

# Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

Antonio Calcagni

Dipartimento di Psicologia dello Sviluppo e della Socializzazione (DPSS)  
Università di Padova

A.A. 2019/2020

Copyright © 2019 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

Nel modulo C abbiamo introdotto la TCT come modello statistico per rappresentare la misurazione di attributi psicologici.

Questa si basa sull'**assunzione di unidimensionalità**: dato un insieme di variabili osservate  $(X_1, \dots, X_j, \dots, X_p)$ , l'equazione di base

$$X_j = T + E_j$$

definisce ciascuna delle osservabili in relazione ad un unico misurando  $T$  che *forma* gli items. In altri termini, una scala di misurazione deve sottendere un solo misurando (tratto latente). Sebbene un test possa essere formato da più scale e dunque anche da più misurandi latenti, una scala può misurare un solo misurando.

Ciò è anche alla base della definizione dell'attendibilità come *coerenza interna* di una scala: questa, se definita "bene", deve avere items coerenti tra loro, nel senso - come abbiamo visto nel modulo C - che  $\text{Var}[T] > \text{Var}[E]$  (sempre), caso che si ha quando  $\text{Cor}[X_1, \dots, X_p] \approx 1$  (caso ideale).

# Introduzione

La costruzione di un test, dunque, necessita di costruire buone scale di misura nel senso dato dall'assunzione di unidimensionalità. L'**analisi della dimensionalità** di un test è la procedura statistica che permette di valutare la dimensionalità dei modelli di misura (dimensionalità delle scale). Una scala monodimensionale assicura che essa sia un buon *proxy* per quantificare il misurando oggetto di interesse.

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

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

All'interno di questo modulo ci concentreremo sulla CFA. Per chi fosse interessato all'EFA, maggiori dettagli sono disponibili in BN(4.1).

# Introduzione

L'analisi fattoriale confermativa (CFA), come anche quella esplorativa (EFA), utilizza come *dato di partenza* la **matrice di correlazione** delle variabili osservate

$\text{Cov}[\mathbf{X}] = \text{Cov}[X_1, \dots, X_j, \dots, X_p]$ , la quale viene scomposta ed analizzata per determinare se la variabilità delle osservabili sia riconducibile ad una (ed una sola) *variabile latente* (misurando).

Mentre nel *modello di regressione lineare* semplice e multiplo il dato di partenza è il vettore dei dati osservati  $\mathbf{y}$  ed una matrice di predittori  $\mathbf{X}$ , posti in relazione secondo il modello  $\mathbf{y} = \alpha + \mathbf{X}\beta + \epsilon$ , la CFA utilizza solamente una *sintesi dei dati* (una statistica) che è la covarianza tra le osservabili  $\text{Cov}[\mathbf{X}]$ , in assenza di una variabile risposta  $\mathbf{y}$  da porre in relazione a  $\mathbf{X}$ .

Per tale motivo, nel linguaggio del *data mining*, spesso questo tipo di tecnica è anche definita come non-supervisionata (*unsupervised method*).

# Introduzione

Richiami di probabilità: Covarianza e correlazione

[https://en.wikipedia.org/wiki/Correlation\\_and\\_dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence)

Date due v.a.  $X_1$  e  $X_2$  con media e varianza finita

$$\mathbb{E}[X_1] = \mu_{X_1} \quad \mathbb{E}[X_2] = \mu_{X_2} \quad \text{Var}[X_1] = \sigma_{X_1}^2 \quad \text{Var}[X_2] = \sigma_{X_2}^2$$

la **covarianza** tra  $X_1$  e  $X_2$  è definita come di seguito:

$$\text{Cov}[X_1, X_2] = \sigma_{X_1 X_2} = \mathbb{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]$$

mentre la *matrice di covarianza* di  $X = (X_1, X_2)$  è definita come di seguito:

$$\text{Cov}[X] = \Sigma_X = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 \end{bmatrix}$$

dove:

- o la diagonale contiene le varianze
- o i triangoli sotto e sopra la diagonale sono simmetrici
- o se  $X_1 \perp\!\!\!\perp X_2$  (variabili indipendenti):  $\sigma_{X_1 X_2} = \sigma_{X_2 X_1} = 0$

# Introduzione

Richiami di probabilità: Covarianza e correlazione

[https://en.wikipedia.org/wiki/Correlation\\_and\\_dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence)

La **correlazione** tra  $X_1$  e  $X_2$  è ottenuta per trasformazione dalla covarianza:

$$\text{Cor}[X_1, X_2] = \rho_{X_1 X_2} = \frac{\text{Cov}[X_1, X_2]}{\sqrt{\text{Var}[X_1]} \sqrt{\text{Var}[X_2]}}$$

e la *matrice di correlazione* di  $X = (X_1, X_2)$  è definita come di seguito:

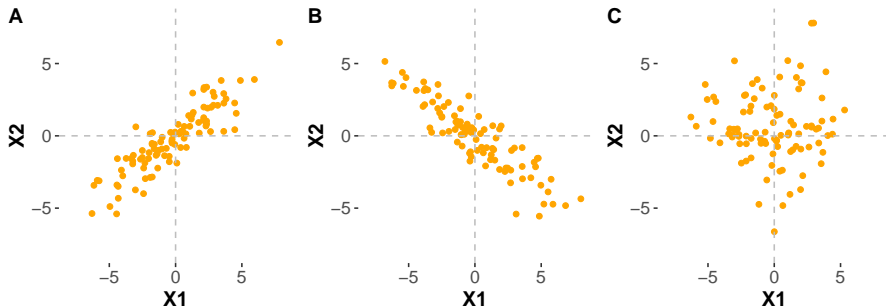
$$\text{Cor}[X] = \mathbf{P}_X = \begin{bmatrix} 1 & \rho_{X_1 X_2} \\ \rho_{X_2 X_1} & 1 \end{bmatrix}$$

dove:

- o la diagonale è unitaria
- o i triangoli sotto e sopra la diagonale sono simmetrici
- o se  $X_1 \perp\!\!\!\perp X_2$  (variabili indipendenti):  $\rho_{X_1 X_2} = \rho_{X_2 X_1} = 0$

# Introduzione

## Richiami di probabilità: Covarianza e correlazione



**Grafico a dispersione bivariato:** Rappresentazione grafica dei valori osservati congiuntamente delle variabili  $X_1$  e  $X_2$ .

**A:** Grafico con associazione lineare positiva  $\sigma_{X_1 X_2} = 6.003$ .

**B:** Grafico con associazione lineare negativa  $\sigma_{X_1 X_2} = -6.003$ .

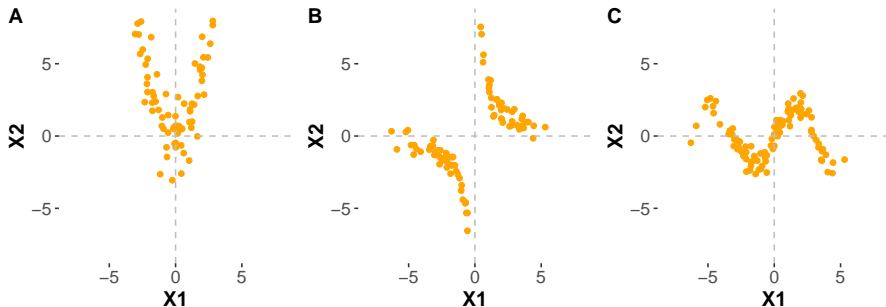
**C:** Grafico con nessuna associazione lineare  $\sigma_{X_1 X_2} \approx 0$ .

In tutti e tre i grafici  $\mu_{X_1} = \mu_{X_2} = 0$  e varianze  $\sigma_{X_1}^2 = 9.140$  e  $\sigma_{X_2}^2 = 5.165$ .



# Introduzione

## Richiami di probabilità: Covarianza e correlazione



**Grafico a dispersione bivariato:** Rappresentazione grafica dei valori osservati congiuntamente delle variabili  $X_1$  e  $X_2$ .

**A:** Grafico con associazione non lineare (parabola)  $\sigma_{X_1 X_2} \approx 0$ .

**B:** Grafico con associazione non lineare (iperbole)  $\sigma_{X_1 X_2} \approx 0$ .

**C:** Grafico con associazione non lineare (sinusoide)  $\sigma_{X_1 X_2} \approx 0$ .

Nota: L'operatore  $\text{Cov}[X_1, X_2]$  (come anche  $\text{Cor}[X_1, X_2]$ ) è lineare: quando tende ad assumere valori prossimi a zero non indica che non vi è *nessuna* associazione tra le variabili ma, al contrario, che non vi è associazione lineare. Nei tre pannelli, difatti, osserviamo indici di covarianza/correlazione pari a zero sebbene le variabili mostrino pattern di associazione non lineare.

# Introduzione

Richiami di probabilità: Covarianza e correlazione

[https://en.wikipedia.org/wiki/Correlation\\_and\\_dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence)

Nel caso  $p$ -variato  $X = (X_1, \dots, X_j, \dots, X_p)$ , la *matrice di covarianza* si generalizza come segue:

$$\text{Cov}[X] = \Sigma_X = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1j} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{j1} & \dots & \sigma_j^2 & \dots & \sigma_{jp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pj} & \dots & \sigma_p^2 \end{bmatrix}$$

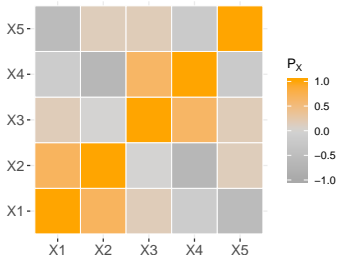
e conserva le proprietà viste per il caso bivariato.

Allo stesso modo si ottiene la *matrice di correlazione* nel caso di  $p$  variabili.

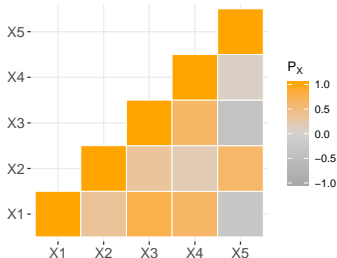
# Introduzione

## Richiami di probabilità: Covarianza e correlazione

**A**



**B**



**Grafico della matrice di correlazione:** Rappresentazione grafica di una matrice di correlazione  $P_X$  mediante gradiente dei colori:  $\rho_{ij} \approx -1$  (grigio scuro),  $\rho_{ij} \approx 0$  (grigio chiaro),  $\rho_{ij} \approx 1$  (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  $\text{diag}[P_X] = 1$ .

# Analisi fattoriale confermativa

## Definizione del modello (unidimensionale)

Dato un insieme di variabili osservate  $(X_1, \dots, X_j, \dots, X_p)$  il modello fattoriale confermativo è un *modello lineare* definito come segue:

$$X_1 = \eta\lambda_1 + \delta_1$$

$$\vdots$$

$$X_j = \eta\lambda_j + \delta_j$$

$$\vdots$$

$$X_p = \eta\lambda_p + \delta_p$$

dove:

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

# Analisi fattoriale confermativa

## Definizione del modello (unidimensionale)

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

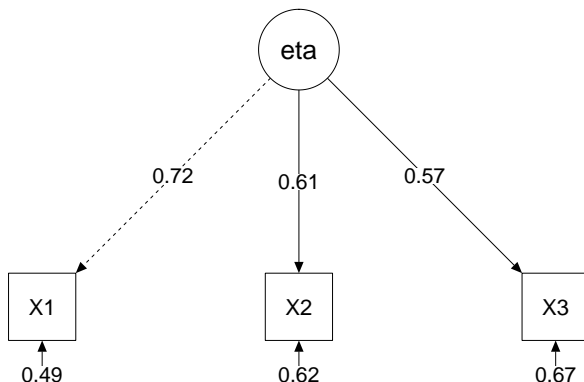
- o  $c$ : componente del modello lineare *in comune* con le  $p$  osservabili
- o  $u_j$ : componente del modello lineare specifica (*unica*) per ciascuna delle  $p$  osservabili

Le misure osservate  $X_1, \dots, X_p$  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: tale modello si dice *riflessivo*.

# Analisi fattoriale confermativa

## Definizione del modello (unidimensionale)



**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  $\lambda_{X1}$ ,  $\lambda_{X2}$ ,  $\lambda_{X3}$  mentre i numeri posti lungo le frecce orientate sui quadrati indicano i residui del modello  $\delta_{X1}$ ,  $\delta_{X2}$ ,  $\delta_{X3}$ .

# Analisi fattoriale confermativa

## Definizione del modello (unidimensionale)

Il modello fattoriale confermativo decompone la varianza osservata degli indicatori manifesti

$$\text{Var}[X_j] = \text{Var}[c] + \text{Var}[\delta_j]$$

in due parti:

(a) **comunalità**: frazione di varianza spiegata dalla variabile latente  $\eta$ :

$$h^2 = \frac{\text{Var}[c]}{\text{Var}[X_j]}$$

(b) **unicità**: frazione di varianza rimanente (*da spiegare*) dopo l'introduzione di  $\eta$ :

$$u^2 = \text{Var}[\delta_j]$$

Se le variabili osservate sono standardizzate si ha che  $h^2 + u^2 = 1$ .

# Analisi fattoriale confermativa

## Definizione del modello (multidimensionale)

Dato un insieme di variabili osservate ( $X_1, \dots, X_j, \dots, X_p$ ) il modello fattoriale confermativo ad un solo fattore latente è generalizzabile al caso con  $q$  variabili latenti ( $q < p$ ).

$$X_1 = \eta_1 \lambda_{11} + \dots + \eta_k \lambda_{k1} + \dots + \eta_q \lambda_{q1} + \delta_1$$

$$\vdots$$

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

$$\vdots$$

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

dove  $\eta_1, \dots, \eta_q$  sono i misurandi latenti del modello.



# Analisi fattoriale confermativa

## Definizione del modello (multidimensionale)

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

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

dove  $c$  in questo caso raccoglie la componente del modello lineare riferita ai fattori latenti  $\eta_1, \dots, \eta_q$ .

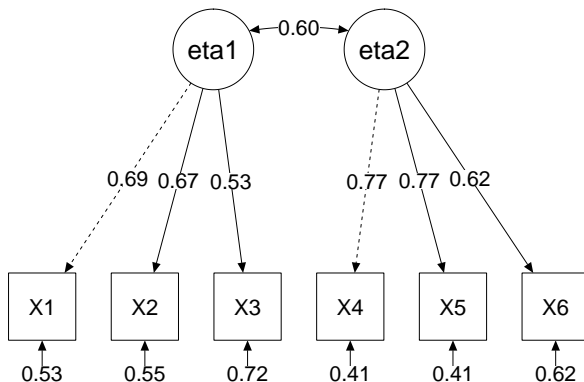
Il modello con  $q$  misurandi può essere scritto in forma compatta usando l'algebra lineare:

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

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).

# Analisi fattoriale confermativa

## Definizione del modello (multidimensionale)



**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.

# Analisi fattoriale confermativa

## Definizione del modello (multidimensionale)

### Esempio slide 18:

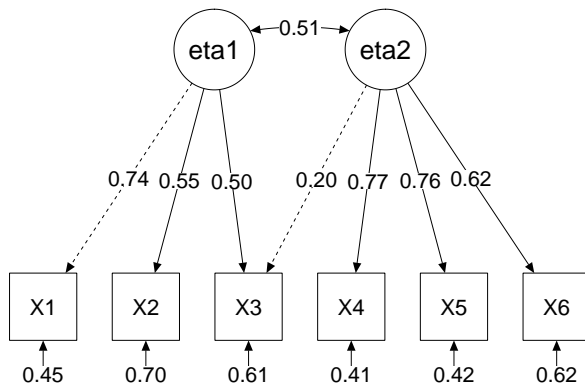
$$\mathbf{X}_{6 \times 1} = \mathbf{\Lambda}_{6 \times 2} \boldsymbol{\eta}_{2 \times 1} + \boldsymbol{\delta}_{6 \times 1}$$

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}}_{\boldsymbol{\delta}}$$

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*).

# Analisi fattoriale confermativa

## Definizione del modello (multidimensionale)



**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  $X_3$ ) 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.

# Analisi fattoriale confermativa

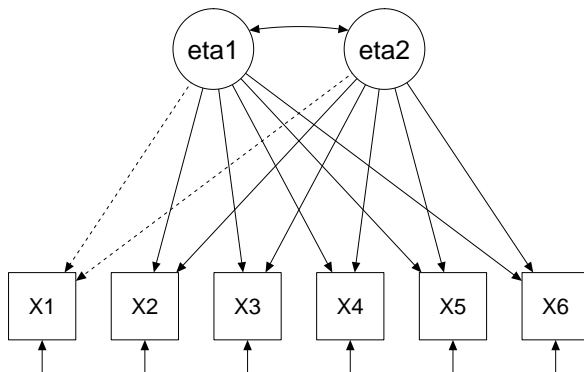
## Definizione del modello (multidimensionale)

Esempio slide 20:

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

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}}_{\boldsymbol{\delta}}$$

# Analisi fattoriale esplorativa



**Modello EFA multidimensionale:** Generalizzazione del modello confermativo al modello fattoriale esplorativo tramite manipolazione della matrice  $\Lambda$ : 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 confermativa

## Definizione del modello (multidimensionale)

### Esempio slide 22:

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

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}}_{\boldsymbol{\delta}}$$

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.

# Analisi fattoriale confermativa

## Definizione del modello (multidimensionale)

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:

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

$$\underbrace{\begin{bmatrix} X_1 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1k} & \dots & \lambda_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{j1} & \lambda_{j2} & \dots & \lambda_{jk} & \dots & \lambda_{jq} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pk} & \dots & \lambda_{pq} \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_k \\ \vdots \\ \eta_q \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_j \\ \vdots \\ \delta_p \end{bmatrix}}_{\boldsymbol{\delta}}$$

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).



# Analisi fattoriale confermativa

## Assunti del modello confermativo

### Il modello lineare dell'analisi fattoriale confermativa

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

assume quanto segue:

- (i)  $\mathbb{E}[X_j] = 0$       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

# Analisi fattoriale confermativa

## Modello confermativo e matrice di covarianza osservata

Nell'introduzione a questo modulo abbiamo affermato che la CFA utilizza quale *dato di input* la matrice di covarianza delle osservabili  $\text{Cov}[\mathbf{X}] = \mathbf{S}_X$  e non la matrice delle variabili osservate  $\mathbf{X}_{n \times p}$ .

Per comprendere, dunque, come il modello della CFA è legato a  $\mathbf{S}_X$  scriviamo la matrice di covarianza attesa del modello:

$$\begin{aligned}\text{Cov}[\mathbf{X}] &= \text{Cov}[(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta}), (\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})] && \text{usiamo la definizione del modello CFA: } \mathbf{X} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta} \\ &= \mathbb{E} \left[ (\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})^T \right] && \text{usiamo la definizione di covarianza (slide 5) notando che } \boldsymbol{\mu}_X = \mathbf{0} \\ \dots &&& \text{usando gli assunti (i)-(iv) del modello, con alcune manipolazioni algebriche otteniamo:} \\ &= \mathbf{\Lambda} (\text{Cov}[\boldsymbol{\eta}, \boldsymbol{\eta}]) \mathbf{\Lambda}^T + \text{Cov}[\boldsymbol{\delta}, \boldsymbol{\delta}] \\ &= \boxed{\mathbf{\Lambda}\boldsymbol{\Phi}\mathbf{\Lambda}^T + \boldsymbol{\Theta}_\delta}\end{aligned}$$

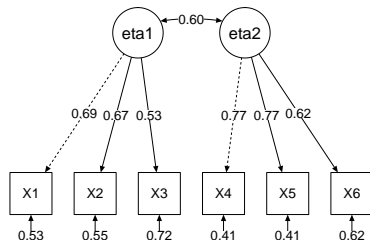
dove:

- o la matrice  $\boldsymbol{\Phi}_{q \times q}$  esprime la covarianza/correlazione tra i misurandi latenti
- o la matrice  $\boldsymbol{\Theta}_{p \times p}$  esprime la covarianza/correlazione tra gli errori

# Analisi fattoriale confermativa

## Modello confermativo e matrice di covarianza osservata

Riprendiamo l'esempio rappresentato nella slide 18:



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

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

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

# Analisi fattoriale confermativa

## Raccordo tra CFA e TCT

CFA e TCT sono fortemente interrelate. L'equazione di base della TCT è la medesima di quella della CFA così come anche alcuni degli assunti (es.: correlazione tra errori e misurando è zero). Sebbene non siano la stessa cosa, la CFA permette di costruire e valutare una scala di misura mediante un modello lineare che lega indicatori a costrutti. Per cui, secondo la CFA, *misurare significa definire e stimare un modello confermativo usando le misurazioni osservabili*.

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

<sup>†</sup> L'espressione si ottiene dalla definizione generale  $\text{Var}[X_j] = \text{Var}[\eta_j] + \text{Var}[\delta_j]$  notando dalla slide 25 che:

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

# Analisi fattoriale confermativa

## Raccordo tra CFA e TCT

Un'elegante corrispondenza tra CFA e TCT può essere rintracciata nella definizione di **attendibilità** di una scala. Mentre nel modulo C abbiamo definito la misura di attendibilità utilizzando misurazioni parallele (e sfruttando dunque l'informazione presente nei dati), nella CFA l'attendibilità è definita utilizzando il modello stesso, come segue:

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

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<sup>†</sup>:

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

<sup>†</sup> L'uguaglianza è valida notando che la definizione di attendibilità è quella definita dalla correlazione intraclasse (vedi modulo C, slide 51).

# Analisi fattoriale confermativa

## Raccordo tra CFA e TCT

Un indice di attendibilità alternativo all' $\alpha$  di Chronbach (vedi modulo C) e coerente con l'approccio CFA è l'**indice**  $\omega$ , definito come:

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

che utilizza direttamente i parametri del modello CFA (una volta determinati/stimati).

L'indice si interpreta in maniera analoga all' $\alpha$  di Chronbach sebbene, rispetto a quest'ultimo, utilizzi un modello dei dati più che i dati direttamente.

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

# Analisi fattoriale confermativa

## CFA in sintesi

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

- (i)  $\mathbb{E}[\mathbf{X}_j] = 0$
- (ii)  $\mathbb{E}[\eta_k] = 0$
- (iii)  $\mathbb{E}[\delta_j] = 0$
- (iv)  $\text{Cov}[\eta_k, \delta_j] = 0$
  
- (v)  $\text{Cov}[\mathbf{X}] = \boldsymbol{\Sigma}_X = \mathbf{\Lambda} \boldsymbol{\Phi} \mathbf{\Lambda}^T + \boldsymbol{\Theta}_\delta$
  
- (vi) comunaltà:  $\text{Var}[c_j] = \sum_{k=1}^q \lambda_{jk}^2$  caso semplice in cui non si hanno correlazione tra misure latenti
- (vii) unicità:  $\text{Var}[\delta_j] = \theta_{\delta_j}$

# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

Il modello CFA è parametrizzato secondo la matrice  $\Lambda$  (coefficienti che legano le osservabili alle latenti), la matrice  $\Phi$  (che contiene le varianze delle latenti e le covarianze quando  $q > 1$ ) e la matrice  $\Theta_\delta$  (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  $\Lambda$ ,  $\Phi$ ,  $\Theta_\delta$  nella loro interezza ma solo quelle componenti di queste che sono diverse da zero (*parametri liberi*).

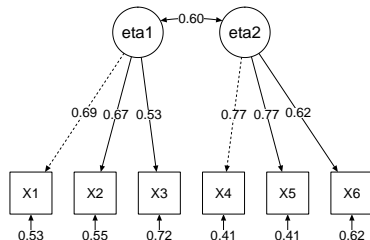
Un modello confermativo deve essere il *più parsimonioso possibile* nel senso di avere pochi parametri liberi e più parametri fissi: questo si realizza, appunto, nella definizione di un modello basato sulle ipotesi del ricercatore.



# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

Nell'esempio della slide 18 abbiamo 13 parametri liberi (da stimare):



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

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

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

# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

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 18 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  $\mathbf{X}$ : a parità di matrice di covarianza osservata  $\mathbf{S}_X$ , infatti, si preferirà il modello con più gradi di libertà, dunque quello più parsimonioso.

# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

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

- o 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 18:  $\lambda_{11} = 1$  e  $\lambda_{21}, \lambda_{31}$  si lasciano liberi di essere stimati.
- o approccio UVI (*unit variance identification*): si fissa ad 1 la varianza delle variabili latenti  $\phi_1, \dots, \phi_q$ . Nell'esempio della slide 18:  $\phi_{\eta_1} = \phi_{\eta_2} = 1$ .

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

Per semplicità nelle applicazioni fatte durante il laboratorio utilizzeremo l'approccio ULI.

# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

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

La procedura può essere implementata mediante diversi approcci, quali ad esempio quello di *massima verosimiglianza* (ML) o dei *minimi quadrati generalizzati* (GLS, DWLS, ecc).

I parametri stimati (tra quelli liberi) alla fine della procedura di stima risultano:

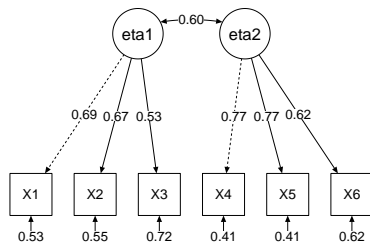
- $\hat{\Lambda}$  matrice dei coefficienti del modello fattoriale (*factor loadings*)
- $\hat{\Phi}$  matrice delle varianze/covarianze tra i fattori latenti
- $\hat{\Theta}_\delta$  matrice delle varianze/covarianze degli errori del modello

i quali vengono utilizzati - tra le altre cose - per *interpretare* il modello finale.

# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

Riprendiamo l'esempio rappresentato nella slide 18:



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

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

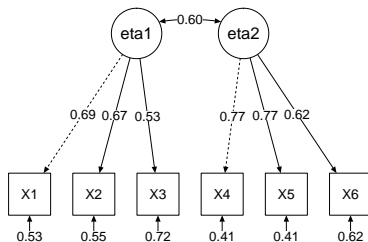
$$\hat{\Theta}_\delta = \begin{bmatrix} 0.53 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.72 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.41 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.62 \end{bmatrix}$$

Esempio di stima dei parametri del modello

# Analisi fattoriale confermativa

## Stime dei parametri del modello CFA

Riprendiamo l'esempio rappresentato nella slide 18:



$$\mathbf{S}_X = \begin{bmatrix} & X1 & X2 & X3 & X4 & X5 & X6 \\ X1 & 1.00 & & & & & \\ X2 & 0.43 & 1.00 & & & & \\ X3 & 0.41 & 0.35 & 1.00 & & & \\ X4 & 0.28 & 0.35 & 0.16 & 1.00 & & \\ X5 & 0.32 & 0.32 & 0.19 & 0.59 & 1.00 & \\ X6 & 0.24 & 0.32 & 0.18 & 0.47 & 0.46 & 1.00 \end{bmatrix}$$

$$\hat{\Sigma}_X = \begin{bmatrix} & X1 & X2 & X3 & X4 & X5 & X6 \\ X1 & 1.00 & & & & & \\ X2 & 0.46 & 1.00 & & & & \\ X3 & 0.36 & 0.36 & 1.00 & & & \\ X4 & 0.31 & 0.31 & 0.24 & 1.00 & & \\ X5 & 0.31 & 0.31 & 0.24 & 0.59 & 1.00 & \\ X6 & 0.25 & 0.25 & 0.19 & 0.47 & 0.47 & 1.00 \end{bmatrix}$$

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

# Analisi fattoriale confermativa

## Valutazione del modello CFA

### Test del $\chi^2$ (chi-quadrato)

Si valuta una statistica  $T = f(\mathbf{S}_X, \Sigma_X)$  che calcola la differenza tra  $\mathbf{S}_X$  e  $\Sigma_X$  e sotto l'ipotesi nulla si distribuisce in probabilità secondo la legge  $\chi^2(df)$ , dove  $df$  sono i gradi di libertà del modello (vedi slide 34). In questo caso, l'ipotesi nulla codifica l'assenza di differenza tra  $\mathbf{S}_X$  e  $\Sigma_X$ . Come al solito, se  $T_{\text{oss}} > \chi^2_{\alpha/2}$  (valore critico della statistica test) allora l'ipotesi nulla è rigettata e dunque il modello teorico implicato da  $\Sigma_X$  non si adatta bene ai dati osservati  $\mathbf{S}_X$ .

Come ogni decisione basata sulla statistica  $\chi^2$ , quando la numerosità campionaria  $n$  è molto elevata il test di ipotesi tende a rigettare spesso il modello teorico  $\Sigma_X$  (ossia  $T_{\text{oss}} > \chi^2_{\alpha/2}$ ). Per questo motivo si preferisce utilizzare altri indici per valutare il modello.

# Analisi fattoriale confermativa

## Valutazione del modello CFA

### RMSEA

Valuta in maniera assoluta il grado di non adattamento del modello ai dati:

$$\text{RMSEA} = \sqrt{\frac{T_{ndf}}{ndf}}$$

In generale,  $\text{RMSEA} < 0.05$  indica adattamento ottimo del modello ai dati,  $0.05 \leq \text{RMSEA} < 0.1$  indica adattamento medio/scarso, mentre  $\text{RMSEA} > 0.10$  indica adattamento non adeguato del modello.



# Analisi fattoriale confermativa

## Valutazione del modello CFA

### 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.

# Analisi fattoriale confermativa

## Valutazione del modello CFA

### 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 e modello nullo (oppure modelli alternativi) è AIC (oppure BIC). Il modello che meglio si adatta ai dati è dunque quello che ottiene AIC più basso (*the lower, the better*).

# Analisi fattoriale confermativa

## Valutazione del modello CFA

### Indici di modifica

Permettono di valutare come la statistica  $T_n$  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.

# Analisi fattoriale confermativa

## Valutazione del modello CFA

### 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  $\delta$  elevati.

L'analisi dei residui si basa sull'osservazione della matrice  $\mathbf{S}_X - \hat{\Sigma}_X$  alla ricerca di quei coefficienti  $\theta_\delta$  che, rispetto a due variabili fissate, risultano elevati. Una volta individuato il residuo, si “libera” un parametro della matrice  $\Theta_\delta$  del modello e si ri-stima 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.

# Analisi fattoriale confermativa

## Correlazione tra gli errori di misura

La matrice degli errori  $\Theta_{\delta}$  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.

# Analisi fattoriale confermativa

## Correlazione tra gli errori di misura

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

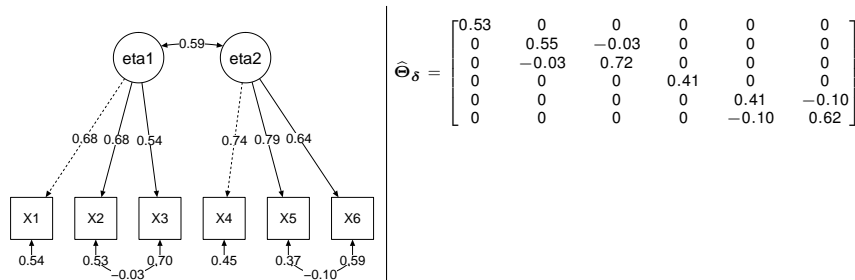
Al contrario, liberare un parametro della matrice - ad esempio  $\theta_{\delta_{2j}}$  - equivale a correlare gli errori  $\delta$  della variabile osservata 2 e  $j$  e ciò equivale a dire che le due osservabili esprimono, in qualche modo, una medesima informazione sul misurando  $\eta$  (e quindi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

# Analisi fattoriale confermativa

## Correlazione tra gli errori di misura

Riprendiamo l'esempio rappresentato nella slide 18:



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

# Analisi fattoriale confermativa

## 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)



# Analisi fattoriale confermativa

## 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 la dispensa della prof.ssa Agnoli (corso Testing, matr. Pari) disponibile nella cartella APPROFONDIMENTI.

# Analisi fattoriale confermativa

## 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<sup>(1)</sup> ( $n = 2200$ ):

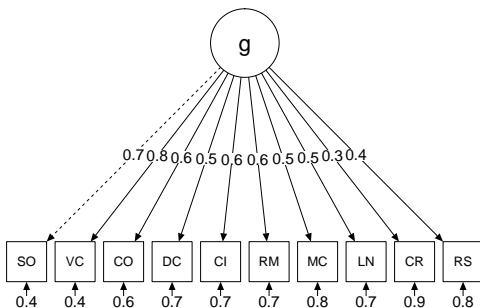
$$\mathbf{S}_X = \begin{bmatrix} DC & SO & MC & CI & CR & VC & LN & RM & CO & RS \\ 1.00 & & & & & & & & & \\ 0.38 & 1.00 & & & & & & & & \\ 0.26 & 0.35 & 1.00 & & & & & & & \\ 0.34 & 0.43 & 0.28 & 1.00 & & & & & & \\ 0.25 & 0.14 & 0.15 & 0.11 & 1.00 & & & & & \\ 0.33 & 0.62 & 0.33 & 0.41 & 0.13 & 1.00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

<sup>(1)</sup> 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.

# Analisi fattoriale confermativa

## Studio della dimensionalità della WISC-IV

### M1: Scala $g$



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

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

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

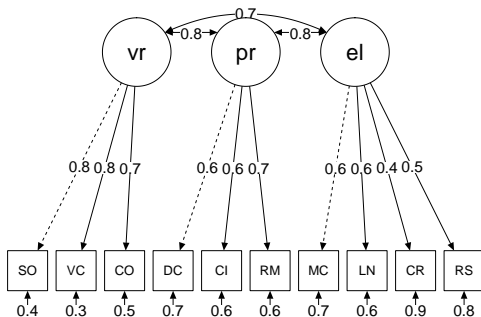
RMSEA = 0.11, CFI = 0.84, AIC = 57359

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

# Analisi fattoriale confermativa

## Studio della dimensionalità della WISC-IV

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



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

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

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

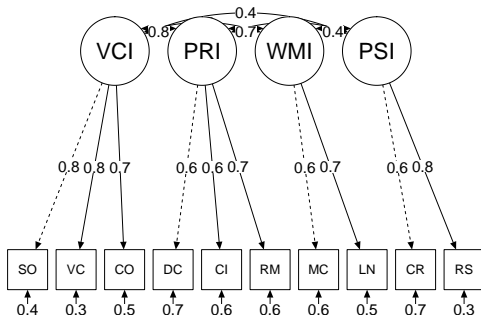
RMSEA = 0.08, CFI = 0.91, AIC = 56904

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

# Analisi fattoriale confermativa

## Studio della dimensionalità della WISC-IV

### M3: Scala VCI, PRI, WMI, PSI



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

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

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

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

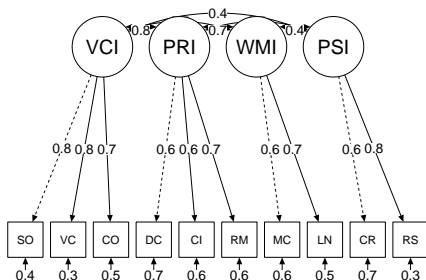
RMSEA = 0.04, CFI = 0.97, AIC = 56548

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

# Analisi fattoriale confermativa

## 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
...	...	...	...	...

$$\mathbf{Z}_X = (\mathbf{S}_X - \Sigma_X) =$$

-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

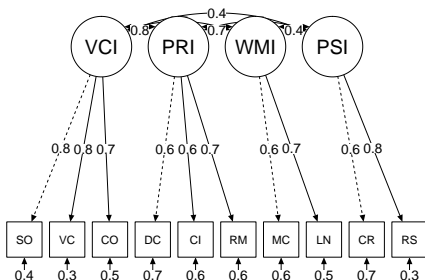
RMSEA = 0.04, CFI = 0.97, AIC = 56548

Nota:  $\mathbf{Z}_X$  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 `lavaan`).

# Analisi fattoriale confermativa

## Studio della dimensionalità della WISC-IV

### M3: Scala VCI, PRI, WMI, PSI



Indici di attendibilità

	VCI	PRI	WMI	PSI	total
$\alpha$	0.80	0.66	0.59	0.63	0.82
$\omega$	0.80	0.67	0.59	0.65	0.86

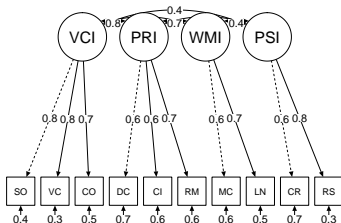
Gli indici sono calcolati separatamente per ciascuna scala della WISC-IV e complessivamente (*total*).

RMSEA = 0.04, CFI = 0.97, AIC = 56548

# Analisi fattoriale confermativa

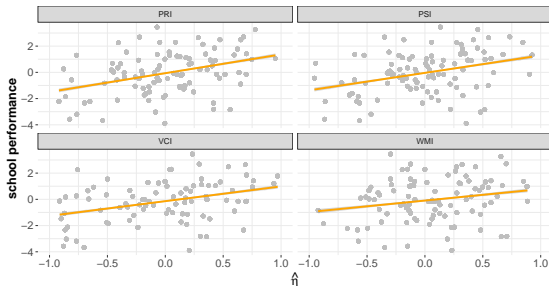
## Studio della dimensionalità della WISC-IV

### 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:



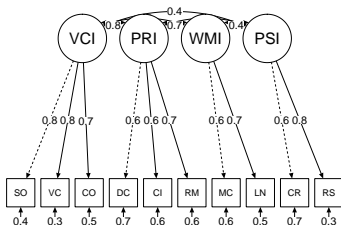
RMSEA = 0.04, CFI = 0.97, AIC = 56548



# Analisi fattoriale confermativa

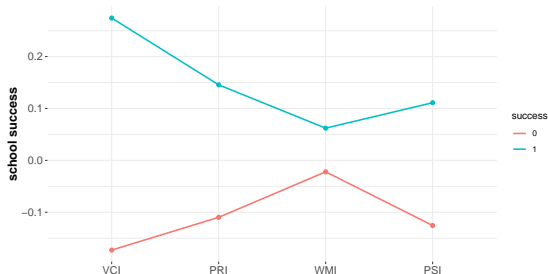
## Studio della dimensionalità della WISC-IV

### 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).



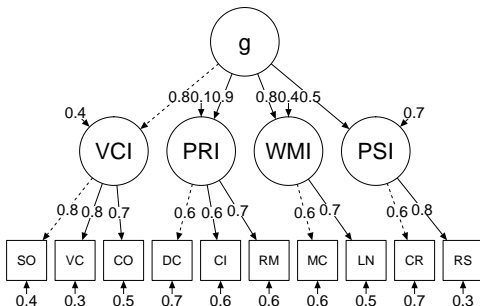
Notiamo come individui con successo scolastico presentino profili WISC-IV differenti da coloro con basso successo scolastico.

RMSEA = 0.04, CFI = 0.97, AIC = 56548

# Analisi fattoriale confermativa

## Studio della dimensionalità della WISC-IV

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



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

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

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

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

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*.

# Analisi fattoriale confermativa

## Suggerimenti per la costruzione 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  $\hat{\Sigma}_X$  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

# Analisi fattoriale confermativa

## Risorse utili e sitografia minima: CFA con lavaan

- o Sito web di riferimento per la libreria lavaan:

<http://lavaan.ugent.be/>

- o Tutorial sull'utilizzo della libreria lavaan:

<http://lavaan.ugent.be/tutorial/tutorial.pdf>

- o Corso su CFA (Weeks 1-4) con lavaan:

<http://sachaepskamp.com/SEM2017>

- o CFA e SEMs con lavaan:

[https://personality-project.org/r/tutorials/summerschool.14/rosseel\\_sem\\_intro.pdf](https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_intro.pdf)

- o CFA in a nutshell with lavaan:

[https://www.yourpersonality.net/R/R\\_Notes\\_3.html](https://www.yourpersonality.net/R/R_Notes_3.html)

- o lavaan cheat sheet:

<https://github.com/jeromyanglim/lavaan-examples/blob/master/cheat-sheet-lavaan/cheat-sheet-lavaan.md>

- o lavaan google group (forum):

<https://groups.google.com/forum/#!forum/lavaan>

# 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:

- o confronto tra gruppi in analisi successive  
(es.: gruppi hanno medesima struttura di intelligenza)
- o analisi delle dimensioni cognitive tra gruppi  
(es.: bambini tipici e atipici presentano medesima componente VCI)
- o 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:  $\chi^2$ , RMSEA, CFI, AIC.

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

- (i) configurale
- (ii) debole
- (iii) forte
- (iv) 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  $\mathbf{S}_X^{(1)}$  e  $\mathbf{S}_X^{(2)}$ . Dati due generici e distinti livelli di invarianza, la procedura procede come segue:

- Si adattano due modelli CFA,  $\mathcal{M}_1$  su  $\mathbf{S}_X^{(1)}$  e  $\mathcal{M}_2$  su  $\mathbf{S}_X^{(2)}$ , per il primo livello di invarianza ottenendo i parametri CFA stimati per entrambi i modelli adattati
- Si calcola la statistica complessiva  $\chi_0^2 = \chi_{\mathcal{M}_1}^2 + \chi_{\mathcal{M}_2}^2$  e i gradi di libertà  $df_0$  (vedi slide 34)
- Si adattano due modelli CFA,  $\mathcal{M}'_1$  su  $\mathbf{S}_X^{(1)}$  e  $\mathcal{M}'_2$  su  $\mathbf{S}_X^{(2)}$ , per il secondo livello di invarianza ottenendo i parametri CFA stimati per entrambi i modelli adattati
- Si calcola la statistica complessiva  $\chi_1^2 = \chi_{\mathcal{M}'_1}^2 + \chi_{\mathcal{M}'_2}^2$  e i gradi di libertà  $df_1$
- Si valuta la statistica  $(\chi_1^2 - \chi_0^2)$  mediante il test del  $\chi^2$  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  $\chi_{\alpha}^2$  calcolato rispetto all' $\alpha$  scelto, il test rigetta l'ipotesi di equivalenza tra i due modelli.

In generale:

- o Se  $H_0$  è rigettata la procedura di valutazione dell'invarianza si arresta
- o 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<sup>†</sup>:

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

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

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).

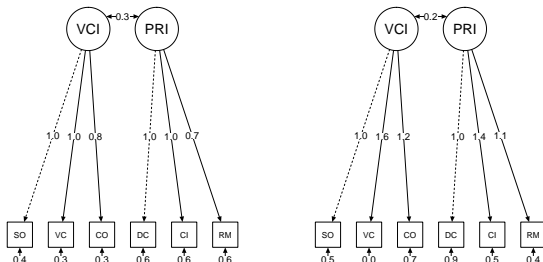
Nota: rispetto al modello precedente, abbiamo aggiunto il termine  $\boldsymbol{\tau}_{p \times 1}$  che rappresenta le intercette del modello per ciascuna variabile osservata. Queste hanno utilità quando, ad esempio, gli item (variabili osservate) hanno diversa media (es.: difficoltà diversa) o quando,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

---

<sup>†</sup> I pedici indicanti le dimensioni delle matrici sono stati omessi per semplicità sebbene questi siano gli stessi di quelli presentati nel modello delle slide 16-17.

# Invarianza di misurazione

## Invarianza configurale



$\chi^2$	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:

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

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

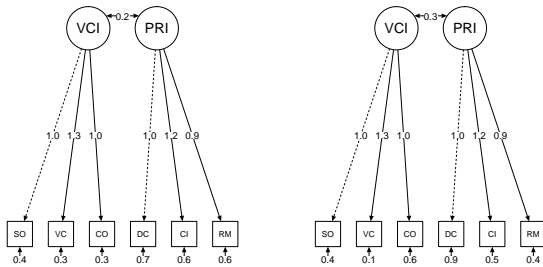
dove ciascun gruppo ha un proprio modello fattoriale sotto il vincolo:

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

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  $\Lambda$ .

# Invarianza di misurazione

## Invarianza forte

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

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

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

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

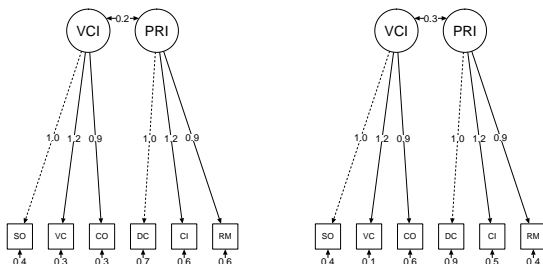
$$\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$$

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

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 items 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 items 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  $\Lambda$ , tendono a rispondere allo stesso modo agli items (questi hanno lo stesso significato in entrambi i gruppi) e non vi è bias indotto dalle differenze degli items  $\tau$  nei due gruppi.



# Invarianza di misurazione

## Invarianza esatta

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

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

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

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

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

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

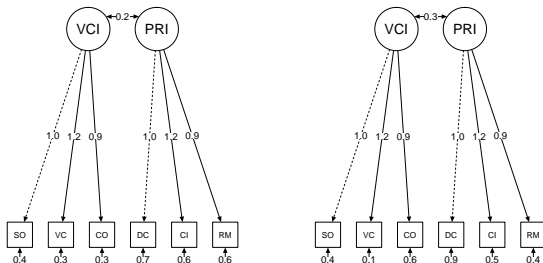
$$\boldsymbol{\delta}^{(1)} = \boldsymbol{\delta}^{(2)}$$

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 items non producono bias nei gruppi e i due gruppi possono essere confrontati in termini di differenze tra costrutti latenti.

Nota: questo tipo di invarianza è piuttosto rara nella pratica di ricerca.

# 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  $\Lambda$ , non vi è bias indotto dalle differenze degli items  $\tau$  nei due gruppi ed inoltre i residui del modello  $\delta$  sono gli stessi. I due gruppi possono essere confrontati tra loro in termini di punteggi fattoriali  $\hat{\eta}^{(1)}$  e  $\hat{\eta}^{(2)}$ .

## Modelli CFA di invarianza in sintesi

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

- |  |                        |
|--|------------------------|
| (i) $\mathbf{X}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ | invarianza configurale |
| (ii) $\mathbf{X}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$      | invarianza debole      |
| (iii) $\mathbf{X}^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$           | invarianza forte       |
| (iv) $\mathbf{X}^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}$                  | invarianza esatta      |

# Invarianza di misurazione

Risorse utili e sitografia minima: invarianza con lavaan

- o Tutorial sull'invarianza con lavaan:

<https://tinyurl.com/yy3rutpz>

- o Breve talk sull'invarianza con lavaan:

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

- o Ebook sull'invarianza pubblicato su *Frontiers in Psychology* (free):

<https://www.frontiersin.org/research-topics/1695/measurement-invariance>