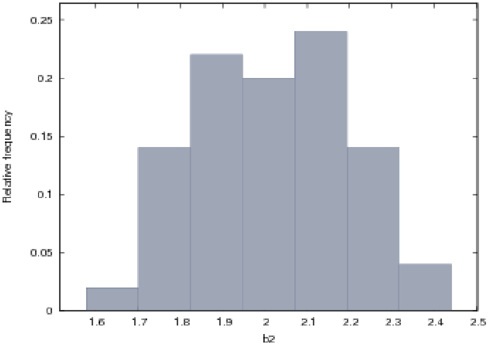




Experiments de Monte Carlo amb Gretl (III)



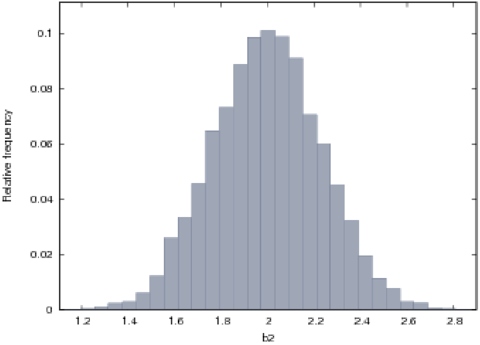

```
nulldata 50
series b2=0
smpl 1 30
genr x=uniform(0,20)
loop j=1..50 --quiet
    genr u=normal(0,6)
    genr y=5+2*x+u
    ols y const x --quiet
    genr b2[j]=$coeff(x)
endloop
smpl 1 50
```



Experiments de Monte Carlo amb Gretl (III)

gretl: MonteCarlo_Tema2.inp

```
nulldata 100
series b2=0
smpl 1 30
genr x=uniform(0,20)
loop j=1..100 --quiet
    genr u=normal(0,6)
    genr y=5+2*x+u
    ols y const x --quiet
    genr b2[j]=$coeff(x)
endloop
smpl 1 100
```

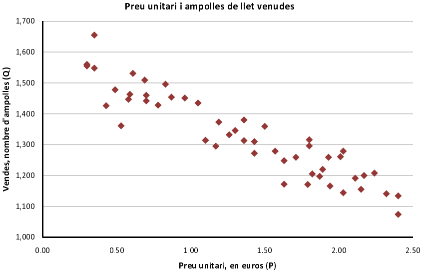


Continguts

TEMA 2. EL MODEL DE REGRESSIÓ SIMPLE

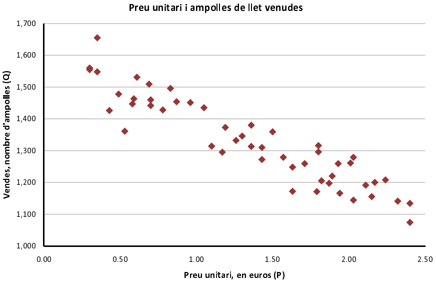
Els nostres tres exemples (del Tema 1)

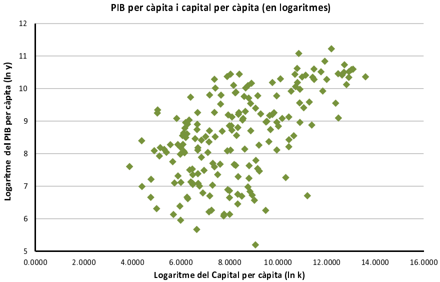
Demanda de llet: $Q_i = \beta_0 + \beta_1 P_i + u_i$



Els nostres tres exemples (del Tema 1)

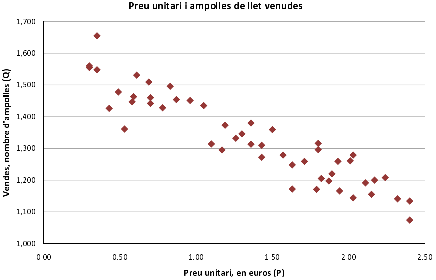
Demanda de llet: $Q_i = \beta_0 + \beta_1 P_i + u_i$ Cobb-Douglas: $\ln Y_i = \beta_0 + \beta_1 \ln K_i + u_i$



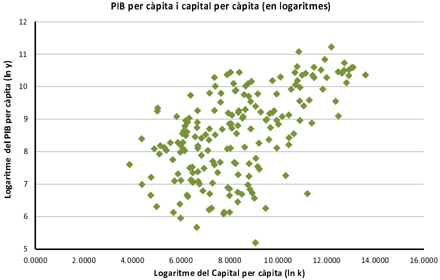


Els nostres tres exemples (del Tema 1)

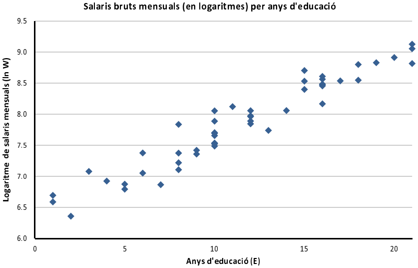
Demanda de llet: $Q_i = \beta_0 + \beta_1 P_i + u_i$



Cobb-Douglas: $\ln Y_i = \beta_0 + \beta_1 \ln K_i + u_i$



Salaris i educació: $\ln W_i = \beta_0 + \beta_1 E_i + u_i$



Exemple I: Demanda de llet $Q_i = \beta_0 + \beta_1 P_i + u_i$

La nostra base de dades és la següent:

Dades_Aplicacions.xls

	A	B
1	Preu	Vendes
2	0,3	1555
3	0,3	1560
4	0,35	1655
5	0,35	1548
6	0,43	1426
7	0,49	1478
8	0,53	1361
9	0,58	1447
10	0,59	1463
11	0,61	1531
12	0,69	1509
13	0,7	1460
14	0,7	1442
15	0,78	1428
16	0,83	1496
17	0,87	1454
18	0,96	1451
19	1,05	1435
20	1,1	1314
21	1,17	1295

Regressió (manualment)—coeficients

$$X = \begin{bmatrix} 1 & 0.3 \\ 1 & 0.3 \\ 1 & 0.35 \\ 1 & 0.35 \\ \vdots & \vdots \end{bmatrix} \quad y = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ 1548 \\ \vdots \end{bmatrix}$$

Regressió (manualment)—coeficients

$$X = \begin{bmatrix} 1 & 0.3 \\ 1 & 0.3 \\ 1 & 0.35 \\ 1 & 0.35 \\ \vdots & \vdots \end{bmatrix} \quad y = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ 1548 \\ \vdots \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + \ldots & 1 \cdot 0.3 + 1 \cdot 0.3 + 1 \cdot 0.35 + \ldots \\ 0.3 \cdot 1 + 0.3 \cdot 1 + 0.35 \cdot 1 + \ldots & 0.3 \cdot 0.3 + 0.3 \cdot 0.3 + 0.35 \cdot 0.35 + \ldots \end{bmatrix} = \begin{bmatrix} 50 & 67.45 \\ 67.45 & 111.28 \end{bmatrix}$$

Regressió (manualment)—coeficients

$$X'X = \qquad = \begin{bmatrix} 50 & 67.45 \\ 67.45 & 111.28 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{50 \cdot 111.28 - 67.45^2} \begin{bmatrix} 111.28 & -67.45 \\ -67.45 & 50 \end{bmatrix} = \begin{bmatrix} 0.1097 & -0.0665 \\ -0.0665 & 0.0493 \end{bmatrix}$$

Regressió (manualment)—coeficients

$$X = \begin{bmatrix} 1 & 0.3 \\ 1 & 0.3 \\ 1 & 0.35 \\ 1 & 0.35 \\ \vdots & \vdots \end{bmatrix} \quad y = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ 1548 \\ \vdots \end{bmatrix}$$

$$X'y = \begin{bmatrix} 1 \cdot 1555 + 1 \cdot 1560 + 1 \cdot 1655 + \ldots \\ 0.3 \cdot 1555 + 0.3 \cdot 1560 + 0.35 \cdot 1655 + \ldots \end{bmatrix} = \begin{bmatrix} 66568 \\ 85797.43 \end{bmatrix}$$

Regressió (manualment)—coeficients

$$(X'X)^{-1} = \qquad = \begin{bmatrix} 0.1097 & -0.0665 \\ -0.0665 & 0.0493 \end{bmatrix}$$

$$X'y = \qquad = \begin{bmatrix} 66568 \\ 85797.43 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0.1097 \cdot 66568 - 0.0665 \cdot 85797.43 \\ -0.0665 \cdot 66568 + 0.0493 \cdot 85797.43 \end{bmatrix} = \begin{bmatrix} 1597.55 \\ -197.32 \end{bmatrix}$$

Regressió (manualment)—coeficients

$$X = \begin{bmatrix} 1 & 0.3 \\ 1 & 0.3 \\ 1 & 0.35 \\ 1 & 0.35 \\ \vdots & \vdots \end{bmatrix} \qquad y = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ 1548 \\ \vdots \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1\cdot 1 + 1\cdot 1 + 1\cdot 1 + \ldots & 1\cdot 0.3 + 1\cdot 0.3 + 1\cdot 0.35 + \ldots \\ 0.3\cdot 1 + 0.3\cdot 1 + 0.35\cdot 1 + \ldots & 0.3\cdot 0.3 + 0.3\cdot 0.3 + 0.35\cdot 0.35 + \ldots \end{bmatrix} = \begin{bmatrix} 50 & 67.45 \\ 67.45 & 111.28 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{50\cdot 111.28 - 67.45^2} \begin{bmatrix} 111.28 & -67.45 \\ -67.45 & 50 \end{bmatrix} = \begin{bmatrix} 0.1097 & -0.0665 \\ -0.0665 & 0.0493 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 1\cdot 1555 + 1\cdot 1560 + 1\cdot 1655 + \ldots \\ 0.3\cdot 1555 + 0.3\cdot 1560 + 0.35\cdot 1655 + \ldots \end{bmatrix} = \begin{bmatrix} 66568 \\ 85797.43 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0.1097\cdot 66568 - 0.0665\cdot 85797.43 \\ -0.0665\cdot 66568 + 0.0493\cdot 85797.43 \end{bmatrix} = \begin{bmatrix} 1597.55 \\ -197.32 \end{bmatrix}$$

Regressió (manualment)—s.e.'s i R^2

$$\hat{u} = y - X\hat{\beta} = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ \vdots \end{bmatrix} - \begin{bmatrix} 1597.55 + (-197.32) \cdot 0.3 \\ 1597.55 + (-197.32) \cdot 0.3 \\ 1597.55 + (-197.32) \cdot 0.35 \\ \vdots \end{bmatrix} = \begin{bmatrix} 16.65 \\ 21.65 \\ 126.52 \\ \vdots \end{bmatrix}$$

Regressió (manualment)—s.e.'s i R^2

$$\hat{u} = y - X\hat{\beta} = \qquad\qquad = \begin{bmatrix} 16.65 \\ 21.65 \\ 126.52 \\ \vdots \end{bmatrix}$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{50-2} = \frac{16.65^2+21.65^2+126.52^2+\ldots}{48} = 2665.57$$

Regressió (manualment)—s.e.'s i R^2

$$\hat{\sigma}^2 = \qquad\qquad\qquad\qquad = 2665.57$$

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = 2665.57 \cdot \begin{bmatrix} 0.1097 & -0.0665 \\ -0.0665 & 0.0493 \end{bmatrix} = \begin{bmatrix} 292.44 & -177.26 \\ -177.26 & 131.40 \end{bmatrix}$$

Regressió (manualment)—s.e.'s i R^2

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \qquad = \begin{bmatrix} 292.44 & -177.26 \\ -177.26 & 131.40 \end{bmatrix}$$

$$s.e.(\hat{\beta}_0) = \sqrt{292.44} = 17.10 \quad s.e.(\hat{\beta}_1) = \sqrt{131.4} = 11.46$$

Regressió (manualment)—s.e.'s i R^2

$$\bar{y} = \frac{1555+1560+1655+...}{50} = 1331.36$$

Regressió (manualment)—s.e.'s i R^2

$$\bar{y} = \qquad\qquad\qquad\qquad = 1331.36$$

$$y - \bar{y}\iota = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ \vdots \end{bmatrix} - 1331.36 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1555 - 1331.36 \cdot 1 \\ 1560 - 1331.36 \cdot 1 \\ 1655 - 1331.36 \cdot 1 \\ \vdots \end{bmatrix} = \begin{bmatrix} 223.64 \\ 228.64 \\ 323.64 \\ \vdots \end{bmatrix}$$

Regressió (manualment)—s.e.'s i R^2

$$\hat{u} = y - X\hat{\beta} = \qquad = \begin{bmatrix} 16.65 \\ 21.65 \\ 126.52 \\ \vdots \end{bmatrix}$$

$$y - \bar{y}\iota = \qquad = \begin{bmatrix} 223.64 \\ 228.64 \\ 323.64 \\ \vdots \end{bmatrix}$$

$$R^2 = 1 - \frac{\hat{u}'\hat{u}}{(y-\bar{y}\iota)'(y-\bar{y}\iota)} = \frac{16.65^2+21.65^2+126.52^2+\ldots}{223.64^2+228.64^2+323.64^2+\ldots} = 0.8606$$

Regressió (manualment)—s.e.'s i R^2

$$\hat{u} = y - X\hat{\beta} = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ \vdots \end{bmatrix} - \begin{bmatrix} 1597.55 + (-197.32) \cdot 0.3 \\ 1597.55 + (-197.32) \cdot 0.3 \\ 1597.55 + (-197.32) \cdot 0.35 \\ \vdots \end{bmatrix} = \begin{bmatrix} 16.65 \\ 21.65 \\ 126.52 \\ \vdots \end{bmatrix}$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{50-2} = \frac{16.65^2+21.65^2+126.52^2+\ldots}{48} = 2665.57$$

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} = 2665.57 \cdot \begin{bmatrix} 0.1097 & -0.0665 \\ -0.0665 & 0.0493 \end{bmatrix} = \begin{bmatrix} 292.44 & -177.26 \\ -177.26 & 131.40 \end{bmatrix}$$

$$s.e.(\hat{\beta}_0) = \sqrt{292.44} = 17.10 \quad s.e.(\hat{\beta}_1) = \sqrt{131.4} = 11.46$$

$$\bar{y} = \frac{1555+1560+1655+\ldots}{50} = 1331.36$$

$$y - \bar{y}\iota = \begin{bmatrix} 1555 \\ 1560 \\ 1655 \\ \vdots \end{bmatrix} - 1331.36 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1555 - 1331.36 \cdot 1 \\ 1560 - 1331.36 \cdot 1 \\ 1655 - 1331.36 \cdot 1 \\ \vdots \end{bmatrix} = \begin{bmatrix} 223.64 \\ 228.64 \\ 323.64 \\ \vdots \end{bmatrix}$$

$$R^2 = 1 - \frac{\hat{u}'\hat{u}}{(y-\bar{y}\iota)'(y-\bar{y}\iota)} = \frac{16.65^2+21.65^2+126.52^2+\ldots}{223.64^2+228.64^2+323.64^2+\ldots} = 0.8606$$

Presentació dels resultats

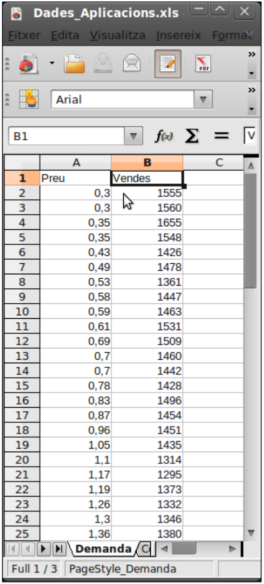
La forma estàndard de presentar els resultats d'una regressió és la següent:

$$\widehat{Q}_i = \underset{(17.10)}{1597.55} - \underset{(11.46)}{197.32}P_i \qquad R^2 = 0.8606$$

$$SQR = 127,947.22 \text{ (Opcional)}$$

Interpretació?

Importar dades



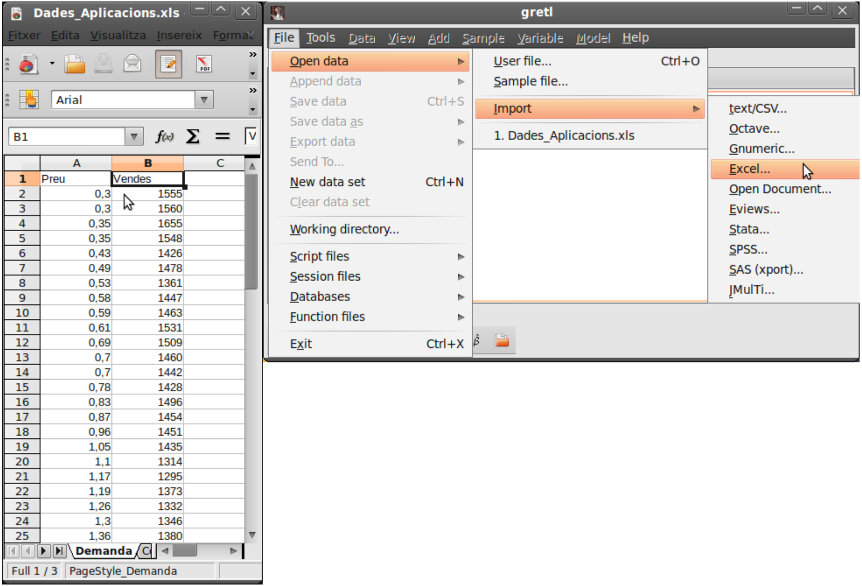
Dades_Aplicacions.xls

	A	B	C
1	Preu	Vendes	
2	0,3	1555	
3	0,3	1560	
4	0,35	1655	
5	0,35	1548	
6	0,43	1426	
7	0,49	1478	
8	0,53	1361	
9	0,58	1447	
10	0,59	1463	
11	0,61	1531	
12	0,69	1509	
13	0,7	1460	
14	0,7	1442	
15	0,78	1428	
16	0,83	1496	
17	0,87	1454	
18	0,96	1451	
19	1,05	1435	
20	1,1	1314	
21	1,17	1295	
22	1,19	1373	
23	1,26	1332	
24	1,3	1346	
25	1,36	1380	

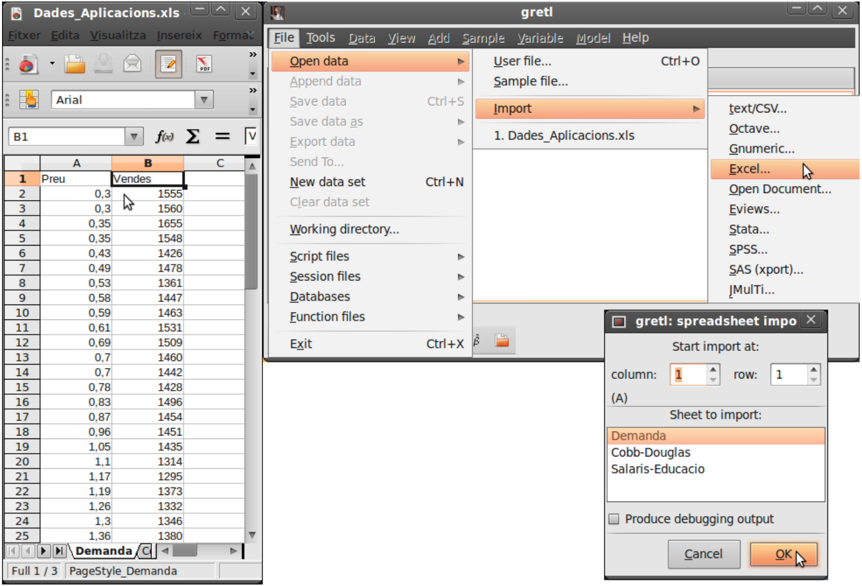
Demanda

Full 1 / 3 PageStyle_Demanda

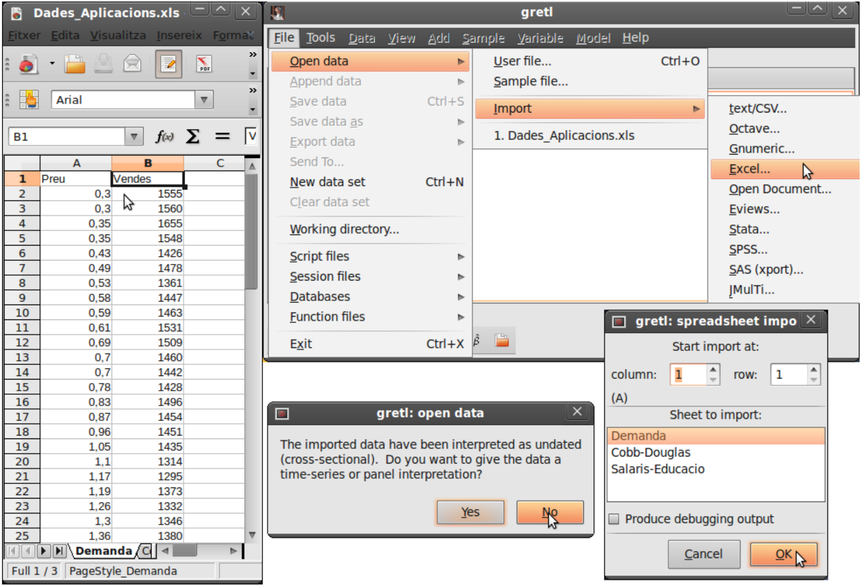
Importar dades



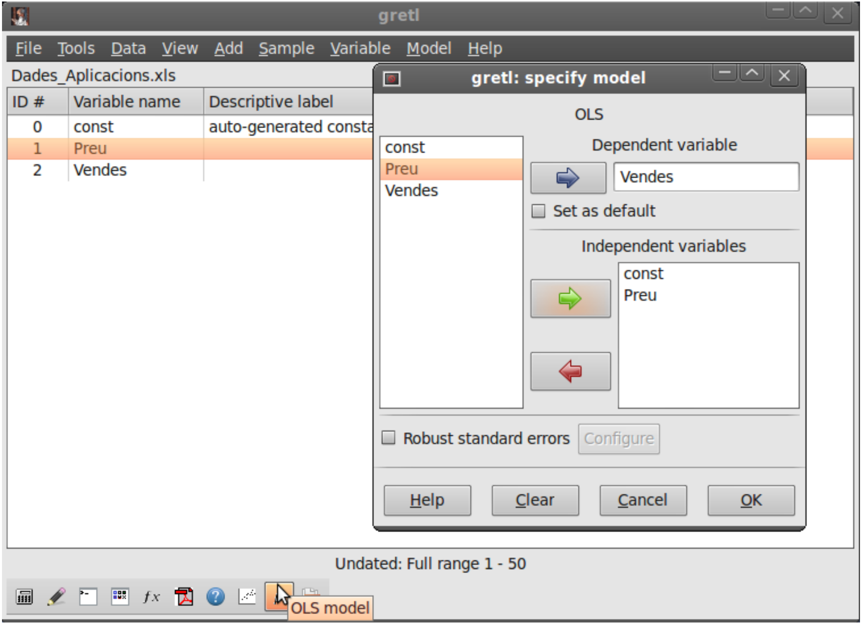
Importar dades



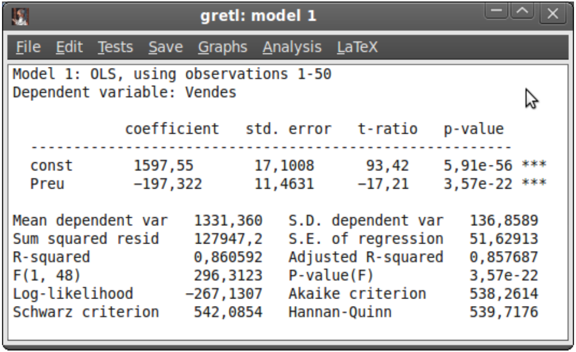
Importar dades



Regressió (Gretl, via menús)



Regressió (Gretl): Demanda de llet




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

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       1597,55       17,1008      93,42    5,91e-56 ***
  Preu        -197,322      11,4631     -17,21    3,57e-22 ***

Mean dependent var   1331,360   S.D. dependent var   136,8589
Sum squared resid    127947,2   S.E. of regression   51,62913
R-squared            0,860592   Adjusted R-squared   0,857687
F(1, 48)             296,3123   P-value(F)           3,57e-22
Log-likelihood      -267,1307   Akaike criterion     538,2614
Schwarz criterion    542,0854   Hannan-Quinn         539,7176
```

Regressió (Gretl): Demanda de llet

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        1597,55      17,1008      93,42    5,91e-56 ***
  Preu         -197,322     11,4631     -17,21    3,57e-22 ***

Mean dependent var    1331,360   S.D. dependent var    136,8589
Sum squared resid     127947,2   S.E. of regression    51,62913
R-squared             0,860592   Adjusted R-squared    0,857687
F(1, 48)              296,3123   P-value(F)            3,57e-22
Log-likelihood       -267,1307   Akaike criterion      538,2614
Schwarz criterion     542,0854   Hannan-Quinn          539,7176
```

Regressió (Gretl): Demanda de llet

gretl: model 1

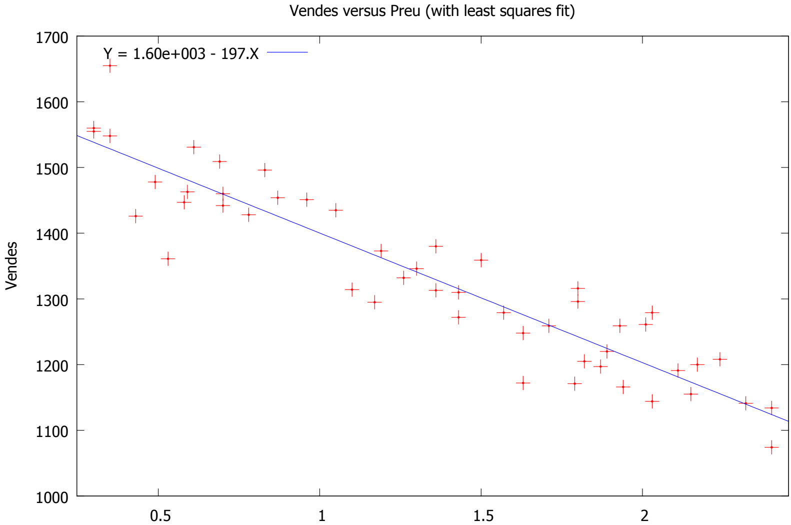
File Edit Tests Save Graphs Analysis LaTeX

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

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       1597,55       17,1008      93,42    5,91e-56 ***
  Preu       -197,322       11,4631     -17,21    3,57e-22 ***

Mean dependent var   1331,360   S.D. dependent var   136,8589
Sum squared resid    127947,2   S.E. of regression   51,62913
R-squared            0,860592   Adjusted R-squared   0,857687
F(1, 48)             296,3123   P-value(F)           3,57e-22
Log-likelihood      -267,1307   Akaike criterion     538,2614
Schwarz criterion    542,0854   Hannan-Quinn         539,7176
```

$$\widehat{Q}_i = \underset{(17.10)}{1,597.55} - \underset{(11.46)}{197.32} P_i \qquad R^2 = 0.8606$$

$$SQR = 127,947.22 \text{ (Opcional)}$$

Regressió (Gretl): Demanda de llet—Presentació gràfica



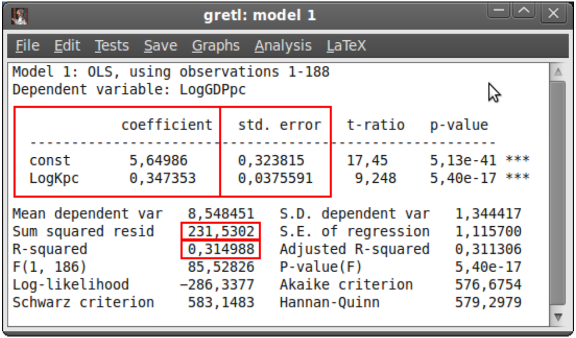
Funció de producció Cobb-Douglas: $\ln y_i = \beta_0 + \beta_1 \ln k_i + u_i$

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

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

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       5,64986      0,323815     17,45     5,13e-41 ***
  LogKpc      0,347353     0,0375591     9,248    5,40e-17 ***

Mean dependent var   8,548451   S.D. dependent var   1,344417
Sum squared resid    231,5302   S.E. of regression   1,115700
R-squared            0,314988   Adjusted R-squared   0,311306
F(1, 186)            85,52826   P-value(F)           5,40e-17
Log-likelihood      -286,3377   Akaike criterion     576,6754
Schwarz criterion    583,1483   Hannan-Quinn         579,2979
```

Funció de producció Cobb-Douglas: $\ln y_i = \beta_0 + \beta_1 \ln k_i + u_i$

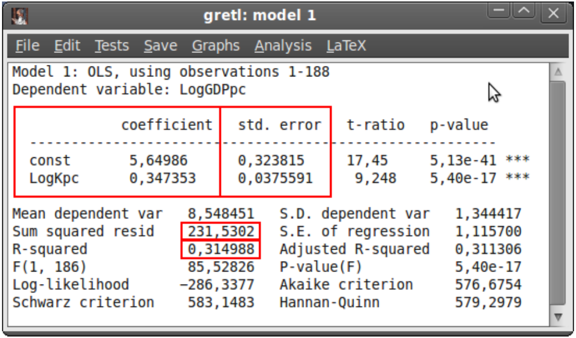


Model 1: OLS, using observations 1-188
Dependent variable: LogGDPpc

	coefficient	std. error	t-ratio	p-value	
const	5,64986	0,323815	17,45	5,13e-41	***
LogKpc	0,347353	0,0375591	9,248	5,40e-17	***

Mean dependent var	8,548451	S.D. dependent var	1,344417
Sum squared resid	231,5302	S.E. of regression	1,115700
R-squared	0,314988	Adjusted R-squared	0,311306
F(1, 186)	85,52826	P-value(F)	5,40e-17
Log-likelihood	−286,3377	Akaike criterion	576,6754
Schwarz criterion	583,1483	Hannan-Quinn	579,2979

Funció de producció Cobb-Douglas: $\ln y_i = \beta_0 + \beta_1 \ln k_i + u_i$




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

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       5,64986      0,323815     17,45     5,13e-41 ***
  LogKpc      0,347353     0,0375591     9,248    5,40e-17 ***

Mean dependent var   8,548451   S.D. dependent var   1,344417
Sum squared resid    231,5302   S.E. of regression   1,115700
R-squared            0,314988   Adjusted R-squared   0,311306
F(1, 186)            85,52826   P-value(F)           5,40e-17
Log-likelihood      -286,3377   Akaike criterion     576,6754
Schwarz criterion    583,1483   Hannan-Quinn         579,2979
```

$$\ln \widehat{y}_i = \underset{(0.32)}{5.65} + \underset{(0.04)}{0.35} \ln k_i \qquad R^2 = 0.3150$$

$$SQR = 231.53 \text{ (Opcional)}$$

Interpretació?

Rendiments de l'educació: $\ln W_i = \beta_0 + \beta_1 E_i + u_i$

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

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

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       6,37104      0,0672001     94,81    2,92e-56 ***
  Educacio    0,129172     0,00532663    24,25    1,38e-28 ***

Mean dependent var    7,853941   S.D. dependent var   0,709951
Sum squared resid     1,863741   S.E. of regression   0,197048
R-squared             0,924537   Adjusted R-squared   0,922965
F(1, 48)              588,0748   P-value(F)           1,38e-28
Log-likelihood        11,28901   Akaike criterion    -18,57801
Schwarz criterion    -14,75397   Hannan-Quinn        -17,12180
```

Rendiments de l'educació: $\ln W_i = \beta_0 + \beta_1 E_i + u_i$

```
gretl: model 1
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

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

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        6,37104      0,0672001     94,81    2,92e-56 ***
  Educacio     0,129172     0,00532663    24,25    1,38e-28 ***

Mean dependent var   7,853941   S.D. dependent var   0,709951
Sum squared resid    1,863741   S.E. of regression   0,197048
R-squared            0,924537   Adjusted R-squared   0,922965
F(1, 48)             588,0748   P-value(F)           1,38e-28
Log-likelihood       11,28901   Akaike criterion    -18,57801
Schwarz criterion   -14,75397   Hannan-Quinn        -17,12180
```

Rendiments de l'educació: $\ln W_i = \beta_0 + \beta_1 E_i + u_i$

```
gretl: model 1
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

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

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       6,37104      0,0672001     94,81    2,92e-56 ***
  Educacio    0,129172     0,00532663    24,25    1,38e-28 ***

Mean dependent var   7,853941   S.D. dependent var   0,709951
Sum squared resid    1,863741   S.E. of regression   0,197048
R-squared            0,924537   Adjusted R-squared   0,922965
F(1, 48)             588,0748   P-value(F)           1,38e-28
Log-likelihood       11,28901   Akaike criterion    -18,57801
Schwarz criterion   -14,75397   Hannan-Quinn        -17,12180
```

$$\ln \widehat{W}_i = \underset{(0.07)}{6.37} + \underset{(0.01)}{0.13}\, E_i \qquad R^2 = 0.9245$$

$$SQR = 1.86 \text{ (Opcional)}$$

Interpretació?