Principi di Econometria

lezione 10

AA 2016-2017

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

Redditi svedesi

sex	age	edu	y_gross
Male	69	4	21213.2
Male	70	12	88388.35
Female	67	13	27000
Male	26	15	21500
Female	49	16	36950.42
Male	70	11	33941.13
Male	51	15	45000
Female	54	17	87681.24
Male	31	18	35355.34
Male	34	16	56568.54
Female	64	8	22000
Male	25	16	10000
Male	48	12	35000
Female	57	17	16970.56
Female	36	22	45033.32
Female	45	16	40000
Female	60	13	26000
Female	58	13	53033.01
Male	33	14	42500
Male	32	19	28284.27
Female	65	12	42426.41

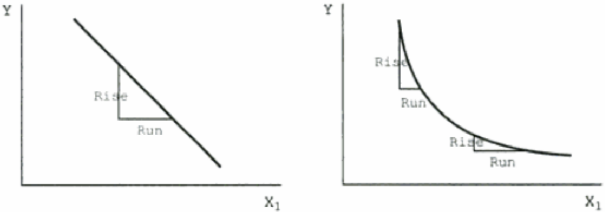
l'effetto di X su Y è fisso?

- 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

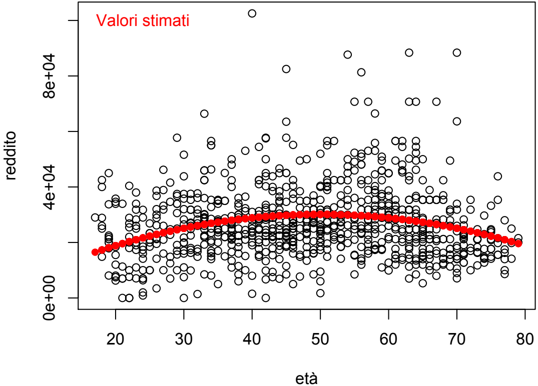
- il modello che spiega il reddito
  $Y = \beta_0 + \beta_1 \text{età} + \beta_2 \text{età}^2$
- ci aspettiamo $\beta_1 > 0$
- $\beta_2 > 0$ significa che il tasso di crescita del salario all'aumentare dell'esperienza è crescente
- $\beta_2 < 0$ significa che il tasso di crescita del salario all'aumentare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



attenzione: un outlier è stato rimosso dal grafico

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 $= 12370$

l'effetto di una variazione di X quando il regressore è una trasformazione non lineare di X

- nel modello:

$$Y = \beta_0 + \beta_1 X^2$$

- dalla stima ottengo $\beta_0 = 0.38, \beta_1 = 0.005$ come varia Y a seguito della variazione ΔX?

$$\Delta Y = [\beta_0 + \beta_1 \times (X + \Delta X)^2] - [\beta_0 + \beta_1 \times X^2]$$

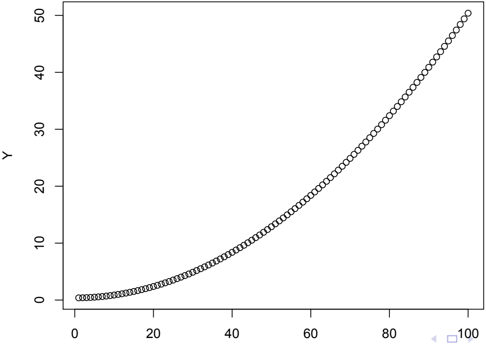
$$\Delta Y = \Delta X^2 \beta_1 = \Delta X^2 \times 0.005$$

l'effetto di una variazione di X quando il regressore è una trasformazione non lineare di X

- ma una variazione unitaria di X implica una variazione non unitaria di X^2
- ΔX^2 dovuto ad una variazione unitaria di X dipende dal valore iniziale di X
- ad esempio
  - se X varia da 3 a 4: $\Delta X^2 = 16 - 9 = 7$
  - se X varia da 7 a 8: $\Delta X^2 = 64 - 49 = 25$

l'effetto di una variazione di X quando il regressore è una trasformazione non lineare di X

- tornando al nostro esempio: $\Delta Y = \Delta X^2 \beta_1 = \Delta X^2 \times 0.005$
- come varia Y al variare unitario di X?



l'effetto di una variazione di X quando il regressore è una trasformazione non lineare di X

- nel caso dei redditi svedesi X è presente sia linearmente che al quadrato
- ci sono due effetti: lineare positivo e non lineare negativo

$$\Delta Y = \beta_1 \Delta X - \beta_2 \Delta X^2$$

$$\Delta Y = 1119.13 \Delta X - 10.407 \Delta X^2$$

- per uno svedese che passa da 20 a 21 anni questo implica:

$$\Delta Y = 1119.13 \times (21 - 20) - 10.407 \times (441 - 400) = 692.45$$

l'effetto di una variazione di X quando il regressore è una trasformazione non lineare di X

- per uno svedese che passa da 65 a 66 anni questo implica:

$$\Delta Y = 1119.13 \times (66-65) - 10.407 \times (4356-4225) = -1252.43$$

l'effetto di una variazione di X quando il regressore è una trasformazione non lineare di X

- col tempo l'interpretazione diventa automatica ma se avete dei dubbi dovete mettere per scritto la variazione di Y dovuta ad una variazione di Y
- calcolate Y prima e dopo la variazione e poi semplificate per ottenere ΔY dovuto a una variazione di ΔX

come procedere?

1. identificare una possibile relazione non lineare
2. stimare un OLS con funzione non lineare
3. testare se costituisce un miglioramento rispetto a quella lineare
4. rappresentare quello che stiamo facendo graficamente
5. stimare l'effetto di una variazione di X su Y

Regressione polinomiale

- se la relazione è non lineare potrebbe essere meglio approssimata da una funzione più complessa di quella quadratica
- il modello polinomiale di grado r:

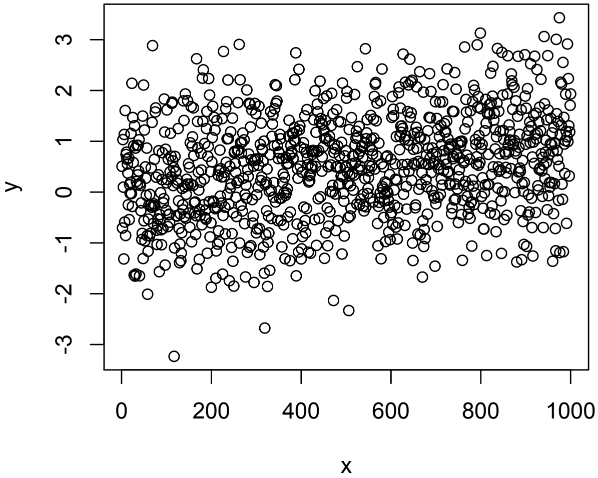
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_r X_i^r + u_i$$

- stesse tecniche di inferenza usate nel caso lineare

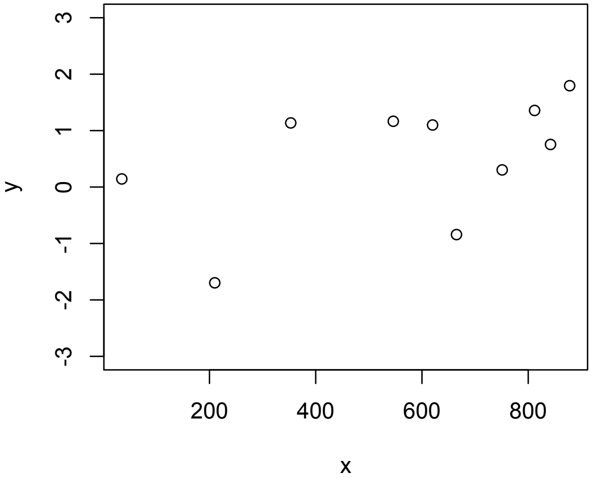
quale grado di polinomio usare?

- sappiamo che se usiamo $r = n - 1$ avremo un'interpolazione perfetta dei dati (r regressori $+\beta_0$=n)
- d'altra parte stiamo usando tutte le informazioni a disposizione
- restiamo senza gradi di libertà: l'errore di stima sul campione è zero ma non sappiamo nulla riguardo quanto accurata sia la nostra stima!
- ma non abbiamo variabilità per valutare l'affidabilità delle stime
- mi devo chiedere che modello sceglierei se potessi osservare tutta la popolazione e non solo un campione

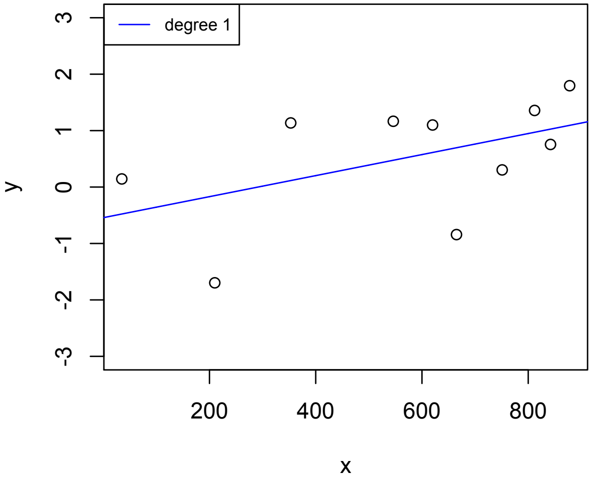
Popolazione



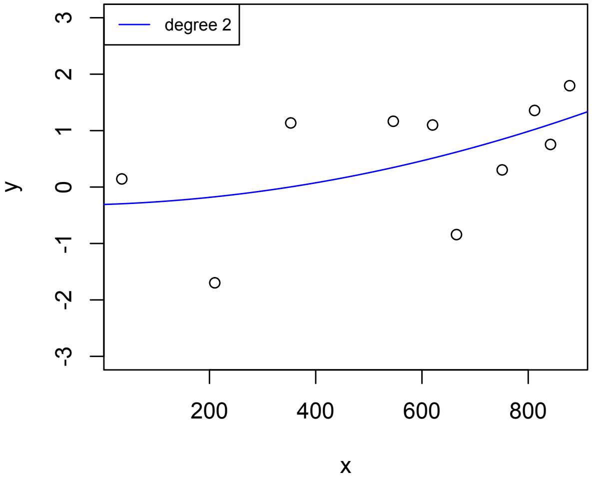
Campione $n = 10$



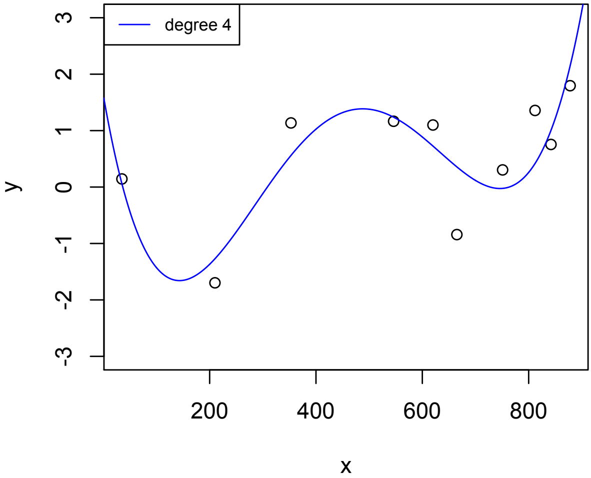
$$Y = \beta_0 + \beta_1 X$$



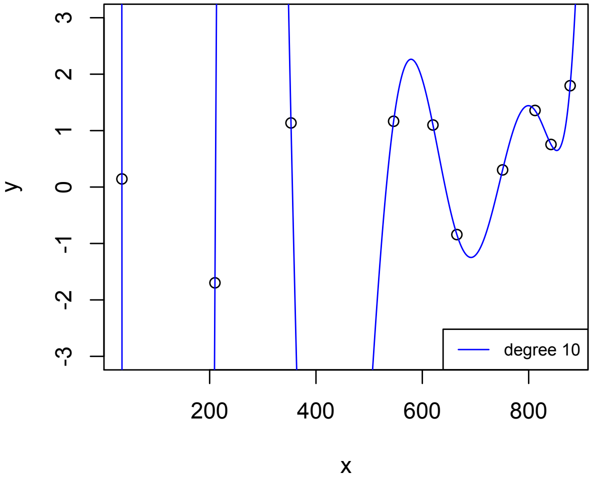
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2$$



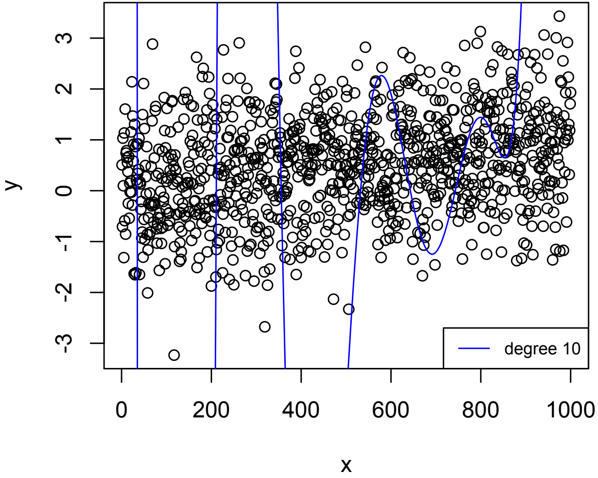
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_4 X^4$$



$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + ... + \beta_n X^n$$



$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_n X^n$$



Quale grado di polinomio usare?

- un'interpolazione perfetta sul campione non implica una buona capacità di spiegare la relazione nella popolazione
- interpolare perfettamente un campione significa interpretare anche la componente casuale come parte della relazione
- il potere predittivo di un modello ottenuto così è inferiore al potere predittivo di un modello polinomiale di grado inferiore
- quindi non dobbiamo cercare il polinomio che meglio interpola i dati del campione!

Polinomi

	coefficiente	errore standard	t	$p - value$
β_0	1.0812	0.2509	4.309	0.00258
β_X	0.8836	0.7934	1.114	0.29777

- $R^2 = 0.1342$, R^2 corretto $= 0.026$
- errore standard di regressione $= \sqrt{\frac{1}{n-k}\sum_i^n u_i^2} = 0.7934$

Polinomi

	coefficiente	errore standard	t	$p - value$
β_0	1.0812	0.2509	4.096	0.00459
β_X	0.8836	0.8347	1.059	0.32494
β_{X^2}	0.3990	0.8347	0.478	0.64718

- $R^2 = 0.1616$, R^2 corretto $= -0.07795$
- errore standard di regressione $= \sqrt{\frac{1}{n-k} \sum_i^n u_i^2}$ $=0.8347$

Polinomi

	coefficiente	errore standard	t	$p - value$
β_0	1.0812	NA	NA	NA
β_X	-0.8836	NA	NA	NA
β_{X^2}	0.3990	NA	NA	NA
β_{X^3}	-0.3463	NA	NA	NA
β_{X^4}	-1.3649	NA	NA	NA
β_{X^5}	-0.1688	NA	NA	NA
β_{X^6}	0.2984	NA	NA	NA
β_{X^7}	-0.6287	NA	NA	NA
β_{X^8}	-0.8308	NA	NA	NA
β_{X^9}	1.3003	NA	NA	NA

- $R^2 = 1.0000$, R^2 corretto $= NAN$
- errore standard di regressione $= \sqrt{\frac{1}{n-k}\sum_i^n u_i^2}$ =NAN

Regressione multipla redditi svedesi: quale polinomio usare?

	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^2corretto $= 0.1339$
- errore standard di regressione $= \sqrt{\frac{1}{n-k} \sum_i^n u_i^2}$=12370

Regressione multipla redditi svedesi: quale polinomio usare?

	coefficiente	errore standard	t	$p - value$
β_0	19625.42	10123.7622	1.939	0.05289
β_{age}	-1486.15	722.42	-2.057	0.0399
β_{age^2}	48.51	15.97	3.037	0.00247
β_{age^3}	-0.412	0.1111	-3.711	0.00022
β_{edu}	1094.39	122.23	8.953	0.0000
β_{sex}	1851.75	850.32	2.178	0.02971

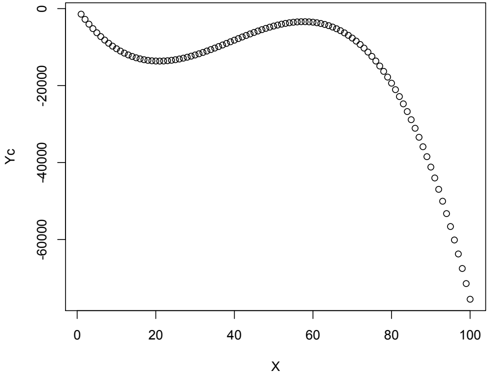
- $R^2 = 0.1521,\ R^2\text{corretto} = 0.147$
- errore standard di regressione $= \sqrt{\frac{1}{n-k}\sum_i^n u_i^2}$=12270

Redditi svedesi, l'effetto di una variazione unitaria dell'età con stima polinomiale di terzo grado

- ora abbiamo tre coefficienti che si riferiscono all'età
- in questi casi è quasi obbligatorio scrivere la variazione e chiedersi come cambia al variare di X

$$\Delta Y = -1486.15\Delta X + 48.51\Delta X^2 - 0.412\Delta X^3$$

Redditi svedesi, l'effetto di una variazione unitaria dell'età con stima polinomiale di terzo grado



Redditi svedesi, l'effetto di una variazione unitaria dell'età con stima polinomiale di terzo grado: all'interno dell'intervallo di età nel campione