La correlazione

Lezioni di Psicometria

Giovanni Battista Flebus

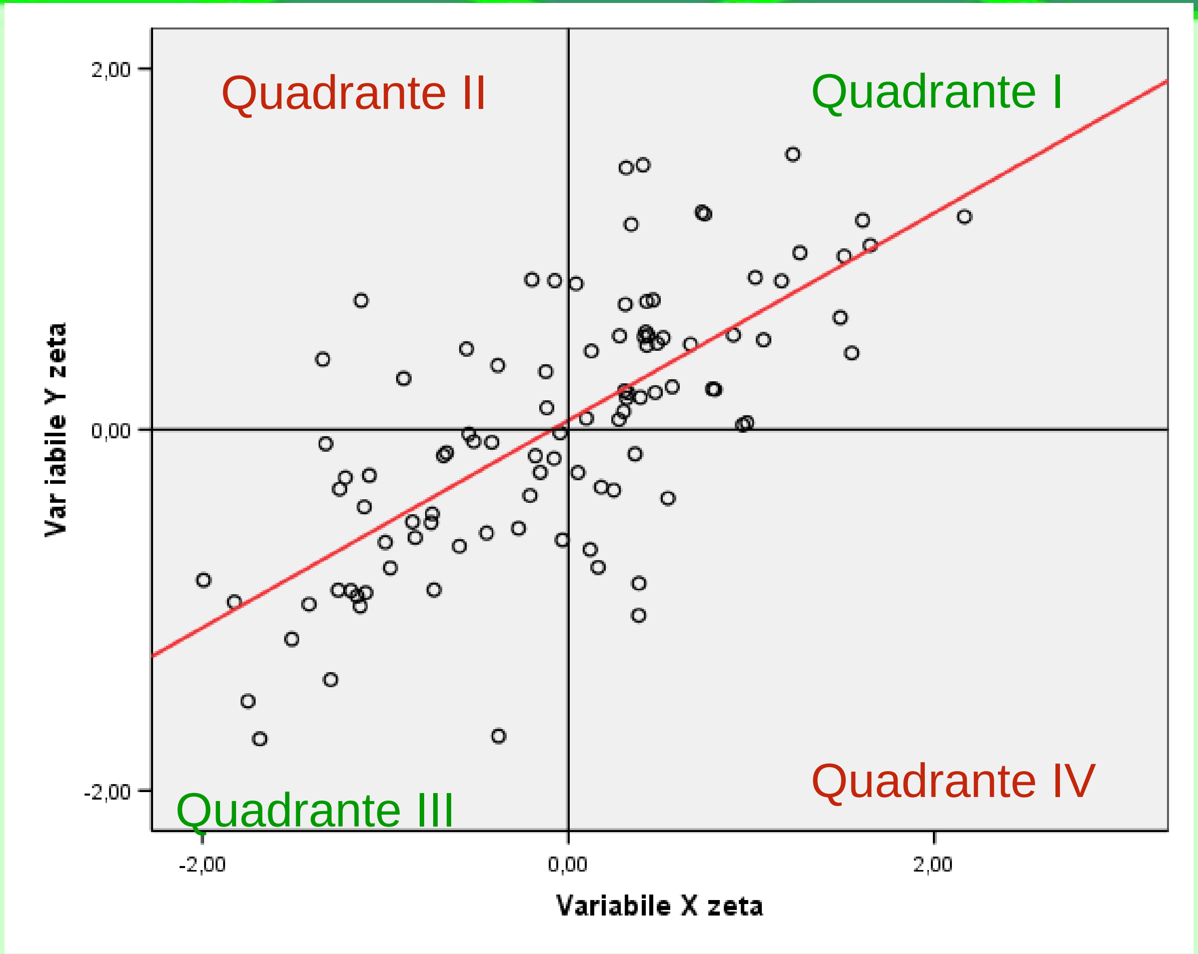
La correlazione

- È lo studio della relazione di due variabili misurate su scala a intervalli
- Può essere diretta (valori alti della prima variabile sono associati a valori alti nella seconda variabile, a valori bassi della prima sono associati valori bassi nella seconda)
- Oppure inversa (valori alti della prima variabile sono associati a valori bassi nella seconda variabile, a valori bassi della prima sono associati valori alti nella seconda

- La relazione positiva indica che all’aumentare di una aumenta anche l’altra
- la relazione negativa o inversa indica che all’aumentare dell’una l’altra diminuisce,

- Per esaminare la relazione, è comodo riportare in un grafico le coppie di osservazioni.
- Usiamo variabili standardizzate (M=0, ds =1) , in modo da eliminare l’effetto dovuta all’unità di misura.

Correlazione **positiva**



È possibile tracciare una retta che passi vicino a molti punti corrispondenti alle osservazioni compiute (ovvero che **approssimi** bene i punteggi). La retta passerà nel **primo** e **terzo** quadrante degli assi cartesiani.

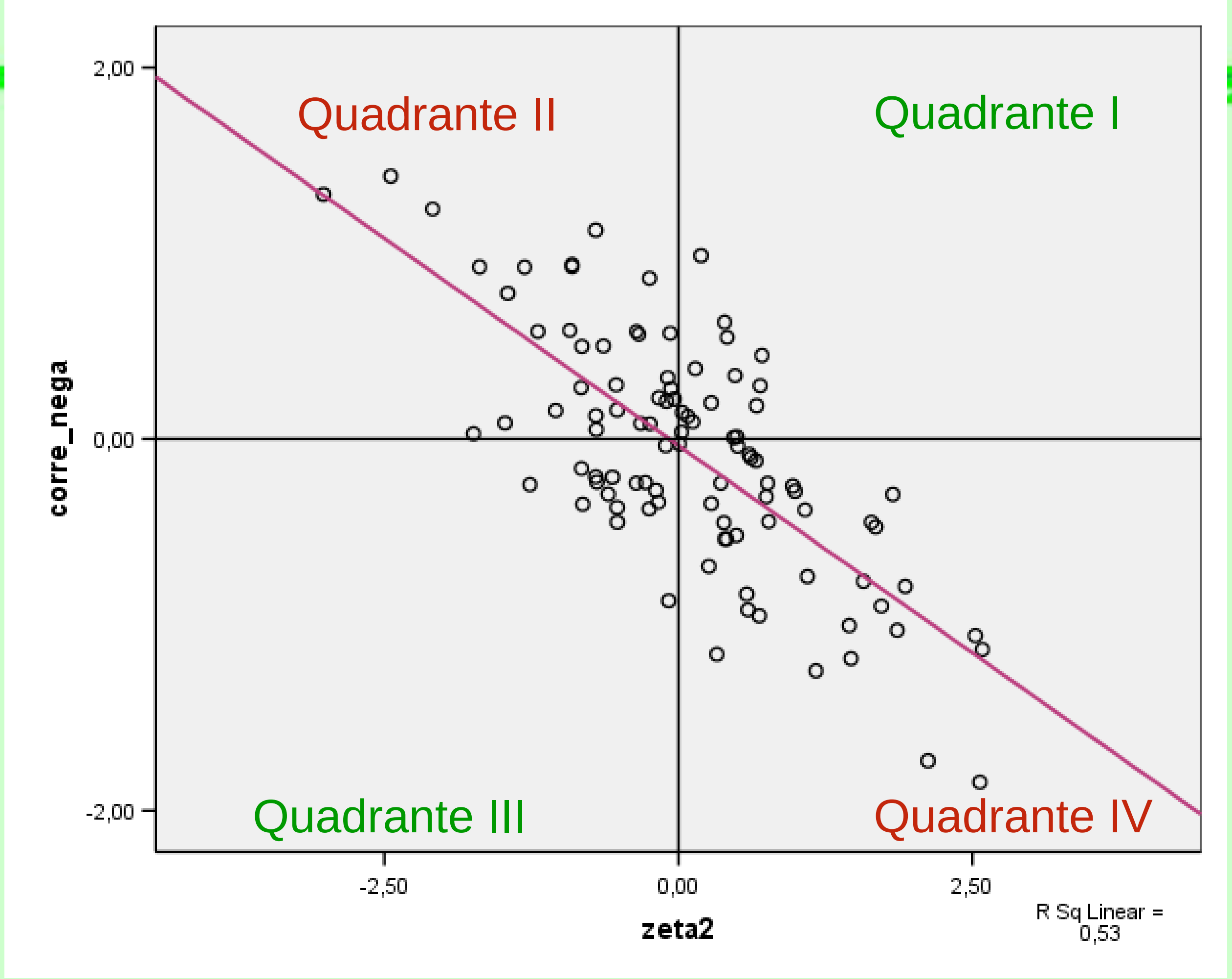
Esempi di correlazione positiva

- Peso e altezza nei bambini
- Peso e statura nei bambini
- Peso di un automobile e consumo di carburante per ogni km di spostamento

In campo psicologico:

- Abilità numerica e conoscenza di vocabolario
- Ore di studio e voti scolastici

Correlazione **negativa**



È possibile tracciare una retta che passi vicino a molti punti corrispondenti alle osservazioni compiute (ovvero che **approssimi** bene i punteggi). La retta passerà nel **secondo** e **quarto** quadrante degli assi cartesiani.

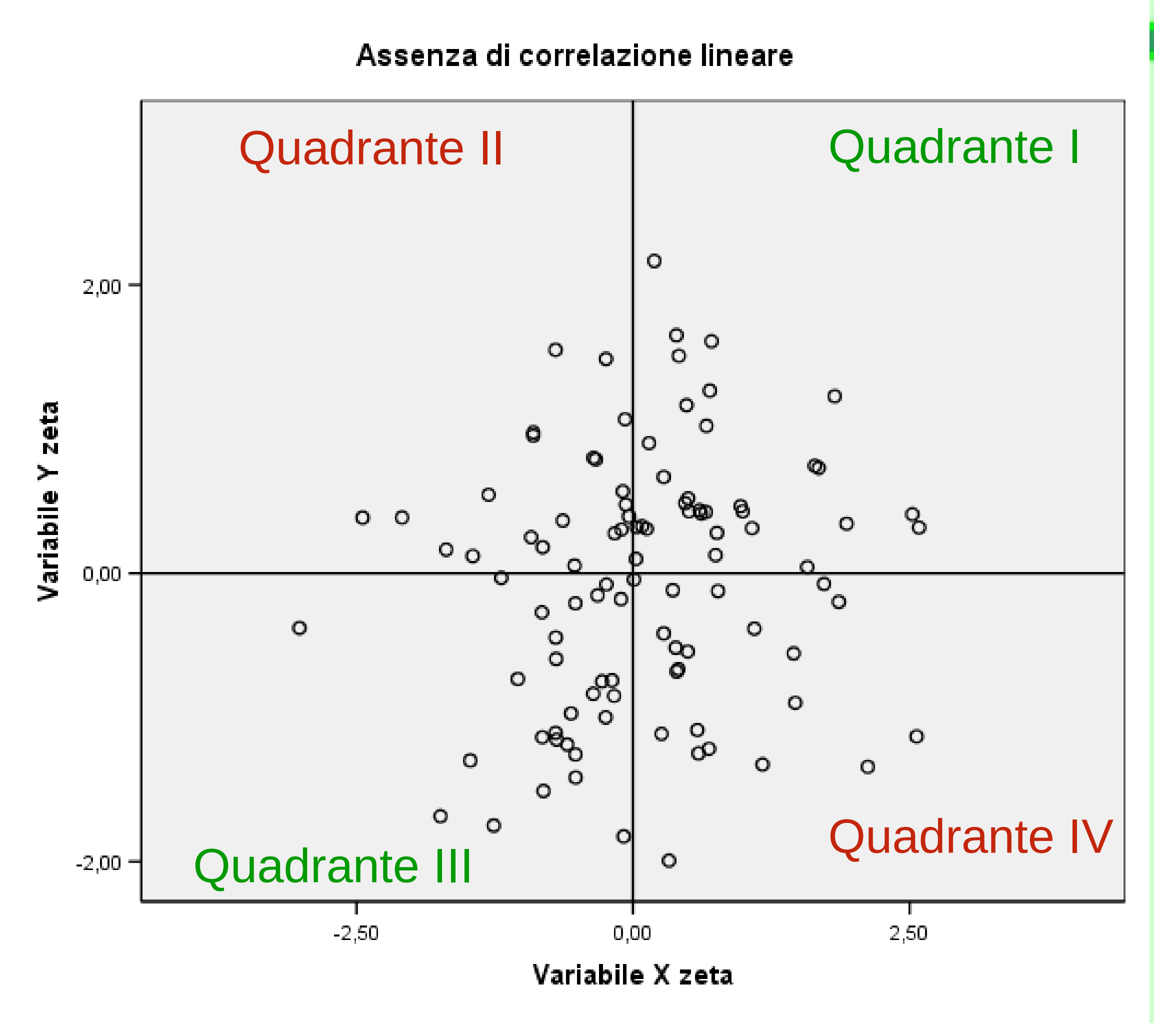
Esempi di correlazione negativa

- Anzianità di un'auto e risparmio per il trasporto
- Peso di un'auto e numero di chilometri per litro

In campo psicologico:

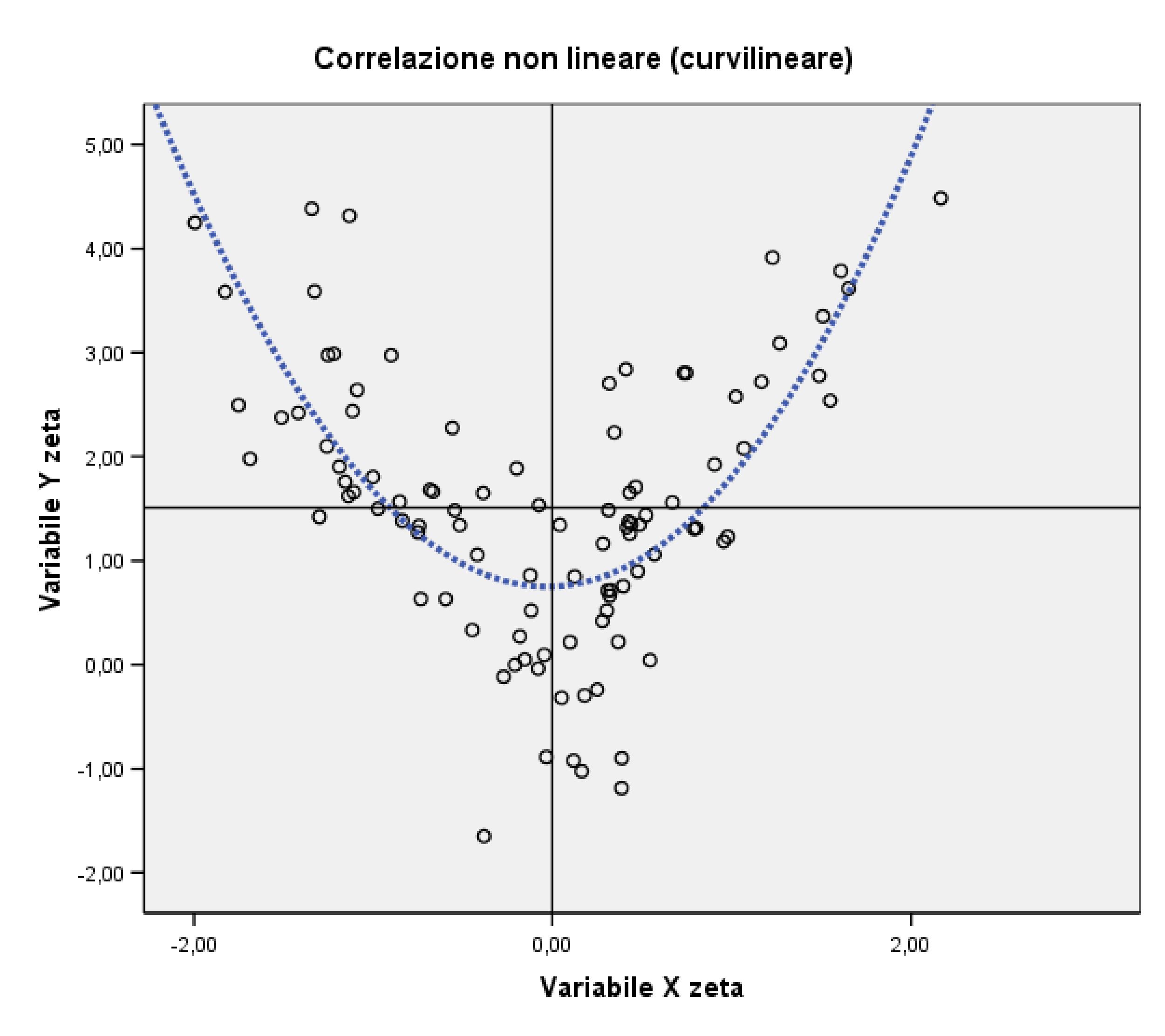
- Punteggio in un test di abilità e numero di errori commessi
- Ore di studio e ore di attività lusorie negli studenti
- Giornate di assenze da scuola e voti scolastici

Assenza di correlazione



Non è possibile tracciare una retta che passi molto vicino ai punti corrispondenti alle osservazioni compiute (ovvero che **approssimi** bene i punteggi). I punti sono sparsi più o meno equamente nei quattro quadranti degli assi cartesiani.

Correlazione non lineare (curvilineare)



Per approssimare in modo adeguato i punteggi occorre tracciare non una retta, come negli esempi precedenti, ma una curva.

Questo esempio di correlazione non sarà trattato.

Esempi di correlazione non lineare

- Gradevolezza dello zucchero in una tazza di caffè
- Sapidità dovuta al sale in una pietanza
- Motivazione ad un compito e efficienza nel compito stesso
- Età e funzioni biologiche negli umani: reazioni alle malattie, efficienza menta: basse nei bambini, crescono negli adulti, diminuiscono negli anziani
- Comportamenti antisociali negli adolescenti: crescono nella preadolescenza, hanno un picco verso i 16-18 anni e poi diminuiscono.

Perché le rette sono inclinate in questo modo?

- Proviamo a moltiplicare tra loro i segni delle variabili dipendente e indipendente nei vari quadranti:

- + = -	+ + = +
- - = +	+ - = -

0

+ → correlazione positiva

- → correlazione negativa

Come quantificare questa relazione?

Con la media dei prodotti dei punti zeta

$$r_{xy} = \frac{\sum_{i=1}^{N} z_{xi} z_{yi}}{N}$$

r_{xy} = coefficiente di correlazione di Bravais - Pearson
z_{xi} = punteggio standardizzato relativo alla variabile *x* ottenuto da un soggetto *i*
z_{yi} = punteggio standardizzato relativo alla variabile *y* ottenuto da un soggetto *i*
N = numerosità del campione

Esempio di calcolo

soggetto	Test R	Test Q
Anna	1	9
Brigida	2	7
Carlo	4	2
Delia	7	4
Enrico	8	3

Esempio di calcolo

soggetto	Test R	punto zeta R [z_x]	Test Q
Anna	1	-1,25	9
Brigida	2	-0,88	7
Carlo	4	-0,15	2
Delia	7	0,95	4
Enrico	8	1,32	3

Esempio di calcolo

soggetto	Test R	punto zeta R [z_x]	Test Q	punto zeta Q [z_y]
Anna	1	-1,25	9	1,53
Brigida	2	-0,88	7	0,77
Carlo	4	-0,15	2	-1,15
Delia	7	0,95	4	-0,38
Enrico	8	1,32	3	-0,77

Esempio di calcolo

soggetto	Test R	punto zeta R $[z_x]$	Test Q	punto zeta Q $[z_y]$	Prodotto $[z_x \cdot z_y]$
Anna	1	-1,25	9	1,53	-1,91
Brigida	2	-0,88	7	0,77	-0,67
Carlo	4	-0,15	2	-1,15	0,17
Delia	7	0,95	4	-0,38	-0,37
Enrico	8	1,32	3	-0,77	-1,01

Esempio di calcolo

soggetto	Test R	punto zeta R $[z_x]$	Test Q	punto zeta Q $[z_y]$	Prodotto $[z_x \cdot z_y]$
Anna	1	-1,25	9	1,53	-1,91
Brigida	2	-0,88	7	0,77	-0,67
Carlo	4	-0,15	2	-1,15	0,17
Delia	7	0,95	4	-0,38	-0,37
Enrico	8	1,32	3	-0,77	-1,01
somma	22	0	25	0	-3,80

Esempio di calcolo

soggetto	Test R	punto zeta R [z_x]	Test Q	punto zeta Q [z_y]	Prodotto [$z_x \cdot z_y$]
Anna	1	-1,25	9	1,53	-1,91
Brigida	2	-0,88	7	0,77	-0,67
Carlo	4	-0,15	2	-1,15	0,17
Delia	7	0,95	4	-0,38	-0,37
Enrico	8	1,32	3	-0,77	-1,01
somma	22	0	25	0	-3,80
media	4,4	0	5	0	**-0,76**

Esempio di calcolo

soggetto	Test R	punto zeta R [z_x]	Test Q	punto zeta Q [z_y]	Prodotto [$z_x \cdot z_y$]
Anna	1	-1,25	9	1,53	-1,91
Brigida	2	-0,88	7	0,77	-0,67
Carlo	4	-0,15	2	-1,15	0,17
Delia	7	0,95	4	-0,38	-0,37
Enrico	8	1,32	3	-0,77	-1,01
somma	22	0	25	0	-3,80
media	4,4	0	5	0	**-0,76**
varianza	7,44	1	6,8	1	
dev stand	2,73	1	2,61	1	

[correlazione= sommatoria($z_x \cdot z_y$)/ N]

Caratteristiche del coefficiente di correlazione di Bravais-Pearson (r_{xy})

- Varia fra -1 (correlazione negativa perfetta) e +1 (correlazione positiva perfetta).
- Se r = 0 : la correlazione lineare è assente.
- E' un indice standardizzato, e non una misura, quindi senza dimensione

Bravais e Pearson sono i due autori che lo hanno inventato in modo indipendente.

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R			Test Q	
Anna	1			9	
Brigida	2			7	
Carlo	4			2	
Delia	7			4	
Enrico	8			3	

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

N = 5

Somme necessarie per il calcolo di r con i valori grezzi

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

$$r_{xy} = \frac{N\sum XY - (\sum X) \bullet (\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2] \bullet [N\sum Y^2 - (\sum Y)^2]}}$$

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

$$r_{xy} = \frac{N\sum XY - (\sum \boxed{X}) \bullet (\sum Y)}{\sqrt{[N\sum X^2 - (\sum \boxed{X})^2] \bullet [N\sum Y^2 - (\sum Y)^2]}}$$

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

$$r_{xy} = \frac{N\sum XY - (\sum \boxed{X}) \bullet (\sum Y)}{\sqrt{\left[N\sum \boxed{X^2} - (\sum \boxed{X})^2\right] \bullet \left[N\sum Y^2 - (\sum Y)^2\right]}}$$

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

$$r_{xy} = \frac{N\sum XY - (\sum X)\bullet(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2]\bullet[N\sum Y^2 - (\sum Y)^2]}}$$

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

$$r_{xy} = \frac{N\sum XY - (\sum X)\bullet(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2]\bullet[N\sum Y^2 - (\sum Y)^2]}}$$

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

$$r_{xy} = \frac{N\sum XY - (\sum X)\bullet(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2]\bullet[N\sum Y^2 - (\sum Y)^2]}}$$

Calcolo del coefficiente r usando i dati **grezzi**

soggetto	Test R	Quadrato di R	Prodotto di R · Q	Test Q	Quadrato di Q
Anna	1	1	9	9	81
Brigida	2	4	14	7	49
Carlo	4	16	8	2	4
Delia	7	49	28	4	16
Enrico	8	64	24	3	9
somma	**22**	**134**	**83**	**25**	**159**

N = 5

$$r_{xy} = \frac{N\sum XY - (\sum X)\bullet(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2]\bullet[N\sum Y^2 - (\sum Y)^2]}}$$

Formula che usa le cinque somme

$$r_{xy} = \frac{N\sum XY - (\sum X)\bullet(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2]\bullet[N\sum Y^2 - (\sum Y)^2]}}$$

Dall'esempio precedente:

$\Sigma X=22$ $\Sigma X^2=134$

$\Sigma Y=25$ $\Sigma Y^2=159$

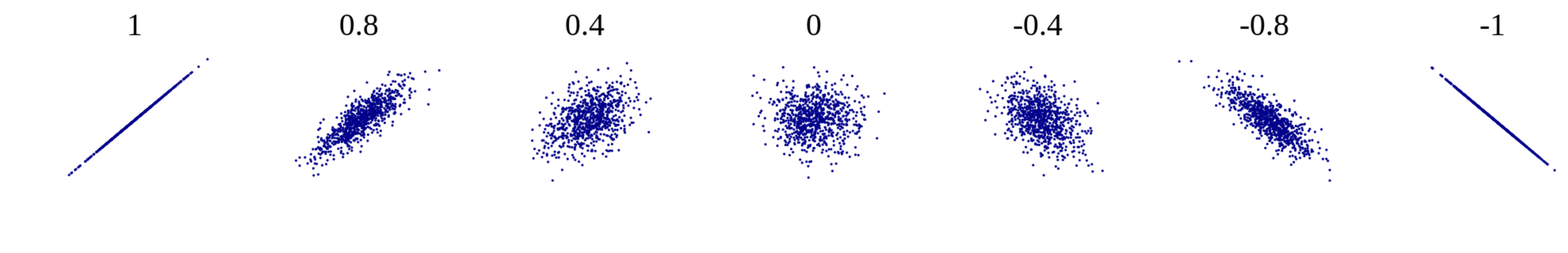
$\Sigma XY=83$

$N = 5$

Quindi, applicando la formula, troverò:

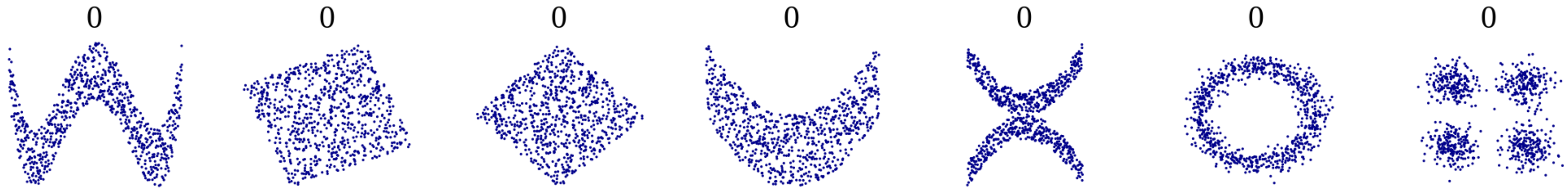
$$r_{xy} = \frac{5\cdot 83 - 22\cdot 25}{\sqrt{\left[5\cdot 134 - 22^2\right]\cdot\left[5\cdot 159 - 25^2\right]}} = \boxed{-0,7591}$$

Coefficienti di correlazione di situazioni particolari e grafici corrispondenti



Coefficienti di correlazione di situazioni particolari e grafici corrispondenti

Coefficienti di correlazione di situazioni particolari e grafici corrispondenti



Il coefficiente di determinazione

- Il quadrato del coefficiente di correlazione (r2) indica la quota di varianza comune fra le due variabili.
- Se moltiplicato per 100, indica la percentuale di varianza comune fra le due variabili.

Esempio:

- r_{xy} =-0,50, r^2_{xy} = 0,25 = 25 %

= varianza comune

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso
0,40	discreto

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso
0,40	discreto
0,50,-0,60	Buono o molto buono

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso
0,40	discreto
0,50,-0,60	Buono o molto buono
0,70	Eccellente

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso
0,40	discreto
0,50,-0,60	Buono o molto buono
0,70	Eccellente
,80	Fantastico

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso
0,40	discreto
0,50,-0,60	Buono o molto buono
0,70	Eccellente
,80	Fantastico
0,90	Sospetto

Valori di r e interpretazione possibile

I coefficienti di correlazione in psicologia sono suscettibili di questa interpretazione

<0,30	Inifluente, importante solo per ragioni teoriche
0,30	basso
0,40	discreto
0,50,-0,60	Buono o molto buono
0,70	Eccellente
,80	Fantastico
0,90	Sospetto
0,9 - 0,99	Stessa variabili, correlazione fra somme delle stesse variabili

Altri coefficienti di correlazione 1

Coefficiente punto-biseriale

Coefficiente fi (φ)

- Sono strettamente equivalenti al coefficiente di correlazione prodotto-momento di Bravais-Pearson.
- Si usano quando una (pb) o entrambe (fi) le misurazioni sono **dicotomiche**.
- La differenza di nome e di formula aveva senso quando era importante disporre di forme abbreviate di calcolo.
- Non c'è nessuna differenza nei risultati se non la formula e il nome.

Altri coefficienti di correlazione 2

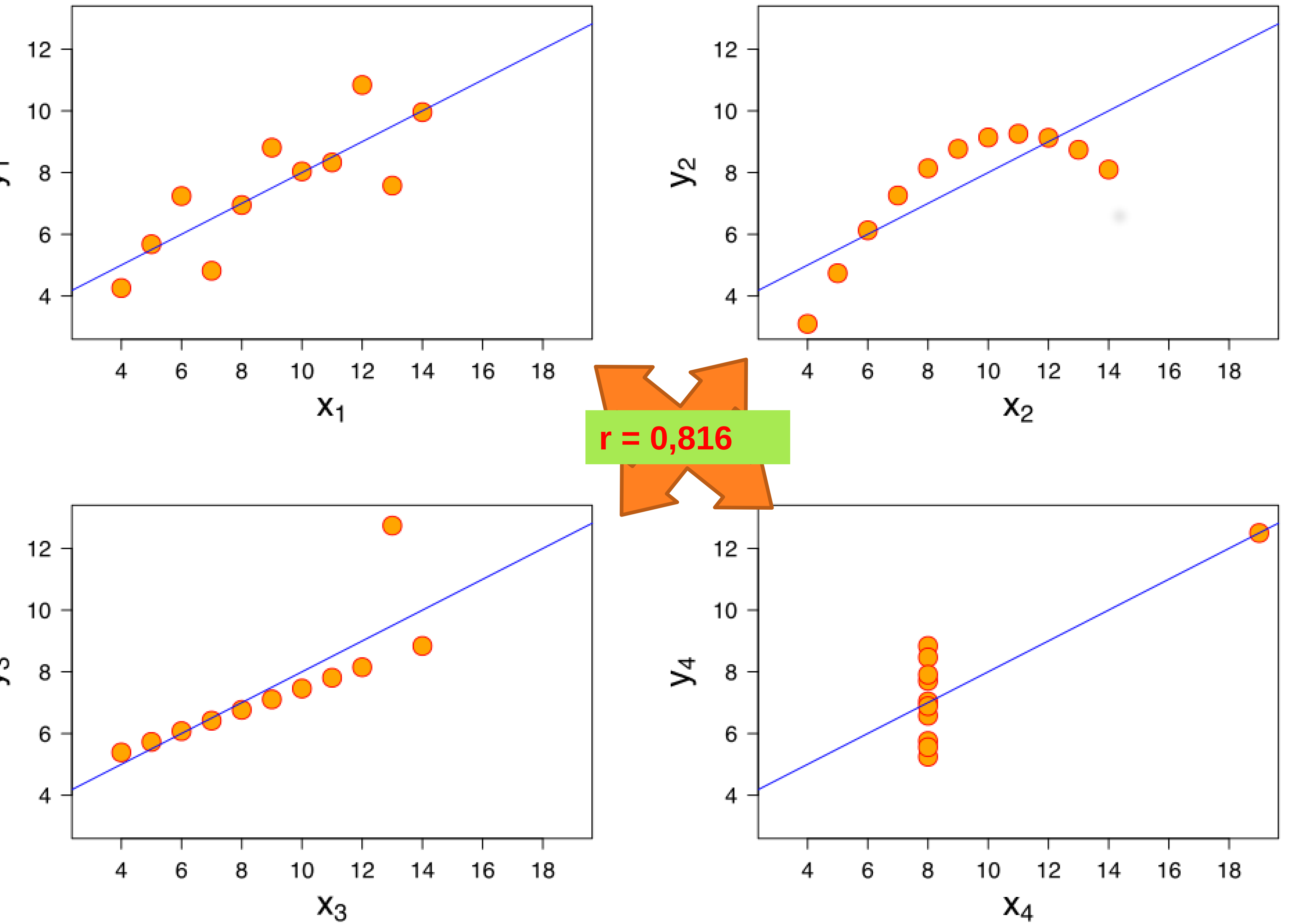
- Quando una o entrambe le variabili sono dicotomiche, ma presuppongono una distribuzione continua e normale (es: accordo per un item) si usano i coefficienti:
  - Biseriale
  - Tetracorico (entrambe dicotomiche).

Il coefficiente policorico

- Si usa con le variabili continue che presuppongono una partizione in più parti (non solo in due, come per il coefficiente tetracorico), tipica degli item di un questionario.
- Si usa generalmente nei programmi di modellistica strutturale (SEM, LISREL).
- Richiede molte centinaia di casi per il calcolo

Il quartetto di Anscombe

- Sono quattro insiemi di coppie, la cui correlazione è sempre pari a 0,816, ma hanno una relazione molto diversa fra di loro.
- Mette in evidenza la necessità di esaminare sempre il grafico dei punti per individuare valori anomali e distribuzioni particolari


r = 0,816
x_1
y_2
x_2
x_3
y_4
x_4