

LA CORRELAZIONE LINEARE

Correlazione

Nella ricerca psicologica si è spesso orientati allo studio delle relazioni che legano due variabili rilevate sullo stesso gruppo di soggetti

Alla base del concetto di relazione tra variabili c'è quello di covarianza

Date due variabili X e Y, si dice covarianza la tendenza di X e Y a “variare insieme”

Bisogna distinguere:

- a. La forma della relazione
- b. l'intensità della relazione
- c. la direzione della relazione

Correlazione

a. La forma della relazione: lineare vs. non lineare

b. L'intensità della relazione: riguarda l'entità del legame tra le due variabili. Nei casi estremi si parla di:

- associazione nulla (le variabili sono indipendenti)
- associazione perfetta

c. La direzione della relazione può essere:

- positiva: le due variabili variano nella stessa direzione (es. all'aumentare dell'una, anche l'altra aumenta)
- negativa: le due variabili variano in direzione opposta (es. all'aumentare dell'una, l'altra diminuisce)

Correlazione

Un primo modo per esaminare l'associazione tra due variabili è costruire un diagramma di dispersione

Questo tipo di grafico consente di visualizzare la tendenza di due variabili a “variare insieme”

Il diagramma di dispersione consente di collocare ogni soggetto in uno spazio bidimensionale, le cui coordinate corrispondono ai punteggi nelle due variabili (gli assi x e y del diagramma)

Il modo in cui i soggetti si dispongono (la forma della «nuvola di punti») consente una valutazione visiva dell'associazione tra le due variabili (forma, direzione e intensità)

Correlazione

Esempio 1

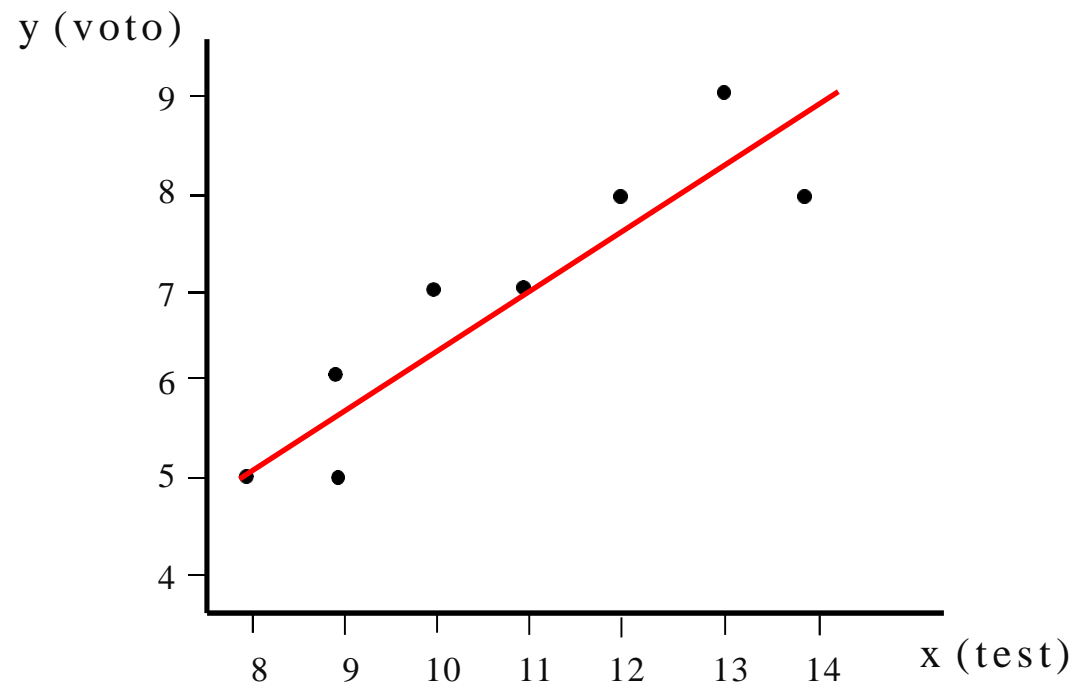
In un gruppo di 8 soggetti sono stati rilevati i risultati ottenuti in un test di attitudine per la matematica (x) e i voti ottenuti alla fine dell'anno scolastico (y)

SOGGETTO	TEST (x)	VOTO (y)
A	12	8
B	10	7
C	14	8
D	9	5
E	9	6
F	13	9
G	11	7
H	8	5

Correlazione

Esempio 1

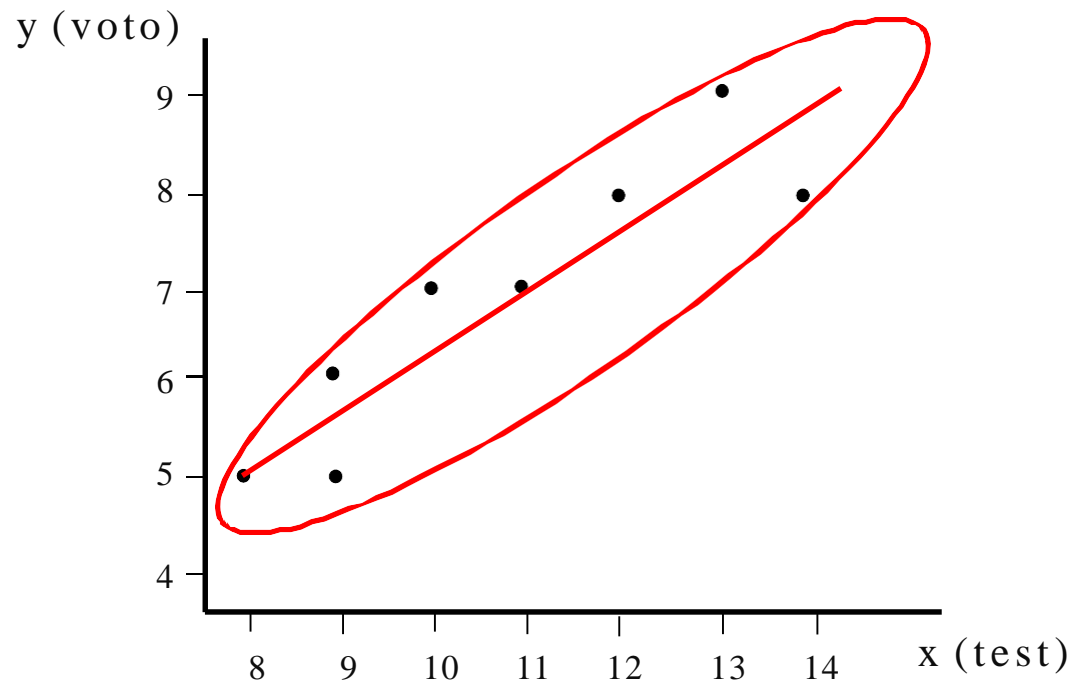
1. Forma della relazione: le due variabili sono legate da una relazione di tipo lineare (la forma della nuvola di punti è quella di una retta)



Correlazione

Esempio 1

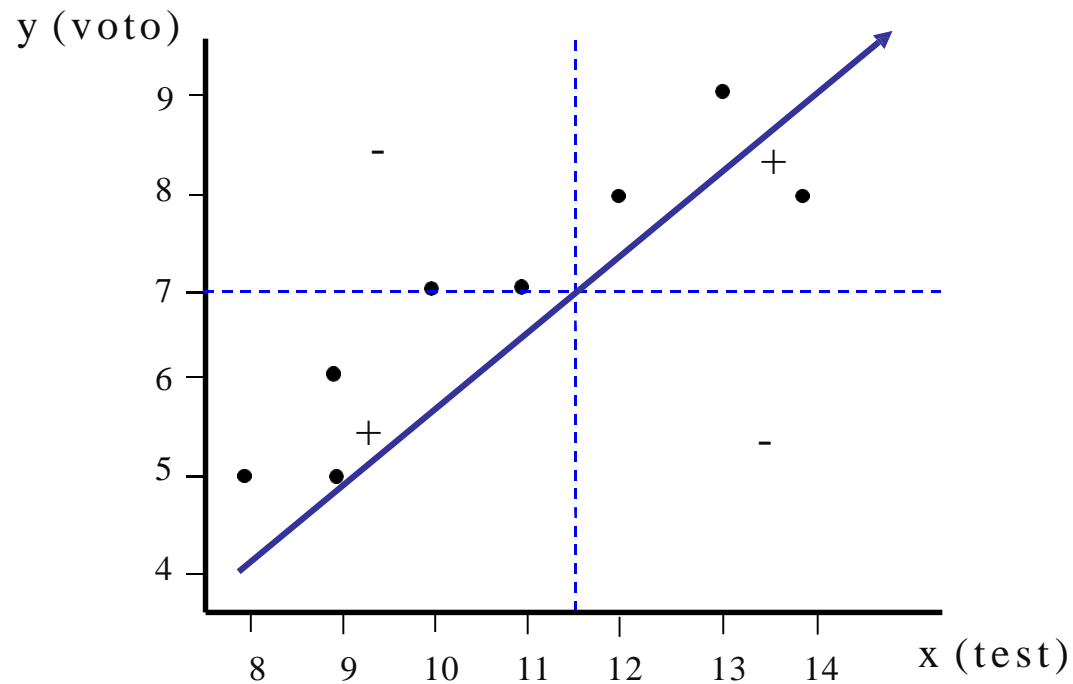
2. Intensità: la relazione è forte (i punti tendono ad addensarsi attorno all'aretta)



Correlazione

Esempio 1

3. Direzione: la relazione è positiva (i punti vanno da sinistra in basso a destra in alto). Al crescere dei valori di x aumentano sistematicamente i valori di y



Correlazione

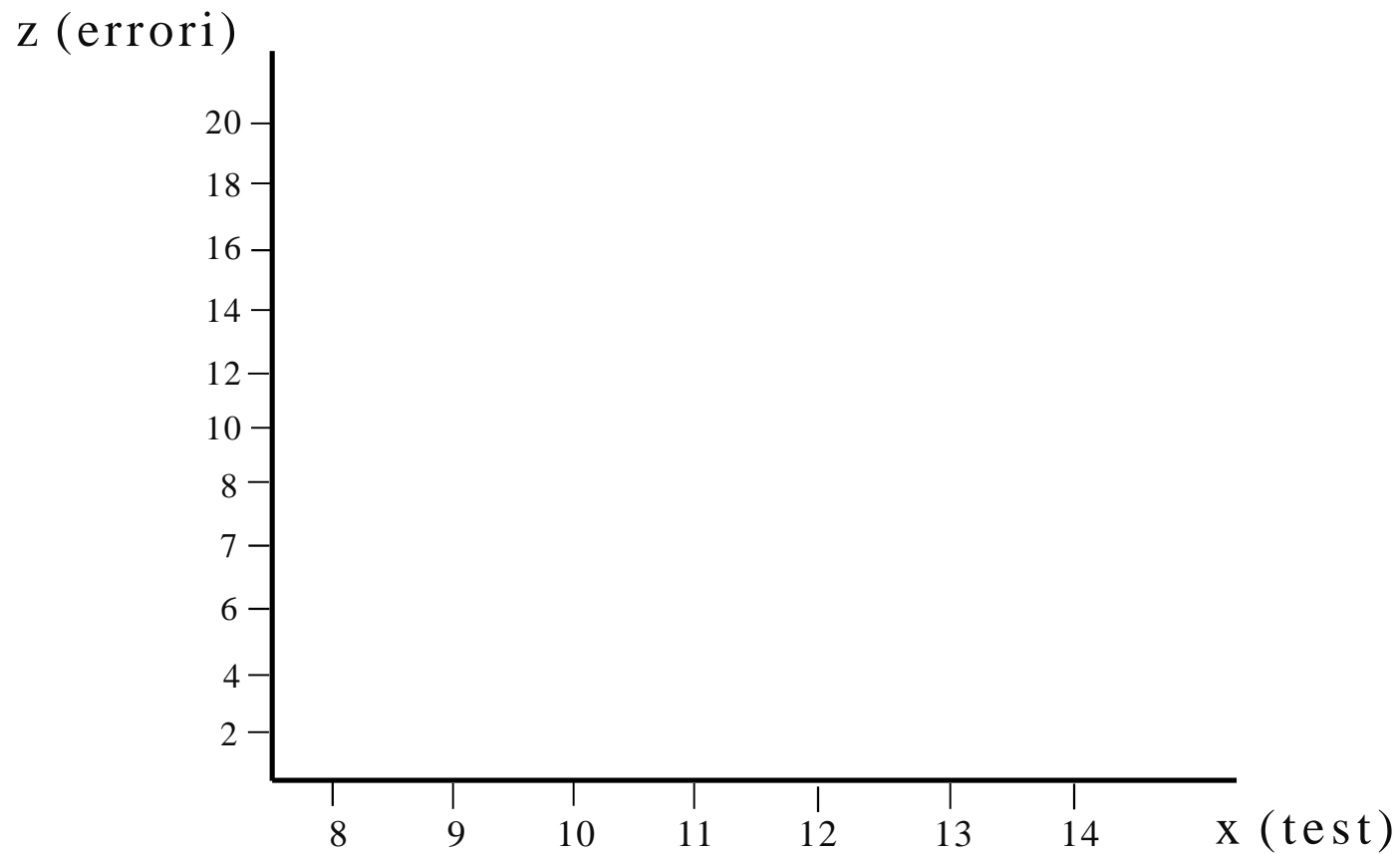
Esempio 2

Sugli stessi 8 soggetti è stato rilevato il numero di errori commessi in un'altra prova di tipo matematico (z)

SOGGETTO	TEST (x)	ERRORI (z)
A	12	10
B	10	14
C	14	6
D	9	14
E	9	18
F	13	6
G	11	16
H	8	18

Correlazione

Esempio 2 Costruiamo, come nell'esempio precedente, il diagramma di dispersione:

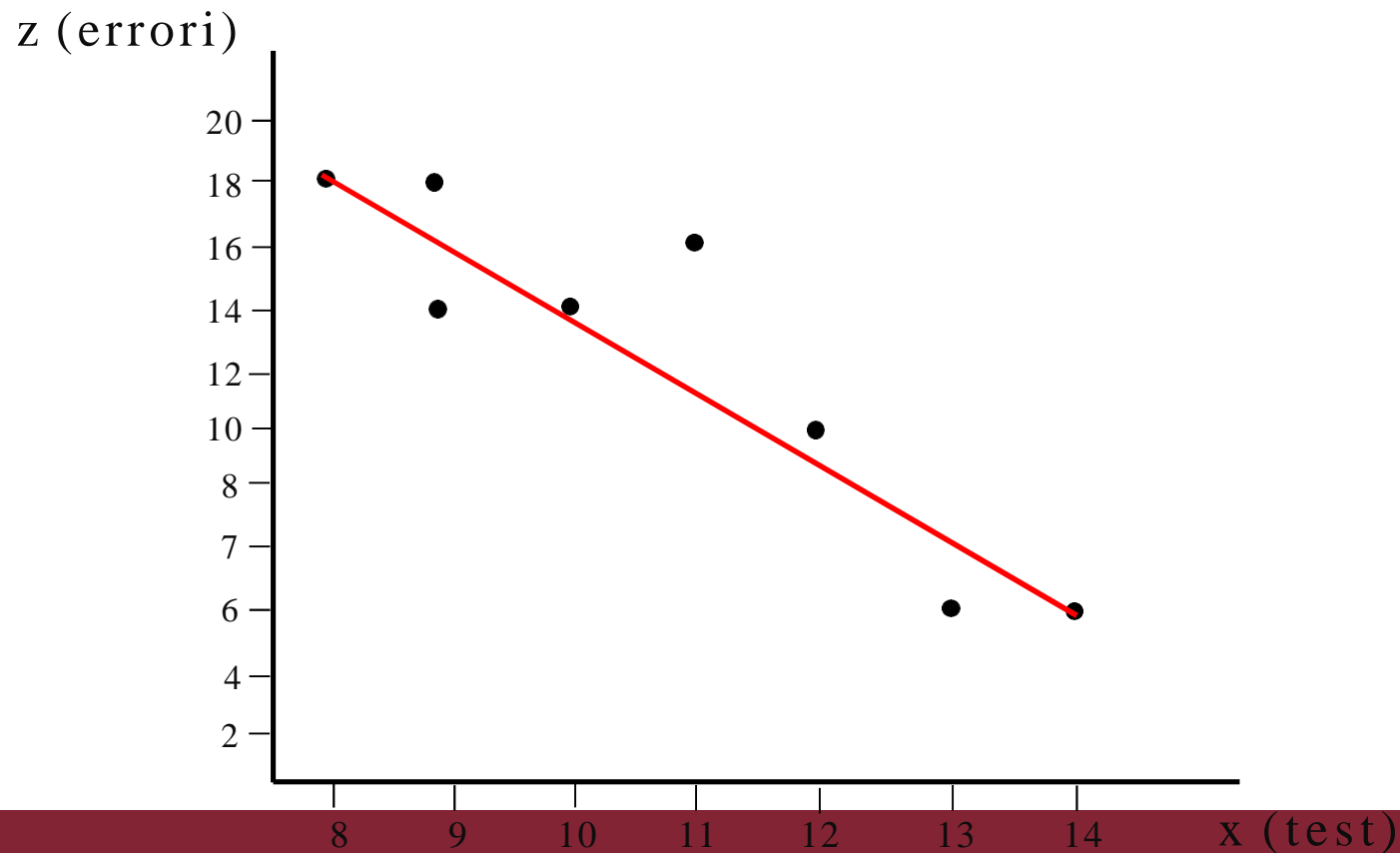


Cosa possiamo concludere?

Correlazione

Esempio 2

1. Forma della relazione: anche in questo caso, la relazione è lineare (la forma della nuvola di punti è quella di una retta)

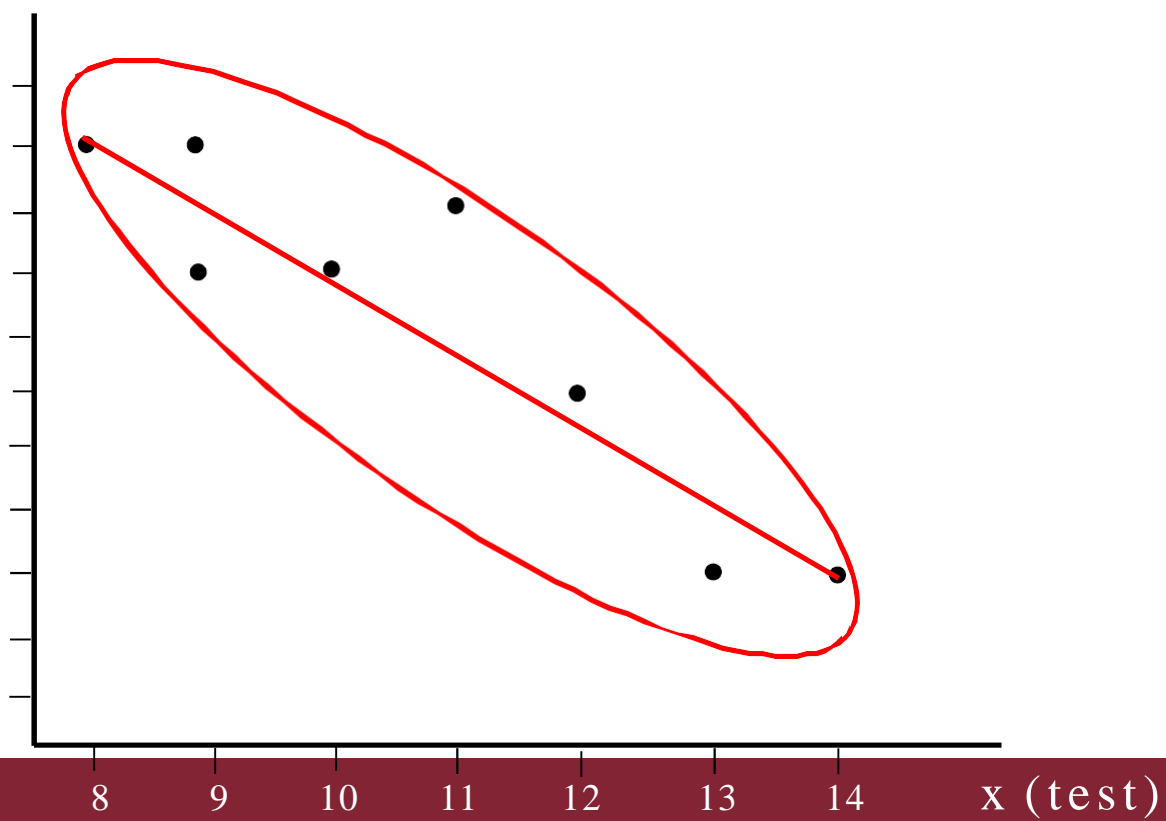


Correlazione

Esempio 2

Intensità:

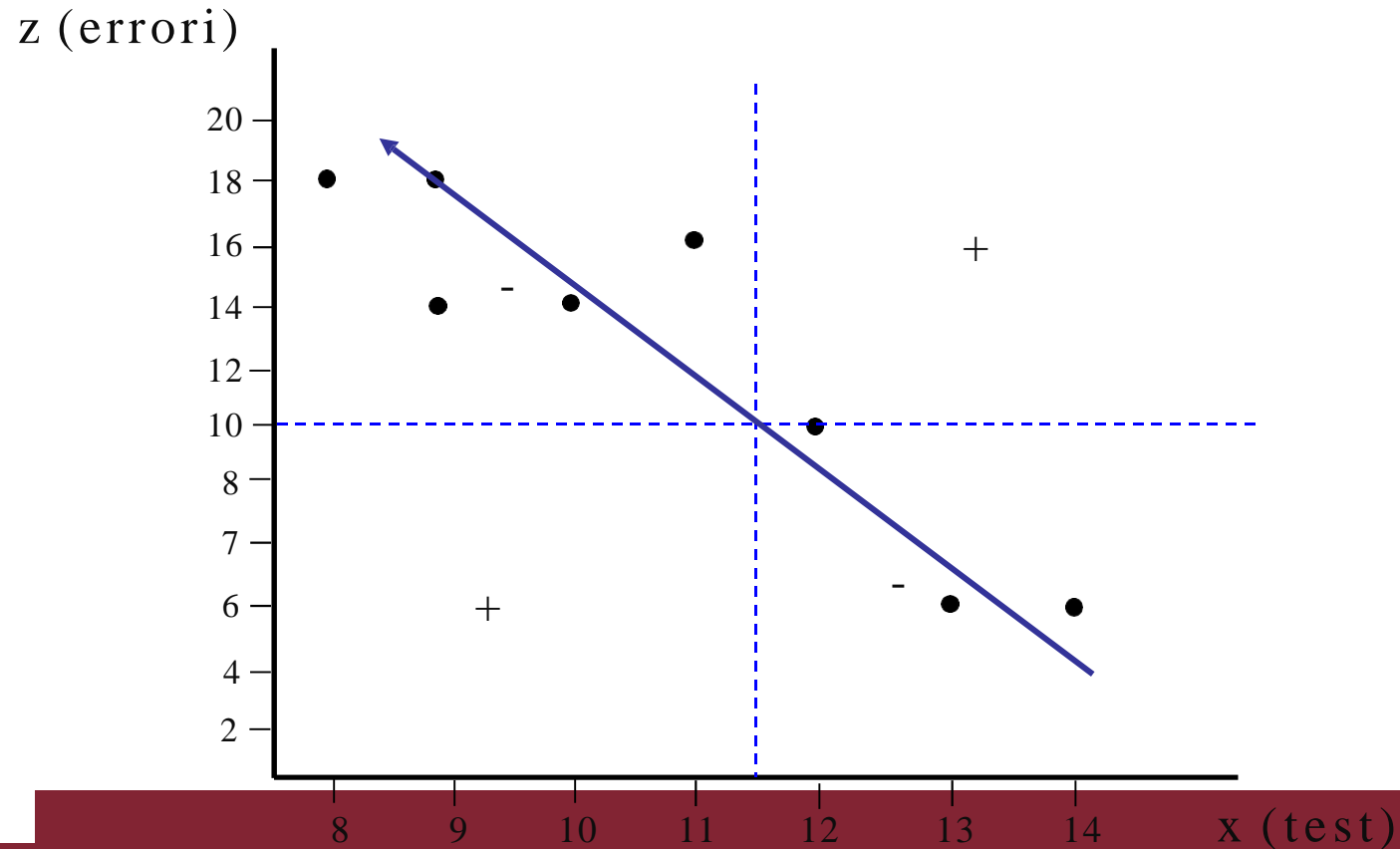
forte



Correlazione

Esempio 2

3. Direzione: la relazione è negativa (i punti vanno da sinistra in alto a destra in basso. Al crescere dei valori di x diminuiscono sistematicamente i valori di z



Correlazione

Esempio 3

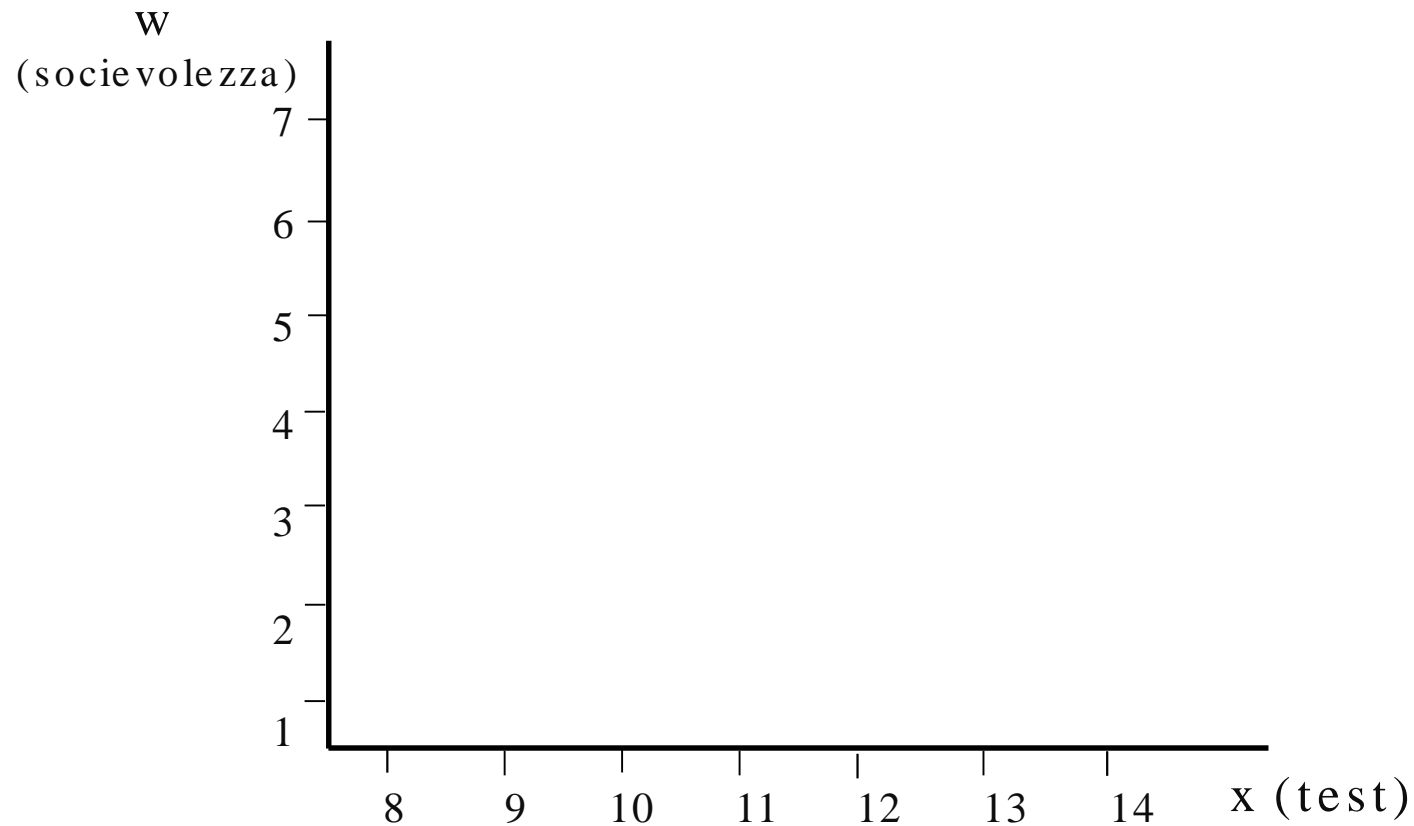
Sugli stessi 8 soggetti è stato rilevato il grado di socievolezza, tramite un apposito test (w)

SOGGETTO	TEST (x)	SOCIEVOLEZZA (w)
A	12	4
B	10	3
C	14	2
D	9	1
E	9	7
F	13	7
G	11	5
H	8	4

Correlazione

Esempio 3

Mettiamo in relazione la socievolezza (w) con il test in matematica (x) costruendo il diagramma:

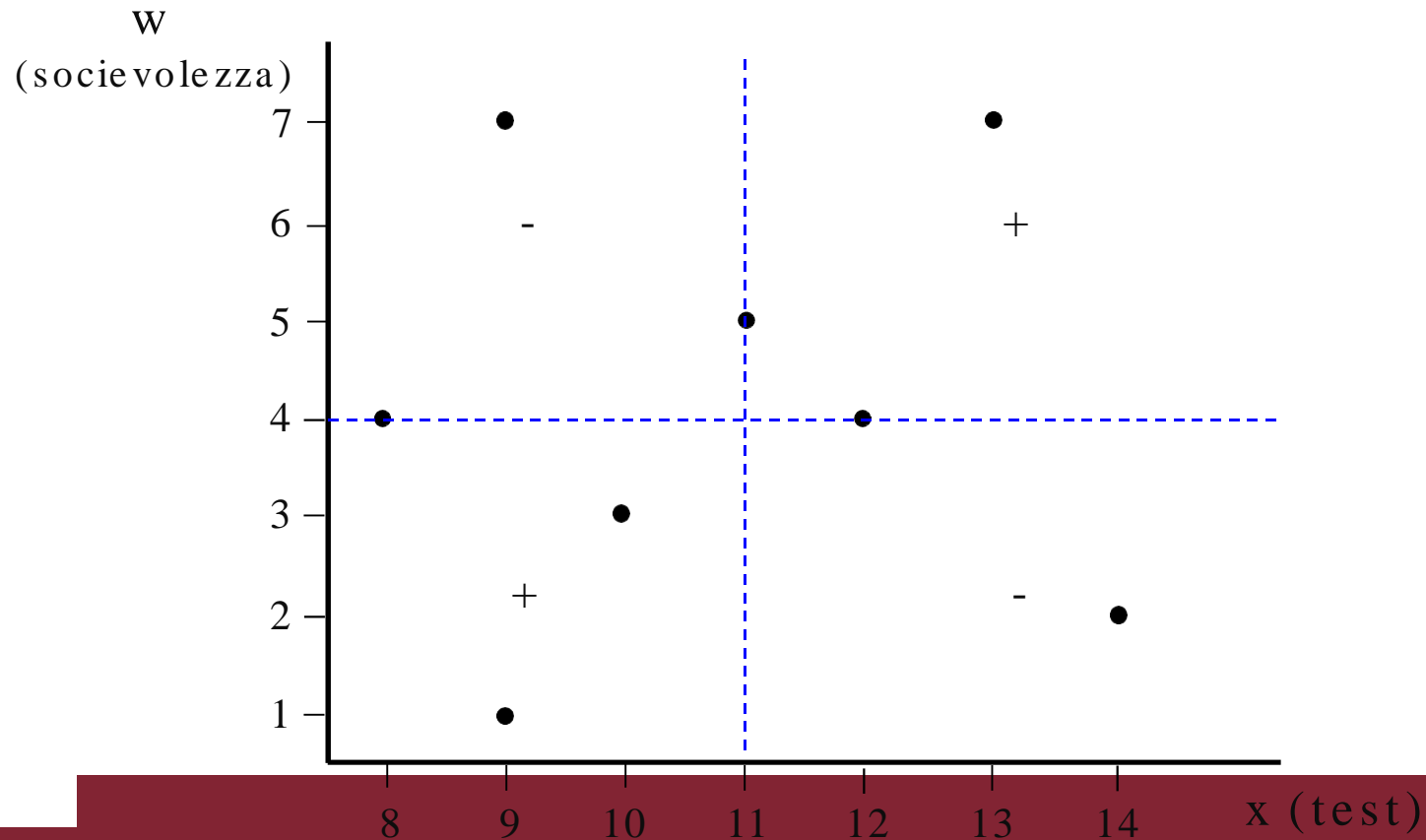


Questa volta la disposizione dei punti non suggerisce
l'esistenza di una relazione

Correlazione

Esempio 3

Al crescere dei valori di x , i valori di w oscillano indifferentemente tra incremento e decremento. Non c'è relazione tra attitudine per la matematica e socievolezza



Correlazione

Esempio 4

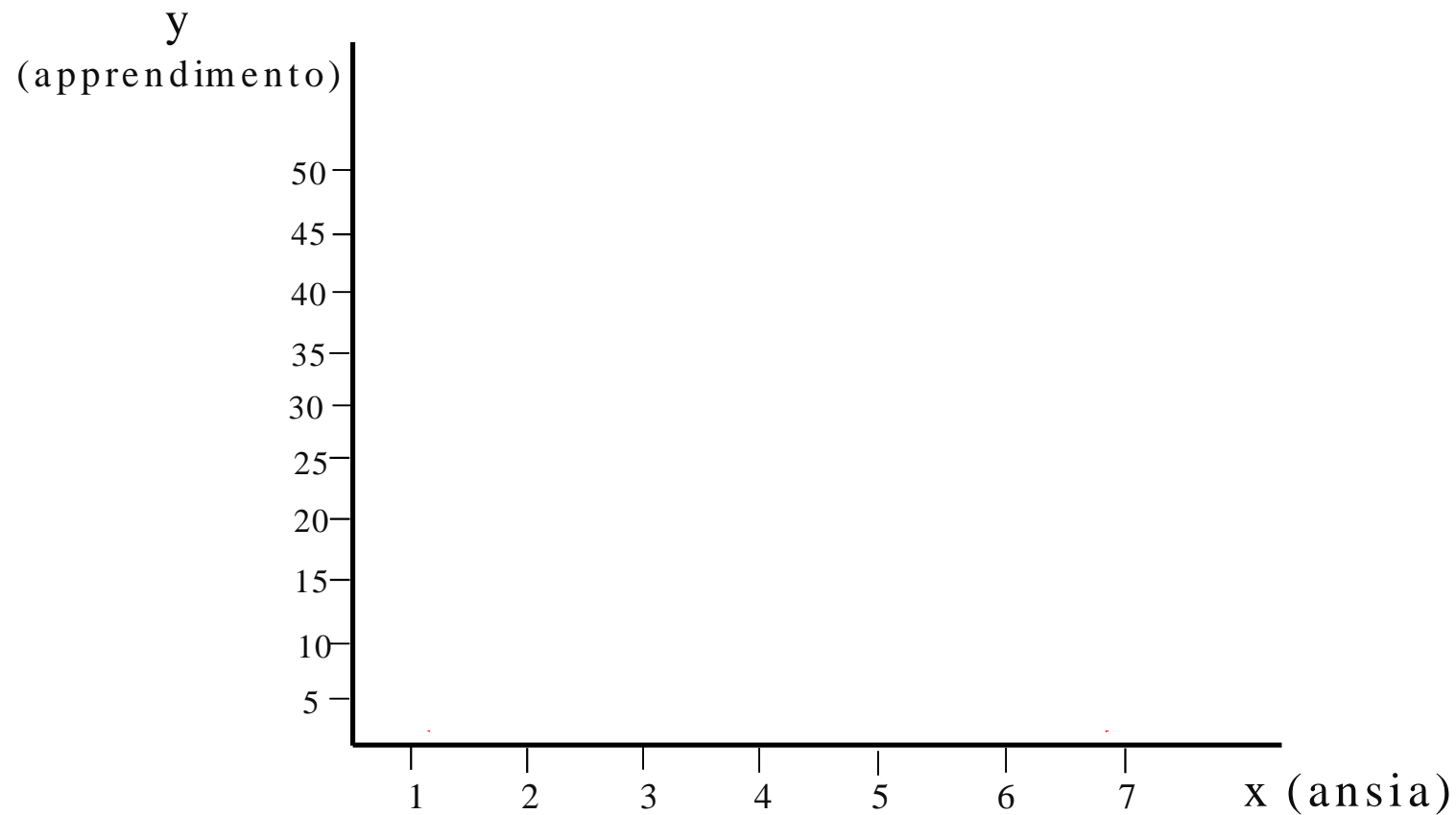
Su 12 soggetti sono state misurate l'ansia (x) e la capacità di apprendimento (y)

SOGGETTO	ANSIA (x)	APPRENDIMENTO (y)
1	1	10
2	2	20
3	3	35
4	2	25
5	4	40
6	5	35
7	3	30
8	6	25
9	7	10
10	7	15
11	6	30
12	4	45

Correlazione

Esempio 4

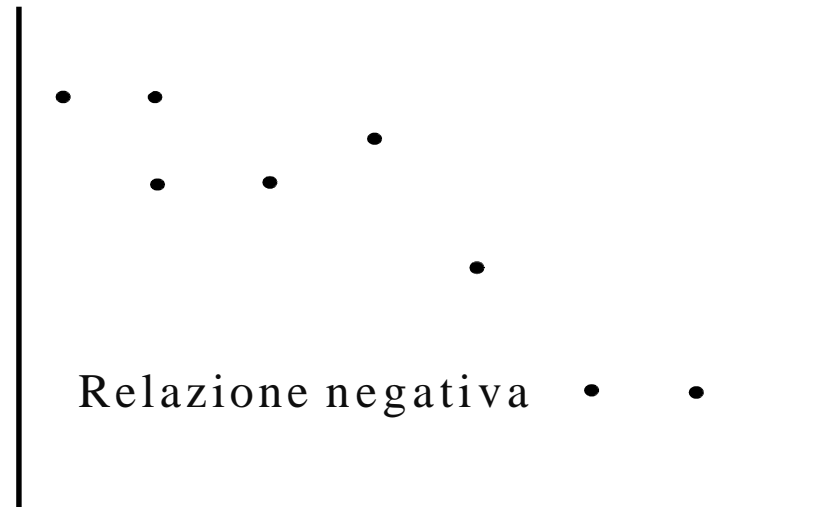
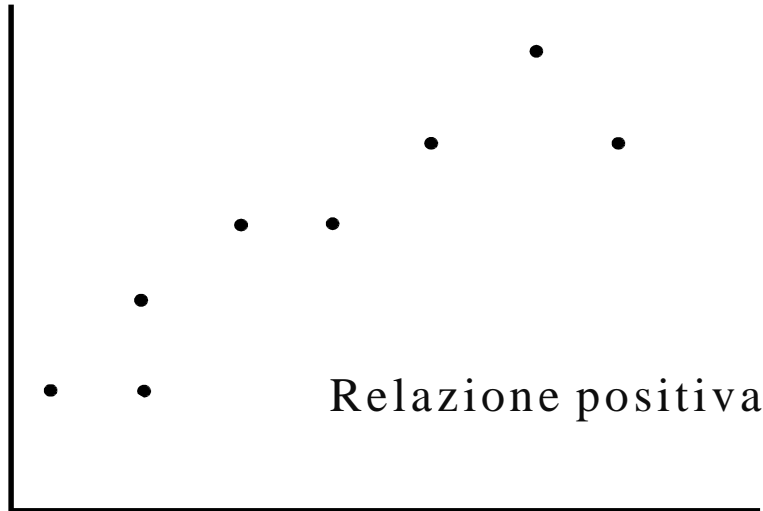
Esaminiamo la relazione tra queste due variabili, costruendo il diagramma di dispersione:



la relazione è di tipo non lineare

Correlazione

Per riassumere:



Correlazione

Il diagramma di dispersione in SPSS

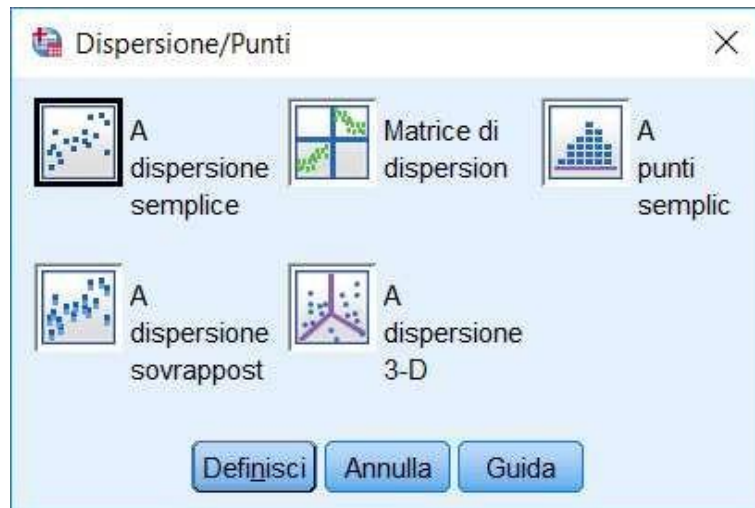
The screenshot shows the SPSS software interface. The 'Grafici' (Charts) menu is open, and the 'Dispersione/Punti...' (Scatter/Punt) option is highlighted. The background data table is as follows:

	Test	Voto	Errori	Sc
1	12,00	8,00	10,00	
2	10,00	7,00	14,00	
3	14,00	8,00	6,00	2,00
4	9,00	5,00	14,00	1,00
5	9,00	6,00	18,00	7,00
6	13,00	9,00	6,00	7,00
7	11,00	7,00	16,00	5,00
8	8,00	5,00	18,00	4,00
9				
10				
11				
12				
13				
14				
15				
16				

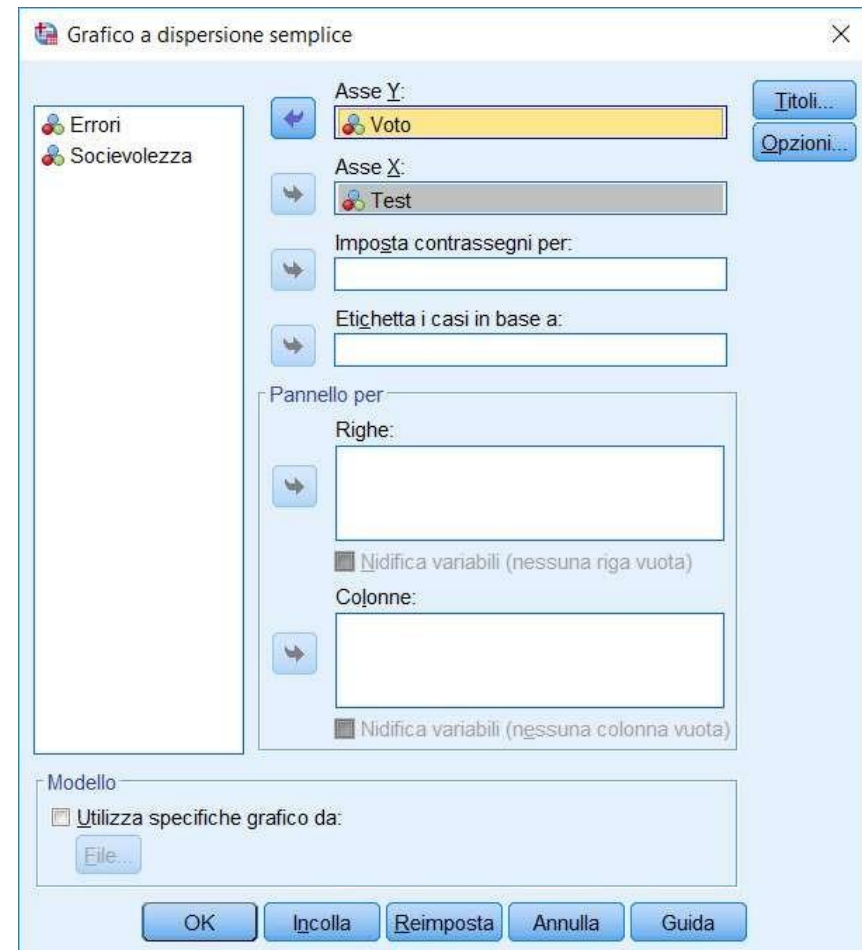
Dispersione/Punti

Correlazione

Il diagramma di dispersione in SPSS



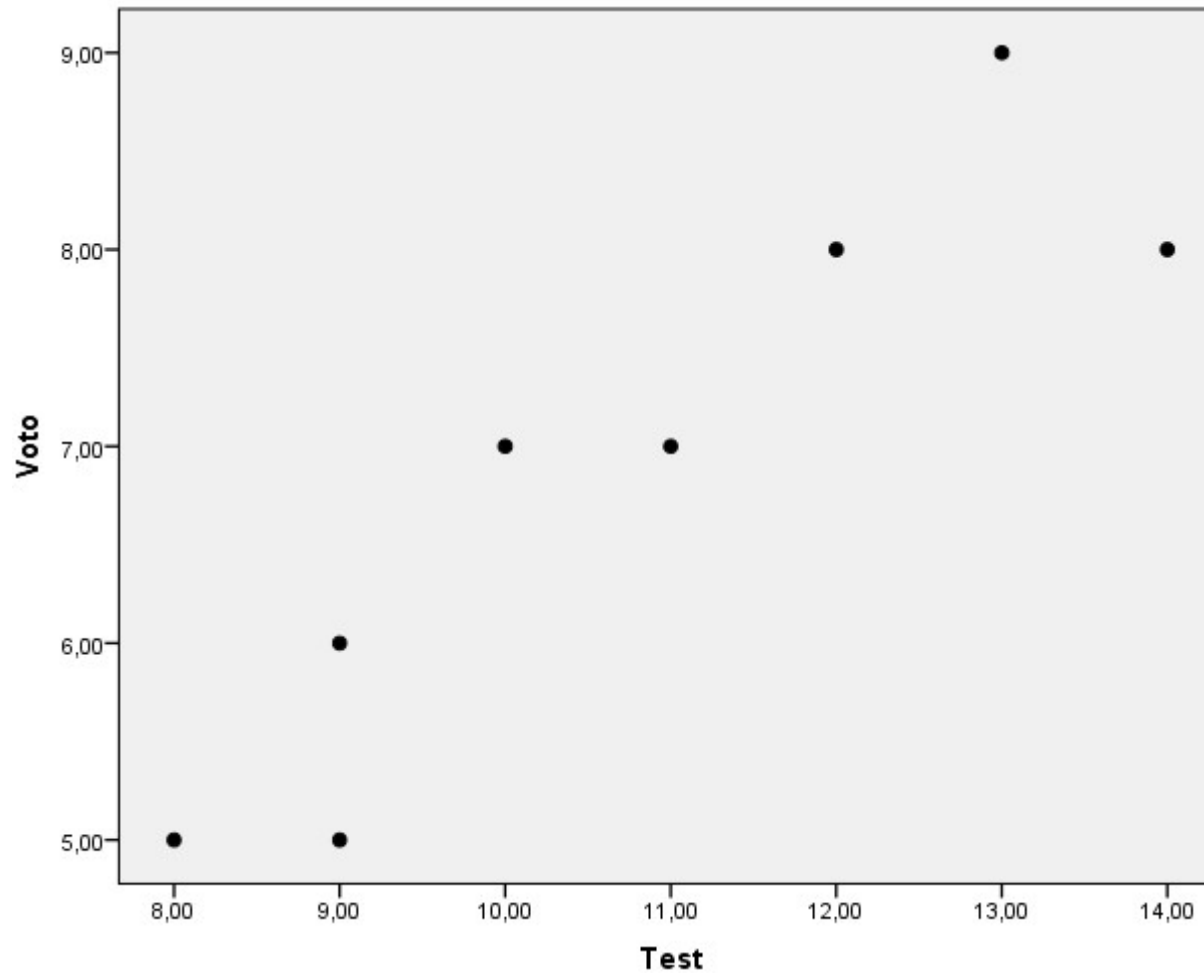
A dispersione semplice



Voto) e Asse X» (es. Test)

Correlazione

Il diagramma di dispersione in SPSS



Correlazione

Il coefficiente di correlazione

Per quantificare l'associazione tra due variabili è possibile calcolare un coefficiente

Il coefficiente di correlazione è un “numero” che sintetizza la relazione le due variabili in termini di intensità e direzione

La maggior parte dei coefficienti di correlazione (es. Pearson) consentono di cogliere solo relazioni di tipo lineare

Correlazione

Il coefficiente di correlazione

Il segno indica la direzione dell'associazione:

- positivo quando le due variabili aumentano (o diminuiscono) insieme
- negativo quando all'aumento dell'una corrisponde una diminuzione dell'altra (e viceversa)

Il valore assoluto indica l'intensità della relazione
(varia da 0 a 1):

- è pari a 0 quando l'associazione è nulla
- aumenta al crescere dell'intensità della relazione
- è pari ad 1 quando c'è un'associazione perfetta tra due le variabili

Correlazione

Il coefficiente di correlazione

E' importante ricordare che:

-un valore assoluto basso o nullo di correlazione non va interpretato come assenza di una qualsiasi forma di relazione tra le due variabili: è assente solo una relazione di tipo lineare

-un'elevata correlazione fra due variabili NON implica una relazione causa-effetto

Correlazione

Tipi di coefficienti

- r di Pearson:
- rho di Spearman (r_s):
- tau di Kendall:
- r phi:
- r punto-biseriale:

Il coefficiente di correlazione r di Pearson

Correlazione

Il coefficiente di correlazione
r di Pearson

-Si applica a variabili misurate su scala a intervalli o a rapporti equivalenti (variabili continue)

-È chiamato coefficiente di correlazione di Pearson perché è stato presentato per la prima volta da Karl Pearson nel 1895

-Viene chiamato anche coefficiente di correlazione di Bravais-Pearson, per ricordare lo statistico Bravais (1846), che aveva presentato alcuni concetti importanti di tale metodo cinquanta anni prima di Pearson

- Viene indicato con il simbolo r

Correlazione

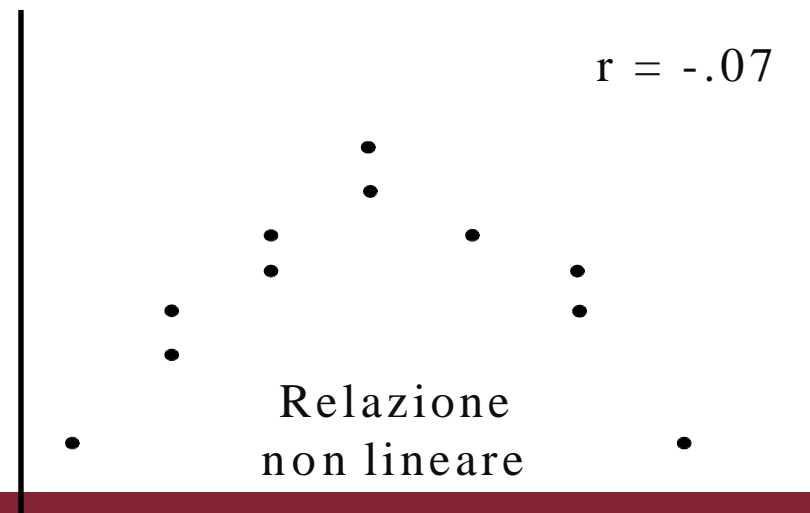
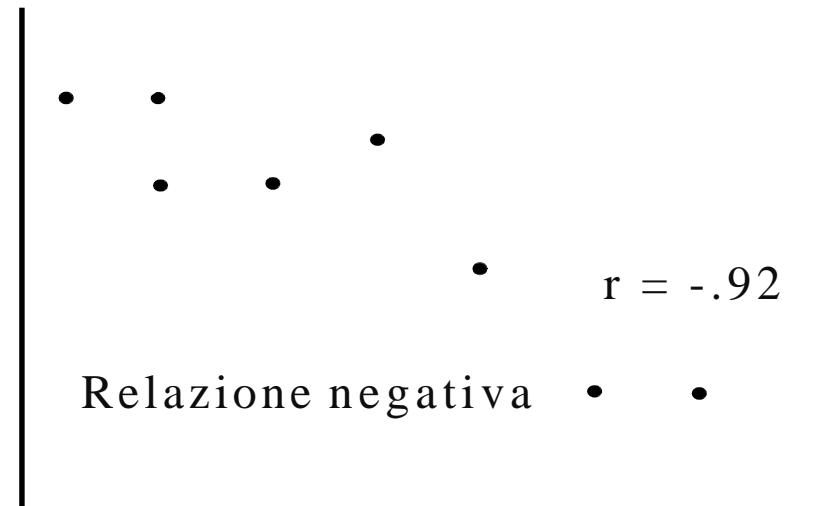
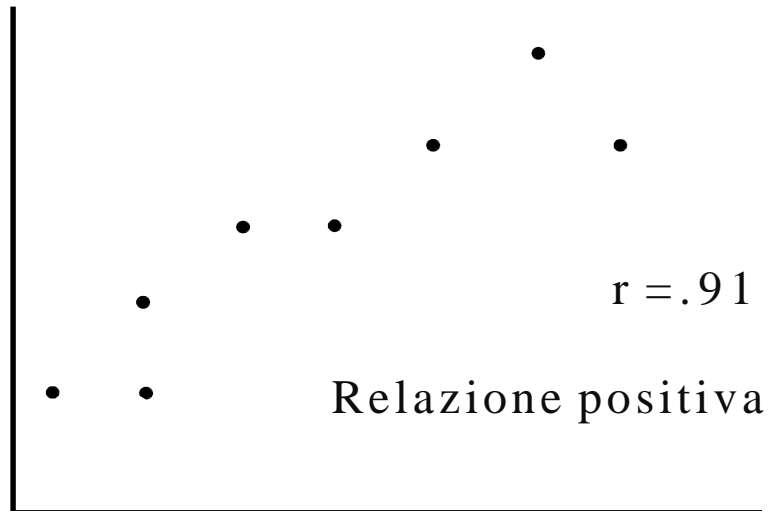
Il coefficiente di correlazione r di Pearson
interpretazione

Cohen's Rules Of Thumb For Effect Size

Effect size	Correlation coefficient	Difference between means
"Small effect"	$r = 0.1$	$d = 0.2$ standard deviations
"Medium effect"	$r = 0.3$	$d = 0.5$ standard deviations
"Large effect"	$r = 0.5$	$d = 0.8$ standard deviations

Correlazione

coefficiente r e diagramma di dispersione:



Correlazione

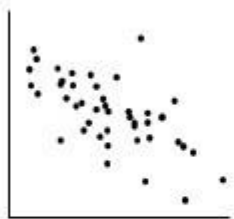
coefficiente r e diagramma di dispersione:

- r has a value between -1 and +1:

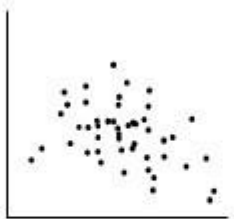


$r = -1$

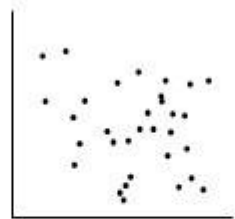
Points fall exactly
on a straight line



$r = -0.7$

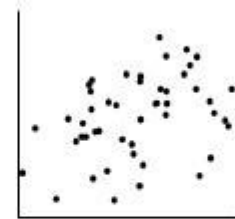


$r = -0.4$

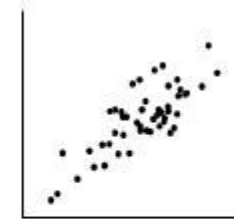


$r = 0$

No linear
relationship



$r = 0.3$



$r = 0.8$

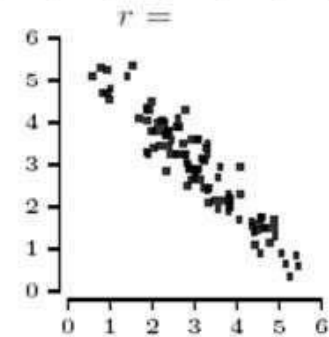
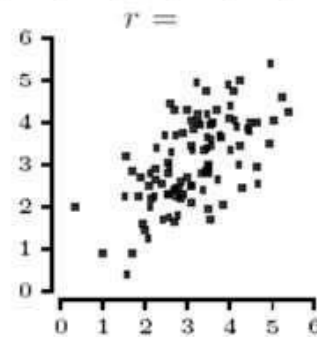
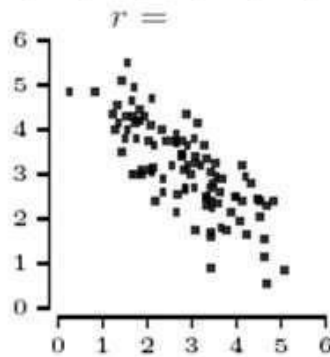
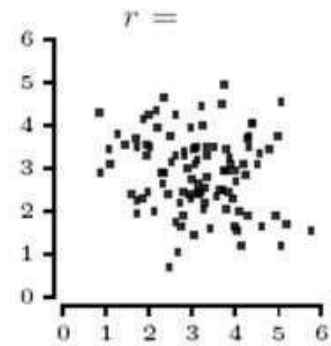
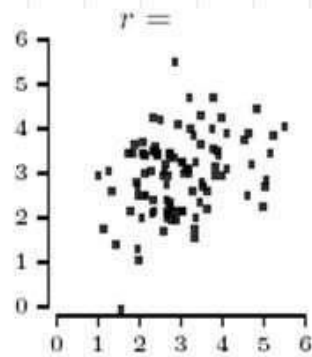
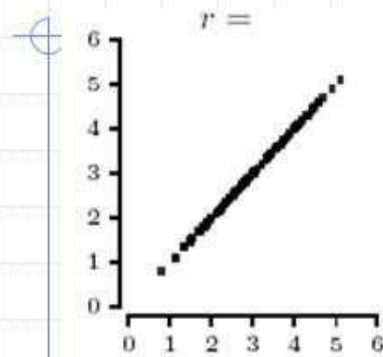


$r = 1$

Points fall exactly
on a straight line

Correlazione

Guess the correlation



Match the diagrams with the following correlations:

-0.93 **-0.75** **-0.20** **0.27** **0.63** **1.0**

Correlazione

Il coefficiente di correlazione r di Pearson formule (1)

$$r = \frac{\sum z_x z_y}{n - 1}$$

dove:

- al numeratore c'è la somma dei prodotti dei punteggi standardizzati z_x e z_y
- al denominatore c'è il numero di soggetti ($n - 1$, per la precisione)

Correlazione

Il coefficiente di correlazione r di Pearson
esempio di calcolo (1)

$X =$ test di creatività

$Y =$ voto in matematica

Soggetto	X	Y	z_x	z_y	$z_x z_y$
A	10	7	-1.32	0.63	-0.84
B	12	5	-0.22	-0.63	0.14
C	15	8	1.43	1.26	1.81
D	13	6	0.33	0.00	0.00
E	12	4	-0.22	-1.26	0.28
M	12.40	6.00			$\Sigma = 1.39$
SD	1.82	1.58			

Il coefficiente di correlazione r di Pearson formule (2)

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\frac{\sum (x - \bar{x})^2}{n-1} \cdot \frac{\sum (y - \bar{y})^2}{n-1}}}$$

Covarianza tra X e Y

Deviazione standard di X

Deviazione standard di Y

- al numeratore c'è la covarianza tra x e y
- al denominatore c'è il prodotto tra le deviazioni standard di x e y

Il coefficiente di correlazione r di Pearson esempio di calcolo (2)

X = test di creatività

Y = voto in matematica

Soggetto	X	Y	(X- \bar{X})	(Y- \bar{Y})	(X- \bar{X})(Y- \bar{Y})
A	10	7	-2.40	1.00	-2.40
B	12	5	-0.40	-1.00	0.40
C	15	8	2.60	2.00	5.20
D	13	6	0.60	0.00	0.00
E	12	4	-0.40	-2.00	0.80
M	12.40	6.00			$\Sigma = 4.00$
SD	1.82	1.58			

$$r = \frac{\frac{\sum (x - \bar{x})(y - \bar{y})}{n-1}}{\sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} \sqrt{\frac{\sum (y - \bar{y})^2}{n-1}}} = \frac{4/4}{(1.82)(1.58)} = \frac{1}{2.88} = 0.35$$

Correlazione

Il coefficiente di correlazione r di Pearson
verifica delle ipotesi

In genere il coefficiente di correlazione di Pearson si calcola su un campione (r). Questo coefficiente è una stima del parametro della popolazione ρ (r)

Ciò che ci interessa sapere è se il coefficiente di correlazione è diverso da zero nella popolazione

Ovvero: possiamo sostenere che le variabili sono correlate nella popolazione da cui proviene il campione?

Per rispondere a questa domanda dobbiamo effettuare la verifica delle ipotesi sul coefficiente r

Ricorda: i parametri (popolazione) sono indicati con lettere greche, le statistiche (campione) sono indicati con lettere latine

Correlazione

Il coefficiente di correlazione r di Pearson
verifica delle ipotesi

Le ipotesi statistiche sono le seguenti:

$$H_0: r = 0$$

le due variabili non sono correlate nella popolazione. Il coefficiente r osservato nel campione è diverso da zero solo perché r è una stima imprecisa di ρ

$$H_1: r \neq 0$$

le due variabili sono correlate nella popolazione (la correlazione tra le due variabili è statisticamente significativa)

L'ipotesi alternativa può essere anche monodirezionale
es. $H_1: r < 0$

Correlazione

Il coefficiente di correlazione r di Pearson
verifica delle ipotesi

La verifica delle ipotesi sul coefficiente di correlazione r può essere esaminata con due modalità, che forniscono risultati identici:

1– la tabella dei valori di r , che riporta i valori critici di r , in funzione di α e dei gradi di libertà (gdl) (Tav. G, pag. 250)

2 – il test t di Student

Correlazione

Il coefficiente di correlazione r di Pearson
verifica delle ipotesi

La tavola G riporta il valore critico di r per ipotesi alternative monodirezionali (dato un certo valore di alfa e determinati gradi di libertà)

- Se $r > r\text{-critico}$ si rifiuta H_0

TAV. G. Valori critici del coefficiente di correlazione r di Pearson per diversi livelli di probabilità (ipotesi monodirezionale)

gdl	p			
	.050	.025	0.010	.005
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487

Consideriamo l'esempio precedente ($n = 5$, $r = .35$). Se fissiamo il livello di alfa a .05 il valore critico è .805: $.35 < .805$, quindi non possiamo rifiutare H_0

Correlazione

Il coefficiente di correlazione r di Pearson
verifica delle ipotesi

Se si utilizza un'ipotesi alternativa bidirezionale, si deve consultare la colonna per livello di alfa pari a $\alpha/2$

Ad esempio, se si stabilisce un livello di alfa pari a .05 si deve consultare la colonna .025 (.05/2). In questo caso il valore critico è pari a .878

TAV. G. Valori critici del coefficiente di correlazione r di Pearson per diversi livelli di probabilità (ipotesi monodirezionale)

gdl	p			
	.050	.025	0.010	.005
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487

Il coefficiente di correlazione r di Pearson verifica delle ipotesi

La significatività statistica del coefficiente r può essere verificata utilizzando la t di Student

Esiste infatti una relazione tra r di Pearson e t di Student:

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

In particolare, la quantità $r \sqrt{\frac{n-2}{1-r^2}}$ si distribuisce approssimativamente come una t di Student con $n-2$ gradi di libertà

Correlazione

Il coefficiente di correlazione r di Pearson verifica delle ipotesi

Dopo aver operato questa trasformazione si può utilizzare la tavola E (pag. 245) per individuare il valore t critico

Il valore osservato di t viene quindi confrontato con il valore critico della distribuzione t di Student ...

Torniamo all'esempio ($r = .35$, $n = 5$, $\alpha = .05$):

$$t = r \sqrt{\frac{n-2}{1-r^2}} = .35 \sqrt{\frac{5-2}{1-.1225}} = .65$$

t osservata (o «empirica»)

Correlazione

Il coefficiente di correlazione r di Pearson
verifica delle ipotesi

Con $n = 5$ i gradi di libertà
sono pari a 3 ($n - 2$)

Con $\alpha = .05$ e H_1
bidirezionale la t critica è
pari a 3.182



gdl	Ipotesi monodirezionale					
	.10	.05	.025	.01	.005	.0005
	Ipotesi bidirezionale					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073

... poiché $.65 < 3.182$ non possiamo rifiutare H_0 : la r di
Pearson non è significativamente diversa da zero

Correlazione

Il coefficiente di correlazione r di Pearson Esercizio 1

Su un campione di 10 studenti vengono rilevate le variabili «Comprensione Verbale» e «Memoria di Lavoro»:

	 Compr_Verb	 Mem_Lav	var
1	12	8	
2	17	11	
3	11	8	
4	19	14	
5	21	11	
6	15	13	
7	20	7	
8	13	10	

		Compr_Verb	Mem_Lav
N	Valido	8	8
	Mancante	0	0
Media		16,00	10,25
Deviazione std.		3,817	2,493

1. Costruisci il diagramma di dispersione ($x = \text{Compr_Verb}$)
2. Calcola il coefficiente di correlazione
3. Formula le ipotesi statistiche (con H_1 bidirezionale) e verifica la significatività di r , per $\alpha = .05$

Correlazione

Il coefficiente di correlazione r di Pearson
Esercizio 2

Immaginiamo di ripetere la misura delle due variabili, questa volta su 100 studenti, e di calcolare nuovamente il coefficiente r di Pearson. Anche in questo caso otteniamo un valore di r pari a .31

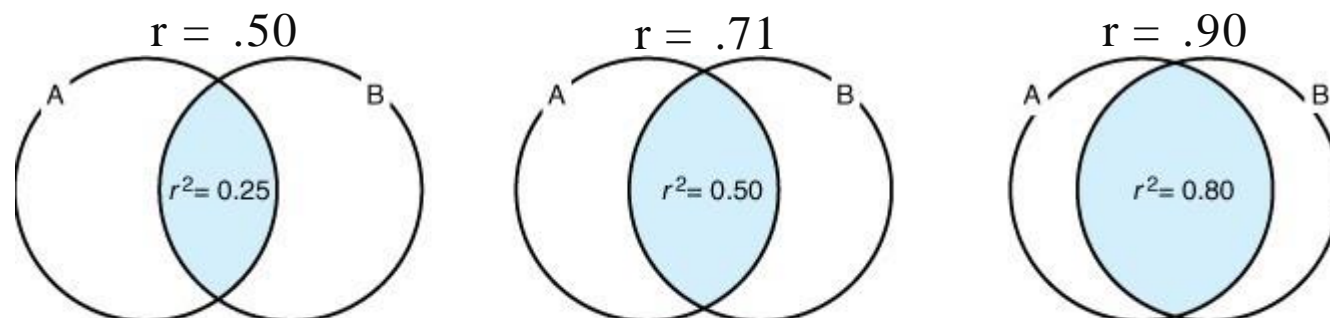
Cosa cambia nel processo di verifica delle ipotesi?

Correlazione

Il coefficiente di correlazione r di Pearson
coefficiente di determinazione

Il quadrato del coefficiente di correlazione (r^2) viene
definito coefficiente di determinazione

Misura l'ammontare di varianza che le due variabili
hanno in comune. Ad esempio, se il coefficiente di
correlazione è pari a $.50$, le due variabili condividono
il 25% di varianza

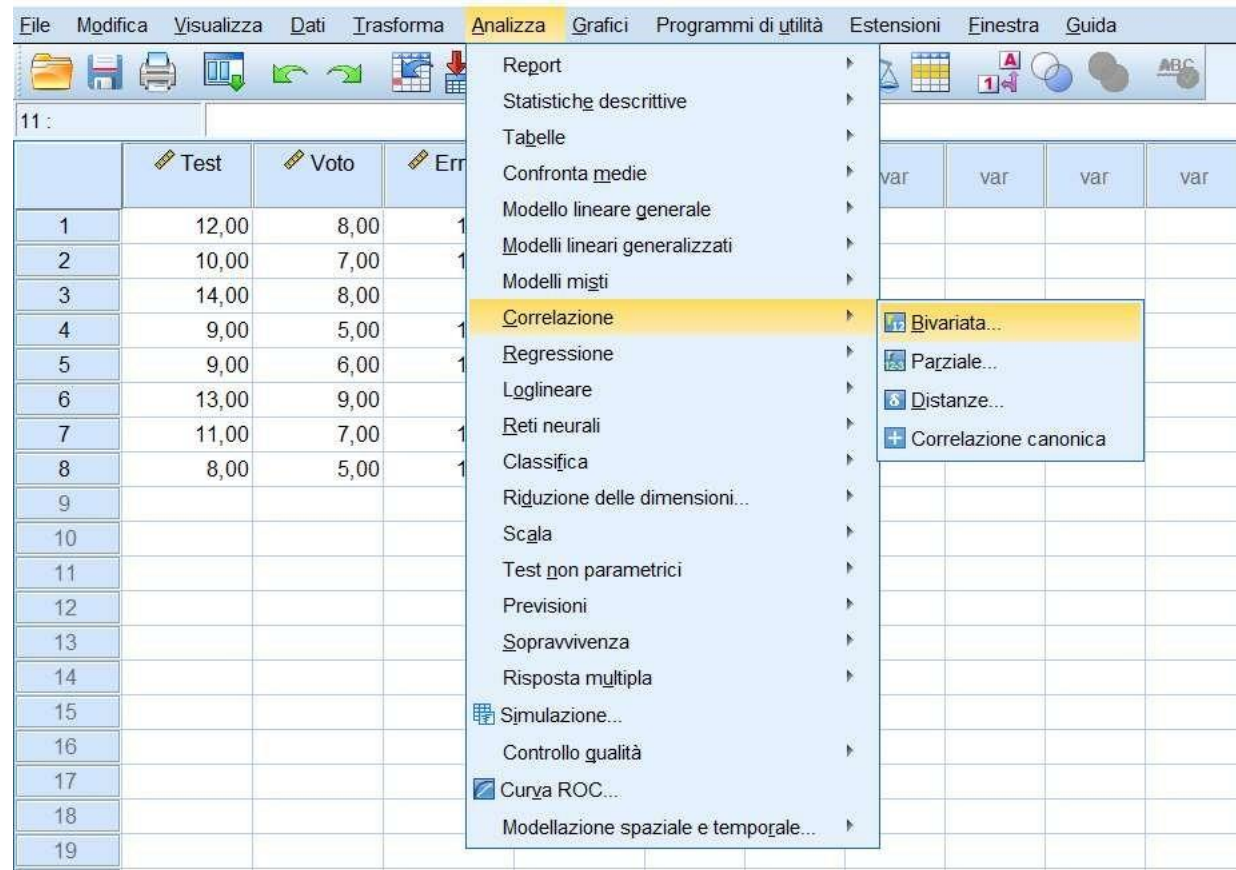


Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition:
<http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Correlazione

Il coefficiente di correlazione di Pearson in SPSS



Si può calcolare dalla finestra di dialogo «Bivariata» (selezionabile dal menu «Analizza», procedura «Correlazione»)

Correlazione

Il coefficiente di correlazione di Pearson in SPSS

1. Selezionare le variabili di cui si vuole calcolare la correlazione (es. Test e voto) e spostarle nel menu delle variabili attive

2. Selezionare l'opzione «Pearson» (è l'impostazione di default)

3. Cliccare su «OK»



Matrice di correlazione

Correlazioni

		Test	Voto
Test	Correlazione di Pearson	1	,912**
	Sign. (a due code)		,002
	N	8	8
Voto	Correlazione di Pearson	,912**	1
	Sign. (a due code)	,002	
	N	8	8

r di Pearson

** La correlazione è significativa a livello 0,01 (a due code).

Correlazione

Matrici di correlazione in SPSS

I coefficienti di correlazione calcolati su un insieme di variabili (es. punteggi in 5 item di un test) vengono disposti in una matrice dicorrelazione

Sulla diagonale principale ci sono le varianze delle variabili

Correlazioni

		Test	Voto	Errori	Socievolezza
Test	Correlazione di Pearson	1	,912**	-,916**	,039
	Sign. (a due code)		,002	,001	,927
	N	8	8	8	8
Voto	Correlazione di Pearson	,912**	1	-,826*	,322
	Sign. (a due code)	,002		,011	,436
	N	8	8	8	8
Errori	Correlazione di Pearson	-,916**	-,826*	1	,098
	Sign. (a due code)	,001	,011		,818
	N	8	8	8	8
Socievolezza	Correlazione di Pearson	,039	,322	,098	1
	Sign. (a due code)	,927	,436	,818	
	N	8	8	8	8

Coefficiente r

Livello di significatività

Numero di soggetti

** . La correlazione è significativa a livello 0,01 (a due code).

* . La correlazione è significativa a livello 0,05 (a due code).

Correlazione

La matrice di correlazione Esercizio 3

E =
Estroversione

A =
Amicalità

C =
Coscienziosità

N =
Nevroticismo

O =
Apertura mentale

		Correlazioni				
		E	A	C	N	O
E	Correlazione di Pearson	1	,048	,185**	-,137*	,333**
	Sign. (a due code)		,378	,001	,013	,000
	N	330	330	330	330	330
A	Correlazione di Pearson	,048	1	,232**	-,231**	,063
	Sign. (a due code)	,378		,000	,000	,251
	N	330	330	330	330	330
C	Correlazione di Pearson	,185**	,232**	1	-,255**	,049
	Sign. (a due code)	,001	,000		,000	,376
	N	330	330	330	330	330
N	Correlazione di Pearson	-,137*	-,231**	-,255**	1	-,092
	Sign. (a due code)	,013	,000	,000		,095
	N	330	330	330	330	330
O	Correlazione di Pearson	,333**	,063	,049	-,092	1
	Sign. (a due code)	,000	,251	,376	,095	
	N	330	330	330	330	330

** La correlazione è significativa a livello 0,01 (a due code).

* La correlazione è significativa a livello 0,05 (a due code).

1. Quale è la correlazione più elevata nella matrice? _____
2. La correlazione tra «N» e «O» è pari a: _____ È significativa (a = .05)? _____
3. Quale è la correlazione più bassa nella matrice? _____
4. Quale è la % di varianza che «E» e «C» hanno in comune? _____
5. Quanti coefficienti sono statisticamente significativi con a .05? _____
6. E con a .01? _____