

Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

PARTE II-B: CFA E ANALISI DELL'INVARIANZA

Antonio Calcagni

DPSS, Università di Padova

A.A. 2022/2023

Copyright © 2022 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

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

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

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 psicomatria applicata, è detto **schema riflesso-vo**, in contrasto con lo **schema formativo** (tipico, ad esempio, nella costruzione degli indicatori sociali) dove invece \mathbf{Y}_p precede \mathbf{Y}_q .

Introduzione

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

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

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

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

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

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 $\theta^{(2)}$, $\theta^{(1)}$ 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: Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*. John Wiley & Sons.

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 policorica*).

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.

CFA come modello a variabili latenti

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

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

$$\boldsymbol{\eta}_i \sim \mathcal{N}_q(\boldsymbol{\mu}_i, \boldsymbol{\Phi})$$

avente media $\boldsymbol{\mu}_i$ (vettore $q \times 1$ di reali) e matrice di covarianza $\boldsymbol{\Phi}$ (di ordine $q \times q$).

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

$$\mathbf{y}_i = \boldsymbol{\tau}_i + \boldsymbol{\Lambda} \boldsymbol{\eta}_i + \boldsymbol{\delta}_i$$

dove $\boldsymbol{\tau}_i$ è il vettore $p \times 1$ di intercette, $\boldsymbol{\Lambda}$ è una matrice $p \times q$ di coefficienti reali (c.d. *coefficienti fattoriali*) che legano le p osservate alle q latenti, $\boldsymbol{\delta}_i$ è il vettore $p \times 1$ di errore del modello.

CFA come modello a variabili latenti

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:

$$\mathbf{y}_i \sim \mathcal{N}_p(\underbrace{\boldsymbol{\tau}_i + \mathbf{\Lambda}\boldsymbol{\mu}_i}_{\boldsymbol{\mu}_y}, \underbrace{\mathbf{\Lambda}\Phi\mathbf{\Lambda}^T + \Theta_\delta}_{\boldsymbol{\Sigma}_y})$$

dove $\boldsymbol{\mu}_y$ è il vettore $p \times 1$ delle medie marginali delle variabili osservate mentre $\boldsymbol{\Sigma}_y$ ne è la matrice di covarianza di ordine $p \times p$.

Note:

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

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

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

CFA come modello a variabili latenti

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

$$\boldsymbol{\eta}_i \sim \mathcal{N}_q(\boldsymbol{\mu}_i, \boldsymbol{\Phi})$$

$$\boldsymbol{\delta}_i \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Theta}_\delta)$$

$$\mathbf{y}_i = \boldsymbol{\tau}_i + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\delta}_i$$

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

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

$$\boldsymbol{\mu} = \mathbf{0}_q$$

$$\boldsymbol{\tau} = \mathbf{0}_p$$

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

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

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

- **modelli unidimensionali** ($q = 1$):

$\Phi_{q \times q} = \phi$ contiene solo la varianza ϕ^2 dell'unica v.a. latente η

- **modelli bidimensionali** ($q = 2$):

$\Phi_{q \times q} = \Phi_{2 \times 2}$ 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** ($q > 2$):

$\Phi_{q \times q}$ contiene q varianze sulla diagonale relative alle variabili latenti e $\frac{1}{2}(q \times q - q)$ termini di covarianza

CFA: 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:

$$Y_1 = \eta\lambda_1 + \delta_1$$

$$\vdots$$

$$Y_j = \eta\lambda_j + \delta_j$$

$$\vdots$$

$$Y_p = \eta\lambda_p + \delta_p$$

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:

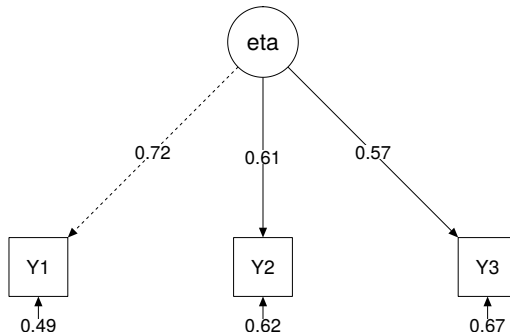
- c : componente del modello lineare *in comune* con le p osservabili
- u_j : componente del modello lineare specifica (*unica*) per ciascuna delle p osservabili

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 λ_{Y1} , λ_{Y2} , λ_{Y3} mentre i numeri posti lungo le frecce orientate sui quadrati indicano i residui del modello δ_{Y1} , δ_{Y2} , δ_{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 ($q < p$).

$$Y_1 = \eta_1 \lambda_{11} + \dots + \eta_k \lambda_{k1} + \dots + \eta_q \lambda_{q1} + \delta_1$$

$$\vdots$$

$$Y_j = \eta_1 \lambda_{1j} + \dots + \eta_k \lambda_{kj} + \dots + \eta_q \lambda_{qj} + \delta_j$$

$$\vdots$$

$$Y_p = \eta_1 \lambda_{1p} + \dots + \eta_k \lambda_{kp} + \dots + \eta_q \lambda_{qp} + \delta_p$$

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

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

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

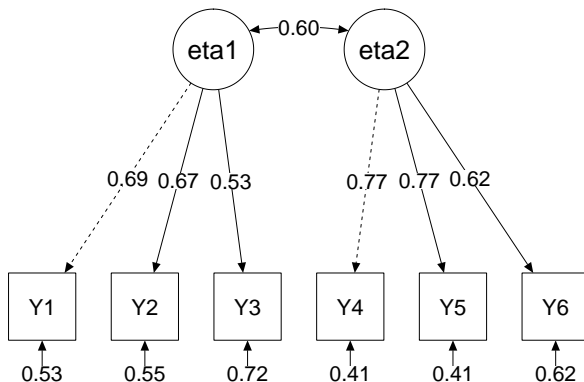
Il modello con q misurandi può essere scritto in forma compatta usando la notazione algebrica:

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

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 20:

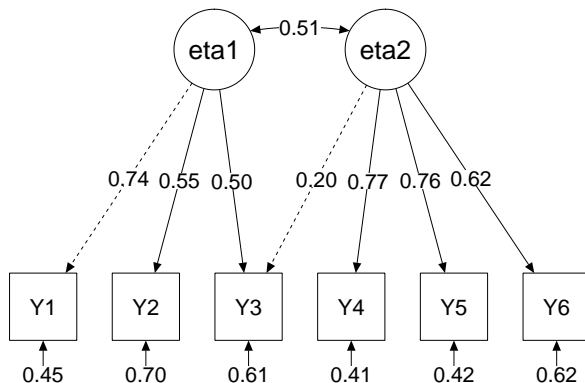
$$\mathbf{y}_{6 \times 1} = \mathbf{\Lambda}_{6 \times 2} \boldsymbol{\eta}_{2 \times 1} + \boldsymbol{\delta}_{6 \times 1}$$

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

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)

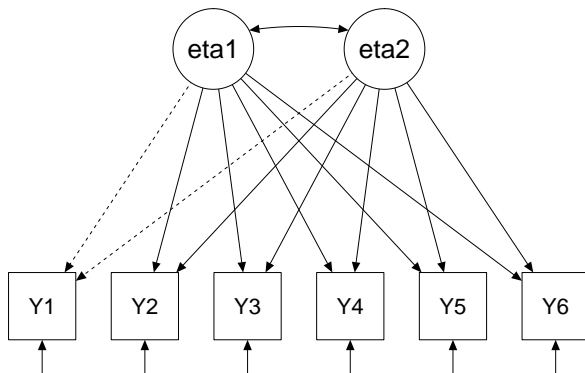
Esempio slide 22:

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

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

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 24:

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

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

Si noti come la struttura della matrice $\mathbf{\Lambda}$ contiene nel caso esplorativo tutti *parametri liberi*, ossia parametri da determinare (stimare) a partire dai dati.

CFA: modello multidimensionale

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

La rappresentazione algebrica vista negli esempi precedenti può essere agevolmente estesa ad includere il caso con p variabili manifeste e q variabili latenti, come segue:

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

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_j \\ \vdots \\ Y_p \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1k} & \dots & \lambda_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{j1} & \lambda_{j2} & \dots & \lambda_{jk} & \dots & \lambda_{jq} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pk} & \dots & \lambda_{pq} \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_k \\ \vdots \\ \eta_q \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_j \\ \vdots \\ \delta_p \end{bmatrix}}_{\boldsymbol{\delta}}$$

Come in precedenza, la matrice $\mathbf{\Lambda}$ (detta anche *factor pattern matrix*) è molto importante poiché esprime l'ipotesi del ricercatore circa la formazione delle scale di misura (essa contiene difatti i coefficienti che legano le osservabili ai misurandi).

CFA: assunzioni del modello

FONTI: CRB(VII.3.0)

Il modello lineare dell'analisi fattoriale confermativa,

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

sottende le seguenti assunzioni:

- (i) $\mathbb{E} [Y_j] = 0$ valore atteso delle v.a. osservate è uguale a zero
- (ii) $\mathbb{E} [\eta_k] = 0$ valore atteso delle v.a. latenti è uguale a zero
- (iii) $\mathbb{E} [\delta_j] = 0$ valore atteso delle v.a. errori è uguale a zero - vedi *legge degli errori accidentali*
- (iv) $\text{Cov} [\eta_k, \delta_j] = 0$ v.a. degli errori e misurandi sono incorrelati

CFA: matrice di covarianza riprodotta

FONTI: CRB(VI.2.1)

In precedenza abbiamo visto che la densità delle variabili osservate (Y_1, \dots, Y_p) segue in legge il modello (marginale) della densità Normale:

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

dove

$$\Sigma_y = \mathbf{\Lambda}\Phi\mathbf{\Lambda}^T + \Theta_\delta$$

è la matrice di varianze-covarianze del modello che ne rappresenta il parametro da stimare. La matrice Σ_y è detta **matrice di covarianza riprodotta** (o *matrice attesa*) dal modello.

La decomposizione ammessa da Σ_y permette di rappresentare tutte le varianze e covarianze del modello marginale di \mathbf{y} rispetto ai legami con le variabili latenti $\boldsymbol{\eta}$.

CFA: matrice di covarianza riprodotta

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

La CFA utilizza quale dato di input la matrice di varianze-covarianze delle osservabili $\text{Cov}[\mathbf{Y}] = \mathbf{S}_Y$ e non la matrice delle variabili osservate $\mathbf{Y}_{n \times p}$. Difatti:

$$\begin{aligned}\text{Cov}[\mathbf{Y}] &= \text{Cov}[(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta}), (\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})] && \text{usiamo la definizione del modello CFA: } Y = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta} \\ &= \mathbb{E} \left[(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})^T \right] && \text{usiamo la definizione di covarianza notando che } \tau = 0 \\ \dots &&& \text{usando gli assunti (i)-(iv) del modello, con alcune manipolazioni algebriche otteniamo:} \\ &= \mathbf{\Lambda} (\text{Cov}[\boldsymbol{\eta}, \boldsymbol{\eta}]) \mathbf{\Lambda}^T + \text{Cov}[\boldsymbol{\delta}, \boldsymbol{\delta}] \\ &= \boxed{\mathbf{\Lambda}\boldsymbol{\Phi}\mathbf{\Lambda}^T + \boldsymbol{\Theta}_{\boldsymbol{\delta}}}\end{aligned}$$

dove:

- $\boldsymbol{\eta} = (\eta_1, \dots, \eta_q)$ è il vettore aleatorio q -variato per le q variabili latenti
- $\mathbf{Y} = (Y_1, \dots, Y_p)$ è il vettore aleatorio p -variato per le p variabili osservate
- $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)$ è il vettore aleatorio p -variato per le p componenti di errore

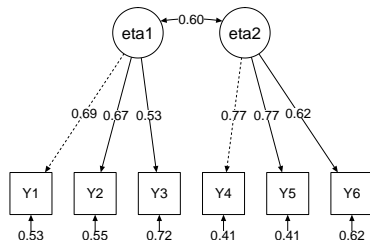


per ulteriori dettagli, si veda `slide29.pdf` degli Approfondimenti

CFA: matrice di covarianza riprodotta

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

Riprendiamo l'esempio rappresentato nella slide 20:



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

$$\mathbf{\Lambda} = \begin{bmatrix} 0.69 & 0 \\ 0.67 & 0 \\ 0.53 & 0 \\ 0 & 0.77 \\ 0 & 0.77 \\ 0 & 0.62 \end{bmatrix} \quad \mathbf{\Phi} = \begin{bmatrix} 1 & 0.60 \\ 0.60 & 1 \end{bmatrix}$$

$$\mathbf{\Theta}_{\delta} = \begin{bmatrix} 0.53 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.72 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.41 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.62 \end{bmatrix}$$

CFA: decomposizione delle covarianze osservate

FONTI: CRB(VI.2)

Data una matrice di covarianza osservata \mathbf{S}_Y un modello CFA esprime le varianze-covarianze osservate in funzione dei parametri (ignoti) del modello.

Si consideri la scala $\eta^{(1)}$ del modello riportato nelle slide 22-23 (modello con due fattori latenti ed un *cross-loading*).

CFA: decomposizione delle covarianze osservate

FONTI: CRB(VI.2)

Le equazioni (scalari) sono le seguenti:

$$Y_{1i} = \eta_{1i}\lambda_{11} + \delta_{1i}$$

$$Y_{2i} = \eta_{1i}\lambda_{21} + \delta_{2i}$$

$$Y_{3i} = \eta_{1i}\lambda_{31} + \eta_{2i}\lambda_{32} + \delta_{3i}$$

con la seguente matrice di covarianza:

$$\mathbf{S}_Y = \begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

CFA: decomposizione delle covarianze osservate

FONTI: CRB(VI.2)

Le varianze delle v.a. $\{Y_1, Y_2, Y_3\}$ (diagonale di \mathbf{S}_Y) sono espresse come di seguito (con medie nulle):

$$\begin{aligned}\sigma_1^2 &= \text{Var}[Y_{1i}] = \mathbb{E}[Y_{1i} Y_{1i}] = \mathbb{E}[(\eta_{1i} \lambda_{11} + \delta_{1i})(\eta_{1i} \lambda_{11} + \delta_{1i})] \\&= \lambda_{11}^2 \text{Var}[\eta_{1i}] + 2\lambda_{11} \mathbb{E}[\eta_{1i} \delta_{1i}] + \text{Var}[\delta_{1i}] \\&= \lambda_{11}^2 \text{Var}[\eta_{1i}] + 2\lambda_{11} \text{Cov}[\eta_{1i}, \delta_{1i}] + \text{Var}[\delta_{1i}] \\&= \underbrace{\lambda_{11}^2 \phi_{11}}_{\text{varianza spiegata da } \eta_1} + \underbrace{\theta_{\delta_{11}}}_{\text{varianza non spiegata da } \eta_1}\end{aligned}$$

Analogamente si procede per σ_2^2 .

CFA: decomposizione delle covarianze osservate

FONTI: CRB(VI.2)

$$\begin{aligned}\sigma_3^2 &= \text{Var}[Y_{3i}] = \mathbb{E}[(\eta_{1i}\lambda_{31} + \eta_{2i}\lambda_{32} + \delta_{3i})(\eta_{1i}\lambda_{31} + \eta_{2i}\lambda_{32} + \delta_{3i})] \\&= \lambda_{31}^2 \text{Var}[\eta_{1i}] + 2\lambda_{31}\lambda_{32}\text{Cov}[\eta_{1i}, \eta_{2i}] + \lambda_{32}^2 \text{Var}[\eta_{2i}] + \\&\quad + 2\lambda_{31}\cancel{\text{Cov}[\eta_{1i}, \delta_{3i}]}^{\rightarrow=0} + 2\lambda_{32}\cancel{\text{Cov}[\eta_{2i}, \delta_{3i}]}^{\rightarrow=0} + \text{Var}[\delta_{3i}] \\&= \lambda_{31}^2 \text{Var}[\eta_{1i}] + 2\lambda_{31}\lambda_{32}\text{Cov}[\eta_{1i}, \eta_{2i}] + \lambda_{32}^2 \text{Var}[\eta_{2i}] + \text{Var}[\delta_{3i}] \\&= \underbrace{\lambda_{31}^2\phi_{11} + \lambda_{32}^2\phi_{22} + 2\phi_{12}\lambda_{31}\lambda_{32}}_{\text{varianza spiegata}} + \underbrace{\theta_{\delta_{33}}}_{\text{varianza non spiegata}}\end{aligned}$$

CFA: decomposizione delle covarianze osservate

FONTI: CRB(VI.2)

$$\begin{aligned}\sigma_{12} &= \text{Cov}[Y_{1i} Y_{2i}] = \mathbb{E}[(\eta_{1i}\lambda_{11} + \delta_{1i})(\eta_{1i}\lambda_{21} + \delta_{2i})] \\&= \lambda_{11}\lambda_{21}\mathbb{E}[\eta_{1i}\eta_{1i}] + \lambda_{11}\mathbb{E}[\eta_{1i}\delta_{1i}] + \lambda_{21}\mathbb{E}[\eta_{1i}\delta_{2i}] + \mathbb{E}[\delta_1\delta_2] \\&= \lambda_{11}\lambda_{21}\phi_{11} + \lambda_{11}\cancel{\text{Cov}[\eta_{1i}, \delta_{1i}]} = 0 + \\&+ \lambda_{21}\cancel{\text{Cov}[\eta_{1i}, \delta_{2i}]} = 0 + \text{Cov}[\delta_1, \delta_2] \\&= \lambda_{11}\lambda_{21}\phi_{11} + \theta_{\delta_{21}}\end{aligned}$$

Il termine $\theta_{\delta_{21}}$ è la covarianza tra gli errori delle v.a. $\{Y_1, Y_2\}$. Per semplificare il modello spesso è usato l'assunto per cui Θ_δ è una *matrice diagonale* dove si ha che $\text{Cov}[\delta_j, \delta_h] = \theta_{\delta_{jh}} = 0$ (per ogni $j \neq k$).

CFA: decomposizione delle covarianze osservate

FONTI: CRB(VI.2)

Nota: se $\phi_{11} = 1$ e $\theta_{\delta_{21}} = 0$ (caso spesso frequente) allora

$$\begin{aligned}\sigma_{12} = \text{Cov}[Y_{1i} Y_{2i}] &= \lambda_{11} \lambda_{21} \phi_{11} + \theta_{\delta_{21}} \\ &= \lambda_{11} \lambda_{21}\end{aligned}$$

la covarianza/correlazione tra Y_1 e Y_2 è uguale al prodotto dei rispettivi *factor loadings*.

Analogamente si procede per σ_{13} e σ_{32} .

La decomposizione della matrice \mathbf{S}_Y in funzione dei parametri del modello è di cruciale importanza per studiare se un modello CFA ipotizzato risulta **identificabile** o meno (tale aspetto verrà studiato in seguito).

CFA: stime dei parametri del modello

FONTE: CRB(IV.1,IV.2.1,IV.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).

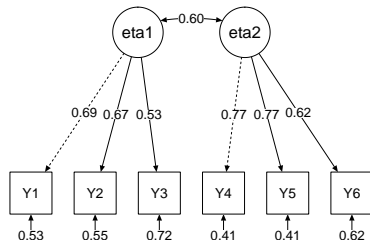
L'approccio confermativo - proprio perché forza il ricercatore a fare delle ipotesi sul modello di misura - non necessita di stimare/determinare le matrici Λ , Φ , Θ_δ nella loro interezza ma solo quelle componenti di queste che sono diverse da zero (*parametri liberi*).

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

FONTI: CRB(IV.1,IV.2.1,IV.3)

Nell'esempio della slide 20 abbiamo 13 parametri liberi (da stimare):



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

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

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

CFA: stime dei parametri del modello

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

La parsimoniosità 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 20 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

FONTE: 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 20: $\lambda_{11} = 1$ 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 20: $\phi_{\eta_1} = \phi_{\eta_2} = 1$.

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)

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 $\Sigma_Y = \Lambda\Phi\Lambda^T + \Theta_\delta$:

- **massima verosimiglianza:**

riprendendo la densità marginale del modello (vedi slide 11), 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)

- **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

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

CFA: stime dei parametri del modello

FONTE: CRB(IV.1,IV.2.1,IV.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)

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: identificabilità del modello

FONTI: CRB(IV.6)

In alcune situazioni il modello CFA ipotizzato rispetto ad una matrice di covarianza osservata \mathbf{S}_Y potrebbe non essere identificabile.

Formalmente un modello CFA è identificato se tutti i suoi parametri sono esprimibili in funzioni dei dati osservati o delle statistiche di sintesi (es.: varianze-covarianze). Riprendendo l'esempio della decomposizione delle covarianze osservate (slide 31), il modello CFA è identificato se è possibile esprimere i parametri $\Lambda_{3 \times 1}$, $\Phi_{1 \times 1}$, $\Theta_{\delta_{3 \times 3}}$ in funzione di $\mathbf{Y}_{3 \times 3}$ (si veda la Tabella 6.2 di CRB(IV.6, p. 139) per un esempio svolto di identificabilità).

Una condizione necessaria affinché un modello CFA sia identificato è quella per cui $df > 0$.

Un modo per migliorare la condizione di non identificabilità è quella di introdurre vincoli al modello CFA (vale a dire: fissare dei parametri).

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

FONTI: CRB(VII.2.1-VII.2.3)

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)

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 $\hat{\Lambda}$

FONTE: CRB(VII.2.1-VII.2.3)

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

È facile identificare tre gruppi distinti (non sovrapposti) di indicatori manifesti che legano ai 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)

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).

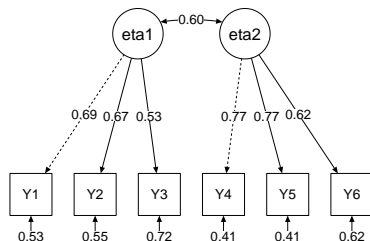
Esistono diverse scelte per la rotazione di $\hat{\Lambda}$:

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

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

FONTI: CRB(VII.2.1-VII.2.3)

Riprendiamo l'esempio rappresentato nella slide 20:



$$\Sigma_Y = \Lambda \Phi \Lambda^T + \Theta_\delta$$

$$\hat{\Lambda} = \begin{bmatrix} 0.69 & 0 \\ 0.67 & 0 \\ 0.53 & 0 \\ 0 & 0.77 \\ 0 & 0.77 \\ 0 & 0.62 \end{bmatrix} \quad \hat{\Phi} = \begin{bmatrix} 1 & 0.60 \\ 0.60 & 1 \end{bmatrix}$$

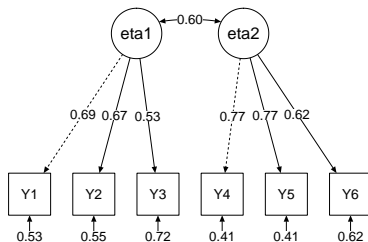
$$\hat{\Theta}_\delta = \begin{bmatrix} 0.53 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.72 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.41 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.62 \end{bmatrix}$$

Esempio di stima dei parametri del modello

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

FONTI: CRB(VII.2.1-VII.2.3)

Riprendiamo l'esempio rappresentato nella slide 20:



$$\mathbf{S}_Y = \begin{bmatrix} & Y1 & Y2 & Y3 & Y4 & Y5 & Y6 \\ Y1 & 1.00 & & & & & \\ Y2 & 0.43 & 1.00 & & & & \\ Y3 & 0.41 & 0.35 & 1.00 & & & \\ Y4 & 0.28 & 0.35 & 0.16 & 1.00 & & \\ Y5 & 0.32 & 0.32 & 0.19 & 0.59 & 1.00 & \\ Y6 & 0.24 & 0.32 & 0.18 & 0.47 & 0.46 & 1.00 \end{bmatrix}$$

$$\hat{\Sigma}_Y = \begin{bmatrix} & Y1 & Y2 & Y3 & Y4 & Y5 & Y6 \\ Y1 & 1.00 & & & & & \\ Y2 & 0.46 & 1.00 & & & & \\ Y3 & 0.36 & 0.36 & 1.00 & & & \\ Y4 & 0.31 & 0.31 & 0.24 & 1.00 & & \\ Y5 & 0.31 & 0.31 & 0.24 & 0.59 & 1.00 & \\ Y6 & 0.25 & 0.25 & 0.19 & 0.47 & 0.47 & 1.00 \end{bmatrix}$$

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 $\text{diag}(\Sigma_V)$, 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)

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 36). In particolare, per la stessa matrice di covarianza osservata 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 34)
- (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

FONTI: CRB(I.3,VI.1,VI.3,VI.4)

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 36). 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 34)
- (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

FONTI: CRB(I.3,VI.1,VI.3,VI.4)

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 36). 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 34)
- (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)

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)

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 $\text{Cov}[\mathbf{y}] = \phi \mathbf{I}_{p \times p} + \Theta_\delta$, 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 $\mathbf{\Lambda} = \mathbf{0}_{p \times q}$.

CFA: adeguatezza dell'adattamento del modello

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 34).

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

FONTI: CRB(VI.5.2)

Indici di modifica

Permettono di valutare come la statistica χ^2_1 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 $\mathbf{S}_Y - \hat{\Sigma}_Y$ alla ricerca di quei coefficienti $\theta_{\delta_{jj'}}$ 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)

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

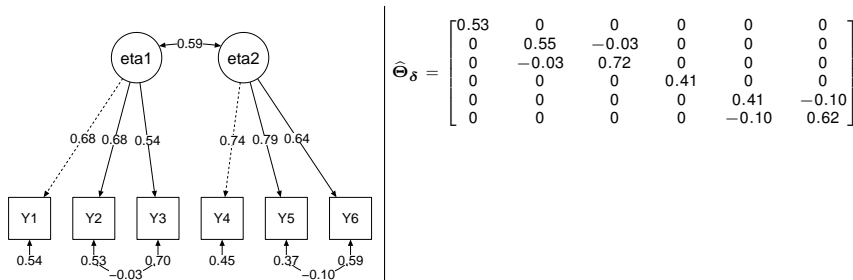
Al contrario, liberare un parametro della matrice - ad esempio $\theta_{\delta_{2j}}$ - equivale a correlare gli errori δ della variabile osservata 2 e j e ciò equivale a dire che le due osservabili esprimono, in qualche modo, una medesima informazione sul misurando η (e quindi sono “accomunabili”, convogliano 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 20:



Esempio di modello confermativo con errori correlati: i parametri in oggetto sono $\theta_{\delta_{32}}$ e $\theta_{\delta_{56}}$.

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, utilizzi 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	$\text{Cov}[\mathbf{y}] = \lambda \mathbf{1}_p \phi \mathbf{1}_p^T + \mathbf{\Theta}_\delta$
congenerico	$\text{Cov}[\mathbf{y}] = \lambda \phi \lambda^T + \mathbf{\Theta}_\delta$

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: calcolo dei punteggi fattoriali

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 costrutto 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, Thurtstone 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):

- ❶ VCI: similarities (SO), vocabulary (VC), comprehension (CO)
- ❷ PRI: block design (DC), picture concepts (CI), matrix reasoning (RM)
- ❸ WMI: digit span (MC), letter-number sequencing (LN)
- ❹ 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 sostengono 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⁾ ($n = 2200$):

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

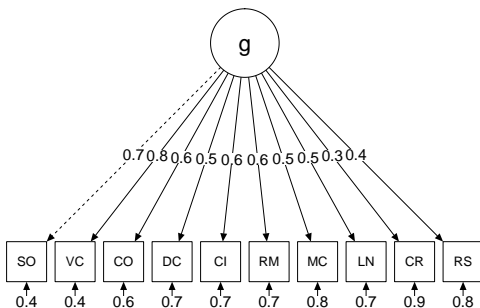
Procederemo esplorando quattro modelli fattoriali diversi (M1-M4) e valuteremo la loro adeguatezza rispetto ai dati campionari.

⁽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



$$\Sigma_Y = \Lambda \Phi \Lambda^T + \Theta_\delta$$

$$\hat{\Lambda} = \begin{bmatrix} 1.00 \\ 1.01 \\ 0.85 \\ 0.70 \\ 0.79 \\ 0.78 \\ 0.64 \\ 0.72 \\ 0.37 \\ 0.54 \end{bmatrix} \quad \hat{\Phi} = [0.56]$$

$$\hat{\Theta}_\delta = \begin{bmatrix} 0.44 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.43 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.59 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.72 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.65 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.65 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.77 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.71 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.92 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.84 \end{bmatrix}$$

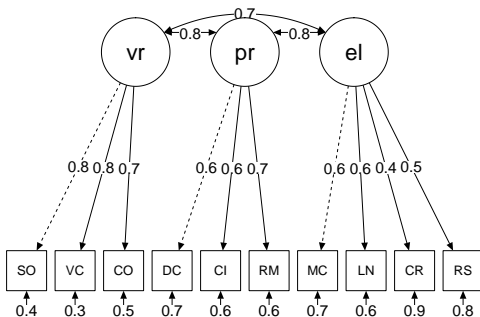
RMSEA = 0.11, CFI = 0.84, AIC = 57359

Nota: i coefficienti rappresentati 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)



$$\Sigma_Y = \Lambda \Phi \Lambda^T + \Theta_\delta$$

$$\hat{\Lambda} = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 1.06 & 0.00 & 0.00 \\ 0.89 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 1.08 & 0.00 \\ 0.00 & 1.14 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.14 \\ 0.00 & 0.00 & 0.68 \\ 0.00 & 0.00 & 0.86 \end{bmatrix} \quad \hat{\Phi} = \begin{bmatrix} 0.59 & & \\ 0.35 & 0.35 & \\ 0.29 & 0.27 & 0.31 \end{bmatrix}$$

$$\hat{\Theta}_\delta = \begin{bmatrix} 0.41 & & & & & & & & & \\ 0.00 & 0.33 & & & & & & & & \\ 0.00 & 0.00 & 0.53 & & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.65 & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.59 & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.55 & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.69 & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.59 & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.86 & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.77 \end{bmatrix}$$

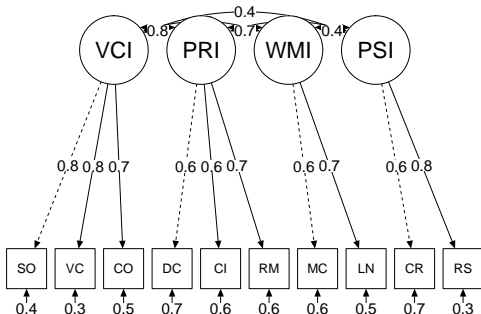
RMSEA = 0.08, CFI = 0.91, AIC = 56904

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



$$\Sigma_Y = \Lambda \Phi \Lambda^T + \Theta_\delta$$

$$\hat{\Lambda} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 1.06 & 0.00 & 0.00 & 0.00 \\ 0.89 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 1.08 & 0.00 & 0.00 \\ 0.00 & 1.14 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.16 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 0.00 & 1.45 \end{bmatrix}$$

$$\hat{\Phi} = \begin{bmatrix} 0.59 & & & \\ 0.35 & 0.35 & & \\ 0.31 & 0.26 & 0.36 & \\ 0.16 & 0.17 & 0.14 & 0.32 \end{bmatrix}$$

$$\hat{\Theta}_\delta = \begin{bmatrix} 0.41 & & & & & & & & & \\ 0.00 & 0.33 & & & & & & & & \\ 0.00 & 0.00 & 0.53 & & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.65 & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.59 & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.55 & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.64 & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.51 & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.68 & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.33 \end{bmatrix}$$

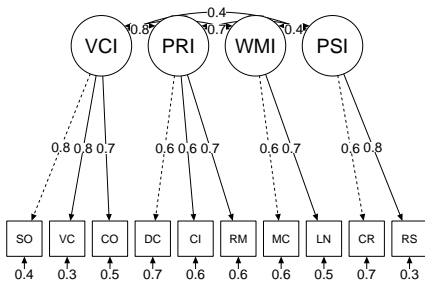
RMSEA = 0.04, CFI = 0.97, AIC = 56548

Nota: i coefficienti rappresentati 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



Modification indices (mi)

lhs	op	rhs	mi	epc
PRI	=~	SO	44.66	0.53
VC	~~	CO	40.93	0.11
PSI	=~	DC	28.84	0.29
CI	~~	CR	21.42	-0.08
VCI	=~	CI	21.24	0.31
DC	~~	CR	19.66	0.07
PSI	=~	CI	13.60	-0.20
SO	~~	VC	13.40	-0.08
DC	~~	CI	13.13	-0.07
PRI	=~	VC	12.76	-0.30
...

$$\mathbf{Z}_Y = (\mathbf{S}_Y - \Sigma_Y) =$$

-0.00										
-0.90	0.00									
0.95	-0.15	-0.00								
0.07	-2.49	-1.45	0.00							
2.43	2.73	3.00	-3.87	0.00						
-2.31	-0.23	-2.20	2.70	0.66	0.00					
2.45	0.63	-1.79	-0.34	-0.35	-0.75	0.00				
-2.18	0.76	-0.35	-1.29	1.90	0.49	-0.00	0.00			
-2.84	-2.66	1.03	5.42	-5.27	0.77	0.44	1.84	-0.00		
0.86	-0.07	1.89	4.02	-2.13	-1.77	-0.84	-0.28	-0.00	-0.00	

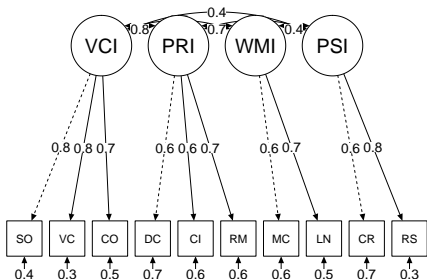
RMSEA = 0.04, CFI = 0.97, AIC = 56548

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

CFA: caso studio

Studio della dimensionalità della WISC-IV

M3: Scala VCI, PRI, WMI, PSI



Indici di attendibilità

	VCI	PRI	WMI	PSI	total
α	0.80	0.66	0.59	0.63	0.82
ω	0.80	0.67	0.59	0.65	0.86

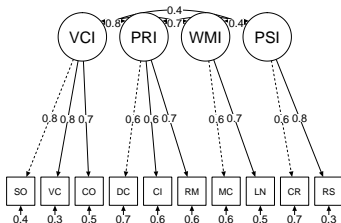
Gli indici sono calcolati separatamente per ciascuna scala della WISC-IV e complessivamente (*total*).

RMSEA = 0.04, CFI = 0.97, AIC = 56548

CFA: caso studio

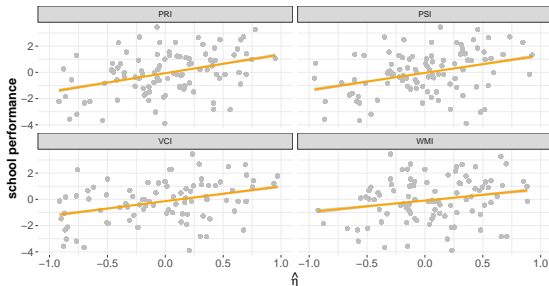
Studio della dimensionalità della WISC-IV

M3: Scala VCI, PRI, WMI, PSI



Dopo aver adattato il modello CFA ai dati, è possibile calcolare i valori dei misurandi $\hat{\eta}_i$ (factor scores) per ciascuna unità statistica su cui sono stati rilevati i dati. In questo caso, possiamo calcolare i valori dei costrutti $\hat{\eta}_{VCI}$, $\hat{\eta}_{PRI}$, $\hat{\eta}_{WMI}$, $\hat{\eta}_{PSI}$ per ciascun soggetto.

Questi possono essere usati per analisi successive, ad esempio per predire la performance scolastica dei soggetti sottoposti a WISC-IV:

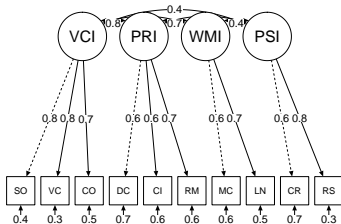


RMSEA = 0.04, CFI = 0.97, AIC = 56548

CFA: caso studio

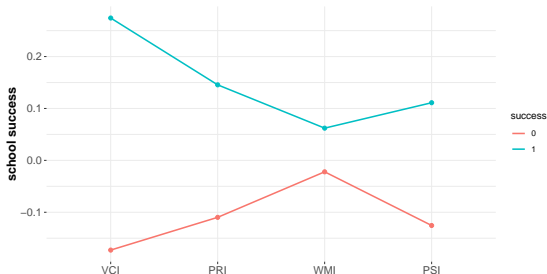
Studio della dimensionalità della WISC-IV

M3: Scala VCI, PRI, WMI, PSI



Dopo aver adattato il modello CFA ai dati, è possibile calcolare i valori dei misurandi $\hat{\eta}_i$ (factor scores) per ciascuna unità statistica su cui sono stati rilevati i dati. In questo caso, possiamo calcolare i valori dei costrutti $\hat{\eta}_{VCI}$, $\hat{\eta}_{PRI}$, $\hat{\eta}_{WMI}$, $\hat{\eta}_{PSI}$ per ciascun soggetto.

Possono altresì essere usati per creare *profili* rispetto ad una variabile concomitante, ad esempio il successo scolastico (1=successo, 0=insuccesso).



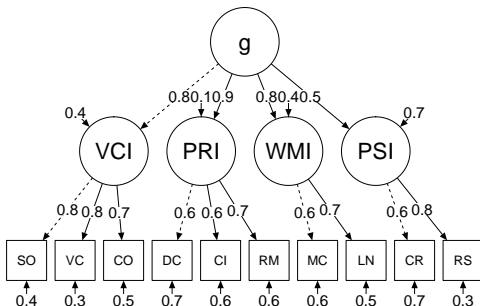
Notiamo come individui con successo scolastico presentino profili WISC-IV differenti da coloro con basso successo scolastico.

RMSEA = 0.04, CFI = 0.97, AIC = 56548

CFA: caso studio

Studio della dimensionalità della WISC-IV

M4: Scala VCI, PRI, WMI, PSI con **fattore sovraordinato** g



$$\Sigma_Y = \Lambda \Phi \Lambda^T + \Theta_\delta$$

$$\hat{\Lambda} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.06 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.89 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.09 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.14 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.16 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.50 & 0.00 \end{bmatrix}$$

$$\hat{\Phi} = \begin{bmatrix} 0.21 & & & & \\ 0.00 & 0.04 & & & \\ 0.00 & 0.00 & 0.13 & & \\ 0.00 & 0.00 & 0.00 & 0.23 & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.39 \end{bmatrix}$$

$$\hat{\Theta}_\delta = \begin{bmatrix} 0.41 & & & & & & & & & \\ 0.00 & 0.33 & & & & & & & & \\ 0.00 & 0.00 & 0.53 & & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.66 & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.59 & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.55 & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.64 & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.51 & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.69 & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.31 \end{bmatrix}$$

RMSEA = 0.04, CFI = 0.98, AIC = 56553

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

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

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>

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.

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[†]:

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

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

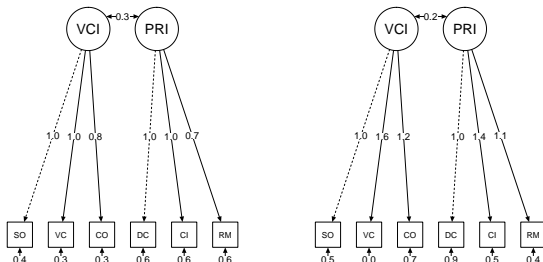
dove ciascun gruppo ha un proprio modello fattoriale sotto il vincolo che i parametri fissi e quelli liberi della matrice $\boldsymbol{\Lambda}$ siano i medesimi tra i due gruppi. Le stime $\hat{\boldsymbol{\Lambda}}^{(1)}$ e $\hat{\boldsymbol{\Lambda}}^{(2)}$ possono differire ma le strutture $\boldsymbol{\Lambda}^{(1)}$ 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 omissso per semplicità notazionale.

Invarianza di misurazione

Invarianza configurale



χ^2	df	m	AIC
15.46	16	38	1554.74

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

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

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

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

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

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

Invarianza di misurazione

Invarianza debole

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

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

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

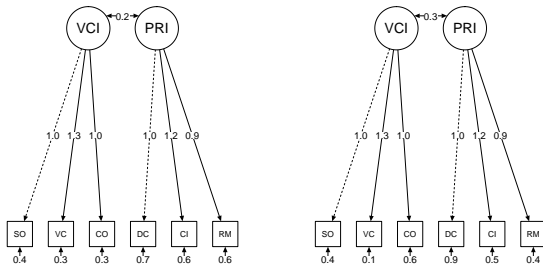
dove ciascun gruppo ha un proprio modello fattoriale sotto il vincolo:

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

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



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

$$\hat{\Phi}^{(1)} = \begin{bmatrix} 0.34 & \\ & 0.21 & 0.35 \end{bmatrix} \quad \hat{\Phi}^{(2)} = \begin{bmatrix} 0.52 & \\ & 0.26 & 0.31 \end{bmatrix}$$

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

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

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

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

	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

Conclusione: 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:

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

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

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

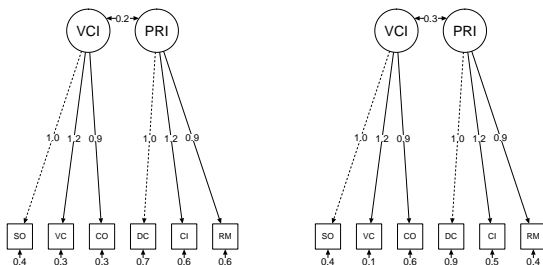
$$\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$$

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

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



$$\hat{\Lambda}^{(1)} = \hat{\Lambda}^{(2)} = \begin{bmatrix} 1.00 & 0.00 \\ 1.22 & 0.00 \\ 0.93 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.18 \\ 0.00 & 0.92 \end{bmatrix}$$

$$\hat{\Phi}^{(1)} = \begin{bmatrix} 0.38 & \\ & 0.22 & 0.33 \end{bmatrix} \quad \hat{\Phi}^{(2)} = \begin{bmatrix} 0.57 & \\ & 0.27 & 0.29 \end{bmatrix}$$

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

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

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

	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

Conclusione: 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 τ nei due gruppi.

Invarianza di misurazione

Invarianza esatta

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

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

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

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

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

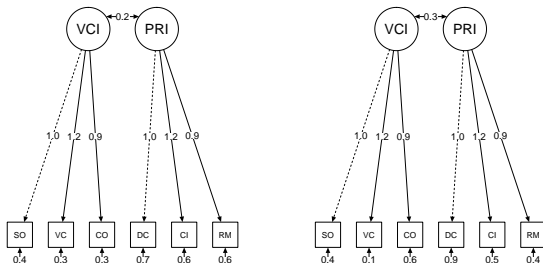
$$\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$$

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

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



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

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

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

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

	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

Conclusione: 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 $\hat{\eta}^{(1)}$ e $\hat{\eta}^{(2)}$.

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 | $\mathbf{y}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ | invarianza configurale |
| 2 | $\mathbf{y}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ | invarianza debole |
| 3 | $\mathbf{y}^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ | invarianza forte |
| 4 | $\mathbf{y}^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}$ | 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>