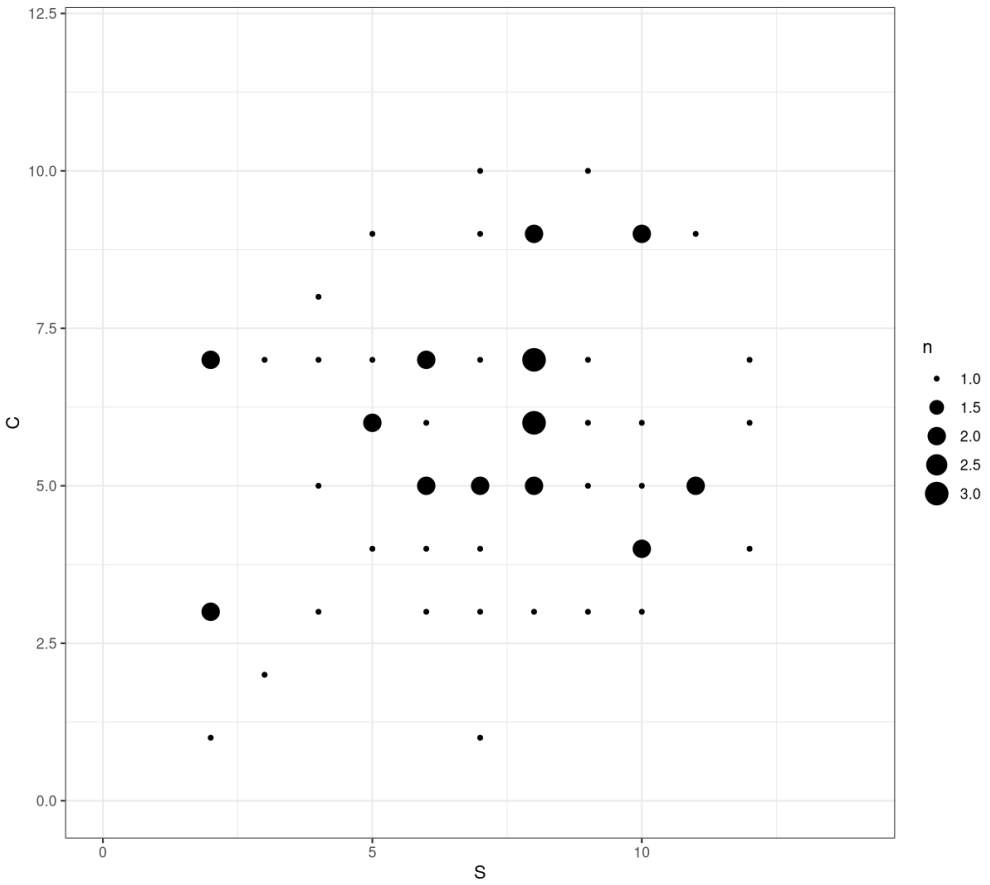
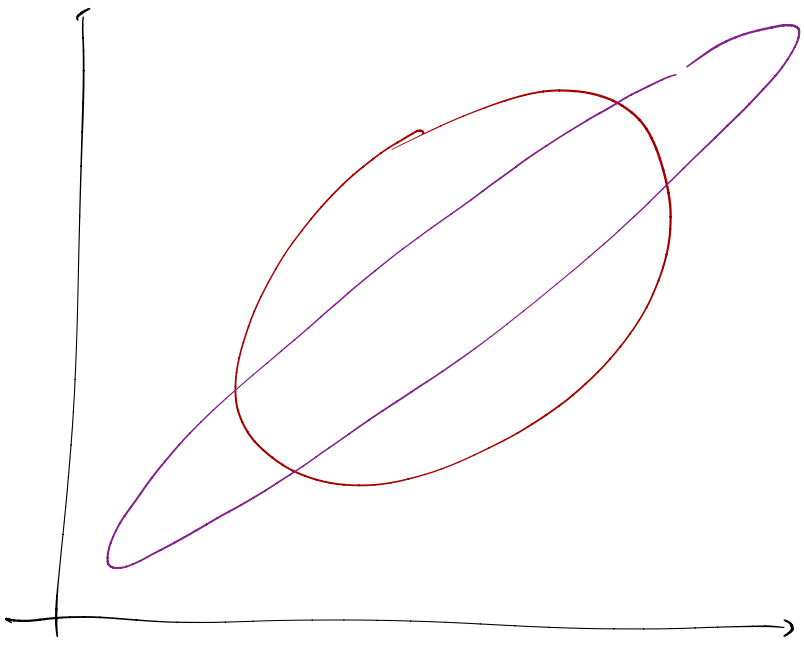
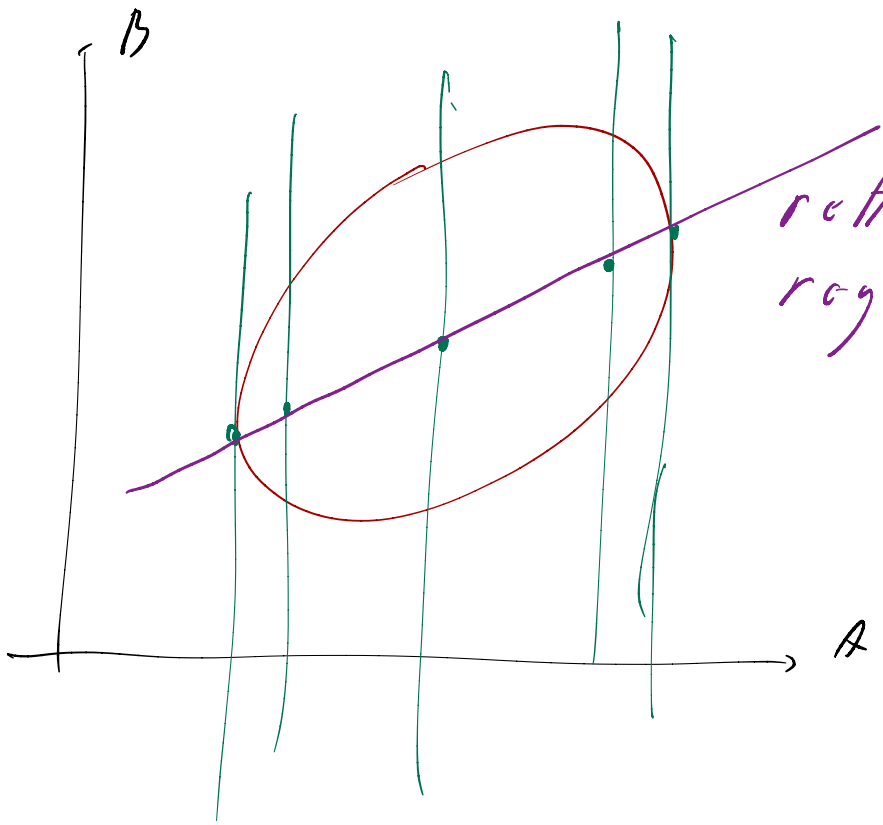
Regressione lineare







- coeff. di correlazione r
- retta di regressione
- p-value

$-1 \leq r \leq 1$

Calcolo di r

A, B v. statistiche.

$(a_1, b_1), (a_2, b_2), \ldots$ — campione

$$A^* = \frac{A - \mu_A}{\sigma_A} \qquad B = \frac{B - \mu_B}{\sigma_B}$$

$(a_1^*, b_1^*), (a_2^*, b_2^*), \ldots$

$$r = \frac{a_1^* \cdot b_1^* + a_2^* \cdot b_2^* + \ldots}{n}$$

n ↑ numerosità del campione.

	1	0	16	16	1	4	0	1		1
	-1	0	4	4	-1	2	0	-1	0	1
a_i	4	5	1	9	4	7	5	4	5	6
b_i	5	6	8	2	4	1	5	3	2	4

$\mu_A = 5.0$ $\quad \sigma_A^2 = 4$ $\quad \sigma_A = 2$

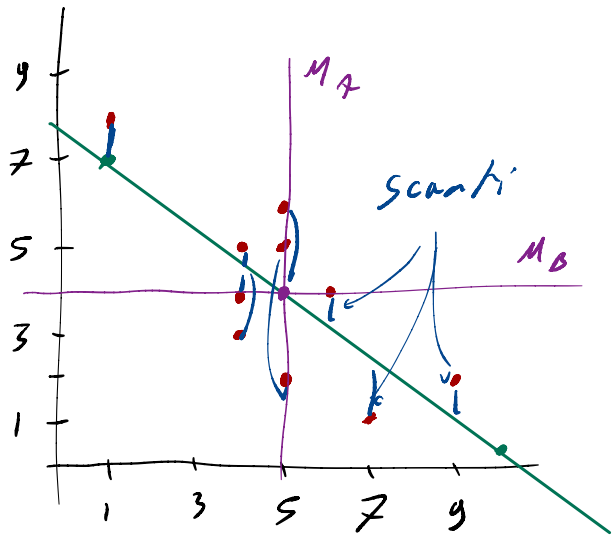
$\mu_B = 4.0$ $\quad \ldots$ $\quad \sigma_B = 2$

$A^* = \frac{A-5}{2}$

	-0.5	0	-2	2	-0.5	1	0	-0.5	0	0.5
a_i	4	5	1	9	4	7	5	4	5	6
b_i	5	6	8	2	4	1	5	3	2	4
$B^* = \frac{B-4}{2}$	0.5	1	2	-1	0	-1.5	0.5	-0.5	-1	0
$A^* \cdot B^*$	-0.25	0	-4	-2	0	-1.5	0	0.25	0	0	-7.5

$$r = \frac{a_1^* b_1^* + \ldots}{10} = -0.75$$

a_i	4	5	1	9	4	7	5	4	5	6
b_i	5	6	8	2	4	1	5	3	2	4



Retta di regressione

$$B^* = r\, A^*$$

$$\frac{B - \mu_B}{\sigma_B} = r\,\frac{A - \mu_A}{\sigma_A}$$

$$B = A \cdot r\,\frac{\sigma_B}{\sigma_A} + \mu_B - \mu_A\,\frac{\sigma_B}{\sigma_A}$$

Es. (continua)

$$B^* = -0.75\, A^*$$

$$\frac{B - 4}{2} = -0.75\,\frac{A - 5}{2}$$

$$B = 4 - 0.75\,(A - 5)$$

$$B = -0.75\,A + 7.75$$

p-value

Q: c'è evidenza statistica di una correlazione reale?

Ipotesi nulla: "le variabili sono scorrelate"

I. alternativa: "c'è una correlazione reale"

p-value = probabilità di un valore di $|r| \geq$ di quello osservato SOTTO L'IPOTESI NULLA

C'è evidenza di una corr. se

$$p\text{-value} < \alpha$$

α = livello di significatività

(tipici 0.05, 0.01)

5% 1%

α MINORE $\rightsquigarrow$ significatività MAGGIORE

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

è distribuita secondo una t di Student con $n-2$ gradi di libertà.

Es. (continua)

$r = -0.75 \qquad n = 10$

$$t = -0.75 \sqrt{\frac{8}{1-0.75^2}} = -3.21$$

$|t| > t_{0.025,8}$ ✓

3.21

2.31

$\frac{5\%}{2}$