

PRINCIPI DI ECONOMETRIA

lezione 7

AA 2015-2016

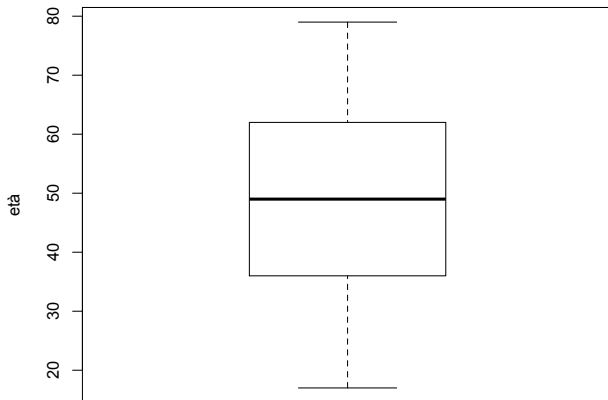
Paolo Brunori

Redditi svedesi

- il dataset contiene i dati di reddito di 838 individui
- il dataset contiene le variabili:
 - ▶ *sex* = sesso
 - ▶ *age* = età
 - ▶ *edu* = anni di istruzione
 - ▶ *y_gross* = reddito lordo

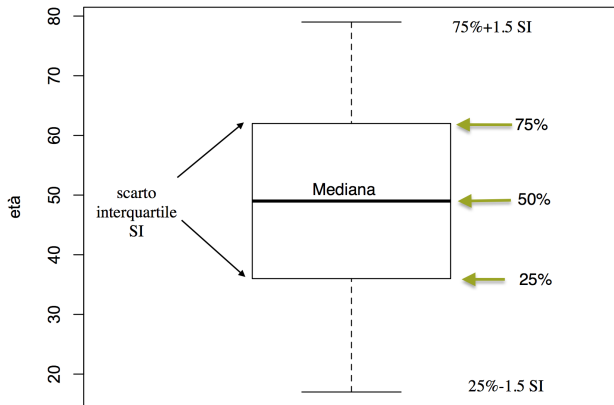
Blox plot

BOX PLOT



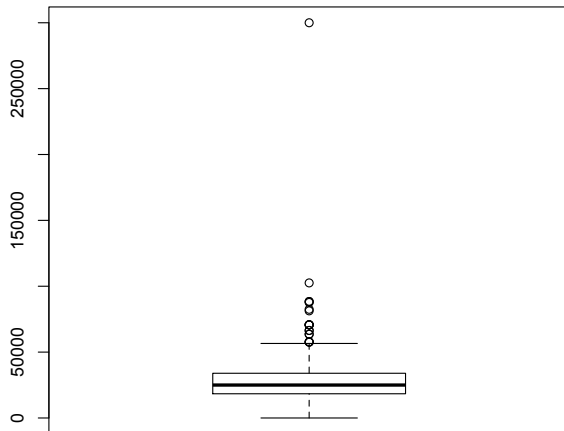
Blox plot

BOX PLOT



Redditi svedesi: outlier

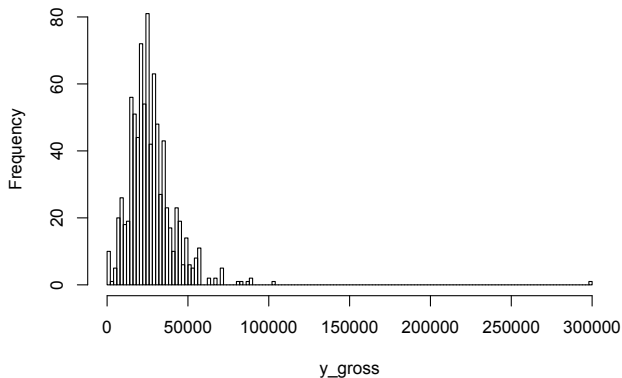
Redditi svedesi



outlier

la distribuzione del reddito: istogramma

Histogram of y_gross



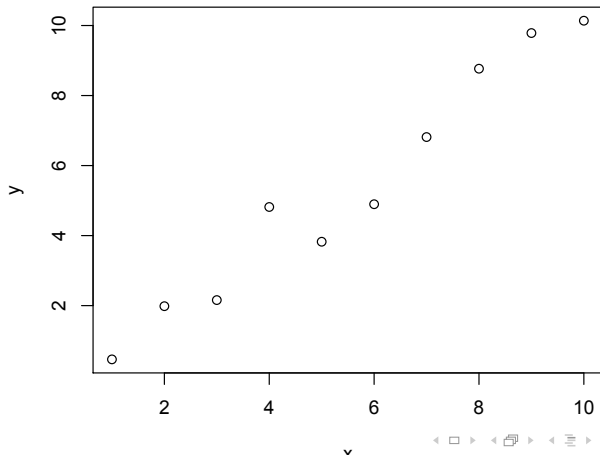
stimatore delle minime deviazioni assolute

- se gli outlier non sono rari
- è più efficiente uno stimatore meno sensibile ai valori u_i elevati
- lo stimatore delle minime deviazioni assolute (Least Absolute Deviations - LAD) ottenuto derivando b_0, b_1 che minimizzano:

$$\sum_{i=1}^n |Y_i - b_0 - b_1 X_i|$$

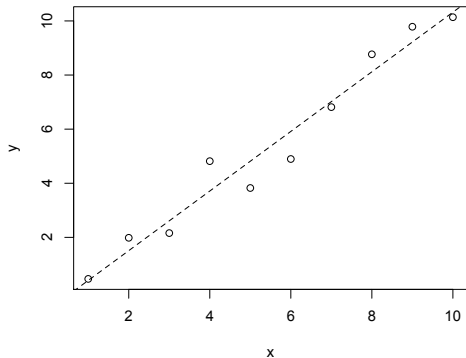
L'effetto di un outlier sulla retta di regressione

Immaginate due variabili legate linearmente



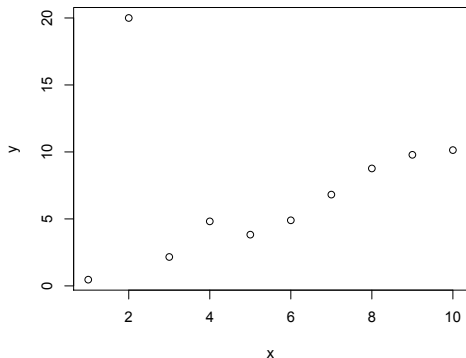
L'effetto di un outlier sulla retta di regressione

La retta di regressione interpola bene i dati



L'effetto di un outlier sulla retta di regressione

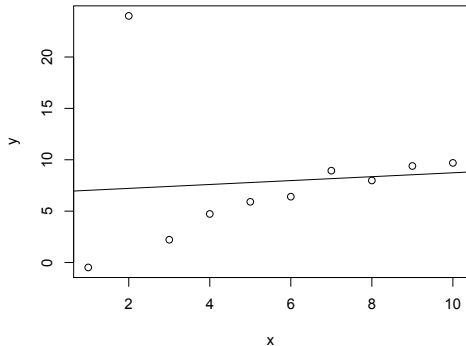
manipoliamo i dati in modo che il valore dell'osservazione che ha valore $x=2$ abbia valore $y=20$



Si tratta chiaramente di un outlier

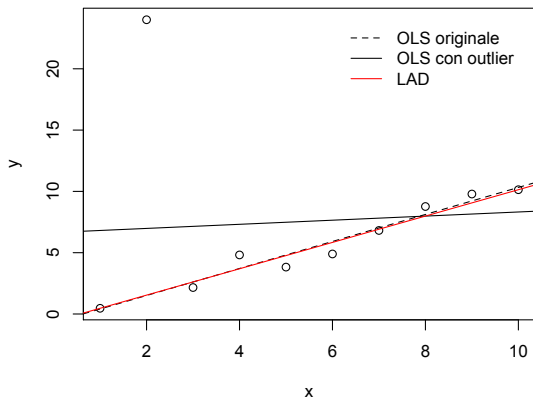
L'effetto di un outlier sulla retta di regressione

la retta stimata è estremamente sensibile alla presenza dell'outlier



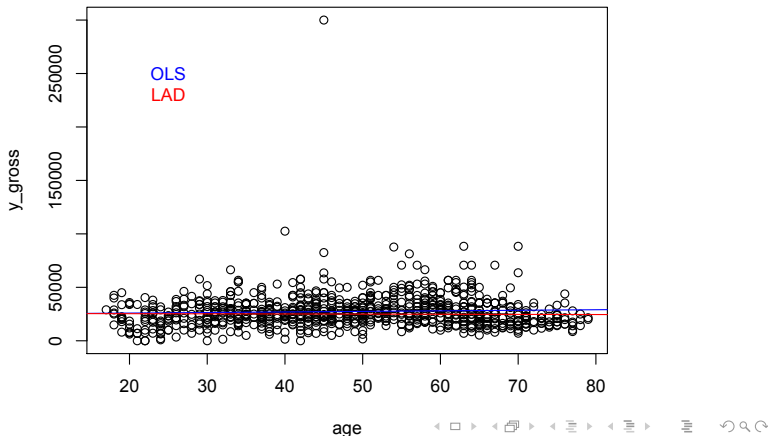
L'effetto di un outlier sulla retta di regressione

Se invece si utilizza la stima LAD si ottiene una retta praticamente identica a quella stimata sui dati originari



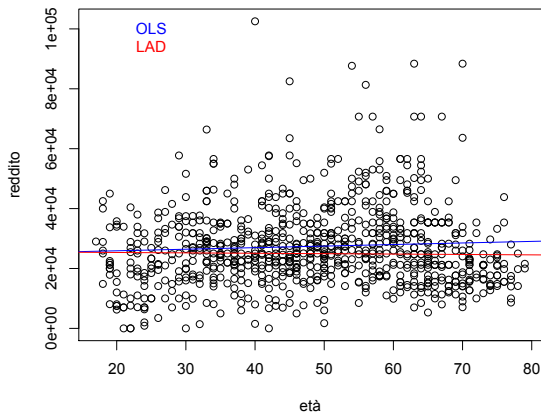
Redditi svedesi: outlier

Il caso dei dati di reddito svedesi è simile qui la distorsione dei coefficienti è molto meno marcato perché l'outlier è solo uno su oltre 800 osservazioni.



Redditi svedesi: outlier

la differenza è meglio apprezzabile se si mostra nel grafico la nuvola di punti escludendo l'outlier



attenzione: un outlier è stato rimosso dal grafico

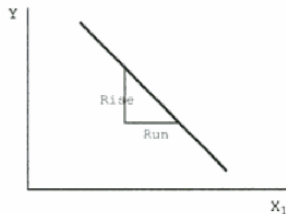
l'effetto di X su Y è fisso?

- nelle realtà economiche spesso l'effetto di una variabile indipendente dipende dalla variabile stessa
- la produttività di un fattore produttivo
- l'utilità del consumo di un'unità di bene
- l'aumento di qualità di un bene dovuto ad un euro in più speso per produrlo
- in questi casi l'approssimazione lineare potrebbe essere insoddisfacente

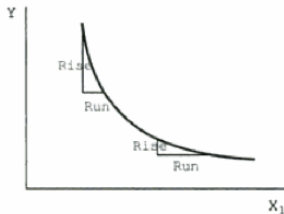
funzioni di regressione non lineari

- è possibile che Y sia funzione non lineare di X
- in questo caso ΔY dato ΔX cambia a seconda del valore di X
- quindi in teoria il β dovrebbe dipendere dal valore di X

Andamenti lineari e non lineari



(a) Pendenza costante



(b) Pendenza funzione del valore di X_1

Il legame età - reddito

- la relazione fra età e reddito è tipicamente considerata non lineare
- le retribuzioni crescono nella prima fase della vita lavorativa
- ma col passare del tempo la crescita rallenta
- e durante la pensione tipicamente diminuisce
- quindi sarebbe meglio utilizzare una funzione non lineare del regressore per interpolare i dati

Regressioni con trasformazione non lineare della variabile indipendente

- un modo per catturare un andamento non-lineare consiste nell'utilizzare come regressore una trasformazione dello stesso
- supponete di regredire Y come funzione di X e poi di $\tilde{X} = 12X$ quale coefficienti ottenete?
- se la relazione originaria è $Y = \beta_0 + \beta_1 X$
- la relazione diventa $Y = \beta_0 + \frac{\beta_1}{12} \tilde{X}$
- se la retribuzione (Y) aumenta di 120 (β_1) euro per ogni anno di anzianità (X)
- allora Y aumenta di 10 ($\frac{\beta_1}{12}$) euro per ogni mese di anzianità (\tilde{X})

Regressioni con trasformazione non lineare della variabile indipendente

- trasformare un regressore quindi implica modificare il suo coefficiente
- se si modifica il regressore aumentandone il valore il β diminuisce, se si diminuisce il valore del regressore il β aumenta
- se si utilizza una trasformazione non lineare del regressore ad esempio creando la variabile $\tilde{X} = X^2$ si modifica il coefficiente
- β non coglie più la relazione fra Y e X ma fra Y e una variabile che cresce con il quadrato di X (è minore di X se $X < 1$ ed è maggiore di X per valori maggiori di 1)

Regressioni con trasformazione non lineare della variabile indipendente

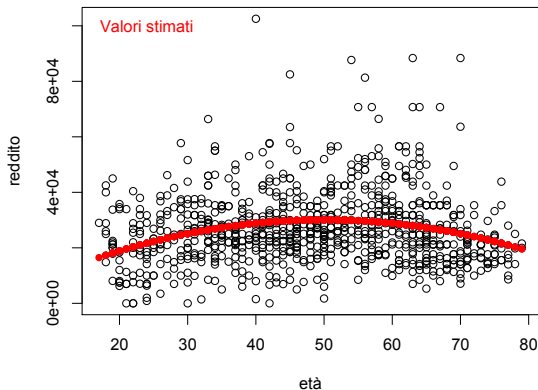
- il modello che spiega il reddito $Y = \beta_0 + \beta_1 \text{età} + \beta_2 \text{età}^2$
- ci aspettiamo $\beta_1 > 0$
- $\beta_2 > 0$ significa che il tasso di crescita del salario all'aumentare dell'esperienza è crescente
- $\beta_2 < 0$ significa che il tasso di crescita del salario all'aumentare dell'esperienza è decrescente

Regressione multipla redditi svedesi

	coefficiente	errore standard	t	$p - value$
β_0	-1166.809	4030.44	-0.289	0.7720
β_{age}	1251.04	178.30	7.016	0.0000
β_{age^2}	-12.509	1.843	-6.786	0.0000

- $R^2 = 0.05652$, R^2 corretto = 0.05426
- errore standard di regressione = 12920

Redditi svedesi: regressioni con trasformazione non lineare della variabile indipendente



Regressione multipla redditi svedesi

	coefficiente	errore standard	t	$p - value$
β_0	-14693.729	4149.33	-3.541	0.0004
β_{age}	1119.13	171.52	6.522	0.0000
β_{age^2}	-10.407	1.785	-5.831	0.0000
β_{sex}	1054.07	122.67	8.592	0.0000
β_{edu}	1776.77	856.584	2.074	0.0383

- $R^2 = 0.138$, $R^2_{\text{corretto}} = 0.1339$
- errore standard di regressione = 12370