



**Analisi dell'associazione tra
due caratteri:
correlazione**

Introduzione

- ❑ **Analizzare due caratteri quantitativi**
- ❑ **Qualificare e quantificare la relazione lineare**
- ❑ **Utilizzare insieme il coefficiente di correlazione e il Chi-quadrato**

Grafico di dispersione

Reddito familiare (X) e spesa per consumi alimentari (Y)

	A	B	C	D	E	F	G	H	I	J
X	39	41	37	28	20	25	42	20	30	60
Y	9	8,4	7,2	13,4	12	8	11,4	6,5	9,6	15,7
	K	L	M	N	O	P	Q	R	S	T
X	18	24	42	45	36	15	21	34	25	58
Y	6	7,5	12	15	12,2	6	7,5	10,5	8	14,1

A



Ascissa = 39 ; Ordinata = 9

B



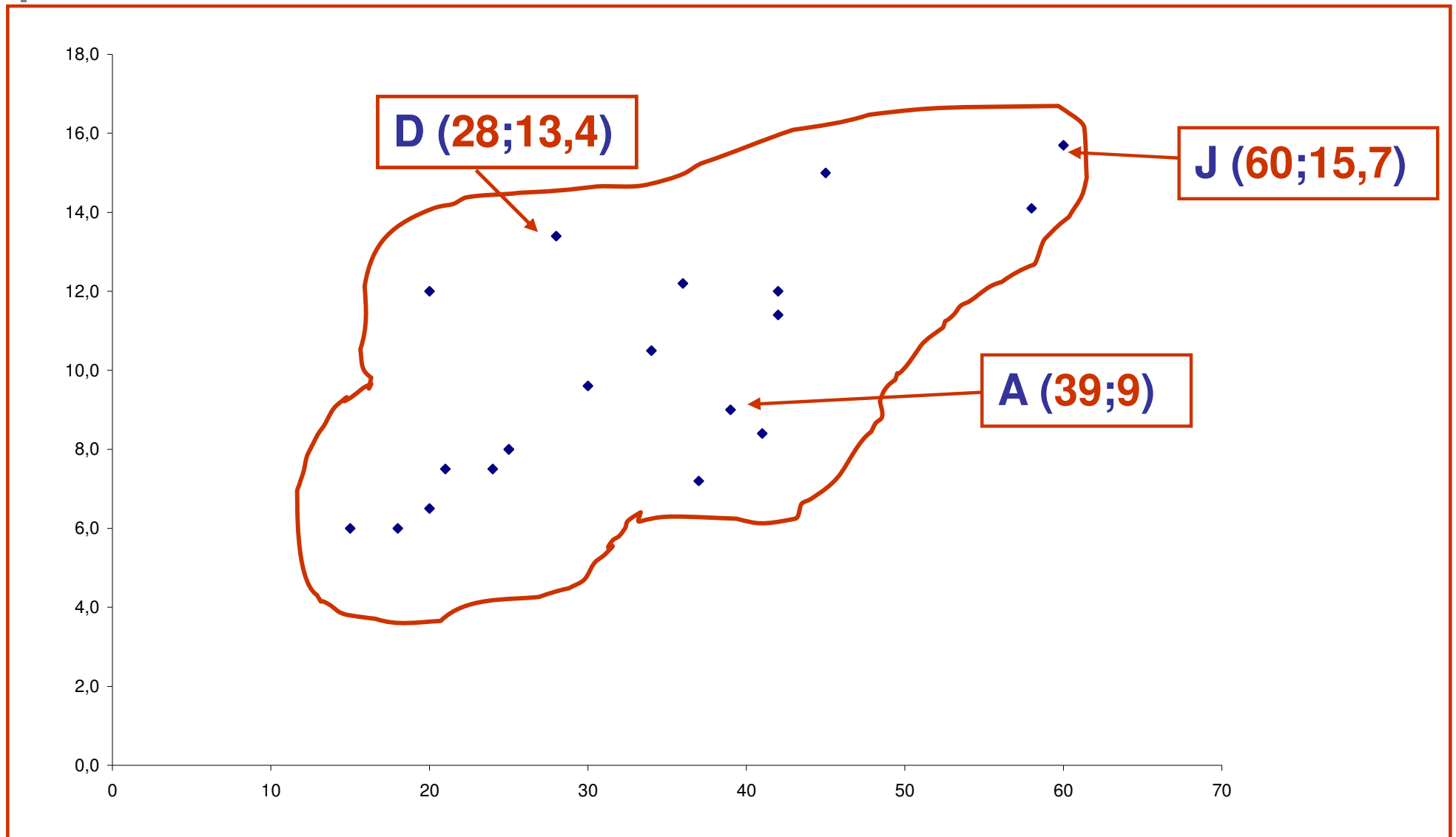
Ascissa = 41 ; Ordinata = 8,4

C

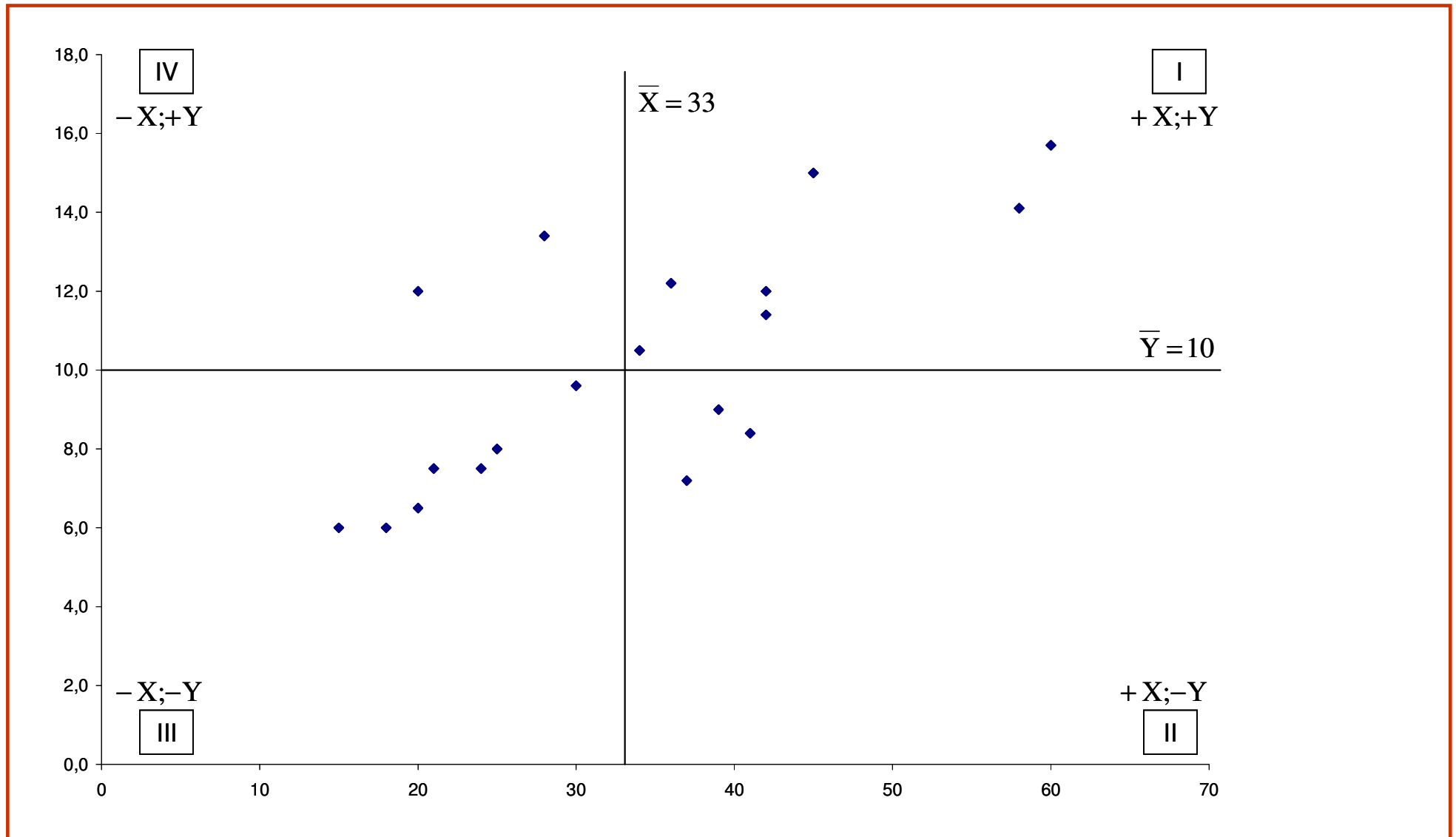


Ascissa = 37 ; Ordinata = 7,2

Grafico di dispersione - 2



La relazione lineare



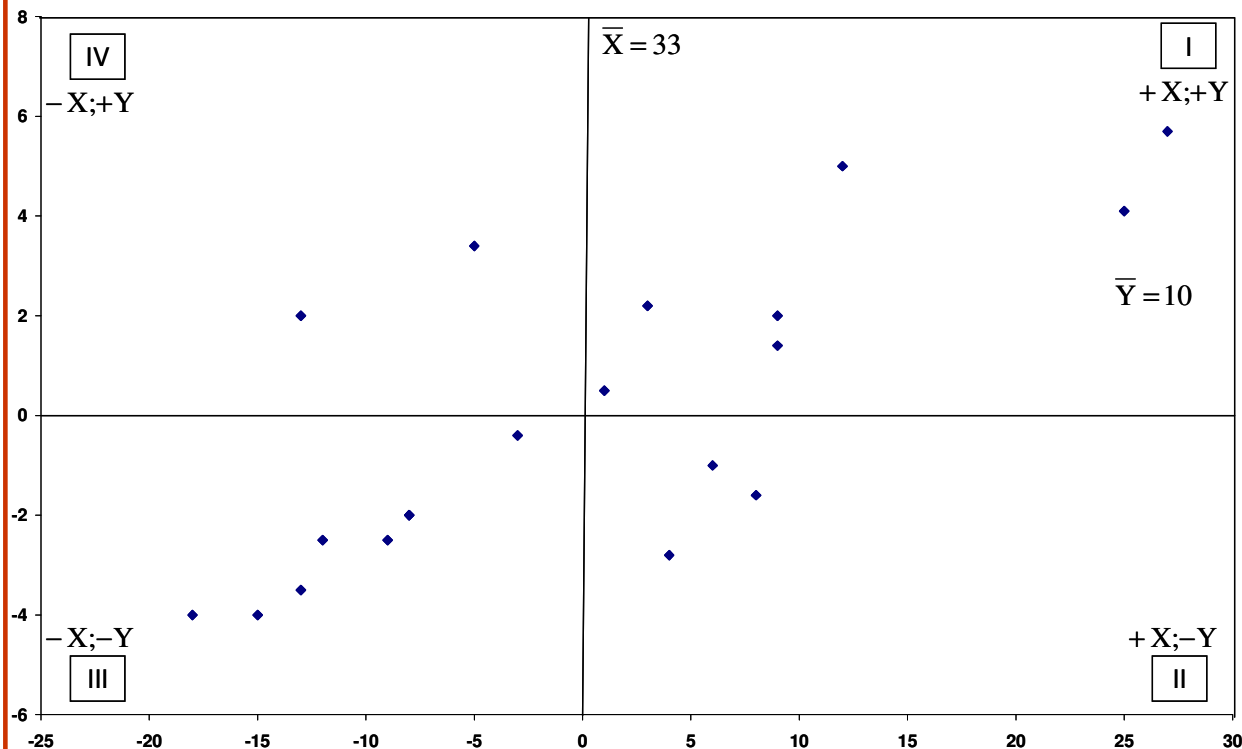
La relazione lineare

$$(X_i - \bar{X})$$

rappresenta lo scarto tra ognuno dei valori della variabile X e la propria media aritmetica

$$(y_i - \bar{y})$$

rappresenta lo scarto tra ognuno dei valori della variabile Y e la propria media aritmetica



Nel I e III quadrante scarti con stesso segno (**concordi**)

Nel II e IV quadrante scarti con segno diverso (**discordi**)

La relazione lineare - 2

Se la maggior parte dei punti si trova nel I e nel III quadrante, allora la maggior parte degli scarti che andremo a calcolare saranno **CONCORDI**



Esiste una relazione **di tipo DIRETTO** tra le variabili

Se la maggior parte dei punti si trova nel II e nel IV quadrante, allora la maggior parte degli scarti che andremo a calcolare saranno **DISCORDI**



Esiste una relazione **di tipo INVERSO** tra le variabili

La Covarianza

$$\sigma_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

COVARIANZA

media di tutti i prodotti tra gli scostamenti delle variabili X ed Y riscontrati in tutte le unità statistiche considerate

La covarianza è una misura della dipendenza (o interdipendenza) lineare tra due variabili

Valori positivi



CONCORDANZA

Valori negativi



DISCORDANZA

Zero



INDIPENDENZA (LINEARE!)

La Covarianza - 2

La covarianza ci dice se esiste una relazione lineare e ci dice se questa è di tipo diretto o inverso.

Purtroppo la covarianza non è un indice relativo, e, pertanto, non è in grado di dirci quanto è forte la dipendenza.

Possiamo vedere “come varia” la covarianza

$$-\sigma_X \sigma_Y \leq \sigma_{XY} \leq +\sigma_X \sigma_Y$$

Il Coefficiente di Correlazione

Possiamo, quindi, costruire un indice relativo sulla base delle conoscenze emerse in precedenza

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

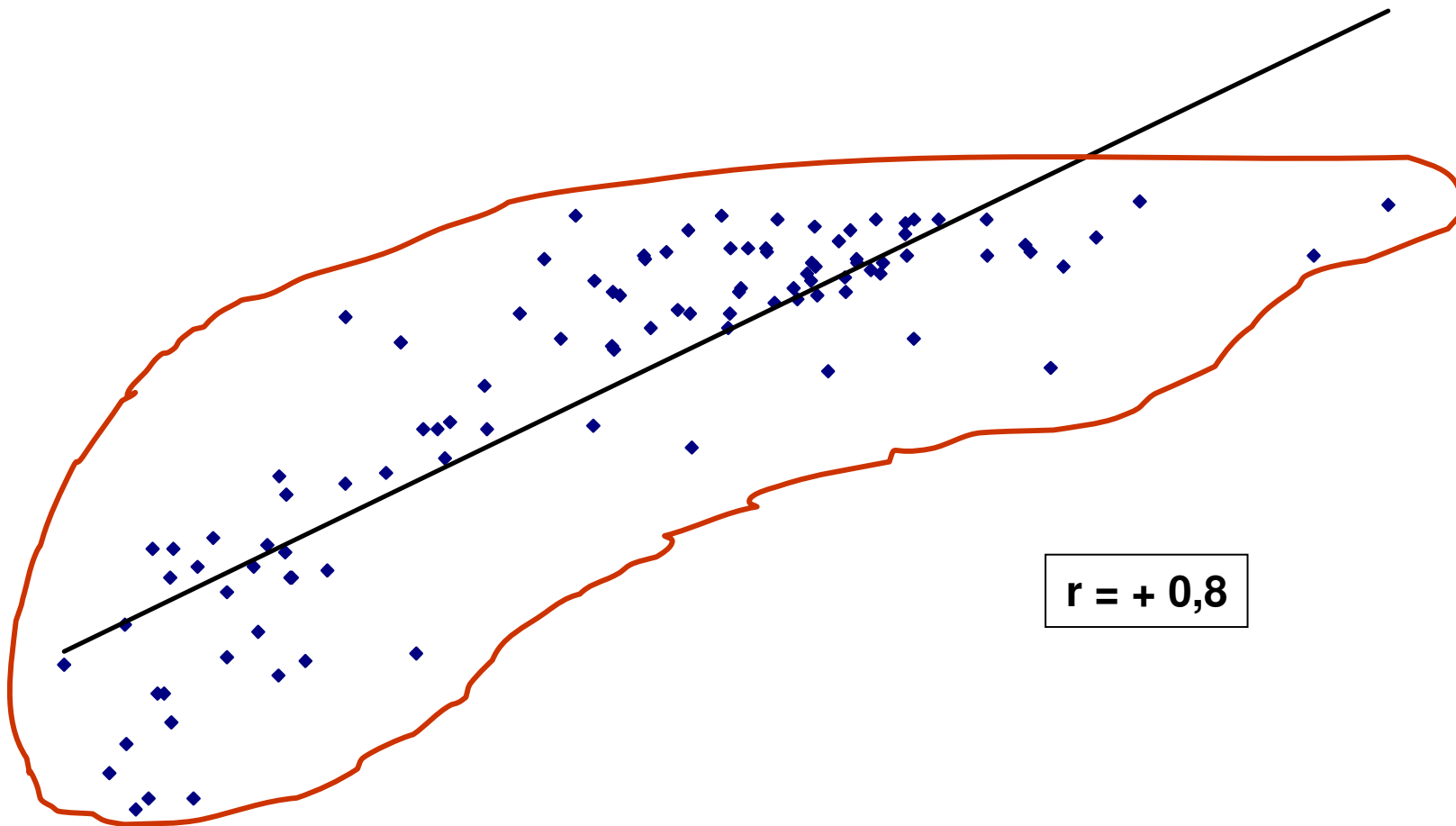
Coefficiente di correlazione lineare di Bravais Pearson

misura l'intensità del legame di tipo lineare tra due variabili



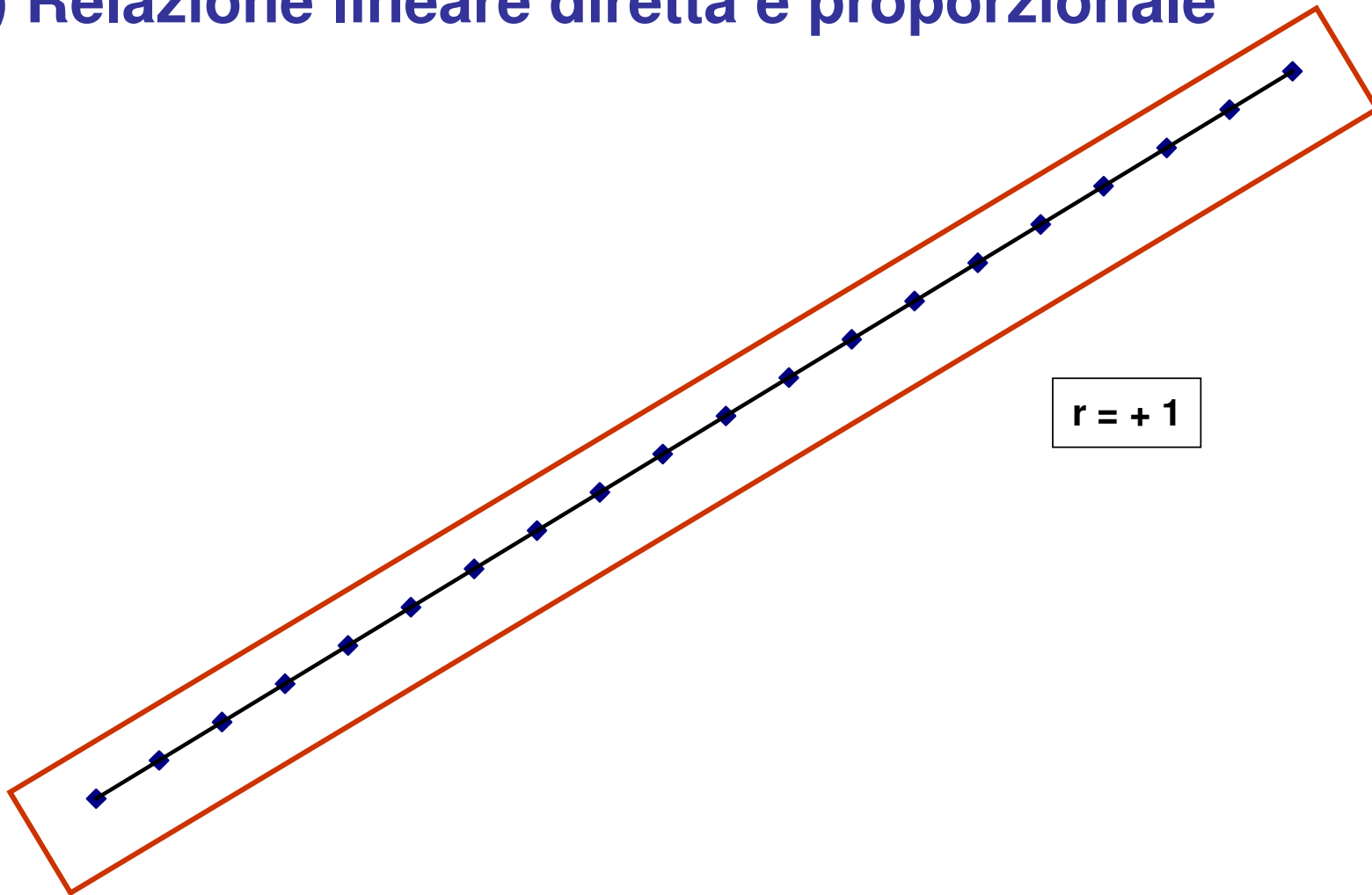
La relazione lineare - 1

a) Relazione lineare diretta e intensa



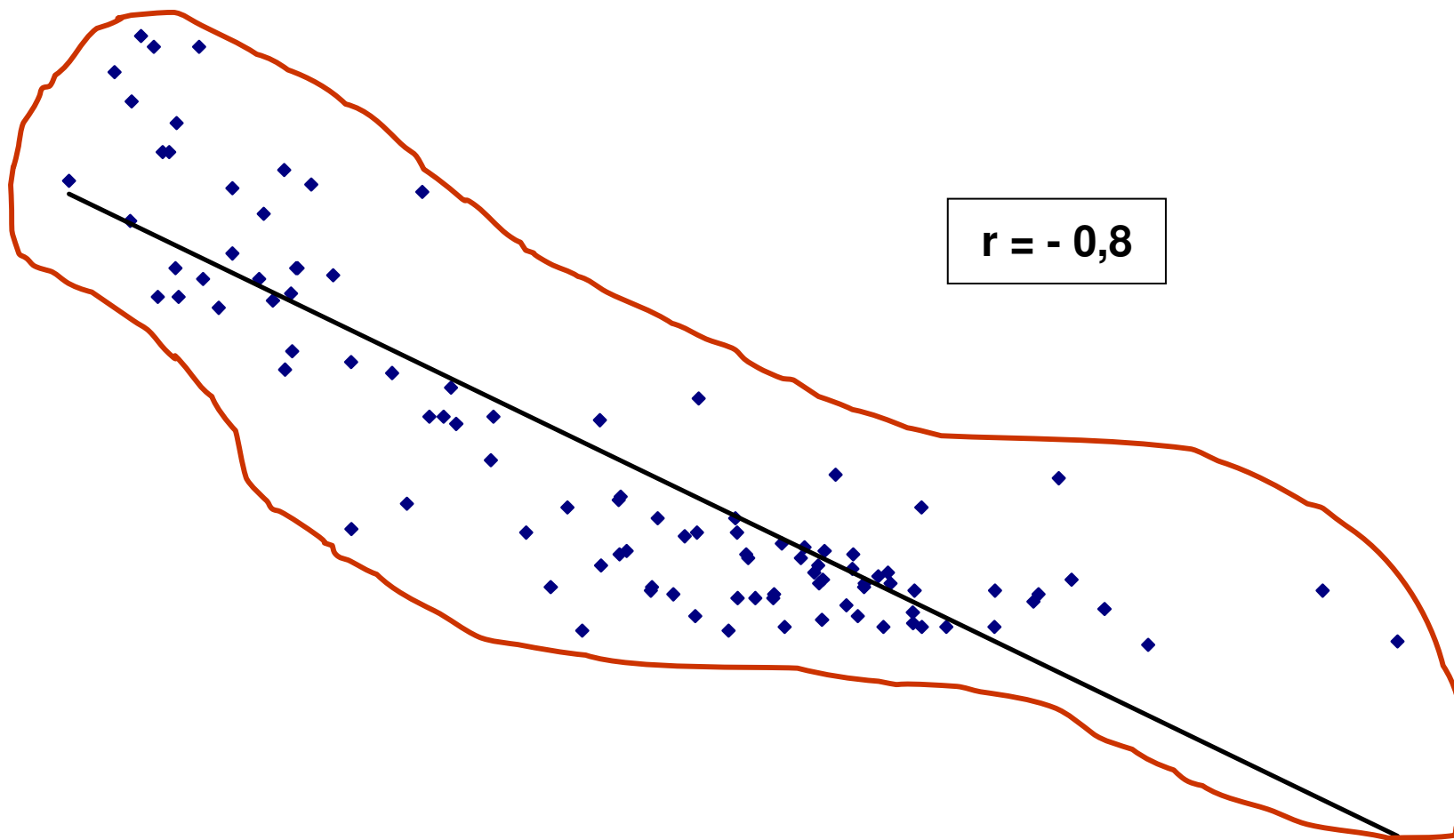
La relazione lineare - 2

b) Relazione lineare diretta e proporzionale



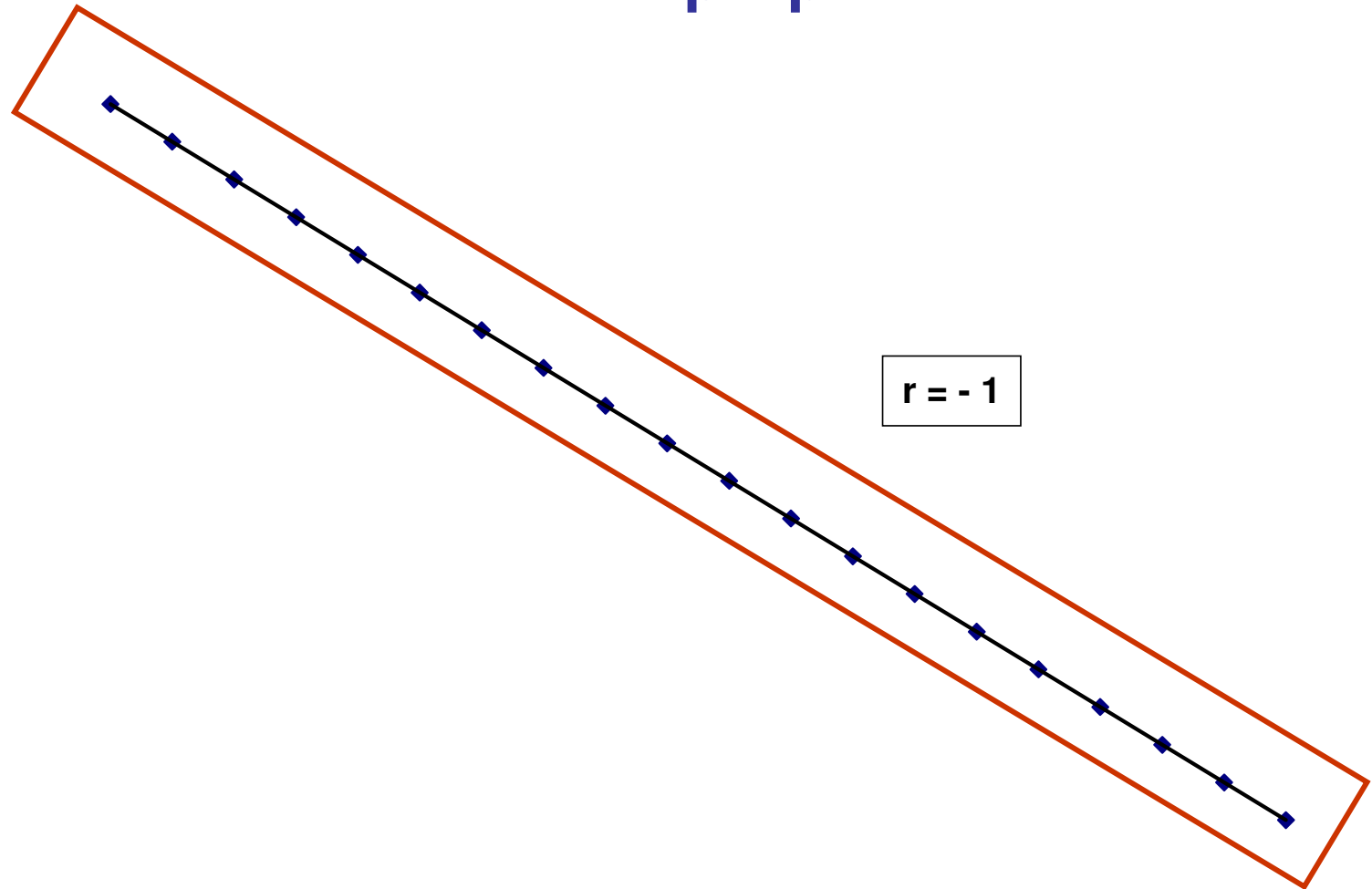
La relazione lineare - 3

c) Relazione lineare inversa e intensa



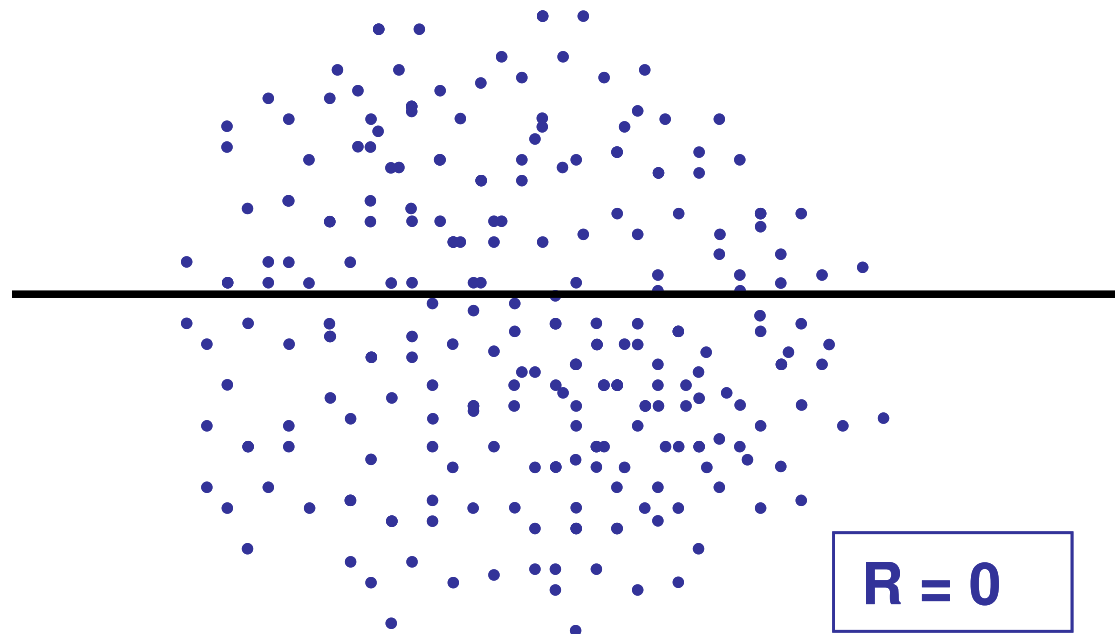
La relazione lineare - 4

d) Relazione lineare inversa e proporzionale



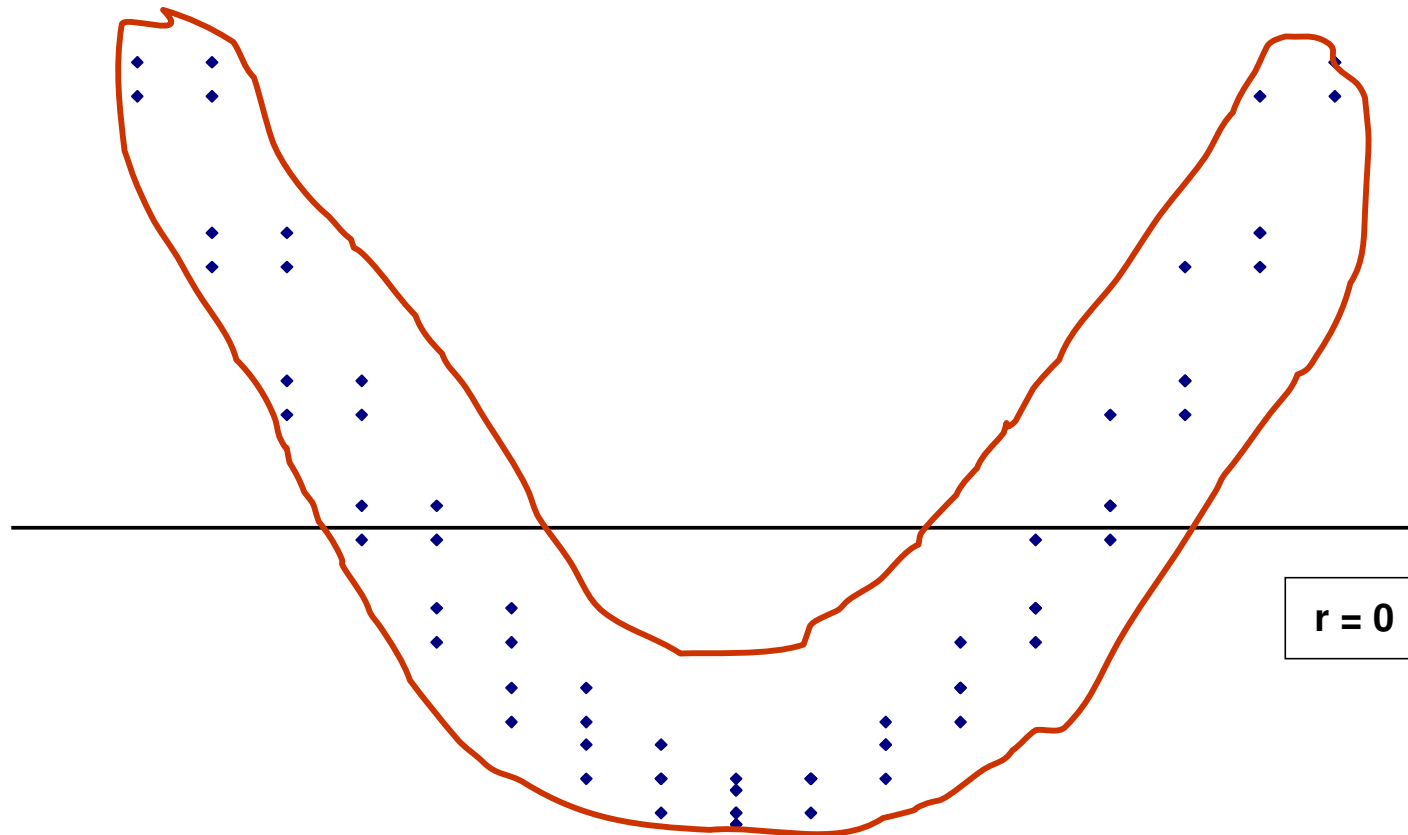
La relazione lineare - 5

e) Assenza di relazione lineare o di altro tipo



La relazione lineare - 5

e) Assenza di relazione lineare ma presenza di relazione di altro tipo



Esempio

Si vuole accertare l'esistenza di un legame lineare tra le variabili prodotto interno lordo pro capite e tasso di disoccupazione in dieci province italiane, delle quali riportiamo i valori. Misurare il coefficiente di correlazione tra le due variabili considerate.

	PIL pro capite	Tasso di disoccupazione
TORINO	24.602	4,8
MILANO	31.985	4,2
BOLOGNA	29.245	2,7
FIRENZE	28.046	4,5
ROMA	27.843	7,3
NAPOLI	13.995	17,1
L'AQUILA	16.738	5,9
BARI	14.871	13,5
PALERMO	14.347	19,2
CAGLIARI	16.455	12,9



$$\bar{x}_{pil} = \frac{\sum_{i=1}^n x_i}{n} = \frac{218.127}{10} = 21.812,7$$

$$\bar{y}_{dis} = \frac{\sum_{i=1}^n y_i}{n} = \frac{92,10}{10} = 9,21$$

Esempio - 2

Varianza e S.Q.M. del Pil pro capite (X)

	PIL pro capite	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
TORINO	24.602	2.789,3	7.780.194,49
MILANO	31.985	10.172,3	103.475.687,29
BOLOGNA	29.245	7.432,3	55.239.083,29
FIRENZE	28.046	6.233,3	38.854.028,89
ROMA	27.843	6.030,3	36.364.518,09
NAPOLI	13.995	-7.817,7	61.116.433,29
L'AQUILA	16.738	-5.074,7	25.752.580,09
BARI	14.871	-6.941,7	48.187.198,89
PALERMO	14.347	-7.465,7	55.736.676,49
CAGLIARI	16.455	-5.357,7	28.704.949,29
		0,0	461.211.350,10

$$\bar{x}_{pil} = 21.812,7$$

$$\sigma_{pil}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{461.211.350,10}{10} = 46.121.135,01$$

$$\sigma_{pil} = \sqrt{\sigma_{pil}^2} = \sqrt{46.121.135,01} = 6.791,25$$

Esempio - 3

Varianza e S.Q.M. del Tasso di Disoccupazione (Y)

	Tasso di disoccupazione.	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
TORINO	4,8	-4,41	19,45
MILANO	4,2	-5,01	25,10
BOLOGNA	2,7	-6,51	42,38
FIRENZE	4,5	-4,71	22,18
ROMA	7,3	-1,91	3,65
NAPOLI	17,1	7,89	62,25
L'AQUILA	5,9	-3,31	10,96
BARI	13,5	4,29	18,40
PALERMO	19,2	9,99	99,80
CAGLIARI	12,9	3,69	13,62
		0,0	317,79

$$\bar{y}_{dis} = 9,21$$

$$\begin{aligned}\sigma_{disl}^2 &= \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \\ &= \frac{317,79}{10} = 31,78\end{aligned}$$

$$\begin{aligned}\sigma_{disl} &= \sqrt{\sigma_{disl}^2} = \\ &= \sqrt{31,78} = 5,64\end{aligned}$$

Esempio - 4

Covarianza e correlazione tra X ed Y

	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
TORINO	2.789,3	-4,41	-12.300,81
MILANO	10.172,3	-5,01	-50.963,22
BOLOGNA	7.432,3	-6,51	-48.384,27
FIRENZE	6.233,3	-4,71	-29.358,84
ROMA	6.030,3	-1,91	-11.517,87
NAPOLI	-7.817,7	7,89	-61.681,65
L'AQUILA	-5.074,7	-3,31	16.797,26
BARI	-6.941,7	4,29	-29.779,89
PALERMO	-7.465,7	9,99	-74.582,34
CAGLIARI	-5.357,7	3,69	-19.769,91
	0,0	0,0	-321.541,54

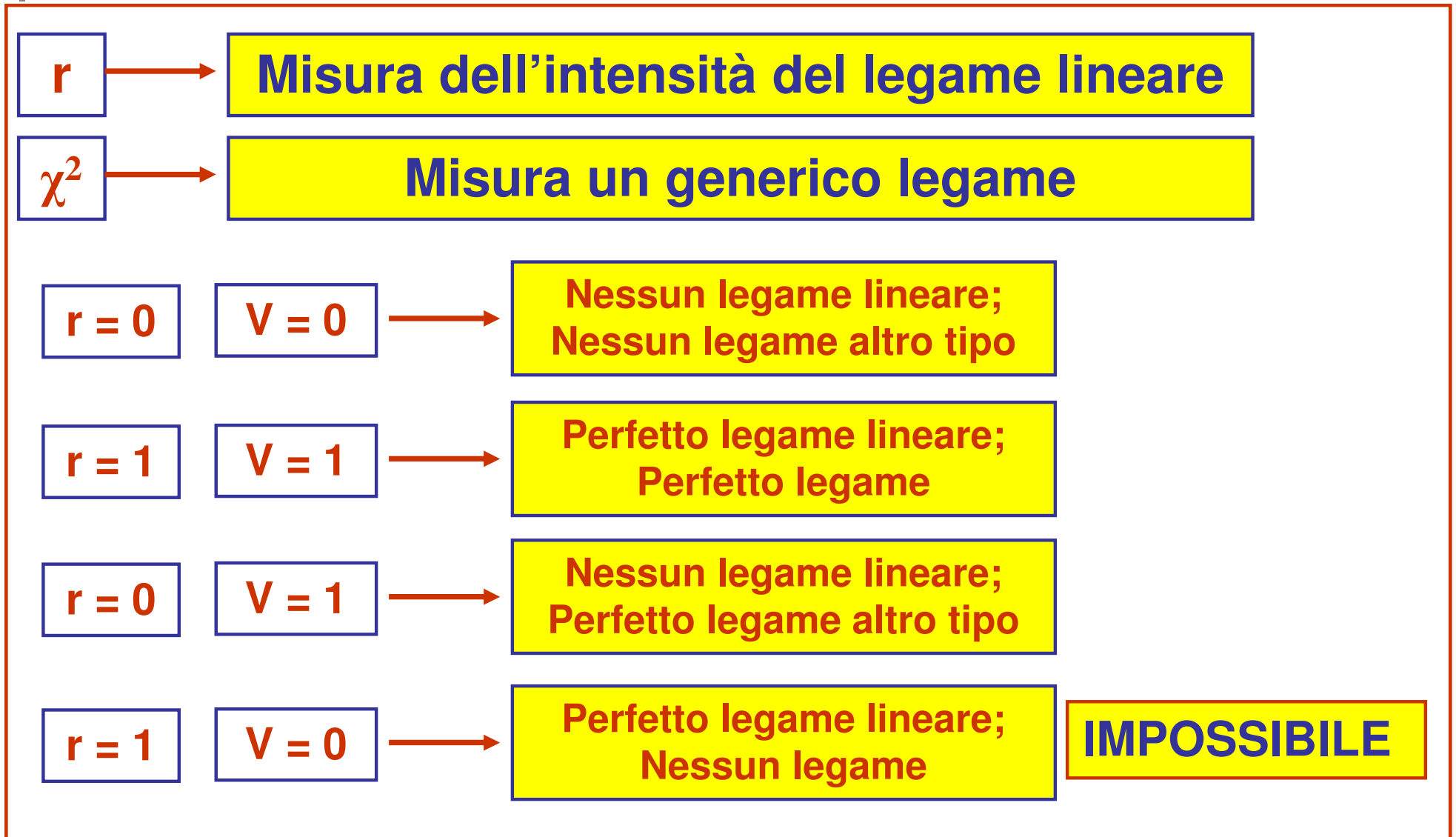
$$\bar{x}_{pil} = 21.812,7$$

$$\bar{y}_{dis} = 9,21$$

$$\begin{aligned} \sigma_{xy} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \\ &= \frac{-321.541,54}{10} = -32.154,15 \\ r &= \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{-32.154,15}{6.791,25 \cdot 5,64} = \\ &= \frac{-32.154,15}{38.302,65} = -0,839 \end{aligned}$$

**Relazione lineare
inversa molto intensa**

Correlazione e Chi-Quadrato



Correlazione – competenze acquisite

Cosa abbiamo imparato?

- Relazione lineare
- Confrontare relazione lineare con altri tipi di relazione

Quali strumenti usare?

- Per “osservare” relazione tra 2 caratteri → Dispersione
- Per misurare la relazione lineare → r