

# Elementi di Psicometria

## Analisi fattoriale: Spiegazione intuitiva

(April 28, 2010)

Germano Rossi<sup>1</sup>

`germano.rossi@unimib.it`

<sup>1</sup>Dipartimento di Psicologia, Università di Milano-Bicocca

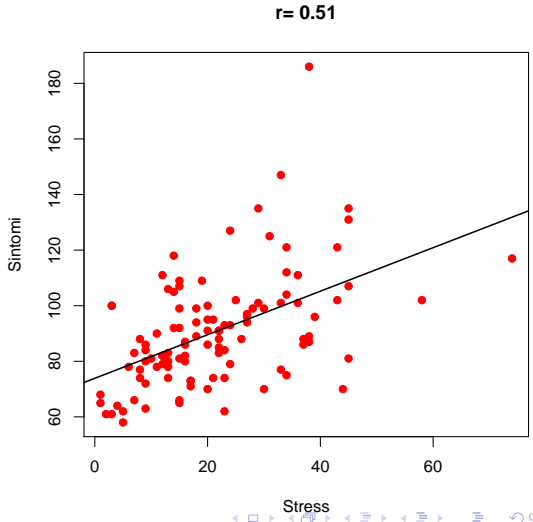
2009-2010

# Correlazione

C'è un legame fra i sintomi fisici e lo stress?

Parlando di grafica abbiamo visto che si può rappresentare la relazione fra due variabili tramite un grafico a dispersione.

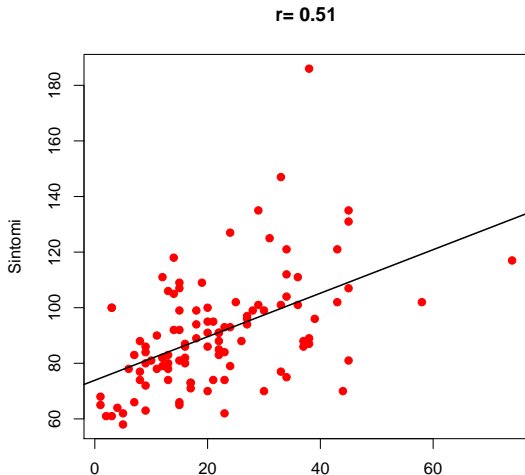
Nel grafico, ogni punto rappresenta un individuo con i suoi valori sulle due variabili. La correlazione stima il legame fra due variabili con un valore che oscilla fra -1 e 1. 0 significa mancanza di legami, 1 indica un legame matematico



# Regressione

Tramite una funzione matematica ci chiediamo se una variabile quantitativa possa essere spiegata da un'altra variabile quantitativa (o da altre variabili quantitative)

Se riusciamo ad avvicinarci a questa funzione matematica, siamo in grado di prevedere il valore di una variabile a partire dall'altra (e viceversa). In questo caso, la funzione prescelta è la retta.

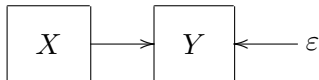


# Causa-effetto

- Sostanzialmente, quando diciamo che una variabile “predice” o “spiega” un’altra variabile, intendiamo dire che stiamo cercando di stabilire una relazione di causa-effetto. La variabile “spiegata” è l’effetto, quella che spiega (o che influenza) è la causa.
- La regressione lineare utilizza il principio matematico della retta per “cercare” questa relazione causale.

$$Y = bX + a$$

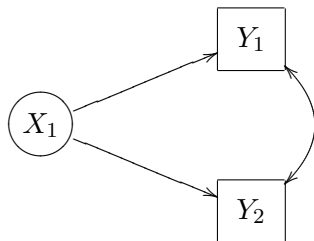
- Possiamo rappresentare graficamente l’“influenza” in questo modo:



# Analisi fattoriale

Quando studierete la correlazione, vedrete come possa essere interpretata come un indice di qualcosa che le due variabili hanno in comune.

E vedrete anche come sia possibile che due variabili correlino fra loro, anche senza una connessione logica perché tale correlazione può invece dipendere da una terza variabile. Se questa terza variabile non l'abbiamo mai effettivamente misurata, è una variabile latente (o fattore) e dovrà essere stimata sulla base dei suoi effetti.



# Definizioni

**Variabile osservata o misurata** una caratteristica che si può misurare in modo diretto (es.: età, reddito, numero di errori)

**Variabile latente** una caratteristica che non può essere misurata direttamente ma che si ipotizza avere dei legami con altre variabili osservabili (es.: intelligenza, cultura, inconscio)

La maggior parte dei costrutti psicologici sono “variabili latenti” che studiamo tramite le loro manifestazioni.

Per motivi storici, le variabili latenti furono chiamate “fattori”

# Analisi fattoriale

L'analisi fattoriale esplorativa è una tecnica di analisi dei dati che cerca di

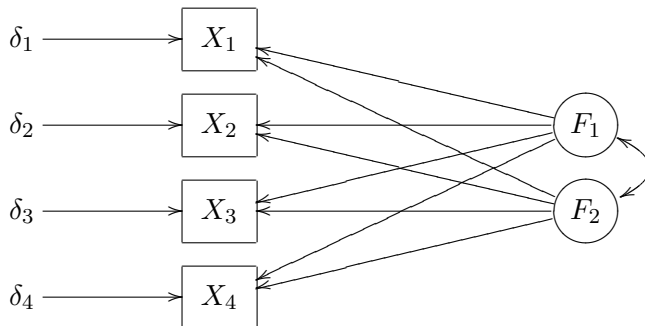
- associare una o più variabili latenti (che non si conoscono) ad un gruppo di variabili osservate. Si presuppone che le osservate abbiano qualche cosa in comune fra di loro (correlazioni) e che questo qualcosa dipenda dalla variabile latente.
- dato un determinato costrutto (psicologico) trovare delle variabili osservabili che servano per misurarlo

# AF Esplorativa: esempio 1

- In un prova (test) sono presenti item relativi a: 1) *Contare e ordinare sequenze numeriche*, 2) *Soluzione di problemi aritmetici*, 3) *Trovare errori di sintassi*, 4) *Trovare errori grammaticali*
- Tutti (o alcuni) di questi item hanno qualcosa in comune? Se sì, cosa? e con chi?
- Posso ipotizzare la presenza di due variabili latenti, una che ha qualcosa a che fare con i numeri e l'altra con la scrittura?
- Se effettuo una analisi fattoriale esplorativa posso calcolare la “relazione” che ciascuna prova ha con 1 o più fattori (var. latenti)



# Analisi fattoriale esplorativa

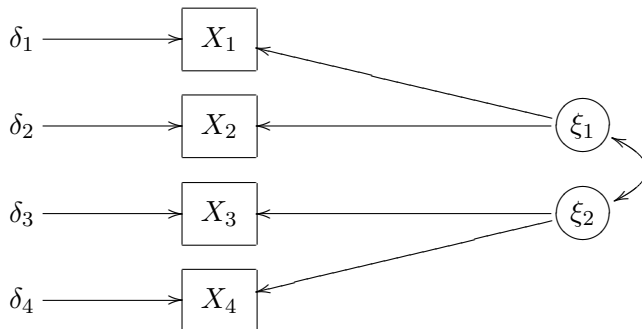


Tutte le  $X$  (ma in grado diverso) vengono spiegate da tutti i fattori (che possono anche essere correlati fra loro).

# Analisi fattoriale confermativa

- Serve per verificare che una o più variabili latenti che abbiamo ipotizzato siano effettivamente legate alle relative variabili osservate che gli sono associate
- Nell'esempio di prima, potremmo essere noi stessi ad ipotizzare che 1 e 2 sottostiano ad una variabile latente di *abilità numerica* e che 3 e 4 indichino una variabile latente di *abilità del linguaggio scritto*.
- L'analisi ci dirà quanto “buono” è il modello che abbiamo ipotizzato.

# Analisi fattoriale confermativa



Solo alcune X vengono spiegate dai Fattori (variabili latenti)

Anche in questo caso, ogni osservata (in quanto “riceve” una freccia) sarà oggetto di una equazione:

$$X_1 = b_1\xi_1 + \delta_1$$

e così via

# Scopo

Sia l'*Analisi fattoriale esplorativa* sia il suo correlato matematico (l'*Analisi delle componenti principali*) hanno gli stessi scopi

- 1 Ridurre un insieme di **variabili osservate** ad un insieme inferiore di **variabili non osservate** o **latenti** (fattori, componenti, dimensioni)
- 2 Trasformare le variabili osservate in una **struttura più semplice** che contenga però le stesse informazioni dell'originale

# Aspetti storici

## ■ 1904, Charles Spearman: Teoria bifattoriale

sosteneva che le misure di abilità mentale relative ad un test potevano essere spiegate come attribuibili ad un'abilità generale comune a tutte le abilità e ad un'abilità specifica e queste abilità dipendono ciascuna da un "fattore", chiamati da Spearman "Fattore generale" (G) e "fattore specifico o unico" (U).

## ■ 1945, Thurstone: Teoria multifattoriale

propose di sostituire il fattore generale con dei "fattori comuni" (F).

- La differenza è che i fattori comuni sono relativi solo ad alcuni item, quello generale li prendeva in considerazione tutti contemporaneamente.

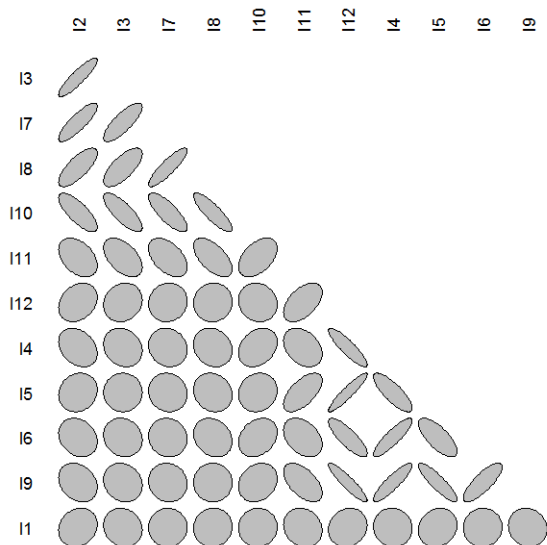
# Esempio matrice di correlazioni

Correlazione di Pearson

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1										
X2	,891**									
X3	-,300	-,177								
X4	,126	,000	-,804**							
X5	-,221	-,124	,876**	-,721**						
X6	,842**	,799**	-,127	-,086	-,074					
X7	,747**	,704**	-,161	-,150	-,206	,878**				
X8	-,258	-,151	,885**	-,880**	,798**	-,084	-,001			
X9	-,763**	-,787**	,280	,106	,273	-,787**	-,850**	,201		
X10	,264	,141	-,884**	,928**	-,820**	,076	,036	-,915**	-,109	

\*\*.

# Ispezione visiva della matrice di correlazione



# % di varianza spiegata

## Varianza totale spiegata

Fattore	Autovalori iniziali		
	Totale	% di varianza	% cumulata
1	4,961	49,611	49,611
2	3,743	37,429	87,041
3	,532	5,319	92,359
4	,272	2,724	95,084
5	,177	1,767	96,850
6	,102	1,020	97,870
7	,078	,785	98,655
8	,071	,710	99,365
9	,048	,480	99,845
10	,015	,155	100,000

Metodo di estrazione: Fattorizzazione dell'asse principale.



# Significato della rotazione

Non ruotata

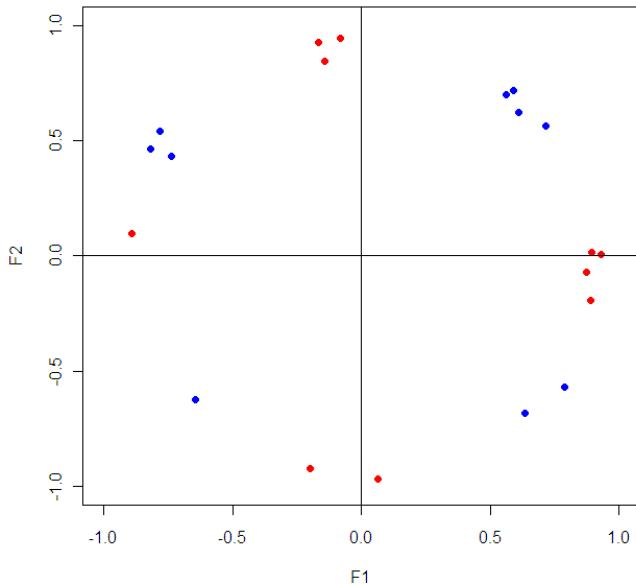
	1	2
x1	,716	,563
x2	,612	,625
x3	-,819	,464
x4	,633	-,681
x5	-,739	,431
x6	,588	,719
x7	,561	,698
x8	-,780	,541
x9	-,642	-,624
x10	,788	-,571

Ruotata

	1	2
x1	-,191	,890
x2	-,071	,872
x3	,926	-,166
x4	-,922	-,119
x5	,843	-,141
x6	,007	,929
x7	,014	,895
x8	,946	-,083
x9	,096	-,891
x10	-,970	,064

Tentativo di trovare una “struttura semplice” (Thurstone, 1947): saturazioni elevate in un fattore e il più basse possibili negli altri (focalizziamoci su X4 e X10)

blu=non ruotati; rosso=ruotati



## Grafico fattoriale nello spazio fattoriale ruotato

