

Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

PARTE II-A: ELEMENTI DI TCT

Antonio Calcagni

DPSS, Università di Padova

A.A. 2021/2022

Copyright © 2021 Antonio Calcagni. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation. A copy of the license is available at: <https://www.gnu.org/licenses/fdl-1.3.html>.

Teoria Classica dei test

Diversi sono i modelli formali utilizzabili nello sviluppo, validazione ed analisi delle proprietà dei test psicologici. Il primo di questi ad essere sviluppato, ed ancora oggi utilizzato, è il c.d. **modello del punteggio vero** (o teoria classica dei test - TCT). Altri modelli ricadono all'interno della **teoria della risposta all'item** (*Item Response Theory* - IRT) e sono stati sviluppati per diverse tipologie di item (es.: dicotomici, politomici) e diverse tipologie di test (es.: test di personalità, cognitivi, di performance).

Ciascuno di tali modelli ha sviluppati propri metodi per l'analisi dei test e lo studio delle loro proprietà, tra cui l'*attendibilità*, la *validità*, lo *scoring*, il *bias*, l'*analisi degli item*.

In questo corso affronteremo alcuni aspetti importanti per l'analisi dei test utilizzando principalmente la **TCT** come modello formale di riferimento. Quest'ultima non verrà trattata in maniera esaustiva ma si farà riferimento solo ad alcuni argomenti.

Per gli interessati, una sintesi introduttiva ai modelli IRT è disponibile in BN(3) mentre una comparazione tra TCT e IRT dal punto di vista applicativo è altresì disponibile in BN(5).

Teoria Classica dei Test

FONTI: BN(2.1,2.2), CRA(6)

Consideriamo un insieme di misure (o item) indicizzate tramite $j = 1, \dots, p$ e un insieme di rispondenti (estratti casualmente da una popolazione di riferimento) indicizzati tramite $i = 1, \dots, n$ su cui le misurazioni sono effettuate mediante un test.

La TCT stabilisce che

$$Y_j = T_j + E_j$$

dove:

Y_j è la v.a. che governa la realizzazione della j -esima misura relativamente alla componente **osservabile**

T_j è la v.a. che governa la realizzazione della j -esima misura relativamente alla componente **vera** (non osservabile)

E_j è la v.a. che governa la realizzazione della j -esima misura relativamente alla componente di **errore**

Teoria Classica dei Test

FONTI: BN(2.1,2.2), CRA(6)

Consideriamo un insieme di misure (o item) indicizzate tramite $j = 1, \dots, p$ e un insieme di rispondenti (estratti casualmente da una popolazione di riferimento) indicizzati tramite $i = 1, \dots, n$ su cui le misurazioni sono effettuate mediante un test.

La TCT stabilisce che

$$Y_j = T_j + E_j$$

dove:

- (i) $\mathbb{E}[E_j] = 0$ l'errore ha media nulla (errore non sistematico)
- (ii) $\text{Cor}[E_j, T_j] = 0$ l'errore e il valore vero della misura sono non correlati
- (iii) $\text{Cor}[E_j, E_{j'}] = 0$ due misurazioni differenti j e j' hanno errori non correlati

Teoria Classica dei Test

FONTI: BN(2.1-2.4), CRA(6)

Risultati generali per la TCT

$$Y_j = T + E_j$$

- (i) $\mathbb{E}[E_j] = 0$
- (ii) $\text{Cor}[E_j, T] = 0$
- (iii) $\text{Cor}[E_j, E_{j'}] = 0$

- (iv) $\mathbb{E}[Y_j] = \mathbb{E}[T]$
- (v) $\text{Var}[Y_j] = \text{Var}[T] + \text{Var}[E_j]$
- (vi) $\text{Var}[T] = \text{Cor}[Y_j, Y_{j'}]^\dagger \text{Var}[Y_j]$
- (vii) $\text{Var}[E_j] = \text{Var}[Y_j] (1 - \text{Cor}[Y_j, Y_{j'}]^\dagger)$

[†] La quantità $\text{Cor}[Y_j, Y_{j'}]$ è ottenuta utilizzando una coppia di **misure (osservabili) parallele**.

Due misure Y e Y' si dicono parallele quando misurano la stessa quantità latente T con la stessa quantità di errore E .

Si ha in questo caso $\mathbb{E}[Y_j] = \mathbb{E}[Y_{j'}]$ e $\text{Var}[Y_j] = \text{Var}[Y_{j'}]$.

Teoria Classica dei Test

FONTI: BN(2.3), CRA(6)

Attendibilità

Una misura importante che lega T a Y è la correlazione tra tali quantità, ossia ρ_{YT} , che esprime la forza con cui si esprimono le realizzazioni delle due v.a. Valori alti di ρ_{YT} indicano una notevole precisione nella quantificazione di T mediante Y . Nella realtà della rilevazione dei dati, solo Y è conoscibile mentre T (analogamente, E) sono stimabili solo utilizzando una coppia di misure parallele.

Teoria Classica dei Test

FONTI: BN(2.3), CRA(6)

Attendibilità

Consideriamo due misure (Y, Y') parallele. La loro correlazione lineare

$$\text{Cor}[Y, Y'] = \frac{\text{Cov}[Y, Y']}{\sqrt{\text{Var}[Y]} \sqrt{\text{Var}[Y']}}$$

può essere allora semplificata come segue:

$$\begin{aligned} \text{Cov}[Y, Y'] &= \text{Cov}[(T + E_Y), (T + E_{Y'})] \\ &= \text{Cov}[T, T] + \text{Cov}[T, E_{Y'}] + \text{Cov}[T, E_Y] + \text{Cov}[E_Y, E_{Y'}] \\ &= \text{Var}[T] \end{aligned}$$

Analogamente il denominatore:

$$\sqrt{\text{Var}[Y']} \sqrt{\text{Var}[Y]} = \text{Var}[Y]$$

Teoria Classica dei Test

FONTI: BN(2.3), CRA(6)

Attendibilità

L'attendibilità di un test o misura è dunque definita come segue:

$$\rho_{YT} = \frac{\text{Var}[T]}{\text{Var}[Y]} = 1 - \frac{\text{Var}[E]}{\text{Var}[Y]}, \quad \rho_{YT} \in [0, 1]$$

che esprime la variazione del misurando T (non osservato) rispetto al misurabile Y (osservato) utilizzando un criterio parallelo Y' .

Se $\text{Var}[E] = 0$, $\rho_{YT} = 1$, viceversa se $\text{Var}[E] = \text{Var}[Y]$ allora $\rho_{YT} = 0$.

L'attendibilità esprime informazione rispetto alla precisione/imprecisione della misura (nella definizione è coinvolta difatti la varianza): in particolare, $\text{Var}[E]$ indica l'imprecisione della misura mentre ρ_{YT} esprime al contrario la precisione del test. Quando $\text{Var}[E] \rightarrow 0$, la precisione della misura cresce.



Teoria Classica dei Test

FONTI: BN(2.3), CRA(6)

Attendibilità

Intepretazioni di ρ_{YT}

Ipotizziamo che il manuale di un certo test riporti $\rho_{YT} = 0.81$ con $\text{Var}[Y] = 16$. Ciò indicherebbe che la varianza osservata delle misurazioni $\text{Var}[Y]$ è attribuibile per l'81% alle variazioni nella componente del punteggio vero T (e non alla componente d'errore).

Infine, è possibile asserire che la correlazione tra punteggio vero T e misura Y (o precisione di Y) è pari a 0.81.



Teoria Classica dei Test

FONTI: BN(2.3), CRA(6)

Attendibilità

La definizione di attendibilità è basata sull'utilizzo di misure parallele, aspetto raramente incontrato nella realtà della somministrazione di test.

Un modo per approssimare misure parallele è quello di agire sul disegno della somministrazione del test:

- somministrazione dello stesso test sullo stesso campione in occasioni differenti: in questo caso, ρ_{YT} misura la *stabilità temporale* di Y
- somministrazione di due forme differenti dello stesso test sullo stesso campione: in questo caso, ρ_{YT} misura l'*equivalenza del test alternativo*

Nella pratica della psicometria applicata, la quantità ρ_{YT} è spesso stimata utilizzando diversi procedimenti statistici.



Teoria Classica dei Test

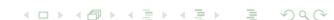
FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità

Dopo aver descritto il modello teorico per la formulazione dell'attendibilità, di seguito si riportano alcuni metodi per la stima dell'attendibilità in contesti reali di somministrazione di test.

Due approcci in generale:

1. Procedure che richiedono due somministrazioni del test
 - ▶ metodo delle forme parallele
 - ▶ metodo test-retest
2. Procedure che richiedono una sola somministrazione del test
 - ▶ metodo dello split-half
 - ▶ metodo basato sulla covarianza degli item (c.d. *coerenza interna*)



Teoria Classica dei Test

Major Error Source	Reliability Coefficient	Data-Collection Procedure	Statistical Treatment of Data
1. Change in examinee's overtime	1. Stability coefficient	1. Test, wait, retest	1. Compute Pearson product moment coefficient, ρ_{12}
2. Content sampling from form to form	2. Equivalence coefficient	2. Give form 1, give form 2	2. Compute Pearson product moment coefficient, ρ_{12}
3. Content sampling, or flawed items	3. Internal consistency coefficient	3. Give one form on one occasion	3a. Divide test into halves; correlate half-test, ρ_{AB} ; use Spearman Brown correction: $\rho_{xx'} = \frac{2\rho_{AB}}{1 + \rho_{AB}}$ b. Divide test into halves; use Guttman's or Rulon's formula: $\rho_{xx'} = 1 - \frac{\sigma_b^2}{\sigma_x^2}$ c. Compute item variances; compute coefficient alpha: $\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$

Approcci per la stima dell'attendibilità: Sintesi dei metodi per il calcolo empirico dell'attendibilità. Fonte: CRA(7, p.141)

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Split-Half

Dopo aver somministrato il test Y al campione, la procedura si basa sulla suddivisione del test in due parti di medesima lunghezza (ad esempio, divisione tra item pari e dispari, assegnazione casuale degli item alle due forme) - Y_A e Y_B - che vengono considerate forme parallele del medesimo test.

Fatta l'assunzione che le misure composite Y_A e Y_B siano parallele, l'attendibilità è stimata utilizzando la formula di Spearman-Brown:

$$\hat{\rho}_Y^2 = \frac{2\hat{\rho}_{AB}^2}{1 + \hat{\rho}_{AB}^2}$$

dove $\hat{\rho}_{AB}^2$ è la correlazione tra le due parti di Y calcolata tra i punteggi totali y_{totA} e y_{totB} .

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Split-Half

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	y_{totA}	y_{totB}	y_{tot}
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}_Y^2 = \frac{2\hat{\rho}_{AB}^2}{1 + \hat{\rho}_{AB}^2} = \frac{2 \cdot 0.34}{1 + 0.34} = 0.507$$

Il risultato ci indica che l'attendibilità del test Y, calcolata usando la divisione a metà del test, non è particolarmente alta. Un modo, ad esempio, per aumentarne il valore è quello di aggiungere ulteriori item (paralleli e coerenti in contenuto) affinché le forme composite Y_A e Y_B risultino più lunghe.

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Split-Half

Un altro modo per calcolare l'attendibilità del test totale Y nel caso in cui Y_A e Y_B non siano parallele, è quello di utilizzare l'indice di Rulon come stima dell'attendibilità del test complessivo:

$$\hat{\rho}_Y^2 = 2 \left(1 - \frac{\sigma_{y_A}^2 + \sigma_{y_B}^2}{\sigma_{y_{tot}}^2} \right)$$

dove con $\sigma_{y_{(.)}}^2$ si intende in generale la varianza del vettore di dati $y_{(.)}$.

In generale si ha la seguente relazione:

$$\hat{\rho}_{Y_{Sp-Brown}}^2 \geq \hat{\rho}_{Y_{Rulon}}^2$$

ed i due indici tendono ad equivalersi quando:

$$\frac{\sigma_{y_A}^2}{\sigma_{y_B}^2} \in [0.90, 1.10]$$

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Split-Half

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	y_{totA}	y_{totB}	y_{tot}
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}_Y^2 = 2 \left(1 - \frac{\sigma_{yA}^2 + \sigma_{yB}^2}{\sigma_{y_{tot}}^2} \right) = 2 \left(1 - \frac{1.82 + 1.56}{4.54} \right) = 0.508$$

Il risultato ottenuto con l'indice di Rulon è pressoché simile a quello ottenuto con la formula di Spearman-Brown. Il rapporto tra le varianze delle due metà del test è pari a $\frac{1.82}{1.56} = 1.163$.

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Coerenza interna

Consideriamo una collezione di m misure $(Y_1, T_1, E_1), \dots, (Y_m, T_m, E_m)$ che formano la *misura composta* $X = \sum_{j=1}^m Y_j$. Uno stimatore dell'attendibilità della misura composta $\hat{\rho}_{XX'}^2$ è noto come α di Cronbach:

$$\hat{\rho}_{XX'}^2 = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \text{Var}[Y_j]}{\sum_{j=1}^m \text{Var}[Y_j] + \sum_{j \neq h} \text{Cov}[Y_j, Y_h]} \right)$$

dove maggiore è la componente di covarianza tra le misure componenti, maggiore sarà l'indice α finale. Quando invece la covarianza tra le misure è prossima allo zero, l'indice assumerà anch'esso valore prossimo allo zero.

⁽¹⁾ Nota: la sommatoria è sulle misure componenti che sono diverse. Infatti quando $j = h$ abbiamo $\text{Cov}[Y_j, Y_h] = \text{Var}[Y_j, Y_h]$ già calcolata nella sommatoria precedente. Tale formulazione permette di operare sulle covarianze delle coppie di misure realmente differenti tra loro.

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Coerenza interna

$$\hat{\rho}_{XX'}^2 = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \text{Var}[Y_j]}{\sum_{j=1}^m \text{Var}[Y_j] + \sum_{j \neq h} \text{Cov}[Y_j, Y_h]} \right)$$

La componente di covarianza tra le misure componenti $\text{Cov}[Y_j, Y_h]$ fornisce un'indicazione sulla *coerenza* delle misure tra loro. Per tale ragione l'indice α si dice quantifichi la **coerenza interna** di una scala (o misura composta).

Nota: tale formulazione fornisce un'interpretazione dell'attendibilità in termini di *coerenza interna* della misura composta. A differenza di altre tecniche per valutare l'attendibilità di un test (ad esempio, *split-half*), l' α di Cronbach utilizza solo l'informazione derivante dalla covarianza dalle misure componenti (o item).

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Coerenza interna

Lo stimatore α :

- interpreta l'attendibilità di una misura composta in termini di coerenza interna rispetto alle misure componenti che la formano
- presuppone che le misure componenti *formino bene* la misura composta, nel senso dato dalla intercorrelazione tra le misure componenti
- è interpretato usando i seguenti valori di riferimento:

valore soglia	attendibilità
$\alpha > 0.9$	ottima
$0.8 \leq \alpha \leq 0.9$	buona
$0.7 \leq \alpha < 0.8$	discreta
$0.6 \leq \alpha < 0.7$	sufficiente
$\alpha < 0.6$	insufficiente

Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Coerenza interna

Un altro indice per la stima dell'attendibilità della scala Y quando le misure componenti (item) sono di tipo dicotomico è noto come indice di Kuder-Richardson (KR20):

$$\hat{\rho}_{XX'}^2 = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \pi_j(1-\pi_j)}{\sigma_{Y_{tot}}^2} \right)$$

L'indice si ottiene dalla sostituzione della generica $\text{Var}[Y]$ nella formula dell' α di Chronbach con la varianza della distribuzione binomiale.

L'interpretazione segue quella adottata per l'indice α di Chronbach.



Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Coerenza interna

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	Y_{totA}	Y_{totB}	Y_{tot}
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3
π	0.4	0.3	0.6	0.7	0.6	0.3			

$$\hat{\rho}_{XX'}^2 = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \pi_j(1-\pi_j)}{\sigma_{Y_{tot}}^2} \right) = \frac{6}{6-1} \left(1 - \frac{1.35}{4.54} \right) = 0.703$$

Il risultato differisce da quello ottenuto usando il metodo dello Split-Half: ciò dipende dalla particolare scelta della divisione a metà del test fatta con quest'ultimo metodo. Difatti, facendo variare le possibili suddivisioni a metà (es.: usando la suddivisione casuale) si ottengono diversi valori per $\hat{\rho}_{XX'}^2$.



Teoria Classica dei Test

FONTI: BN(2.8.0,2.9), CRA(7)

Procedure per la stima dell'attendibilità: Correlazione intraclasse

Utilizzando i risultati della TCT, è possibile fornire un calcolo dell'attendibilità come rapporto di varianze:

$$\hat{\rho}_{YT}^2 = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_E^2}$$

Tale formula è denominata coefficiente di correlazione intraclasse e rapporta la quantità di varianza della componente vera alla varianza complessiva.



Teoria Classica dei Test

FONTI: BN(2.5), CRA(7)

Procedure per la stima dell'attendibilità: Correlazione intraclasse

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	Y_{totA}	Y_{totB}	Y_{tot}
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}_{YT}^2 = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_E^2} = \hat{\rho}_{YT}^2 = \frac{1.787^2}{1.787^2 + 1.161^2} = 0.703$$

Il risultato è analogo a quello ottenuto precedente ed evidenzia il calcolo della precisione del test Y come rapporto tra la varianza del misurando T e la varianza complessiva del test dovuta sia alla componente dell'errore E che a quella della componente latente T .



Teoria Classica dei Test

FONTI: BN(2.6), CRA(6)

Attendibilità e lunghezza del test

In generale, data una misura composita iniziale X (o test) avente attendibilità $\rho_{XX'}^2$, possiamo valutare come questa vari in funzione dell'aggiunta di m nuovi item.

Questo può essere utile, ad esempio, quando si vuole valutare quanti item occorra aggiungere per avere una desiderata attendibilità.

La risposta a tale quesito è nota come *formula profetica di Spearman-Brown*:

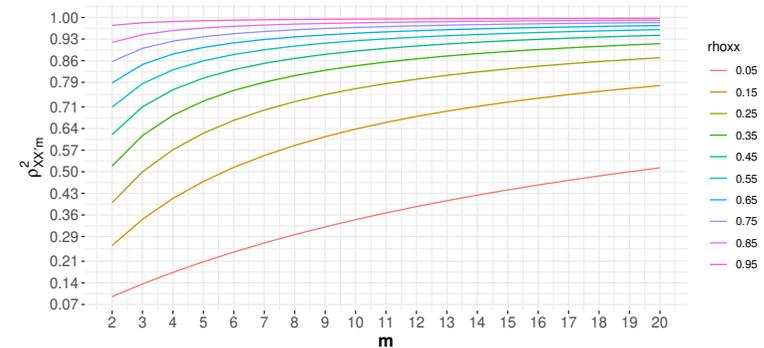
$$\rho_{XX',m}^2 = \frac{m\rho_{XX'}^2}{1 + (m-1)\rho_{XX'}^2}$$

dove $\rho_{XX'}$ indica l'attendibilità del test iniziale (o test di origine).



Teoria Classica dei Test

FONTI: BN(2.6), CRA(6)



Attendibilità di una misura composita e lunghezza: Curve di attendibilità di una misura composita (o test) quando questa è allungata per un intero m . Le curve sono in funzione dell'attendibilità iniziale $\rho_{XX'}^2$ della misura composita (colore delle curve) e mostrano come questa, tenendo fisso tale valore, cambia in precisione $\rho_{XX',m}^2$ (ordinata) quando m nuove misure componenti (ascissa) sono aggiunte a formare il test.



Teoria Classica dei Test

FONTI: BN(2.10), CRA(7)

Fattori che influenzano l'attendibilità

- qualità del campione di individui su cui le misurazioni sono effettuate (rappresentatività)
- qualità delle misure componenti (o item)
- campionamento realmente casuale ed indipendenza delle osservazioni
- condizione di somministrazione del test
- lunghezza del test adeguata (vedi slide 61-62)
- aspetti cognitivi degli individui testati: fatica, ricordo, bassa compliance



si approfondisca BN(2.10), CRA(7,pp.143-146)



Teoria Classica dei Test

FONTI: BN(2.8.0,2.9), CRA(7)

Stima dei punteggi veri ad un test

Dato il test (o misura composita) Y , abbiamo precedentemente calcolato \mathbf{y}_{tot} come somma dei punteggi su p item (vedi slide 66). Il vettore \mathbf{y}_{tot} contiene i punteggi osservati del campione di n individui testati. Per ottenere una stima dei punteggi veri $\tau_{n \times 1}$ è possibile utilizzare lo stimatore lineare:

$$\tau_i = \hat{\rho}_{YT}^2 y_{tot_i} + (1 - \hat{\rho}_{YT}^2) \mu_x$$

dove $\mu_x = \frac{1}{n} \sum_{i=1}^n y_{tot_i}$ è la media complessiva delle misurazioni effettuate (stima della media dei punteggi veri nella popolazione). Lo stimatore utilizza una stima dell'attendibilità/precisione del test: quando questa è bassa, esso tende ad attribuire maggior peso alla componente μ_x rispetto alla misura osservata y_{tot_i} .

Nota: lo stimatore utilizzato per τ potrebbe non essere adeguato quando la relazione tra Y e T è non lineare.



Teoria Classica dei Test

FONTI: BN(2.8.0,2.9), CRA(7)

Stima dei punteggi veri ad un test

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	Y_{totA}	Y_{totB}	Y_{tot}	τ
1	0	0	0	0	0	0	0	0	0	0.861
2	0	0	0	0	1	0	1	0	1	1.564
3	1	0	1	1	1	0	3	1	4	3.673
4	1	1	1	1	1	1	3	3	6	5.079
5	1	1	1	1	1	1	3	3	6	5.079
6	0	0	1	0	0	0	1	0	1	1.564
7	0	0	1	1	1	0	2	1	3	2.970
8	0	0	0	1	0	0	0	1	1	1.564
9	1	0	1	1	1	0	3	1	4	3.673
10	0	1	0	1	0	1	0	3	3	2.973

Caso con alta attendibilità: Per la stima di τ si è utilizzata la stima di ρ_{YT}^2 ottenuta tramite il KR20 (vedi slide 74). Si noti come i punteggi veri siano stimati rispetto al tratto/costrutto latente continuo sotteso agli item dicotomici. Inoltre, il calcolo di τ non richiede nessuna informazione aggiuntiva sul singolo individuo (es.: variazioni individuali dei punteggi) se non il suo punteggio osservato.

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9), CRA(7)

Stima dei punteggi veri ad un test

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	Y_{totA}	Y_{totB}	Y_{tot}	τ
1	0	0	0	0	0	0	0	0	0	2.871
2	0	0	0	0	1	0	1	0	1	2.881
3	1	0	1	1	1	0	3	1	4	2.911
4	1	1	1	1	1	1	3	3	6	2.931
5	1	1	1	1	1	1	3	3	6	2.931
6	0	0	1	0	0	0	1	0	1	2.881
7	0	0	1	1	1	0	2	1	3	2.901
8	0	0	0	1	0	0	0	1	1	2.881
9	1	0	1	1	1	0	3	1	4	2.911
10	0	1	0	1	0	1	0	3	3	2.901

Caso con bassa attendibilità: In questo caso (estremo) si è ipotizzato un valore di precisione prossimo allo zero $\rho_{YT}^2 = 0.01$. Si noti come i punteggi veri tendano ad assumere lo stesso valore della media della popolazione ignorando i punteggi osservati. Ciò è in linea con l'impostazione della TCT che formalizza il test (e le sue proprietà) come oggetto primario rispetto al comportamento di risposta dei partecipanti.

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9), CRA(7)

Stima dei punteggi veri ad un test

Una volta calcolato $\hat{\tau}_i$ è possibile quantificare l'incertezza della stima mediante un intervallo di confidenza simmetrico $C_i = (\tau_i^{lb}, \tau_i^{ub}) \in \mathbb{R}^2$ che utilizza il modello probabilistico della normale standardizzata come riferimento per τ :

$$\mathbb{P}(\tau_i \in C_i) = 1 - \alpha$$

$$\mathbb{P}(\hat{\tau}_i - z_{1-\alpha}\sigma_E \leq \tau_i \leq \hat{\tau}_i + z_{1-\alpha}\sigma_E) = 1 - \alpha$$

dove $z_{1-\alpha}$ è il valore critico della distribuzione normale standardizzata corrispondente alla probabilità α scelta (es: $\alpha = 0.05$, $z_{1-\alpha} = 1.96$).

La componente $\sigma_E = \sqrt{\text{Var}[E]}$ e può essere calcolata utilizzando il risultato (vii) della TCT e la stima della precisione del test $\hat{\rho}_{YT}^2$ (calcolata opportunamente con uno dei procedimenti visti in precedenza):

$$\hat{\sigma}_E = \hat{\sigma}_{Y_{tot}} \sqrt{(1 - \hat{\rho}_{YT}^2)}$$

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9), CRA(7)

Stima dei punteggi veri ad un test

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	Y_{totA}	Y_{totB}	Y_{tot}	τ^{lb}	τ	τ^{ub}
1	0	0	0	0	0	0	0	0	0	-1.416	0.861	3.138
2	0	0	0	0	1	0	1	0	1	-0.713	1.564	3.841
3	1	0	1	1	1	0	3	1	4	1.396	3.673	5.950
4	1	1	1	1	1	1	3	3	6	2.802	5.079	7.356
5	1	1	1	1	1	1	3	3	6	2.802	5.079	7.356
6	0	0	1	0	0	0	1	0	1	-0.713	1.564	3.841
7	0	0	1	1	1	0	2	1	3	0.693	2.970	5.247
8	0	0	0	1	0	0	0	1	1	-0.713	1.564	3.841
9	1	0	1	1	1	0	3	1	4	1.396	3.673	5.950
10	0	1	0	1	0	1	0	3	3	0.693	2.970	5.247

Intervallo di confidenza per il punteggio vero: Per la stima di τ si è utilizzata la stima di ρ_{YT}^2 ottenuta tramite il KR20 (vedi slide 74). Si noti come l'ampiezza degli intervalli di confidenza sia dipendente dall'attendibilità del test: maggiore è il valore di precisione ρ_{YT}^2 , minore sarà l'ampiezza di C_i .

Teoria Classica dei Test

FONTI: BN(6.1), CRA(10)

Validità

Quando il nostro interesse è rivolto a capire *cosa* un determinato test misuri siamo nell'ambito della validità di un test (o misura). Diversi sono i metodi per valutare la validità di un test e diverse sono anche le accezioni di riferimento:

- validità di contenuto (riferita al contenuto o semantica degli item/scale)
- validità fattoriale (riferita alla coerenza interna degli item, valutata mediante *analisi della dimensionalità*)
- validità di criterio (riferita alla coerenza di un test rispetto ad un criterio/test esterno)
- validità di costrutto (riferita alla convergenza o divergenza rispetto ad altre misure per un medesimo costrutto da quantificare)



Teoria Classica dei Test

FONTI: BN(6.1), CRA(10)

Validità



studio individuale: BN(6.1.1), BN(6.1.2), BN(6.1.4)



Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2), CRA(19)

Punteggi e norme

Quando una scala è somministrata ad un campione di n individui, le misure componenti y_{i1}, \dots, y_{ip} rilevate sul campione sono utilizzate per formare il *punteggio grezzo* $y_{tot} = \sum_{j=1}^p y_{ij}$.

Il processo di calcolo del punteggio grezzo individuale usando gli item di un test è detto **scoring**.

L'interpretazione di y_{tot} viene effettuata per confronto con i **valori normativi** del test (solitamente riportati nel manuale associato al test/questionario). I valori normativi sono ottenuti durante la *fase di taratura del test* ed esprimono le caratteristiche di Y a livello di popolazione, queste ultime quantificate utilizzando i momenti di Y (solitamente media e varianza).



Taratura di un test: BN(2.12.3)



Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2), CRA(19)

Punteggi e norme

Per valutare/confrontare i punteggi ottenuti dagli individui ad una scala o tra più scale è opportuno trasformare i punteggi grezzi in *punteggi standardizzati*. Questi possono essere impiegati altresì per la costruzione dei *profili individuali*.

Diversi sono i modi per calcolare i punteggi standardizzati, ad esempio mediante:

- punteggio z
- punteggio T



Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2), CRA(19)

Punteggi e norme

Punteggio z

$$z_i = \frac{y_i - \bar{y}}{s_y}$$

dove:

- \bar{y} è la media del campione
- s_y è lo scarto quadratico medio (standard deviation) del campione

I punteggi z hanno media nulla e varianza unitaria poiché utilizzano la Normale standardizzata come legge di riferimento.

I punteggi z si esprimono su una scala centrata sullo zero e con unità di misura lo scarto quadratico medio: le distanze tra punteggi infatti sono espresse in termini di $\pm 1, 2, \dots, K$ deviazioni standard.



Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2), CRA(19)

Punteggi e norme

Punteggio t

Si ottiene per trasformazione lineare dei punteggi z:

$$t_i = 50 + 10z_i$$

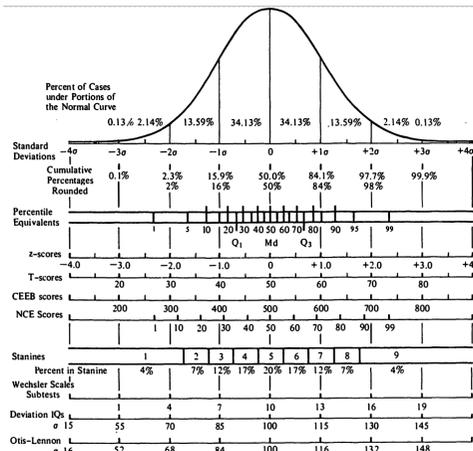
I punteggi t hanno media pari a 50 e deviazione standard pari a 10 e vengono utilizzati preferibilmente per rendere i punteggi finali tutti positivi (i punteggi z infatti possono assumere valori negativi).



Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2), CRA(19)

Punteggi e norme



Tipologie di punteggi standardizzati: Punteggi standardizzati rispetto al modello normale standard (con media e dev. standard). Fonte: CRA(19,p.446).



Teoria Classica dei Test

Note di sintesi

- La TCT decompone la variabilità della misurazione Y in una componente di errore E (accidentale) ed una componente attribuita al misurando T (latente)
- Obiettivo della TCT è di costruire un test di misura che abbia alta precisione/attendibilità, ossia quando $\text{Var}[T] > \text{Var}[E]$
- L'attendibilità viene costruita utilizzando misure parallele coerenti tra loro
- Le componenti di variabilità individuale (differenze individuali) non vengono modellate se non attraverso il campione/gruppo
- La TCT si focalizza maggiormente sulle caratteristiche del test più che su quelle degli item



Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

PARTE II-B: CFA

Antonio Calcagni

DPSS, Università di Padova

A.A. 2021/2022

Copyright © 2021 Antonio Calcagni. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation A copy of the license is available at: <https://www.gnu.org/licenses/fdl-1.3.html>.

Introduzione

La costruzione di un test/questionario necessita di definire buone scale di misura, ossia scale che siano valide e attendibili. Nel modulo II-A abbiamo definito la *coerenza interna* di una scala e la *validità fattoriale* ed abbiamo più volte richiamato all'**analisi della dimensionalità** come metodo per studiare tali aspetti.

Diverse sono le tecniche statistiche utilizzate per studiare la dimensionalità di un test/questionario:

- Analisi fattoriale confermativa (CFA)
- Analisi fattoriale esplorativa (EFA)
- Generalized components analysis
- Partial least squares (PLS) components analysis
- ...

All'interno di questo modulo ci concentreremo sulla CFA. Per un'eventuale introduzione ad EFA, si veda BN(4.1).

Introduzione

FONTI: BKM(1.4)

La CFA è un modello statistico multivariato per lo studio simultaneo di più variabili. Appartiene alla famiglia più generale dei **modelli a variabili latenti** nei quali, dato un insieme di variabili aleatorie (spesso continue) Y_1, \dots, Y_J , la distribuzione congiunta $f(Y_1, \dots, Y_J; \theta)$ viene studiata condizionando un sottoinsieme $p < J$ di variabili, Y_1, \dots, Y_p , ad un altro sottoinsieme di $q < J$, Y_1, \dots, Y_q (con $q < p$).

Quest'ultimo sottoinsieme è inteso essere **non osservato** nelle sue realizzazioni mentre Y_1, \dots, Y_p sono solitamente osservate e misurate.

Introduzione

FONTI: BKM(1.4)

Così facendo, si induce una *gerarchia* tra le J v.a. per cui $\mathbf{Y}_q = (Y_1, \dots, Y_q)$ precedono le variabili $\mathbf{Y}_p = (Y_1, \dots, Y_p)$ nel senso del meccanismo generatore dei dati:

- 1 si realizzano gli esiti delle variabili latenti $\mathbf{Y}_q \sim f(\mathbf{Y}_q; \boldsymbol{\theta}^{(2)})$
- 2 si realizzano gli esiti delle variabili osservate condizionatamente agli esiti delle variabili latenti $\mathbf{Y}_p \sim f(\mathbf{Y}_p | \mathbf{Y}_q; \boldsymbol{\theta}^{(1)})$

Lo schema di condizionamento fa sì che le v.a. \mathbf{Y}_p non possano essere osservate senza che le v.a. latenti \mathbf{Y}_q (condizionanti) si siano realizzate.

Questo schema, nel contesto della psicometria applicata, è detto **schema riflessivo**, in contrasto con lo **schema formativo** (tipico, ad esempio, nella costruzione degli indicatori sociali) dove invece \mathbf{Y}_p precede \mathbf{Y}_q .

Introduzione

FONTI: BKM(1.4)

Sovente nella CFA si denotano le v.a. latenti con le lettere greche, a sottolineare che le loro realizzazioni non sono osservate. Ad esempio, indicando con $\boldsymbol{\eta} = \mathbf{Y}_q$ e assumendo che le realizzazioni $i = 1, \dots, n$ delle v.a. osservate e latenti siano indipendenti e con identica distribuzione, avremo il modello seguente:

$$\boldsymbol{\eta}_i \sim f(\boldsymbol{\eta}_i; \boldsymbol{\theta}^{(1)})$$

$$\mathbf{y}_i | \boldsymbol{\eta}_i \sim f(\mathbf{y}_i | \boldsymbol{\eta}_i; \boldsymbol{\theta}^{(2)})$$

dove $\boldsymbol{\theta}^{(1)} \subset \mathbb{R}^k$ è il vettore $k \times 1$ di parametri che governa la densità delle v.a. latenti mentre $\boldsymbol{\theta}^{(2)} \subset \mathbb{R}^h$ è il vettore $h \times 1$ di parametri della densità delle osservate.

Nota: nella terminologia della CFA, le v.a. latenti $\boldsymbol{\eta}$ sono detti fattori o costrutti.

Introduzione

FONTI: BKM(1.4)

La **densità congiunta** è la seguente ($\Omega = \boldsymbol{\theta}^{(1)} \cup \boldsymbol{\theta}^{(2)}$):

$$f(\boldsymbol{\eta}, \mathbf{y}; \Omega) = \prod_{i=1}^n f(\mathbf{y}_i | \boldsymbol{\eta}_i; \boldsymbol{\theta}^{(2)}) f(\boldsymbol{\eta}_i; \boldsymbol{\theta}^{(1)})$$

mentre la **densità marginale** è ottenuta per integrazione delle quantità $\eta_{i1}, \dots, \eta_{iq}$:

$$f(\mathbf{y}; \Omega) = \prod_{i=1}^n \int_{\mathcal{R}(\boldsymbol{\eta})} f(\mathbf{y}_i | \boldsymbol{\eta}_i; \boldsymbol{\theta}^{(2)}) f(\boldsymbol{\eta}_i; \boldsymbol{\theta}^{(1)}) d\boldsymbol{\eta}_i$$

dove $\mathcal{R}(\boldsymbol{\eta})$ è il supporto delle v.a. latenti. La soluzione dell'integrale q -variato permette di ottenere la densità di probabilità per le sole v.a. osservate.

Nota: nel problema statistico di tipo inferenziale, dopo aver osservato il campione di dati $\mathbf{Y}_{n \times p}$ siamo interessati a stimare $\boldsymbol{\theta}^{(2)}$, $\boldsymbol{\theta}^{(1)}$ e a determinare/predire $\eta_{i1}, \dots, \eta_{iq}$.

Introduzione

Modelli a variabili latenti

		Manifest variables	
		Metrical	Categorical
Latent variables	Metrical	Factor analysis	Latent trait analysis
	Categorical	Latent profile analysis	Latent class analysis

Modelli a variabili latenti: Sulla base della tipologia di variabile (continua - *metrical*; ordinale o nominale - *categorical*) è possibile definire diverse tipologie di modelli a variabili latenti, agendo sulle densità $f(\boldsymbol{\eta}_i; \boldsymbol{\theta}^{(1)})$ e $f(\mathbf{y}_i | \boldsymbol{\eta}_i; \boldsymbol{\theta}^{(2)})$ che definiscono il modello. Scegliendo, difatti, opportuni modelli continui o discreti per le due densità (es.: normale, binomiale, esponenziale) - possibilmente appartenenti tutti alla *famiglia esponenziale* - è possibile definire modelli generalizzati per ogni tipologia di dati campionari $\mathbf{Y}_{n \times p}$ a disposizione. Fonte: BKM(p.11).

Introduzione

Trasformazioni di variabili

Il caso dell'analisi fattoriale confermativa ricade nella prima cella della tabella precedente e richiede che entrambe le v.a. η e \mathbf{Y} siano rappresentate come variabili continue. Quando ciò non è possibile, un approccio alternativo all'utilizzo di modelli generalizzati è quello di utilizzare tecniche che trasformano la metrica delle variabili, a livello dei dati osservati (c.d. *optimal scaling*) o di statistiche di essi (es.: *correlazione poliorica*).

Tali approcci permettono di lasciare invariata la struttura statistica del modello a variabili latenti - cioè non agiscono sulle densità $f(\eta_i; \theta^{(1)})$ e $f(\mathbf{y}_i | \eta_i; \theta^{(2)})$ che rimangono continue - mentre usano tecniche di trasformazione (quantificazione delle variabili) cercando di conservare, del dato originario, quanta più informazione possibile.

Introduzione



introduzione storica a CFA: BKM(1.8)

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

Per derivare la CFA come caso particolare di un modello a variabili latenti che usa la densità normale per $f(\eta_i; \theta^{(1)})$ e $f(\mathbf{y}_i | \eta_i; \theta^{(2)})$, procediamo come segue.

Assumiamo che le v.a. latenti seguano in distribuzione il modello Gaussiano q -variato

$$\eta_i \sim \mathcal{N}_q(\mu_i, \Phi)$$

avente media μ_i (vettore $q \times 1$ di reali) e matrice di covarianza Φ (di ordine $q \times q$).

Rappresentiamo la relazione tra v.a. latenti e v.a. osservate mediante il *modello lineare*:

$$\mathbf{y}_i = \tau_i + \Lambda \eta_i + \delta_i$$

dove τ_i è il vettore $p \times 1$ di intercette, Λ è una matrice $p \times q$ di coefficienti reali (c.d. *coefficienti fattoriali*) che legano le p osservate alle q latenti, δ_i è il vettore $p \times 1$ di errore del modello.

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

Assumendo un modello Gaussiano per l'errore delle misurazioni (variabili osservate),

$$\delta_i \sim \mathcal{N}_p(\mathbf{0}_p, \Theta_\delta)$$

dove $\mathbf{0}_p$ è un vettore $p \times 1$ di zeri (medie nulle) mentre Θ_δ è la matrice $p \times p$ di varianze-covarianze solitamente *diagonale* (ossia, solo valori non nulli sulla diagonale e zero al di fuori di questa), il modello probabilistico *marginale* delle v.a. che governano le misurazioni è il seguente:

$$\mathbf{y}_i \sim \mathcal{N}_p(\underbrace{\tau_i + \Lambda \mu_i}_{\mu_y}, \underbrace{\Lambda \Phi \Lambda^T + \Theta_\delta}_{\Sigma_y})$$

dove μ_y è il vettore $p \times 1$ delle medie marginali delle variabili osservate mentre Σ_y ne è la matrice di covarianza di ordine $p \times p$.

Note:

$$\mu_y = \mathbb{E}[\mathbf{y}] = \mathbb{E}[\tau + \Lambda \eta + \delta] = \tau + \Lambda \mathbb{E}[\eta] + \mathbb{E}[\delta] = \tau + \Lambda \mu$$

$$\Sigma_y = \text{Var}[\mathbf{y}] = \text{Var}[\tau + \Lambda \eta + \delta] = \text{Var}[\Lambda \eta] + \text{Var}[\delta] = \Lambda \text{Var}[\eta] \Lambda^T + \Theta_\delta = \Lambda \Phi \Lambda^T + \Theta_\delta$$

In generale, si ricordi la proprietà dell'operatore varianza per cui $\text{Var}[Ya] = a^2 \text{Var}[Y]$ (caso scalare) e $\text{Var}[YA] = A \text{Var}[Y] A^T$ (caso con A matrice) dove AA^T equivale al quadrato del caso scalare (T indica il *trasposto di una matrice*).

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

Il modello congiunto $(\mathbf{y}, \boldsymbol{\eta})_i$ risulta:

$$\begin{bmatrix} \mathbf{y}_i \\ \boldsymbol{\eta}_i \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\tau}_i + \boldsymbol{\Lambda} \boldsymbol{\mu}_i \\ \boldsymbol{\mu}_i \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda} \boldsymbol{\Phi} \boldsymbol{\Lambda}^T + \boldsymbol{\Theta}_\delta & \boldsymbol{\Phi} \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} \boldsymbol{\Phi} & \boldsymbol{\Phi} \end{bmatrix} \right)$$

dove:

- $\boldsymbol{\tau}_i + \boldsymbol{\Lambda} \boldsymbol{\mu}_i = \mathbb{E}[\mathbf{y}_i]$
- $\boldsymbol{\mu}_i = \mathbb{E}[\boldsymbol{\eta}_i]$
- $\boldsymbol{\Lambda} \boldsymbol{\Phi} \boldsymbol{\Lambda}^T + \boldsymbol{\Theta}_\delta = \text{Cov}[\mathbf{y}_i]$
- $\boldsymbol{\Phi} = \text{Cov}[\boldsymbol{\eta}_i]$
- $\boldsymbol{\Lambda} \boldsymbol{\Phi} = \boldsymbol{\Phi} \boldsymbol{\Lambda}^T = \text{Cov}[\boldsymbol{\eta}_i, \mathbf{y}_i]$

◀ ▶ ⏪ ⏩ 🔍 ↺

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

La CFA riscritta in termini di modello lineare a variabili latenti risulta infine:

$$\begin{aligned} \boldsymbol{\eta}_i &\sim \mathcal{N}_q(\boldsymbol{\mu}_i, \boldsymbol{\Phi}) \\ \boldsymbol{\delta}_i &\sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Theta}_\delta) \\ \mathbf{y}_i &= \boldsymbol{\tau}_i + \boldsymbol{\Lambda} \boldsymbol{\eta}_i + \boldsymbol{\delta}_i \end{aligned}$$

da cui si evince la relazione lineare tra variabili osservate (misurazioni) \mathbf{y} e fattori latenti $\boldsymbol{\eta}$, la cui intensità è rappresentata dalla matrice dei coefficienti $\boldsymbol{\Lambda}$ perturbata dalle componenti di errore casuale $\boldsymbol{\delta}_i$.

◀ ▶ ⏪ ⏩ 🔍 ↺

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

Solitamente si assume che le v.a. latenti $\boldsymbol{\eta}$ ed osservate \mathbf{y} abbiano media nulla, vale a dire:

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{0}_q \\ \boldsymbol{\tau} &= \mathbf{0}_p \end{aligned}$$

Ciò comporta una semplificazione del modello marginale, quest'ultimo diventando:

$$\mathbf{y}_i \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Lambda} \boldsymbol{\Phi} \boldsymbol{\Lambda}^T + \boldsymbol{\Theta}_\delta)$$

da cui si evince che gli unici parametri del modello delle osservate sono le componenti della matrice di varianze-covarianze

$$\boldsymbol{\Sigma}_y = \boldsymbol{\Lambda} \boldsymbol{\Phi} \boldsymbol{\Lambda}^T + \boldsymbol{\Theta}_\delta$$

che diventa l'oggetto delle procedure di stima ed inferenza statistica data la statistica osservata $\text{Cov}[\mathbf{y}_i]$ (nota: al posto del dato osservato \mathbf{y}_i).

◀ ▶ ⏪ ⏩ 🔍 ↺

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_{11}^2 & \dots & \phi_{1k} & \dots & \phi_{1q} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{k1} & \dots & \phi_{kk}^2 & \dots & \phi_{kq} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{q1} & \dots & \phi_{qk} & \dots & \phi_{qq}^2 \end{bmatrix}$$

Sulla base dell'ordine della matrice $\boldsymbol{\Phi}_{q \times q}$ è possibile avere:

- **modelli unidimensionali** ($q = 1$):
 $\boldsymbol{\Phi}_{q \times q} = \phi$ contiene solo la varianza ϕ^2 dell'unica v.a. latente η
- **modelli bidimensionali** ($q = 2$):
 $\boldsymbol{\Phi}_{q \times q} = \boldsymbol{\Phi}_{2 \times 2}$ contiene le varianze ϕ_1^2 e ϕ_2^2 della v.a. latenti η_1 ed η_2 più il termine di covarianza ϕ_{12} (o ϕ_{21}) tra le due componenti
- **modelli multidimensionali** ($q > 2$):
 $\boldsymbol{\Phi}_{q \times q}$ contiene q varianze sulla diagonale relative alle variabili latenti e $\frac{1}{2}(q \times q - q)$ termini di covarianza

◀ ▶ ⏪ ⏩ 🔍 ↺

CFA come modello a variabili latenti

FONTI: BKM(3.1-3.2)

Assumendo per semplicità $\Phi = \mathbf{I}_{q \times q}$ (*matrice identità*)[†], i vincoli $\mu = \mathbf{0}_q$ e $\tau = \mathbf{0}_p$ permettono di notare un risultato importante per quanto riguarda le varianze della variabili osservate. Fissando i e j , si ha infatti che

$$\begin{aligned}\text{Var}[Y_j] &= \lambda_j \lambda_j^T + \Theta_{\delta_{jj}} \\ &= \sum_{k=1}^q \lambda_k^2 + \delta_j\end{aligned}$$

la varianza dell'osservato è decomposta in due componenti denominate parte comune o **comunalità** ($\sum_{k=1}^q \lambda_k^2$) ed una parte residua o **unica** (δ_j) che raccoglie la parte di varianza della variabile Y_j non spiegata dalla presenza delle (*non riflessa nelle*) variabili latenti η .

[†] Tale vincolo sulla matrice Φ fa sì che le variabili latenti abbiano varianze pari a 1 e covarianze pari a zero. Questo modello è detto **ortogonale**, poiché presenta variabili latenti tra loro non correlate.

CFA: modello unidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)

Fissando i , dato un insieme di variabili osservate ($Y_1, \dots, Y_j, \dots, Y_p$) il modello fattoriale confermativo unidimensionale è un *modello lineare* definito dalle seguenti equazioni:

$$\begin{aligned}Y_1 &= \eta \lambda_1 + \delta_1 \\ &\vdots \\ Y_j &= \eta \lambda_j + \delta_j \\ &\vdots \\ Y_p &= \eta \lambda_p + \delta_p\end{aligned}$$

dove:

- η è la variabile latente (fattore) che sottende le osservabili
- $\delta_1, \dots, \delta_p$ sono le variabili di errore della misurazione
- $\lambda_1, \dots, \lambda_p$ sono i coefficienti che legano le osservabili al misurando latente (coefficienti fattoriali o *factor loadings*)

CFA: modello unidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)

In generale, come visto in precedenza,

$$\begin{aligned}Y_j &= \underbrace{\eta \lambda_j}_{\text{parte comune } c} + \underbrace{\delta_j}_{\text{parte unica } u_j} \\ &= c + u_j\end{aligned}$$

dove:

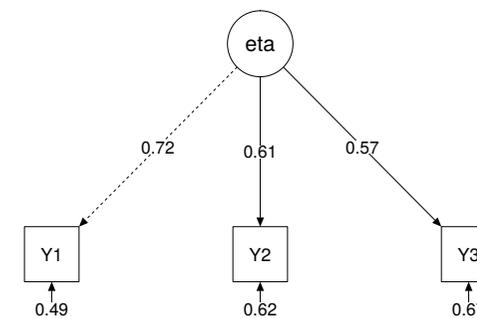
- c : componente del modello lineare *in comune* con le p osservabili
- u_j : componente del modello lineare specifica (*unica*) per ciascuna delle p osservabili

Le misure osservate Y_1, \dots, Y_p sono linearmente dipendenti da una parte comune c (pesata secondo $\lambda_1, \dots, \lambda_p$) ed una parte unica/residuale u_j specifica per ciascuna di esse.

In questa definizione, che ricalca quella della TCT, l'osservabile è in dipendenza lineare dal misurando e non viceversa (*modello riflessivo*).

CFA: modello unidimensionale

FONTI: CRB(VII.2.0,VII.2.1)



Modello CFA unidimensionale: Esempio di modello fattoriale (scala) unidimensionale con una sola variabile latente (misurando) e tre variabili osservate (indicatori). Nota: le variabili latenti sono rappresentate mediante cerchi mentre quelle osservate mediante quadrati, i numeri posti lungo le frecce orientate sono i coefficienti $\lambda_{Y1}, \lambda_{Y2}, \lambda_{Y3}$ mentre i numeri posti lungo le frecce orientate sui quadrati indicano i residui del modello $\delta_{Y1}, \delta_{Y2}, \delta_{Y3}$.

CFA: modello multidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)

Fissando i , dato un insieme di variabili osservate ($Y_1, \dots, Y_j, \dots, Y_p$) il modello fattoriale confermativo ad un solo fattore latente è generalizzabile al caso con q variabili latenti ($q < p$).

$$\begin{aligned} Y_1 &= \eta_1 \lambda_{11} + \dots + \eta_k \lambda_{k1} + \dots + \eta_q \lambda_{q1} + \delta_1 \\ &\vdots \\ Y_j &= \eta_1 \lambda_{1j} + \dots + \eta_k \lambda_{kj} + \dots + \eta_q \lambda_{qj} + \delta_j \\ &\vdots \\ Y_p &= \eta_1 \lambda_{1p} + \dots + \eta_k \lambda_{kp} + \dots + \eta_q \lambda_{qp} + \delta_p \end{aligned}$$

dove η_1, \dots, η_q sono i misurandi latenti del modello.



CFA: modello multidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)

Anche nel modello a più dimensioni latenti *vale* la decomposizione della varianza in comunaltà ed unicità, posto che

$$Y_j = \underbrace{\eta_1 \lambda_{1j} + \dots + \eta_k \lambda_{kj} + \dots + \eta_q \lambda_{qj}}_{\text{parte comune } c} + \underbrace{\delta_j}_{\text{parte unica } u_j} = c + u_j$$

dove c in questo caso raccoglie la componente del modello lineare riferita ai fattori latenti η_1, \dots, η_q .

Il modello con q misurandi può essere scritto in forma compatta usando la notazione algebrica (vedi slide 11-15):

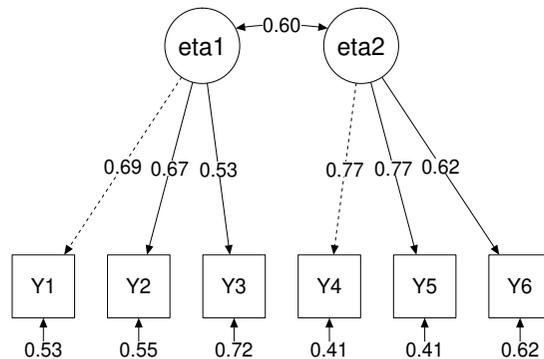
$$\mathbf{y}_{p \times 1} = \mathbf{\Lambda}_{p \times q} \boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{p \times 1}$$

dove in questo caso la matrice $\mathbf{\Lambda}_{p \times q}$ contiene i coefficienti che congiungono le variabili manifeste (in riga) alle variabili latenti (in colonna).



CFA: modello multidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)



Modello CFA multidimensionale: Esempio di modello fattoriale con due variabili latenti (misurandi) e sei variabili osservate (indicatori). Il grafico riporta due misurandi quantificati in modo unidimensionale: ciascuna delle due scale ha uno (ed uno solo) misurando. Nota: l'arco tra i due cerchi indicanti le variabili latenti rappresenta la correlazione tra i due misurandi.



CFA: modello multidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)

Esempio slide 22:

$$\mathbf{y}_{6 \times 1} = \mathbf{\Lambda}_{6 \times 2} \boldsymbol{\eta}_{2 \times 1} + \boldsymbol{\delta}_{6 \times 1}$$

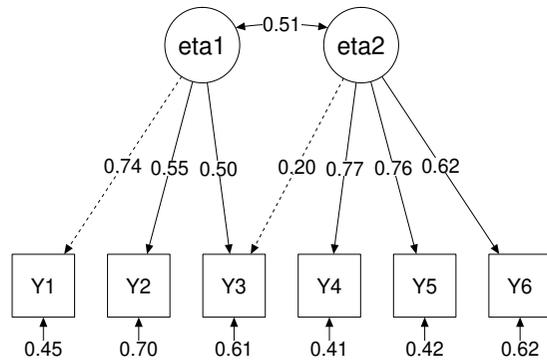
$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

La struttura della matrice $\mathbf{\Lambda}$ è tale per cui $\lambda = 0$ indica che l'indicatore non è legato al misurando (*parametro fissato*) mentre $\lambda \neq 0$ indica il legame da stimare tra misurando e indicatore (*parametro libero*).



CFA: modello multidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)



Modello CFA multidimensionale: Esempio di modello fattoriale con due variabili latenti (misurandi) e sei variabili osservate (indicatori). Il grafico riporta due misurandi quantificati in modo non unidimensionale: alcuni indicatori (in questo caso Y_3) sono condivisi da più di un misurando e la scala non rispetta l'ipotesi della TCT di unidimensionalità. Ciò può essere dovuto, ad esempio, ad una non ottimale scelta degli indicatori (bassa coerenza interna) oppure al fatto che i due misurandi siano espressione di un unico misurando di secondo livello.

CFA: modello multidimensionale

FONTI: CRB(III.5,VII.2.0,VII.2.1)

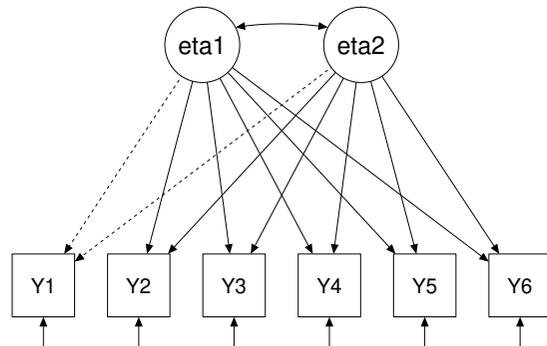
Esempio slide 24:

$$\mathbf{y}_{6 \times 1} = \mathbf{\Lambda}_{6 \times 2} \boldsymbol{\eta}_{2 \times 1} + \boldsymbol{\delta}_{6 \times 1}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

Analisi fattoriale esplorativa

FONTI: CRB(VII.2.2)



Modello EFA multidimensionale: Generalizzazione del modello confermativo al modello fattoriale esplorativo tramite manipolazione della matrice Λ : notiamo come il modello non esprima nessuna ipotesi specifica circa i legami tra le osservabili e i misurandi mentre, al contrario, tutti i misurandi sono legati a tutte le variabili latenti. Tale modello, sebbene utile in alcuni contesti, non è un buon modello nel senso della TCT ed apporta scarso valore interpretativo.

Analisi fattoriale esplorativa

FONTI: CRB(VII.2.2)

Esempio slide 26:

$$\mathbf{y}_{6 \times 1} = \mathbf{\Lambda}_{6 \times 2} \boldsymbol{\eta}_{2 \times 1} + \boldsymbol{\delta}_{6 \times 1}$$

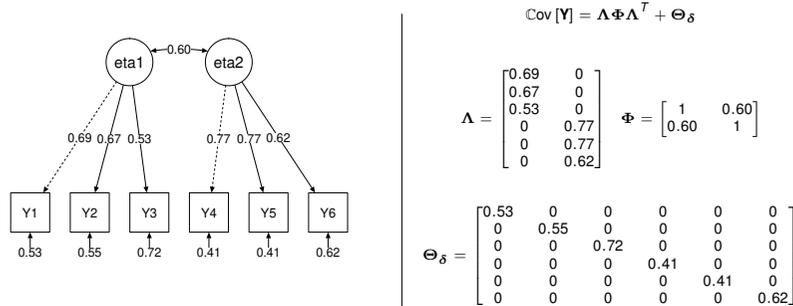
$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{bmatrix} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

Si noti come la struttura della matrice Λ contiene nel caso esplorativo tutti *parametri liberi*, ossia parametri da determinare (stimare) a partire dai dati.

CFA: matrice di covarianza riprodotta

FONTI: CRB(IV.2.1,VI.2.1)

Riprendiamo l'esempio rappresentato nella [slide 22](#):



CFA e TCT

FONTI: CRB(VII.3.0)

CFA e TCT sono fortemente interrelate. La CFA permette di costruire e valutare una scala di misura mediante un modello lineare che lega indicatori osservati a costrutti latenti. In questo senso, nell'ottica della CFA, *misurare* significa definire un modello confermativo e adattarlo ai dati, stimandone i parametri e valutandone l'adattamento complessivo.

CFA e TCT

FONTI: CRB(VII.3.0)

Rispetto alle **assunzioni** dei due modelli, la seguente corrispondenza può essere effettuata:

TCT	CFA
$\mathbb{E} [E_j] = 0$	$\mathbb{E} [\delta_j] = 0$
$\text{Cor} [E_j, T] = 0$	$\text{Cor} [\eta_j, \delta_j] = 0$
$\text{Var} [Y_j] = \text{Var} [T] + \text{Var} [E_j]$	$\text{Var} [Y_j] = \lambda^2 \phi + \theta_{\delta}^{\dagger}$

[†] L'espressione si ottiene dalla definizione generale $\text{Var} [Y_j] = \text{Var} [\eta_j] + \text{Var} [\delta_j]$ notando dalla slide 32 che:

$$\text{Var} [\delta_j] = \theta_{\delta} \text{ e } \text{Var} [\eta_j] = \lambda \phi \lambda = \lambda^2 \phi$$

CFA e TCT

FONTI: CRB(VII.3.0)

Rispetto alla definizione di **attendibilità** di una scala, nella CFA l'attendibilità è definita utilizzando i parametri del modello mentre la TCT ricorre all'utilizzo di misurazioni parallele e della covarianza tra loro, sfruttando l'informazione presente nei dati. In particolare, notiamo quanto segue:

$$\rho_{YT}^2 = \frac{\lambda^2 \text{Var} [\eta]}{\text{Var} [Y]} = \frac{\lambda^2 \phi}{\lambda^2 \phi + \theta_{\delta}}$$

da cui si può dedurre l'uguaglianza tra misurando secondo la TCT e misurando secondo la CFA:

$$T = \lambda \phi$$

Notiamo inoltre che quando $\lambda = 1$, la definizione dell'attendibilità secondo la CFA uguaglia quella fornita dalla TCT[†]:

$$\rho_{YT}^2 = \frac{\text{Var} [\eta]}{\text{Var} [Y]} = \frac{\phi}{\phi + \theta_{\delta}}$$

[†] L'uguaglianza è valida notando che la definizione di attendibilità è quella definita dalla correlazione intraclassa (vedi modulo II-A).

CFA e TCT

FONTI: CRB(VII.3.0)

Nel contesto della CFA, un indice di attendibilità alternativo all' α di Chronbach è l'**indice** ω , il quale utilizza direttamente i parametri del modello CFA una volta che questi siano stati stimati:

$$\omega_{YT} = \frac{\sum_{j=1}^p \lambda_j^2}{\sum_{j=1}^p \lambda_j^2 + \theta_{\delta_j}^2}$$

L'indice si interpreta in maniera analoga all' α di Chronbach sebbene, rispetto a quest'ultimo, utilizza un modello dei dati e non questi direttamente mediante la statistica di covarianza.

In alcuni casi, ad esempio quando vale la τ -equivalenza, $\omega_{YT} = \alpha$. Si veda BN(4.3.0) per ulteriori dettagli.



CFA e TCT

CFA come verifica dei modelli di misurazione TCT

Rispetto ai modelli di misurazione implicati dalla TCT (si veda BN(2.4)), nel caso unidimensionale $q = 1$ con $\mu = 0$ e $\tau = \mathbf{0}_p$, otteniamo i seguenti sotto-modelli CFA.

Test paralleli: p misure parallele \mathbf{y} hanno identico punteggio vero ed identica varianza di errore. Il sotto-modello CFA corrispondente è il seguente:

$$\text{Cov}[\mathbf{y}] = \lambda \mathbf{1}_p \phi \mathbf{1}_p^T + \mathbf{I} \theta_{\delta}$$

Note:

- (i) ϕ e λ sono scalari, $\mathbf{1}_p$ è un vettore $p \times 1$ di tutti 1
- (ii) \mathbf{I} è una matrice $p \times p$ identità (solo 1 lungo la diagonale e 0 altrimenti)
- (iii) \mathbf{y} denota un vettore aleatorio di dimensione $p \times 1$



CFA e TCT

CFA come verifica dei modelli di misurazione TCT

Rispetto ai modelli di misurazione implicati dalla TCT (si veda BN(2.4)), nel caso unidimensionale $q = 1$ con $\mu = 0$ e $\tau = \mathbf{0}_p$, otteniamo i seguenti sotto-modelli CFA.

Test τ -equivalenti: p misure parallele \mathbf{y} hanno identico punteggio vero ed diversa varianza di errore. Il sotto-modello CFA corrispondente è il seguente:

$$\text{Cov}[\mathbf{y}] = \lambda \mathbf{1}_p \phi \mathbf{1}_p^T + \Theta_{\delta}$$

Note:

- (i) ϕ e λ sono scalari, $\mathbf{1}_p$ è un vettore $p \times 1$ di tutti 1
- (ii) \mathbf{y} denota un vettore aleatorio di dimensione $p \times 1$



CFA e TCT

CFA come verifica dei modelli di misurazione TCT

Rispetto ai modelli di misurazione implicati dalla TCT (si veda BN(2.4)), nel caso unidimensionale $q = 1$ con $\mu = 0$ e $\tau = \mathbf{0}_p$, otteniamo i seguenti sotto-modelli CFA.

Test congenerici: p misure parallele \mathbf{y} hanno identico punteggio vero ed diversa varianza di errore. Il sotto-modello CFA corrispondente è il seguente:

$$\text{Cov}[\mathbf{y}] = \lambda \phi \lambda^T + \Theta_{\delta}$$

Note:

- (i) ϕ è uno scalare, $\mathbf{1}_p$ è un vettore $p \times 1$ di tutti 1
- (ii) \mathbf{y} denota un vettore aleatorio di dimensione $p \times 1$



CFA e TCT

CFA come verifica dei modelli di misurazione TCT

Rispetto ai modelli di misurazione implicati dalla TCT (si veda BN(2.4)), nel caso unidimensionale $q = 1$ con $\mu = 0$ e $\tau = \mathbf{0}_p$, otteniamo i seguenti sotto-modelli CFA.

modello TCT	modello CFA
parallelo	$\text{Cov}[\mathbf{y}] = \lambda \mathbf{1}_p \phi \mathbf{1}_p^T + \mathbf{I} \theta_\delta$
τ -equivalente	$\text{Cov}[\mathbf{y}] = \lambda \mathbf{1}_p \phi \mathbf{1}_p^T + \Theta_\delta$
congenerico	$\text{Cov}[\mathbf{y}] = \lambda \phi \lambda^T + \Theta_\delta$

CFA e TCT

CFA come verifica dei modelli di misurazione TCT



per un approfondimento tra CFA e modelli di misurazione implicati dalla TCT si veda in Approfondimenti al corso:

- ▶ Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate behavioral research*, 26(3), 479-497

CFA: stime dei parametri del modello

FONTE: CRB(IV.1,IV.2.1,IV.3), BKM(3.4,8.6.3)

Il modello CFA è parametrizzato secondo la matrice Λ (coefficienti che legano le osservabili alle latenti), la matrice Φ (che contiene le varianze delle latenti e le covarianze quando $q > 1$) e la matrice Θ_δ (che contiene le varianze e le covarianze dei residui del modello).

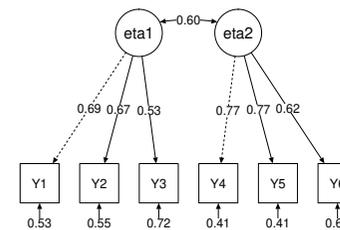
L'approccio confermativo - proprio perché forza il ricercatore a fare delle ipotesi sul modello di misura - non necessita di stimare/determinare le matrici Λ , Φ , Θ_δ nella loro interezza ma solo quelle componenti di queste che sono diverse da zero (*parametri liberi*).

Un modello confermativo deve essere il *più parsimonioso possibile* nel senso di avere pochi parametri liberi e più parametri fissi: questo si realizza, appunto, nella definizione di un modello basato sulle ipotesi del ricercatore.

CFA: stime dei parametri del modello

FONTE: CRB(IV.1,IV.2.1,IV.3), BKM(3.4,8.6.3)

Nell'esempio della slide 22 abbiamo 13 parametri liberi (da stimare):



$$\text{Cov}[\mathbf{Y}] = \Lambda \Phi \Lambda^T + \Theta_\delta$$

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{12} \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \end{bmatrix} \quad \Phi = \begin{bmatrix} 1 & 0 \\ \phi_{21} & 1 \end{bmatrix}$$

$$\Theta_\delta = \begin{bmatrix} \theta_{\delta_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\delta_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{\delta_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{\delta_4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{\delta_5} & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_6} \end{bmatrix}$$

CFA: stime dei parametri del modello

FONTI: CRB(IV.1,IV.2.1,IV.3,VI.4), BKM(3.4,8.6.3)

La parsimonia del modello è quantificata dai **gradi di libertà** che, nel caso della CFA, si esprimono come di seguito:

$$df = \frac{p(p+1)}{2} - m$$

dove p indica il numero di indicatori del modello mentre m esprime il numero di parametri liberi del modello.

Nell'esempio della [slide 22](#) abbiamo $m = 13$ e $\frac{p(p+1)}{2} = 21$ per un totale di $df = 8$ gradi di libertà.

I gradi di libertà spesso possono essere utilizzati per la scelta di diversi modelli adattati su uno stesso insieme di dati \mathbf{Y} : a parità di matrice di covarianza osservata \mathbf{S}_Y , infatti, si preferirà il modello con più gradi di libertà, dunque quello più parsimonioso.



CFA: stime dei parametri del modello

FONTI: CRB(VI.6)

Per agevolare la stima dei parametri del modello confermativo deve essere definita la **metrica dei misurandi**:

- approccio ULI (*unit loading identification*): si fissa ad 1 il primo dei coefficienti $\lambda_1, \dots, \lambda_p$ per ciascuna delle q variabili latenti. Nell'esempio della [slide 22](#): $\lambda_{11} = 1$ e $\lambda_{21}, \lambda_{31}$ si lasciano liberi di essere stimati.
- approccio UVI (*unit variance identification*): si fissa ad 1 la varianza delle variabili latenti ϕ_1, \dots, ϕ_q . Nell'esempio della [slide 22](#): $\phi_{\eta_1} = \phi_{\eta_2} = 1$.

I due approcci sono mutualmente esclusivi: non è possibile applicarli entrambi.



CFA: stime dei parametri del modello

FONTI: CRB(IV.1,IV.2.1,IV.3), BKM(3.4,8.6.3)

La stima dei parametri avviene mediante una procedura iterativa che cerca di minimizzare la distanza tra la matrice osservata \mathbf{S}_Y e la matrice di covarianza del modello Σ_Y : minore è la distanza, meglio il modello riproduce i dati osservati.

Diversi sono gli approcci utilizzabili rispetto ai parametri ignoti $\Sigma_Y = \mathbf{\Lambda}\Phi\mathbf{\Lambda}^T + \Theta_\delta$:

- **massima verosimiglianza**: riprendendo la densità marginale del modello (vedi slide 12), con $\mu = \mathbf{0}_q$ e $\tau = \mathbf{0}_p$, si massimizza la funzione di verosimiglianza

$$l(\Sigma_Y) \approx \frac{n}{2} \left(\ln |\Sigma_Y^{-1}| - \text{trace} \left(\Sigma_Y^{-1} \mathbf{S}_Y \right) \right)$$



CFA: stime dei parametri del modello

FONTI: CRB(IV.1,IV.2.1,IV.3), BKM(3.4,8.6.3)

- **minimi quadrati non pesati** (ULS): il metodo minimizza il funzionale

$$d_{ULS} = \frac{1}{2} \text{trace} \left((\mathbf{S}_Y - \Sigma_Y)^2 \right)$$

- **minimi quadrati generalizzati** (GLS): il metodo minimizza il funzionale

$$d_{ULS} = \frac{1}{2} \text{trace} \left(\left(\mathbf{S}_Y^{-\frac{1}{2}} (\mathbf{S}_Y - \Sigma_Y) \mathbf{S}_Y^{-\frac{1}{2}} \right)^2 \right)$$



CFA: stime dei parametri del modello

FONTI: CRB(IV.1,IV.2.1,IV.3), BKM(3.4,8.6.3)

- altri approcci basati sul razionale dei minimi quadrati (es.: **WLS**, minimi quadrati pesati) o della massima verosimiglianza (es.: **EM**, expectation-maximization)
- approcci Monte-Carlo (es.: **MCMC**, markov-chain monte-carlo)

CFA: stime dei parametri del modello

FONTI: CRB(IV.1,IV.2.1,IV.3), BKM(3.4,8.6.3)

Alla fine della procedura di stima i parametri stimati (tra quelli liberi) risultano:

- $\hat{\Lambda}$ matrice dei coefficienti del modello fattoriale (*factor loadings*)
- $\hat{\Phi}$ matrice delle varianze/covarianze tra i fattori latenti
- $\hat{\Theta}_\delta$ matrice delle varianze/covarianze degli errori del modello

i quali vengono utilizzati fare inferenza, come la valutazione di test di ipotesi mediante l'utilizzo di test asintotici, e per interpretare il modello finale.

Vedremo la parte di inferenza sui parametri direttamente nei laboratori mentre nelle slide che seguono ci concentreremo sull'interpretazione del modello finale.

CFA: interpretazione di $\hat{\Lambda}$

FONTI: CRB(VII.2.1-VII.2.3), BKM(2.12)

Dopo aver stimato i parametri del modello, l'interpretazione della matrice Λ fornisce informazioni circa l'abilità della scala di quantificare il costrutto indagato. Ciò si lega alla validità fattoriale della scala nonché all'attendibilità, intesa questa nei termini della coerenza interna.

In pratica, si procede ricercando quegli elementi di Λ che presentano alti valori (in termine di magnitudine): rispetto alla k -esima variabile latente (k -esima colonna della matrice Λ), maggiore è il valore dei coefficienti fattoriali $\lambda_{1k}, \dots, \lambda_{pk}$, maggiore sarà il contributo che il k -esimo fattore eserciterà sugli indicatori manifesti ad esso collegati.

Più alti saranno i valori dei parametri $\lambda_{1k}, \dots, \lambda_{pk}$, più forte sarà l'abilità che i p indicatori avranno nel riflettere il k -esimo costrutto indagato.

Si adotta il principio semplice del *looking for what is in common*.

CFA: interpretazione di $\hat{\Lambda}$

FONTI: CRB(VII.2.1-VII.2.3), BKM(2.12)

Consideriamo la seguente *factor pattern matrix* che lega 10 indicatori osservati a 3 variabili latenti:

$$\hat{\Lambda}_{10 \times 3} = \begin{bmatrix} + & . & . \\ + & . & . \\ + & . & . \\ . & + & . \\ . & + & . \\ . & + & . \\ . & + & . \\ . & . & + \\ . & . & + \\ . & . & + \end{bmatrix}$$

dove “+” indica un alto valore del coefficiente fattoriale mentre “.” indica un valore trascurabile (es.: prossimo allo zero). L'interpretazione si concentra su quali variabili osservate (in riga) sono in comune con i fattori (in colonna).

Ad esempio, η_2 (seconda colonna) condivide le osservate Y_4, Y_5, Y_6, Y_7 (righe 4-7).

CFA: interpretazione di $\hat{\Lambda}$

FONTI: CRB(VII.2.1-VII.2.3), BKM(2.12)

$$\hat{\Lambda}_{10 \times 3} = \begin{bmatrix} + & . & . \\ + & . & . \\ + & . & . \\ . & + & . \\ . & + & . \\ . & + & . \\ . & + & . \\ . & . & + \\ . & . & + \\ . & . & + \end{bmatrix}$$

È facile identificare tre gruppi distinti (non sovrapposti) di indicatori manifesti che legano a tre costrutti indagati.

Nel caso in cui $\hat{\Phi} = \mathbf{I}$ (variabili latenti incorrelate, *ortogonali*), la matrice $\hat{\Lambda}$ rappresenta quello che è denominato *modello fattoriale a struttura semplice (simple structure)* dove ciascuna variabile latente è associata ad un solo gruppo di variabili manifeste.



CFA: interpretazione di $\hat{\Lambda}$

FONTI: CRB(VII.2.1-VII.2.3), BKM(2.12)

Sebbene tale situazione sia pressoché rara nella pratica dell'analisi dei dati (si vedano gli esempi esposti fino ad ora), essa identifica una situazione interpretativa ottimale e si coniuga con il criterio dell'*unidimensionalità delle scale* evocato dalla TCT.

Per ragioni di semplicità interpretativa dei risultati, spesso la matrice stimata $\hat{\Lambda}$ è sottoposta ad una procedura di **rotazione** che permette ad essa di raggiungere una struttura che approssima quella semplice (favorendo l'interpretazione dei parametri del modello e della struttura fattoriale).

Esistono diverse scelte per la rotazione di $\hat{\Lambda}$:

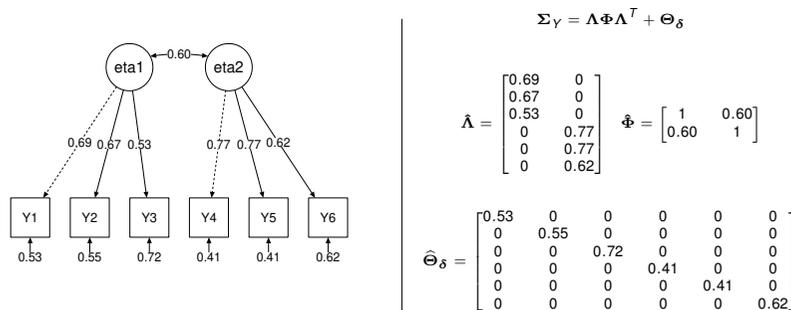
- ortogonale (es.: *varimax*, *quartimax*)
- obliqua (es.: *oblimin*)



CFA: interpretazione di $\hat{\Lambda}$

FONTI: CRB(VII.2.1-VII.2.3), BKM(2.12)

Riprendiamo l'esempio rappresentato nella [slide 22](#):



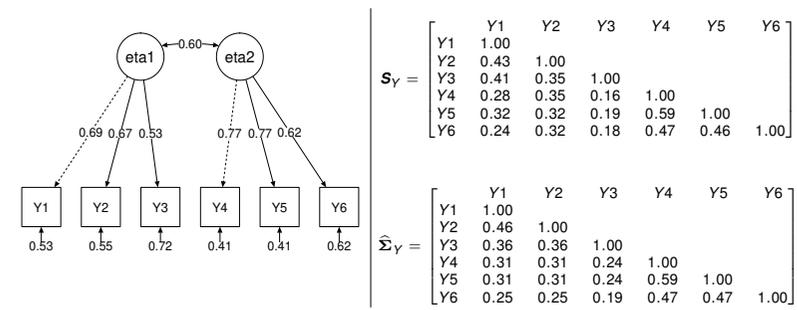
Esempio di stima dei parametri del modello



CFA: interpretazione di $\hat{\Lambda}$

FONTI: CRB(VII.2.1-VII.2.3), BKM(2.12)

Riprendiamo l'esempio rappresentato nella [slide 22](#):



Esempio di matrice di covarianza riprodotta dal modello $\hat{\Sigma}_Y$ dopo aver stimato i parametri rispetto alla matrice di covarianza osservata \mathbf{S}_Y .



CFA: standardizzazione dei parametri del modello

Una volta stimate le matrici dei parametri del modello CFA, è possibile facilitare l'interpretazione dei parametri mediante **standardizzazione**. Questa ha l'effetto di trasformare i coefficienti in modo che questi si esprimano nell'intervallo unitario.

La standardizzazione solitamente avviene utilizzando $\text{diag}(\Phi)$ e $\text{diag}(\Sigma_Y)$, ossia le varianze delle variabili latenti e quelle delle osservate, quale fattore di standardizzazione.

Maggiori dettagli su questa procedura sono disponibili nella cartella Approfondimenti nel file `standardizzazione_parametri_CFA.pdf`.

CFA: adeguatezza dell'adattamento del modello

FONTE: CRB(I.3,VI.1,VI.3,VI.4), BKM(8.7)

Test del χ^2 (chi-quadrato)

Per valutare l'adeguatezza complessiva del modello ai dati si utilizza una statistica basata sulla funzione di verosimiglianza $l(\Sigma_Y)$ (vedi slide 43). In particolare, per la stessa matrice di covarianza osservata \mathbf{S}_Y si definiscono e si adattano due modelli CFA:

- (a) il *modello corrente* \mathcal{M}_1 con df_1 gradi di libertà (vedi slide 41)
- (b) il *modello saturo* \mathcal{M}_{sat} con $df_{\text{sat}} = 0$ gradi di libertà (baseline)

Nota: un modello CFA si definisce *saturo* quando contiene tante variabili latenti quante sono le variabili osservate ($q = p$). In questo caso si ha che $\Lambda = \mathbf{I}_{p \times p}$, Θ_δ è diagonale e i gradi di libertà sono pari a zero.



CFA: adeguatezza dell'adattamento del modello

FONTE: CRB(I.3,VI.1,VI.3,VI.4), BKM(8.7)

Test del χ^2 (chi-quadrato)

Per valutare l'adeguatezza complessiva del modello ai dati si utilizza una statistica basata sulla funzione di verosimiglianza $l(\Sigma_Y)$ (vedi slide 43). In particolare, per la stessa matrice di covarianza osservata \mathbf{S}_Y si definiscono e si adattano due modelli CFA:

- (a) il *modello corrente* \mathcal{M}_1 con df_1 gradi di libertà (vedi slide 41)
- (b) il *modello saturo* \mathcal{M}_{sat} con $df_{\text{sat}} = 0$ gradi di libertà (baseline)

Successivamente si calcola la seguente statistica basata sulle funzioni di verosimiglianza dei due modelli:

$$\chi_1^2 = -2(l(\Sigma_Y)_1 - l(\Sigma_Y)_{\text{sat}})$$

che sotto l'ipotesi nulla di assenza di differenza tra i due modelli si distribuisce come:

$$\chi_1^2 \stackrel{H_0}{\sim} \chi_{(df_1)}^2$$



CFA: adeguatezza dell'adattamento del modello

FONTE: CRB(I.3,VI.1,VI.3,VI.4), BKM(8.7)

Test del χ^2 (chi-quadrato)

Per valutare l'adeguatezza complessiva del modello ai dati si utilizza una statistica basata sulla funzione di verosimiglianza $l(\Sigma_Y)$ (vedi slide 43). In particolare, per la stessa matrice di covarianza osservata \mathbf{S}_Y si definiscono e si adattano due modelli CFA:

- (a) il *modello corrente* \mathcal{M}_1 con df_1 gradi di libertà (vedi slide 41)
- (b) il *modello saturo* \mathcal{M}_{sat} con $df_{\text{sat}} = 0$ gradi di libertà (baseline)

Come al solito, se $\chi_1^2 > \chi_{1-\alpha/2}^2$ (valore critico della statistica test) allora H_0 è rigettata e dunque il modello saturo non si adatta bene a \mathbf{S}_Y .

Nota: come ogni decisione basata sulla statistica χ^2 , quando la numerosità campionaria n è molto elevata il test di ipotesi tende a favorire il modello corrente. Per questo motivo si preferisce altresì utilizzare altri indici per valutare il modello.



CFA: adeguatezza dell'adattamento del modello

FONTE: CRB(I.3,VI.1,VI.3,VI.4), BKM(8.7)

Altri indici

Altri indici vengono spesso utilizzati insieme ed in maniera comparativa a RMSEA, tra cui CFI (Bentler, 1990), TLI/NNFI, AGFI/GFI. Questi ultimi classificano un modello come ottimo quando il loro valore supera 0.95 mentre considerano un modello come accettabile quando il loro valore è superiore a 0.90.

CFA: adeguatezza dell'adattamento del modello

FONTE: CRB(I.3,VI.1,VI.3,VI.4), BKM(8.7)

Altri indici: CFI

In particolare, l'indice CFI è calcolato mediante la statistica χ_1^2 , comparando il modello corrente \mathcal{M}_1 con il modello nullo $\mathcal{M}_{\text{null}}$ come segue:

$$\text{CFI} = \frac{(\chi_1^2 - \text{df}_1)}{\max(\chi_1^2 - \text{df}_1, \chi_{\text{sat}}^2 - \text{df}_{\text{sat}})}$$

Nota: un modello CFA saturo è del tipo $\text{Cov}[\mathbf{y}] = \phi \mathbf{1}_{p \times p} + \Theta_\delta$, ossia un modello che ha come parametri liberi solo varianze e covarianze (*worst-fitting model*) mentre la matrice dei coefficienti fattoriali è fissata a zero $\Lambda = \mathbf{0}_{p \times q}$.

CFA: adeguatezza dell'adattamento del modello

FONTE: BKM(3.7.1)

AIC e comparazione tra modelli

Un ulteriore modo per valutare l'adattamento del modello ai dati è di confrontare quest'ultimo con un modello di base (*null model*) e/o con più modelli annidati (*nested models*). L'indice per valutare la discrepanza tra modello teorico corrente \mathcal{M}_1 e modello nullo $\mathcal{M}_{\text{null}}$ (oppure più modelli alternativi $\mathcal{M}_2, \mathcal{M}_3, \dots$) è AIC (*Akaike Information Index*):

$$\text{AIC} = -2l(\Sigma_y) + 2m$$

dove m è il numero di parametri liberi del modello (vedi slide 41).

Il modello che meglio si adatta ai dati è dunque quello che ottiene AIC più basso (*the lower, the better*).

CFA: adeguatezza dell'adattamento del modello

FONTE: CRB(VI.5.2)

Indici di modifica

Permettono di valutare come la statistica χ_1^2 cambia se i parametri fissi (cioè quelli non stimati) del modello venissero liberati (cioè stimati). Indici sensibilmente superiori a 4 indicano che l'adattamento complessivo del modello migliorerebbe se quel parametro venisse incluso (stimato) nel modello ipotizzato.

La procedura si basa essenzialmente su una logica incrementale dove un parametro viene liberato alla volta, il modello con il nuovo parametro incluso viene ri-stimato e gli indici di modifica vengono nuovamente calcolati e ciascuno di essi valutati secondo la regola precedente.

CFA: adeguatezza dell'adattamento del modello

FONTI: CRB(VI.5.3)

Analisi dei residui

Permette di individuare quelle variabili osservate del modello che, sebbene incluse all'interno di una scala, non sono coerenti con il misurando con la conseguenza di ottenere $\lambda \approx 0$ e residui δ elevati.

L'analisi dei residui si basa sull'osservazione della matrice $\mathbf{S}_Y - \widehat{\Sigma}_Y$ alla ricerca di quei coefficienti $\theta_{\delta_{j'}}$ che, rispetto a due variabili fissate j e j' , risultano elevati. Una volta individuato il residuo, si libera il parametro della matrice Θ_δ corrispondente e si ristima il modello con il nuovo parametro incluso. Ciò equivale alla **correlazione tra gli errori delle variabili osservate**.

Nota: questa procedura - come anche quella basata sugli indici di modifica - deve essere condotta con molta cautela: difatti è sempre possibile migliorare un modello correlando gli errori delle variabili osservate. Il risultato è avere modelli con alto indice di adattamento (*fit* elevato) ma *scarsa* capacità predittiva e interpretativa.



CFA: adeguatezza dell'adattamento del modello

FONTI: CRB(VI.5.3)

La matrice degli errori Θ_δ contiene l'informazione necessaria per effettuare l'analisi dei residui:

$$\Theta_\delta = \begin{bmatrix} \theta_{\delta_{11}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\delta_{22}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{\delta_{jj}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{pp}} \end{bmatrix}$$

In generale, in un modello di misura coerente con la TCT, gli unici parametri liberi sono quelli posti lungo la diagonale della matrice, ossia le varianze della componente osservata residua.



CFA: adeguatezza dell'adattamento del modello

FONTI: CRB(VI.5.3)

$$\Theta_\delta = \begin{bmatrix} \theta_{\delta_{11}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\delta_{22}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & \theta_{\delta_{j2}} & 0 & \theta_{\delta_{jj}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{pp}} \end{bmatrix}$$

Al contrario, liberare un parametro della matrice - ad esempio $\theta_{\delta_{2j}}$ - equivale a correlare gli errori δ della variabile osservata 2 e j e ciò equivale a dire che le due osservabili esprimono, in qualche modo, una medesima informazione sul misurando η (e quindi sono "accomunabili", convogliano una medesima informazione).

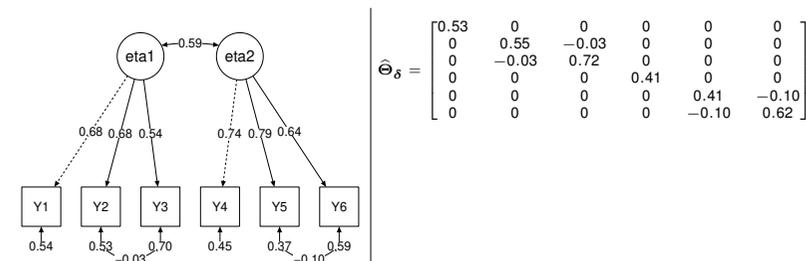
Il risultato pratico è che avere due variabili manifeste con errori correlati è come affermare - indirettamente - che una delle due è *ridondante*.



CFA: adeguatezza dell'adattamento del modello

FONTI: CRB(VI.5.3)

Riprendiamo l'esempio rappresentato nella [slide 22](#):



Esempio di modello confermativo con errori correlati: i parametri in oggetto sono $\theta_{\delta_{32}}$ e $\theta_{\delta_{66}}$.



CFA: calcolo dei punteggi fattoriali

FONTI: BKM(3.14)

Dopo aver stimato i parametri del modello $\hat{\Lambda}$, $\hat{\Phi}$, $\hat{\Theta}_\delta$ ed aver valutata l'adeguatezza del modello fattoriale, è possibile predire i valori (ignoti) degli n individui sulle variabili latenti

$$\hat{\eta}_i = \mathbb{E}[\eta_i | \mathbf{y}_i]$$

che prendono il nome di **punteggi fattoriali** (*factor scores*). Questi possono essere utilizzati in svariati modi, ad esempio per classificare le unità statistiche sulla base della quantità posseduta di costruito latente. Allo stesso modo, possono essere utilizzati per ulteriori analisi dei dati, ad esempio possono essere predetti da altre variabili concomitanti mediante modelli di regressione.

Diversi sono gli stimatori lineari che possono essere impiegati per stimare $\hat{\eta}_i$, tra cui i metodi di Bartlett, Thurstone e Anderson. Ciò che li accomuna è l'utilizzo della matrice del modello riprodotta $\hat{\Sigma}_y$.

Nota: il calcolo dei punteggi fattoriali equivale al calcolo dei punteggi veri nel caso della TCT (vedi modulo II-A).

CFA: caso studio

Studio della dimensionalità della WISC-IV

La WISC-IV (2012) rappresenta ad oggi uno strumento per la quantificazione dell'intelligenza di bambini di età 6-11 anni. Il modello teorico sotteso allo strumento comprende concettualizzazioni dell'intelligenza che la rappresentano in termini multifattoriali, con enfasi sul ragionamento fluido, memoria di lavoro e velocità di elaborazione degli stimoli.

Da un punto di vista psicometrico, lo strumento quantifica il misurando *intelligenza* mediante la misura QI (quoziente intellettivo) totale utilizzando quattro scale:

- 1 Comprensione verbale (VCI)
- 2 Ragionamento percettivo (PRI)
- 3 Memoria di lavoro (WMI)
- 4 Velocità di elaborazione (PSI)

CFA: caso studio

Studio della dimensionalità della WISC-IV

Ciascuna scala è costituita utilizzando le seguenti variabili osservate (indicatori principali):

- 1 VCI: similarities (SO), vocabulary (VC), comprehension (CO)
- 2 PRI: block design (DC), picture concepts (CI), matrix reasoning (RM)
- 3 WMI: digit span (MC), letter-number sequencing (LN)
- 4 PSI: coding (CR), symbol search (RS)

Le variabili osservate sono quantificate mediante svolgimento di compiti di performance cognitiva. Ciascuna scala, inoltre, possiede anche alcuni indicatori secondari che possono essere usati all'occorrenza.



Per maggiori dettagli sulla storia, la codifica, la validazione, le teorie che sottendono la WISC-IV si consulti il file `Agnoli_WISC_IV.pdf` disponibile negli Approfondimenti al corso.

CFA: caso studio

Studio della dimensionalità della WISC-IV

Il nostro obiettivo è quello di studiare lo strumento WISC-IV mediante la tecnica della CFA. Punto di partenza è la matrice di covarianza delle variabili osservate⁽¹⁾ ($n = 2200$):

$$S_y = \begin{bmatrix} DC & SO & MC & CI & CR & VC & LN & RM & CO & RS \\ 1.00 & & & & & & & & & \\ 0.38 & 1.00 & & & & & & & & \\ 0.26 & 0.35 & 1.00 & & & & & & & \\ 0.34 & 0.43 & 0.28 & 1.00 & & & & & & \\ 0.25 & 0.14 & 0.15 & 0.11 & 1.00 & & & & & \\ 0.33 & 0.62 & 0.33 & 0.41 & 0.13 & 1.00 & & & & \\ 0.29 & 0.35 & 0.42 & 0.35 & 0.19 & 0.38 & 1.00 & & & \\ 0.42 & 0.41 & 0.29 & 0.43 & 0.20 & 0.40 & 0.35 & 1.00 & & \\ 0.27 & 0.51 & 0.24 & 0.35 & 0.15 & 0.59 & 0.30 & 0.30 & 1.00 & \\ 0.30 & 0.27 & 0.20 & 0.24 & 0.46 & 0.24 & 0.24 & 0.26 & 0.22 & 1.00 \end{bmatrix}$$

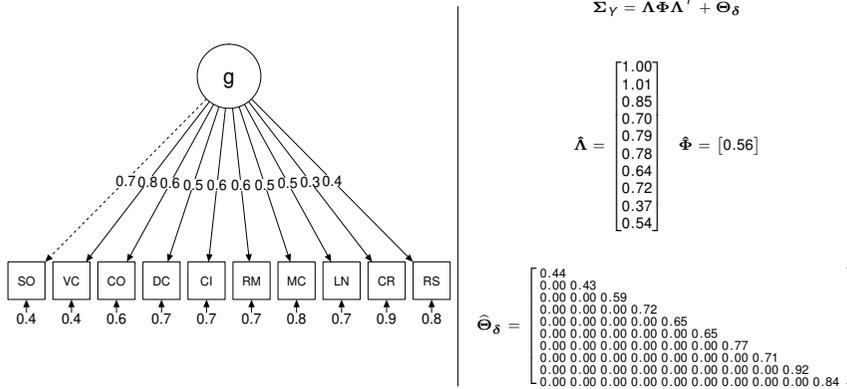
Procederemo esplorando quattro modelli fattoriali diversi (M1-M4) e valuteremo la loro adeguatezza rispetto ai dati campionari.

⁽¹⁾ Orsini A., Pezzuti L., & Hulbert S. (2015) Beyond the floor effect on the Wechsler Intelligence Scale for Children – 4th Ed. (WISC-IV): calculating IQ and indexes of subjects presenting a floored pattern of results. *Journal of Intellectual Disability Research* 59, 468–73.

CFA: caso studio

Studio della dimensionalità della WISC-IV

M1: Scala g



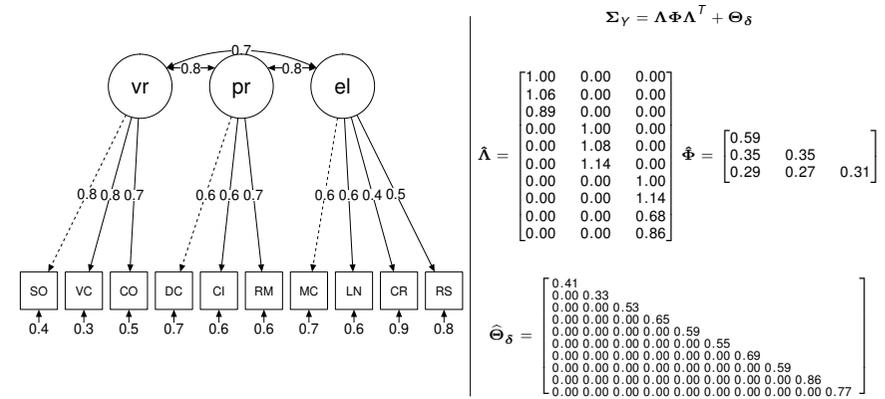
RMSEA = 0.11, CFI = 0.84, AIC = 57359

Nota: i coefficienti rappresentanti nel grafico sono *standardizzati* mentre quelli riportati nelle matrici a destra sono *assoluti*.

CFA: caso studio

Studio della dimensionalità della WISC-IV

M2: Scala verbale (vr), percettiva (pr) ed elaborazione informazione (el)



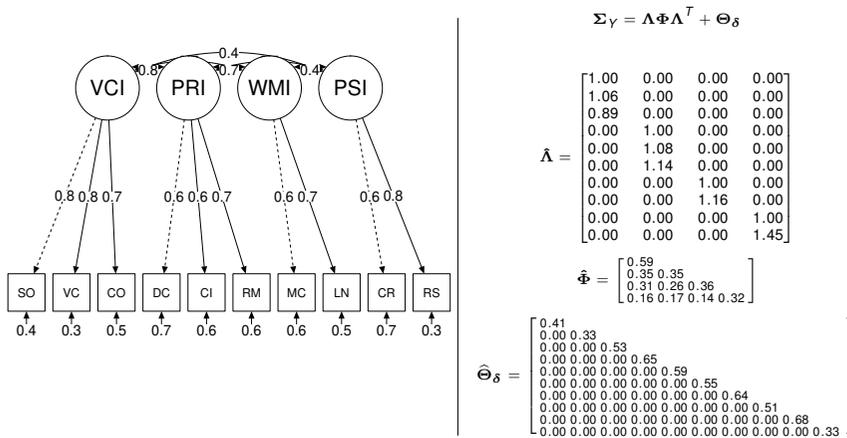
RMSEA = 0.08, CFI = 0.91, AIC = 56904

Nota: i coefficienti rappresentanti nel grafico sono *standardizzati* mentre quelli riportati nelle matrici a destra sono *assoluti*.

CFA: caso studio

Studio della dimensionalità della WISC-IV

M3: Scala VCI, PRI, WMI, PSI



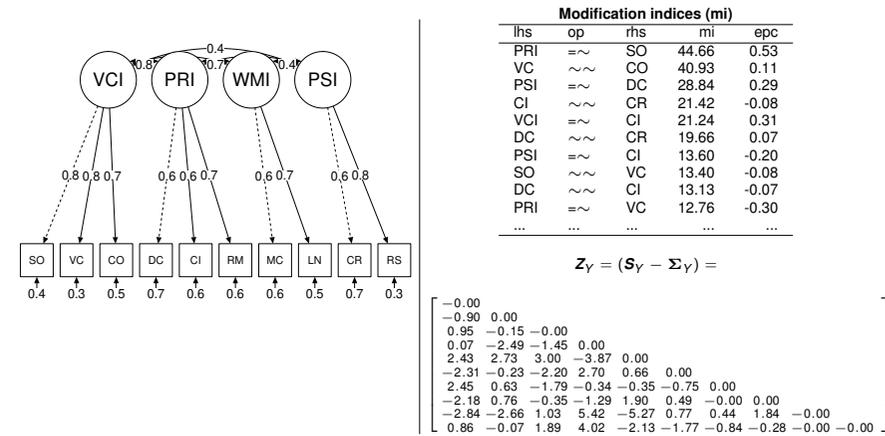
RMSEA = 0.04, CFI = 0.97, AIC = 56548

Nota: i coefficienti rappresentanti nel grafico sono *standardizzati* mentre quelli riportati nelle matrici a destra sono *assoluti*.

CFA: caso studio

Studio della dimensionalità della WISC-IV

M3: Scala VCI, PRI, WMI, PSI



RMSEA = 0.04, CFI = 0.97, AIC = 56548

Nota: Z_Y indica la matrice dei *residui* $\hat{\Theta}_\delta$ standardizzata per il loro errore standard. Il simbolo \sim indica dipendenza mentre $\sim\sim$ correlazione (si veda la simbologia usata dalla libreria di R *lavaan* - vedi laboratori).

CFA: caso studio

CFA e EFA



si consideri il caso studio per la EFA descritto in BKM(3.17) come caso generale per la CFA



A. Calcagni

Testing psicologico – PARTE II (B)

76/96

CFA: note conclusive

Suggerimenti per la definizione di un modello CFA

- 1 Valutare la matrice di correlazione delle variabili osservate prima di definire modelli CFA
- 2 Definire modelli semplici, da far crescere a poco a poco includendo nuove variabili osservate e/o latenti oppure “liberando” parametri
- 3 Valutare i modelli in maniera integrata, utilizzando gli indici a disposizione, interpretandoli in maniera comparata
- 4 Valutare, quando possibile, la matrice di covarianza $\hat{\Sigma}_Y$ riprodotta
- 5 Valutare un modello in ottica comparativa, rispetto ad un insieme di modelli alternativi tra loro
- 6 Definire o ampliare un modello già definito sulla base di *considerazioni teoriche* oltre che statistiche: il modello finale, infatti, è un modello di misura che deve essere interpretato ed usato
- 7 Diversi modelli possono essere adattati bene agli stessi dati: guidare l’analisi statistica del modello da considerazioni teoriche adeguate



A. Calcagni

Testing psicologico – PARTE II (B)

77/96

CFA: note conclusive

Risorse e sitografia minima per la CFA mediante la libreria di R `lavaan`

- 1 Sito web di riferimento per la libreria `lavaan`:
<http://lavaan.ugent.be/>
- 2 Tutorial sull'utilizzo della libreria `lavaan`:
<http://lavaan.ugent.be/tutorial/tutorial.pdf>
- 3 Corso su CFA (Weeks 1-4) con `lavaan`:
<http://sachaepskamp.com/SEM2017>
- 4 CFA e SEMs con `lavaan`:
https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_intro.pdf
- 5 CFA in a nutshell with `lavaan`:
https://www.yourpersonality.net/R/R_Notes_3.html
- 6 `lavaan` cheat sheet:
<https://github.com/jeromyanglim/lavaan-examples/blob/master/cheat-sheet-lavaan/cheat-sheet-lavaan.md>



A. Calcagni

Testing psicologico – PARTE II (B)

78/96

Invarianza di misurazione

Introduzione

Spesso nella ricerca applicata si è soliti confrontare strumenti di misura tra **gruppi** (es.: pazienti clinici vs. controllo, atipici nello sviluppo vs. tipici nello sviluppo) o nel **tempo** (es.: pazienti pre-trattamento vs. trattamento, pre-scolastico vs. post-scolastico).

In questo contesto ci si può chiedere se la struttura fattoriale di uno strumento rimane *invariante* tra gruppi o nel tempo. Le ragioni possono essere molteplici:

- confronto tra gruppi in analisi successive
(es.: gruppi hanno medesima struttura di intelligenza)
- analisi delle dimensioni cognitive tra gruppi
(es.: bambini tipici e atipici presentano medesima componente VCI)
- validazione di uno strumento di misura in gruppi diversi
(es.: WISC-IV potrebbe non essere adeguata in particolari popolazioni)

La valutazione dell’invarianza di misurazione (o equivalenza della misurazione) è spesso anche utilizzata negli studi cross-culturali.



A. Calcagni

Testing psicologico – PARTE II (B)

79/96

Invarianza di misurazione

Introduzione

La valutazione dell'invarianza di misurazione avviene attraverso una **procedura incrementale** dove modelli CFA di tipo annidato (*nested*) vengono valutati utilizzando gli indici di fit visti in precedenza: χ^2 , RMSEA, CFI, AIC.

La procedura incrementale valuta quattro tipologie di invarianza, ordinate dalla più lasca alla più restrittiva:

- 1 configurale
- 2 debole
- 3 forte
- 4 esatta

A ciascuna delle (i)-(iv) tipologie di invarianza corrisponde uno specifico modello CFA.

La procedura compara le tipologie a coppia: (ii) vs. (i), (iii) vs. (ii), (iv) vs. (iii) e si arresta quando il confronto fallisce rispetto ai dati a disposizione.



Invarianza di misurazione

Introduzione

Schema generale

Supponiamo, come in precedenza, di avere p *variabili osservate* su $g = 1, \dots, G$ *gruppi* distinti aventi n_g unità statistiche, con $q < p$ *variabili latenti*. Per semplicità di esposizione, fissiamo $G = 2$ e procediamo con due soli gruppi distinti.



Invarianza di misurazione

Introduzione

Schema generale

Consideriamo le $p \times p$ matrici di covarianza osservata per i due gruppi $\mathbf{S}_Y^{(1)}$ e $\mathbf{S}_Y^{(2)}$. Dati due generici e distinti livelli di invarianza, la procedura procede come segue:

- a) Si adattano due modelli CFA, \mathcal{M}_1 su $\mathbf{S}_Y^{(1)}$ e \mathcal{M}_2 su $\mathbf{S}_Y^{(2)}$, per il primo livello di invarianza ottenendo i parametri CFA stimati per entrambi i modelli adattati
- b) Si calcola la statistica complessiva $\chi_0^2 = \chi_{\mathcal{M}_1}^2 + \chi_{\mathcal{M}_2}^2$ e i gradi di libertà df_0
- c) Si adattano due modelli CFA, \mathcal{M}'_1 su $\mathbf{S}_Y^{(1)}$ e \mathcal{M}'_2 su $\mathbf{S}_Y^{(2)}$, per il secondo livello di invarianza ottenendo i parametri CFA stimati per entrambi i modelli adattati
- d) Si calcola la statistica complessiva $\chi_1^2 = \chi_{\mathcal{M}'_1}^2 + \chi_{\mathcal{M}'_2}^2$ e i gradi di libertà df_1
- e) Si valuta la statistica $(\chi_1^2 - \chi_0^2)$ mediante il test del χ^2 per modelli annidati (*nested chi-square test*)



Invarianza di misurazione

Introduzione

Schema generale

Nested chi-square test:

$$(\chi_1^2 - \chi_0^2) \stackrel{H_0}{\sim} \chi_{(df_1 - df_0)}^2$$

dove l'ipotesi nulla è definita come segue:

$$H_0 : \chi_1^2 - \chi_0^2 = 0$$

Se il valore della statistica test $(\chi_1^2 - \chi_0^2)$ è maggiore del valore critico χ_{α}^2 calcolato rispetto all' α scelto, il test rigetta l'ipotesi di equivalenza tra i due modelli.

In generale:

- Se H_0 è rigettata la procedura di valutazione dell'invarianza si arresta
- Se H_0 non è rigettata la procedura di valutazione dell'invarianza continua per livelli successivi



Invarianza di misurazione

Introduzione

Schema generale

Nota: H_0 formalizza l'ipotesi che i due modelli siano invarianti (equivalenti). Per procedere nella valutazione dell'invarianza occorre dunque che H_0 non venga rigettata. Il rigetto dell'ipotesi nulla indica che i modelli sono non-invarianti (non equivalenti).

Invarianza di misurazione

Invarianza configurale

Il modello CFA di **invarianza configurale** è il seguente[†]:

$$\mathbf{y}^{(1)} = \boldsymbol{\tau}^{(1)} + \boldsymbol{\Lambda}^{(1)}\boldsymbol{\eta}^{(1)} + \boldsymbol{\delta}^{(1)} \quad (\text{gruppo 1})$$

$$\mathbf{y}^{(2)} = \boldsymbol{\tau}^{(2)} + \boldsymbol{\Lambda}^{(2)}\boldsymbol{\eta}^{(2)} + \boldsymbol{\delta}^{(2)} \quad (\text{gruppo 2})$$

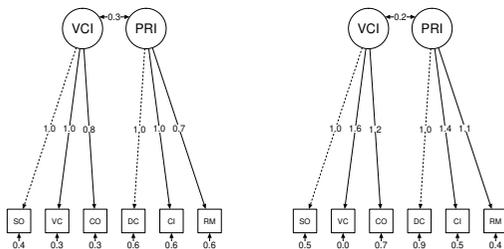
dove ciascun gruppo ha un proprio modello fattoriale sotto il vincolo che i parametri fissi e quelli liberi della matrice $\boldsymbol{\Lambda}$ siano i medesimi tra i due gruppi. Le stime $\hat{\boldsymbol{\Lambda}}^{(1)}$ e $\hat{\boldsymbol{\Lambda}}^{(2)}$ possono differire ma le strutture $\boldsymbol{\Lambda}^{(1)}$ e $\boldsymbol{\Lambda}^{(2)}$ devono essere le stesse (es.: stesso numero di fattori latenti q , stessi zeri nelle stesse posizioni).

Nota: rispetto al modello precedente, abbiamo aggiunto il termine $\boldsymbol{\tau}_{p \times 1}$ che rappresenta le intercette del modello per ciascuna variabile osservata. Queste hanno utilità quando, ad esempio, gli item (variabili osservate) hanno diversa media (es.: difficoltà diversa) o quando, come in questo caso, si conduce l'analisi di invarianza.

[†] Il pedice i è omissis per semplicità notazionale. Inoltre, le dimensioni dei vettori sono le stesse di quelle indicate per la CFA (vedi slide 11 e seguenti).

Invarianza di misurazione

Invarianza configurale



$$\hat{\boldsymbol{\Lambda}}^{(1)} = \begin{bmatrix} 1.00 & 0.00 \\ 0.99 & 0.00 \\ 0.84 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 0.97 \\ 0.00 & 0.73 \end{bmatrix} \quad \hat{\boldsymbol{\Lambda}}^{(2)} = \begin{bmatrix} 1.00 & 0.00 \\ 1.63 & 0.00 \\ 1.17 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.40 \\ 0.00 & 1.11 \end{bmatrix}$$

$$\hat{\boldsymbol{\Phi}}^{(1)} = \begin{bmatrix} 0.49 & \\ 0.30 & 0.43 \end{bmatrix} \quad \hat{\boldsymbol{\Phi}}^{(2)} = \begin{bmatrix} 0.38 & \\ 0.18 & 0.22 \end{bmatrix}$$

$$\hat{\boldsymbol{\tau}}^{(1)} = [0.26 \ 0.17 \ 0.08 \ -0.05 \ -0.12 \ 0.01]$$

$$\hat{\boldsymbol{\tau}}^{(2)} = [-0.32 \ -0.24 \ -0.13 \ 0.05 \ -0.18 \ -0.25]$$

$$\text{diag}(\hat{\boldsymbol{\Theta}}_{\boldsymbol{\delta}}^{(1)}) = [0.36 \ 0.30 \ 0.29 \ 0.64 \ 0.61 \ 0.63]$$

$$\text{diag}(\hat{\boldsymbol{\Theta}}_{\boldsymbol{\delta}}^{(2)}) = [0.46 \ 0.03 \ 0.66 \ 0.94 \ 0.52 \ 0.41]$$

χ^2	df	m	AIC
15.46	16	38	1554.74

Invarianza di misurazione

Invarianza debole

Il modello CFA di **invarianza debole** è il seguente:

$$\mathbf{y}^{(1)} = \boldsymbol{\tau}^{(1)} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(1)} + \boldsymbol{\delta}^{(1)} \quad (\text{gruppo 1})$$

$$\mathbf{y}^{(2)} = \boldsymbol{\tau}^{(2)} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(2)} + \boldsymbol{\delta}^{(2)} \quad (\text{gruppo 2})$$

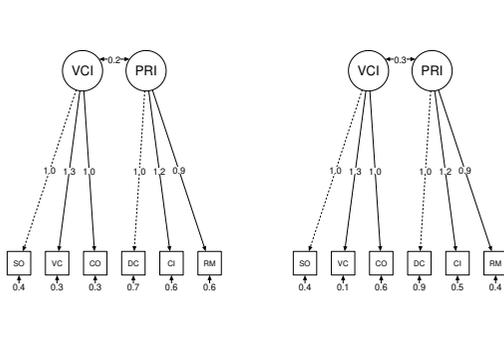
dove ciascun gruppo ha un proprio modello fattoriale sotto il vincolo:

$$\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$$

secondo cui i gruppi hanno i medesimi coefficienti fattoriali (stessi *factor loadings*). Il vincolo traduce l'ipotesi che i due gruppi condividano la stessa struttura fattoriale, ossia gli stessi costrutti.

Invarianza di misurazione

Invarianza debole



$$\hat{\Lambda}^{(1)} = \hat{\Lambda}^{(2)} = \begin{bmatrix} 1.00 & 0.00 \\ 1.31 & 0.00 \\ 1.02 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.15 \\ 0.00 & 0.88 \end{bmatrix}$$

$$\hat{\Phi}^{(1)} = \begin{bmatrix} 0.34 & \\ & 0.35 \end{bmatrix} \quad \hat{\Phi}^{(2)} = \begin{bmatrix} 0.52 & \\ & 0.31 \end{bmatrix}$$

$$\hat{\tau}^{(1)} = [0.26 \ 0.17 \ 0.08 \ -0.05 \ -0.12 \ 0.01]$$

$$\hat{\tau}^{(2)} = [-0.32 \ -0.24 \ -0.13 \ 0.05 \ -0.18 \ -0.25]$$

$$\text{diag}(\hat{\Theta}_{\delta}^{(1)}) = [0.42 \ 0.25 \ 0.29 \ 0.69 \ 0.57 \ 0.61]$$

$$\text{diag}(\hat{\Theta}_{\delta}^{(2)}) = [0.42 \ 0.13 \ 0.63 \ 0.91 \ 0.54 \ 0.42]$$

	Df	AIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
configurale	16	1554.74	15.46			
debole	20	1550.20	18.91	3.46	4	0.48

Conclusion: Il test del chi-quadrato non consente di rigettare l'ipotesi che i due modelli siano equivalenti: i due gruppi sono invarianti rispetto al criterio debole e condividono la medesima struttura fattoriale Λ .

Invarianza di misurazione

Invarianza forte

Il modello CFA di **invarianza forte** è il seguente:

$$\mathbf{y}^{(1)} = \boldsymbol{\tau} + \Lambda \boldsymbol{\eta}^{(1)} + \boldsymbol{\delta}^{(1)} \quad (\text{gruppo 1})$$

$$\mathbf{y}^{(2)} = \boldsymbol{\tau} + \Lambda \boldsymbol{\eta}^{(2)} + \boldsymbol{\delta}^{(2)} \quad (\text{gruppo 2})$$

dove ciascun gruppo ha un proprio modello fattoriale sotto i vincoli:

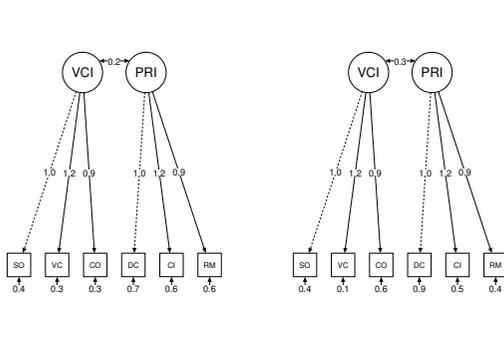
$$\Lambda^{(1)} = \Lambda^{(2)}$$

$$\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$$

secondo cui i gruppi hanno i medesimi coefficienti fattoriali (stessi *factor loadings*) e le stesse intercette. Il vincolo traduce l'ipotesi che i due gruppi condividano la stessa struttura fattoriale unitamente al fatto che gli item hanno lo stesso significato in entrambi i gruppi. In altri termini, un modello con invarianza forte assicura l'assenza di bias indotto dalle differenze tra gli item utilizzati nei due gruppi.

Invarianza di misurazione

Invarianza forte



$$\hat{\Lambda}^{(1)} = \hat{\Lambda}^{(2)} = \begin{bmatrix} 1.00 & 0.00 \\ 1.22 & 0.00 \\ 0.98 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.18 \\ 0.00 & 0.92 \end{bmatrix}$$

$$\hat{\Phi}^{(1)} = \begin{bmatrix} 0.38 & \\ & 0.33 \end{bmatrix} \quad \hat{\Phi}^{(2)} = \begin{bmatrix} 0.57 & \\ & 0.29 \end{bmatrix}$$

$$\hat{\tau}^{(1)} = \hat{\tau}^{(2)} = [0.16 \ 0.20 \ 0.12 \ 0.03 \ -0.10 \ -0.10]$$

$$\text{diag}(\hat{\Theta}_{\delta}^{(1)}) = [0.42 \ 0.26 \ 0.30 \ 0.71 \ 0.57 \ 0.62]$$

$$\text{diag}(\hat{\Theta}_{\delta}^{(2)}) = [0.43 \ 0.15 \ 0.64 \ 0.93 \ 0.54 \ 0.42]$$

	Df	AIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
debole	20	1550.20	18.91			
forte	24	1549.00	25.72	6.80	4	0.15

Conclusion: Il test del chi-quadrato non consente di rigettare l'ipotesi che i due modelli siano equivalenti: i due gruppi sono invarianti rispetto al criterio forte, condividono la medesima struttura fattoriale Λ , tendono a rispondere allo stesso modo agli item (questi hanno lo stesso significato in entrambi i gruppi) e non vi è bias indotto dalle differenze degli item $\boldsymbol{\tau}$ nei due gruppi.

Invarianza di misurazione

Invarianza esatta

Il modello CFA di **invarianza esatta** è il seguente:

$$\mathbf{Y}^{(1)} = \boldsymbol{\tau} + \Lambda \boldsymbol{\eta}^{(1)} + \boldsymbol{\delta} \quad (\text{gruppo 1})$$

$$\mathbf{Y}^{(2)} = \boldsymbol{\tau} + \Lambda \boldsymbol{\eta}^{(2)} + \boldsymbol{\delta} \quad (\text{gruppo 2})$$

dove ciascun gruppo ha un proprio modello fattoriale sotto i vincoli:

$$\Lambda^{(1)} = \Lambda^{(2)}$$

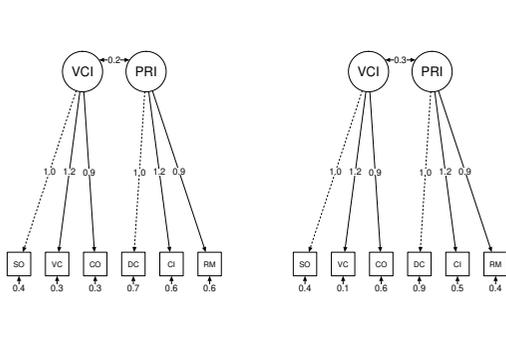
$$\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$$

$$\boldsymbol{\delta}^{(1)} = \boldsymbol{\delta}^{(2)}$$

secondo cui i gruppi hanno i medesimi coefficienti fattoriali (stessi *factor loadings*), stesse intercette e stessi errori. Il vincolo traduce l'ipotesi che i due gruppi condividano la stessa struttura fattoriale, gli item non producono bias nei gruppi e i due gruppi possono essere confrontati in termini di differenze tra costrutti latenti.

Invarianza di misurazione

Invarianza esatta



$$\hat{\Lambda}^{(1)} = \hat{\Lambda}^{(2)} = \begin{bmatrix} 1.00 & 0.00 \\ 1.21 & 0.00 \\ 0.93 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.27 \\ 0.00 & 0.96 \end{bmatrix}$$

$$\Phi^{(1)} = \begin{bmatrix} 0.39 & \\ & 0.20 & 0.31 \end{bmatrix} \quad \Phi^{(2)} = \begin{bmatrix} 0.56 & \\ & 0.26 & 0.26 \end{bmatrix}$$

$$\tau^{(1)} = \tau^{(2)} = [0.16 \ 0.19 \ 0.15 \ 0.04 \ -0.10 \ -0.08]$$

$$\text{diag}(\hat{\Theta}_\delta^{(1)}) = \text{diag}(\hat{\Theta}_\delta^{(2)}) = \begin{bmatrix} 0.42 \\ 0.20 \\ 0.47 \\ 0.84 \\ 0.53 \\ 0.53 \end{bmatrix}$$

	Df	AIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
forte	24	1549.00	25.72			
esatta	30	1544.60	33.32	7.60	6	0.27

Conclusion: Il test del chi-quadrato non consente di rigettare l'ipotesi che i due modelli siano equivalenti: i due gruppi sono invarianti rispetto al criterio forte, condividono la medesima struttura fattoriale Λ , non vi è bias indotto dalle differenze degli item τ nei due gruppi ed inoltre i residui del modello δ sono gli stessi. I due gruppi possono essere confrontati tra loro in termini di punteggi fattoriali $\eta^{(1)}$ e $\eta^{(2)}$.

Invarianza di misurazione: note conclusive

Invarianza parziale

Nelle slide precedenti è stata presentata la procedura generale per valutare l'adeguatezza dell'**invarianza completa**.

Esiste tuttavia la possibilità di valutare la c.d. **invarianza parziale** che permette di stabilire, una volta arrestata la procedura ad un certo grado di invarianza (es.: *debole*), se G gruppi condividono la stessa struttura fattoriale *a meno* di alcuni parametri. In pratica, l'invarianza parziale si valuta vincolando pochi parametri per volta finché la procedura, per un dato livello, non si arresta nuovamente.

Vedremo nei laboratori come valutare l'invarianza parziale una volta che quella completa non sia verificata.

Invarianza di misurazione: note conclusive

Modelli CFA di invarianza in sintesi

per $g = 1, \dots, G$:

- 1 $\mathbf{y}^{(g)} = \tau^{(g)} + \Lambda^{(g)}\eta^{(g)} + \delta^{(g)}$ invarianza configurale
- 2 $\mathbf{y}^{(g)} = \tau^{(g)} + \Lambda\eta^{(g)} + \delta^{(g)}$ invarianza debole
- 3 $\mathbf{y}^{(g)} = \tau + \Lambda\eta^{(g)} + \delta^{(g)}$ invarianza forte
- 4 $\mathbf{y}^{(g)} = \tau + \Lambda\eta^{(g)} + \delta$ invarianza esatta

Invarianza di misurazione: note conclusive

Modelli CFA di invarianza in sintesi



Breve tutorial pratico sull'invarianza (disponibile negli Approfondimenti del corso):

- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492.

Invarianza di misurazione: note conclusive

Risorse e sitografia minima per l'invarianza mediante la libreria di R `lavaan`

1 Tutorial sull'invarianza con `lavaan`:

<https://tinyurl.com/yy3rutpz>

2 Breve talk sull'invarianza con `lavaan`:

<https://users.ugent.be/~yrosseel/lavaan/multiplegroup6Dec2012.pdf>

Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

NOTAZIONE UTILE E RICHIAMI

Antonio Calcagni

DPSS, Università di Padova

A.A. 2021/2022

Elementi di algebra lineare

Matrice: Organizzazione di numeri (es.: naturali, reali) in forma tabellare disposti in I righe e J colonne.

$$\mathbf{X}_{I \times J} = \begin{bmatrix} X_{11} & \dots & X_{1j} & \dots & X_{1J} \\ \vdots & & \vdots & & \vdots \\ X_{i1} & \dots & X_{ij} & \dots & X_{iJ} \\ \vdots & & \vdots & & \vdots \\ X_{I1} & \dots & X_{Ij} & \dots & X_{IJ} \end{bmatrix}$$

Gli elementi di \mathbf{X} possono essere individuati da una coppia di indici (i, j) che prendono valore in un sottoinsieme di \mathbb{N} :

$$i \in \{1, \dots, I\} \subset \mathbb{N}$$

$$j \in \{1, \dots, J\} \subset \mathbb{N}$$

Elementi di algebra lineare

Matrice: Organizzazione di numeri (es.: naturali, reali) in forma tabellare disposti in I righe e J colonne.

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1J} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iJ} \\ \vdots & & \vdots & & \vdots \\ x_{I1} & \dots & x_{Ij} & \dots & x_{IJ} \end{bmatrix}$$

L'elemento nella i -esima riga e j -colonna x_{ij} si dice **scalare**.



Elementi di algebra lineare

Matrice: Organizzazione di numeri (es.: naturali, reali) in forma tabellare disposti in I righe e J colonne.

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1J} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iJ} \\ \vdots & & \vdots & & \vdots \\ x_{I1} & \dots & x_{Ij} & \dots & x_{IJ} \end{bmatrix}$$

Il sottoinsieme di I elementi individuati fissando la j -esima colonna si chiama **vettore colonna**:

$$\mathbf{x}_{I \times 1} = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{Ij} \end{bmatrix}$$



Elementi di algebra lineare

Matrice: Organizzazione di numeri (es.: naturali, reali) in forma tabellare disposti in I righe e J colonne.

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1J} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iJ} \\ \vdots & & \vdots & & \vdots \\ x_{I1} & \dots & x_{Ij} & \dots & x_{IJ} \end{bmatrix}$$

Il sottoinsieme di J elementi individuati fissando la i -esima riga si chiama **vettore riga**:

$$\mathbf{x}_{1 \times J} = [x_{i1}, \dots, x_{ij}, \dots, x_{iJ}]$$



Elementi di algebra lineare

Due matrici $\mathbf{X}_{I \times J}$ e $\mathbf{Y}_{N \times M}$ si dicono **uguali** se $I = N$, $J = M$, $x_{ij} = y_{ij}$.

Una matrice $\mathbf{X}_{I \times J}$ si dice:

- **nulla** se $x_{ij} = 0$ per ogni i, j
- **unitaria** se $x_{ij} = 1$ per ogni i, j
- **quadrata** se $I = J$
- **diagonale** se $x_{ij} = 0$ per ogni $i \neq j$ e $x_{ij} \neq 0$ per ogni $i = j$

$$\mathbf{X}_{I \times J} = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$



Elementi di algebra lineare

La **diagonale principale** di $\mathbf{X}_{I \times J}$ il sottoinsieme di elementi così composti:

$$\text{diag}(\mathbf{X}) = \{x_{ij} : i = j, i = 1, \dots, l; j = 1, \dots, J\}$$

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & x_{jj} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & x_{JJ} \end{bmatrix}$$



Elementi di algebra lineare

La **traccia** di $\mathbf{X}_{I \times I}$ è definita come segue:

$$\text{Tr}(\mathbf{X}) = \sum_{i=1}^I \text{diag}(\mathbf{X})_i$$

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1J} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iJ} \\ \vdots & & \vdots & & \vdots \\ x_{J1} & \dots & x_{Jj} & \dots & x_{JJ} \end{bmatrix}$$

$$\text{Tr}(\mathbf{X}) = x_{11} + \dots + x_{jj} + \dots + x_{JJ}$$



Elementi di algebra lineare

Una matrice $\mathbf{X}_{I \times J}$ si dice **identità** e si indica con \mathbf{I} se $x_{ij} = 0$ per ogni $i \neq j$ e $x_{ij} = 1$ per ogni $i = j$.

$$\mathbf{I}_{I \times J} = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$



Elementi di algebra lineare

Una matrice $\mathbf{X}_{I \times J}$ si dice **triangolare inferiore** se $x_{ij} = 0$ per ogni $i > j$ e $x_{ij} \neq 0$ per ogni $i \leq j$:

$$\text{tril}(\mathbf{X}) = \{x_{ij} : i \leq j, i = 1, \dots, l; j = 1, \dots, J\}$$

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ x_{J1} & \dots & x_{Jj} & \dots & x_{JJ} \end{bmatrix}$$



Elementi di algebra lineare

Una matrice $\mathbf{X}_{I \times J}$ si dice **triangolare superiore** se $x_{ij} = 0$ per ogni $i < j$ e $x_{ij} \neq 0$ per ogni $i \geq j$:

$$\text{triu}(\mathbf{X}) = \{x_{ij} : i \geq j, i = 1, \dots, I; j = 1, \dots, J\}$$

$$\mathbf{X}_{I \times J} = \begin{bmatrix} x_{11} & \dots & x_{1i} & \dots & x_{1J} \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & x_{ij} & \dots & x_{iJ} \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & x_{IJ} \end{bmatrix}$$



Elementi di algebra lineare

Una matrice $\mathbf{X}_{I \times J}$ si dice **simmetrica** se :

$$\text{tril}(\mathbf{X}) = \text{triu}(\mathbf{X})$$

$$\mathbf{X}_{3 \times 3} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

$$\text{tril}(\mathbf{X}) = [1, 2, 3]$$

$$\text{triu}(\mathbf{X}) = [1, 2, 3]$$



Elementi di algebra lineare

Data una matrice $\mathbf{X}_{I \times J}$ si dice **trasposta** e la si indica con \mathbf{X}^T la matrice $\mathbf{Y}_{J \times I}$ le cui righe sono le colonne di \mathbf{X} e le colonne le righe di \mathbf{X} :

$$\mathbf{X}_{3 \times 3} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \mathbf{X}_{3 \times 3}^T = \mathbf{Y}_{3 \times 3} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$



Elementi di algebra lineare

Una matrice $\mathbf{X}_{I \times J}$ si dice **reale** se $\mathbf{X}_{I \times J} \in \mathbb{R}^{I \times J}$, ossia se i suoi elementi sono numeri reali.

Allo stesso modo:

- un vettore colonna reale è del tipo $\mathbf{x}_{J \times 1} \in \mathbb{R}^J$
- un vettore riga reale è del tipo $\mathbf{x}_{1 \times I} \in \mathbb{R}^I$



Elementi di algebra lineare

La **somma** tra due matrici reali $\mathbf{X}_{I \times J}$ e $\mathbf{Y}_{I \times J}$ è definita come segue:

$$\mathbf{Z}_{I \times J} = \mathbf{X}_{I \times J} + \mathbf{Y}_{I \times J}$$

$$\underbrace{\begin{bmatrix} 2 & 8 & 14 \\ 3 & 10 & 11 \\ 3 & 7 & 15 \end{bmatrix}}_{\mathbf{Z}} = \underbrace{\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}}_{\mathbf{X}} + \underbrace{\begin{bmatrix} 1 & 4 & 7 \\ 1 & 5 & 3 \\ 0 & 1 & 6 \end{bmatrix}}_{\mathbf{Y}}$$

Allo stesso modo la differenza tra matrici reali.



Elementi di algebra lineare

Il **prodotto** tra una matrice reale $\mathbf{X}_{I \times J}$ e uno scalare $b \in \mathbb{R}$ è definito come segue:

$$\mathbf{Z}_{I \times J} = \mathbf{X}_{I \times J} \cdot b$$

$$\underbrace{\begin{bmatrix} 1.5 & 4.5 & 7.5 \\ 2.5 & 5.5 & 8.5 \\ 3.5 & 6.5 & 9.5 \end{bmatrix}}_{\mathbf{Z}} = \underbrace{\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}}_{\mathbf{X}} \cdot \underbrace{1.5}_b$$



Elementi di algebra lineare

Il **prodotto** elemento per elemento tra due matrici reali $\mathbf{X}_{I \times J}$ e $\mathbf{Y}_{I \times J}$ è definito come segue:

$$\mathbf{Z}_{I \times J} = \mathbf{X}_{I \times J} \circ \mathbf{Y}_{I \times J}$$

$$\underbrace{\begin{bmatrix} 1 & 16 & 49 \\ 2 & 25 & 24 \\ 0 & 6 & 54 \end{bmatrix}}_{\mathbf{Z}} = \underbrace{\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}}_{\mathbf{X}} \circ \underbrace{\begin{bmatrix} 1 & 4 & 7 \\ 1 & 5 & 3 \\ 0 & 1 & 6 \end{bmatrix}}_{\mathbf{Y}}$$

Tale operazione si dice *prodotto di Hadamard*.



Elementi di algebra lineare

Il **prodotto** riga-colonna tra due matrici reali $\mathbf{X}_{I \times J}$ e $\mathbf{Y}_{I \times J}$ è definito come segue:

$$\mathbf{Z}_{I \times I} = \mathbf{X}_{I \times J} \mathbf{Y}_{I \times J}^T$$

$$\underbrace{\begin{bmatrix} 3 & 11 \\ 3 & 0 \\ 3 & 6 \end{bmatrix}}_{\mathbf{Z}_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{\mathbf{X}_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{\mathbf{Y}_{3 \times 2}^T}$$

Nota: questa operazione esiste solo se il numero delle colonne della matrice che pre-moltiplica (\mathbf{X}) è uguale al numero delle righe che post-moltiplica (\mathbf{Y}).



Elementi di algebra lineare

$$\underbrace{\begin{bmatrix} (1 \cdot 1) + (2 \cdot 0) + (1 \cdot 2) & 11 \\ 3 & 0 \\ 3 & 6 \end{bmatrix}}_{Z_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{X_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{Y_{3 \times 2}}$$



Elementi di algebra lineare

$$\underbrace{\begin{bmatrix} 3 & 11 \\ (3 \cdot 1) + (0 \cdot 0) + (0 \cdot 2) & 0 \\ 3 & 6 \end{bmatrix}}_{Z_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{X_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{Y_{3 \times 2}}$$



Elementi di algebra lineare

$$\underbrace{\begin{bmatrix} 3 & 11 \\ 3 & 0 \\ (1 \cdot 1) + (1 \cdot 0) + (1 \cdot 2) & 6 \end{bmatrix}}_{Z_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{X_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{Y_{3 \times 2}}$$



Elementi di algebra lineare

$$\underbrace{\begin{bmatrix} 3 & 11 \\ 3 & 0 \\ 3 & 6 \end{bmatrix}}_{Z_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{X_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{Y_{3 \times 2}}$$



Elementi di algebra lineare

$$\underbrace{\begin{bmatrix} 3 & & 11 \\ 3 & (3 \cdot 0) + (0 \cdot 5) + (0 \cdot 1) & \\ 3 & & 6 \end{bmatrix}}_{z_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{x_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{y_{3 \times 2}}$$



Elementi di algebra lineare

$$\underbrace{\begin{bmatrix} 3 & & 11 \\ 3 & 0 & \\ 3 & (1 \cdot 0) + (1 \cdot 5) + (1 \cdot 1) & \end{bmatrix}}_{z_{3 \times 2}} = \underbrace{\begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{x_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{bmatrix}}_{y_{3 \times 2}}$$



Elementi di algebra lineare

Allo stesso modo avviene il prodotto tra una matrice ed un vettore, ad esempio:

$$\underbrace{\begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}}_{z_{3 \times 1}} = \underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ 1 & 1 \end{bmatrix}}_{x_{3 \times 2}} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{y_{2 \times 1}}$$



Elementi di algebra lineare

Allo stesso modo avviene il prodotto tra una matrice ed un vettore, ad esempio:

$$\underbrace{[5]}_{z_{1 \times 1}} = \underbrace{[1 \quad 1 \quad 1]}_{y_{1 \times 3}} \underbrace{\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}}_{x_{3 \times 1}}$$

Nota: in questo caso il prodotto equivale a $z = \sum_{i=1}^3 x_i$



Elementi di algebra lineare

Allo stesso modo avviene il prodotto tra una matrice ed un vettore, ad esempio:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \\ 1 & 1 & 1 \end{bmatrix}}_{\mathbf{Z}_{3 \times 3}} = \underbrace{\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}}_{\mathbf{x}_{3 \times 1}} \underbrace{[1 \quad 1 \quad 1]}_{\mathbf{y}_{1 \times 3}}$$



Elementi di algebra lineare

Allo stesso modo avviene il prodotto tra una matrice ed un vettore, ad esempio:

$$\underbrace{[14]}_{\mathbf{z}_{1 \times 1}} = \underbrace{[2 \quad 3 \quad 1]}_{\mathbf{x}_{1 \times 3}} \underbrace{\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}}_{\mathbf{x}_{3 \times 1}}$$

Nota: in questo caso il prodotto equivale a $z = \sum_{i=1}^3 x_i^2$



Elementi di algebra lineare

Due casi particolari:

$$\underbrace{\begin{bmatrix} 1 & 3 & 9 \\ 3 & 0 & 2 \\ 1 & 7 & 1 \end{bmatrix}}_{\mathbf{Z}_{3 \times 3}} = \underbrace{\begin{bmatrix} 1 & 3 & 9 \\ 3 & 0 & 2 \\ 1 & 7 & 1 \end{bmatrix}}_{\mathbf{X}_{3 \times 3}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{I}_{3 \times 3}}$$



Elementi di algebra lineare

Due casi particolari:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Z}_{3 \times 3}} = \underbrace{\begin{bmatrix} 1 & 3 & 9 \\ 3 & 0 & 2 \\ 1 & 7 & 1 \end{bmatrix}}_{\mathbf{X}_{3 \times 3}} \circ \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{I}_{3 \times 3}}$$

Nota: in questo caso il prodotto equivale a $\mathbf{Z} = \text{diag}(\mathbf{X})$



Elementi di algebra lineare

Proprietà della somma

associatività	$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
commutatività	$\mathbf{A} + \mathbf{C} = \mathbf{C} + \mathbf{A}$
elemento neutro	$\mathbf{A} + \mathbf{0} = \mathbf{A}$
distributività	$\beta(\mathbf{A} + \mathbf{B}) = \mathbf{A}\beta + \mathbf{B}\beta$
	$(\alpha + \beta)\mathbf{A} = \mathbf{A}\alpha + \mathbf{A}\beta$



Elementi di algebra lineare

Proprietà del prodotto (riga-colonna)

**le matrici siano tali che tutte le moltiplicazioni scritte siano possibili*

associatività	$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$
distributività	$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{AC} + \mathbf{BC}$
	$\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\mathbf{B}\alpha)$
	$\mathbf{BI} = \mathbf{B}$
	$\mathbf{B0} = \mathbf{0}$



Elementi di algebra lineare

Proprietà delle matrici trasposte

$$\begin{aligned}(\mathbf{A}^T)^T &= \mathbf{A} \\ (\mathbf{A} \pm \mathbf{B})^T &= \mathbf{A}^T \pm \mathbf{B}^T \\ (\mathbf{A}\beta)^T &= \beta\mathbf{A}^T \\ (\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T\end{aligned}$$



Elementi di algebra lineare

Proprietà della traccia

$$\begin{aligned}\text{Tr}(\mathbf{A} + \mathbf{B}) &= \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B}) \\ \text{Tr}(\mathbf{A}\beta) &= \text{Tr}(\mathbf{A})\beta \\ \text{Tr}(\mathbf{AB}) &= \text{Tr}(\mathbf{BA})\end{aligned}$$



Elementi di algebra lineare

Determinante

Il determinante $\det(\mathbf{X})$ (oppure $|\mathbf{X}|$) di una matrice quadrata \mathbf{X} è uno scalare che gioca un ruolo importante in molti utilizzi dell'algebra lineare (es.: trasformazioni lineari, soluzione di sistemi di equazioni lineari, calcolo di polinomi caratteristici).

Diversi sono i metodi per il suo calcolo (es.: metodo di Laplace, algoritmo di Gauss). Nel caso di matrici quadrate di piccole dimensioni può essere utilizzata la definizione costruttiva.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \quad \det(\mathbf{X}) = (x_{11} \cdot x_{22}) - (x_{21} \cdot x_{12})$$



Elementi di algebra lineare

Determinante

Il determinante $\det(\mathbf{X})$ (oppure $|\mathbf{X}|$) di una matrice quadrata \mathbf{X} è uno scalare che gioca un ruolo importante in molti utilizzi dell'algebra lineare (es.: trasformazioni lineari, soluzione di sistemi di equazioni lineari, calcolo di polinomi caratteristici).

Diversi sono i metodi per il suo calcolo (es.: metodo di Laplace, algoritmo di Gauss). Nel caso di matrici quadrate di piccole dimensioni può essere utilizzata la definizione costruttiva.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$\det(\mathbf{X}) = (x_{11} \cdot x_{22} \cdot x_{33}) + (x_{13} \cdot x_{21} \cdot x_{32}) + (x_{12} \cdot x_{23} \cdot x_{31}) - (x_{13} \cdot x_{22} \cdot x_{31}) - (x_{11} \cdot x_{23} \cdot x_{32}) - (x_{12} \cdot x_{21} \cdot x_{33})$$



Elementi di algebra lineare

Alcune proprietà del determinante

Se \mathbf{X} ha una riga/colonna nulla o se ha due righe/colonne uguali allora $\det(\mathbf{X}) = 0$

$$\det(\mathbf{X}^T) = \det(\mathbf{X})$$

$\det(\mathbf{XY}) = \det(\mathbf{X}) \det(\mathbf{Y})$ se \mathbf{Y} ha la stessa dimensione di \mathbf{X}

Nota: una matrice quadrata \mathbf{X} si dice non singolare se $\det(\mathbf{X}) \neq 0$.



Elementi di algebra lineare

Inversione di una matrice quadrata

Una matrice quadrata \mathbf{X} ammette **inversa** e la si indica con \mathbf{X}^{-1} se

$$\mathbf{X} \mathbf{X}^{-1} = \mathbf{I}$$

Non sempre esiste l'inversa di una matrice. In generale, se $\det(\mathbf{X}) = 0$ l'operazione di inversione non può essere fatta (una matrice singolare non può essere invertita).

Diversi sono i metodi per il calcolo della matrice inversa. Nel caso semplice di una matrice 2×2 l'inversione è fatta nel modo seguente:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \quad \mathbf{X}^{-1} = \frac{1}{\det(\mathbf{X})} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix}$$



Proprietà della matrice inversa

$$\begin{aligned}
 (\mathbf{XY})^{-1} &= \mathbf{Y}^{-1}\mathbf{X}^{-1} \\
 (\mathbf{X} + \mathbf{Y})^{-1} &= \mathbf{X}^{-1} + \mathbf{Y}^{-1} \\
 (\mathbf{X}^T)^{-1} &= (\mathbf{X}^{-1})^T \\
 \det(\mathbf{X}^{-1}) &= 1/\det(\mathbf{X})
 \end{aligned}$$

Rango

In generale, data una matrice \mathbf{X} è possibile estrarre da essa dei sottoinsiemi eliminando alcune righe o colonne. Tali sottomatrici si chiamano **minori** di \mathbf{X} .

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}$$

Il **rango** di una matrice $\mathbf{X}_{I \times J}$ è uno scalare $r(\mathbf{X})$ che è dato dal massimo ordine dei minori di \mathbf{X} aventi determinante non nullo. In generale: $0 \leq r(\mathbf{X}) \leq \min(I, J)$.

L'algoritmo di Gauss può essere utilizzato per il calcolo del rango di una matrice. Nei casi semplici di matrici quadrate di dimensioni ridotte, può essere usato il *criterio dei minori*.

Rango

Se $\det(\mathbf{X}_{I \times I}) \neq 0$ allora $r(\mathbf{X}) = I$ (matrici non singolari).

Al contrario, nel caso di matrici singolari, ossia $\det(\mathbf{X}_{I \times I}) = 0$, avremo che $r(\mathbf{X}) \leq I$. In questo caso (i) si andrà alla ricerca delle sottomatrici di ordine inferiore a I , (ii) si calcoleranno i determinanti delle sottomatrici e (iii) non appena una di queste sottomatrici avrà determinante nullo allora si arresterà la procedura di ricerca: il rango sarà pari alla dimensione di tale sottomatrice.

Nota:

Se $\det(\mathbf{X}_{n \times n}) \neq 0$ allora la matrice \mathbf{X} è *non singolare, invertibile* e ha *rango massimo* pari a n .

operazione	comando R
$\mathbf{X} + \mathbf{Y}$	<code>X+Y</code>
$\mathbf{X} \circ \mathbf{Y}$	<code>X*Y</code>
\mathbf{XY}	<code>X%*%Y</code>
\mathbf{Xy}	<code>X%*%y</code>
\mathbf{Xb}	<code>X%*%b</code>
$\text{diag}(\mathbf{X})$	<code>diag(X)</code>
$\text{tril}(\mathbf{X})$	<code>X[lower.tri(X)]</code>
$\text{triu}(\mathbf{X})$	<code>X[upper.tri(X)]</code>
$\text{Tr}(\mathbf{X})$	<code>tr(X)</code>
\mathbf{X}^T	<code>t(X)</code>
\mathbf{X}^{-1}	<code>solve(X)</code>
$\det(\mathbf{X})$	<code>det(X)</code>
$r(\mathbf{X})$	<code>qr(X)\$rank</code>

Variabili aleatorie

$$X \sim F_X(x; \theta)$$

La distribuzione F_X può essere discreta (es.: Binomiale, Poisson) o continua (es.: Gamma, Normale, Chi-quadrato) ed è un modello probabilistico il cui comportamento di posizione, scala o forma è governato dal parametro reale θ (scalare, vettore o matrice).

I valori medi di X sintetizzano alcune caratteristiche della variabile aleatoria (es.: posizione, dispersione).



Variabili aleatorie

$$X \sim F_X(x; \theta)$$

Valore atteso. Denotato mediante $\mathbb{E}[X]$ quantifica il valore medio dell'esperimento aleatorio associato a X :

$$\mathbb{E}[X] = \sum_{x \in \text{sup}(X)} x \cdot f_X(x; \theta) \quad (\text{caso discreto})$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x; \theta) dx \quad (\text{caso continuo})$$



Variabili aleatorie

$$X \sim F_X(x; \theta)$$

Varianza. Denotata mediante $\text{Var}[X]$ quantifica la dispersione dell'esperimento aleatorio associato a X :

$$\text{Var}[X] = \sum_{x \in \text{sup}(X)} (x - \mathbb{E}[X])^2 f_X(x; \theta) \quad (\text{caso discreto})$$

$$\text{Var}[X] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x; \theta) dx \quad (\text{caso continuo})$$



Variabili aleatorie

Principali proprietà del valore atteso

$$\mathbb{E}[\alpha] = \alpha$$

$$\text{Var}[\alpha] = 0$$

$$\mathbb{E}[X_h + X_k] = \mathbb{E}[X_h] + \mathbb{E}[X_k]$$

$$\text{Var}[X_h + X_k] = \text{Var}[X_h] + \text{Var}[X_k] + 2\text{Cov}[X_h, X_k]$$

$$\mathbb{E}[\beta X_h] = \beta \mathbb{E}[X_h]$$

$$\text{Var}[\beta X_h] = \beta^2 \text{Var}[X_h]$$

$$\mathbb{E}[\alpha + \beta X_h] = \alpha + \beta \mathbb{E}[X_h]$$

$$\text{Var}[\alpha + \beta X_h] = \beta^2 \text{Var}[X_h]$$



Variabili aleatorie

Spesso gli esperimenti casuali sono descritti da più di una variabile aleatoria allo stesso tempo (**vettori casuali**):

$$X = (X_1, \dots, X_J) \sim F_{X_1, \dots, X_J}(x_1, \dots, x_J; \theta)$$

In questi casi, la variabile aleatoria e la associata densità è detta multidimensionale.

La **distribuzione marginale** della j -esima v.a. è ottenuta per integrazione (caso continuo) o somma (caso discreto). Ad esempio, per $J = 2$:

$$f_{X_1}(x_1; \theta) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2; \theta) dx_2$$

La **distribuzione condizionale**, invece, è ottenuta condizionando alcuni esiti ad altri:

$$f_{X_1|X_2}(x_1; \theta) = \frac{f_{X_1, X_2}(x_1, x_2; \theta)}{f_{X_2}(x_2; \theta)}$$



Variabili aleatorie

$$X = (X_1, \dots, X_J) \sim F_{X_1, \dots, X_J}(x_1, \dots, x_J; \theta)$$

Quando le due variabili aleatorie sono **indipendenti** $X_1 \perp\!\!\!\perp X_2$ allora:

$$f_{X_1|X_2}(x_1; \theta) = f_{X_1}(x_1; \theta)$$

oppure

$$f_{X_1, X_2}(x_1, x_2; \theta) = f_{X_1}(x_1; \theta) \cdot f_{X_2}(x_2; \theta)$$



Variabili aleatorie

$$X = (X_1, \dots, X_J) \sim F_{X_1, \dots, X_J}(x_1, \dots, x_J; \theta)$$

Il **valori attesi condizionati** sono ottenuti per condizionamento, ad esempio:

$$\mathbb{E}[X_1|X_2 = x_2] = \int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1; \theta) dx_1$$

$$\text{Var}[X_1|X_2 = x_2] = \mathbb{E}\left[(X_1 - \mathbb{E}[X_1|X_2])^2 \middle| X_2 = x_2\right]$$



Variabili aleatorie

$$X = (X_1, \dots, X_J) \sim F_{X_1, \dots, X_J}(x_1, \dots, x_J; \theta)$$

Una caratteristica importante che quantifica il grado di co-variazione (o associazione) tra coppie di variabili aleatorie è la **covarianza**:

$$\begin{aligned} \text{Cov}[X_h, X_k] &= \mathbb{E}[(X_h - \mu_{X_h})(X_k - \mu_{X_k})] \\ &= \mathbb{E}[X_h X_k] - \mu_{X_h} \mu_{X_k} \end{aligned}$$

dove in generale $\mu_X = \mathbb{E}[X]$. Essendo una misura di associazione lineare, abbiamo i seguenti casi:

- $\text{Cov}[X_h, X_k] > 0$: X_h e X_k sono positivamente associati
- $\text{Cov}[X_h, X_k] < 0$: X_h e X_k sono negativamente associati
- $\text{Cov}[X_h, X_k] = 0$: X_h e X_k non sono *linearmente* associati



Variabili aleatorie

$$X = (X_1, \dots, X_J) \sim F_{X_1, \dots, X_J}(x_1, \dots, x_J; \theta)$$

Le variabili aleatorie X_1, \dots, X_J sono dette **indipendenti e identicamente distribuite** (iid) se:

$$f_{X_1, \dots, X_J}(x_1, \dots, x_J; \theta) = f_{X_1}(x_1; \theta) \cdots f_{X_J}(x_J; \theta)$$

$$f_{X_1}(x_1; \theta) = f_{X_2}(x_2; \theta) = \dots = f_{X_J}(x_J; \theta)$$

Variabili aleatorie iid sono alla base del **campionamento casuale semplice** e giocano un ruolo chiave nella costruzione dei teoremi limite alla base dell'inferenza statistica (es.: legge dei grandi numeri).



Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

Consideriamo un vettore di p variabili aleatorie $X = (X_1, \dots, X_j, \dots, X_p)$. Si dice che X segue in legge il modello normale p -variato se

$$X \sim \mathcal{N}_p(\mu, \Sigma)$$

con $\mu \in \mathbb{R}^p$ vettore $p \times 1$ delle medie/posizioni del modello e Σ matrice $p \times p$ delle varianze-covarianze del modello:

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_p \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11}^2 & \dots & \sigma_{1j} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{j1} & \dots & \sigma_{jj}^2 & \dots & \sigma_{jp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pj} & \dots & \sigma_{pp}^2 \end{bmatrix}$$



Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

La matrice di varianze-covarianze

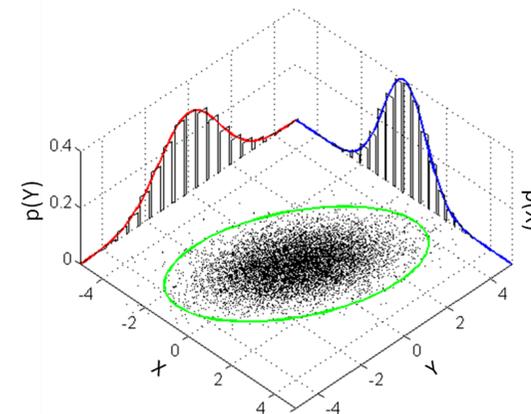
$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \dots & \sigma_{1j} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{j1} & \dots & \sigma_{jj}^2 & \dots & \sigma_{jp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pj} & \dots & \sigma_{pp}^2 \end{bmatrix}$$

- è simmetrica
- $\sigma_{11}^2, \dots, \sigma_{pp}^2$ sono le varianze
- se le p variabili sono tra loro indipendenti $X_h \perp\!\!\!\perp X_k$ ($h \neq k$), Σ diventa matrice diagonale



Modello normale multivariato

Caso bivarato



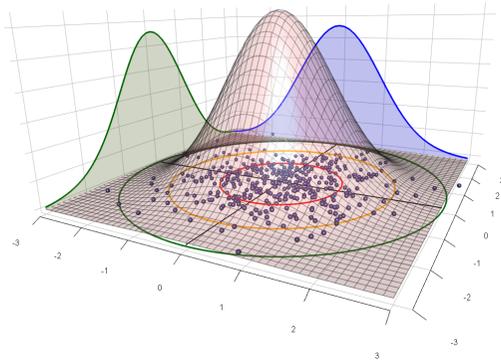
Note: Le densità in rosso e blu sono le densità marginali della distribuzione congiunta, quest'ultima rappresentata mediante i punti sul piano cartesiano $S_X \times X_Y$ dei supporti delle singole distribuzioni. La covarianza è rappresentata dall'ellisse di colore verde. Il valore della covarianza in questo caso è negativo.

Fonte: https://en.wikipedia.org/wiki/Multivariate_normal_distribution



Modello normale multivariato

Caso bivariato



Fonte: http://ballistipedia.com/index.php?title=Closed_Form_Precision

Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

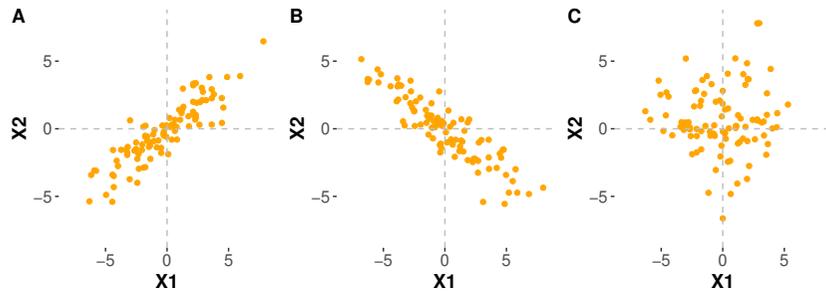
Il termine di covarianza σ_{jp} esprime l'associazione lineare tra la j -esima e la p -esima variabile:

$$\sigma_{jp} = \text{Cov} [X_j, X_p] = \mathbb{E} [(X_j - \mu_{X_j})(X_p - \mu_{X_p})]$$

ed esprime l'ammontare di *variazione congiunta* di X_j e X_p . La covarianza può essere maggiore o minore di 0, ossia $\sigma_{jp} \in \mathbb{R}$.

Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution



Caso bivariato: Rappresentazione grafica dei valori osservati congiuntamente delle variabili X_j e X_p .

A: Grafico con associazione lineare positiva $\sigma_{X_j X_p} = 6.003$.

B: Grafico con associazione lineare negativa $\sigma_{X_j X_p} = -6.003$.

C: Grafico con nessuna associazione lineare $\sigma_{X_j X_p} \approx 0$.

In tutti e tre i grafici $\mu_{X_j} = \mu_{X_p} = 0$ e varianze $\sigma_{X_j}^2 = 9.140$ e $\sigma_{X_p}^2 = 5.165$.

Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

La matrice di correlazione si ottiene per trasformazione da Σ :

$$P = \begin{bmatrix} 1 & \dots & \rho_{1j} & \dots & \rho_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho_{j1} & \dots & 1 & \dots & \rho_{jp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho_{p1} & \dots & \rho_{pj} & \dots & 1 \end{bmatrix}$$

- è simmetrica
- le deviazioni standard sono pari ad 1
- se le p variabili sono tra loro indipendenti $X_h \perp\!\!\!\perp X_k$ ($h \neq k$), P diventa matrice diagonale

Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

Il termine di correlazione ρ_{jp} esprime l'associazione lineare tra la j -esima e la p -esima variabile:

$$\rho_{jp} = \text{Cor}[X_j, X_p] = \frac{\text{Cov}[X_j, X_p]}{\sqrt{\text{Var}[X_j]}\sqrt{\text{Var}[X_p]}}$$

nell'intervallo reale $[-1, 1]$, dove gli estremi indicano rispettivamente l'associazione massima negativa e positiva.

La rappresentazione grafica è analoga a quella della covarianza.

Modello normale multivariato

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

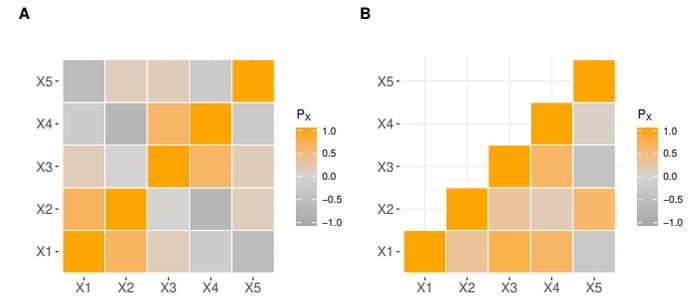


Grafico della matrice di correlazione: Rappresentazione grafica di una matrice di correlazione \mathbf{P} mediante gradiente dei colori: $\rho_{jp} \approx -1$ (grigio scuro), $\rho_{jp} \approx 0$ (grigio chiaro), $\rho_{jp} \approx 1$ (arancione). **A:** matrice di correlazione piena (parte triangolare superiore e inferiore). **B:** matrice di correlazione senza parte triangolare superiore. Si nota che la diagonale è di colore arancione poiché $\text{diag}[\mathbf{P}] = \mathbf{1}$.