TECNICHE DI ANALISI DEI DATI

AA 2018/2019

PROF. V.P. SENESE

Questi materiali sono disponibili per tutti gli studenti al seguente indirizzo:
https://goo.gl/xY15fR

ASSOCIAZIONE

Nelle ricerche spesso si desidera verificare se due serie di punteggi (due misure osservate) stabiliscono tra loro una **relazione**… e, se lo fanno, quale sia il **tipo** e il **grado** di tale relazione.
Ad esempio, che relazione c'è tra la qualità della relazione genitoriale e l'adattamento psicologico

La valutazione del tipo e grado di relazione viene fatta su **campioni** mediante la stima del **parametro di associazione** e l'eventuale relazione riscontrata deve poi essere generalizzata alla popolazione mediante la verifica delle ipotesi (inferenza).

CORRELAZIONE

La statistica che indica il grado di associazione tra variabili è l'**indice di correlazione**.

Associati ai valori dei coefficienti di correlazione vi sono dei ***test statistici*** che consentono di verificare l'ipotesi nulla (H_0) ovvero l'assenza di relazione tra le variabili a livello della popolazione.

L'indice di correlazione viene considerato un indice di **forza dell'effetto** o ***effect size*** dal momento che esprime su una scala standardizzata la forza della associazione o relazione tra variabili.

CORRELAZIONE/CAUSAZIONE

COVARIAZIONE
(Covarianza, Correlazione o Associazione):
quando "semplicemente" osserviamo che due variabili presentano **variazioni concomitanti**.

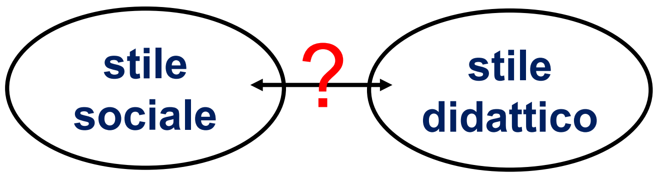
CAUSAZIONE:
quando pensiamo che siano proprio le variazioni della variabile **X a determinare** le variazioni della variabile **Y**. Identifichiamo la **DIREZIONALITÀ** e l'esistenza del **LEGAME DIRETTO** tra le due variabili.

Mentre la covariazione è osservabile la causazione appartiene al dominio della teoria!!!

CORRELAZIONE

Che cos'è la correlazione?

La correlazione tra due variabili è la tendenza delle variabili a "variare insieme" o a "**co-variare**".



Non c'è alcuna implicazione causale.

CORRELAZIONE

La correlazione viene qualificata sulla base di tre elementi:

1. **Tipo** di relazione o forma
2. **Direzione** della relazione
3. **Entità** della relazione

CORRELAZIONE

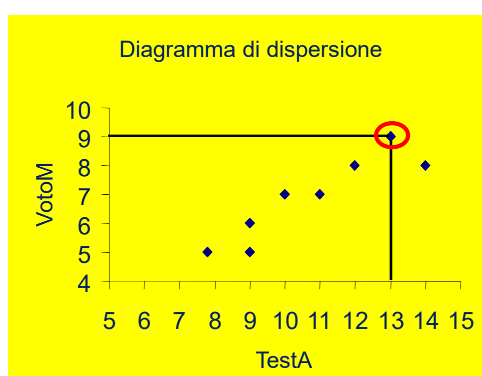
Supponiamo di essere interessati allo studio del tipo di relazione che lega due variabili: **X** (Attitudine in Matematica) e **Y** (Profitto in Matematica) misurate sullo stesso gruppo di soggetti.

Un primo e semplice passo da compiere è rappresentare graficamente la forma assunta dalla relazione tra **X** e **Y**.

CORRELAZIONE

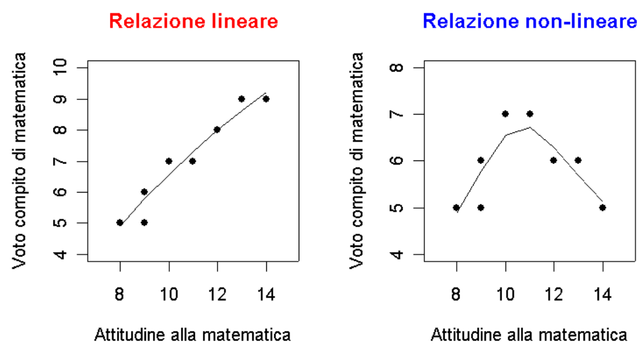
Supponiamo di aver somministrato a un campione di 8 studenti un test di **attitudine alla matematica** (**testA**) e un test di **abilità matematiche** (**votoM**), e di rappresentare graficamente la relazione tra le variabili.

COD	NOME	TEST_A	VOTO_M
1	*antonio*	12	8
2	*roberto*	10	7
3	*ida*	14	8
4	*simona*	9	5
5	*giovanna*	9	6
6	*carla*	13	9
7	*luca*	11	7
8	*melania*	8	5



CORRELAZIONE

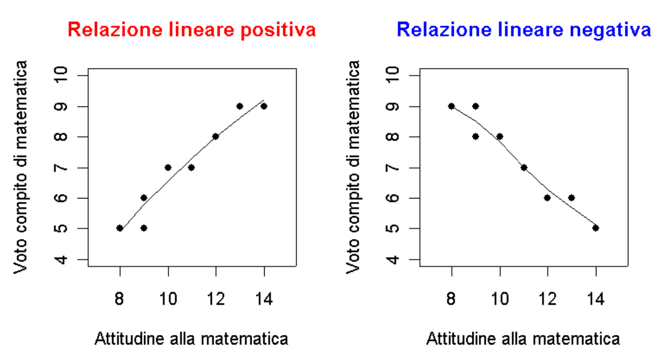
TIPO **DI RELAZIONE O** ***FORMA***



Una relazione si dice *lineare* quando la sua rappresentazione grafica, sugli assi cartesiani, si avvicina alla forma di una retta, *non lineare* quando ha un andamento curvilineo (per es., parabola).

CORRELAZIONE

DIREZIONE **DELLA RELAZIONE**



Una relazione si dice *positiva* quando la sua rappresentazione grafica, sugli assi cartesiani, vista da sinistra a destra tende a "salire", *negativa* quando da sinistra a destra tende a "scendere".

CORRELAZIONE

ENTITÀ **DELLA RELAZIONE**

L'***intensità*** della relazione e la direzione viene espressa attraverso dal ***coefficiente di correlazione***, indicato con la lettera *r*.

$$r_{XY}$$

r può variare tra –1 e +1

CORRELAZIONE

$r = +1$ → relazione "perfetta" **positiva**

"forte"

"sufficiente"

"debole"

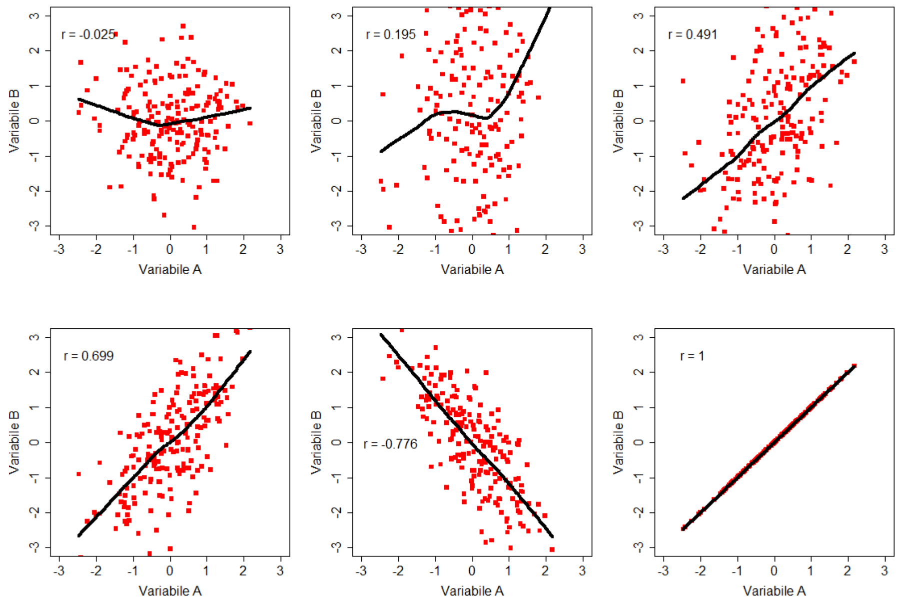
"assente"

$r = 0$ → relazione "**nulla**"

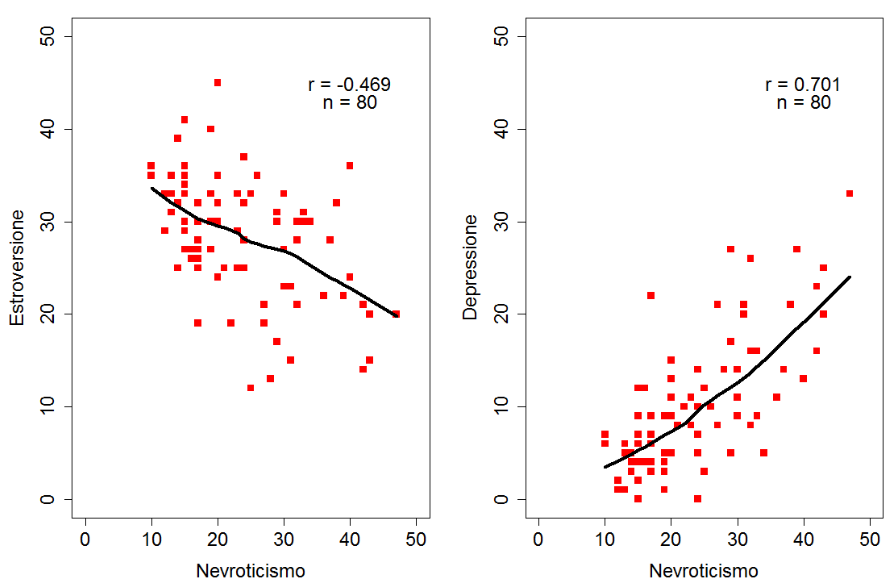
$r = -1$ → relazione "perfetta" **negativa**

Nella **psicologia** un valore di riferimento è ***r*** = |**.30**| che indica una associazione "debole".

CORRELAZIONE



CORRELAZIONE



CORRELAZIONE

L'analisi grafica della relazione tra le variabili è **sempre** consigliata, allo scopo di individuare alcuni dei fattori che se presenti possono essere responsabili di una **stima distorta** dei coefficienti:

1. *outliers* nei dati
2. sottogruppi non omogenei nel campione
3. distribuzioni diverse fra le variabili
4. relazioni non lineari

COEFFICIENTI

A seconda del tipo di variabili si utilizzano <u>diversi coefficienti di correlazione</u>:

quantitative: → *r* di Pearson

ordinali: → r_s (rho) di Spearman
t (tau) di Kendall

nominali: → *V* di Cramèr

una nominale dicotomica, una continua: → *r* punto biseriale

nominali dicotomiche: → *f* (phi)

r DI PEARSON

Quando le variabili sono misurate **almeno** al livello di scala ad **intervalli**, il coefficiente che si utilizza per l'analisi della relazione tra variabili è il coefficiente «prodotto-momento» di Pearson: ***r***.

CODEVIANZA

$$r_{Pearson} = \frac{\sum\left(X-\bar{X}\right)\cdot\left(Y-\bar{Y}\right)}{\sqrt{\sum\left(X-\bar{X}\right)^2\cdot\sum\left(Y-\bar{Y}\right)^2}}$$

SQRT PRODOTTO DELLE DEVIANZE

r DI PEARSON

$$r_{pearson} = \frac{N\sum xy-\sum x\sum y}{\sqrt{[N\sum x^2-(\sum x)^2][N\sum y^2-(\sum y)^2]}} =$$

$$= \frac{\sum z_x z_y}{N}$$

INFERENZA *r* DI PEARSON

Per la verifica delle ipotesi relative all'associazione di due variabili è necessario fare riferimento ad un modello bivariato che consenta di verificare ipotesi sulle relazioni tra variabili.

Tale modello è quello noto come **DISTRIBUZIONE NORMALE BIVARIATA** (μ_x, σ_x, μ_y, σ_y, ρ_{xy}) che consente di stimare la probabilità relative a ciascuna coppia di valori (x, y).

INFERENZA *r* DI PEARSON

Nel caso della correlazione bisogna ricorrere alla distribuzione campionaria del parametro ρ (***rho***) del quale ***r*** è una stima campionaria.

L'ipotesi nulla che si usa più **frequentemente** è quella di mancanza di relazione tra *x* e *y* ⇨ H_0: $\rho_{xy} = 0$.

Con questa ipotesi i valori si distribuiscono normalmente con ***gdl* = *n* – 2**.

Per l'interpretazione è necessario individuare il valore $r_{Critico}$ sulla tavola di distribuzione con le probabilità note, oppure **calcolare il valore della probabilità** associata all'ipotesi $\mathbf{H_0}$ dato il valore osservato ($\mathbf{H_0|r}$) utilizzando un software statistico.

INFERENZA *r* DI PEARSON

In alternativa è possibile impiegare una formula che riporti il valore del parametro ***r*** osservato in termini della distribuzione ***t di Student***, e procedere alla verifica delle ipotesi (con le tavole o un software).

$$t_{test_{r_{xy}}} = \frac{\left|r_{xy}\right|}{\sqrt{1-r_{xy}^2}} \cdot \sqrt{n-2}$$

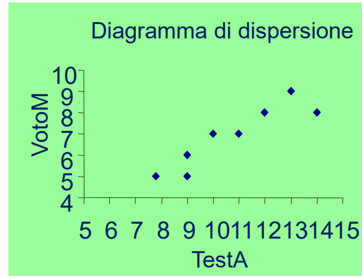
$$gdl = n-2$$

r DI PEARSON

(1) raccolta e codifica dei dati (valori osservati);
(2) inserimento dei dati in una matrice;
(3) definizione ipotesi nulla e ipotesi alternativa;
(4) calcolo dei parametri delle due distribuzioni;
(5) calcolo del *valore* r_{xy};
(6) verifica l'ipotesi in base alla distribuzione teorica *normale bivariata*;
(7) interpretazione dei risultati.

Esempio 1 *r* DI PEARSON

Supponiamo di voler verificare se esiste una **associazione** tra l'attitudine alla matematica (**testA**) e la performance al test di matematica (**votoM**).



L'ipotesi generale è che ci sia una relazione tra l'**attitudine alla matematica** (**VI**, non manipolata) e il **punteggio al test** (**VI**, non manipolata).

$$H_0 \Rightarrow r_{xy} = \rho_{xy} = 0$$

$$H_1 \Rightarrow r_{xy} = \rho_{xy} \neq 0$$

$\alpha = .05$ ➡ $\alpha = .025$

COMPOSTA
BIDIREZIONALE

Esempio 1 *r* DI PEARSON

Cod	TestA	VotoM			
	x	*y*	*xy*	x^2	y^2
1	12	8	96	144	64
2	10	7	70	100	49
3	14	8	112	196	64
4	9	5	45	81	25
5	9	6	54	81	36
6	13	9	117	169	81
7	11	7	77	121	49
8	8	5	40	64	25
$\sum$	86	55	611	956	393

$$\sum x = 86$$
$$\sum y = 55$$
$$\sum xy = 611$$
$$\sum x^2 = 956$$
$$\sum y^2 = 393$$

$$\bar{x} = 10.75; s_x = 1.98$$
$$\bar{y} = 6.88; s_y = 1.36$$

Esempio 1 *r* DI PEARSON

$$r_{pearson} = \frac{N\sum xy - \sum x \sum y}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

$$N = 8$$
$$\sum x = 86$$
$$\sum y = 55$$
$$\sum xy = 611$$
$$\sum x^2 = 956$$
$$\sum y^2 = 393$$

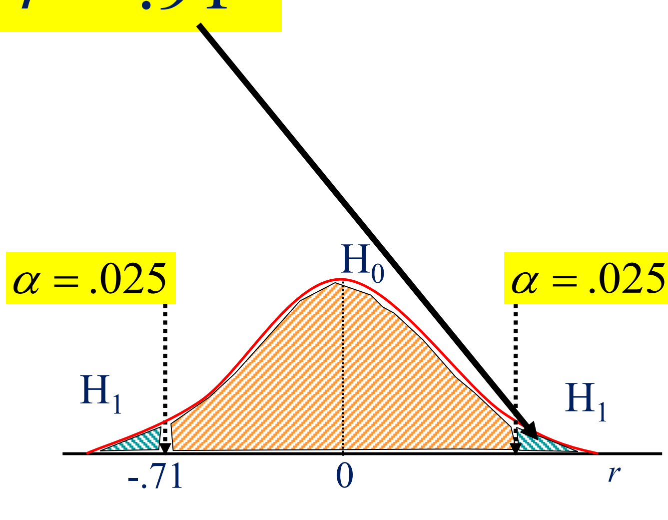
$$r = \frac{8 \cdot 611 - (86 \cdot 55)}{\sqrt{[8 \cdot 956 - 86^2] \cdot [8 \cdot 393 - 55^2]}} =$$

$$r = \frac{158}{173.17} = .912398$$

$$r = .91$$

Esempio 1 *r* DI PEARSON

$$r = .91$$



Esempio 1 *r* DI PEARSON

$$r = .91$$

$$t_{test_{r_{xy}}} = \frac{|r_{xy}|}{\sqrt{1-r_{xy}^2}} \cdot \sqrt{n-2}$$

$$gdl = n - 2$$

$$t_{test_{r_{xy}}} = \frac{|.91|}{\sqrt{1-.83}} \cdot \sqrt{8-2} = 5.406$$

$$gdl = 8 - 2 = 6$$

$$P(t_{test_{r_{xy}}}, gdl = 2) \cdot 2 = 0.001 \cdot 2 = 0.002$$

```
> pt(5.406,6,lower.tail = 0)*2
[1] 0.001654489
```

Esempio 1 *r* DI PEARSON

Correlations

		TestA	VotoM
TestA	Pearson Correlation	1	.912**
	Sig. (2-tailed)	.	.002
	N	8	8
VotoM	Pearson Correlation	.912**	1
	Sig. (2-tailed)	.002	.
	N	8	8

**. Correlation is significant at the 0.01 level

Esempio 1 *r* DI PEARSON

Questo risultato ci porta a **respingere** l'**ipotesi nulla** e a supportare l'**ipotesi alternativa**.

$$H_0 \Rightarrow r_{xy} = \rho_{xy} = 0$$

$$H_1 \Rightarrow r_{xy} = \rho_{xy} \neq 0$$

Come riportare il risultato
L'analisi della correlazione ha evidenziato una relazione lineare, positiva e forte tra l'attitudine alla matematica e il voto riportato al compito in classe. Infatti, maggiore è l'attitudine e maggiore è il voto riportato al compito (**r = .91; p < .01, N = 8**).

COEFF. DI DETERMINAZIONE

Quanto hanno in comune le due variabili?

Elevando al quadrato il coefficiente di correlazione si ha il **coefficiente di determinazione** indicato da r^2.

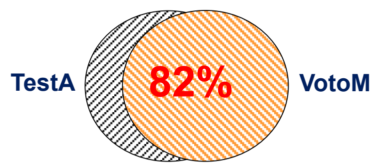
$$r^2_{XY}$$

Il **coefficiente di determinazione** r^2 misura la proporzione di variabilità comune alle due variabili. Se si moltiplica il valore per **100** si ottiene il valore in **percentuale** (**%**). Tale indice viene utilizzato per esprimere l'***effect size*** della associazione.

Esempio 1 *r* DI PEARSON

$$r = .91 \longrightarrow r^2 = .82$$

Nell'esempio, i risultati evidenziano che le due variabili hanno in **comune** l'**82%** di **varianza**.



$r^2_{|.30|} = .09 = 9\%$

Se, da un punto di vista teorico, una delle due variabili ha un "valore causale" (**variabile indipendente**), rispetto alla seconda, allora si può dire che la prima **spiega** l'82% della variabilità della seconda.

MATRICE DI CORRELAZIONE

Se ho più di due variabili (es. A, B e C), posso calcolare per **ciascuna coppia** il coefficiente di correlazione.

Posso perciò organizzarle in una tabella che viene detta matrice di *correlazione* o di *intercorrelazione*.

	A	B	C
A	1	r_{AB}	r_{AC}
B	r_{AB}	1	r_{BC}
C	r_{AC}	r_{BC}	1

MATRICE VARIABILIxVARIABILI (simmetrica rispetto alla diagonale)

MATRICE DI CORRELAZIONE

Correlations Between Perceived Partner Acceptance, Parental Acceptance, and Psychological Adjustment, by Gender of Respondent

Variable	1	2	3	4
1. Partner acceptance		.22†	.39**	.41**
2. Maternal acceptance	.26***		.55***	.26***
3. Paternal acceptance	.25***	.51***		.31*
4. Psychological adjustment	.40***	.31***	.40***	

Note: Correlations above the diagonal are for men (n = 67); correlations below the diagonal are for women (n = 421).
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Cross-Cultural Research
Volume 42 Number 1
February 2008 13-22
© 2008 Sage Publications
10.1177/1069397107309750
http://ccr.sagepub.com
hosted at
http://online.sagepub.com

Intimate Partner Acceptance, Parental Acceptance in Childhood, and Psychological Adjustment Among American Adults in Ongoing Attachment Relationships

Ronald P. Rohner
Tatiana Melendez
Lisa Kraimer-Rickaby
University of Connecticut

N.B. In questa matrice non **sono stati riportati i valori 1** nella diagonale, mentre nella parte superiore e inferiore della diagonale sono state riportate le **correlazioni relative a due sotto-campioni diversi** (maschi e femmine rispettivamente).

MATRICE DI CORRELAZIONE

Table 4. Pearson's correlation coefficients between the 15-item EQ subscales and the criterion measures

	EQ subscales			
Scales	Cognitive Empathy	Emotional Reactivity	Social Skills	EQ total
[a]IRI				
Fantasy scale	.064	.412***	−.156	.174*
Perspective Taking	.128	.433***	.045	.309***
Empathic Concern	.251**	.622***	.093	.489***
Personal Distress	.127	.169*	−.228**	.041
[b]HCL-32	.044	−.058	.075	.027
[c]TAS-20	−.131**	−.163***	−.356***	−.308***

Notes. [a]n = 150; [b]n = 250; [c]n = 413; IRI: Interpersonal Reactivity Index; HCL-32: Hypomania/Mania Symptom Checklist; TAS-20: Toronto Alexithymia scale; Hommel's corrected p-values: *$p < .05$; **$p < .01$; ***$p < .001$.

The Factorial Structure of a 15-Item Version of the Italian Empathy Quotient Scale

Vincenzo Paolo Senese,[1] Annunziata De Nicola,[1] Anna Passaro,[1] and Gennaro Ruggiero[2]
[1]Psychometric Laboratory, Department of Psychology, Second University of Naples, Italy
[2]Laboratory of Cognitive Science & Immersive Virtual Reality, Department of Psychology, Second University of Naples, Italy

N.B. Questa matrice **non è simmetrica** (e manca la diagonale) poiché sono diverse le variabili riportate in riga e colonna. Per tale ragione si dice rettangolare e non quadrata.

European Journal of Psychological Assessment (2016)
DOI: 10.1027/1015-5759/a000348

TECNICHE DI ANALISI DEI DATI

AA 2018/2019

PROF. V.P. SENESE

Questi materiali sono disponibili per tutti gli studenti al seguente indirizzo:
https://goo.gl/xY15fR

r DI SPEARMAN

Quando le variabili sono misurate almeno al livello di scala ad **ordinale**, il coefficiente che si utilizza per l'analisi della relazione tra variabili è il coefficiente di correlazione per **ranghi** di Spearman: r_s (Siegel e Castellan, 1992).

$$r_{s_{xy}} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{n(n^2 - 1)}$$

d_i = differenza tra rango in x e y

INFERENZA r DI SPEARMAN

Per verificare l'ipotesi mediante il coefficiente r_s dobbiamo riferirci come al solito alle tavole che riportano i valori di probabilità associati. Quando ***N* > 20** è possibile trasformare i valori di r_s osservati in *punti z* e confrontarli con la distribuzione normale (valendo H_0) mediante la seguente formula:

$$z_{r_{s_{xy}}} = r_{s_{xy}} \sqrt{N-1}$$

Quando ***N* < 20** è possibile trasformare i valori di r_s osservati in *punti t* mediante la seguente formula (***gdl*** = *N* – 2):

$$t_{r_{s_{xy}}} = r_{s_{xy}} \sqrt{\frac{N-2}{1-r^2_{s_{xy}}}}$$

Esempio 2 *r* DI SPEARMAN

Supponiamo di voler verificare se esiste una relazione tra l'autoritarismo (**TestA**) e i pregiudizi sociali (**PregS**).

L'ipotesi generale è che ci sia una relazione tra autoritarismo (**VI**, non manipolata) e i pregiudizi sociali (**VI**, non manipolata).

$$H_0 \Rightarrow r_{xy} = \rho_{xy} = 0$$

$$H_1 \Rightarrow r_{xy} = \rho_{xy} \neq 0$$

$\alpha = .05$ ➡ $\alpha = .025$

COMPOSTA
BIDIREZIONALE

Esempio 2 *r* DI SPEARMAN

$$r_{s_{xy}} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{n(n^2-1)}$$

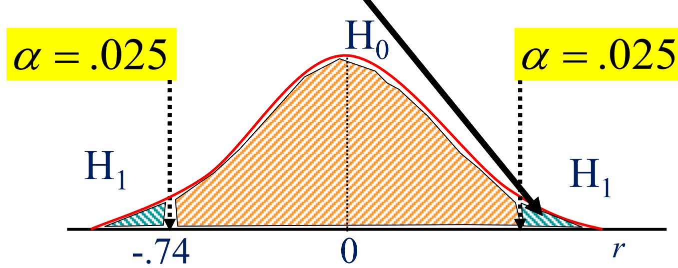
Cod	TestA	PregS				
	x	*y*	R_x	R_y	$R_x - R_y$	$(R_x - R_y)^2$
1	82	42	2	3	-1	1
2	98	46	5	4	1	1
3	87	39	4	2	2	4
4	40	37	1	1	0	0
5	116	65	8	6	2	4
6	113	88	7	8	-1	1
7	111	86	6	7	-1	1
8	83	56	3	5	-2	4

$$\sum d^2 = 16$$

$$r_{s_{xy}} = 1 - \frac{6 \cdot (16)}{8 \cdot (63)} = .80952$$

Esempio 2 *r* DI SPEARMAN

$$N = 8 \quad r_{s_{xy}} = .810$$



Esempio 2 *r* DI SPEARMAN

$$t_{r_{s_{xy}}} = r_{s_{xy}} \sqrt{\frac{N-2}{1-r^2_{s_{xy}}}}$$

$$t_{r_{s_{xy}}} = .81 \sqrt{\frac{8-2}{1-.81^2}} = 3.383$$

$$gdl = 8 - 2 = 6$$

```
> pt(3.383,6,lower.tail = 0)*2
[1] 0.01480289
```

Correlazioni

			TestA	PregS
Rho di Spearman	TestA	Coefficiente di correlazione	1,000	,810*
		Sig. (2-code)	.	,015
		N	8	8
	PregS	Coefficiente di correlazione	,810*	1,000
		Sig. (2-code)	,015	.
		N	8	8

*. La correlazione è significativa al livello 0,05 (2-code).

Esempio 2 *r* DI SPEARMAN

Correlazioni

			TestA	PregS
Rho di Spearman	TestA	Coefficiente di correlazione	1,000	,810*
		Sig. (2-code)	.	,015
		N	8	8
	PregS	Coefficiente di correlazione	,810*	1,000
		Sig. (2-code)	,015	.
		N	8	8

*. La correlazione è significativa al livello 0,05 (2-code).

Questo risultato ci porta a **respingere** l'**ipotesi nulla** e a supportare l'**ipotesi alternativa**.

Come riportare il risultato

L'analisi della correlazione ha evidenziato una relazione positiva e forte tra l'autoritarismo e i pregiudizi sociali. Infatti, maggiore è l'autoritarismo e maggiore è il numero di pregiudizi sociali, **r_s = .81; *p* < .05, *N* = 8**.

TECNICHE DI ANALISI DEI DATI

AA 2018/2019

PROF. V.P. SENESE

Questi materiali sono disponibili per tutti gli studenti al seguente indirizzo:
https://goo.gl/xY15fR

V DI CRAMÈR

Quando le variabili sono misurate su scala **nominale**, il coefficiente che si utilizza per l'analisi della relazione tra variabili è il coefficiente V di Cramèr: ***V*** (Siegel e Castellan, 1992). Il coefficiente **V** varia tra **0** e **1**.

$$V = \sqrt{\frac{\chi^2}{N(L-1)}}$$

$$\chi^2 = chi - quadrato$$
$$N = osservazioni$$
$$L = \min(n.righe, n.colonne)$$

INFERENZA *V* DI CRAMÈR

Per verificare l'ipotesi mediante il coefficiente ***V*** di Cramèr dobbiamo riferirci alla significatività della statistica χ^2 calcolata sulla tabella, con i relativi gradi di libertà. Il valore viene interpretato utilizzando la relativa distribuzione.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(f_{o_{ij}} - f_{a_{ij}}\right)^2}{f_{a_{ij}}}$$

f_o = frequenza osservata
f_a = frequenza attesa
r = numero di righe, $1, 2, \ldots, i$
c = numero di colonne, $1, 2, \ldots, j$

$$gdl_{\chi^2} = (r-1)(c-1)$$

Assunzioni:
(1) osservazioni indipendenti;
(2) nessuna freq. osservata = 0;
(3a) se dicotomica, nessuna freq. teorica < 5;
(3b) se politomica, nessuna freq. teorica < 1 e meno del 20% < 5.

Esempio 3 *V* DI CRAMÈR

Supponiamo che uno psicologo voglia valutare l'associazione tra la laurea e la scelta professionale. A tal scopo osserva un campione di 108 psicologi e un campione di 72 economisti e confronta le scelte professionali fatte (settore privato o università).

L'ipotesi generale è che la laurea (non manipolata, N) sia associata alla preferenza per il settore professionale (N).

$$H_0 \Rightarrow p_{f_{\Pr P}} = p_{f_{\Pr E}} \text{ e } p_{f_{PuP}} = p_{f_{PuE}}$$

$$H_1 \Rightarrow p_{f_{\Pr P}} \neq p_{f_{\Pr E}} \text{ e } p_{f_{PuP}} \neq p_{f_{PuE}}$$

$$\alpha = .05$$

CHE LA DISTRIBUZIONE DI FREQUENZE SIA DIVERSA TRA I DUE GRUPPI, QUINDI CHE LE VARIABILI SIANO DIPENDENTI o ASSOCIATE

Esempio 3 *V* DI CRAMÈR

f_o

	Privato	Pubblico	TOT
Psicologi	55	53	108
Economisti	45	27	72
TOT	100	80	180
$Modello\_H_0$	0.556	0.444	1

$$H_0 \Rightarrow p_{f_{PrP}} = p_{f_{PrE}} \text{ e } p_{f_{PuP}} = p_{f_{PuE}}$$

f_a

	Privato	Pubblico
Psicologi	60	48
Economisti	40	32

$108(.556) = 60$

$108(.444) = 48$

$72(.556) = 40$

$72(.444) = 32$

Esempio 3 *V* DI CRAMÈR

f_o

	Privato	Pubblico	TOT
Psicologi	55	53	108
Economisti	45	27	72
TOT	100	80	180

$$H_0 \Rightarrow p_{f_{PrP}} = p_{f_{PrE}} \text{ e } p_{f_{PuP}} = p_{f_{PuE}}$$

$fa_{PrP} : 108 = 100 : 80$

f_a

	Privato	Pubblico
Psicologi	60	48
Economisti	40	32

$$fa_{PrP} = \frac{108 \cdot 100}{180} = 60$$

$$fa_{PrE} = \frac{72 \cdot 100}{180} = 40$$

$$fa_{PuP} = \frac{108 \cdot 80}{180} = 48$$

$$fa_{PuE} = \frac{72 \cdot 80}{180} = 32$$

Esempio 3 *V* DI CRAMÈR

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(f_{o_{ij}} - f_{a_{ij}}\right)^2}{f_{a_{ij}}}$$

$$gdl_{\chi^2} = (r-1)(c-1)$$

f_o = frequenza osservata
f_a = frequenza attesa
r = numero di righe, 1, 2,..., i
c = numero di colonne, 1, 2,..., j

f_o	f_a	$f_o - f_a$	$(f_o - f_a)^2$	$\frac{(f_o - f_a)^2}{f_a}$
45	40	5	25	0.625
55	60	-5	25	0.417
27	32	-5	25	0.781
53	48	5	25	0.521

$$\chi^2 = 2.344$$
$$gdl = (2-1)(2-1) = 1$$

Esempio 3 *V* DI CRAMÈR

	Privato	Pubblico	TOT
Psicologi	55	53	108
Economisti	45	27	72
TOT	100	80	180

$$\chi^2 = 2.344$$
$$gdl = (2-1)(2-1) = 1$$

$$V = \sqrt{\frac{\chi^2}{N(L-1)}}$$

χ^2 = chi-quadrato
N = osservazioni
L = min(n.righe, n.colonne)

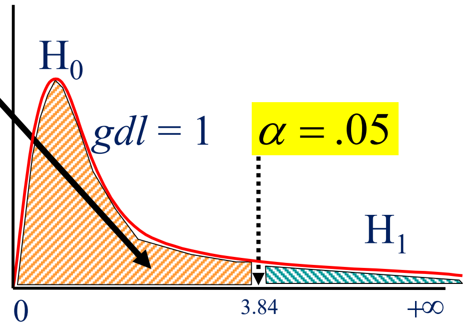
$$L = \min(2,2) = 2$$

$$V = \sqrt{\frac{2.344}{180(2-1)}} = 0.1141$$

Esempio 3 *V* DI CRAMÈR

$$\chi^2 = 2.344$$
$$gdl = (2-1)(2-1) = 1$$

```
> pchisq(2.344,1,lower.tail = 0)
[1] 0.1257662
```



Questo risultato ci porta ad **accettare** l'**ipotesi nulla** e a rifiutare l'**ipotesi alternativa**.

Come riportare il risultato
L'analisi della correlazione **non** ha evidenziato una associazione significativa tra la laurea e le scelte professionali, **$\chi^2(1)$ = 2.344, *p* = .126, *N* = 180, *V* = .114**.