Microeconometria
Day # 3

`L. Cembalo`

Regressione con due variabili
e metodo dei minimi quadrati

SRF: sample regression function

- Il passaggio dalla regressione sulla popolazione a quella sul campione è cruciale
- A partire dai dati della popolazione riportati nella tabella precedente, estraiamo un campione (due volte)

SRF: sample regression function

Esempio di regressione: Y spesa familiare per consumo; X reddito familiare settimanale										
X	80	100	120	140	160	180	200	220	240	260
Y										
Spesa settimanale familiare per consumo (Y in $)	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
		88		113	125	140		160	189	185
				115				162		191
Totale	325	462	445	707	678	750	685	1043	966	1211
Media condizionata di Y, E(Y|X)	65	77	89	101	113	125	137	149	161	173

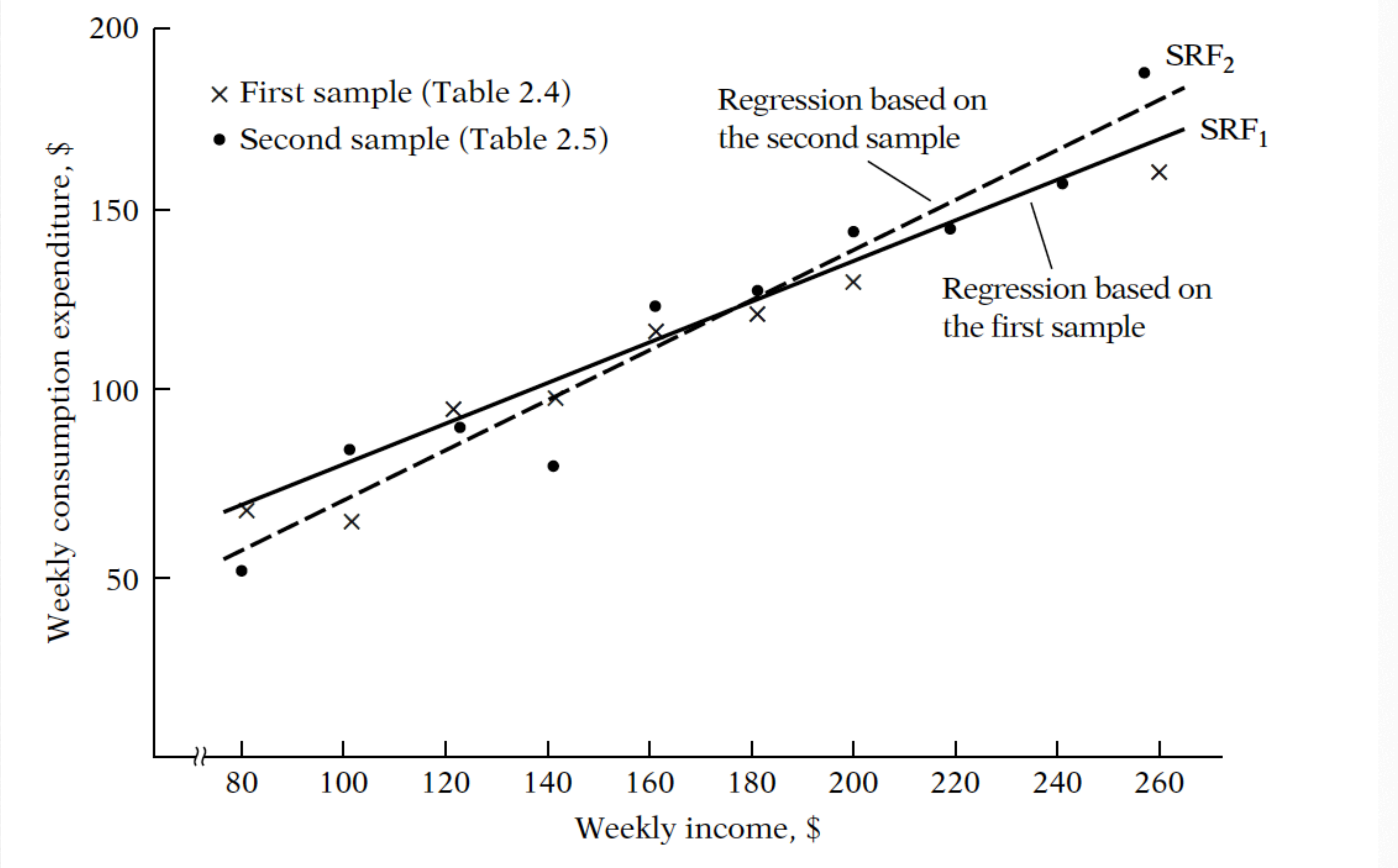
Estrazione casuale del campione

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Estrazione casuale del campione

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

SRF: sample regression function



SRF: sample regression function

Quale delle due linee di regressione rappresenta meglio la popolazione? In realtà non possiamo dirlo! Quello che possiamo fare è sviluppare il concetto di funzione di regressione campionaria (SRF)

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Dove la Y-cappello (Y-hat o Y-cap), è lo stimatore di E(YlXi)
I coefficienti beta cappello, invece, sono gli stimatori dei coefficienti della popolazione

E' bene ricordare che uno stimatore, anche noto come (sample) statistic, è una regola o formula o metodo che ci dice come stimare i parametri della popolazione a partire dalle informazioni contenute nel campione a disposizione. Un valore numerico specifico ottenuto con lo stimatore è una applicazione nota come stima.

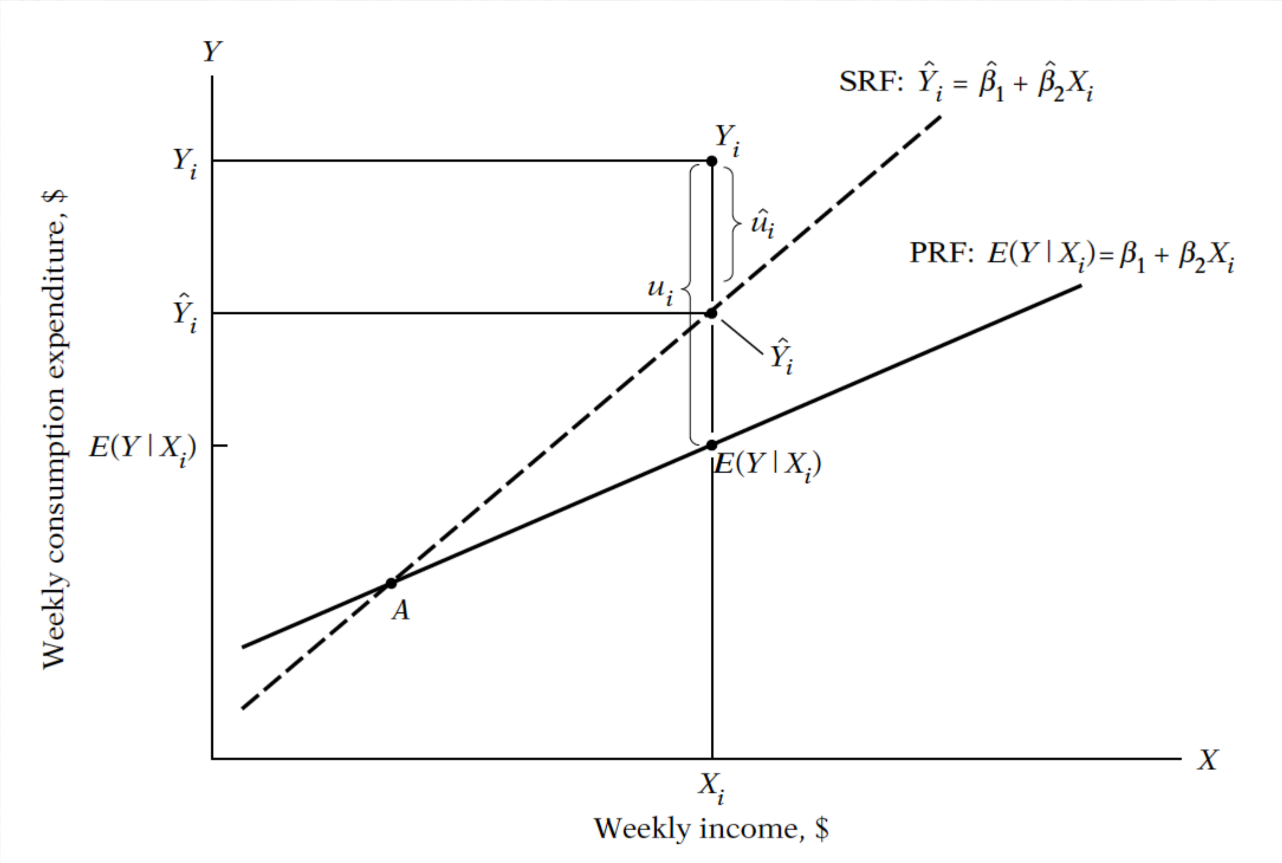
La forma stocastica della funzione sopra descritta è:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

u cappello è il term residuale campionario

SRF: sample regression function

In che relazione sono il PRF e il SRF (population regressione function e sample regression function)?



$$Y_i = \hat{Y}_i + \hat{u}_i$$

o, equivalentemente,

$$Y_i = E(Y \mid X_i) + u_i$$

Modello di regressione multipla

Ordinary Least Squares - Minimi Quadrati Ordinari

Richiamiamo il modello a due variabili PRF: $Y_i = \beta_1 + \beta_2 X_i + u_i$

In condizioni reali questo modello (PRF) non è direttamente osservabile. Lo è, invece, una sua manifestazione campionaria (SRF):

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

Dove $\hat{Y}_i$ è la stima (media condizionata) del valore di *Yi*.

A questo punto, tuttavia, la domanda è: come è determinato l'SRF? Esprimiamo la funzione precedente come:

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \end{aligned}$$

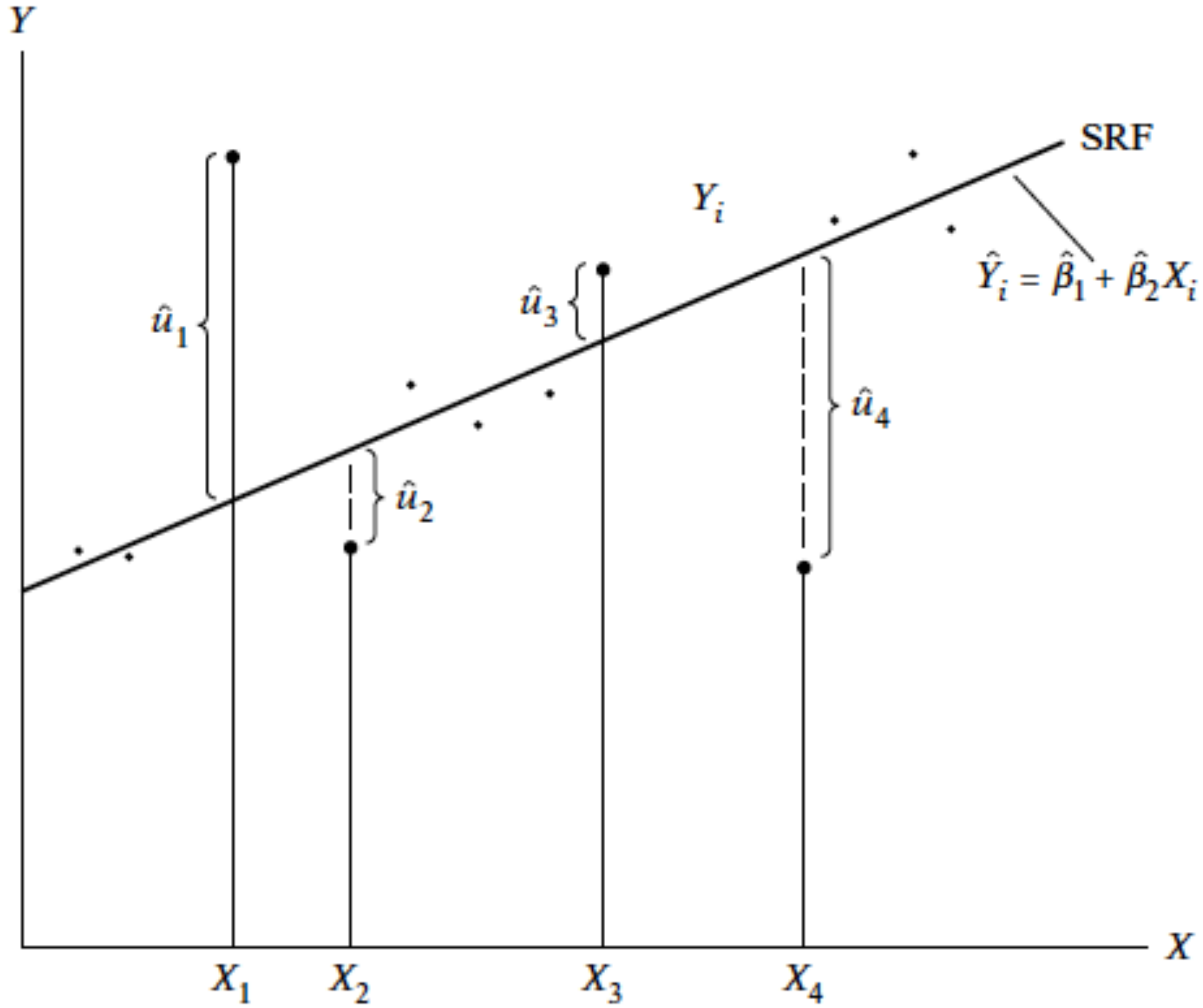
Questo mostra che i residui sono la differenza tra il valore "reale" di Y e la sua stima

Date le coppie X e Y di osservazioni, il tentativo è di determinare l'SRF in un modo tale da ottenere la stima di Y quanto più possibile vicina a quella reale. Detto in altri termini, dobbiamo trovare un criterio che minimizzi la somma dei residui

$$\sum \hat{u}_i = \sum (Y_i - \hat{Y}_i)$$

Modello di regressione multipla

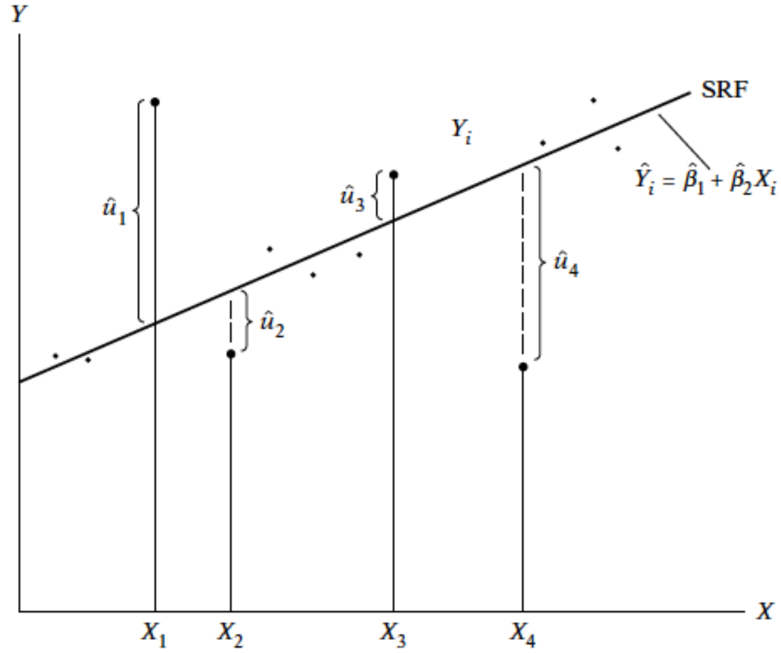
Ordinary Least Squares - Minimi Quadrati Ordinari



Least-squares criterion.

Modello di regressione multipla

Ordinary Least Squares - Minimi Quadrati Ordinari



Least-squares criterion.

Se minimizziamo la somma dei residui, così come sono, è come se stessimo pesando allo stesso modo tutte le deviazioni dalla retta: esempio di 10 -2 +2 -10.

Una soluzione possibile è quella di minimizzare la somma dei QUADRATI dei residui:

$$\sum \hat{u}_i^2 = f(\hat{\beta}_1, \hat{\beta}_2)$$

Si veda esempio numerico

Modello di regressione multipla

Ordinary Least Squares - Minimi Quadrati Ordinari

Consideriamo il seguente modello di regressione lineare multipla:

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 v_1 + e_1$$

$$y_2 = \beta_0 + \beta_1 x_2 + \beta_2 v_2 + e_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$y_n = \beta_0 + \beta_1 x_n + \beta_2 v_n + e_n$$

in formula matriciale è:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & v_1 \\ 1 & x_2 & v_2 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ 1 & x_n & v_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ \dots \\ e_n \end{bmatrix} \qquad \Leftrightarrow$$

$$\boldsymbol{y} \quad = \quad \boldsymbol{X} \qquad \boldsymbol{\beta} \quad + \quad \boldsymbol{e}$$

dove $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{e},$ sono, rispettivamente, vettori di variabile dipendete, esplicative coefficienti non-noti e vettore dei residui

Modello di regressione multipla

Ordinary Least Squares - Minimi Quadrati Ordinari

$y = X\beta + e$ iè il modello di regressione lineare riscritto in termini matriciali dove **y** e **e** sono vettori con *n* elementi, **X** è una matrice con *n* righe e *k+1* colonne, e β contiene *k+1* coefficienti ignoti.

ricordiamo che una ipotesi del modello di regressione classico (**X** deterministica e $E(e) = 0$), quindi sarà

$$E(y) = X\beta.$$

nota: A^{-1} è l'inversa di una matrice quadrata. Una delle sue proprietà è che $AA^{-1} = I$. A' è la trasposta della matrice quadrata.

Se premoltiplichiamo $E(y) = X\beta.$ per *X', abbiamo*

$$X'E(y) = (X'X)\beta$$

Poichè $(X'X)$ è quadrata e invertibile, per le ipotesi prima enunciate, può essere riscritta come:

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

$E(y)$ è ignoto, così possiamo rimpiazzarlo con una stima o, pragmaticamente, con le osservazione campionarie, ottenendo così:

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

Modello di regressione multipla

Ordinary Least Squares - Minimi Quadrati Ordinari

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

per costruzione questo è una stima basata sulla media condizionata che minimizza la somma dei quadrati dei residui e conserva le stesse proprietà della media

$\hat{\beta}$ è non distort su β. Ricordiamo che $y = X\beta + e$

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'[X\beta + e] =$$

$$= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'e = \beta + (X'X)^{-1}X'e$$

poichè $E(e) = 0$ e che Xè deterministica

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(e) = \beta$$

per quanto riguarda la varianza

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

lascio a voi la dimostrazione formale