

# STATISTICA

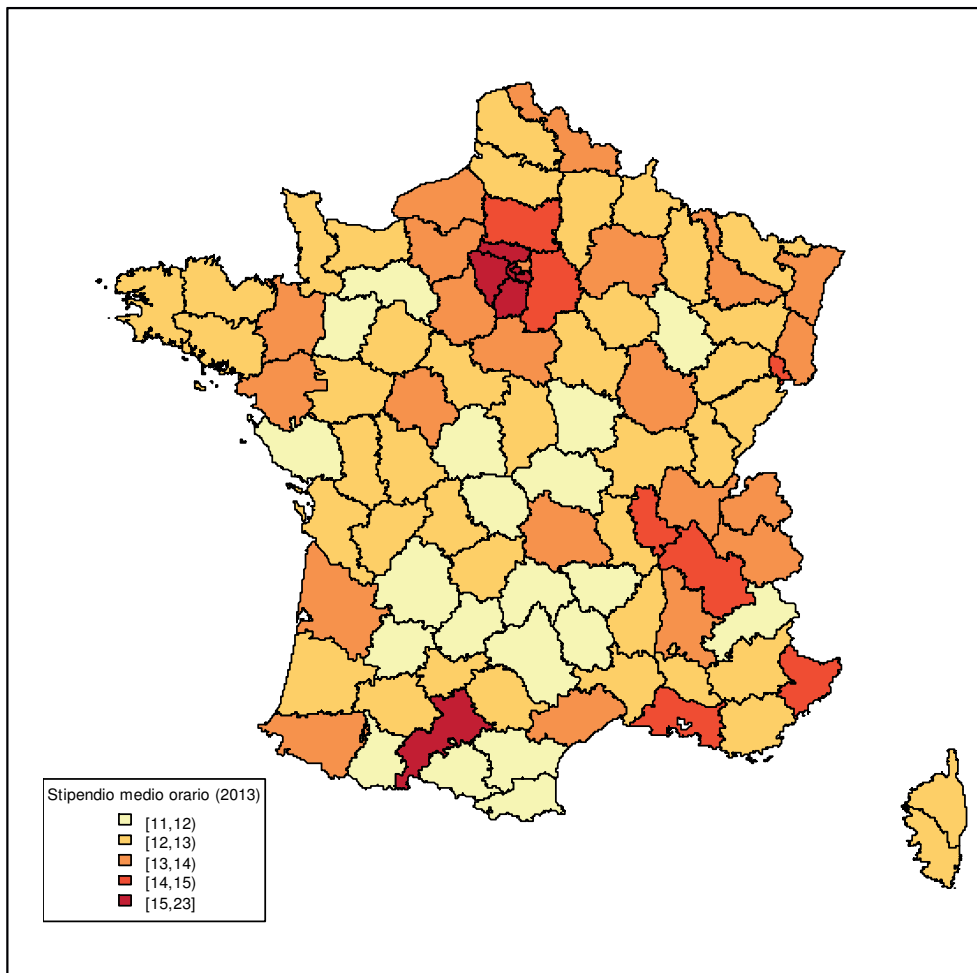
---

Regressione-4

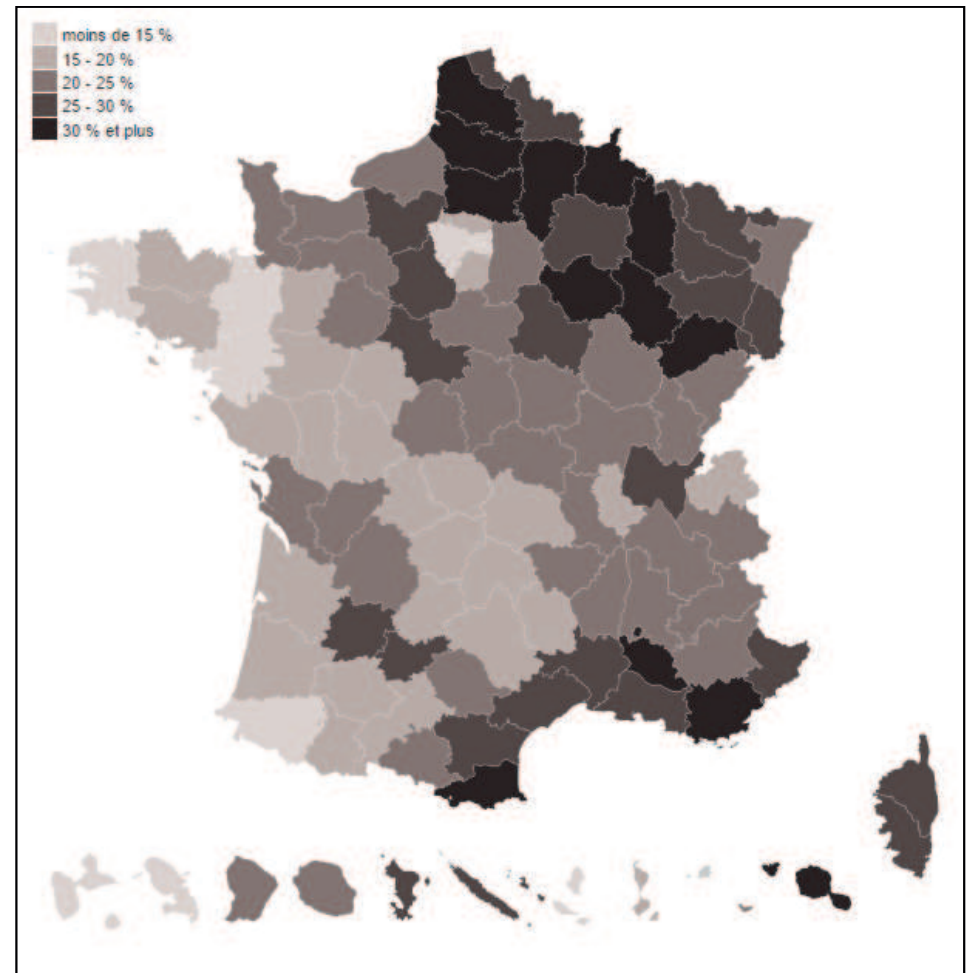
ovvero... Macron!

# Eravamo partiti da qui...

Stipendio medio orario 2013



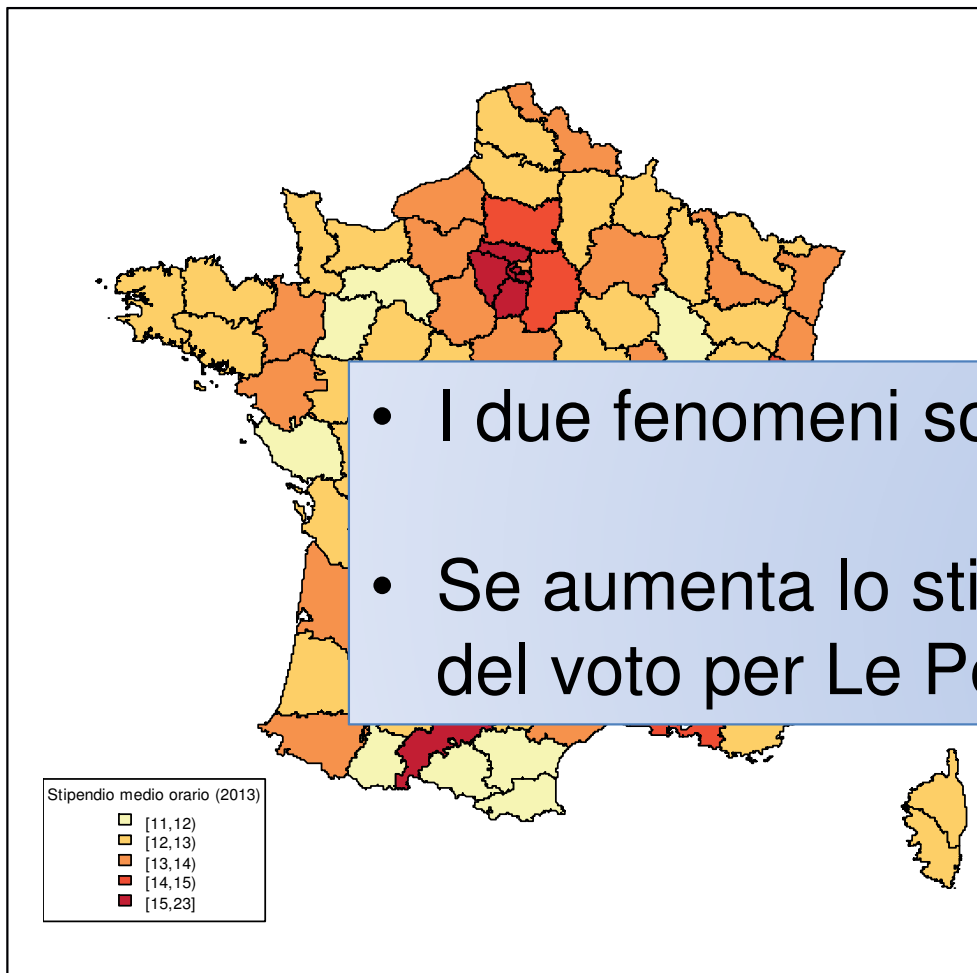
Voto per Le Pen



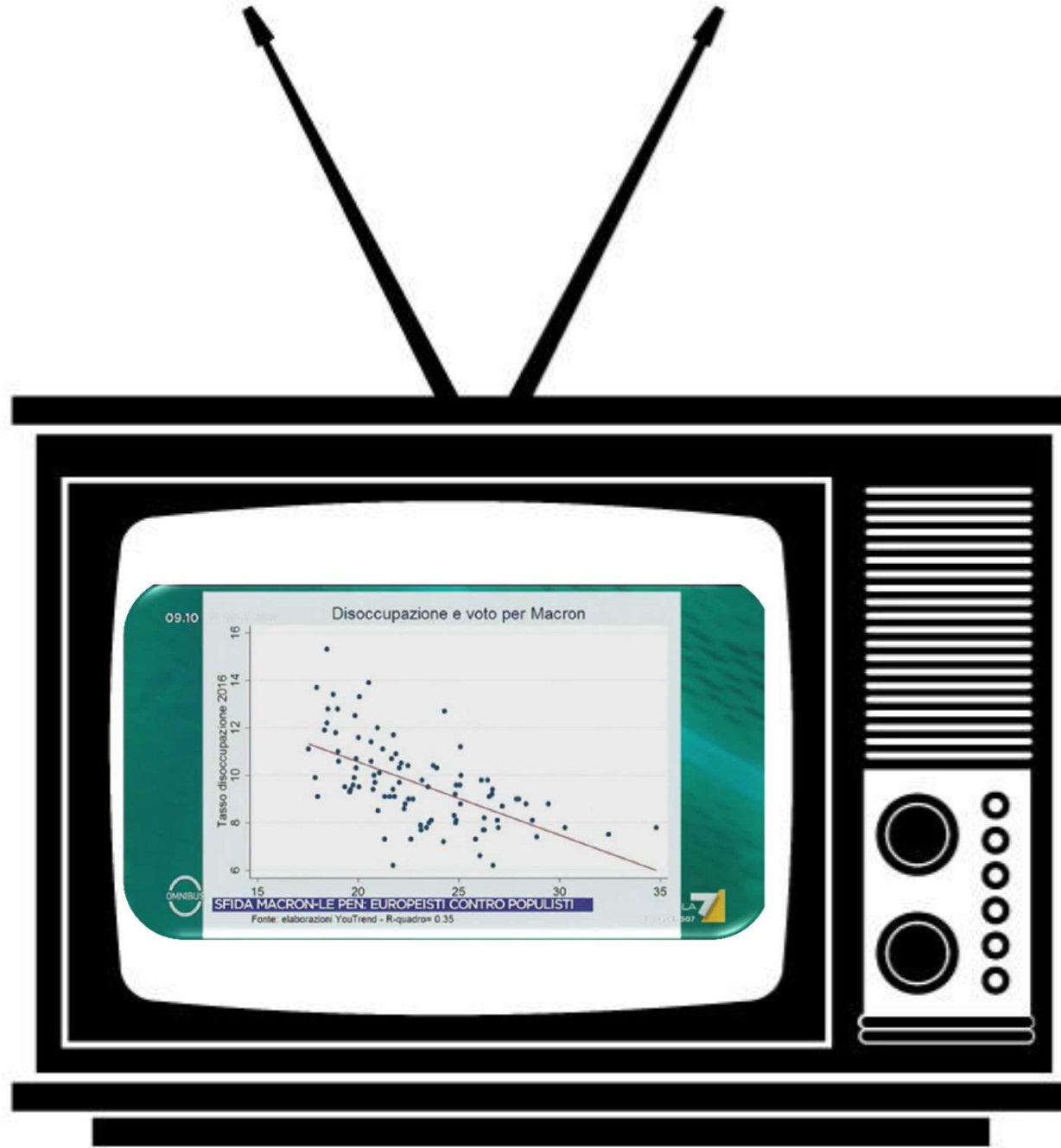
# Eravamo partiti da qui...

Stipendio medio orario 2013

Voto per Le Pen



- I due fenomeni sono collegati?
- Se aumenta lo stipendio, che ne è del voto per Le Pen?



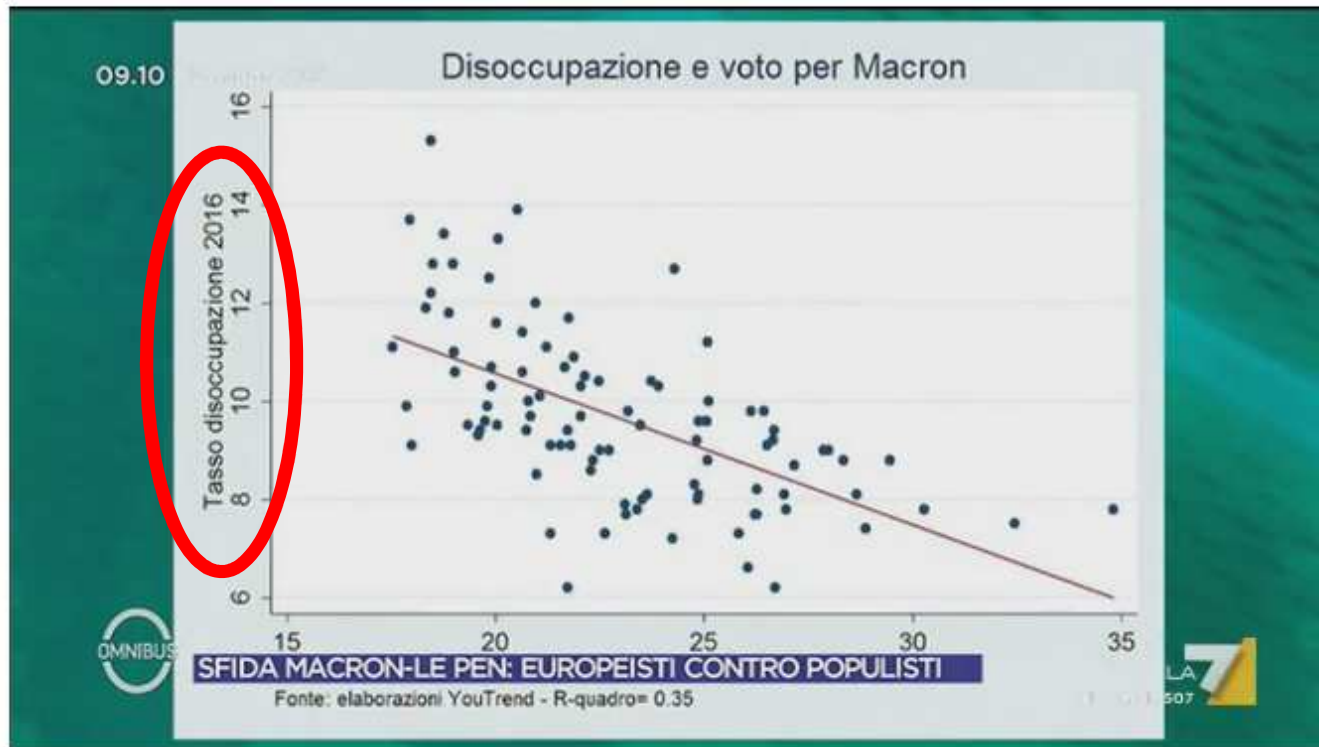


I TEMI DELLA POLITICA

RIVEDILA7

WEB SYSTEM 24

## Francia 2017, chi ha votato per Macron? L'analisi di Lorenzo Pregliasco (YouTrend)



24/04/2017



EMBED

Lorenzo Pregliasco (YouTrend) analizza il voto presidenziale francese. Macron ha ottenuto maggiori consensi a Parigi e tra chi ha un reddito e una posizione stabile

La tua opinione c  
La tua partecipazione  
anonima e le risposte  
la massima riservatezza  
nostro impegno a tutel  
Grazie!

Pa

Informati

RICETTE

CURIOSI



Idee e sugger  
giorni di festa

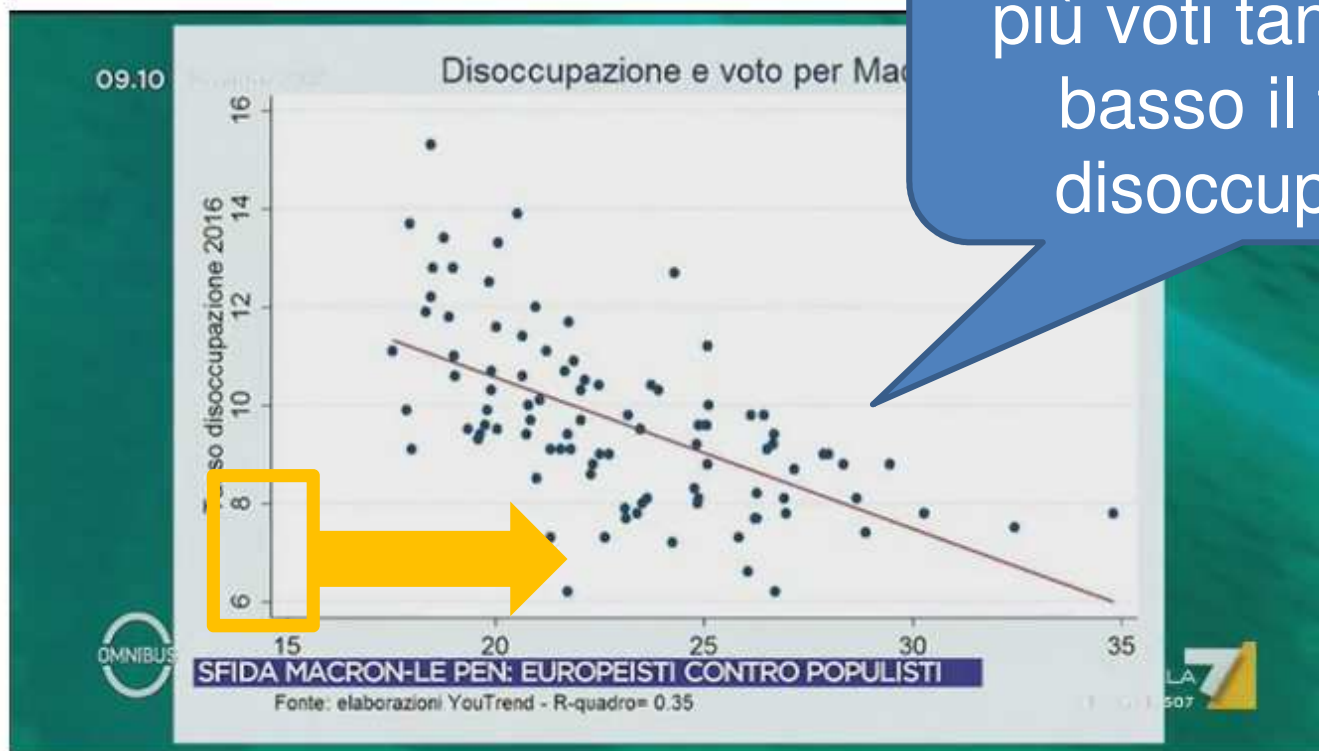


I TEMI DELLA POLITICA

RIVEDILA7

WEB SYSTEM 24

## Francia 2017, chi ha votato per Macron? L'analisi di Lorenzo Pregliasco (YouTrend)



“Macron ha avuto tanti più voti tanto più era basso il tasso di disoccupazione”

24/04/2017

in G+ Twitter f </> EMBED

Lorenzo Pregliasco (YouTrend) analizza il voto presidenziale francese. Macron ha ottenuto maggiori consensi a Parigi e tra chi ha un reddito e una posizione stabile

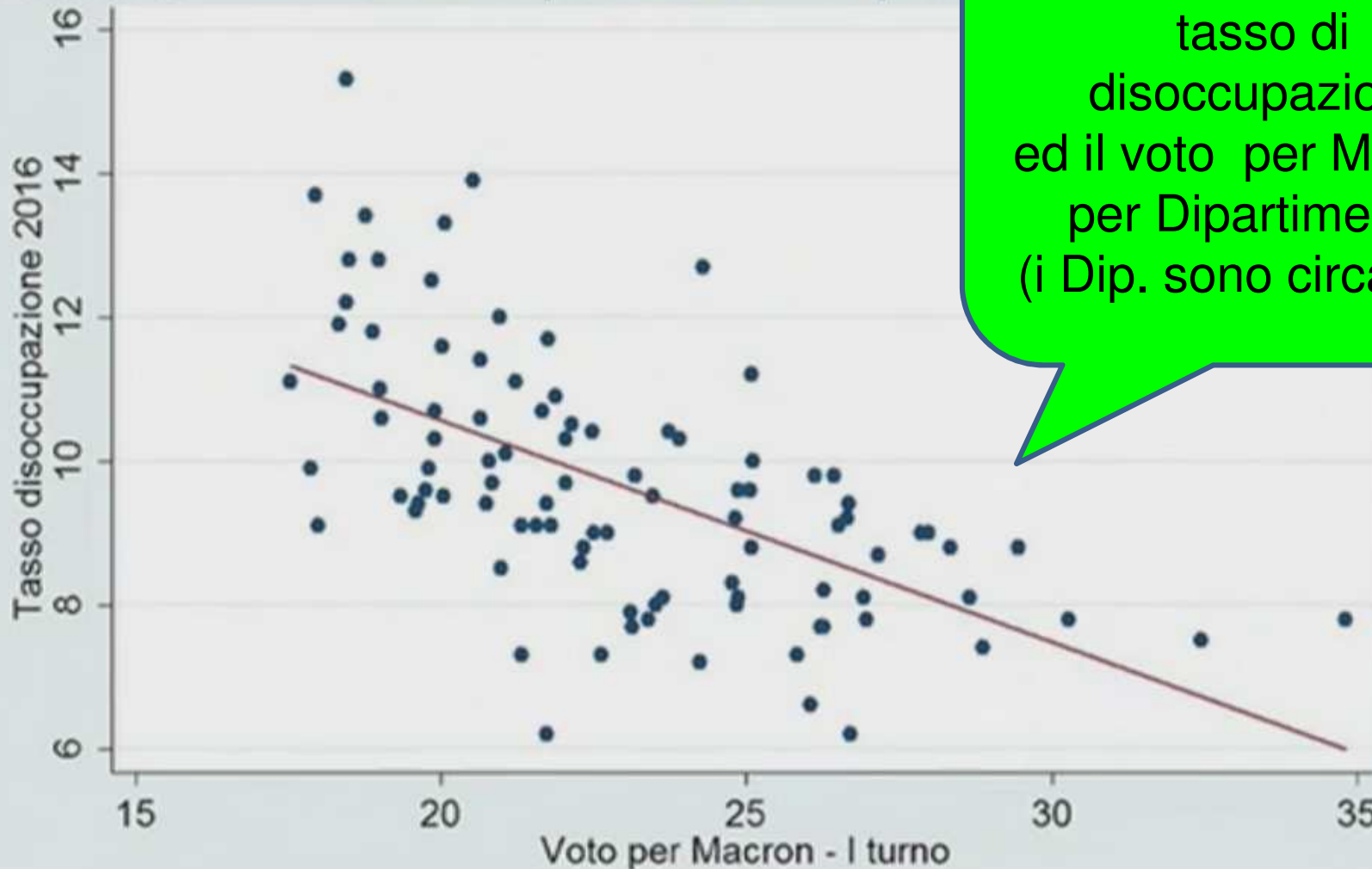
RICETTE CURIOSITÀ



Idee e suggerimenti per i giorni di festa

09.10

## Disoccupazione e voto per



“La **correlazione** tra il tasso di disoccupazione ed il voto per Macron, per Dipartimento”  
(i Dip. sono circa 100)



Fonte: elaborazioni YouTrend - R-quadro= 0.35



09.10

## Disoccupazione e voto per



“La **correlazione** tra il tasso di disoccupazione ed il voto per Macron, per Dipartimento”

$$R^2 = 0.35 \Leftrightarrow \rho = 0.59$$





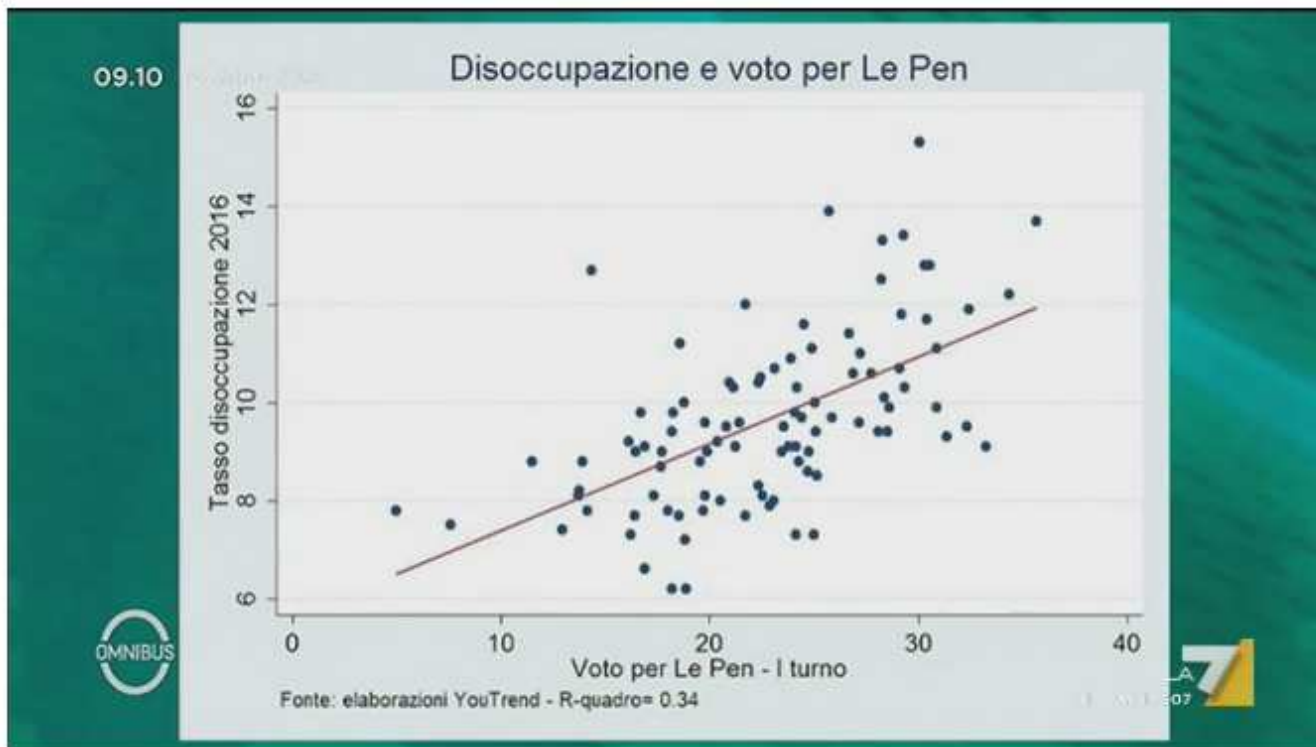


I TEMI DELLA POLITICA

RIVEDILA7

WEB SYSTEM

## Francia 2017, chi ha votato per Macron? L'analisi di Lorenzo Pregliasco (YouTrend)



24/04/2017



</> EMBED

Lorenzo Pregliasco (YouTrend) analizza il voto presidenziale francese. Macron ha ottenuto maggiori consensi a Parigi e tra chi ha un reddito e una posizione stabile

La tua opinio

Ti preghiamo di a  
online più interes

Ti chiediamo qua  
rispondere a alcu

[Info](#)

RICETTE

CUR



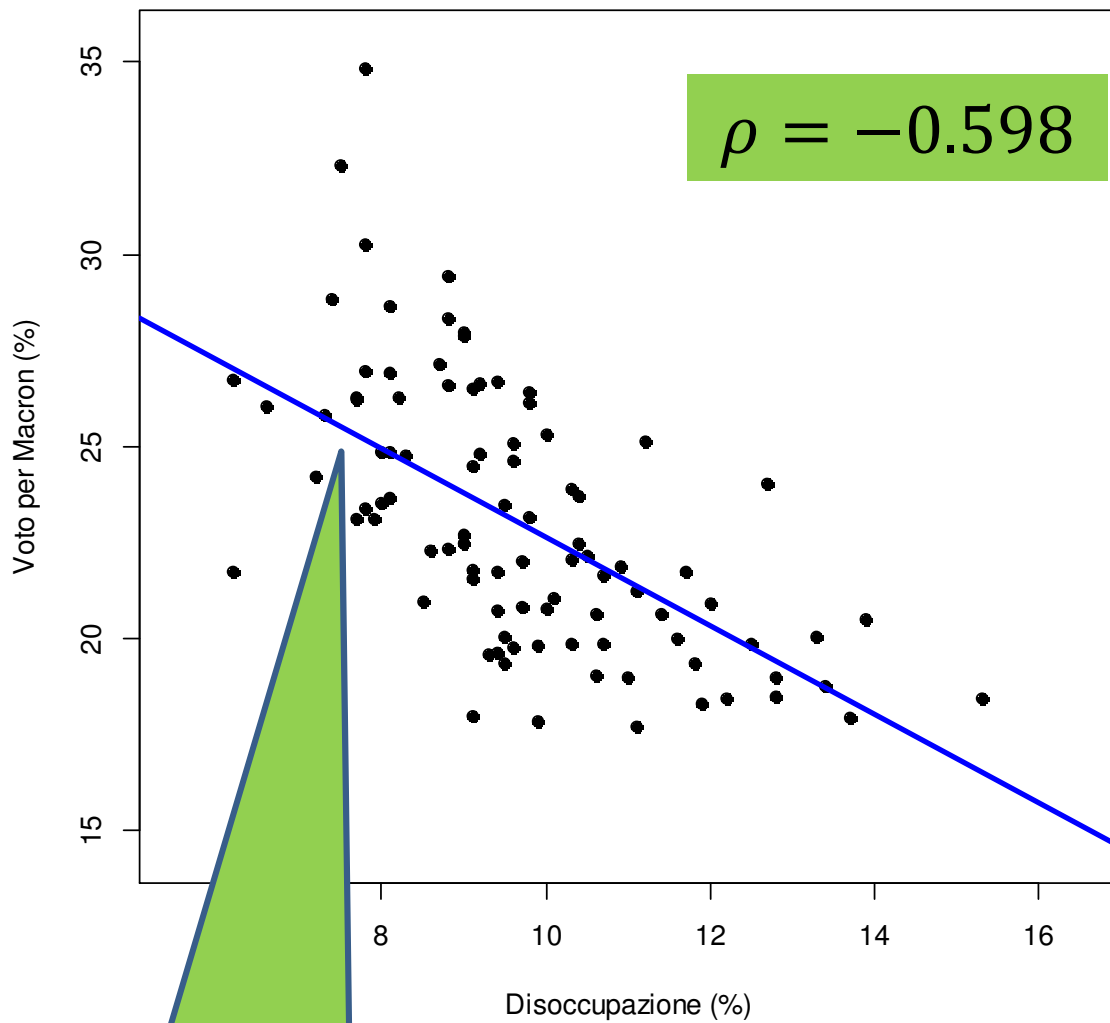
Idee e sug  
giorni di fe

“Nei Dipartimenti dove la disoccupazione è sopra il 12% Marine Le Pen va intorno al 30% dei voti.”

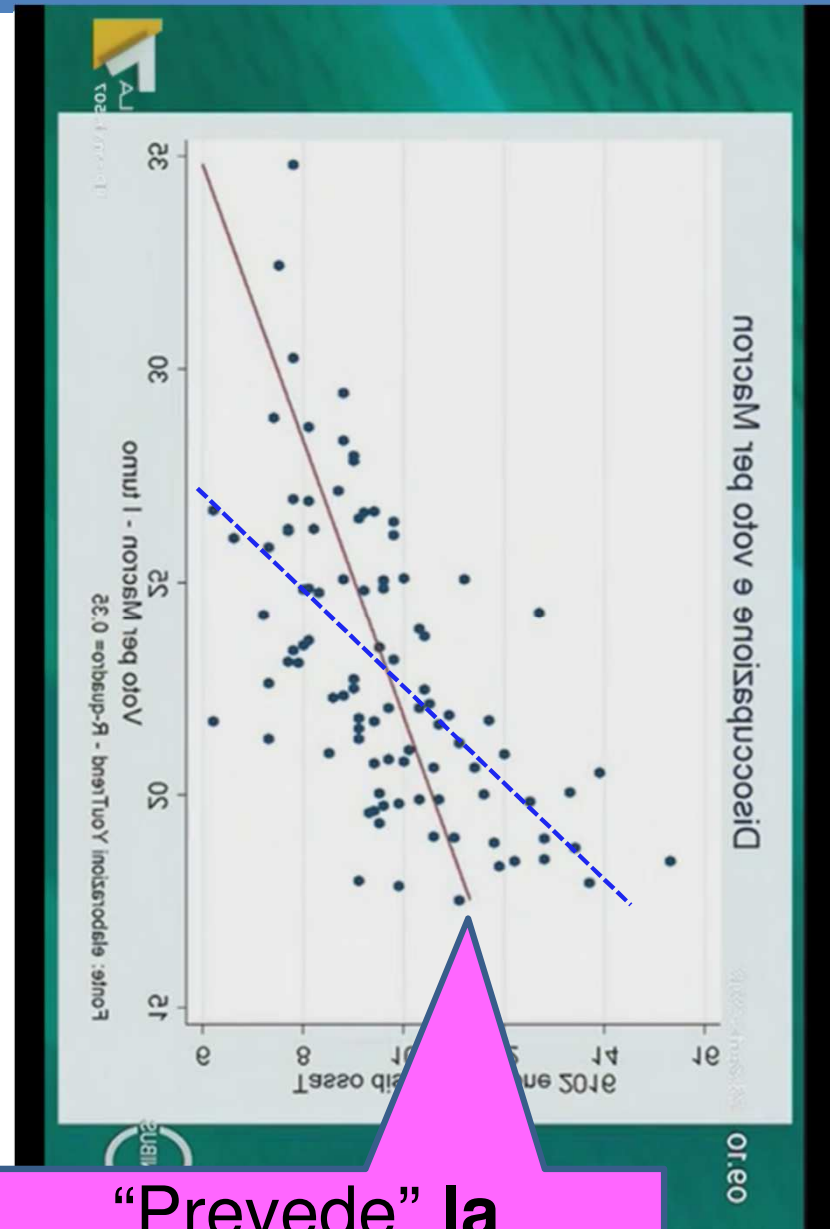


LA 7  
11/05/2017 10:07

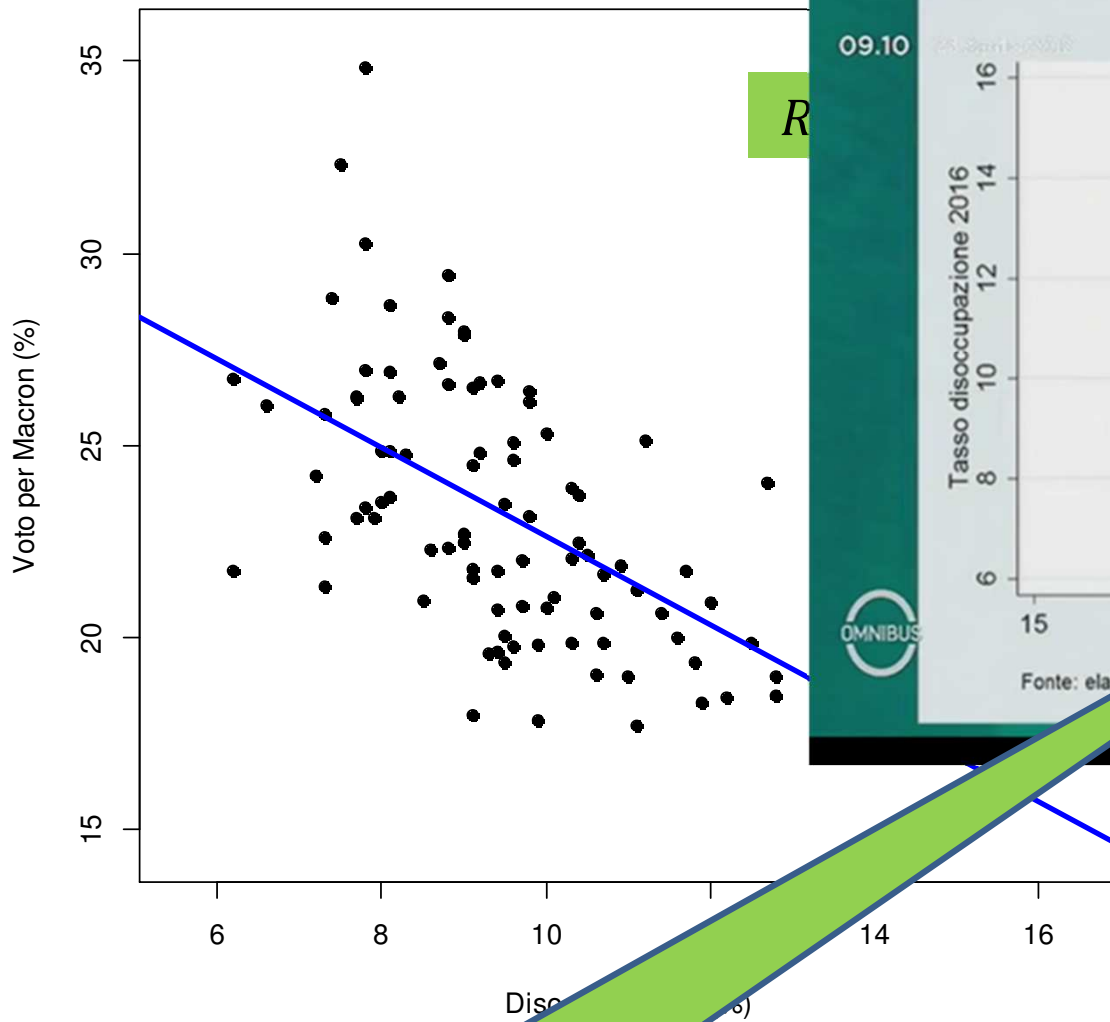




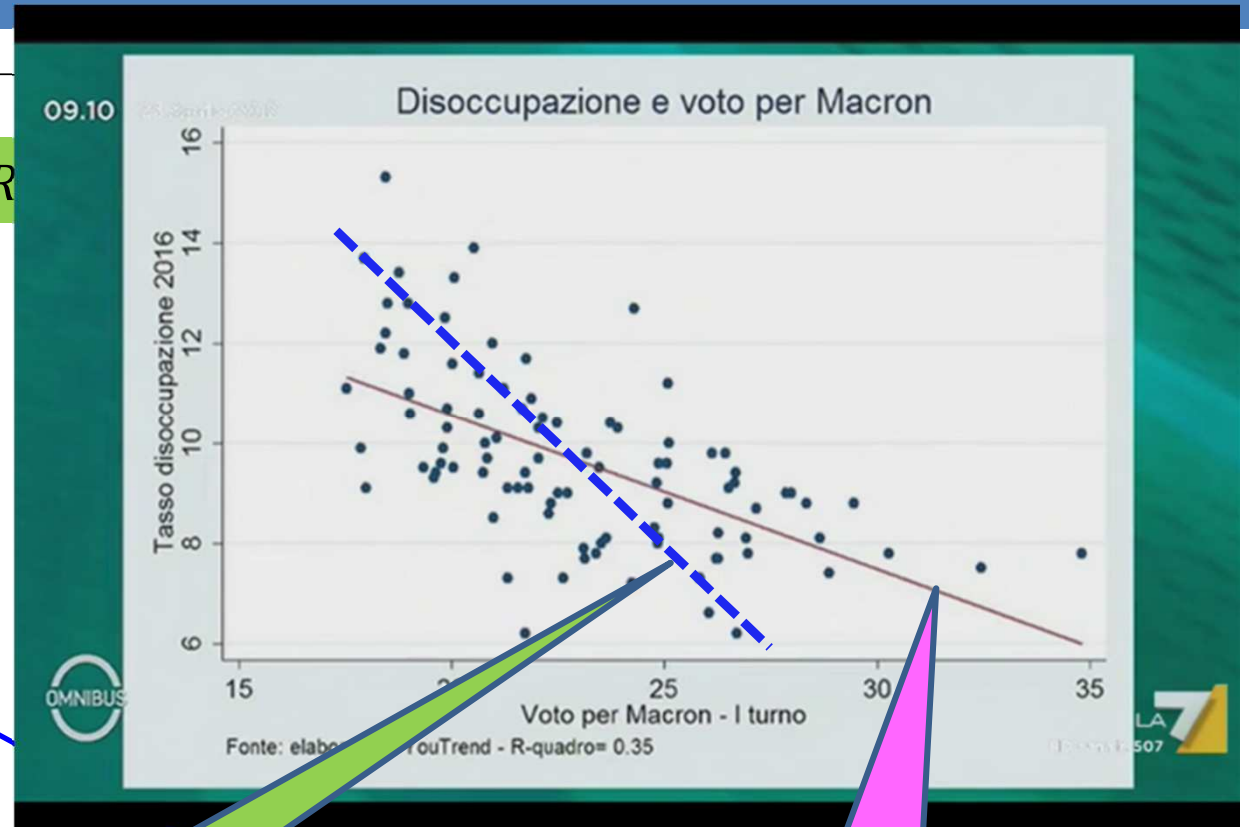
“Prevede” il voto in funzione del tasso di disoccupazione



“Prevede” la disoccupazione in funzione del voto...



R



“Prevede” il **voto** in funzione del tasso di disoccupazione

“Prevede” la **disoccupazione** in funzione del voto...

# I dati

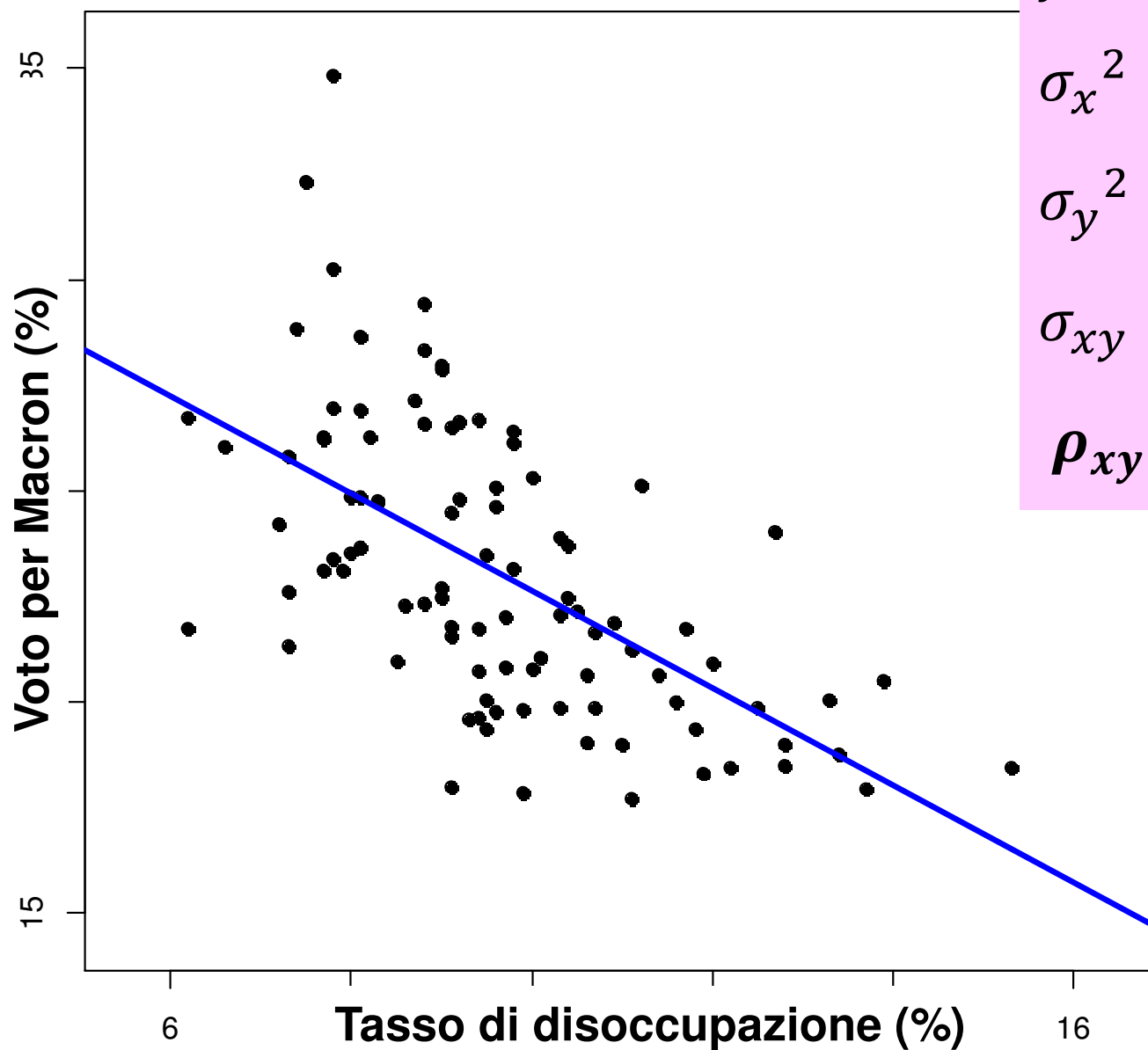
- Il voto, per dipartimenti:

<http://www.la-croix.com/France/Politique/Election-presidentielle-2017-carte-resultats-2017-04-23-1200841661>

- Tasso di disoccupazione, 4° trimestre 2016:

[https://www.insee.fr/fr/statistiques/2012804#tableau-TCRD\\_025\\_tab1\\_departements](https://www.insee.fr/fr/statistiques/2012804#tableau-TCRD_025_tab1_departements)

# Analisi completa



$$\bar{x} = 9.65$$

$$\bar{y} = 23.05$$

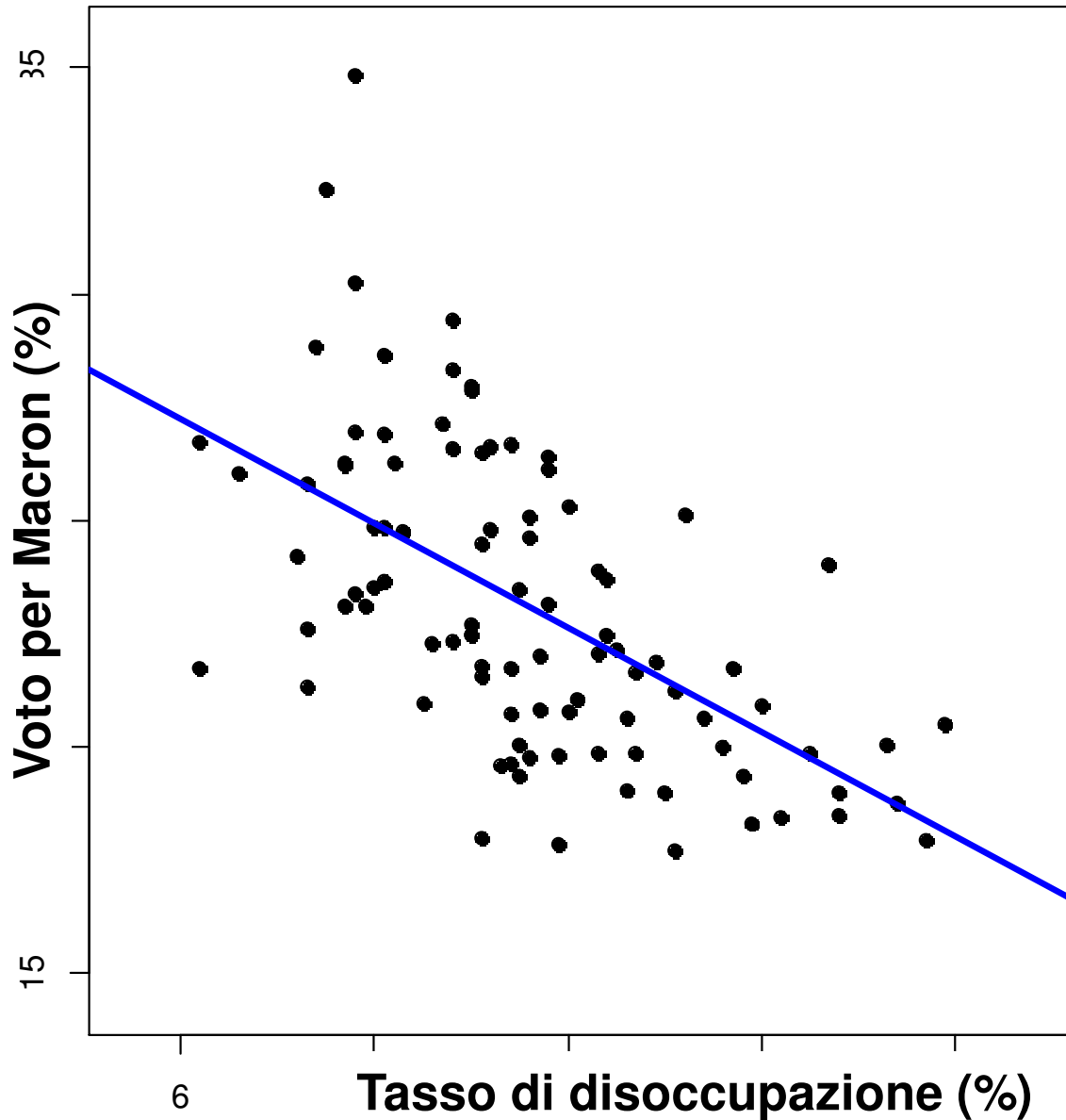
$$\sigma_x^2 = 3.17$$

$$\sigma_y^2 = 11.78$$

$$\sigma_{xy} = -3.73$$

$$\rho_{xy} = -0.61$$

# Analisi completa



$$\bar{x} = 9.65$$

$$\bar{y} = 23.05$$

$$\sigma_x^2 = 3.17$$

$$\sigma_y^2 = 11.78$$

$$\sigma_{xy} = -3.73$$

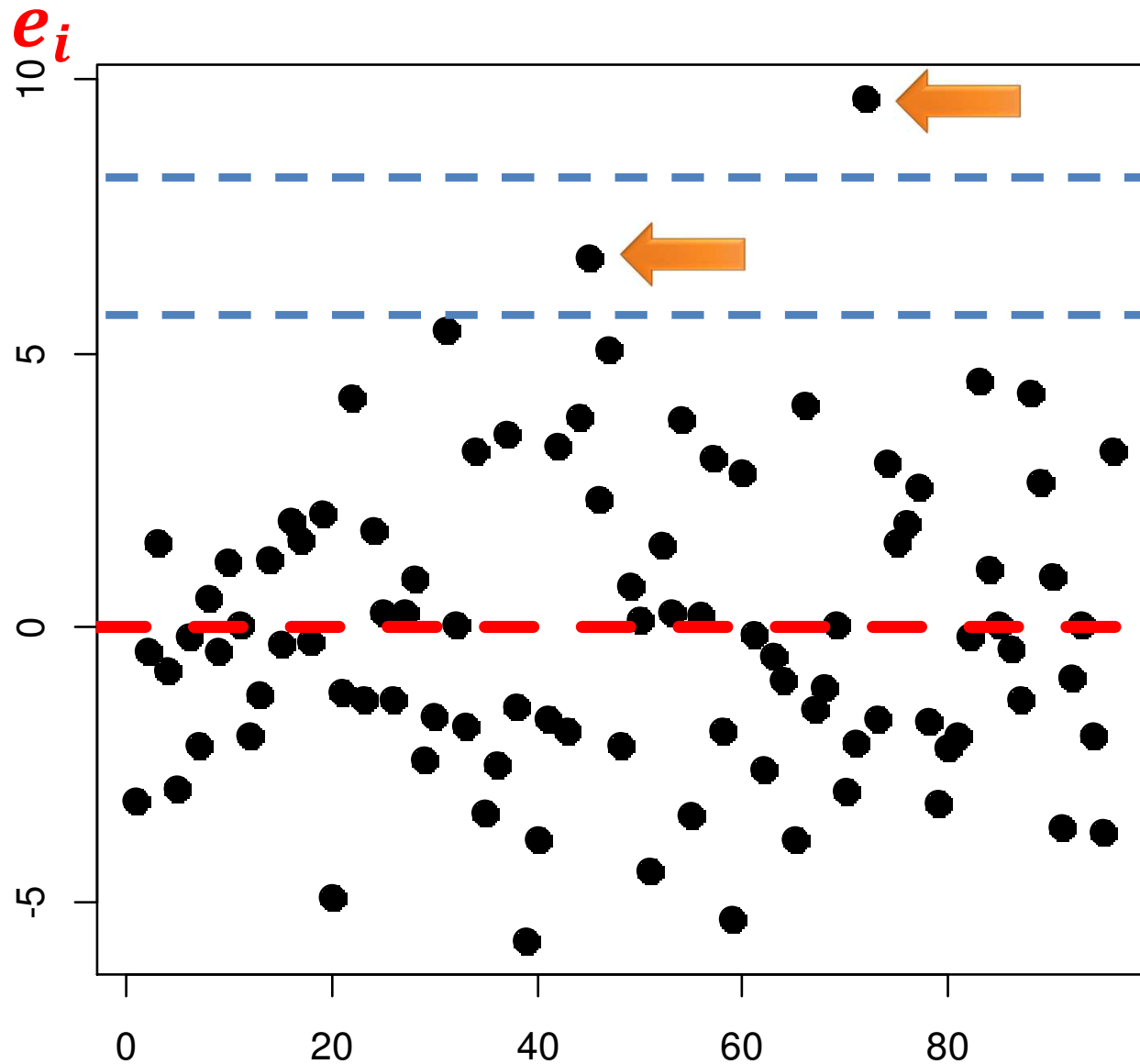
$$\rho_{xy} = -0.61$$

$$R^2 = (-0.61)^2 = 0.37$$

$$\hat{b} = \frac{-3.73}{3.17} = -1.18$$

$$\hat{a} = 23.05 + 1.18 \times 9.65 = 34.44$$

# Analisi completa



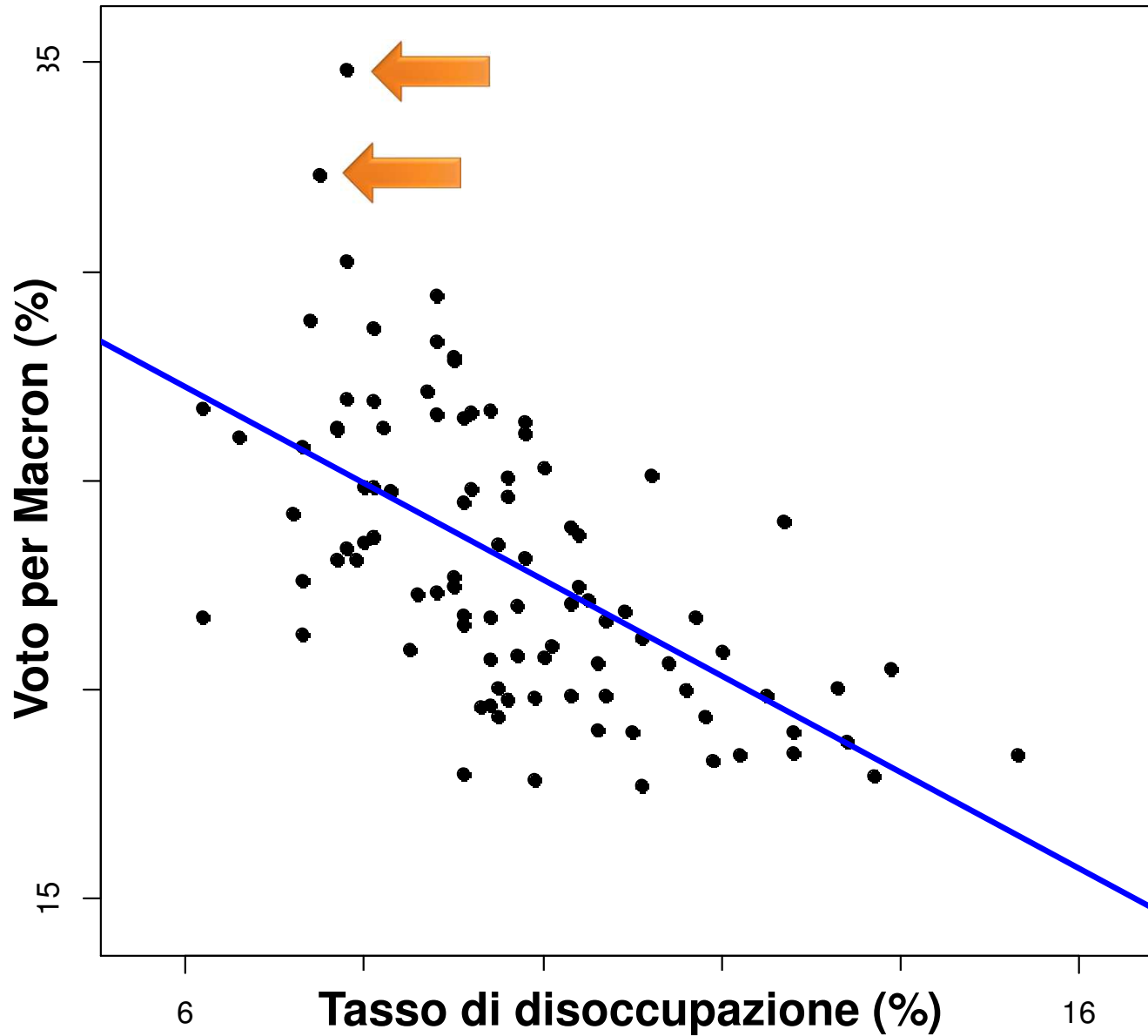
$$s^2 = 7.72$$

$$3s = 8.34$$

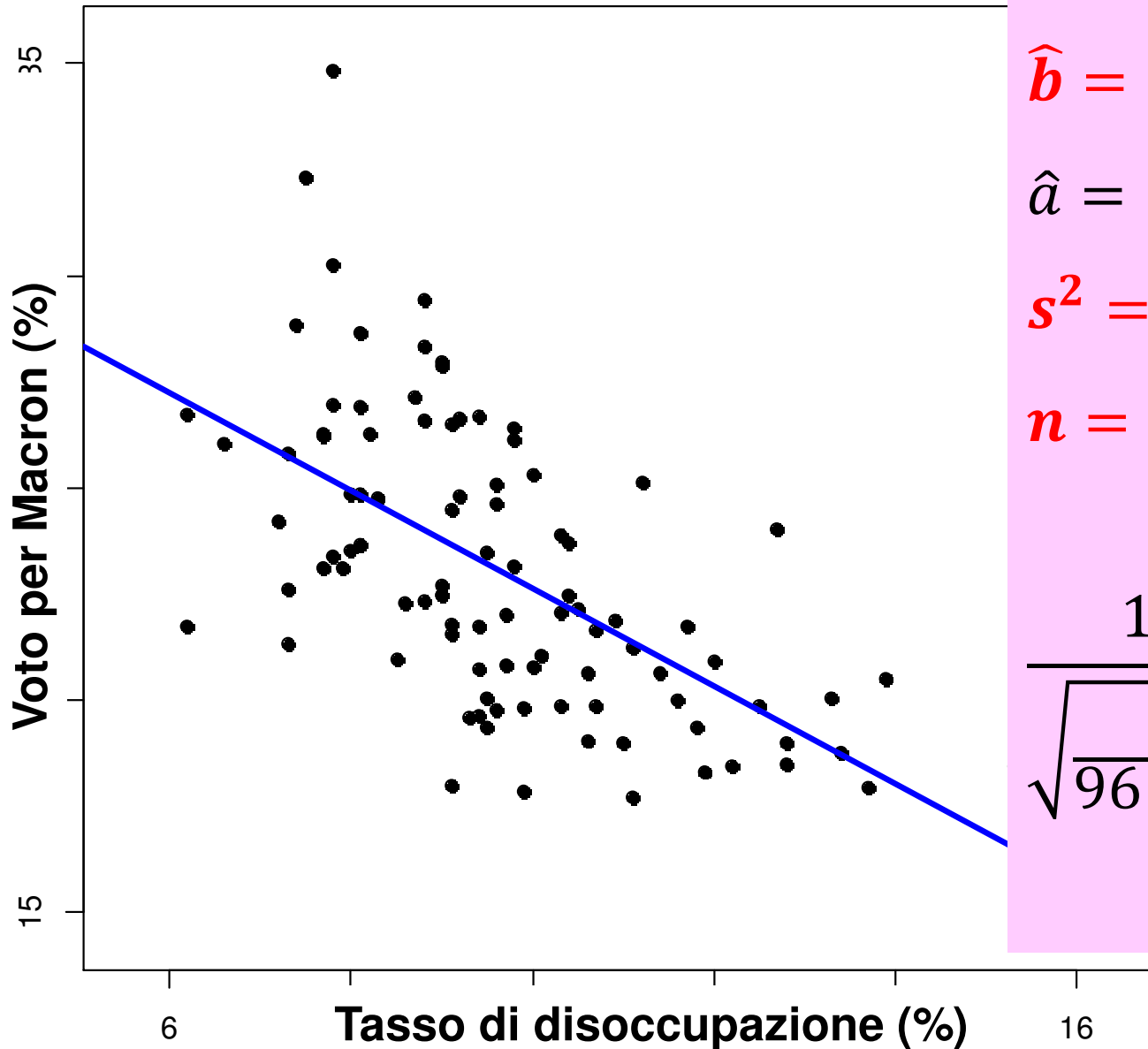
$$2s = 5.56$$



# Analisi completa



# Analisi completa



$$\sigma_x^2 = 3.17$$

$$R^2 = 0.37$$

$$\hat{b} = -1.18$$

$$\hat{a} = 34.44$$

$$s^2 = 7.72$$

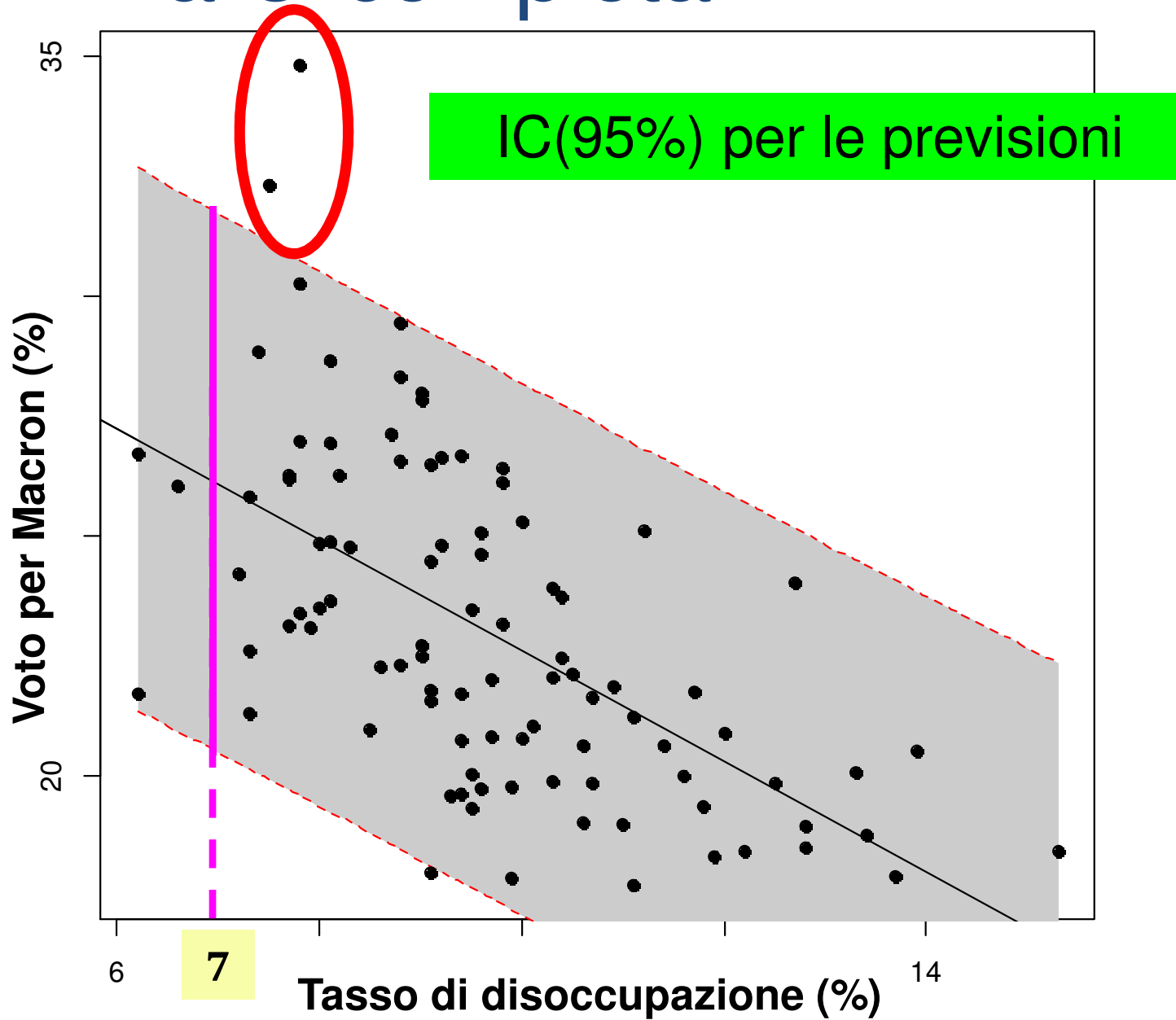
$$n = 96$$

$$H_0 : b = 0$$

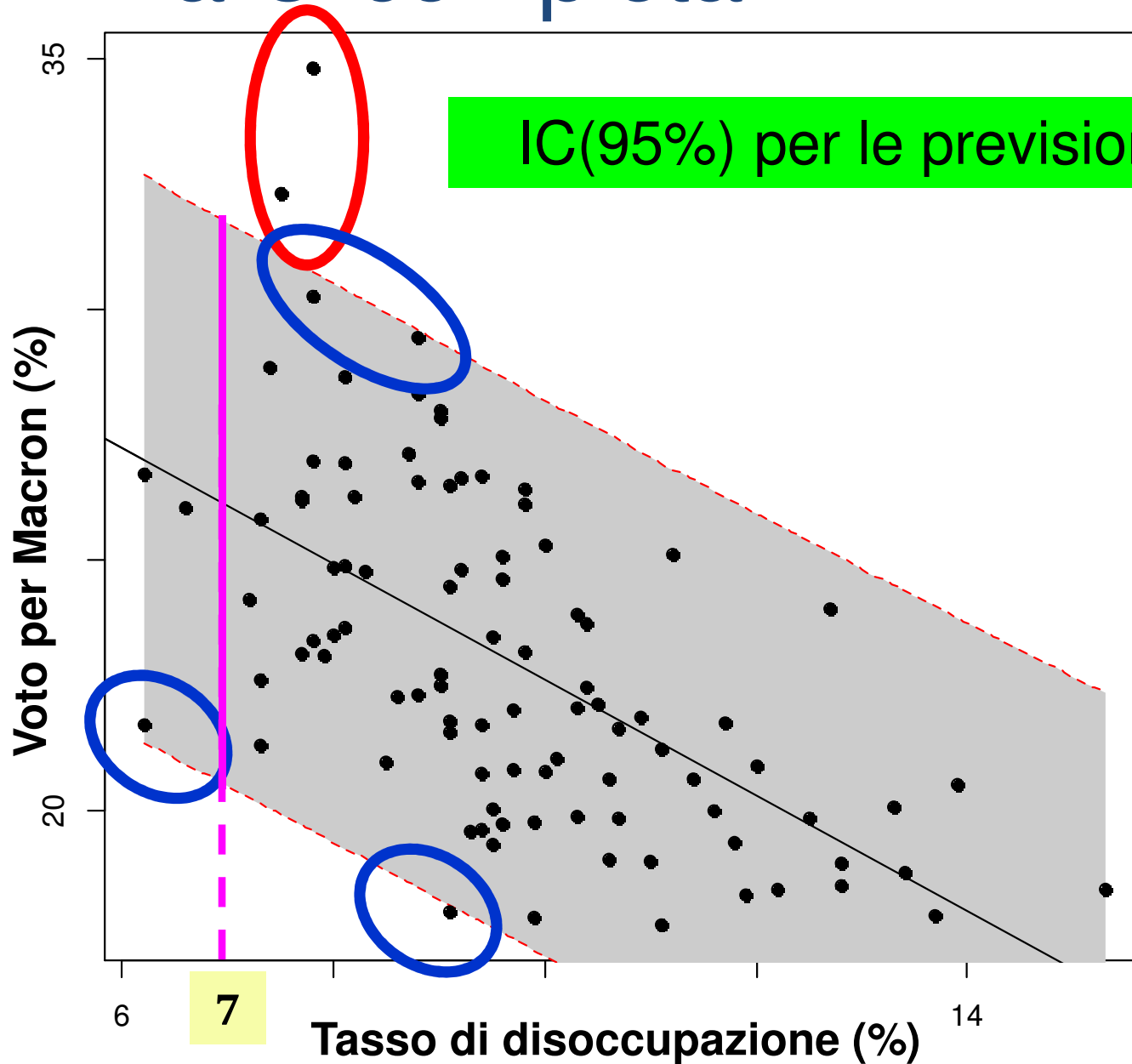
$$\frac{1.18}{\sqrt{\frac{7.72}{96 \times 3.17}}} = 7.41$$

$$> t(94)_{1-\frac{\alpha}{2}} \approx z_{1-\frac{\alpha}{2}} !!$$

# Analisi completa



# Analisi completa

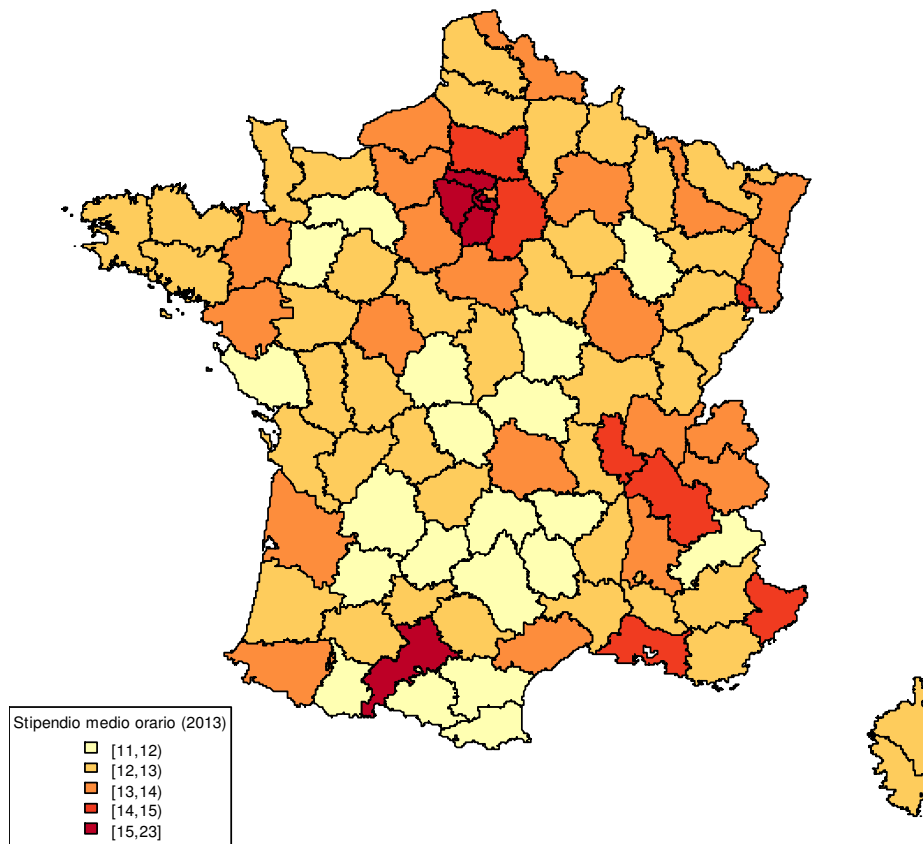


IC(95%) per le previsioni

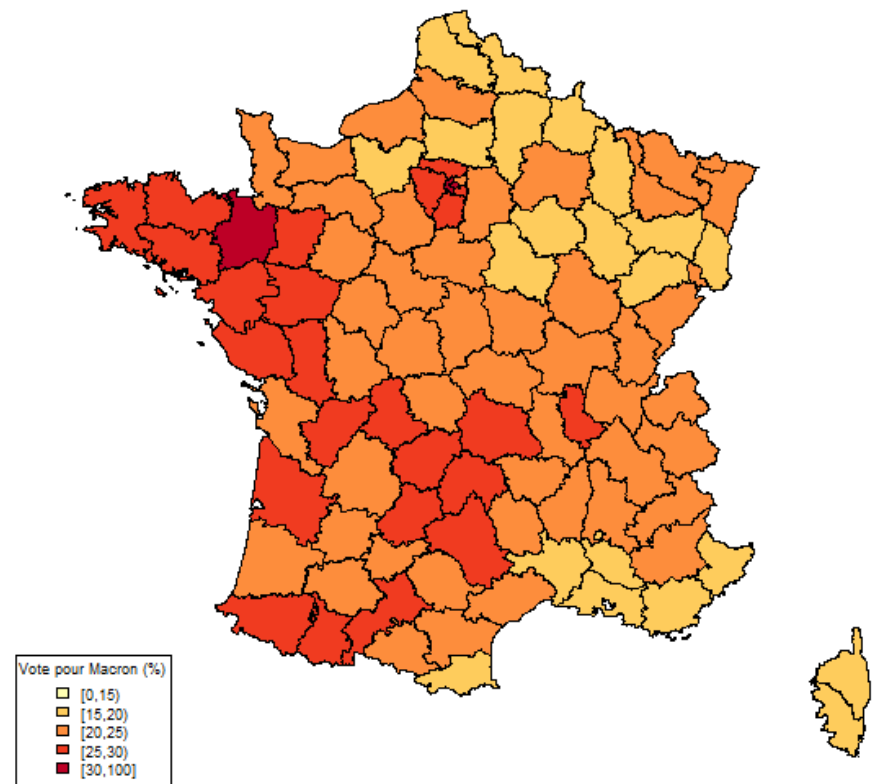
$$R^2 = 0.37 !!!!$$

# Macron e lo stipendio!

## Stipendio



## Voto per Macron



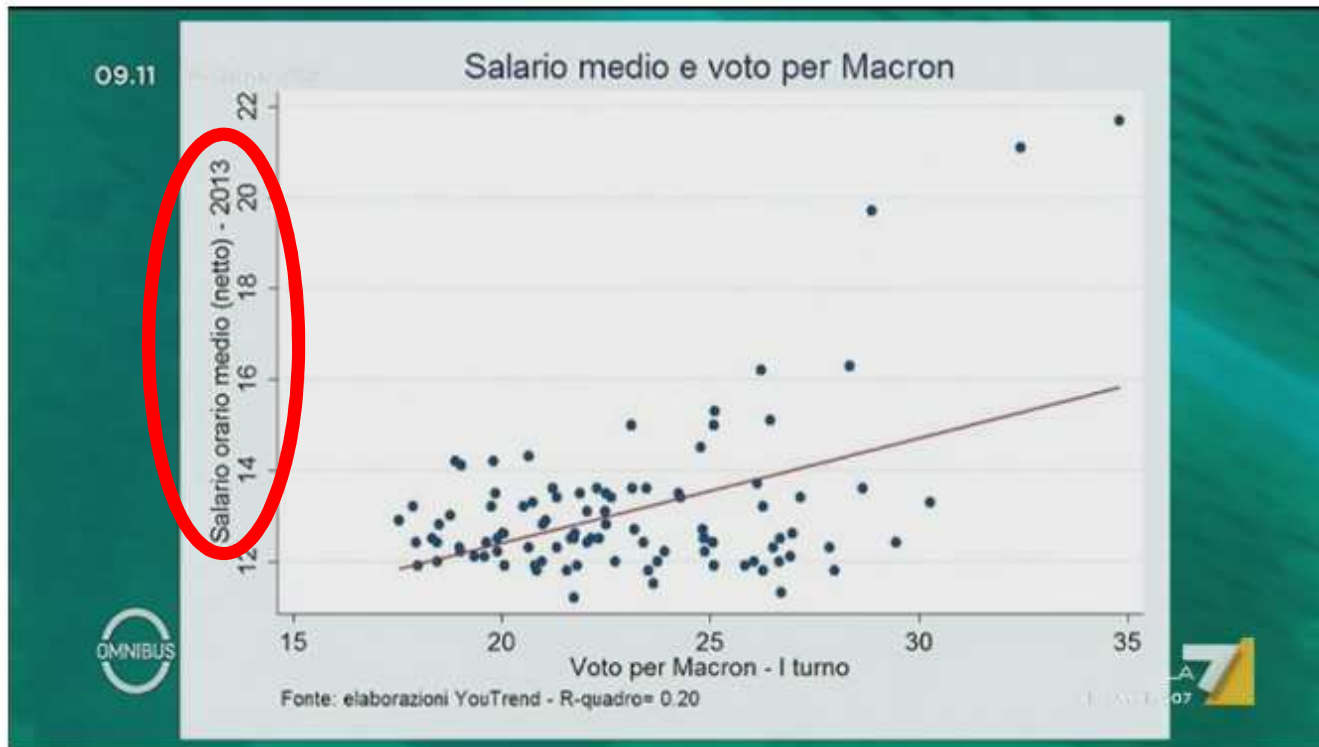


I TEMI DELLA POLITICA

RIVEDILA7

WEB SYSTEM

## Francia 2017, chi ha votato per Macron? L'analisi di Lorenzo Pregliasco (YouTrend)



24/04/2017



EMBED

Lorenzo Pregliasco (YouTrend) analizza il voto presidenziale francese. Macron ha ottenuto maggiori consensi a Parigi e tra chi ha un reddito e una posizione stabile

### La tua opinione

La tua partecipazione anonima e le risposte la massima riservatezza il nostro impegno a te. Grazie!

Informa

RICETTE

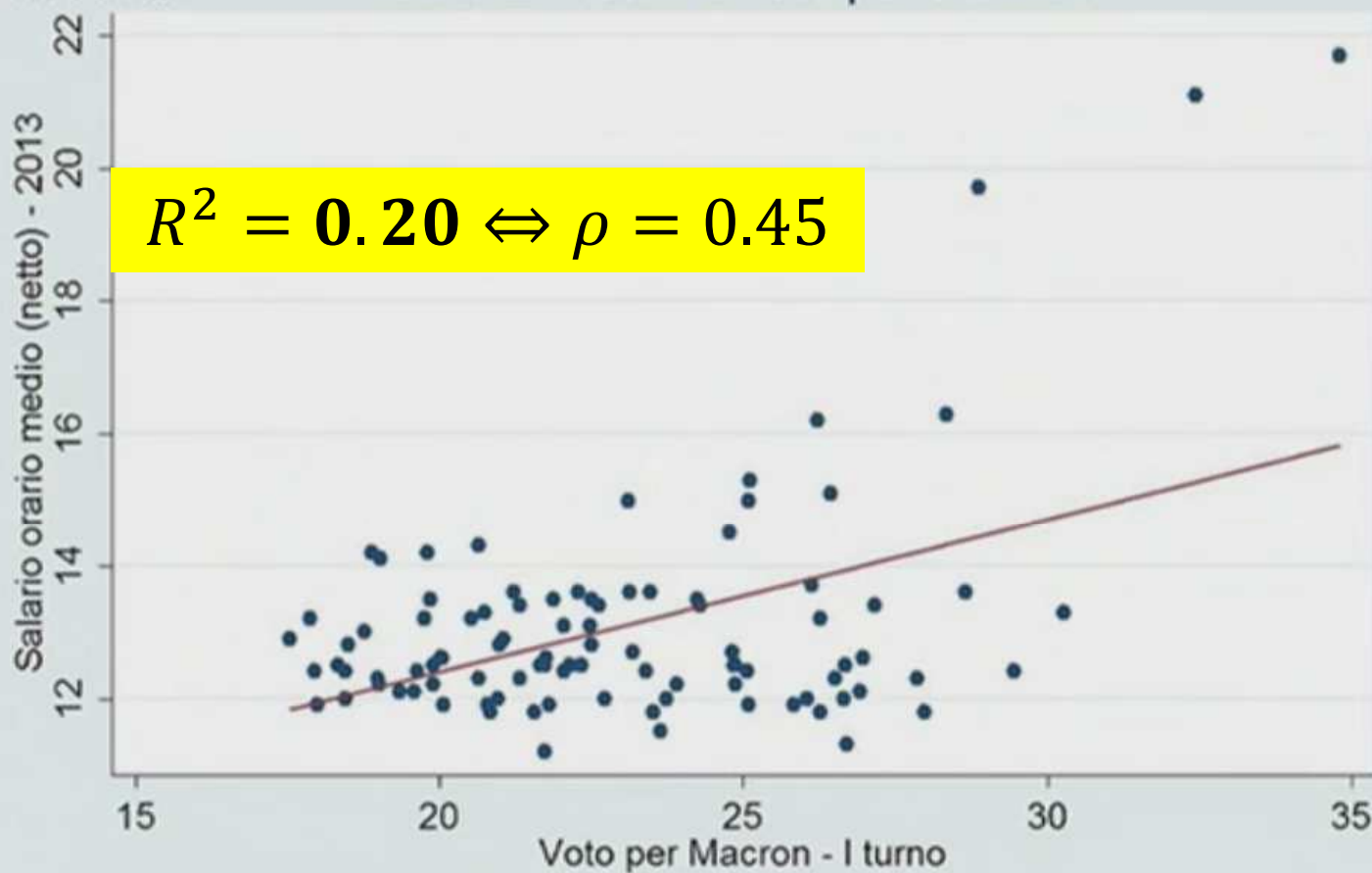
CURIO



Idee e suggerimenti per i giorni di festa

09.11

## Salario medio e voto per Macron



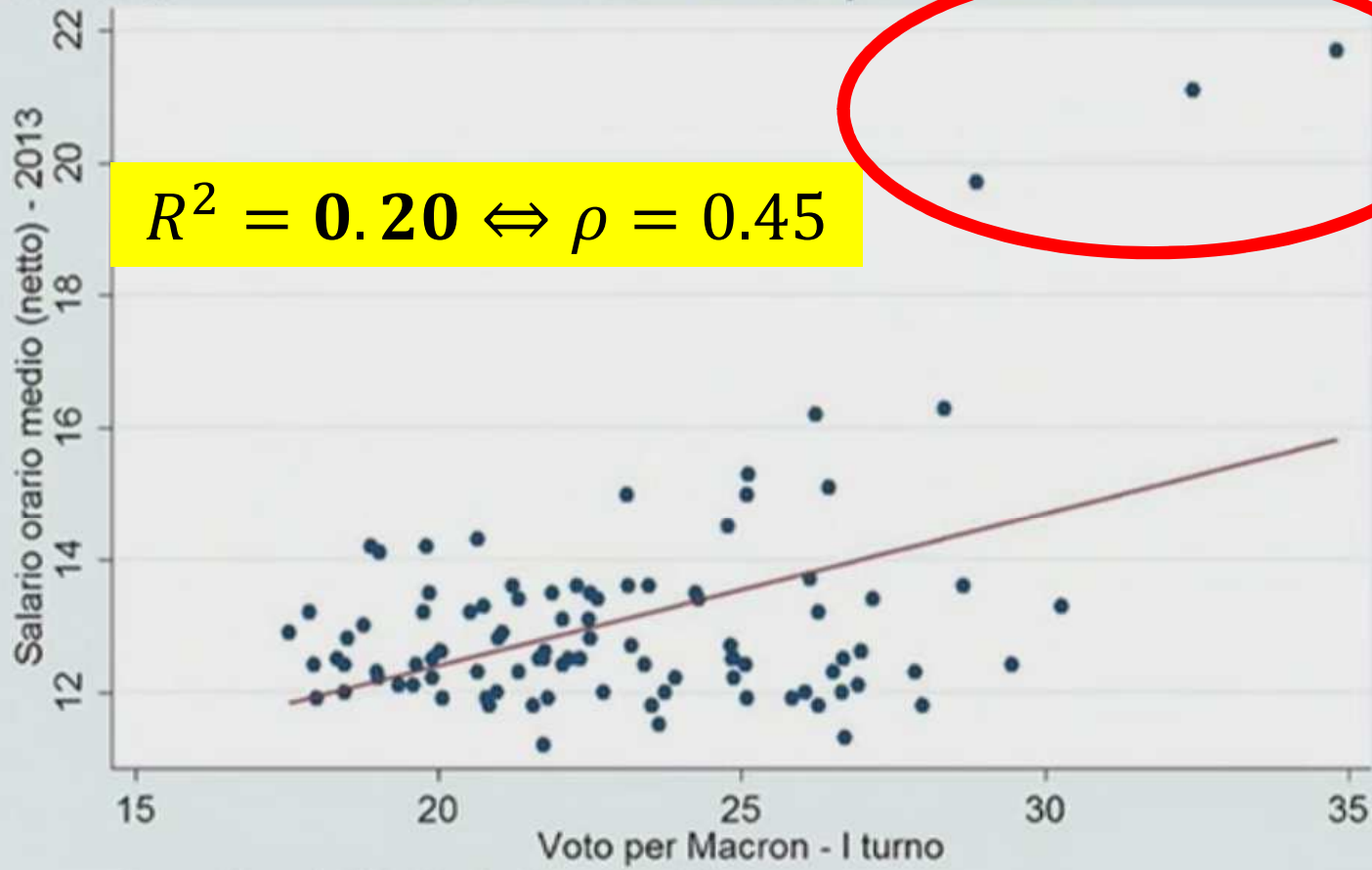
Fonte: elaborazioni YouTrend - R-quadro= 0.20



1:22 / 2:12

09.11

### Salario medio e voto per Macron



$R^2 = 0.20 \Leftrightarrow \rho = 0.45$



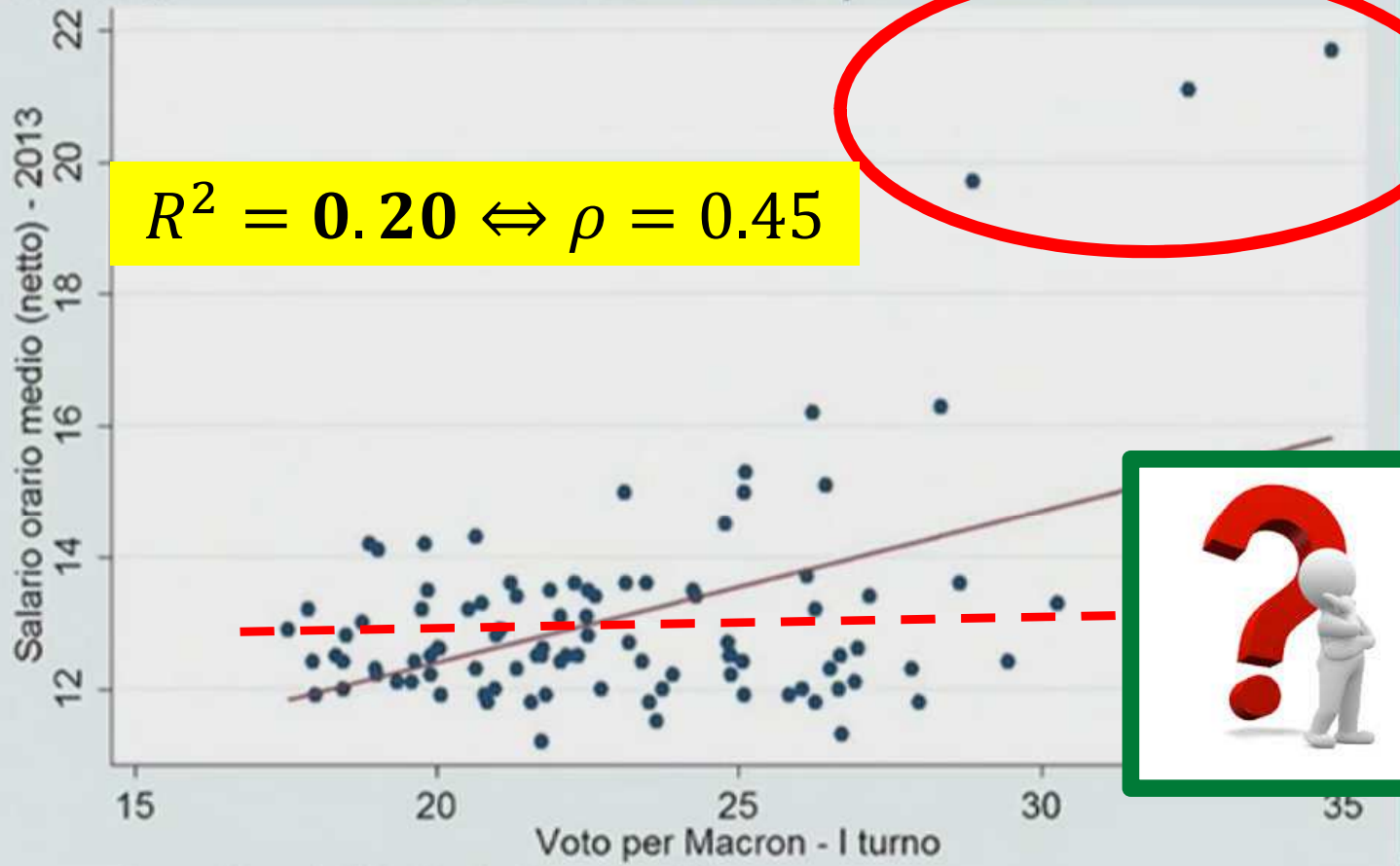
Fonte: elaborazioni YouTrend - R-quadro= 0.20





09.11

# Salario medio e voto per Macron



$R^2 = 0.20 \Leftrightarrow \rho = 0.45$

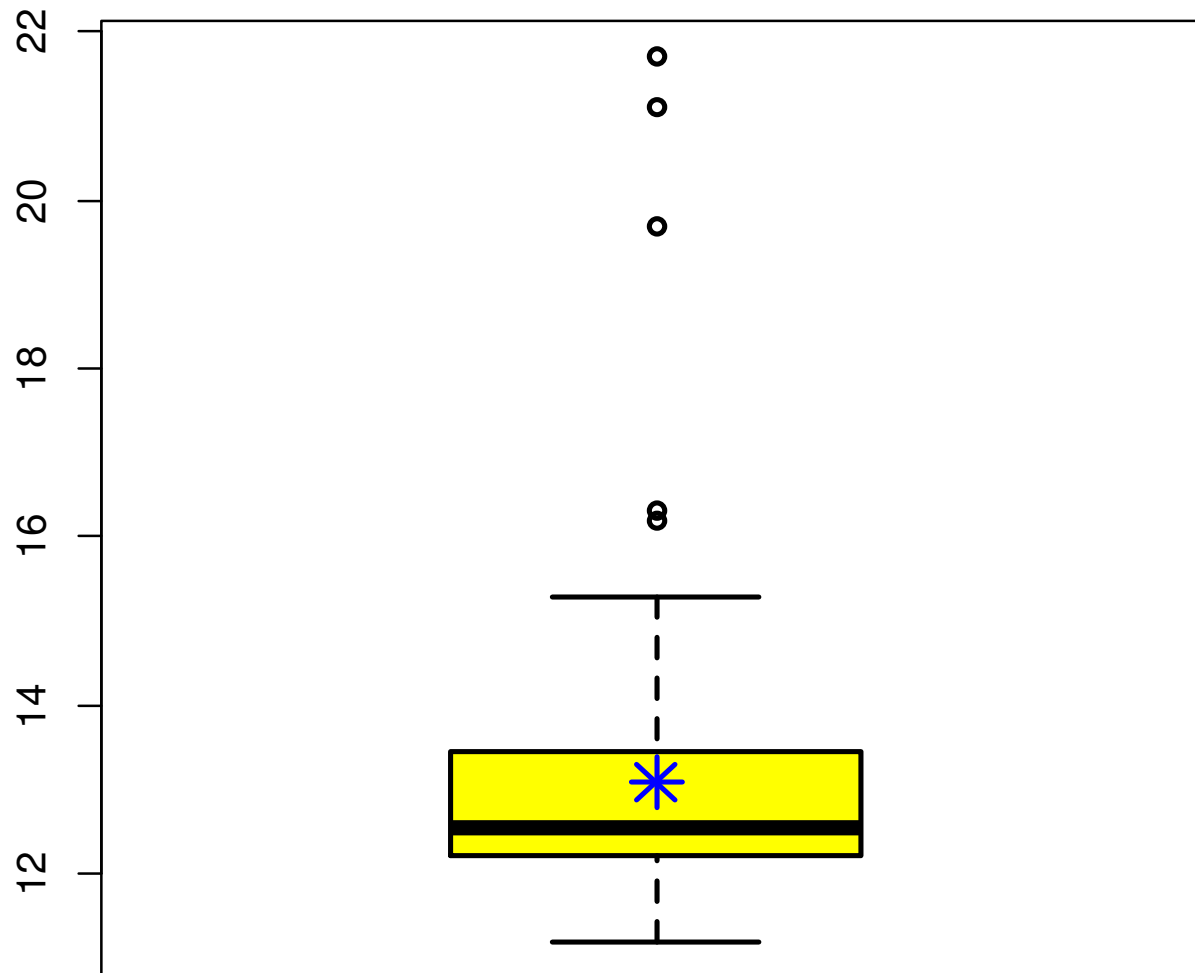


Fonte: elaborazioni YouTrend - R-quadro= 0.20

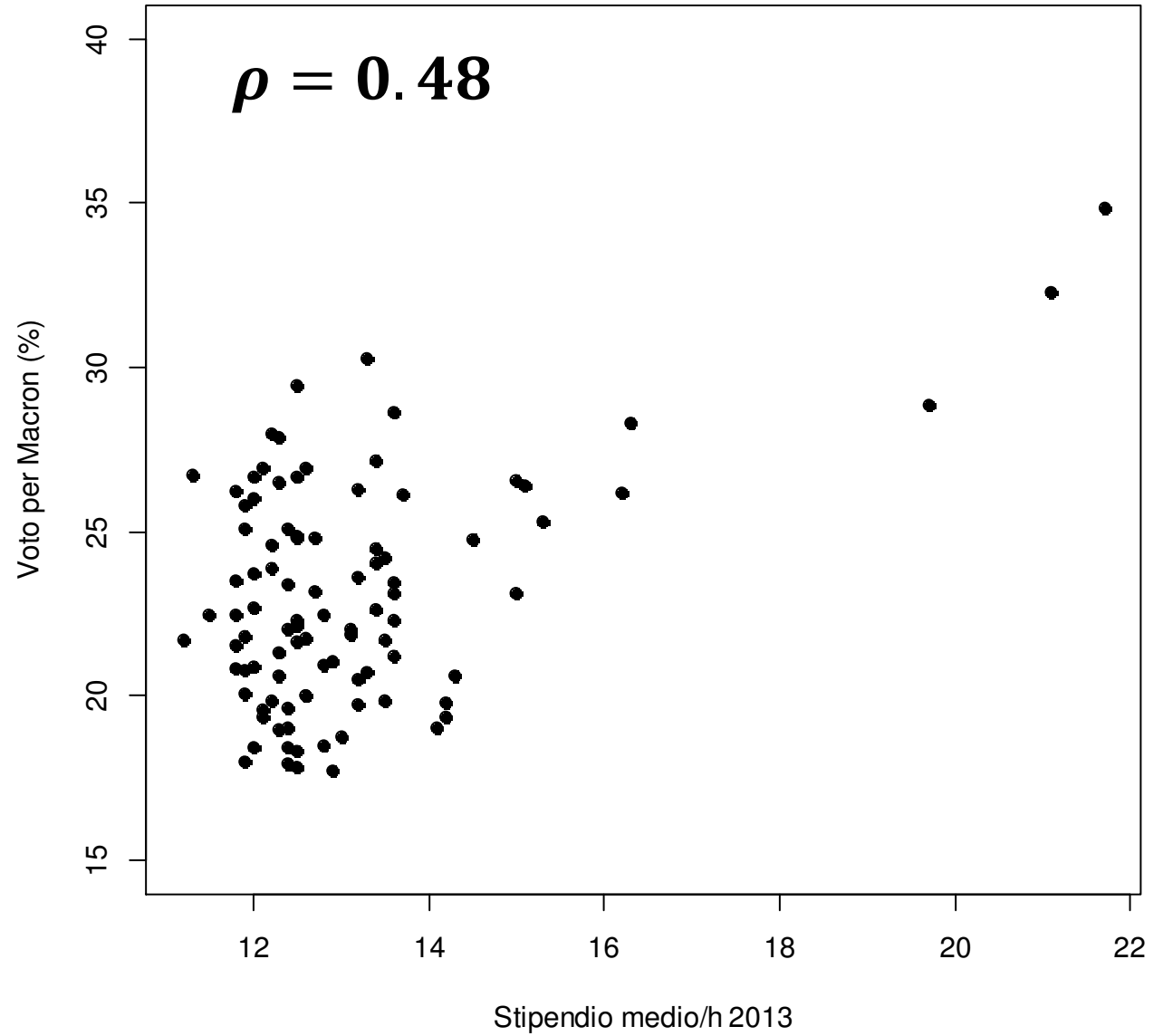
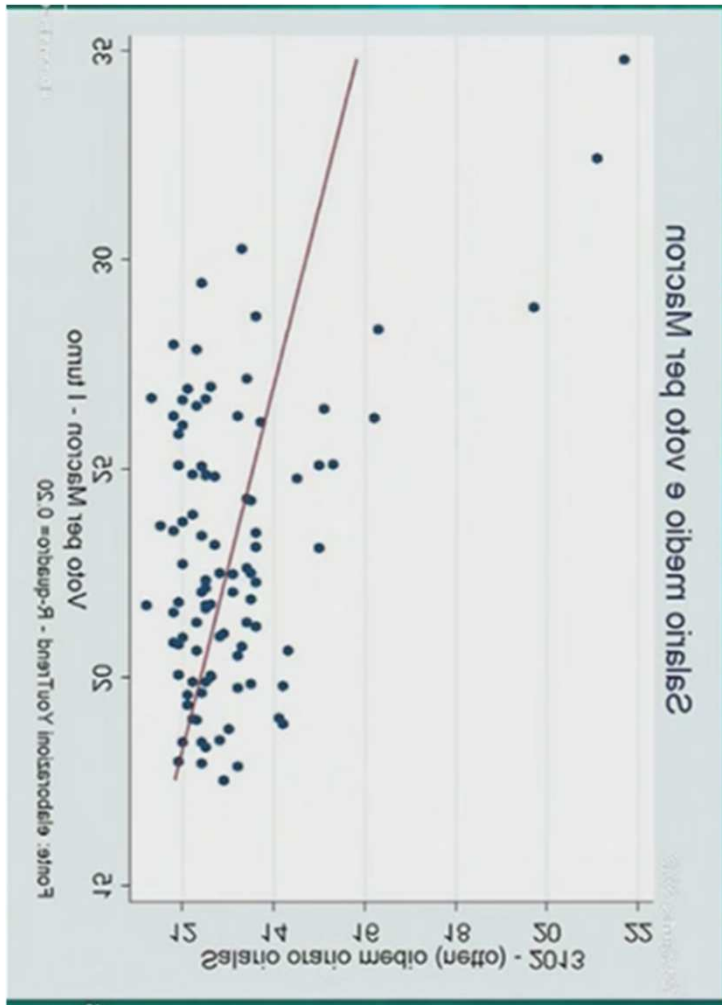


# Macron e lo stipendio

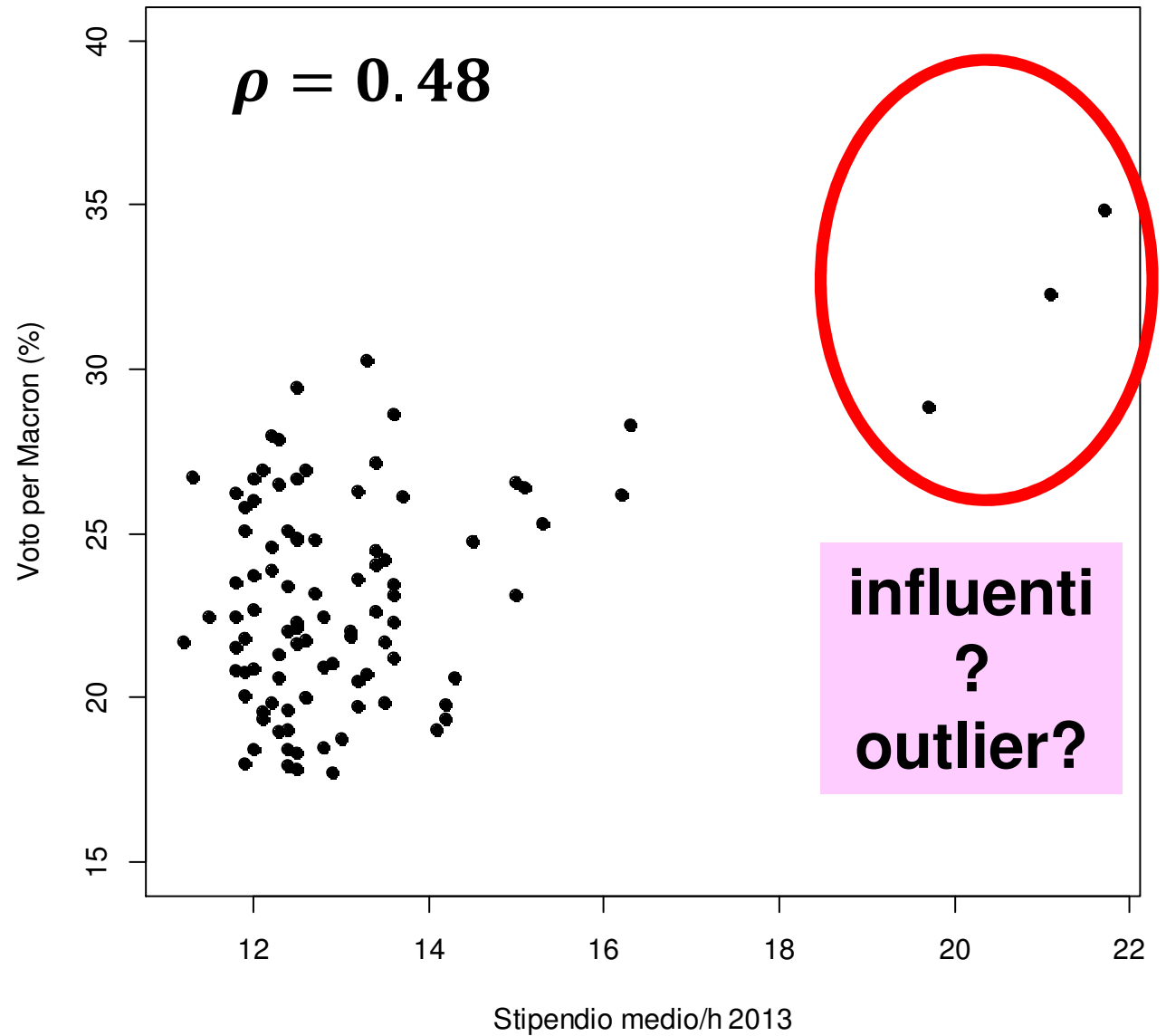
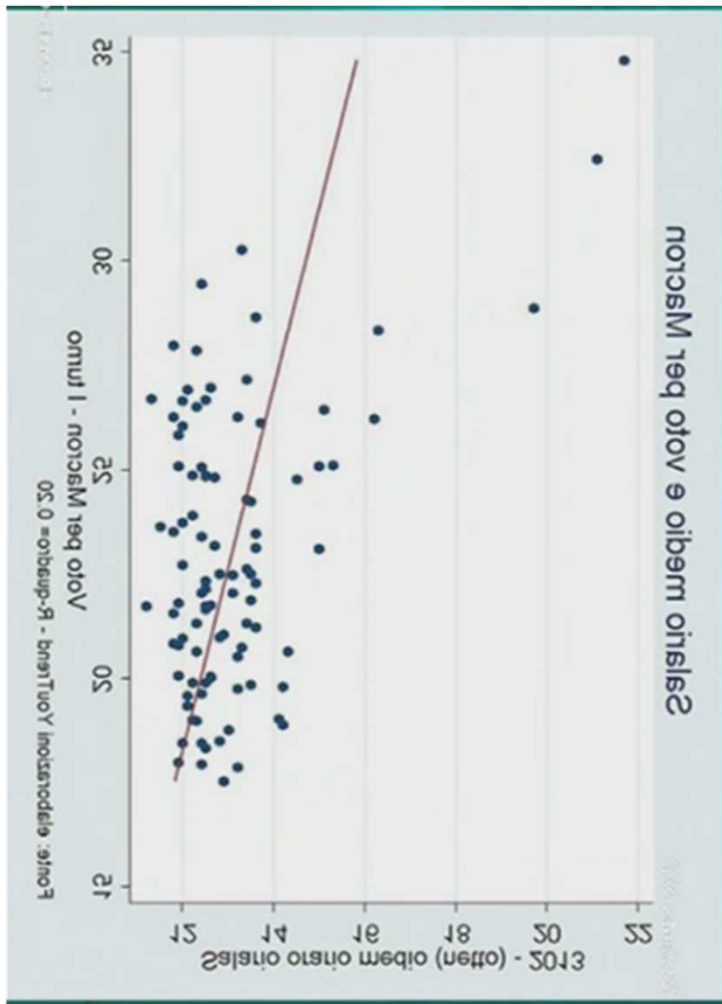
**Stipendio  
medio orario  
2013**



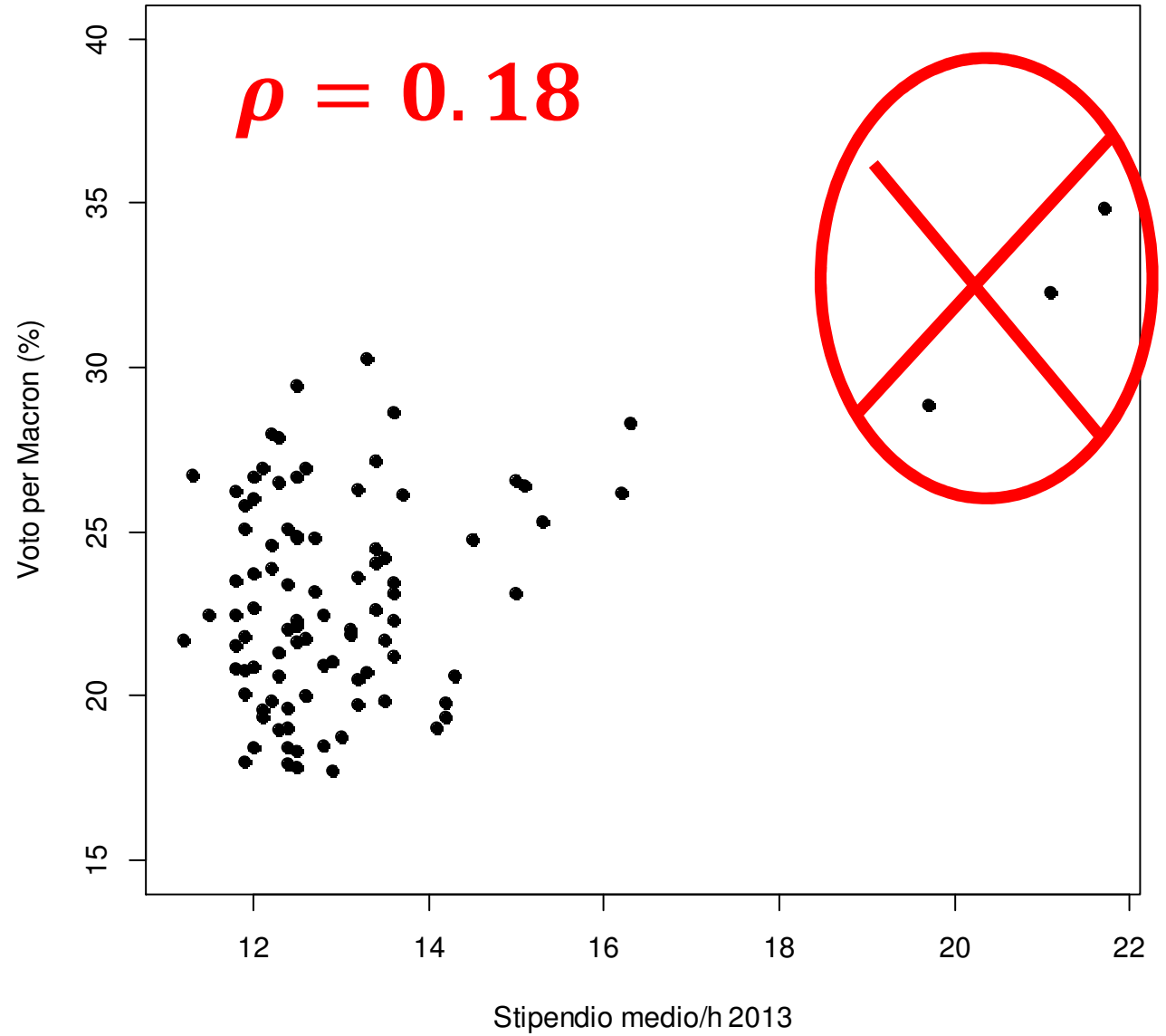
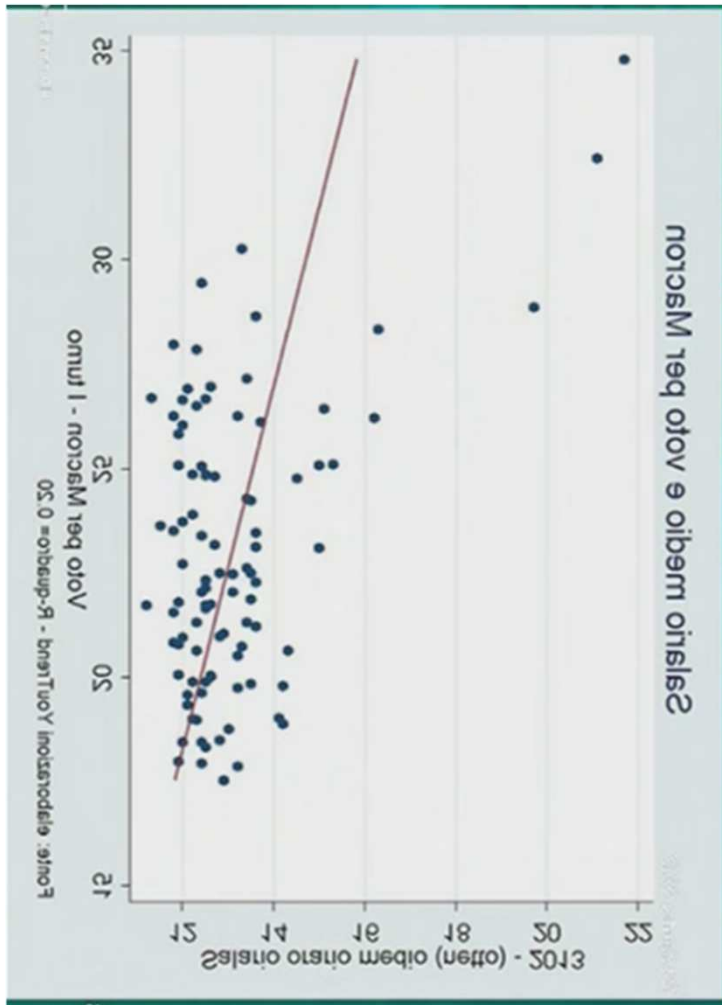
# Macron e lo stipendio



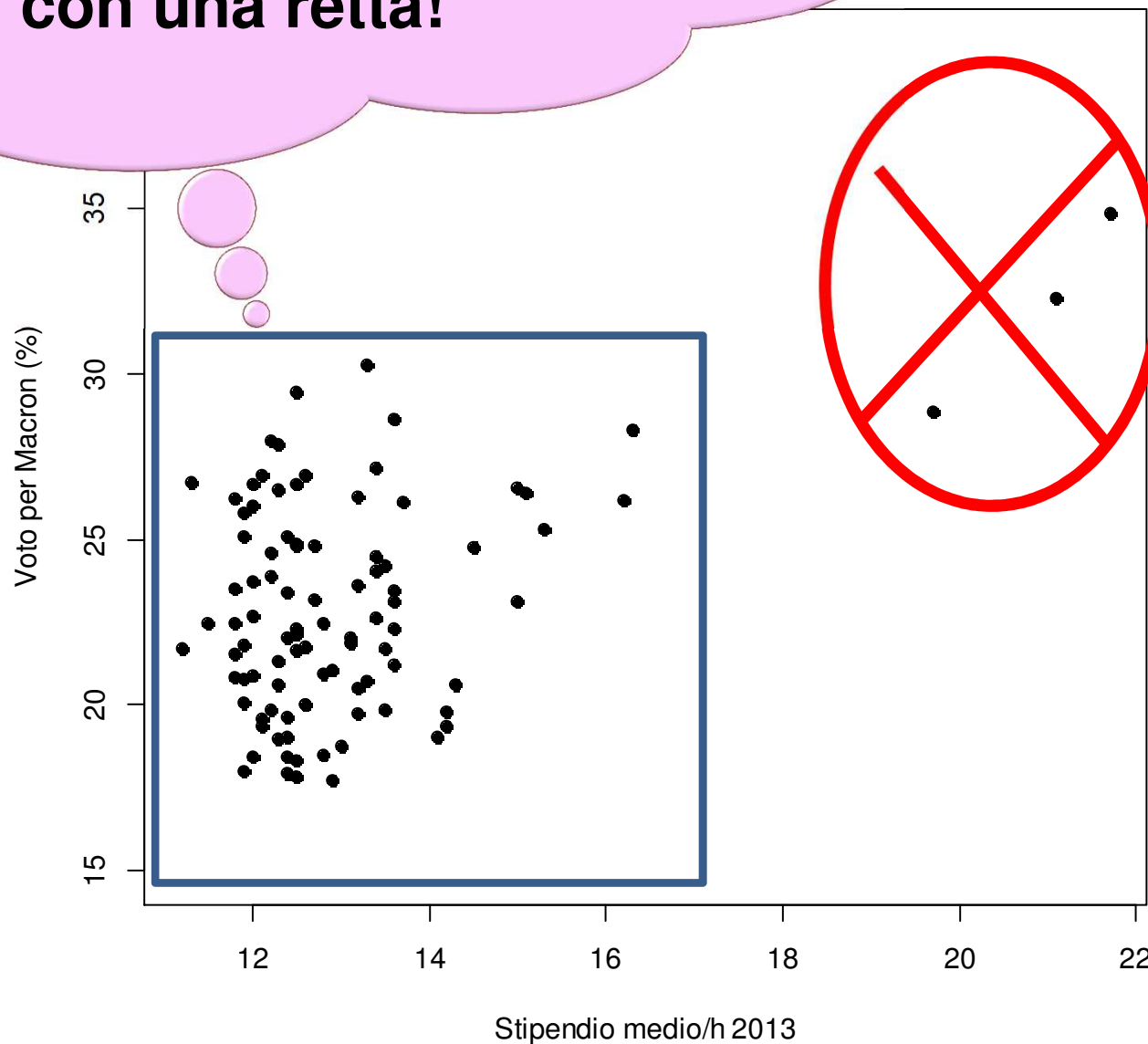
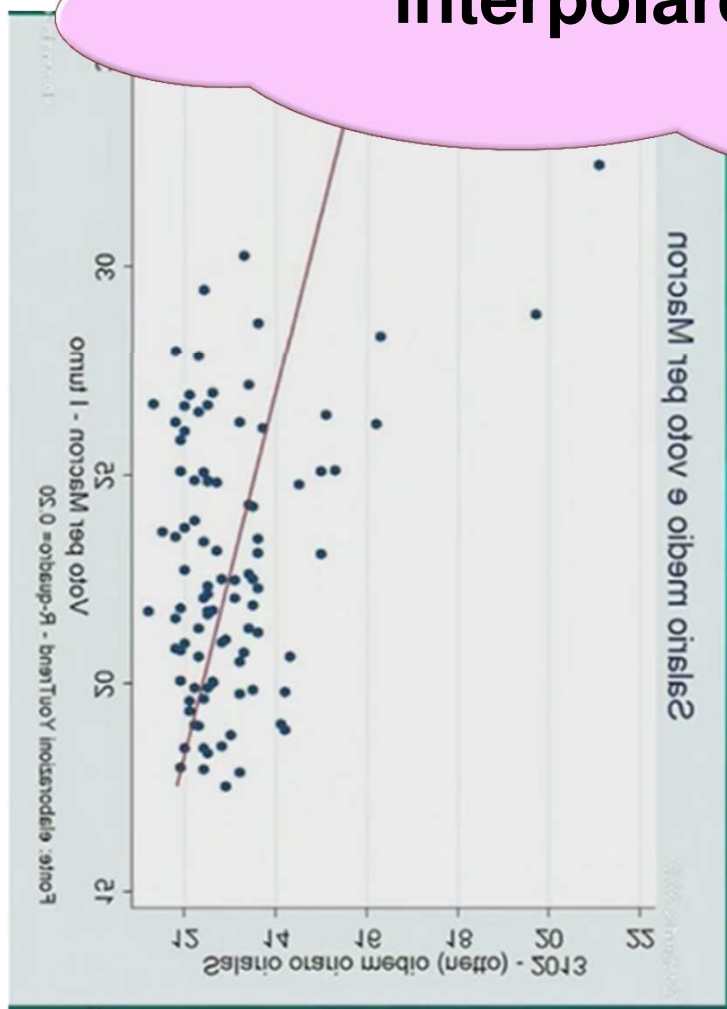
# Macron e lo stipendio



# Macron e lo stipendio



**Non è un diagramma di dispersione che suggerisca di interpolare con una retta!**



# Esercizio 3

<b>Variabile</b>	<b>Coeff.</b>	<b>Dev. std.</b>	<b>Statistica <i>t</i></b>	<b><i>p-value</i></b>
Intercetta	3.8199	9.0891	0.420	0.677
$X$	2.0642	0.3029	6.816	0

## Esercizio 3

Variabile	Coeff.	Dev. std.	Statistica <i>t</i>	<i>p-value</i>
Intercetta	3.8199	9.0891	0.420	0.677
<i>X</i>	2.0642	0.3029	6.816	0

$$y_i = 3.8199 + 2.0642x_i + \varepsilon_i$$



# Esercizio 3

Variabile	Coeff.	Dev. std.	Statistica <i>t</i>	<i>p-value</i>
Intercetta	3.8199	9.0891	0.420	0.677
<i>X</i>	2.0642	0.3029	6.816	0

$$y_i = 3.8199 + 2.0642x_i + \varepsilon_i$$

valori della statistica per i due test d'ipotesi

$$H_0 : a = 0$$

e

$$H_0 : b = 0 :$$

$$\hat{a}$$

$$\hat{b}$$

$$\frac{\hat{a}}{\sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}} = 0.42$$

$$\frac{\hat{b}}{\sqrt{\frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = 6.816$$

# Esercizio 3

Variabile	Coeff.	Dev. std.	Statistica $t$	$p$ -value
Intercetta	3.8199	9.0891	0.420	0.677
$X$	2.0642	0.3029	6.816	0

valori del *denominatore* nella statistica per i due test

$$H_0 : a = 0$$

e

$$H_0 : b = 0 :$$

$$\hat{a}$$

$$\hat{b}$$

9.0891

$$s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

0.3029

$$\frac{s^2}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

# Esercizio 3

Variabile	Coeff.	Dev. std.	Statistica $t$	$p$ -value
Intercetta	3.8199	9.0891	0.420	0.677
$X$	2.0642	0.3029	6.816	0

$$\frac{\text{Coeff}}{\text{Dev. Std}} = \text{Statistica } t$$

valori del denominatore per i due test

$H_0 : a = 0$  e  $H_0 : b = 0 :$

$\hat{a}$   $\hat{b}$

$$\sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$$

$$\sqrt{\frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

# Esercizio 3

Variabile	Coeff.	Dev. std.	Statistica $t$	$p$ -value
Intercetta	3.8199	9.0891	0.420	0.677
$X$	2.0642	0.3029	6.816	0

$$y_i = 3.8199 + 2.0642x_i + \varepsilon_i$$

valori della statistica per i due test d'ipotesi

$$H_0 : a = 0$$

e

$$H_0 : b = 0 :$$

$$\hat{a}$$

$$\hat{b}$$

$$\frac{\hat{a}}{\sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}} = 0.42$$

$$\frac{\hat{b}}{\sqrt{\frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = 6.816$$

## Esercizio 3

Variabile	Coeff.	Dev. std.	Statistica $t$	$p$ -value
Intercetta	3.8199	9.0891	0.420	0.677
$X$	2.0642	0.3029	6.816	0

p-value per i due test d'ipotesi  
 $H_0 : a = 0$  e  $H_0 : b = 0$

1. Intercetta: **non** rifiutiamo  $H_0 : a = 0$

2.  $X$  : **rifiutiamo**  $H_0 : b = 0$  a qualunque livello di significatività, la regressione è significativa

## Esercizio 2, ripreso da lezione del 31/05

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi:  $\bar{x}_U = 115$  e  $\bar{x}_D = 111.9$ , con le rispettive varianze:  $s_U^2 = 624.31$  e  $s_D^2 = 561.04$ .

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

due campioni indipendenti

$$H_0 : \mu_U = \mu_D \quad H_1 : \mu_U > \mu_D$$

$X_i \sim \mathbf{N}(\mu_U, \sigma^2)$  QI uomini

$Y_i \sim \mathbf{N}(\mu_D, \sigma^2)$  QI donne

$$\bar{x}_U = 115 > \bar{x}_D = 111.9 \Rightarrow \text{test}$$

## Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi:  $\bar{x}_U = 115$  e  $\bar{x}_D = 111.9$ , con le rispettive varianze:  $s_U^2 = 624.31$  e  $s_D^2 = 561.04$ .

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

$$H_0 : \mu_U = \mu_D \quad H_1 : \mu_U > \mu_D$$

$$s_p^2 = 592.675$$

$$\frac{\bar{x}_U - \bar{x}_D}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{115 - 111.9}{\sqrt{\frac{592.675 \times 2}{20}}} = 0.40$$

Calcoliamo il p-valore

## Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi:  $\bar{x}_U = 115$  e  $\bar{x}_D = 111.9$ , con le rispettive varianze:  $s_U^2 = 624.31$  e  $s_D^2 = 561.04$ .

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

$$H_0 : \mu_U = \mu_D \quad H_1 : \mu_U > \mu_D$$

$$s_p^2 = 592.675$$

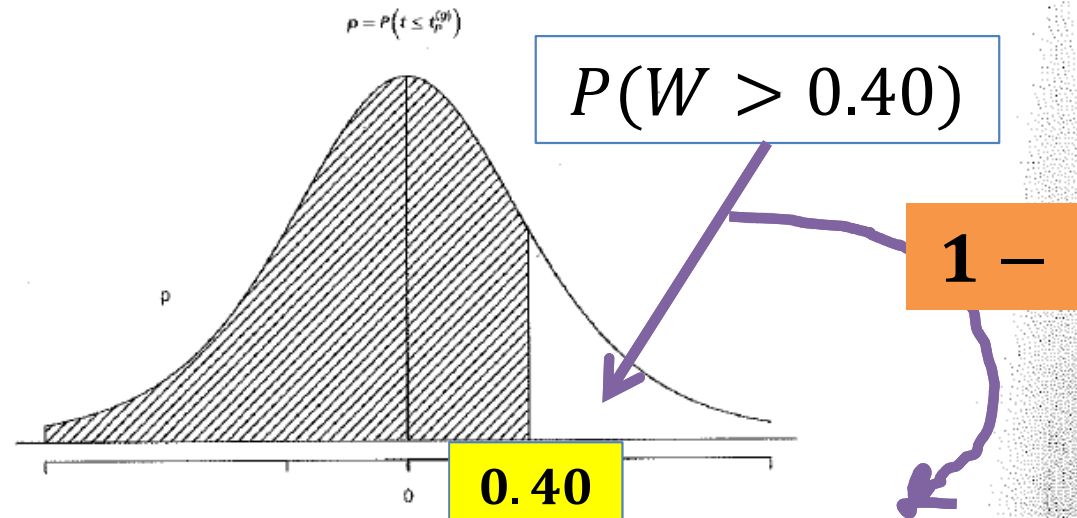
$$\frac{\bar{x}_U - \bar{x}_D}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{115 - 111.9}{\sqrt{\frac{592.675 \times 2}{20}}} = 0.40$$

Indichiamo la  
Statistica test con  $W$ :  
si ha  $W \sim t(38)$

$$\Rightarrow p - \text{valore: } P(W > 0.40)$$



Con riferimento a  $t(40)$ :



p	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	636.61925
2	0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	31.59905
3	0.76489	1.63775	2.35338	3.18245	4.54070	5.84091	12.92398
4	0.74070	1.53321	2.13185	2.77645	3.74695	4.60410	8.61030
5	0.72669	1.47588	2.01505	2.57058	3.36493	4.03216	6.86883
6	0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.95882
7	0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	5.40790
8	0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	5.04131
9	0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.78091
10	0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.58689
11	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.43698
12	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	4.31779
13	0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	4.22083
14	0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	4.14045
15	0.69120	1.34061	1.75305	2.13145	2.60248	2.94671	4.07277
16	0.69013	1.33676	1.74588	2.11991	2.58349	2.92078	4.01500
17	0.68920	1.33338	1.73961	2.10982	2.56693	2.89823	3.96513
18	0.68836	1.33039	1.73406	2.10092	2.55238	2.87844	3.92165
19	0.68762	1.32773	1.72913	2.09302	2.53948	2.86093	3.88341
20	0.68695	1.32534	1.72472	2.08596	2.52798	2.84534	3.84952
21	0.68635	1.32319	1.72074	2.07961	2.51765	2.83136	3.81928
22	0.68581	1.32124	1.71714	2.07387	2.50832	2.81876	3.79213
23	0.68531	1.31946	1.71387	2.06866	2.49987	2.80734	3.76763
24	0.68485	1.31784	1.71088	2.06390	2.49216	2.79694	3.74540
25	0.68443	1.31635	1.70814	2.05954	2.48511	2.78744	3.72514
26	0.68404	1.31497	1.70562	2.05553	2.47863	2.77871	3.70661
27	0.68368	1.31370	1.70329	2.05183	2.47266	2.77068	3.68959
28	0.68335	1.31253	1.70113	2.04841	2.46714	2.76326	3.67391
29	0.68304	1.31143	1.69913	2.04523	2.46202	2.75639	3.65941
30	0.68276	1.31042	1.69726	2.04227	2.45726	2.75000	3.64596
40	0.68067	1.30308	1.68385	2.02108	2.42326	2.70446	3.55097
60	0.67860	1.29582	1.67065	2.00030	2.39012	2.66028	3.46020
120	0.67654	1.28865	1.65765	1.97993	2.35782	2.61742	3.37345
∞	0.67449	1.28155	1.64485	1.95996	2.32635	2.57583	3.29053

Tabella A3. - Tavola della  $t$  di Student. La tavola restituisce i valori di  $t_p^{(g)}$  dove  $g$  sono i gradi di libertà. Si tenga sempre conto della relazione  $t_p^{(g)} = -t_{1-p}^{(g)}$ .

Con riferimento a  $t(40)$ :

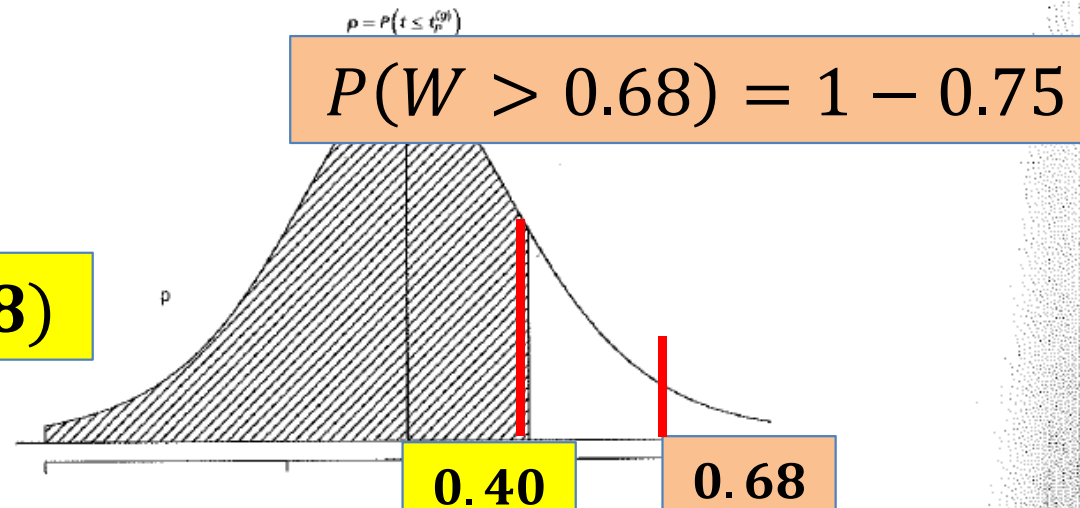
$$P(W > 0.40) > P(W > 0.68)$$



$$P(W > 0.68) > 0.25$$

p-valore > 0.25

Quindi non si può rifiutare l'ipotesi nulla che QI uomini = QI donne, in media



n	p	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1		1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	636.61925
2		0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	31.59905
3		0.76489	1.63775	2.35338	3.18245	4.54070	5.84091	12.92398
4		0.74070	1.53321	2.13185	2.77645	3.74695	4.60410	8.61030
5		0.72669	1.47588	2.01505	2.57058	3.36493	4.03216	6.86883
6		0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.95882
7		0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	5.40790
8		0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	5.04131
9		0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.78091
10		0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.58689
11		0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.43698
12		0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	4.31779
13		0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	4.22083
14		0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	4.14045
15		0.69120	1.34061	1.75305	2.13145	2.60248	2.94671	4.07277
16		0.69013	1.33676	1.74588	2.11991	2.58349	2.92078	4.01500
17		0.68920	1.33338	1.73961	2.10982	2.56693	2.89823	3.96513
18		0.68836	1.33039	1.73406	2.10092	2.55238	2.87844	3.92165
19		0.68762	1.32773	1.72913	2.09302	2.53948	2.86093	3.88341
20		0.68695	1.32534	1.72472	2.08596	2.52798	2.84534	3.84952
21		0.68635	1.32319	1.72074	2.07961	2.51765	2.83136	3.81928
22		0.68581	1.32124	1.71714	2.07387	2.50832	2.81876	3.79213
23		0.68531	1.31946	1.71387	2.06866	2.49987	2.80734	3.76763
24		0.68485	1.31784	1.71088	2.06390	2.49216	2.79694	3.74540
25		0.68443	1.31635	1.70814	2.05954	2.48511	2.78744	3.72514
26		0.68404	1.31497	1.70562	2.05553	2.47863	2.77871	3.70661
27		0.68368	1.31370	1.70329	2.05183	2.47266	2.77068	3.68959
28		0.68335	1.31253	1.70113	2.04841	2.46714	2.76326	3.67391
29		0.68304	1.31143	1.69913	2.04523	2.46202	2.75639	3.65941
30		0.68276	1.31042	1.69726	2.04227	2.45726	2.75000	3.64596
40		0.68067	1.30308	1.68385	2.02108	2.42326	2.70446	3.55097
60		0.67860	1.29582	1.67065	2.00030	2.39012	2.66028	3.46020
120		0.67654	1.28865	1.65765	1.97993	2.35782	2.61742	3.37345
∞		0.67449	1.28155	1.64485	1.95996	2.32635	2.57583	3.29053

Tabella A3. - Tavola della  $t$  di Student. La tavola restituisce i valori di  $t_p^{(g)}$  dove  $g$  sono i gradi di libertà. Si tenga sempre conto della relazione  $t_p^{(g)} = -t_{1-p}^{(g)}$ .

## Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi:  $\bar{x}_U = 115$  e  $\bar{x}_D = 111.9$ , con le rispettive varianze:  $s_U^2 = 624.31$  e  $s_D^2 = 561.04$ .

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

$$H_0 : \mu_U = \mu_D \quad H_1 : \mu_U > \mu_D$$

$$s_p^2 = 592.675$$

$$\frac{\bar{x}_U - \bar{x}_D}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{115 - 111.9}{\sqrt{\frac{592.675 \times 2}{20}}} = 0.40$$

Indichiamo la Statistica test con  $W$ : si ha  $W \sim t(38)$ . Essendo  $38 > 30$  posso approssimare la distribuzione di  $W$  con una normale standard:  $P(Z > 0.40) = 1 - P(Z < 40) = 0.34458$

## Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine
Liceo	47	63
Altra scuola	61	51

- Scelto a caso uno degli studenti del campione, calcolare la probabilità che sia un maschio sapendo che proviene da un liceo.
- Sottoporre a verifica l'ipotesi nulla che il genere e la scuola di provenienza siano indipendenti al livello del 2.5% di significatività
- Calcolare il  $p$ -valore del test al punto b)

# Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47	63	<b>110</b>
Altra scuola	61	51	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

- Scelto a caso uno degli studenti del campione, calcolare la probabilità che sia un maschio sapendo che proviene da un liceo.
- Sottoporre a verifica l'ipotesi nulla che il genere e la scuola di provenienza siano indipendenti al livello del 2.5% di significatività.
- Calcolare il  $p$ -valore del test al punto b)

## Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47	63	<b>110</b>
Altra scuola	61	51	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

- a) Scelto a caso uno degli studenti del campione, calcolare la probabilità che sia un maschio sapendo che proviene da un liceo.
- b) Sottoporre a verifica l'ipotesi nulla che il genere e la scuola di provenienza siano indipendenti al livello del 2.5% di significatività.  $\frac{47}{110} = 0.43$
- c) Calcolare il  $p$ -valore del test al punto b)

## Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47	63	<b>110</b>
Altra scuola	61	51	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

- Scelto a caso uno degli studenti del campione, calcolare la probabilità che sia un maschio sapendo che proviene da un liceo.
- Sottoporre a verifica l'ipotesi nulla che il genere e la scuola di provenienza siano indipendenti al livello del 2.5% di significatività.

# Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47	63	<b>110</b>
Altra scuola	61	51	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

- Scelto a caso uno degli studenti del campione, calcolare la probabilità che sia un maschio sapendo che proviene da un liceo.
- Sottoporre a verifica l'ipotesi nulla che il genere e la scuola di provenienza siano indipendenti al livello del 2.5% di significatività.

$$H_0 : \chi^2 = 0 \text{ vs } H_1 : \chi^2 > 0$$



# Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47; <b>53.5</b>	63; <b>56.5</b>	<b>110</b>
Altra scuola	61; <b>54.5</b>	51; <b>57.5</b>	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

$$\frac{110 \times 108}{222} = 53.5,$$

$$\frac{110 \times 114}{222} = 56.5,$$

$$\frac{112 \times 108}{222} = 54.5,$$

$$\frac{112 \times 114}{222} = 57.5$$

## Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47; <b>53.5</b>	63; <b>56.5</b>	<b>110</b>
Altra scuola	61; <b>54.5</b>	51; <b>57.5</b>	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

$$\begin{aligned}\chi^2 &= \frac{(47-53.5)^2}{53.5} + \frac{(63-56.5)^2}{56.5} + \frac{(61-54.5)^2}{54.5} + \frac{(51-57.5)^2}{57.5} \\ &= \mathbf{0.79 + 0.75 + 0.77 + 0.73 = 3.04}\end{aligned}$$

## Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47; <b>53.5</b>	63; <b>56.5</b>	<b>110</b>
Altra scuola	61; <b>54.5</b>	51; <b>57.5</b>	<b>112</b>
	<b>108</b>	<b>114</b>	<b>222</b>

$$\chi^2 = \frac{(47-53.5)^2}{53.5} + \frac{(63-56.5)^2}{56.5} + \frac{(61-54.5)^2}{54.5} + \frac{(51-57.5)^2}{57.5}$$

$$= 0.79 + 0.75 + 0.77 + 0.73 = 3.04$$

$$\chi^2(1)_{1-0.025} = 5.02389$$

# Esercizio 8

Gli studenti di un corso di laurea vengono classificati sulla base di due caratteristiche: il sesso e il tipo di scuola superiore di provenienza. I dati sono riportati nella seguente tabella:

	Maschi	Femmine	
Liceo	47; 53.5	63; 56.5	110
	51; 54.5	61; 57.5	112
			222

**NON POSSIAMO  
RIFIUTARE L'IPOTESI DI  
INDIPENDENZA AL  
LIVELLO DEL 2.5%**

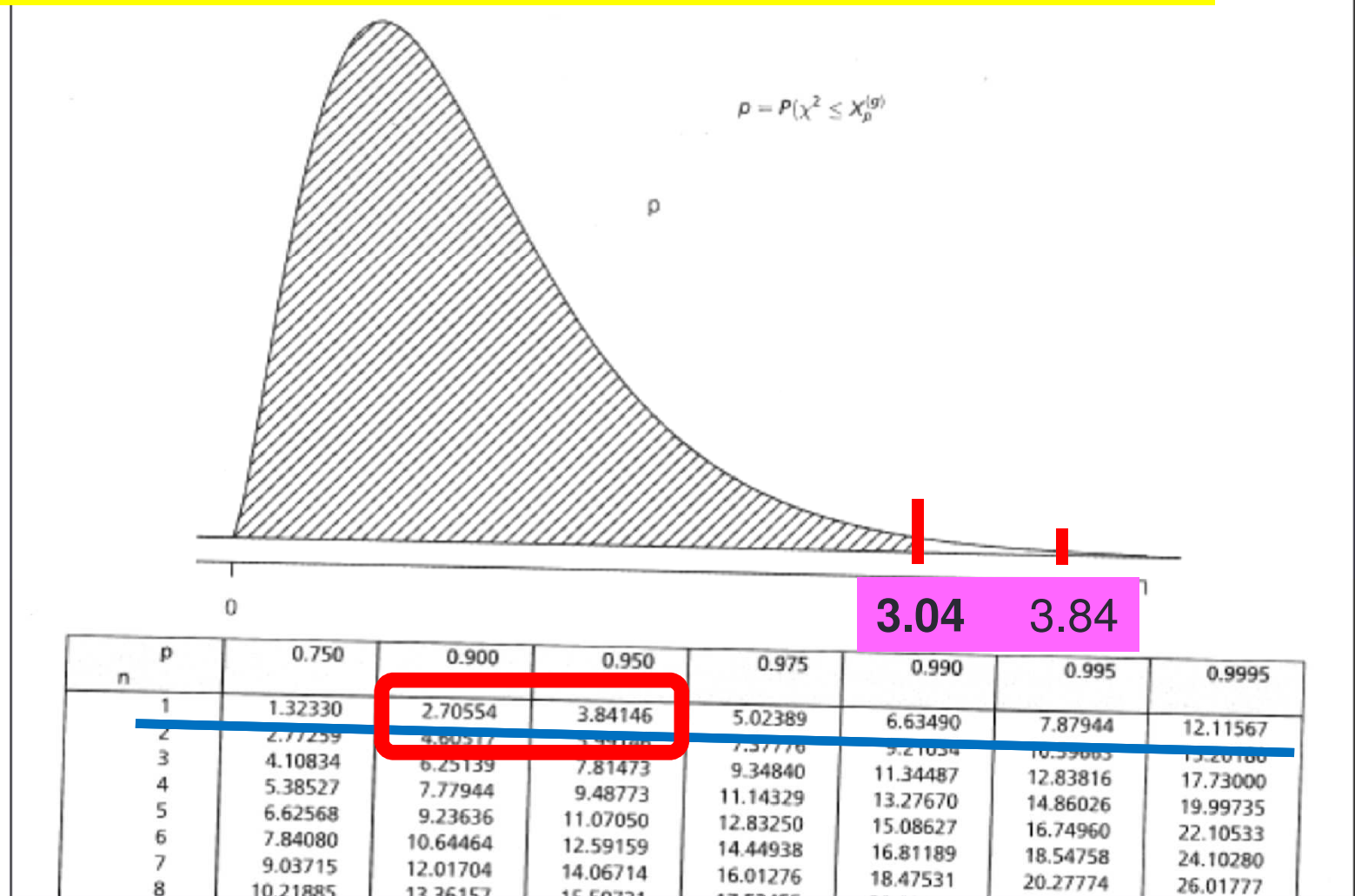
$$\chi^2 = \frac{(47-53.5)^2}{53.5} + \frac{(63-56.5)^2}{56.5} + \frac{(51-54.5)^2}{54.5} + \frac{(61-57.5)^2}{57.5}$$

$$= 0.79 + 0.75 + 0.77 + 0.73 = 3.04$$

$$\chi^2(1)_{1-0.025} = 5.02389$$

# Esercizio 8

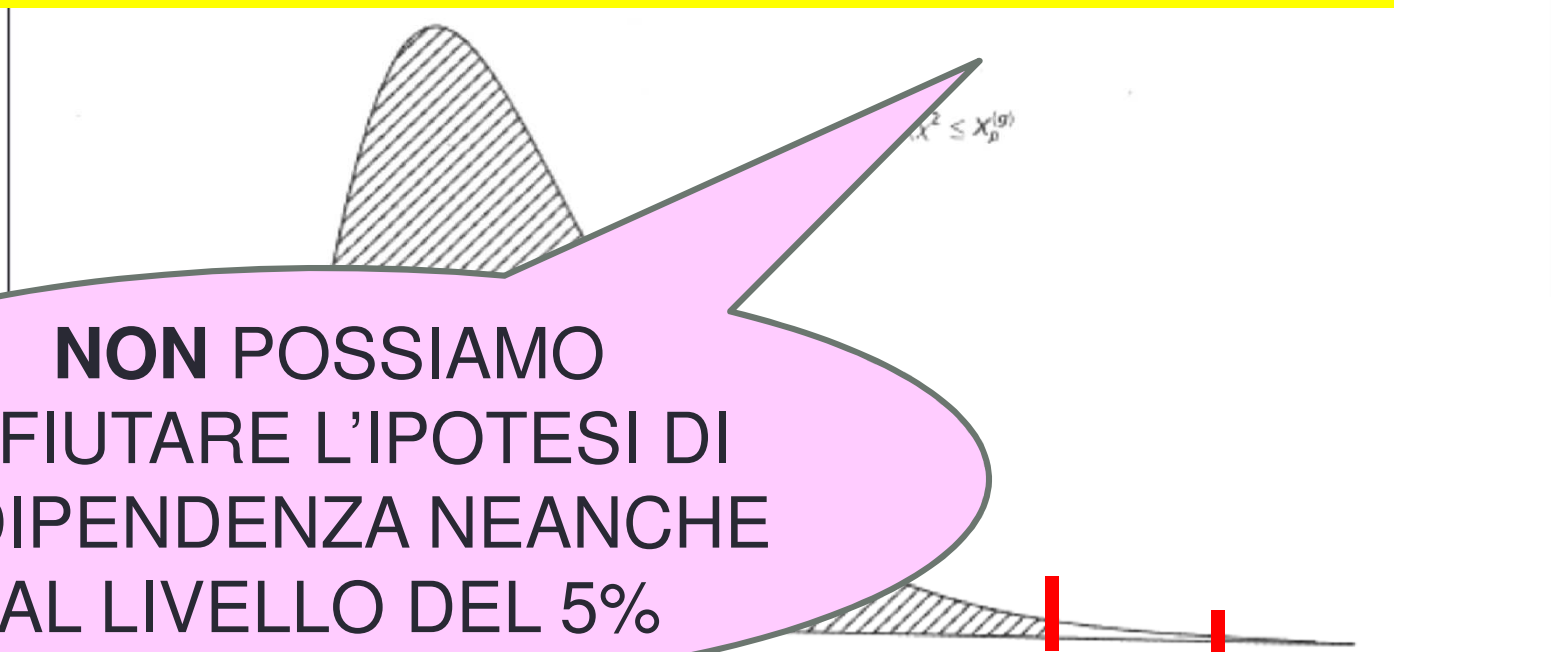
$$P(\chi^2 > 3.04) > P(\chi^2 > 3.84186) = 1 - 0.95 = 0.05$$



# Esercizio 8

$$P(\chi^2 > 3.04) > P(\chi^2 > 3.84186) = 1 - 0.95 = 0.05$$

**NON POSSIAMO  
RIFIUTARE L'IPOTESI DI  
INDIPENDENZA NEANCHE  
AL LIVELLO DEL 5%**



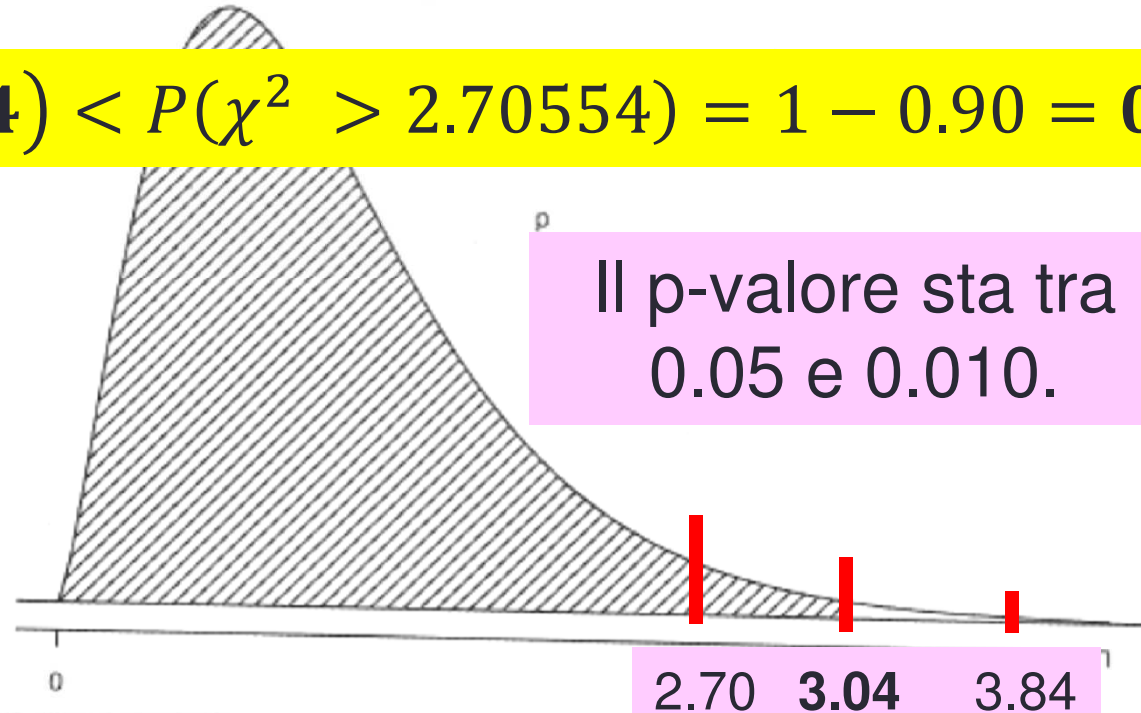
3.04 3.84

n	p	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1		1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2		2.77259	4.60517	5.99148	7.37776	9.21034	10.59653	15.20108
3		4.10834	6.25139	7.81473	9.34840	11.34487	12.83816	17.73000
4		5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5		6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6		7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7		9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8		10.21885	13.36157	15.50771	17.53455	19.97953	22.02777	27.87841

# Esercizio 8

$$P(\chi^2 > 3.04) > P(\chi^2 > 3.84186) = 1 - 0.95 = 0.05$$

$$\& P(\chi^2 > 3.04) < P(\chi^2 > 2.70554) = 1 - 0.90 = 0.10$$



n	p	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1		1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2		2.77259	4.60517	5.99148	7.37776	9.21034	10.59653	15.20108
3		4.10834	6.25139	7.81473	9.34840	11.34487	12.83816	17.73000
4		5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5		6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6		7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7		9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8		10.21885	13.36157	15.50771	17.53455	19.97953	22.02777	27.87841

# Domanda 1

Sia  $X_1, \dots, X_n$  un campione casuale con  $X_i \sim N(\mu, \sigma^2)$  e varianza non nota. Allora, la statistica test per la verifica d'ipotesi  $H_0 : \mu = 0$  vs  $H_1 : \mu \neq 0$  è:

a) $\bar{X}_n$	b) $\bar{X}_n / (\sigma^2 / n)$
c) $ \bar{X}_n  / \sqrt{S_n^2 / n}$	d) $\bar{X}_n / \sqrt{\sigma^2 / n}$



# Domanda 1

Sia  $X_1, \dots, X_n$  un campione casuale con  $X_i \sim N(\mu, \sigma^2)$  e varianza non nota. Allora, la statistica test per la verifica d'ipotesi  $H_0 : \mu = 0$  vs  $H_1 : \mu \neq 0$  è:

a) $\bar{X}_n$	b) $\bar{X}_n / (\sigma^2 / n)$
c) $ \bar{X}_n  / \sqrt{S_n^2 / n}$	d) $\bar{X}_n / \sqrt{\sigma^2 / n}$

## Domanda 2

In un test per la verifica d'ipotesi con livello di significatività 5% si rifiuta l'ipotesi nulla. Allora:

a) Il p-valore è $> 0.05$	b) Il p-valore è $< 0.05$
c) $\mu = 0$	d) $\mu \neq 0$

## Domanda 2

In un test per la verifica d'ipotesi con livello di significatività 5% si rifiuta l'ipotesi nulla. Allora:

a) Il p-valore è $> 0.05$	b) Il p-valore è $< 0.05$
c) $\mu = 0$	d) $\mu \neq 0$

## Domanda 3

Sia  $X_1, \dots, X_n$  un campione casuale di dimensione  $n > 30$  con distribuzione non nota. Si rifiuta l'ipotesi nulla  $H_0 : \mu = 0$  a favore di  $H_1 : \mu > 0$  al livello  $\alpha$

a) se la statistica test è $> z_{1-\frac{\alpha}{2}}$	b) se la statistica test è $> z_{1-\alpha}$
c) se la statistica test è $> t(n-1)_{1-\frac{\alpha}{2}}$	d) se la statistica test è $> t(n-1)_{1-\alpha}$

## Domanda 3

Sia  $X_1, \dots, X_n$  un campione casuale di dimensione  $n > 30$  con distribuzione non nota. Si rifiuta l'ipotesi nulla  $H_0 : \mu = 0$  a favore di  $H_1 : \mu > 0$  al livello  $\alpha$

a) se la statistica test è $> z_{1-\frac{\alpha}{2}}$	b) se la statistica test è $> z_{1-\alpha}$
c) se la statistica test è $> t(n-1)_{1-\frac{\alpha}{2}}$	d) se la statistica test è $> t(n-1)_{1-\alpha}$

# Domanda 4

Per valutare la bontà di adattamento di un modello lineare ai dati si utilizza

a) l'indice di dispersione	b) l'indice $\chi^2$
c) l'indice $R^2$	d) la pendenza

# Domanda 4

Per valutare la bontà di adattamento di un modello lineare ai dati si utilizza

a) l'indice di dispersione	b) l'indice $\chi^2$
c) l'indice $R^2$	d) la pendenza

## Domanda 5

Se due variabili qualitative osservate congiuntamente risultano essere indipendenti, allora

a) l'indice di correlazione vale 1	b) l'indice $\chi^2 = 0$
c) l'indice $R^2 = 0$	d) la pendenza della retta di regressione vale 0



## Domanda 5

Se due variabili qualitative osservate congiuntamente risultano essere indipendenti, allora

a) l'indice di correlazione vale 1	b) l'indice $\chi^2 = 0$
c) l'indice $R^2 = 0$	d) la pendenza della retta di regressione vale 0

# Domanda 6

L'errore di prima specie è

a) La probabilità di rifiutare l'ipotesi nulla quando questa è vera	b) La probabilità di rifiutare l'ipotesi nulla quando questa è falsa
c) La probabilità di non rifiutare l'ipotesi nulla quando questa è vera	d) La probabilità di rifiutare l'alternativa quando questa è vera

# Domanda 6

L'errore di prima specie è

a) La probabilità di rifiutare l'ipotesi nulla quando questa è vera	b) La probabilità di rifiutare l'ipotesi nulla quando questa è falsa
c) La probabilità di non rifiutare l'ipotesi nulla quando questa è vera	d) La probabilità di rifiutare l'alternativa quando questa è vera

# Domanda 7

Se il  $p$ -valore di una verifica d'ipotesi vale 0.015, allora quali di queste affermazioni sono vere?

a) Si rifiuta l'ipotesi nulla al livello del 5%	b) Si rifiuta l'ipotesi nulla al livello dell'1%
c) Non si può rifiutare l'ipotesi nulla al livello del 5%	d) Non si può rifiutare l'ipotesi nulla al livello dell'1%

# Domanda 7

Se il  $p$ -valore di una verifica d'ipotesi vale 0.015, allora quali di queste affermazioni sono vere?

a) Si rifiuta l'ipotesi nulla al livello del 5%	b) Si rifiuta l'ipotesi nulla al livello dell'1%
c) Non si può rifiutare l'ipotesi nulla al livello del 5%	d) Non si può rifiutare l'ipotesi nulla al livello dell'1%

# Domanda 8

Il coefficiente di correlazione lineare  $\rho$  :

a) Indica il grado di associazione tra due variabili qualitative	b) Indica il grado di dipendenza lineare tra due variabili quantitative
c) Indica la connessione tra due variabili	d) Coincide con la covarianza

# Domanda 8

Il coefficiente di correlazione lineare  $\rho$  :

a) Indica il grado di associazione tra due variabili qualitative	b) Indica il grado di dipendenza lineare tra due variabili quantitative
c) Indica la connessione tra due variabili	d) Coincide con la covarianza

## Domanda 9

Sia  $X \sim N(5, \sigma^2)$ , allora  $P(X < 5)$  vale :

a) Non si può calcolare	b) 0
c) 1	d) 0.5



# Domanda 9

Sia  $X \sim N(5, \sigma^2)$ , allora  $P(X < 5)$  vale :

a) Non si può calcolare	b) 0
c) 1	d) 0.5

# Domanda 10

Siano  $A$  e  $B$  due eventi indipendenti con la stessa probabilità, pari a 0.5. Allora

a) $P(A \cap B) = 0$	b) $P(A \cup B) = 1$
c) $P(A \cap B) = 0.25$	d) $P(A B) = 1$

# Domanda 10

Siano  $A$  e  $B$  due eventi indipendenti con la stessa probabilità, pari a 0.5. Allora

a) $P(A \cap B) = 0$	b) $P(A \cup B) = 1$
c) $P(A \cap B) = 0.25$	d) $P(A B) = 1$