

Principi di Econometria

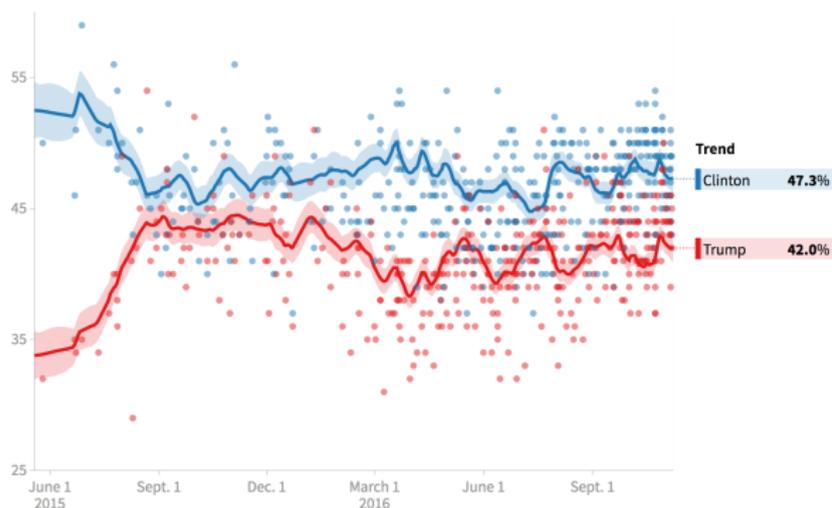
lezione 11

AA 2016-2017

Paolo Brunori

Cos'è successo in US?

- andamento dei sondaggi



Come si prevede un risultato elettorale?

- $Y = \beta_i X_i + u$
- quali variabili includereste nel modello?
- perchè hanno sbagliato?

Cos'è successo in UK?

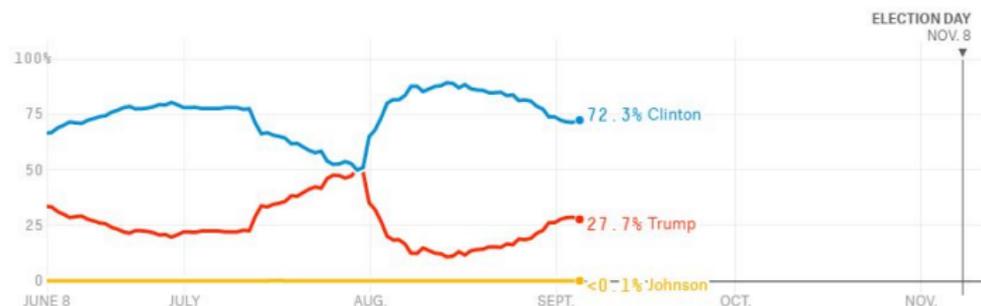
- andamento dei sondaggi prima del referendum sulla permanenza in EU

Remain %	Leave %	Undecided %	Date	Pollster	Sample
55	45	0	Jun 22, 2016	Populus	4,700
48	42	11	Jun 22, 2016	ComRes	1,032
41	43	11	Jun 22, 2016	TNS	2,320
44	45	11	Jun 22, 2016	Opinium	3,000
51	49	0	Jun 22, 2016	YouGov	3,766
52	48	0	Jun 22, 2016	Ipsos MORI	-
45	44	11	Jun 20, 2016	Survation	1,003
42	44	13	Jun 19, 2016	YouGov	1,652
53	46	2	Jun 19, 2016	ORB	800
45	42	13	Jun 18, 2016	Survation	1,004
44	44	12	Jun 17, 2016	Opinium	2,006
42	44	14	Jun 17, 2016	YouGov	1,694
44	43	13	Jun 16, 2016	Yougov	1,734
46	43	11	Jun 15, 2016	BMG Research	1,064
42	45	13	Jun 15, 2016	Survation	1,104
41	51	9	Jun 15, 2016	BMG Research	1,468
43	49	3	Jun 14, 2016	Ipsos MORI	1,257
45	50	6	Jun 13, 2016	ICM	1,000
46	45	9	Jun 13, 2016	Comres	1,002

Come si prevede un risultato elettorale?

How the odds have changed

We'll be updating our forecasts every time new data is available, every day through Nov. 8.



dove siamo arrivati?

- la ricerca del miglior modello si basa sull'utilizzo di tutti i regressori che hanno un impatto su Y
- poichè l'impatto potrebbe essere non lineare si devono valutare anche trasformazioni non lineari dei regressori
- per questi regressori si possono usare le stesse formule di inferenza dell OLS
- il numero dei regressori deve essere però valutato attentamente
- molti regressori e molte trasformazioni non lineari dei regressori finiranno per interpolare perfettamente i dati
- ma saremo totalmente all'oscuro del livello di affidabilità delle stime ottenute

Utilizzo di variabili dicotomiche

- il coefficiente di una variabile dicotomica X spiega qual'è la differenza fra il valore atteso di Y per la categoria per la quale $X=1$
- immaginiamo di voler spiegare il reddito con il genere e il titolo di studio (laurea o meno) di un lavoratore:

$$Y_i = \beta_0 + \beta_1 D_{1,i} + \beta_2 D_{2,i}$$

- dove $D_{1,i}$ = laureata/o e $D_{2,i}$ = sesso
- come interpretate i coefficienti?

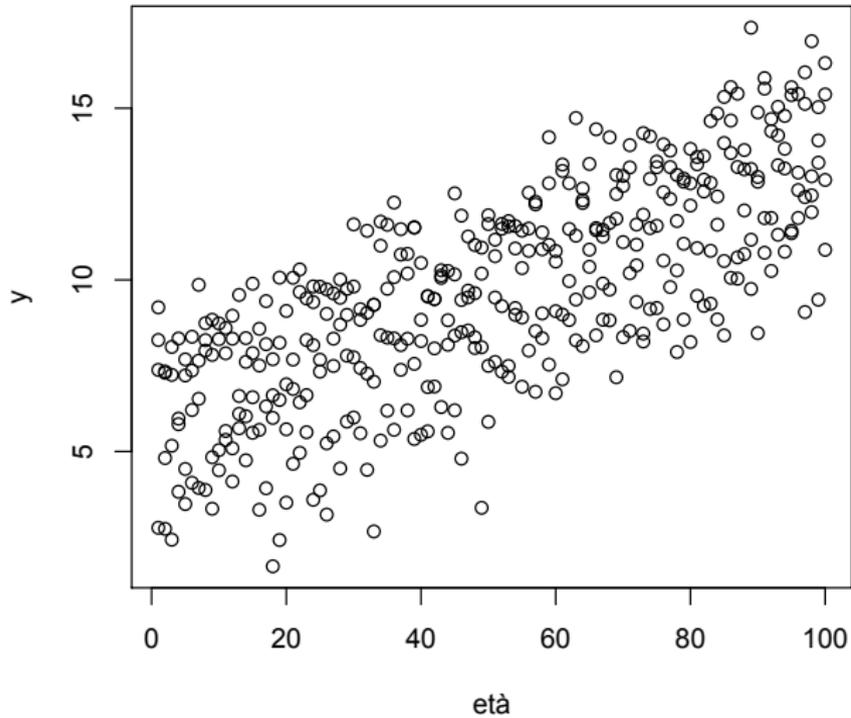
- in questo modello la laurea ha un effetto identico per i due gruppi
- è possibile considerare un possibile effetto di interazione fra le due variabili?
- ho bisogno di un regressore che mi segnali quando entrambe le variabili dicotomiche valgono 1

l'effetto di avere una laurea

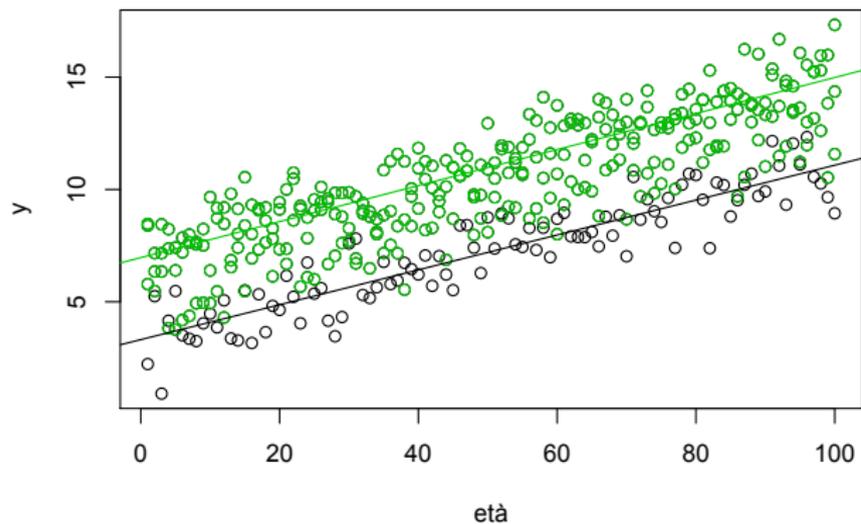
Qual è l'effetto della laurea $D_{1,i}$? Ipotizziamo che il sesso non sia determinato $D_{2,i} = d_2$

- $E(Y|D_{1,i} = 0, D_{2,i} = d_2) = \beta_0 + \beta_1 \times 0 + \beta_2 d_2 + \beta_3 \times 0 \times d_2 = \beta_0 + \beta_2 d_2$
- $E(Y|D_{1,i} = 1, D_{2,i} = d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2 = \beta_0 + \beta_1 + (\beta_2 + \beta_3) d_2$
- l'effetto di avere o meno una laurea D_1 dipende dal sesso (D_2): $\beta_1 + \beta_3 d_2$

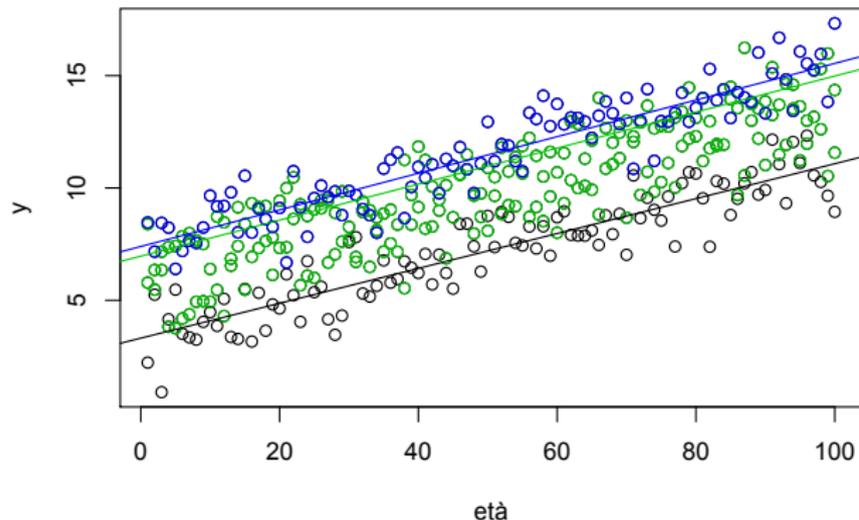
premio di laurea e sesso



premio di laurea e sesso



Regressione con variabile dicotomica sesso



Regressione con variabile dicotomica sesso e interazione fra sesso e laurea

- $D =$ uomo, $X =$ esperienza lavorativa.
- modello base:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$

- in pratica stiamo stimando due rette parallele di regressione
- ma nella realtà le prospettive di carriera potrebbero dipendere dal sesso
- anche qui si può aggiungere un termine di interazione

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + u_i$$

- se si tratta di un uomo la funzione di regressione è:

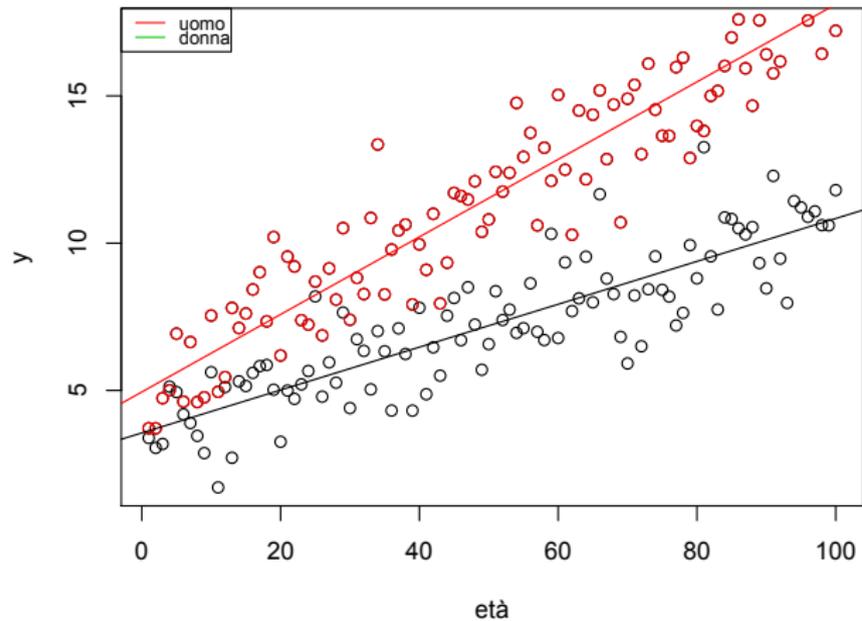
$$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_i + u_i$$

- se donna:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- per cui ho intercette diverse e pendenze diverse

premio di laurea e sesso



variabili interagite

- quando introduciamo variabili interagite otteniamo una variabile artificiale il cui valore dipende da entrambe
- nel caso delle variabili dicotomiche l'interpretazione è facile (ulteriore riduzione/aumento dovuto alla compresenza di entrambe le caratteristiche)
- quando le variabili sono continue il coefficiente coglie un andamento congiunto
- ad esempio se EDU sono gli anni di istruzione e AGE è l'età

$$Y = \beta_0 + \beta_1 EDU + \beta_2 AGE + \beta_3 EDU \times AGE + u$$

- β_1, β_2 colgono l'effetto dell'aumento di uno dei due regressori tenendo fermi tutti gli altri
- β_3 coglie l'effetto di un aumento congiunto

- qual è l'effetto di un anno di istruzione in più?

$$\begin{aligned}\Delta Y &= [\beta_0 + \beta_1(EDU + 1) + \beta_2AGE + \beta_3(EDU + 1) \times AGE] \\ &\quad - [\beta_0 + \beta_1EDU + \beta_2AGE + \beta_3EDU \times AGE] \\ \Delta Y &= \beta_1 + \beta_3 \times AGE\end{aligned}$$

- dipende da AGE

- qual è l'effetto di un anno di età in più?

$$\begin{aligned}\Delta Y &= [\beta_0 + \beta_1 EDU + \beta_2 (AGE + 1) + \beta_3 EDU \times (AGE + 1)] \\ &\quad - [\beta_0 + \beta_1 EDU + \beta_2 AGE + \beta_3 EDU \times AGE]\end{aligned}$$

$$\Delta Y = \beta_1 + \beta_3 \times EDU$$

- dipende da EDU
- β_3 coglie in che modo l'effetto dell'istruzione dipenda (anche) dall'età e in che modo l'effetto dell'età dipenda (anche) dall'istruzione

interazione fra sesso ed età nei redditi svedesi

Call:

```
lm(formula = y_gross ~ sex + age + edu + age2 + age:edu, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-31637	-8173	-1528	6302	75214

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1694.349	7644.765	-0.222	0.824652
sexMale	1673.626	856.516	1.954	0.051037 .
age	847.697	217.519	3.897	0.000105 ***
edu	158.683	459.223	0.346	0.729771
age2	-9.817	1.805	-5.438	7.08e-08 ***
age:edu	16.798	8.303	2.023	0.043389 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12340 on 831 degrees of freedom
Multiple R-squared: 0.1423, Adjusted R-squared: 0.1371
F-statistic: 27.57 on 5 and 831 DF, p-value: < 2.2e-16