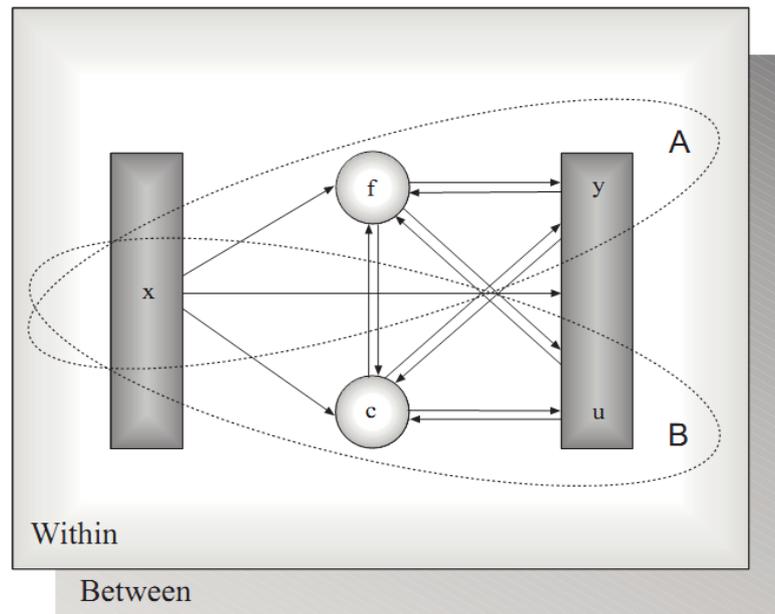


# MODELLI DI EQUAZIONI STRUTTURALI

## *Mplus*





## ELEMENTI DI BASE

- **Cosa sono i Modelli di Equazioni Strutturali (SEM)**
- **Le componenti dei SEM**
- **Ipotesi di base e modelli matematici**
- **Fasi dei SEM**
- **Condizioni di applicabilità**



## I MODELLI DI EQUAZIONI STRUTTURALI

**I Modelli di Equazioni Strutturali (SEM, Structural Equation Modeling) rappresentano una classe di modelli statistici che permettono di esprimere in maniera semplificata e formalizzata le relazioni tra i costrutti considerati in una determinata teoria (o in parti di essa).**

**I SEM consentono di esaminare se un modello in cui vengono ipotizzate determinate relazioni tra un insieme di variabili *è consistente con i dati empirici.***



# I MODELLI DI EQUAZIONI STRUTTURALI

**Il punto di partenza** é rappresentato da una matrice di varianze/covarianze che riassume le relazioni tra le variabili osservate.

4 .02

1 .17 3 .68

1 .01 1 .26 3 .30

1 .47 1 .53 1 .37 3 .99

.13 .55 .39 .49 3 .16

.28 .20 .46 .25 .91 2 .57

.50 .24 .49 .65 .90 .68 2 .81

.22 .31 .57 .18 .76 .81 .73 2 .20

.01 -.02 .02 -.01 .17 .07 .13 .13 .15

.15 .09 .00 .06 -.05 .03 -.08 -.12 -.03 .16

# I MODELLI DI EQUAZIONI STRUTTURALI

**Il punto di arrivo é rappresentato da:**

- a) un **insieme di parametri** che quantificano le relazioni specificate nel modello  $(\lambda, \theta, \beta, \psi)$ ;
- b) una **statistica** associata ad ognuno di questi parametri che consente di esaminarne la significatività statistica  $(t, z)$ ;
- c) una **matrice delle varianze/covarianze** tra le variabili osservate del modello **riprodotta** tramite i parametri del modello  $(S^{\wedge}, \Sigma(\theta^{\wedge}))$ .
- d) uno o più **indici** che misurano la **bontà dell'adattamento** del modello ai dati, cioè la corrispondenza del modello con i dati osservati  $(\chi^2, RMSEA, SRMR, CFI, ecc.)$ ;



## Rappresentazioni e Formalizzazioni nei SEM

Le relazioni tra le variabili in un modello di equazioni strutturali possono essere rappresentate in 4 modi differenti:

a) Una descrizione **verbale**

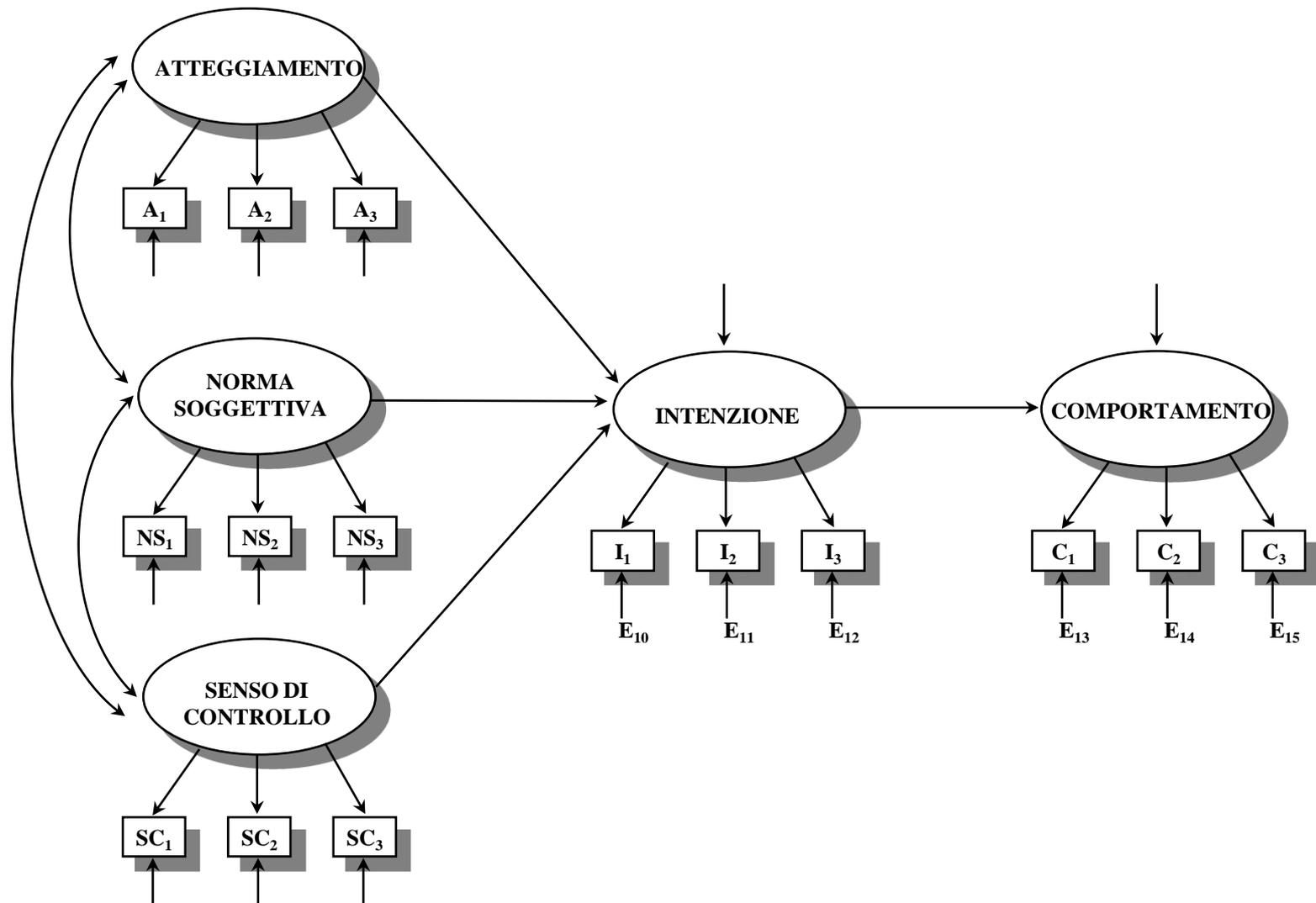
b) Una rappresentazione **grafica** (path diagram)

c) Una formalizzazione **matematica** dove le variabili *dipendenti* sono espresse come *equazioni algebriche lineari*, ovvero combinazioni lineari di altre variabili incluse nel modello, es.:  $y_1 = \lambda_{11}\eta_1 + \varepsilon_1$

d) Un insieme di **comandi** scritti nella sintassi di un linguaggio di programmazione (es., f1 by y1 y2 y3;)

**Queste quattro rappresentazioni DEVONO coincidere**

# Formalizzazione diagrammatica di un modello strutturale



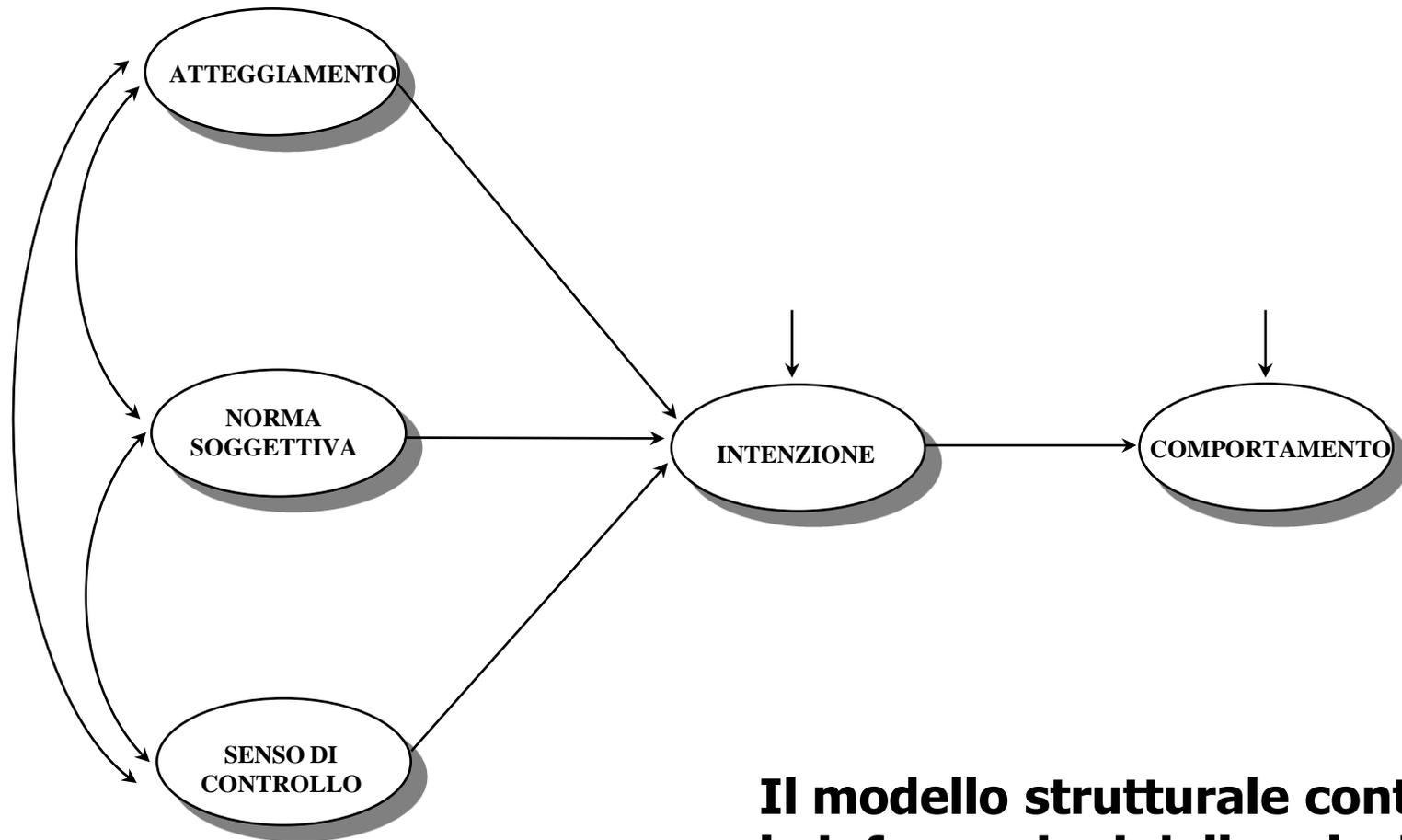


## I Modelli di Equazioni Strutturali prevedono:

- una serie di equazioni in cui vengono specificate le relazioni tra i costrutti chiave della teoria (definibile **modello strutturale**);
  - una serie di equazioni che specificano le relazioni tra le variabili latenti e le variabili osservate (definibile **modello di misura**).
- \* la presenza soltanto del modello di misurazione definisce un modello di *analisi fattoriale confermativa* (***Confirmatory Factor Analysis, CFA***);
- \* la presenza soltanto del modello strutturale, definisce un modello di ***path analysis*** su variabili osservate.

# L'analisi strutturale

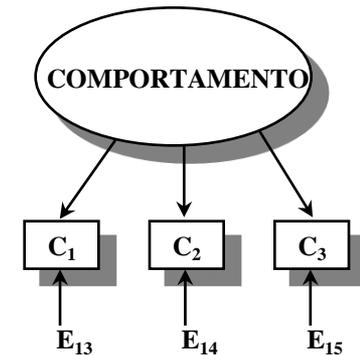
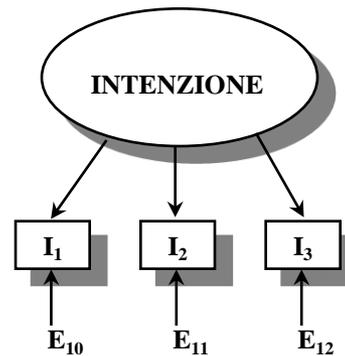
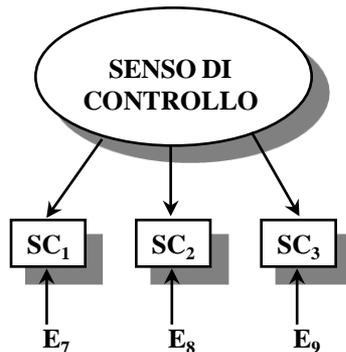
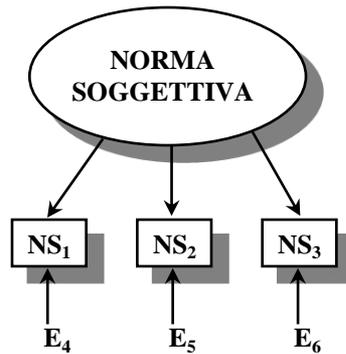
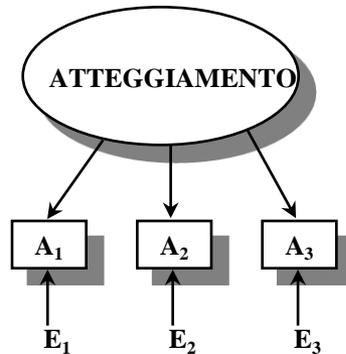
## Il modello strutturale



**Il modello strutturale contiene le informazioni delle relazioni tra i costrutti**

# L'analisi strutturale

## Il modello di misura



**Il modello di misura descrive le relazioni tra le variabili osservate e i costrutti (o variabili latenti)**



## Nei modelli di Equazioni strutturali abbiamo:

### \* **Tipi differenti di variabili**

**Indipendenti (Esogene)/Dipendenti (Endogene)**  
**Latenti/Misurate**

### \* **Tipi differenti di relazioni tra le variabili**

**Associazione (covarianza/relazione simmetrica)**

**Effetto diretto (influenza diretta/relazione asimmetrica)**

**Effetto indiretto (influenza indiretta)**



## Nei modelli di Equazioni strutturali abbiamo:

- \* **Tipi differenti di residui associati alle variabili dipendenti:**

**Variabili misurate:** errore di misurazione

**Variabili latenti:** errore di specificazione

- \* **Tipi differenti di parametri:**

**Liberi:** parametri per cui si calcola una stima

**Fissi:** parametri il cui valore è stato fissato (di solito a 0 o a 1)

**Vincolati:** parametri liberi le cui stime sono vincolate ad assumere soltanto certi valori (ad esempio, 2 parametri vincolati ad assumere lo stesso valore)



## **I parametri di un modello di equazioni strutturali**

**Sono i termini delle equazioni per i quali viene prodotta una stima nella soluzione:**

- a) le relazioni di influenza tra le variabili (ovvero gli "effetti diretti" di una variabile su un'altra variabile)**
- b) le varianze e le covarianze delle variabili indipendenti**
- c) le varianze e le covarianze dei residui**



## I parametri di un modello di equazioni strutturali

- \* I parametri che quantificano l'influenza diretta sono i coefficienti strutturali (pesi beta)
- \* I parametri che quantificano l'associazione non direzionale tra le variabili sono i coefficienti di covarianza
- \* Le varianze e le covarianze delle variabili dipendenti **non sono** parametri del modello ma vengono spiegate dal modello



## I parametri di un modello di equazioni strutturali – i coefficienti strutturali:

- \* Corrispondono ai pesi beta (*beta weights*) della regressione
- \* Esprimono la quantità di variazione che ci si attende nella variabile dipendente in concomitanza di un cambiamento di una unità nelle variabili indipendenti ad essi associate, *mantenendo costanti* le altre variabili
- \* Sono i pesi che misurano l'influenza parziale di una VI su una VD, al netto delle correlazioni della VI con le altre variabili in analisi.



## I parametri di un modello di equazioni strutturali – i residui delle variabili dipendenti **osservate**:

**Riflettono diverse componenti:**

- componente **stocastica**: discrepanza tra dati campionari e dati della popolazione;
- errore di **misurazione**: le variabili non sono misurate perfettamente;
- componente di **specificità** della variabile: varianza unica sistemica ma non condivisa con le altre variabili;
- errore di **specificazione**: modello fattoriale inadeguato (troppi pochi fattori);
- errore di **specificazione**: forma della relazione diversa da quella lineare.



## **I parametri di un modello di equazioni strutturali – i residui delle variabili dipendenti **latenti**:**

**Riflettono soprattutto l'errore di **specificazione**:**

- predittori importanti della variabile esclusi dal modello;**
- predittori irrilevanti inclusi nel modello;**
- forma della relazione diversa da quella lineare.**



# FORMALIZZAZIONE MATEMATICA: MODELLI MATEMATICI NEI SEM

**Esistono diversi modelli matematici che definiscono come specificare le relazioni tra le variabili.**

**\* Muthén**

$$y = \Lambda\eta + \varepsilon$$
$$\eta = B\eta + \zeta$$

**\* Jöreskog, Keesling & Wiley**

$$y = \Lambda_y\eta + \varepsilon$$
$$x = \Lambda_x\xi + \delta$$
$$\eta = B\eta + \Gamma\xi + \zeta$$

**\* Bentler & Weeks**

$$\eta = B\eta + \Gamma\xi$$



## Modello di Muthén

**Il modello generale di equazioni strutturali di Muthén è rappresentabile con le seguenti equazioni di base:**

$$\mathbf{y} = \mathbf{v} + \mathbf{A}\eta + [\mathbf{K}\mathbf{x} +] \varepsilon \quad (\mathbf{a})$$

$$\eta = \alpha + \mathbf{B}\eta + [\mathbf{\Gamma}\mathbf{x} +] \zeta \quad (\mathbf{b})$$

**Queste due equazioni comprendono sei matrici di parametri che definiscono un modello completo MPLUS.**



# Modello di Muthén

## Le variabili del modello

- 1.  $y$  (ipsilon):** sono le variabili osservate che rappresentano gli indicatori delle variabili ( $\eta$ ) nell'equazione relativa al modello di misura.
- 2.  $\eta$  (eta):** sono le variabili latenti misurate dalle  $y$  nell'equazione relativa al modello di misura. Possono essere sia indipendenti che dipendenti nel modello strutturale.
- 3.  $\varepsilon$  (epsilon):** sono i termini residuali associati alle variabili  $y$ , e non sono correlati con nessuna altra variabile del modello.
- 4.  $\zeta$  (zeta):** sono i residui o termini di disturbo associati alle variabili latenti  $\eta$  che risultano dipendenti.
- 5.  $x$  (ics):** sono le variabili indipendenti osservate.

## Le matrici del modello di Muthén

- 1.  $\Lambda$  (Lambda):** matrice dei coefficienti di regressione per esprimere le variabili  $y$  come funzione delle variabili latenti (le  $\eta$ ) nell'equazione relativa al modello di misura.
- 2.  $\Theta_\varepsilon$  (Theta) =  $E(\varepsilon\varepsilon')$ :** matrice delle varianze e covarianze dei termini di errore  $\varepsilon$  (epsilon), associati alle variabili  $y$ .
- 3.  $\nu$  (nu):** vettore di intercette delle variabili osservate dipendenti  $y$ .
- 4.  $\Psi$  (Psi):** matrice che contiene le varianze e le covarianze delle variabili latenti indipendenti,  $E(\eta\eta')$ , e dei termini di disturbo  $\zeta$  associati alle variabili latenti dipendenti,  $E(\zeta\zeta')$ .
- 5.  $\alpha$  (alfa):** vettore di intercette delle variabili latenti  $\eta$ .

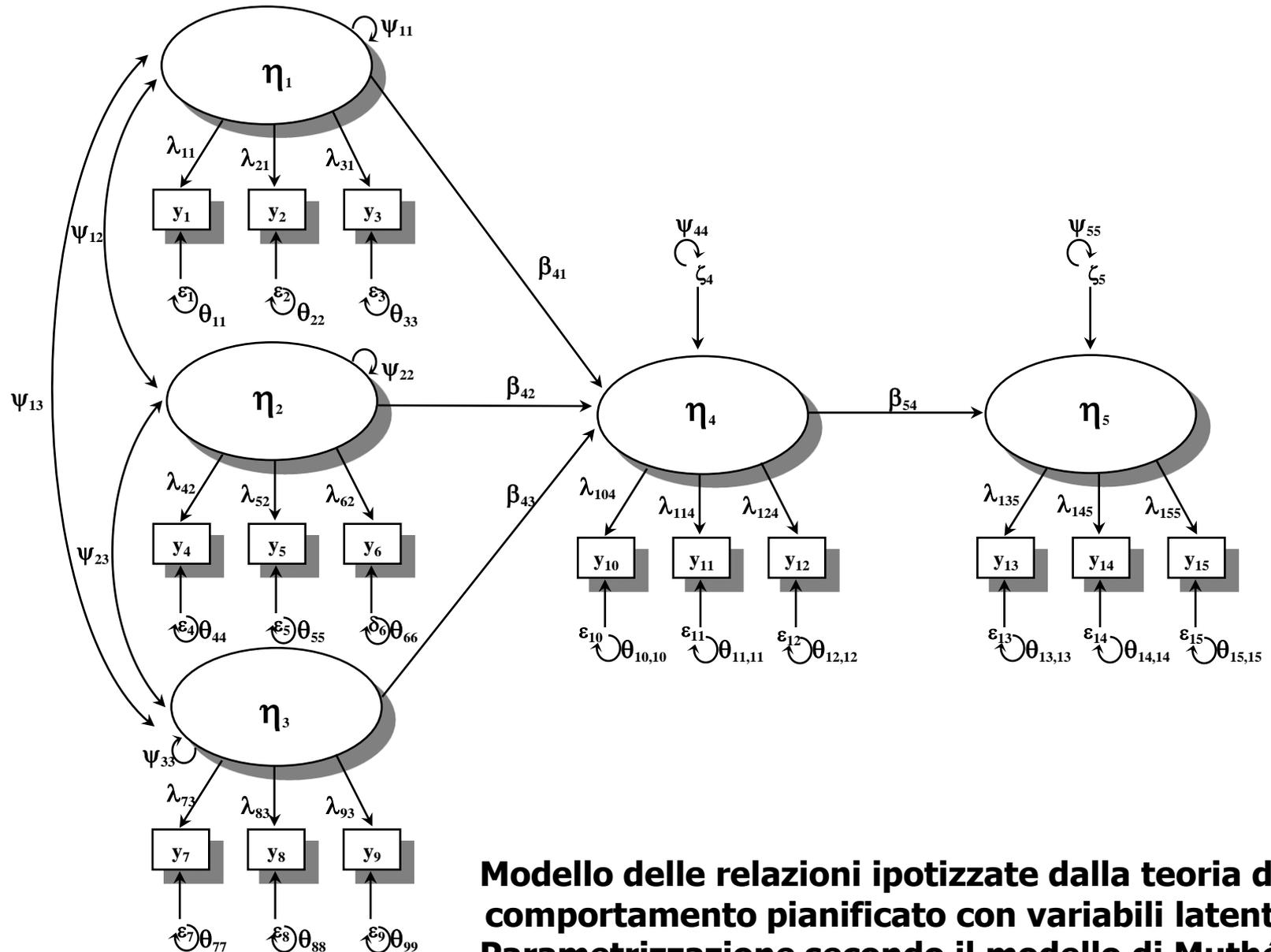
## Le matrici del modello di Muthén

**6. B (Beta):** quando tutte le **variabili osservate** sono **continue** la matrice Beta contiene i seguenti parametri:

- i coefficienti di regressione ( $\beta$ , beta) per predire le variabili latenti ( $\eta$ ) dalle  $\eta$  stesse
- i coefficienti di regressione ( $\gamma$ , gamma) per predire le variabili latenti *dipendenti* ( $\eta$ ) dalle variabili *indipendenti* osservate ( $x$ )
- i coefficienti di regressione ( $\kappa$ , kappa) per predire le variabili osservate dipendenti  $y$  dalle variabili *indipendenti* osservate ( $x$ )

**Indipendentemente dall'etichetta utilizzata, tutti questi parametri sono elementi della matrice beta, quindi sono etichettati con tale lettera nel modello di Muthén.**

# Rappresentazione grafica del modello di Muthén



**Modello delle relazioni ipotizzate dalla teoria del comportamento pianificato con variabili latenti: Parametrizzazione secondo il modello di Muthén**



$$\mathbf{y} = \mathbf{\Lambda} \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad \textcircled{H}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_6 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 & 0 \\ 0 & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{73} & 0 & 0 \\ 0 & 0 & \lambda_{83} & 0 & 0 \\ 0 & 0 & \lambda_{93} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{104} & 0 \\ 0 & 0 & 0 & \lambda_{114} & 0 \\ 0 & 0 & 0 & \lambda_{124} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{135} \\ 0 & 0 & 0 & 0 & \lambda_{145} \\ 0 & 0 & 0 & 0 & \lambda_{155} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \end{bmatrix}$$

$$\begin{bmatrix} \theta_{11}^\varepsilon \\ 0 & \theta_{22}^\varepsilon \\ 0 & 0 & \theta_{33}^\varepsilon \\ 0 & 0 & 0 & \theta_{44}^\varepsilon \\ 0 & 0 & 0 & 0 & \theta_{55}^\varepsilon \\ 0 & 0 & 0 & 0 & 0 & \theta_{66}^\varepsilon \\ \dots \end{bmatrix}$$

Modello di  
misura



$$\eta = B \eta + \zeta$$

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{41} & \beta_{42} & \beta_{43} & 0 & 0 \\ 0 & 0 & 0 & \beta_{54} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \end{bmatrix}$$

$$\begin{bmatrix} \psi_{11} & & & & \\ \psi_{21} & \psi_{22} & & & \\ \psi_{31} & \psi_{32} & \psi_{33} & & \\ 0 & 0 & 0 & \psi_{44} & \\ 0 & 0 & 0 & 0 & \psi_{55} \end{bmatrix} \Psi$$

Modello  
strutturale



## Parametri del modello di Muthén

**I parametri di un modello di equazioni strutturali sono:**

**Gli effetti diretti di una variabile su un'altra ( $B, \Lambda$ )**

**Le varianze delle variabili indipendenti e le covarianze tra le variabili indipendenti ( $\Psi$ )**

**Le varianze dei residui e le covarianze tra i residui ( $\Psi, \Theta$ )**



## L'ipotesi delle Strutture di Covarianza nei SEM

**Un modello di equazioni strutturali va sottoposto a verifica confrontandolo con i dati osservati.**

**Dai parametri del modello è possibile “ricostruire” la matrice delle varianze/covarianze tra le variabili osservate.**

**Questa corrispondenza consente di valutare l'adeguatezza del modello teorico, che definisce i parametri, rispetto ai dati osservati (cioè la bontà dell'adattamento del modello ai dati).**



## L'ipotesi delle Strutture di Covarianza nei SEM

**L'adeguatezza del modello rispetto ai dati viene valutata tramite la seguente ipotesi formale:**

$$\Sigma = \Sigma(\theta)$$

**Secondo questa equivalenza, è possibile definire un modello che specifica le relazioni tra le variabili del modello, in modo che i parametri ( $\theta$ ) del modello consentano di esprimere/ricostruire la matrice di covarianze  $\Sigma$  tra le variabili osservate.**

## L'ipotesi delle Strutture di Covarianza nei SEM

**Poiché gli elementi di  $\Sigma$  possono essere espressi come funzioni dei parametri del modello, questa ipotesi consente di valutare quanto il modello è consistente con i dati empirici.**

**Se il modello è corretto e i parametri sono noti la matrice viene riprodotta esattamente.**

**$\Sigma = \Sigma(\theta)$  rappresenta l'ipotesi nulla da verificare con i dati campionari, attraverso le stime dei parametri  $\theta$ .**

**Poiché sotto determinate assunzioni  $S$  e  $\Sigma(\hat{\theta})$  sono stimatori consistenti rispettivamente di  $\Sigma$  e  $\Sigma(\theta)$ , l'ipotesi nulla verrà accettata se  $S = \Sigma(\hat{\theta})$ .**



## Parametri del modello e matrice $\Sigma$

I parametri sono messi in relazione con la matrice di varianze e covarianze tra le variabili osservate  $S$ , tramite la seguente espressione matriciale:

$$\Sigma = \Lambda (I - B)^{-1} \Psi (I - B)^{-1'} \Lambda' + \Theta$$

che si semplifica nei casi seguenti:

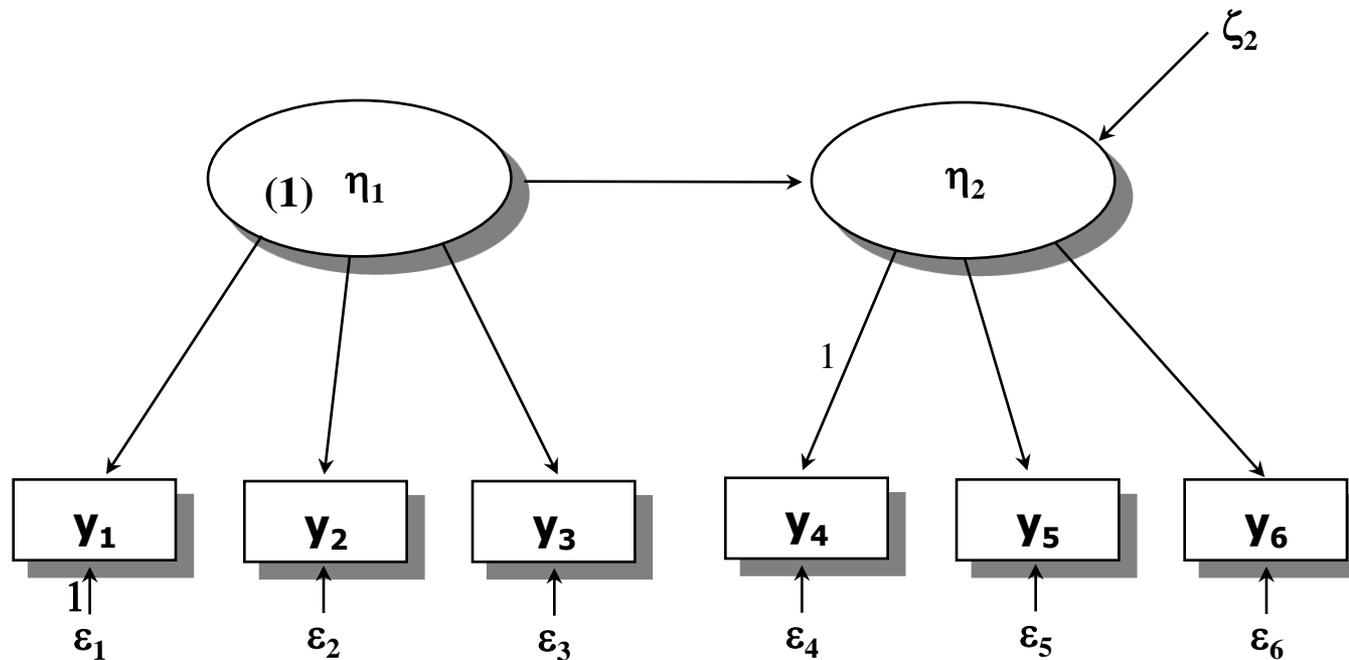
**CFA con variabili osservate continue**

$$\Sigma = \Lambda \Psi \Lambda' + \Theta$$

**Path analysis con variabili osservate continue**

$$\Sigma = (I - B)^{-1} \Psi (I - B)^{-1'}$$

## ESERCIZIO 1: TIPI DI VARIABILI/ RELAZIONI/ ERRORI - CALCOLO DEL NUMERO DEI PARAMETRI



**Quante sono le VI?**

**Quante sono le variabili misurate?**

**Quanti sono gli effetti diretti?**

**Quante sono le VD?**

**Quante sono le variabili latenti?**

**Ci sono delle covarianze tra variabili?**

**Quanti sono gli errori di specificazione e quanti quelli di misurazione ?**

**Quanti sono i parametri del modello che vengono stimati ?**



## Fasi dei modelli di equazioni strutturali

- a) specificazione (formulazione) del modello**
- b) identificazione del modello e dei suoi parametri**
- c) stima dei parametri del modello**
- d) valutazione del modello**
- e) modifica del modello**



## Specificazione del modello

- a) definire le variabili latenti e osservate che compongono le diverse equazioni;**
- b) definire quali variabili saranno indipendenti (o "esogene") e quali dipendenti (o endogene);**
- c) definire le relazioni "direzionali" e "non-direzionali" che legano le variabili;**
- d) definire i vincoli (constraints) tra i parametri.**



## Identificazione dei parametri e del modello

Un modello si dice **identificato** se tutti i suoi parametri sono identificati, cioè se per tutti i suoi parametri esiste una **soluzione numerica *unica***.

Un modello **NON identificato** é un modello in cui la stessa matrice di covarianza riprodotta é compatibile con più insiemi di stime numeriche per gli stessi parametri:

$$\Sigma(\theta_1^\wedge) = \Sigma(\theta_2^\wedge), \text{ ma } \theta_1^\wedge \neq \theta_2^\wedge$$

dove  $\theta_1^\wedge$  e  $\theta_2^\wedge$  sono due vettori che contengono valori differenti per gli stessi parametri, ovvero per uno **stesso modello** esaminato su uno stesso campione.



## Un modello non identificato: l'EFA

Un caso noto di modello non identificato è l'Analisi Fattoriale Esplorativa. Le soluzioni fattoriali possono essere ruotate: ogni rotazione non cambia le proprietà matematiche della soluzione, ma ne cambia solo l'interpretazione *concettuale*.

	Sol. Iniziale (A)		Sol. Ruotata (B)	
	F1	F2	F1	F2
Determinato	.68	.51	.17	.83
Dinamico	.74	.48	.24	.85
Energico	.78	.33	.36	.77
Affidabile	.80	-.41	.87	.23
Responsabile	.84	-.43	.91	.24
Scrupoloso	.82	-.33	.83	.30

La % di varianza spiegata dalle due soluzioni è la stessa (circa 78%)





## Un modello non identificato: l'EFA

**Le due matrici sono perfettamente equivalenti da un punto di vista matematico:**

- spiegano la stessa varianza delle variabili originali**
- riproducono altrettanto bene la matrice delle correlazioni di partenza**

**La differenza dei valori delle saturazioni ha implicazioni profondamente diverse per la interpretazione del modello:**

- Soluzione non ruotata:  
1 fattore generale e 1 bipolare**
- Soluzione ruotata:  
2 fattori, Coscienziosità e Energia**

## Condizioni necessarie per l'identificazione

a) il numero dei coefficienti da stimare ( $t$ ) deve essere inferiore al numero di elementi non ridondanti nella matrice delle covarianze (uguale a  $q(q+1)/2$ , dove  $q$  è il numero di variabili osservate) [**t rule**], ovvero i gradi di libertà devono essere positivi.

$$\text{GDL} = [q(q+1)/2] - t$$

b) la scala di misura delle variabili latenti deve essere fissata

Se tutti i parametri del modello possono essere espressi come funzioni delle varianze e covarianze tra le variabili osservate, allora il modello è sicuramente identificato (es:  $\lambda_{21} = \sqrt{(\sigma_{32} \sigma_{21} / \sigma_{31})}$ )

## Identificazione e gradi di libertà

**Modello "non identificato" (underidentified):** gradi di libertà negativi, il numero di parametri è maggiore del numero di var/cov.

**Modello "appena identificato" (just identified), modello saturo:** ha tante incognite quanti parametri noti, ovvero tanti parametri quante var/cov, per cui ha **0 gradi di libertà**. Ha un fit perfetto e per questo non è interessante: non può essere disconfermato dai dati empirici.

**Modello "sovra identificato" (overidentified):** ha meno parametri che var/cov, quindi ha **gradi di libertà maggiori di 0**. Può essere disconfermato empiricamente.



## Condizioni in cui é realistico considerare il modello come identificato

a) Errori di misurazione non correlati ( $\theta$  è diagonale)

b) Modello **ricorsivo**:

- nessuna relazione di influenza reciproca  
(es,  $\eta_1 \rightleftarrows \eta_2$ )

- nessun loop (es,  $\eta_1 \rightarrow \eta_2 \rightarrow \eta_3 \rightarrow \eta_1$ )

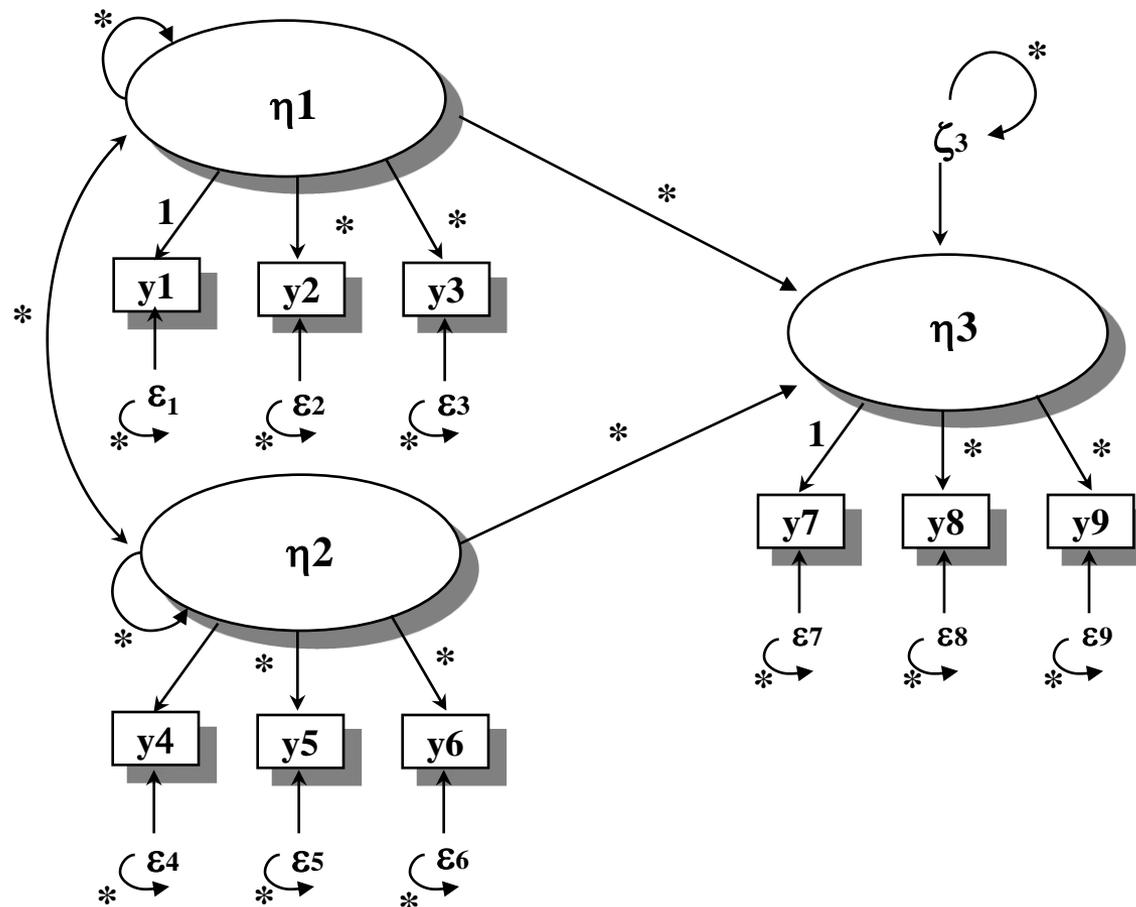
c) Numero di indicatori per ogni variabile latente adeguato (solitamente, almeno 3 indicatori)



## Fenomeni che possono far supporre non identificazione

- a) **Presenza di valori non ammissibili per alcuni parametri:**
- **varianze negative**
  - **correlazioni o coefficienti strutturali maggiori di 1 in valore assoluto, nel caso di una soluzione standardizzata ("Heywood cases")**
- b) **Valori molto elevati per gli errori standard: in un modello identificato l'errore standard dovrebbe avere un valore intorno al valore del parametro diviso  $\sqrt{N}$ , o di 1 diviso  $\sqrt{N}$  (McDonald, 1999).**
- c) **Non convergenza del processo di iterazione**

## ESERCIZIO 2: IDENTIFICAZIONE DEL MODELLO – CALCOLO DEI GRADI DI LIBERTA'

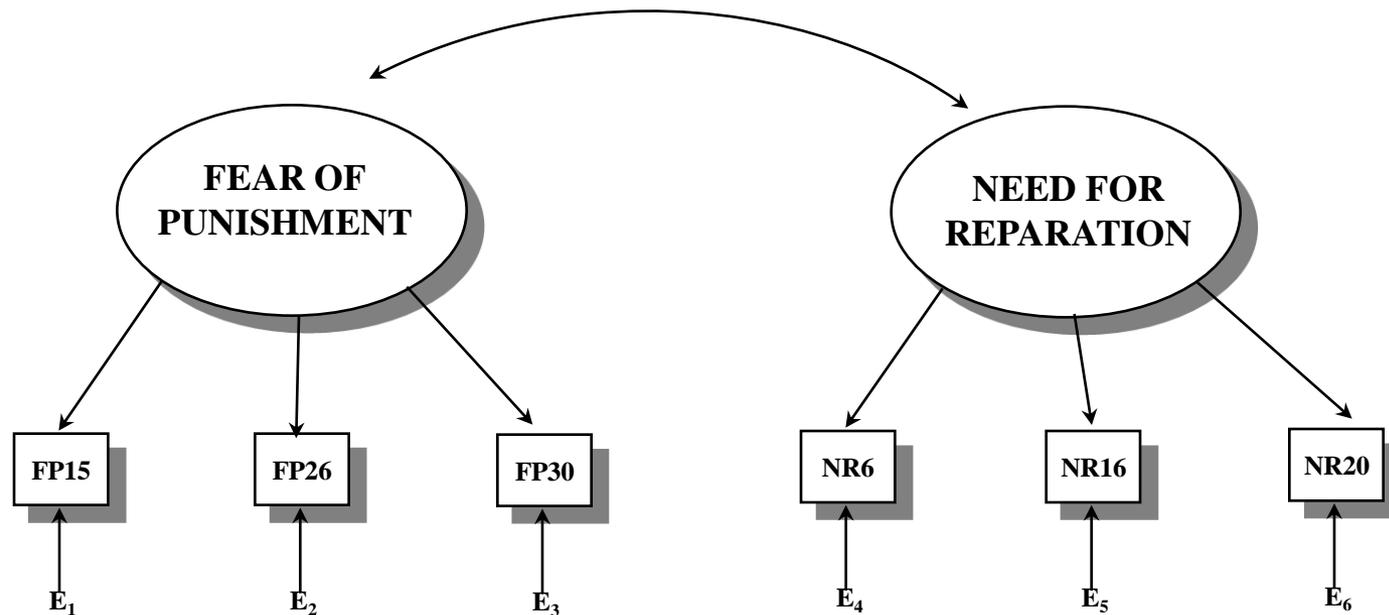


**E' un modello identificato ? Perché ? Quanti sono i suoi gradi di libertà ?**

***NB: gli asterischi indicano dei parametri che devono essere stimati.***

## CFA CON MPLUS

### Modello teorico della struttura fattoriale da esaminare



**Item FP: FP15) Mi capita di sentirmi gravato da sentimenti di colpa; FP26) Mi capita di ripensare con timore alle conseguenze di ciò che ho fatto o detto; FP30) Mi sono sentito come se mi rimordesse la coscienza; Item NR: NR6) Sento il bisogno di riparare ai torti che posso aver procurato ad altri; NR16) Prima o poi i nodi delle proprie colpe vengono al pettine; NR20) Di fronte ai miei errori desidero riparare il prima possibile**



# CFA CON MPLUS

TITLE: CFA

PARIS 2011 CFA.INP

DATA:

FILE IS dati\_efa\_cfa\_grezzi.dat;

VARIABLE:

NAMES ARE

FP10 FP15 FP26 FP30  
NR6 NR16 NR17 NR20 GENDER;

USEV ARE

FP15 FP26 FP30 NR6 NR16 NR20 ;

MISSING ARE ALL (9);

MODEL:

FEARPUN BY FP15 FP26 FP30 ;  
NEEDREP BY NR6 NR16 NR20 ;

OUTPUT: STANDARDIZED SAMPSTAT MODINDICES (3.84) TECH1;



## CFA CON MPLUS

- \* Le righe **TITLE**, **DATA**, **VARIABLE** hanno lo stesso significato delle linee omonime utilizzate nel programma per la EFA.
- \* Le equazioni che seguono la linea **"MODEL:"** mostrano quali variabili devono essere messe in relazione a quali fattori.
- \* Il comando **"BY"** sta per **"MEASURED BY"** e indica che la variabile latente a sinistra è misurata dalle variabili osservate a destra.



## CFA CON MPLUS

**\* Nella riga OUTPUT viene richiesto di aggiungere all'output di default :**

- la soluzione completamente standardizzata [STANDARDIZED]**
- le statistiche campionarie [SAMPSTAT]**
- gli indici di modificazione statisticamente significativi [MODINDICES(3.84)]**
- le matrici del modello di Muthén [TECH1]**



## CFA CON MPLUS

**I settings di default nel modello CFA sono i seguenti**

- \* La saturazione fattoriale della prima variabile dopo il "BY" viene fissata a 1**
- \* Le saturazioni fattoriali delle altre variabili sono stimate**
- \* Le varianze di errore sono stimate**
- \* Le covarianze tra i residui sono fissate a 0**
- \* Le varianze dei fattori sono stimate**
- \* Le covarianze tra le variabili esogene sono stimate**
- \* Il metodo di stima è ML**

# CFA CON MPLUS

## Parametrizzazione alternativa del Modello CFA

MODEL :

FEARPUN BY FP15\* FP26 FP30 ;

NEEDREP BY NR6\* NR16 NR20 ;

FEARPUN@1 NEEDREP@1;

FEARPUN with NEEDREP@0;

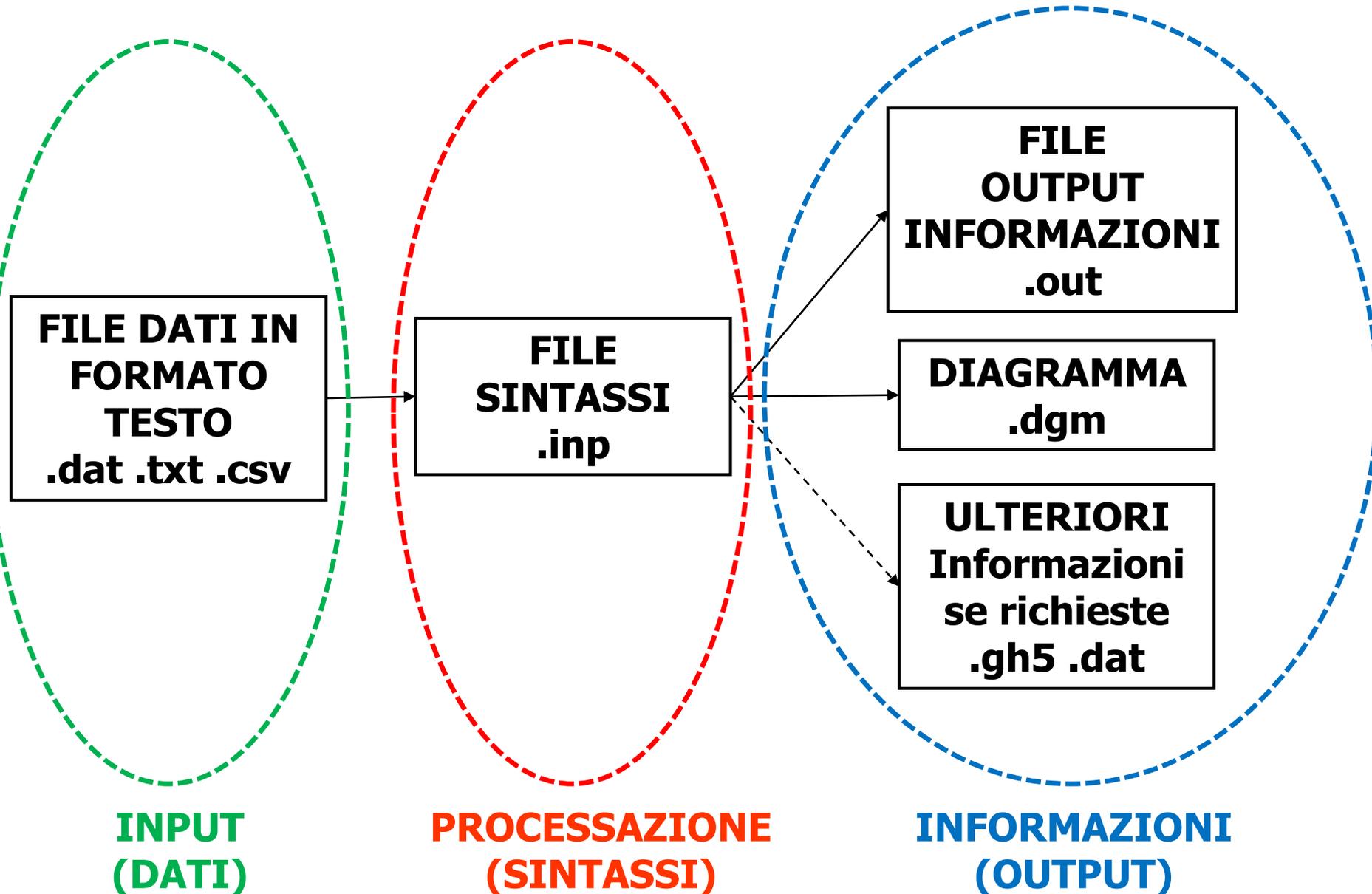
### Nella sintassi di MPLUS

- L'asterisco (\*) segnala che il parametro è libero.
- La chiocciolina (@) segnala che il parametro è fissato al valore che segue.



# **PREPARARE I DATI DA IMPORTARE IN *Mplus***

# LA STRUTTURA DEL PROGRAMMA *Mplus*



## PREPARARE I DATI DA IMPORTARE IN *Mplus*

***Mplus* legge essenzialmente file dati in formato testo (con estensioni .dat. .txt. .csv. ecc).**

**Il file dati può contenere dati di tipo differente: *a*) può contenere i dati "grezzi", ossia una matrice Casi X Variabili; *b*) il vettore delle medie delle variabili (non obbligatorio) considerate, seguito dalla loro matrice di varianze e covarianze; *c*) il vettore delle deviazioni standard (può essere preceduto dal vettore delle medie) delle variabili, seguito dalla loro matrice di correlazioni.**

**L'ordine delle variabili all'interno del file dati, a differenza di altri software, è del tutto indifferente.**

# DATI GREZZI – TYPE IS INDIVIDUAL

File	Modifica	Visualizza	Inserisci	Formato	?											
769	1.00	3.35	3.00	999.00	3.29	3.35	4.00	4.80	999.00	4.58	4.58	2.85	2.69	1.92	2.69	
903	1.00	3.12	2.94	2.88	2.35	999.00	1.80	4.20	3.60	3.25	999.00	2.00	2.31	2.54	2.23	
908	1.00	3.12	2.71	2.94	2.53	2.76	4.40	5.00	4.20	3.67	2.58	2.62	2.77	2.62	2.46	
916	1.00	2.41	2.24	3.06	2.24	999.00	3.40	1.80	4.00	4.00	999.00	2.00	2.08	2.15	2.15	
934	1.00	2.18	2.94	3.35	2.71	1.65	2.00	4.60	4.40	3.50	2.25	1.85	1.54	1.38	2.36	
940	1.00	3.29	3.06	3.12	2.94	2.88	4.40	4.40	3.60	4.50	4.33	1.77	2.31	2.31	2.54	
959	1.00	2.65	2.35	3.24	999.00	3.29	4.20	3.00	4.60	999.00	4.00	1.69	2.23	2.08	2.54	
961	1.00	3.18	3.18	3.00	3.53	3.18	3.80	5.00	5.00	4.67	4.50	2.54	2.77	1.77	2.85	
975	1.00	3.53	2.88	3.59	999.00	3.50	5.00	4.60	4.80	999.00	5.00	2.85	3.00	2.62	2.85	
980	1.00	3.29	2.71	3.38	2.76	2.71	4.00	3.60	3.00	4.17	3.33	1.92	2.31	2.08	2.54	
1168	1.00	3.29	3.71	2.94	3.24	999.00	5.00	4.60	4.60	4.83	999.00	2.46	2.23	2.23	2.36	
2000	1.00	3.65	3.65	3.06	2.88	3.29	5.00	4.80	4.40	4.33	4.25	2.92	2.85	2.69	2.77	
2001	1.00	2.71	2.53	2.88	2.47	999.00	2.40	3.20	4.00	3.00	999.00	2.15	2.00	1.92	2.62	
2003	1.00	3.59	3.41	3.65	2.65	2.65	4.20	4.20	3.80	4.08	4.00	2.54	2.77	2.23	2.85	
2005	1.00	3.59	2.47	3.12	2.59	2.59	4.80	1.25	3.20	3.67	4.75	1.85	2.69	2.62	2.33	
2008	1.00	3.35	3.47	3.53	3.35	999.00	5.00	4.60	4.80	3.67	999.00	2.38	2.62	2.31	2.36	
2012	1.00	1.50	2.88	2.94	2.94	1.82	1.60	3.80	2.80	3.64	2.83	2.46	2.85	2.31	2.17	
2013	1.00	3.00	2.88	3.47	999.00	3.12	2.80	1.20	4.60	999.00	3.75	2.54	2.15	2.62	2.54	
2024	1.00	2.94	2.76	2.35	2.94	3.12	3.40	3.80	5.00	4.50	4.83	2.46	2.38	2.23	2.54	
2025	1.00	3.47	3.53	3.35	3.41	3.00	4.00	4.60	4.20	4.08	3.42	2.50	2.77	2.08	2.77	
2028	1.00	3.53	3.65	3.76	3.88	3.94	4.60	5.00	5.00	5.00	4.92	2.54	2.31	2.08	2.54	
2031	1.00	3.53	3.53	3.24	3.00	3.82	4.40	2.80	2.40	3.00	5.00	2.00	2.54	2.00	2.36	
2040	1.00	3.41	3.76	3.29	3.29	3.29	4.40	5.00	4.00	3.82	4.33	2.31	2.85	2.31	2.23	
2042	1.00	3.82	3.94	4.00	999.00	3.94	5.00	5.00	4.80	999.00	4.42	2.46	2.92	2.69	2.85	
5005	1.00	2.41	2.29	2.24	3.00	999.00	2.80	3.00	1.40	4.25	999.00	1.85	2.08	1.69	2.00	
5019	1.00	2.94	3.65	2.94	3.24	999.00	4.80	5.00	4.00	3.83	999.00	2.38	2.46	2.23	2.65	
5025	1.00	3.06	3.06	2.76	2.76	2.65	4.80	4.60	4.60	3.25	3.50	2.38	2.46	2.31	2.31	
5040	1.00	3.00	2.82	3.06	2.47	999.00	2.60	5.00	3.60	3.17	999.00	2.69	2.69	2.31	2.92	



## DATI GREZZI

- 1) Le righe rappresentano **i soggetti** (o la singola osservazione) e le colonne **le variabili**;
- 2) Tutti i valori contenuti nel file dati debbono essere **preferibilmente numerici** per evitare problemi nella lettura del file da parte del software;
- 3) È più semplice codificare tutti i valori mancanti con un **unico valore comune** a tutte le variabili (es. 9, 99, 999, -9, -99, -999, ecc.);
- 4) Vanno **sostituite tutte le virgole con il punto**, altrimenti *Mplus* non leggerà i dati.



## **DATI GREZZI – Creare il file da SPSS**

- 1) Il suggerimento è salvare un file .sav includendo solo le variabili cui siamo interessati (procedura estremamente utile quando abbiamo un dataset d'origine molto ampio) e, qualora se ne disponga, del codice identificativo del soggetto;**
- 2) Ricodificare tutti i valori mancanti con un unico valore, ricordandosi di specificare anche nel file .sav il valore che abbiamo scelto;**
- 3) Salvare il file nel suo formato originale ed esportarlo come un file dati formato testo, attraverso il percorso: **FILE – SALVA CON NOME – SALVA COME TIPO: Ascii fisso (\*.dat).****

## DATI GREZZI – Creare il file da SPSS

4) Aprire il file dati in comune editor di testo (es. Blocco note o WordPad) e sostituire (qualora ce ne fossero) tutte le virgole con il punto (nel blocco note, il percorso è **MODIFICA – SOSTITUISCI – TROVA(.) – SOSTITUISCI CON(.) – SOSTITUISCI TUTTO**);

5) Ricordarsi di salvare il file dopo aver effettuato tali modifiche.

Il file così creato avrà una **estensione .dat**

# DATI GREZZI – Creare il file da SPSS

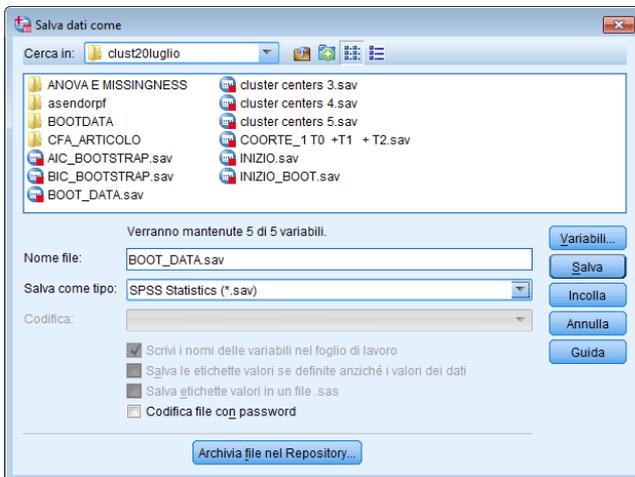
## 1)

	CASENUM	GIUDIZIO	AIUT	EMON	SRL
1	1	2,00	3,33	2,33	3,00
2	2	2,00	3,67	2,67	5,00
3	3	2,25	3,00	3,33	3,00
4	3	2,25	3,00	3,33	3,00
5	5	3,25	3,33	3,33	4,00
6	6	3,25	2,67	3,33	4,00
7	7	2,25	3,00	2,00	2,67
8	7	2,25	3,00	2,00	2,67
9	8	3,50	3,67	4,33	4,67
10	9	2,75	4,00	3,00	4,00
11	9	2,75	4,00	3,00	4,00
12	9	2,75	4,00	3,00	4,00

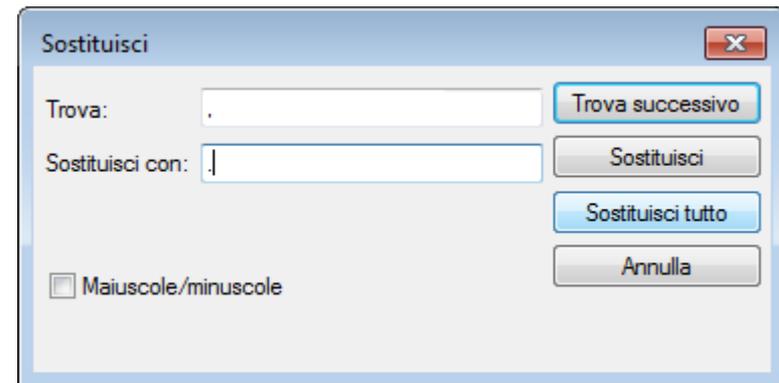
## 2)

Nome	Tipo	Larghezza	Decimali	Etichetta	Valori	Mancante/i
CASENUM	Numerico	8	0		Nessuno	Nessuno
GIUDIZIO	Numerico	8	2		Nessuno	-999,00
AIUT	Numerico	8	2		Nessuno	-999,00
EMON	Numerico	8	2		Nessuno	-999,00
SRL	Numerico	8	2		Nessuno	-999,00

## 3)



## 4)





## **DATI GREZZI – Creare il file da EXCEL**

- 1) Anche qui, ricodificare tutti i valori mancanti con un solo valore che li rappresenti all'interno di tutte le variabili;**
- 2) Esportare il file in formato testo, attraverso il percorso: FILE – SALVA CON NOME – SALVA COME TIPO: TEXT Tab delimited (\*.txt);**
- 3) Aprire il file in un editor di testo e si noterà che se la prima riga del file excel era occupata dalle etichette delle variabili queste verranno trasportate anche nel file testo (**va cancellata!**);**
- 4) Sostituire tutte le virgole con il punto nell'editor di testo.**

# PREPARARE I DATI DA IMPORTARE IN *Mplus*

## *RIASSUMENDO...*

**L'ordine delle variabili all'interno del file dati, a differenza di altri software, è del tutto indifferente.**

**I file che Mplus può leggere sono sempre file di testo.**

**Si possono importare dati grezzi (TYPE IS INDIVIDUAL, consigliata la codifica con un unico valore per i missing all'interno di variabili differenti) oppure file contenenti *summary data* (TYPE IS COVARIANCE, TYPE IS MEANS COVARIANCE, TYPE IS CORRELATIONS, ecc., molto utile per riprodurre modelli con i dati reperibili normalmente all'interno degli articoli).**

## Stima dei parametri

Bisogna stimare i parametri  $\theta$  in modo che la matrice riprodotta  $\Sigma(\theta)$  sia più vicina possibile a quella osservata  $\Sigma$ .

Esistono diversi metodi di stima che danno origine ad altrettante “**funzioni di adattamento**” (fit function) utilizzabili per questo compito.

Le funzioni di adattamento sono rappresentate dall'espressione “ $F(S, \Sigma(\theta^{\wedge}))$ ”.

I metodi di stima più comunemente utilizzati sono:

- \* **la massima verosimiglianza (Maximum Likelihood)**
- \* **i minimi quadrati (Least Squares)**

## Il metodo della Massima Verosimiglianza (ML)

In generale scopo del metodo ML è quello di stimare il parametro della popolazione che sia più vicino al valore campionario osservato, ovvero tale che la probabilità di osservare quel valore campionario data quella stima sia massima, ovvero:

$$P[(\hat{\theta} - \bar{\mathbf{x}}) < \varepsilon] = \max$$

Tra tutte le stime possibili di  $\theta$  la stima ML è quella che massimizza la probabilità o la verosimiglianza (likelihood) che le differenze tra stima e valore osservato siano dovute solo al caso.

## Il metodo della Massima Verosimiglianza (ML) per le stime dei parametri di un modello Mplus

**Nel modello Mplus abbiamo visto che ci sono diverse matrici di parametri. Tramite queste matrici è possibile riprodurre la matrice di covarianze campionaria  $S$  nel modo seguente:**

$$\Sigma(\theta) = \Lambda (I - B)^{-1} \Psi (I - B)^{-1'} \Lambda' + \Theta$$

**Indichiamo la matrice riprodotta in questo modo  $\Sigma(\theta)$ .  
Se il modello funziona perfettamente allora  $S = \Sigma(\theta)$**

**Come facciamo a trovare le migliori stime possibili per i parametri del modello affinché  $S$  e  $\Sigma(\theta)$  siano più simili possibile ?**



## Il metodo della Massima Verosimiglianza (ML) per le stime dei parametri di un modello Mplus

**Il miglior insieme di stime per i parametri  $\theta$  è quello che rende minima la differenza tra  $S$  e  $\Sigma(\theta)$ .**

**Questo vuol dire che le migliori stime sono quelle che forniscono la più elevata probabilità (cioè, verosimiglianza) di osservare  $S$  da un campione casuale di una popolazione che ha matrice di covarianza  $\Sigma$ , la cui stima è rappresentata da  $\Sigma(\theta)$ .**

**Le stime  $\theta$  che determinano tale matrice  $\Sigma(\theta)$  sono le stime di Massima Verosimiglianza.**

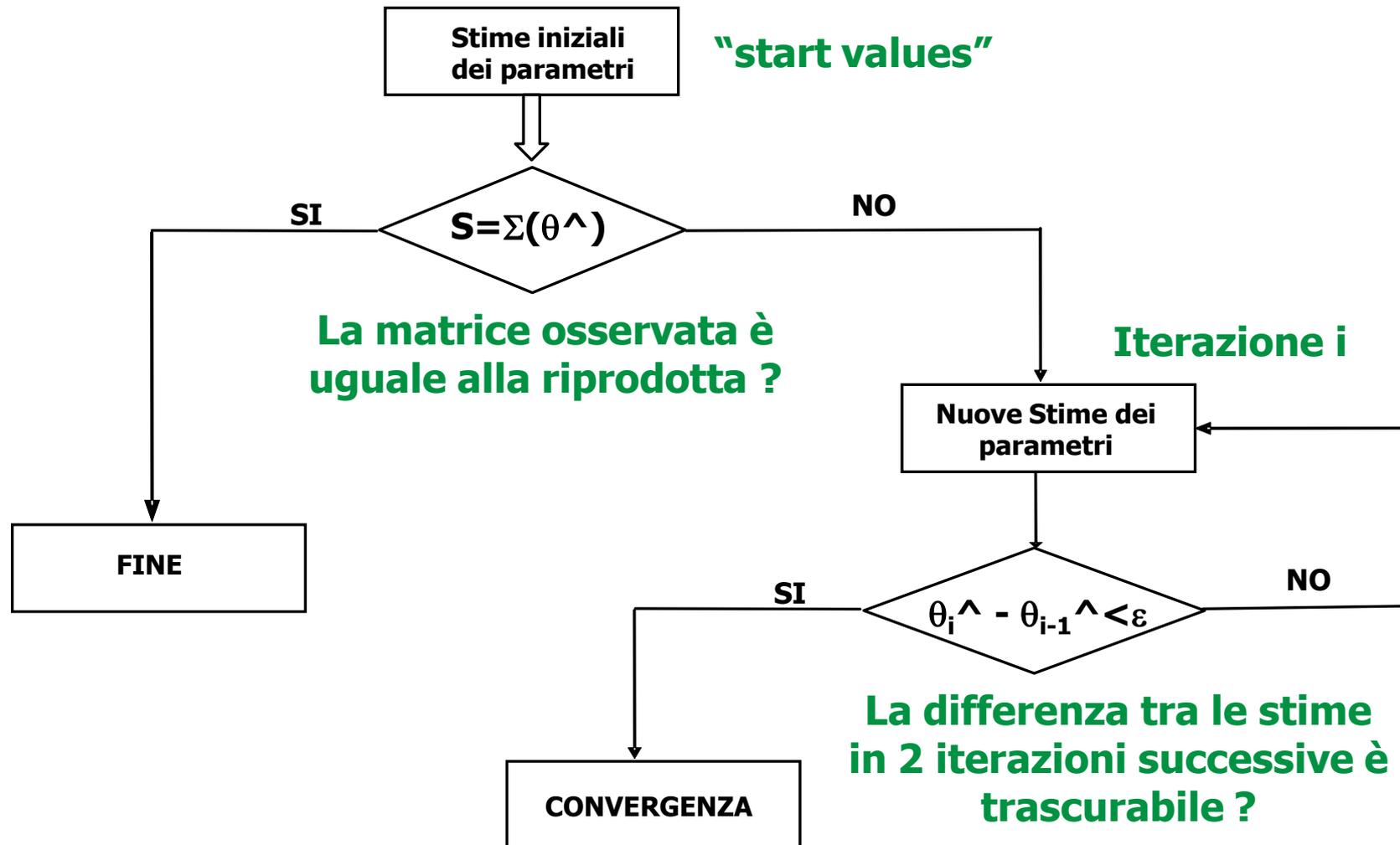
$$F_{ML}(S, \Sigma(\theta^{\wedge})) = \text{tr}(S\Sigma(\theta)^{-1}) + \ln |\Sigma(\theta)| - \ln |S| - q$$

## Il metodo della Massima Verosimiglianza (ML) per le stime dei parametri di un modello Mplus

Minimizzare la funzione  $F_{ML}$  vuol dire trovare dei valori per le stime  $\theta$  che **MINIMIZZANO** la differenza tra  $S$  e  $\Sigma(\theta)$ . Questi valori vengono ricavati attraverso una **procedura iterativa**, basata su alcuni consolidati algoritmi di calcolo numerico.

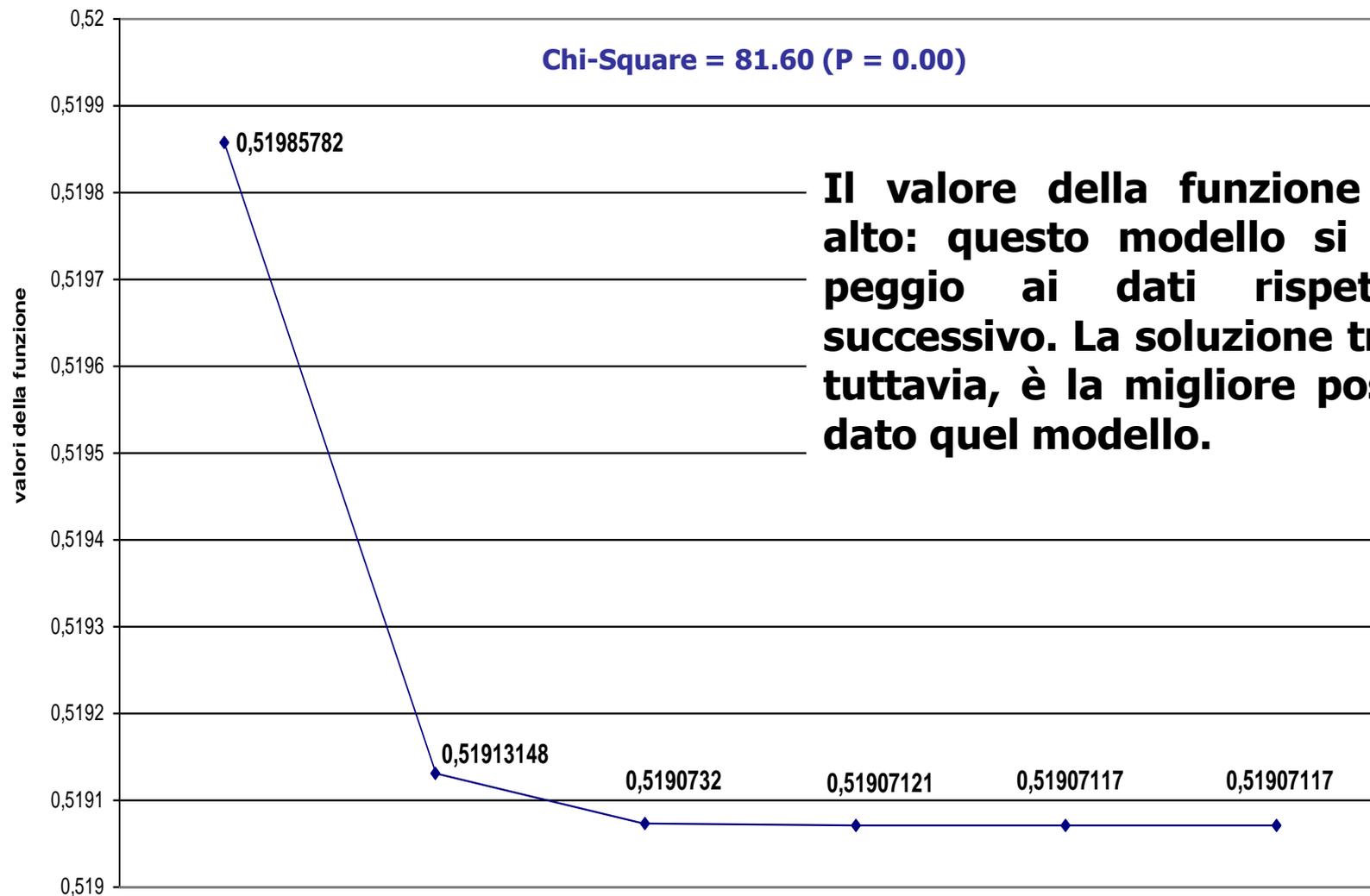
E' necessario utilizzare un metodo iterativo perché la funzione  $F_{ML}$  è una funzione non lineare complessa, per la quale non è possibile arrivare facilmente ad una soluzione esplicita.

# Le stime di Massima Verosimiglianza (ML): Il processo iterativo di stima dei parametri in Mplus

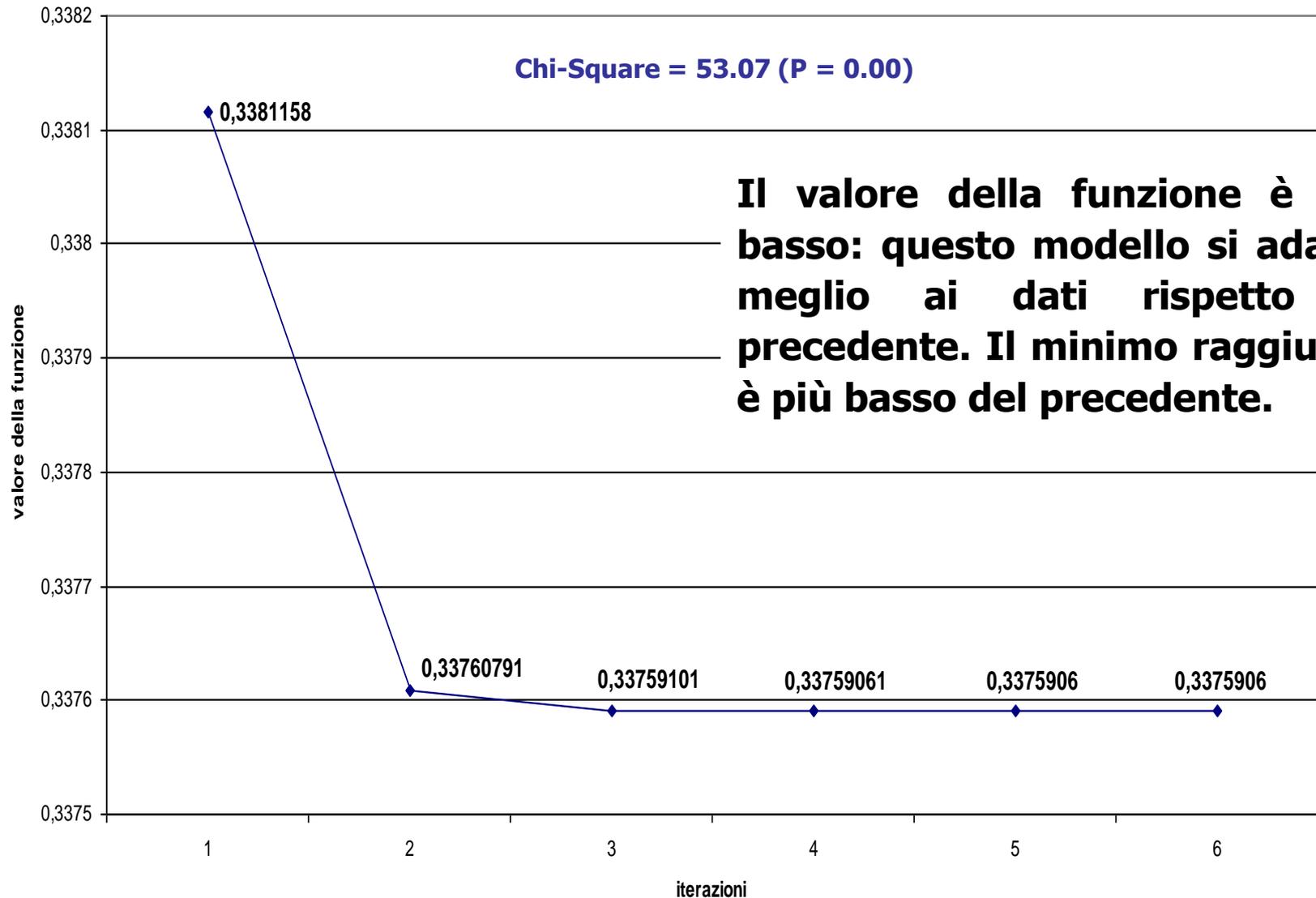


Le stime ottenute sono quelle che rendono minima la differenza tra  $\Sigma(\theta^{\wedge})$  e  $S$

# Il processo di stima dei parametri



# Il processo di stima dei parametri



# CFA IN MPLUS CON VARIABILI CONTINUE

## Metodi di stima dei parametri

- \* **GLS** – minimi quadrati generalizzati
- \* **WLS** – minimi quadrati ponderati (ADF)
- \* **ML** – massima verosimiglianza (DEFAULT)
- \* **MLM, MLMV, MLR, MLF** – metodi robusti che forniscono risultati corretti anche in presenza di violazioni della normalità (solo con dati grezzi)
  - MLM** = Correzione di Satorra-Bentler
  - MLMV** = Correzione rispetto alle varianze-covarianze e alle medie
  - MLR** = Correzione di Yuan-Bentler, robusta anche alla non indipendenza delle osservazioni
  - MLF** = Standard error approssimati, Chi-quadrato tradizionale

## Alcune funzioni di adattamento "speciali" in Mplus

**I metodi robusti che forniscono risultati corretti anche in presenza di violazioni della normalità (solo con dati grezzi)**

***Massima Verosimiglianza Robusta (ML-M):***

**Il chi-quadrato e gli errori standard vengono opportunamente ponderati per fornire stime corrette in presenza di non normalità (Satorra e Bentler).**

***Minimi Quadrati Ponderati Robusti (WLS-MV):***

**Fornisce stime corrette dei parametri e valori corretti del chi-quadrato e degli errori standard in presenza di non normalità e di variabili categoriali (Muthén).**

## Alcune funzioni di adattamento "speciali" in Mplus

**TITLE: CFA**

**DATA:**

**LISTWISE=ON ;**

**FILE IS dati\_efa\_cfa\_grezzi.dat;**

**VARIABLE:**

**NAMES ARE**

**FP10 FP15 FP26 FP30  
NR6 NR16 NR17 NR20 GENDER;**

**USEV ARE**

**FP15 FP26 FP30 NR6 NR16 NR20 ;**

**MISSING ARE ALL (9);**

**ANALYSIS: ESTIMATOR = MLM;**

**MODEL:**

**FEARPUN BY FP15 FP26 FP30 ;**

**NEEDREP BY NR6 NR16 NR20 ;**

**OUTPUT: STANDARDIZED SAMPSTAT MODINDICES(3.84)**

**TECH1;**

**PARIS\_2011\_CFA\_MLM.INP**



# CFA CON MPLUS

## Lettura dell' *Output* : Elementi più importanti

### SAMPLE STATISTICS

#### Covariances

	FP15	FP26	FP30	NR6	NR16
FP15	2.232				
FP26	0.821	1.952			
FP30	1.037	0.910	2.426		
NR6	0.323	0.241	0.264	1.905	
NR16	0.115	0.290	0.155	0.537	1.589
NR20	0.156	0.245	0.087	0.485	0.510

#### Covariances

##### NR20

NR20	1.353
------	-------



# CFA CON MPLUS

## Lettura dell' *Output* : Elementi più importanti

### Correlations

	FP15	FP26	FP30	NR6	NR16
FP15	1.000				
FP26	0.393	1.000			
FP30	0.446	0.418	1.000		
NR6	0.156	0.125	0.123	1.000	
NR16	0.061	0.165	0.079	0.308	1.000
NR20	0.090	0.151	0.048	0.302	0.348

### Correlations

#### NR20

NR20	1.000
------	-------



# CFA CON MPLUS

## Lettura dell' *Output* : Elementi più importanti

### UNIVARIATE SAMPLE STATISTICS

### UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS

Variable/ Sample Size	Mean/ Variance	Skewness/ Kurtosis	Minimum/ Maximum	% with Min/Max	Percentiles 20%/60%	Percentiles 40%/80%	Median
FP15 816.000	3.176 2.234	0.063 -0.956	1.000 6.000	18.50% 6.74%	2.000 4.000	3.000 4.000	3.000
FP26 812.000	3.686 1.952	-0.321 -0.608	1.000 6.000	9.48% 9.11%	2.000 4.000	4.000 5.000	4.000
FP30 813.000	3.148 2.426	0.057 -1.117	1.000 6.000	21.89% 6.40%	1.000 4.000	3.000 5.000	3.000
NR6 817.000	4.376 1.904	-0.856 0.177	1.000 6.000	6.24% 22.40%	3.000 5.000	4.000 6.000	5.000
NR16 820.000	4.676 1.587	-1.096 1.048	1.000 6.000	4.02% 29.76%	4.000 5.000	5.000 6.000	5.000
NR20 816.000	5.045 1.352	-1.510 2.343	1.000 6.000	2.57% 44.85%	4.000 6.000	5.000 6.000	5.000



# CFA CON MPLUS

## Chi-Square Test of Model Fit

Value	21.622
Degrees of Freedom	8
P-Value	0.0057

## RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.046	
90 Percent C.I.	0.023	0.069
Probability RMSEA $\leq$ .05	0.584	

## CFI/TLI

CFI	0.979
TLI	0.961

# CFA CON MPLUS

E' il modello in cui non viene ipotizzata nessuna relazione tra le variabili.

Chi-Square Test of Model Fit for the Baseline Model

Value	663.540
Degrees of Freedom	15
P-Value	0.0000

SRMR (Standardized Root Mean Square Residual)

Value	0.026
-------	-------

## Valutazione globale del modello

**Il valore minimo della funzione di fit moltiplicato per il numero di soggetti meno 1 segue approssimativamente la distribuzione del  $\chi^2$  con gradi di libertà (df) uguali a  $(q(q+1)/2)-t$ , dove  $q$  è il numero di variabili osservate e  $t$  il numero di parametri stimati, e con valore atteso  $E(\chi^2)=df$ .**

**Questo consente di esaminare statisticamente l'adattamento del modello ai dati tramite l'ipotesi nulla  $\Sigma=\Sigma(\theta)$ .**

**Lo standard per questo confronto a livello di campione é che  $\Sigma^{\wedge}$  e  $S$  siano uguali.**



## Valutazione globale del modello

**Se il chi-quadrato risulta statisticamente non significativo, l'ipotesi nulla  $H_0: \Sigma = \Sigma(\theta)$  non può essere rifiutata, dunque il modello teorico risulta compatibile con i dati empirici.**

**Se invece il valore del  $\chi^2$  risulta statisticamente significativo l'ipotesi nulla va rifiutata, dunque il modello teorico non risulta compatibile con i dati empirici.**

# Valutazione globale del modello

## Indici alternativi di fit

Il chi quadrato è fortemente dipendente dalla numerosità del campione, e questo lo rende **conservativo** se il campione è ampio (rifiuta troppo spesso  $H_0$ ) ma **liberale** se il campione è esiguo (rifiuta troppo di rado  $H_0$ ).

Per questo sono stati sviluppati **indici alternativi** per valutare globalmente la bontà dell'adattamento.

- \* Misure di fit nel campione (SRMR, RMR)
- \* Indici incrementali o comparativi (TLI, CFI)
- \* Indici di approssimazione (RMSEA)



# Valutazione globale del modello

## Misure di fit nel campione

**RMR = ROOT MEAN-SQUARE RESIDUAL =**  
**=  $[2\sum_i\sum_j(s_{ij} - \sigma^{ij})^2 / (q(q+1))]^{1/2}$**

**SRMR = STANDARDIZED RMR**

**Indicano la media della varianza e covarianza residua, cioè non spiegata dal modello.**

**L'RMR può essere utilizzato per confrontare l'adattamento di due differenti modelli specificati sugli stessi dati, oppure di uno stesso modello specificato su dati differenti. Allora, il modello da privilegiare è quello che presenta un RMR più basso.**

**L'SRMR può essere interpretato in assoluto. Valori bassi indicano buon fit.**

# Valutazione globale del modello

## Indici incrementali o comparativi

Valutano l'adeguatezza del modello rispetto ad un modello nullo in cui si ipotizza che non ci siano relazioni tra le variabili.

### TUCKER AND LEWIS INDEX (TLI) - (NNFI, NON NORMED FIT INDEX)

$$\frac{\left(\chi^2_{nullo} / df_{nullo}\right) - \left(\chi^2_{target} / df_{target}\right)}{\left(\chi^2_{nullo} / df_{nullo}\right) - 1}$$

# Valutazione globale del modello

## Indici incrementali o comparativi

**Valutano l'adeguatezza del modello rispetto ad un modello nullo in cui si ipotizza che non ci siano relazioni tra le variabili.**

### COMPARATIVE FIT INDEX (CFI)

$$1 - \frac{\max\left[\left(\chi^2_{target} - df_{target}\right), 0\right]}{\max\left[\left(\chi^2_{nullo} - df_{nullo}\right), \left(\chi^2_{target} - df_{target}\right), 0\right]}$$

**Il CFI varia da 0 a 1, il TLI può risultare anche maggiore di 1. Valori intorno a 1 indicano buon fit.**



# Valutazione globale del modello

## Indice di approssimazione

### RMSEA (Root Means Square Error of Approximation)

**Il chi quadrato valuta l'ipotesi nulla che  $\Sigma = \Sigma(\theta)$  e presuppone che esista un modello vero nella popolazione, ovvero un modello che è perfettamente consistente con i dati empirici, ovvero che rappresenta una "fotografia" perfetta della realtà.**

**Spesso questo è irrealistico. I modelli al più possono fornire un'immagine **approssimativa** della realtà.**

# Valutazione globale del modello

## Indice di approssimazione

### RMSEA (Root Means Square Error of Approximation)

Allora, l'RMSEA valuta quanto errore commettiamo nell'approssimare la realtà con il nostro modello.

Se l'errore è contenuto, il nostro modello non sarà una fotografia perfetta della realtà ma almeno la **approssima** sufficientemente bene.

Se l'errore è grande, il nostro modello non può considerarsi **nemmeno un'approssimazione** della realtà.

L'RMSEA valuta l'ipotesi che  $\Sigma \approx \Sigma(\theta)$ .

# Valutazione globale del modello

## Indice di approssimazione

### RMSEA (Root Means Square Error of Approximation)

**E' una misura della discrepanza tra  $\Sigma$  e  $\Sigma(\theta)$  dovuta all'approssimazione, ponderata per i gradi di libertà del modello, e quindi é una misura del fit che tiene in considerazione la parsimonia del modello.**

$$RMSEA = \frac{\hat{F}_0}{df}; \quad \hat{F}_0 = \max \left[ \frac{(\min F - df)}{n}, 0 \right]$$

**RMSEA ≤ .05: errore di approssimazione minimo,  
.05 < RMSEA ≤ .08: errore di appross. accettabile,  
RMSEA > .08: il modello non tiene nella popolazione.**



# Valutazione globale del modello

## Indice di approssimazione

## RMSEA (Root Means Square Error of Approximation)

### Test of Close fit

Esamina l'ipotesi nulla che l'RMSEA sia inferiore a .05 e quindi la seguente ipotesi nulla:  $H_0: \Sigma \approx \Sigma(\theta)$ .

Se  $p(\text{RMSEA} \leq .05) > .05$  non possiamo rifiutare l'ipotesi nulla che il modello sia almeno approssimativamente adeguato.



# Valutazione globale del modello

## Indice di approssimazione

### RMSEA (Root Means Square Error of Approximation)

L'RMSEA viene fornito anche nella versione di **stima intervallare** (con un intervallo di confidenza del 10%).

Il limite inferiore dell'intervallo di confidenza dell'RMSEA deve essere minore di .05, il limite superiore minore di .08: Allora il test of close fit risulta non significativo.



## Valutazione globale del modello

Le indicazioni di Hu e Bentler (1998, 1999)  
sui valori degli indici alternativi

Presentare nei risultati dei SEM almeno due indici:

**SRMR:** è il più sensibile a segnalare modelli errati

un indice tra: **TLI CFI RMSEA**

Valori di cut-off per gli indici "migliori"

**TLI e CFI** maggiore/uguale a .95

**SRMR** minore di .08

**RMSEA** minore di .06 (o di .08)



## Fattori che influenzano il fit

**Ampiezza del Campione:** Modelli testati su campioni più grandi hanno un fit peggiore (soprattutto il chi-quadrato).

**Numero di variabili nel modello/ Gradi di libertà:** Un modello con più variabili generalmente tende ad avere un fit peggiore di un modello con meno variabili/parametri. Più è grande la matrice delle covarianza più è difficile ottenere un buon fit. Modelli con pochi gradi di libertà (a parità di altre condizioni) ottengono più facilmente un buon fit (Marsh, Hau, Balla, & Grayson, 1998).

## Fattori che influenzano il fit

**Relazioni deboli tra le variabili osservate (grandezza delle covarianze):** covarianze piccole aumentano la probabilità di ottenere un basso chi-quadrato, e quindi un buon fit (Dillon & Goldstein, 1985; Fornell, 1983).

**Dati non-normali (specialmente la curtosi):** aumentano il chi quadrato e le misure assolute. Gli indici incrementali e comparativi sono meno influenzati (Muthén & Kaplan, 1985).

**Attendibilità delle variabili osservate:** variabili più attendibili determinano un migliore fit (Jackson, 2001, 2003).



# CFA CON MPLUS

## MODEL RESULTS

## MODEL RESULTS

		<b>Stima non standardizzata</b>	<b>t = Stima/errore</b>		<b>Two-Tailed</b>
		<b>Estimate</b>	<b>S.E.</b>	<b>Est./S.E.</b>	<b>P-Value</b>
			<b>Errore standard</b>		<b>probabilità</b>
<b>FEARPUN</b>	<b>BY</b>				
	FP15	1.000	0.000	999.000	999.000
	FP26	0.903	0.081	11.117	0.000
	FP30	1.086	0.094	11.575	0.000
<b>NEEDREP</b>	<b>BY</b>				
	NR6	1.000	0.000	999.000	999.000
	NR16	1.002	0.126	7.925	0.000
	NR20	0.903	0.113	8.026	0.000
<b>NEEDREP</b>	<b>WITH</b>				
	FEARPUN	0.209	0.046	4.529	0.000
<b>Variances</b>					
	FEARPUN	0.932	0.117	7.949	0.000
	NEEDREP	0.547	0.096	5.697	0.000



# CFA CON MPLUS

	<b>Stima non standardizzata</b>	<b>t = Stima/errore</b>		<b>Two-Tailed</b>
	<b>Estimate</b>	<b>S.E.</b>	<b>Est./S.E.</b>	<b>P-Value</b>
		<b>Errore standard</b>		<b>probabilità</b>
<b>Residual Variances</b>				
FP15	1.300	0.099	13.094	0.000
FP26	1.191	0.087	13.677	0.000
FP30	1.327	0.110	12.017	0.000
NR6	1.358	0.097	14.051	0.000
NR16	1.041	0.084	12.364	0.000
NR20	0.907	0.070	12.889	0.000

# CFA CON MPLUS

## STDYX Standardization

	<b>Stima non standardizzata</b>		<b>t = Stima/errore</b>	<b>Two-Tailed</b>
	<b>Estimate</b>	<b>S.E.</b>	<b>Est./S.E.</b>	<b>P-Value</b>
		<b>Errore standard</b>		<b>probabilità</b>
<b>FEARPUN BY</b>				
FP15	0.646	0.033	19.482	0.000
FP26	0.624	0.034	18.436	0.000
FP30	0.673	0.033	20.261	0.000
<b>NEEDREP BY</b>				
NR6	0.536	0.042	12.691	0.000
NR16	0.588	0.042	13.841	0.000
NR20	0.574	0.042	13.654	0.000
<b>NEEDREP WITH</b>				
FEARPUN	0.292	0.053	5.466	0.000
<b>Variances</b>				
FEARPUN	1.000	0.000	999.000	999.000
NEEDREP	1.000	0.000	999.000	999.000
<b>Residual Variances</b>				
FP15	0.582	0.043	13.581	0.000
FP26	0.610	0.042	14.429	0.000
FP30	0.547	0.045	12.221	0.000
NR6	0.713	0.045	15.745	0.000
NR16	0.655	0.050	13.120	0.000
NR20	0.670	0.048	13.868	0.000



# CFA CON MPLUS

## R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
FP15	0.418	0.043	9.741	0.000
FP26	0.390	0.042	9.218	0.000
FP30	0.453	0.045	10.130	0.000
NR6	0.287	0.045	6.345	0.000
NR16	0.345	0.050	6.921	0.000
NR20	0.330	0.048	6.827	0.000



## Valutazione dei singoli parametri

- \* I valori dei parametri devono essere ammissibili
- \* Significatività statistica delle stime dei parametri:

$$t = (\hat{\theta}_i - \theta_0) / SE$$

dove  $\theta_0 = 0$ , e SE = errore standard del parametro

**Valori maggiori o uguali a  $|1.96|$  indicano una significatività al livello di probabilità di .05.**

- \* **Varianza spiegata (per le variabili dipendenti):**  
**R quadrato**

## Modifica del modello

**Per "migliorare" un modello si possono:**

- \* Fissare a 0 i valori dei parametri che non sono risultati significativamente differenti da 0 [Il test di Wald in EQS calcola il cambiamento determinato nel chi-quadrato dall'aver fissato a 0 uno o più parametri].**
- \* Liberare parametri che erano stati fissati a zero: esistono dei "diagnostici" che permettono di individuare quali parametri modificare [Modification Indexes, Moltiplicatori di Lagrange]. Questi valutano il cambiamento nel  $\chi^2$  determinato dalla "liberazione" dei parametri precedentemente fissati a zero.**



## Modifica del modello

- \* Anche l'esame delle covarianze residue standardizzate può evidenziare difficoltà del modello nel ricostruire alcune specifiche covarianze osservate.
- \* E' fondamentale validare i "nuovi" modelli che risultano da modifiche su un campione diverso. In assenza di campioni diversi si può applicare una strategia di **cross-validation** considerando un sotto-campione per generare le modifiche e un altro validare e generalizzare tali modifiche.
- \* La strategia migliore è comunque quella di specificare modelli alternativi prima dell'analisi dei dati stessa. Questo evita problemi di mancata generalizzabilità e di capitalizzazione sulle caratteristiche di uno specifico campione.

## Modifica del modello

**Spesso gli indici di modificazione suggeriscono di liberare alcune covarianze tra gli errori di misurazione. Secondo Fornell (1983) e Bagozzi (1983) queste possono essere liberate a patto che:**

- a) Siano sensate da un punto di vista teorico o metodologico**
- b) Non alterino le stime dei parametri strutturali**
- c) Non alterino le stime dei parametri del modello di misurazione**

**Anche se queste tre condizioni sono rispettate, l'inclusione delle covarianze tra gli errori di misurazione indebolisce l'interpretazione del modello, perché il modello di misurazione non è corretto (il numero di fattori non basta a rendere ragione delle covarianze tra le variabili osservate).**

## Modifica del modello

**E' diverso il caso della covarianza tra residui  $\zeta$  delle variabili latenti (Kline, 2005).**

**Il residuo di una variabile latente ( $\zeta$ ) riflette l'errore di specificazione: non tutte le "cause" della variabile latente ( $\eta$ ) sono state considerate.**

**Se due  $\zeta$  covariano questo dipende dal fatto che probabilmente c'è almeno una variabile, che rappresenta una "causa" comune per le due  $\eta$ , che è stata omessa dal modello.**

**Questo fenomeno è frequentissimo nella ricerca psicologico-sociale.**



# CFA CON MPLUS

**ATTENZIONE !! SONO INDICI  
UNIVARIATI ! I PARAMETRI FISSI  
VANNO LIBERATI UNO PER VOLTA !!**

**Sono da considerare significativi gli  
indici di modifica che risultano  
maggiore di 3.84 (per  $\alpha=.05$ ) o di 6.63  
(per  $\alpha=.01$ ).**

		M. I.	E. P. C.	Std E. P. C.	StdYX E. P. C.
<b>BY Statements</b>					
NEEDREP	BY FP26	7.744	0.246	0.182	0.130
NEEDREP	BY FP30	4.583	-0.215	-0.159	-0.102
<b>WITH Statements</b>					
FP26	WITH FP15	4.613	-0.401	-0.401	-0.323
FP30	WITH FP15	7.791	0.663	0.663	0.505
NR6	WITH FP15	4.450	0.124	0.124	0.093
NR16	WITH FP15	5.077	-0.120	-0.120	-0.104
NR16	WITH FP26	5.222	0.115	0.115	0.103
NR20	WITH FP30	5.263	-0.118	-0.118	-0.107



## ESERCIZIO: REALIZZAZIONE DI UN MODELLO DI CFA

**Effettuare un modello di analisi fattoriale confermativa.**

**6 item – 2 fattori**

**Y1-Y3**

→ **F1**

**es4.dat**

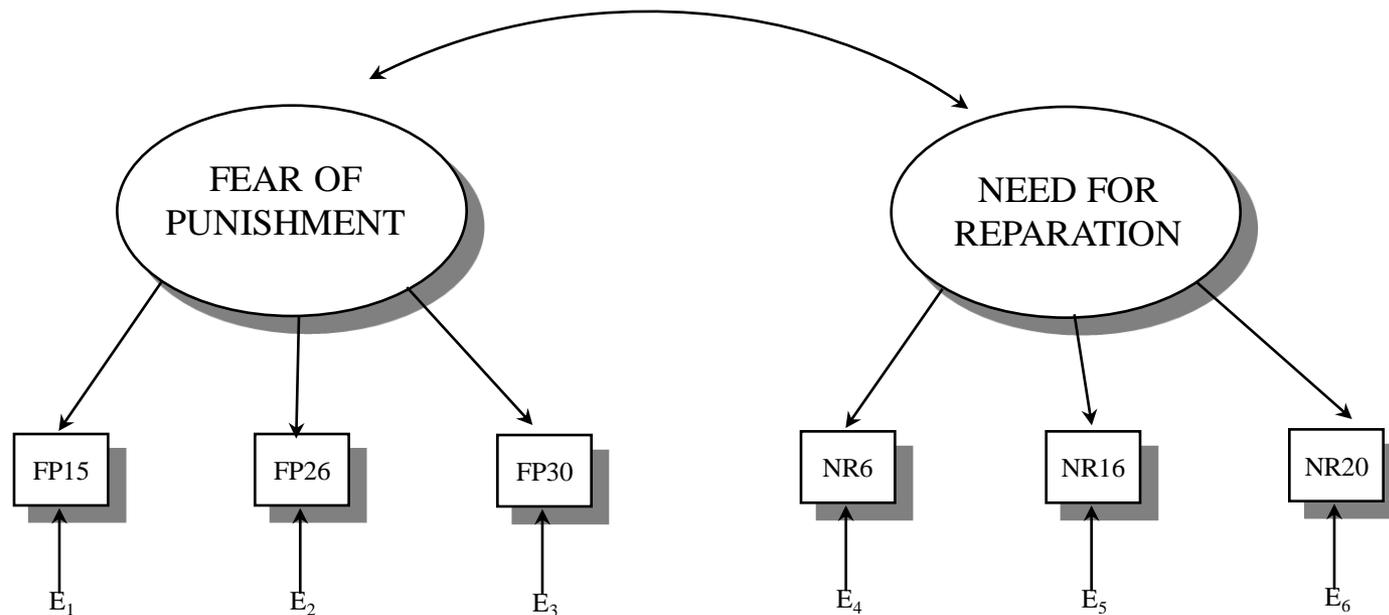
**Y4-Y6**

→ **F2**

**CONFRONTARE IL FIT DEL MODELLO IN CUI I FATTORI SONO  
ORTOGONALI CON IL FIT DEL MODELLO IN CUI I FATTORI SONO  
CORRELATI.**

## Modelli concatenati (*nested*)

Due modelli sono concatenati quando l'uno é ottenibile dall'altro applicando determinate restrizioni. In questo caso i parametri di uno sono un sottoinsieme dei parametri dell'altro.



## Modelli *concatenati* (*nested*)

**Se i modelli sono concatenati é possibile esaminare statisticamente quale modello presenta un indice di adattamento migliore.**

**Infatti la differenza dei loro chi quadrati:**

$$\chi^2_{\text{diff}} = \chi^2_{p-m} - \chi^2_p$$

**si distribuisce ancora come un chi quadrato con gradi di libertà pari a  $df_{p-m} - df_p$ .**

**Se la differenza tra i due chi quadrato ( $\chi^2_{\text{diff}}$ ) é significativa, le restrizioni apportate peggiorano significativamente il modello.**

## Condizioni di applicabilità

### a. Indipendenza delle osservazioni

**Violazione:** Le osservazioni non sono indipendenti

**Test statistici e stime dei parametri non esatti**

**Rimedio:** modelli multilivello

### b. Forma lineare della relazione tra le variabili

### c. Distribuzione normale multivariata delle variabili

**Non linearità, non normalità**

**Sono fenomeni spesso associati. Se presenti i metodi di stima classici (ML) non possono essere applicati perché danno risultati inessatti**

**Rimedio:** metodi di stima "robusti" (correzione di Satorra-Bentler, MLM).



## Condizioni di applicabilità

### d. Livello di misura delle variabili: almeno intervalli

#### Livello di misura "basso"

I metodi sviluppati per variabili continue possono essere utilizzati quando una variabile ha 4 o più categorie "ordinabili" e la distribuzione è sostanzialmente normale. Se le variabili hanno poche categorie, i metodi per variabili continue (es., ML) danno risultati non corretti.

**Rimedio:** Se il numero di categorie é inferiore bisogna ricorrere a metodi di stima alternativi (Mplus costruisce una matrice di correlazioni speciali per il livello di misura).

## Condizioni di applicabilità

**e. La matrice da analizzare é quella delle Covarianze**

**Si analizza la matrice di correlazione**

**Test del chi quadrato ed errori standard non esatti.**

**Rimedio:** convertire la matrice in matrice di covarianza; utilizzare programmi che forniscono stime corrette (es., SEPATH); interpretare con cautela i risultati in assenza di correzioni.



## Condizioni di applicabilità

**f. Il modello testato deve essere specificato a priori rispetto ai dati sui quali viene esaminato**

**Il modello testato viene specificato in funzione dei dati sui quali viene esaminato**

**I risultati non sono generalizzabili e sono soggetti a capitalizzazione sul caso.**

**Rimedio:** mettere in atto procedure di cross-validazione dei risultati.



## Condizioni di applicabilità

**g. I metodi di stima hanno proprietà asintotiche: richiedono campioni molto numerosi**

**Il campione utilizzato è di numerosità modesta**

**Le stime sono inaccurate. Il test di fit è poco potente.**

**Numerosità del campione: come regularsi ?**

## Condizioni di applicabilità

### Numerosità del campione: come regolarsi ?

**Le raccomandazioni relative alla numerosità del campione proposte in letteratura sono varie.**

**Boomsa (1983) ha concluso che campioni di almeno 100 unità rappresentano un requisito minimo ma campioni di più di 200 unità sono preferibili.**

**Geweke e Singleton (1980), al contrario, hanno mostrato che possono essere utilizzati campioni anche più piccoli.**

**Se il campione è inferiore a 100 SS il test di bontà dell'adattamento non segue la distribuzione del chi-2. In generale campioni piccoli tendono a generare risultati instabili.**

## Condizioni di applicabilità

### Numerosità del campione: come regolarci ?

Tanaka (1987) ha evidenziato che il numero assoluto di soggetti non é un problema rilevante, ed ha focalizzato l'attenzione su fattori **come la grandezza del modello (e quindi il numero di parametri da stimare), il numero di variabili osservate e quello di variabili latenti.**

Bentler e Chou (1987) suggeriscono un rapporto minimo di **5 soggetti per ogni parametro libero**, per metodi di stima basati sulla distribuzione normale multivariata, e un rapporto minimo di **10 soggetti per ogni parametro libero**, per metodi di stima "distribution free".

# Condizioni di applicabilità

## Numerosità del campione: come regolarsi ?

**Kline (2005) indica che da 5 a 10 o 20 soggetti per ogni parametro stimato dovrebbe rappresentare un campione sufficiente.**

**A parità di altre condizioni, un numero maggiore di osservazioni determina una **potenza statistica** maggiore.**

**I ricercatori incoraggiano ad utilizzare campioni grandi quando si devono esaminare modelli più complessi (Kim, 2000; McCallum et al., 2006).**

## Condizioni di applicabilità

### Numerosità del campione: come regolarsi ?

**Se non si anticipano problemi nei dati (es., valori mancanti, distribuzioni non normali) si raccomanda un campione minimo di 200 SS per ogni SEM.**

**Tuttavia, se le variabili sono molto attendibili, e molto correlate, gli effetti sono forti e il modello non è troppo complesso, campioni più piccoli possono risultare sufficienti (Bearden, Sharma & Teel 1982; Bollen, 1990).**

**E' stato dimostrato che in questi casi i modelli SEM possono funzionare bene anche con campioni molto piccoli (50 – 100 SS).**



# **I MODELLI DI EQUAZIONI STRUTTURALI: APPLICAZIONI CON MPLUS**

- Analisi Confermativa Vincolata**
- Indici di attendibilità**
- Modelli con variabili osservate**
- SEM completi ("full SEM")**
- L'analisi della mediazione statistica**

# ANALISI FATTORIALE CONFERMATIVA VINCOLATA

**È possibile formulare diverse ipotesi relativamente ai valori assunti dai termini  $\lambda_i$  e  $\theta_i$ .**

**Se si ipotizza che tutti i  $\lambda_i$  abbiano lo stesso valore (cioè,  $\lambda_1 = \lambda_2 = \lambda_3$ ) ma che i  $\theta_i$  possano avere valori differenti, il modello fattoriale specificato viene definito "**tau equivalente**": si ipotizza cioè che ogni indicatore del fattore  $\eta$  abbia soltanto la stessa varianza comune mentre la varianza unica può essere differente per i diversi indicatori.**

## ANALISI FATTORIALE CONFERMATIVA VINCOLATA

Se si ipotizza che tutti i  $\lambda_i$  abbiano lo stesso valore (cioè,  $\lambda_1 = \lambda_2 = \lambda_3$ ) e che anche tutti i  $\theta_i$  abbiano lo stesso valore (cioè,  $\theta_1 = \theta_2 = \theta_3$ ) il modello fattoriale specificato viene definito modello delle “forme parallele”: si ipotizza cioè che ogni indicatore del fattore  $\eta$  abbia la stessa varianza comune e la stessa varianza unica.

Il modello fattoriale in cui sia i  $\lambda_i$  che i  $\theta_i$  possano avere valori differenti, il modello fattoriale specificato viene definito “congenerico”: non viene fatta nessuna assunzione relativamente al valore della varianza comune e della varianza unica, mentre si ipotizza solamente che ogni indicatore del fattore  $\eta$  sia saturo in  $\eta$ .



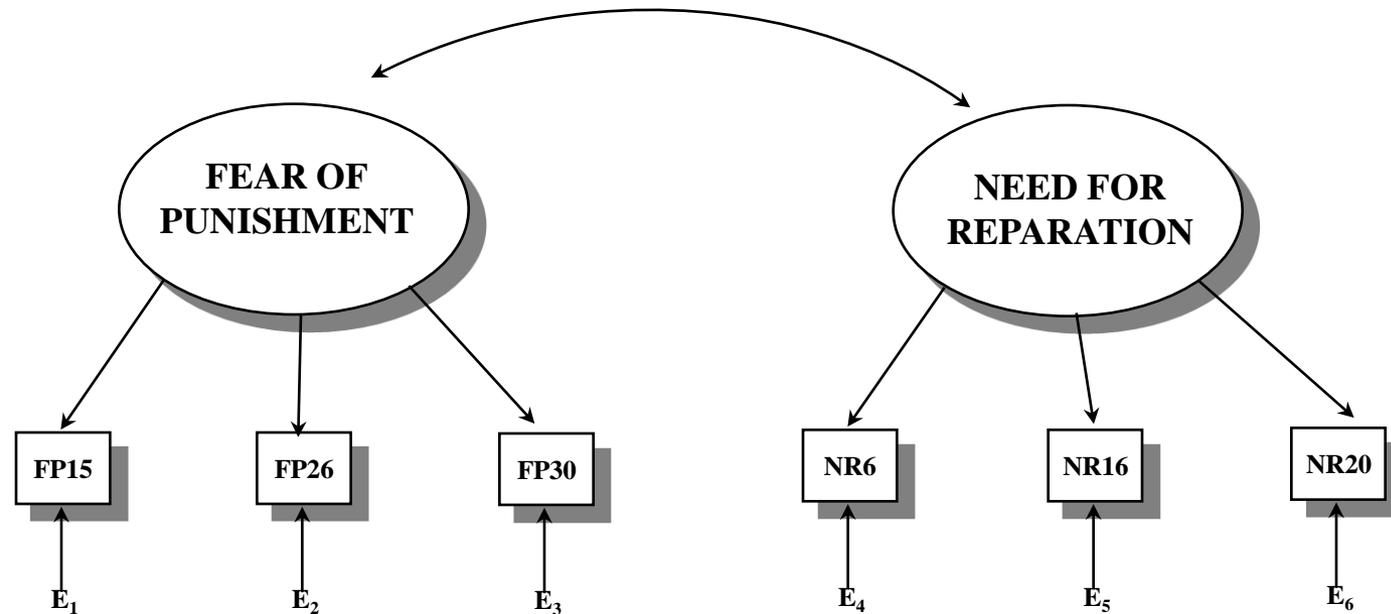
# ANALISI FATTORIALE CONFERMATIVA VINCOLATA

**Questi tre modelli fanno riferimento a tre specifici modelli identificati nella letteratura psicometrica.**

**È possibile sottoporre a verifica empirica ognuna delle tre differenti ipotesi, e vedere quale sia la più consistente rispetto ai dati empirici.**

**Programmi come EQS, LISREL, MPLUS consentono di specificare i diversi vincoli necessari per operationalizzare i diversi modelli, e di ricavare gli indici di bontà dell'adattamento che consentono di optare per un modello piuttosto che per un altro.**

# ANALISI FATTORIALE CONFERMATIVA VINCOLATA



## Modello "tau equivalente"

**Gli indicatori di ciascun fattore comune hanno la stessa saturazione cioè la *stessa varianza comune*.**

## VINCOLI SUI PARAMETRI IN MPLUS

**Come imporre le restrizioni nel modello tau equivalente:**

**FEARPUN BY FP15\* FP26 FP30 (1);**  
**NEEDREP BY NR6\*.5 NR16 NR20 (2);**

**Il numero (1) sta a indicare che le stime delle 3 saturazioni delle variabili FP15 FP26 FP30 vengono vincolate ad essere uguali.**

**Il numero (2) sta a indicare che le stime delle 3 saturazioni delle variabili NR6 NR16 NR20 vengono vincolate ad essere uguali.**

**Vengono specificati due numeri diversi (1) e (2) in modo che le stime siano vincolate all'interno dello stesso fattore ma non attraverso i due fattori.**



# MODELLO "TAU EQUIVALENTE"

**TITLE: CFA**

**CFA\_TAU**

**DATA:**

**FILE IS dati\_efa\_cfa\_grezzi.dat;**

**VARIABLE:**

**NAMES ARE**

**FP10 FP15 FP26 FP30  
NR6 NR16 NR17 NR20 GENDER;**

**USEV ARE**

**FP15 FP26 FP30 NR6 NR16 NR20 ;**

**MISSING ARE ALL (9);**

**MODEL:**

**FEARPUN BY FP15\* FP26 FP30 (1);**

**NEEDREP BY NR6\*.5 NR16 NR20 (2);**

**FEARPUN @1;**

**NEEDREP @1;**

**OUTPUT: STANDARDIZED SAMPSTAT MODINDICES(3.84) TECH1;**



# MODELLO "TAU EQUIVALENTE"

## Chi-Square Test of Model Fit

Value	26.739
Degrees of Freedom	12
P-Value	0.0084

## RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.039	
90 Percent C.I.	0.019	0.059
Probability RMSEA $\leq$ .05	0.811	

## CFI/TLI

CFI	0.977
TLI	0.972

## SRMR (Standardized Root Mean Square Residual)

Value	0.031
-------	-------



# MODELLO "TAU EQUIVALENTE"

## MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>FEARPUN BY</b>				
FP15	0.954	0.036	26.846	0.000
FP26	0.954	0.036	26.846	0.000
FP30	0.954	0.036	26.846	0.000
<b>NEEDREP BY</b>				
NR6	0.713	0.031	22.726	0.000
NR16	0.713	0.031	22.726	0.000
NR20	0.713	0.031	22.726	0.000
<b>NEEDREP WITH FEARPUN</b>	0.301	0.053	5.710	0.000

**NB. I factor loadings sono uguali all'interno dello stesso fattore solo nella soluzione non standardizzata (Estimate) e in quella standardizzata (Std), ma non in quella completamente standardizzata (StdYX), perché le varianze di errore non sono vincolate.**



# MODELLO "TAU EQUIVALENTE"

## STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>FEARPUN BY</b>				
FP15	0.639	0.020	32.440	0.000
FP26	0.672	0.019	34.548	0.000
FP30	0.622	0.020	31.374	0.000
<b>NEEDREP BY</b>				
NR6	0.518	0.021	24.630	0.000
NR16	0.568	0.022	25.396	0.000
NR20	0.608	0.023	26.377	0.000
<b>NEEDREP WITH     FEARPUN</b>	0.301	0.053	5.710	0.000



# MODELLO "TAU EQUIVALENTE"

## Residual Variances

FP15	0.592	0.025	23.526	0.000
FP26	0.548	0.026	20.944	0.000
FP30	0.613	0.025	24.799	0.000
NR6	0.731	0.022	33.536	0.000
NR16	0.678	0.025	26.689	0.000
NR20	0.630	0.028	22.434	0.000

## R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
FP15	0.408	0.025	16.220	0.000
FP26	0.452	0.026	17.274	0.000
FP30	0.387	0.025	15.687	0.000
NR6	0.269	0.022	12.315	0.000
NR16	0.322	0.025	12.698	0.000
NR20	0.370	0.028	13.188	0.000



# MODELLO "TAU EQUIVALENTE"

M.I.            E.P.C.    Std E.P.C.    StdYX E.P.C.

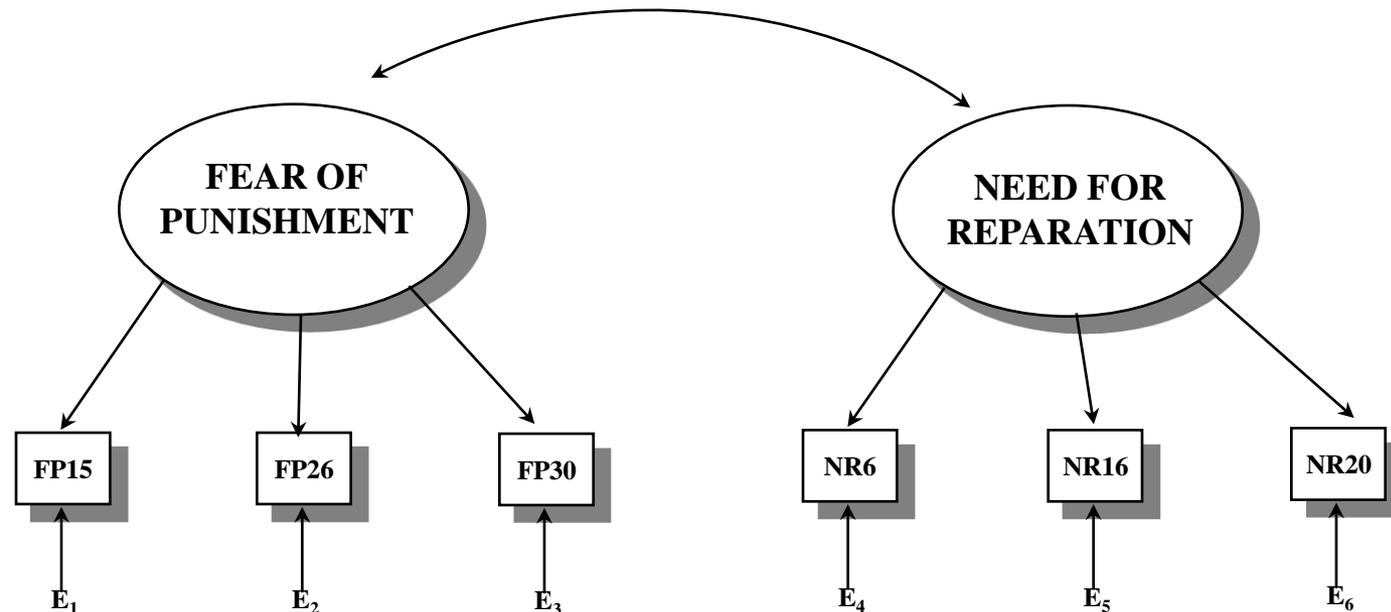
**Cambiamento nel chi-quadrato determinato dalla stima di parametri fissati a zero**

WITH Statements

FP26	WITH FP15	5.583	-0.168	-0.168	-0.139
FP30	WITH FP15	7.262	0.194	0.194	0.141
NR6	WITH FP15	4.669	0.127	0.127	0.094
NR16	WITH FP15	4.939	-0.118	-0.118	-0.099
NR16	WITH FP26	4.822	0.110	0.110	0.101
NR20	WITH FP30	5.219	-0.116	-0.116	-0.104

**Nessuno dei MI relativi al cambiamento nel chi-quadrato determinato dal rilasciamento dei vincoli risulta significativo**

# ANALISI FATTORIALE CONFERMATIVA VINCOLATA



## Modello "forme parallele"

**Gli indicatori di ciascun fattore comune hanno la stessa saturazione cioè la *stessa varianza comune* e la stessa varianza residua (errore di misurazione).**

**Il risultato di questo doppio vincolo è che gli indicatori hanno la stessa attendibilità.**

## VINCOLI SUI PARAMETRI IN MPLUS

**Come imporre le restrizioni nel modello forme parallele:**

**FEARPUN BY FP15\* FP26 FP30 (1);**

**NEEDREP BY NR6\*.5 NR16 NR20 (2);**

**FP15 FP26 FP30 (3);**

**NR6 NR16 NR20 (4);**

**I vincoli indicati con i numeri (1) e (2) sono quelli già imposti nel modello tau equivalente.**

**Il numero (3) sta a indicare che le stime delle 3 varianze residue delle variabili FP15 FP26 FP30 vengono vincolate ad essere uguali.**

**Il numero (4) sta a indicare che le stime delle 3 varianze residue delle variabili NR6 NR16 NR20 vengono vincolate ad essere uguali.**



# MODELLO "FORME PARALLELE"

**TITLE: CFA**

**CFA PARALL**

**DATA:**

**FILE IS dati\_efa\_cfa\_grezzi.dat;**

**VARIABLE:**

**NAMES ARE**

**FP10 FP15 FP26 FP30  
NR6 NR16 NR17 NR20 GENDER;**

**USEV ARE**

**FP15 FP26 FP30 NR6 NR16 NR20 ;**

**MISSING ARE ALL (9);**

**MODEL:**

**FEARPUN BY FP15\* FP26 FP30 (1);  
NEEDREP BY NR6\*.5 NR16 NR20 (2);**

**FEARPUN @1;**

**NEEDREP @1;**

**FP15 FP26 FP30 (3);**

**NR6 NR16 NR20 (4);**

**OUTPUT: STANDARDIZED SAMPSTAT MODINDICES(3.84) TECH1;**



# MODELLO "FORME PARALLELE"

## Chi-Square Test of Model Fit

Value	62.331
Degrees of Freedom	16
P-Value	0.0000

## RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.059	
90 Percent C.I.	0.044	0.075
Probability RMSEA $\leq$ .05	0.146	

## CFI/TLI

CFI	0.929
TLI	0.933

## SRMR (Standardized Root Mean Square Residual)

Value	0.114
-------	-------



# MODELLO "FORME PARALLELE"

## MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>FEARPUN BY</b>				
FP15	0.961	0.036	26.921	0.000
FP26	0.961	0.036	26.921	0.000
FP30	0.961	0.036	26.921	0.000
<b>NEEDREP BY</b>				
NR6	0.714	0.032	22.469	0.000
NR16	0.714	0.032	22.469	0.000
NR20	0.714	0.032	22.469	0.000
<b>NEEDREP WITH</b>				
FEARPUN	0.304	0.053	5.744	0.000
<b>Residual Variances</b>				
FP15	1.281	0.045	28.486	0.000
FP26	1.281	0.045	28.486	0.000
FP30	1.281	0.045	28.486	0.000
NR6	1.105	0.039	28.559	0.000
NR16	1.105	0.039	28.559	0.000
NR20	1.105	0.039	28.559	0.000

# MODELLO "FORME PARALLELE"

I factor loadings sono uguali all'interno dello stesso fattore **anche** nella soluzione completamente standardizzata, perché le varianze di errore sono vincolate.

STDYX Standardization		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>FEARPUN</b>	<b>BY</b>				
FP15		0.647	0.017	38.715	0.000
FP26		0.647	0.017	38.715	0.000
FP30		0.647	0.017	38.715	0.000
<b>NEEDREP</b>	<b>BY</b>				
NR6		0.562	0.020	27.954	0.000
NR16		0.562	0.020	27.954	0.000
NR20		0.562	0.020	27.954	0.000
<b>NEEDREP</b>	<b>WITH</b>				
FEARPUN		0.304	0.053	5.744	0.000
<b>Residual Variances</b>					
FP15		0.581	0.022	26.867	0.000
FP26		0.581	0.022	26.867	0.000
FP30		0.581	0.022	26.867	0.000
NR6		0.684	0.023	30.277	0.000
NR16		0.684	0.023	30.277	0.000
NR20		0.684	0.023	30.277	0.000



# MODELLO "FORME PARALLELE"

## MODEL MODIFICATION INDICES

Minimum M.I. value for printing the modification index 3.840

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
<b>BY Statements</b>				
FEARPUN BY FP26	9.113	-0.118	-0.118	-0.080
FEARPUN BY FP30	7.249	0.105	0.105	0.071
FEARPUN BY NR6	4.919	0.121	0.121	0.095
NEEDREP BY NR6	12.841	0.140	0.140	0.110
NEEDREP BY NR20	12.077	-0.136	-0.136	-0.107

## Variances/Residual Variances

FP26	6.134	-0.173	-0.173	-0.079
FP30	4.710	0.152	0.152	0.069
NR6	23.478	0.274	0.274	0.170
NR20	17.490	-0.237	-0.237	-0.147

**Cambiamento nel chi-quadrato determinato dal  
rilasciamento dei vincoli**



# MODELLO "FORME PARALLELE"

## MODEL MODIFICATION INDICES

Minimum M.I. value for printing the modification index 3.840

		M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
WITH Statements					
NR6	WITH FP15	4.182	0.109	0.109	0.092
NR16	WITH FP15	5.133	-0.121	-0.121	-0.101
NR16	WITH FP26	5.388	0.124	0.124	0.104
NR20	WITH FP30	4.303	-0.111	-0.111	-0.093
NR20	WITH NR16	4.207	0.108	0.108	0.098

**Cambiamento nel chi-quadrato determinato dal  
rilasciamento dei parametri fissati a zero**

## CONFRONTO TRA I CHI<sup>2</sup> DEI MODELLI

**MODELLO CONGENERICO:  $\chi^2(8) = 21.62, p < .01$**

**MODELLO TAU-EQUIVALENTE:  $\chi^2(12) = 26.74, p < .01$**

**MODELLO FORME PARALLELE:  $\chi^2(16) = 62.33, p < .001$**

**TAU-EQUIVALENTE – CONGENERICO:**

**$\chi^2_{\text{diff}}(4) = 5.06, p = .28$**

**FORME PARALLELE - TAU-EQUIVALENTE:**

**$\chi^2_{\text{diff}}(4) = 35.59, p < .001$**

**[CHI 2 DIFF.xls](#)**



## **ESERCIZIO 4: REALIZZAZIONE DI UN MODELLO DI UN MODELLO CFA TAU-EQUIVALENTE E DELLE FORME PARALLELE**

**Realizzare il modello CFA con i dati del file es4.dat**

**Verificare sugli stessi dati i modelli congenerico, tau-equivalente e delle forme parallele, applicando le opportune restrizioni.**

**Confrontare poi i modelli congenerico, tau equivalente e delle forme parallele, sulla base dei loro  $\chi^2$ .**



## **ANALISI FATTORIALE CONFERMATIVA: INDICI DI ATTENDIBILITA' (Fornell e Larcker, 1981)**

**Sia  $x_i$  un indicatore almeno congenerico di una variabile latente  $\xi$ , ovvero  $x_i = \lambda_i \xi + \delta_i$**

**Sia  $\theta_i$  la varianza di errore, ovvero  $E(\delta_i \delta_i) = \theta_i$**

**Si supponga che i termini  $\varepsilon_i$  dei diversi indicatori almeno congenerici di  $\xi$  abbiano media zero e non siano correlati tra di loro né con la variabile latente.**

**Per comodità si assuma che la varianza della variabile latente sia uguale a 1.**

**Allora si possono definire i seguenti indici:**

## ANALISI FATTORIALE CONFERMATIVA: INDICI DI ATTENDIBILITA' (Fornell e Larcker, 1981)

a) **Attendibilità** del singolo indicatore: è data dal rapporto tra la varianza spiegata dal fattore comune e la varianza totale della variabile:

$$\rho_i = \frac{\lambda_i^2}{\lambda_i^2 + \theta_i}$$

Questo indice varia da 0 a 1: vanno considerati come adeguati valori maggiori/uguali a .3.

## ANALISI FATTORIALE CONFERMATIVA: INDICI DI ATTENDIBILITA' (Fornell e Larcker, 1981)

a) **Attendibilità** del singolo indicatore: se il singolo indicatore è influenzato da più di un fattore la formula è un po' più complessa. Ad esempio nel caso di due fattori esse diventa:

$$\rho_i = \frac{\lambda_{i1}^2 + \lambda_{i2}^2 + 2\lambda_{i1}\lambda_{i2}\phi_{21}}{\lambda_{i1}^2 + \lambda_{i2}^2 + 2\lambda_{i1}\lambda_{i2}\phi_{21} + \theta_i}$$

**Si interpreta sempre come rapporto tra la varianza spiegata dai fattori comuni e la varianza totale della variabile**

## ANALISI FATTORIALE CONFERMATIVA: INDICI DI ATTENDIBILITA'

b) **Attendibilità composta** del fattore (simile al **coefficiente omega**): è data dal rapporto tra il quadrato della somma delle saturazioni nel fattore comune e questo stesso numero più la proporzione di varianza totale attribuibile alla unicità delle variabili:

$$\rho_c = \frac{(\sum_i \lambda_i)^2}{(\sum_i \lambda_i)^2 + \sum_i \theta_i}$$

Questo indice varia da 0 a 1: vanno considerati come adeguati valori maggiori/uguali a .6.

## ANALISI FATTORIALE CONFERMATIVA: INDICI DI ATTENDIBILITA'

**c) Varianza media spiegata dal fattore (AVE, Average Extracted Variance):** è data dal rapporto tra la somma della varianza di ogni variabile spiegata dal fattore comune e la varianza totale delle variabili, cioè:

$$\rho_v = \frac{(\sum_i \lambda_i^2)}{(\sum_i \lambda_i^2) + \sum_i \theta_i}$$

**Questo indice varia da 0 a 1. Anche se alcuni ricercatori indicano un valore maggiore/uguale a .50 come ideale, altri considerano il valore di tale indice come "accessorio" rispetto agli indici di fit (globali e dei singoli parametri).**



# ANALISI FATTORIALE CONFERMATIVA: INDICI DI ATTENDIBILITA'

	<b>Attendibilità Indicatore</b>	<b>Attendibilità Composita</b>	<b>Varianza Spiegata Media (AVE)</b>
<b>FP15</b>	<b>0,418</b>	<b>0,685</b>	<b>0,42</b>
<b>FP26</b>	<b>0,391</b>		
<b>FP30</b>	<b>0,453</b>		
<b>NR6</b>	<b>0,287</b>	<b>0,586</b>	<b>0,32</b>
<b>NR16</b>	<b>0,345</b>		
<b>NR20</b>	<b>0,330</b>		

**Reliability.xls**

# ANALISI DI ATTENDIBILITA' CON SPSS

## efa\_dati.sav

efa\_dati.sav [Dataset1] - IBM SPSS Statistics Editor dei dati

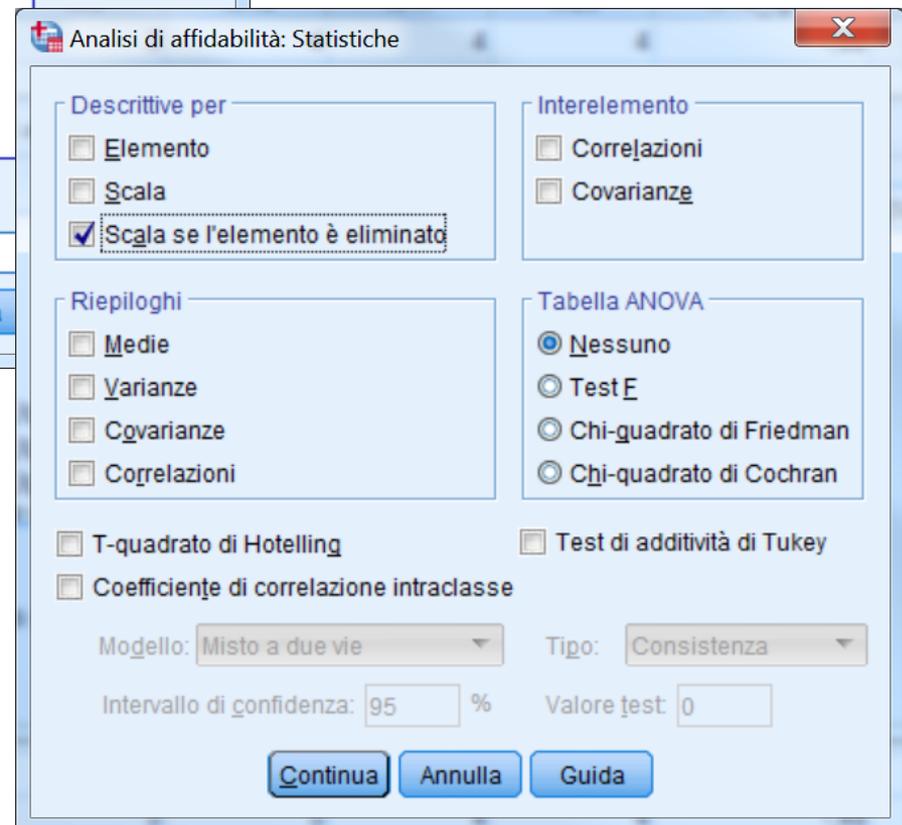
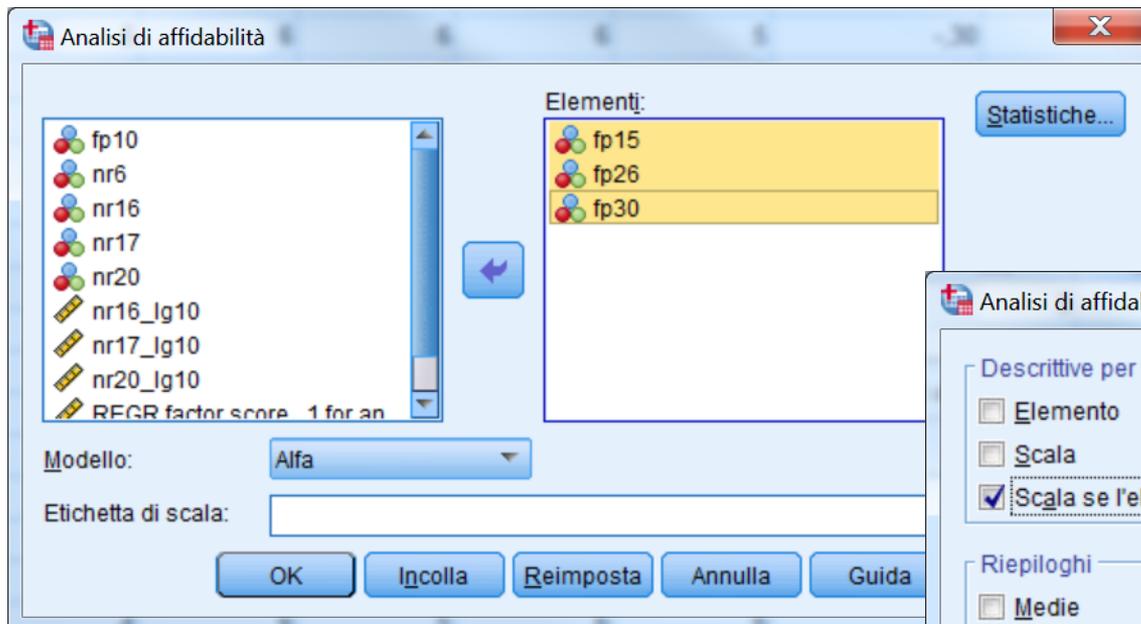
File Modifica Visualizza Dati Trasforma **Analizza** Direct Marketing Grafici Programmi di utilità Finestra Guida

Report  
 Statistiche descrittive  
 Tabelle personalizzate  
 Confronta medie  
 Modello lineare generale  
 Modelli lineari generalizzati  
 Modelli misti  
 Correlazione  
 Regressione  
 Loglineare  
 Reti neurali  
 Classifica  
 Riduzione delle dimensioni...  
**Scala**  
 Test non parametrici  
 Previsioni  
 Sopravvivenza  
 Risposta multipla  
 Analisi valori mancanti...  
 Assegnazione multipla  
 Campioni complessi

	fp10	fp15	fp26	nr17	nr20	nr16_lg10	nr17_
1	3	4		4	4	-,60	
2	1	4		1	5	-,60	
3	2	6		6	5	-,30	
4	1	1		4	6	-,30	
5	4	6		4	5	-,30	
6	1	1		6	6	-,48	
7	3	5		2	6	-,78	
8	2	4		6	5	-,60	
9	3	3		4	4	-,70	
10	1	6		6	6	-,30	
11	4	3					
12	5	3					
13	4	4					
14	2	1					
15	1	1					
16	6	3					
17	1	1					
18	1	2					

Analisi di affidabilità...  
 Unfolding multidimensionale (PREFSCAL)...  
 Scaling multidimensionale (PROXSCAL)...  
 Scaling multidimensionale (ALSCAL)...

# ANALISI DI ATTENDIBILITA' CON SPSS





# ANALISI DI ATTENDIBILITA' CON SPSS

## Statistiche di affidabilità

Alpha di Cronbach	N. di elementi
,682	3

## Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Alpha di Cronbach se viene eliminato l'elemento
fp15	6,83	6,163	,497	,587
fp26	6,32	6,697	,476	,615
fp30	6,85	5,789	,518	,560

# ANALISI DI ATTENDIBILITA' CON SPSS

The image shows two overlapping dialog boxes from the SPSS software interface for conducting a Reliability Analysis.

**Analisi di affidabilità (Main Dialog):**

- Elementi:** nr6, nr16, nr20
- Modello:** Alfa
- Etichetta di scala:** (empty)
- Buttons:** OK, Incolla, Reimposta, Annulla

**Analisi di affidabilità: Statistiche (Sub-dialog):**

- Descrittive per:**
  - Elemento
  - Scala
  - Scala se l'elemento è eliminato
- Interelemento:**
  - Correlazioni
  - Covarianze
- Riepiloghi:**
  - Medie
  - Varianze
  - Covarianze
  - Correlazioni
- Tabella ANOVA:**
  - Nessuno
  - Test E
  - Chi-quadrato di Friedman
  - Chi-quadrato di Cochran
- Altri test:**
  - T-quadrato di Hotelling
  - Test di addittività di Tukey
  - Coefficiente di correlazione intraclassa
- Modello:** Misto a due vie
- Tipo:** Consistenza
- Intervallo di confidenza:** 95 %
- Valore test:** 0
- Buttons:** Continua, Annulla, Guida



# ANALISI DI ATTENDIBILITA' CON SPSS

## Statistiche di affidabilità

Alpha di Cronbach	N. di elementi
,579	3

## Statistiche elemento-totale

	Media scala se viene eliminato l'elemento	Varianza scala se viene eliminato l'elemento	Correlazione elemento-totale corretta	Alpha di Cronbach se viene eliminato l'elemento
nr6	9,71	3,953	,371	,511
nr16	9,41	4,213	,403	,455
nr20	9,05	4,565	,396	,471



## **ESERCIZIO 5: INDICI DI ATTENDIBILITA' RICAVATI DALLA SOLUZIONE CFA**

**Utilizzando i risultati del modello realizzato nell'esercizio 4:**

**Calcolare gli indici a, b e c per la soluzione congenerica.**

**Effettuare l'analisi di attendibilità con SPSS e confrontare la soluzione con gli indici ricavati tramite Mplus.**

# MODELLI CON SOLE VARIABILI OSSERVATE

**Il modello matematico si semplifica in:**

$$\eta = \alpha + B\eta + \zeta$$

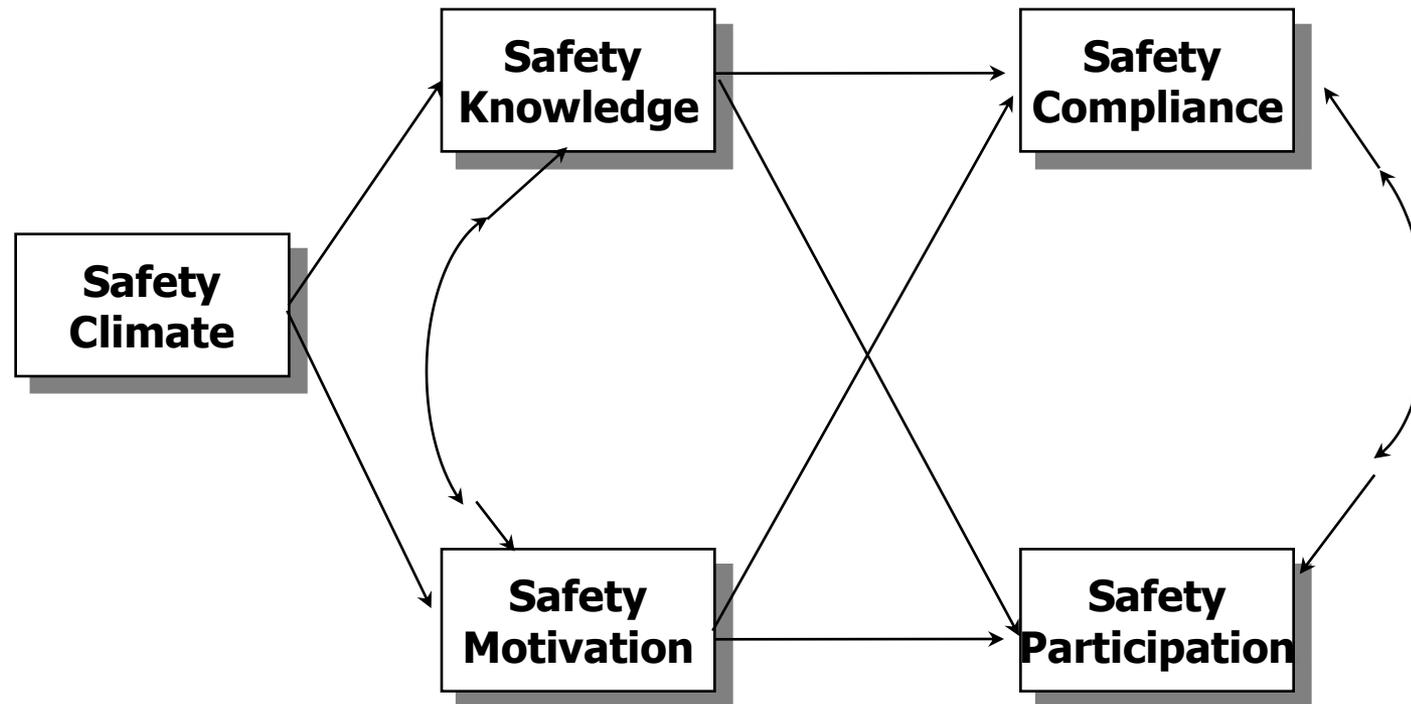
**$\eta$  è il vettore delle variabili dipendenti. Ogni  $\eta$  è misurata da una sola  $y$  la cui saturazione  $\lambda$  è fissata a **1** e la cui varianza d'errore  $\theta$  è fissata a **0**:  $y=1\eta + \varepsilon$ ;  $\theta=0$**

**$\zeta$  sono gli errori di specificazione dell'equazione le cui varianze sono nella matrice  $\Psi$**

**$\alpha$  sono le intercette delle  $y$  (solitamente uguali a **0**)**

**$x$  sono le variabili osservate indipendenti**

# MODELLI CON VARIABILI OSSERVATE





# MODELLI CON VARIABILI OSSERVATE

**TITLE: neal 2011**

[neal 2011 item path](#)

**DATA:**

**FILE IS ITALY\_2012.DAT;  
TYPE IS INDIVIDUAL;**

**VARIABLE:**

**NAMES ARE .....**

**usev are**

**saf\_cli saf\_knw ind\_mot SAF\_COMP SAF\_PART ;**

**define:**

**saf\_cli = SAFCLI1 +SAFCLI2 +SAFCLI3 +SAFCLI4+  
SAFCLI5 +SAFCLI6 +SAFCLI7 +SAFCLI8 +SAFCLI9+  
SAFCLI10 +SAFCLI11 +SAFCLI12 +SAFCLI13+  
SAFCLI14 +SAFCLI15 +SAFCLI16 ;  
saf\_knw =KN\_21 +KN\_22 +KN\_23 +KN\_24 ;  
ind\_mot =MOT\_25 +MOT\_26 +MOT\_27 +MOT\_28;  
SAF\_COMP =COMP\_29+ COMP\_30 +COMP\_31 +COMP\_32;  
SAF\_PART =PART\_33 +PART\_34 +PART\_35 +PART\_36 ;**



# MODELLI CON VARIABILI OSSERVATE

**TITLE: neal 2011**

.....

**analysis:  
estimator=MLMV;**

**MODEL:**

**saf\_knw on saf\_cli;  
ind\_mot on saf\_cli;  
SAF\_COMP ON saf\_knw ind\_mot;  
SAF\_PART ON saf\_knw ind\_mot;**

**saf\_knw WITH ind\_mot;  
SAF\_COMP WITH SAF\_PART;**

**OUTPUT: STANDARDIZED SAMPSTAT MODINDICES(3.84);**

## MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

**Se un costrutto è misurato con una scala composta da più item, si definiscono diversi livelli di aggregazione degli indicatori della variabile latente.**

***Disaggregazione Totale. Ogni item è usato come separato indicatore del costrutto.*** Rappresenta il livello di analisi più dettagliato per l'analisi dei SEM (Bagozzi & Heatherington 1994). Si possono esaminare le proprietà psicometriche di ciascun singolo item. Se il numero di item è molto elevato può risultare una eccessiva potenza della verifica, per cui l'ipotesi nulla viene rifiutata quando invece non dovrebbe esserlo (Bagozzi & Heatherington 1994).

## MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

***Disaggregazione parziale.* Per ogni variabile latente vengono costruiti diversi compositi di item aggregati. Questi compositi vengono chiamati "parcels": i parcels rappresentano il risultato della somma/agggregazione di più item (Bentler & Wu 1995; Dabholkar, Thorpe et al. 1996).**

**Il costrutto viene modellato come una variabile latente misurata dai parcels e non dagli item. I parcels possono essere individuati in modo indipendente dal loro contenuto, e quindi non vanno interpretati direttamente ma solo come indicatori del costrutto. Viceversa i parcels possono essere creati individuando delle "sottodimensioni" (o "facets") del costrutto più globale.**



# MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

La **disaggregazione parziale** può risultare particolarmente utile per la definizione di modelli complessi, poiché:

- riduce l'errore di misurazione
- consente di ottenere stime più stabili riducendo il numero di parametri da stimare
- migliora l'approssimazione alla distribuzione normale

(Bagozzi & Heatherington 1994; Baumgartner & Homburg 1996; Dabholkar, Thorpe et al. 1996).



## MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

***Aggregazione Totale.* Utilizza singole misure per ciascun costrutto, ognuna risultato della aggregazione/somma degli item che compongono la scala di misura del costrutto (Bagozzi & Heatherington 1994).**

**Il principale vantaggio di questo approccio è nella semplicità e nella capacità di catturare gli elementi essenziali del modello concettuale (Bagozzi & Heatherington 1994; Baumgartner & Homburg 1996).**

**Il principale svantaggio è nel fatto che non viene modellato l'errore di misurazione delle variabili.**



# MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

**Esistono due principali tipi di modelli di aggregazione totale:**

**a) Modelli con sole variabili osservate, in cui l'errore di misurazione delle variabili osservate (somma di item) viene fissato a zero.**

**b) Modelli in cui l'errore di misurazione delle variabili osservate (somma di item) viene fissato ad un valore maggiore di 0. Si tratta dei modelli con "single indicator latent variable", o con "reliability correction".**

# MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

## Reliability correction

**In questi modelli si utilizza una stima dell'attendibilità della scala composta dai diversi item che vengono aggregati per misurare il costrutto. Ad esempio si può considerare il valore dell'alfa di Cronbach come stima dell'attendibilità della scala. Quindi si utilizza il valore  $(1-\text{alfa})$  come stima della inattendibilità della scala. Questo valore viene moltiplicato per la varianza della scala  $S^2$  in modo tale da ottenere una stima della varianza unica della scala. Nel modello il valore della varianza residua della scala viene dunque fissato a  $(1-\text{alfa}) * S^2$ .**



# MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

## Reliability correction

**Questa tecnica non stima la varianza unica della variabile come parte del modello, ma non assume neanche che essa sia uguale a zero (come nella “path analysis”). Pertanto, con questa tecnica è possibile ottenere delle stime dei parametri (beta, ecc.) del modello **più corretta** (Stephenson & Holbert, 2003). Rispetto ai modelli della disaggregazione (totale o parziale) tuttavia il modello della reliability correction **è meno preciso** perché la stima della varianza unica sulla quale si basa contiene sia la varianza dovuta all'errore di misurazione sia la varianza specifica della variabile.**



# MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

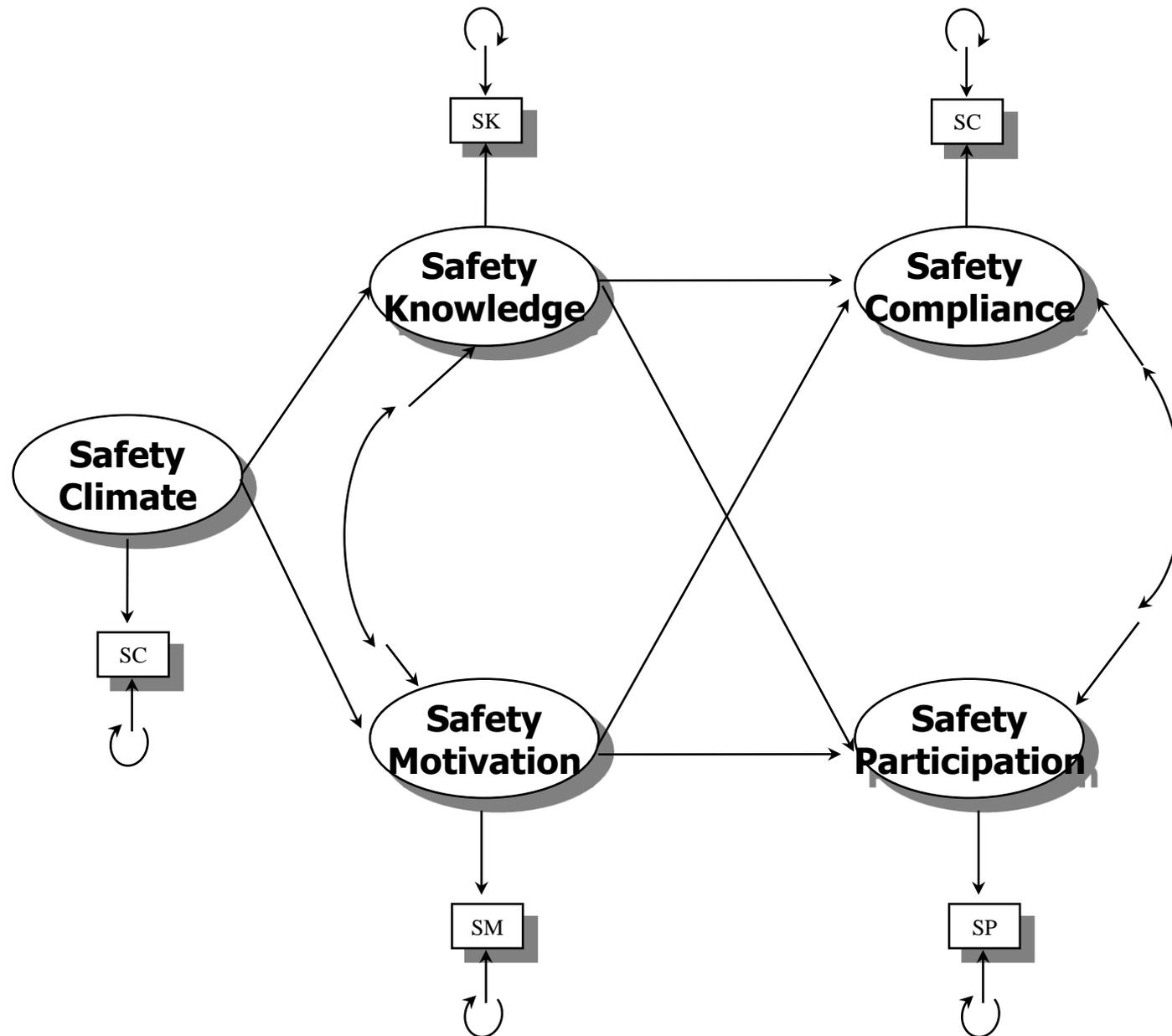
**Ovunque possibile è bene specificare modelli con variabili latenti (disaggregazione totale o parziale) piuttosto che modelli con variabili osservate (aggregazione totale), poiché in questi ultimi si assume che le variabili siano misurate senza errore o si fissa l'errore ad un valore derivante da una stima (Kline, 1998). Questa assunzione non viene fatta quando invece si usano variabili latenti.**



## MODELLI CON SOLE VARIABILI OSSERVATE E LIVELLI DI AGGREGAZIONE

**Nei modelli con variabili latenti la varianza unica viene stimata per ogni indicatore di ogni variabile latenti, e quindi la varianza di errore viene stimata come parte del modello. Utilizzare un modello che assume che le variabili siano prive di errore quando in realtà non lo sono può compromettere le stime dei parametri del modello. Nella maggior parte delle situazioni questa compromissione porta a una sottostima delle relazioni tra le variabili. L'uso di variabili latenti in cui l'errore di misura viene esplicitamente modellato porta invece a stime corrette di tali parametri.**

# MODELLO "RELIABILITY CORRECTION"



# MODELLO "RELIABILITY CORRECTION"

## Stime per la correzione della varianza residua

	<b>alfa</b>	<b>s<sup>2</sup></b>	<b>(1-alfa)s<sup>2</sup></b>
<b>CLIMATE</b>	<b>0,97</b>	<b>458,174</b>	<b>13,745</b>
<b>KNOWL</b>	<b>0,909</b>	<b>23,124</b>	<b>2,104</b>
<b>MOTIV</b>	<b>0,925</b>	<b>21,440</b>	<b>1,608</b>
<b>COMPLI</b>	<b>0,921</b>	<b>25,153</b>	<b>1,987</b>
<b>PARTICI2</b>	<b>0,836</b>	<b>24,893</b>	<b>4,082</b>

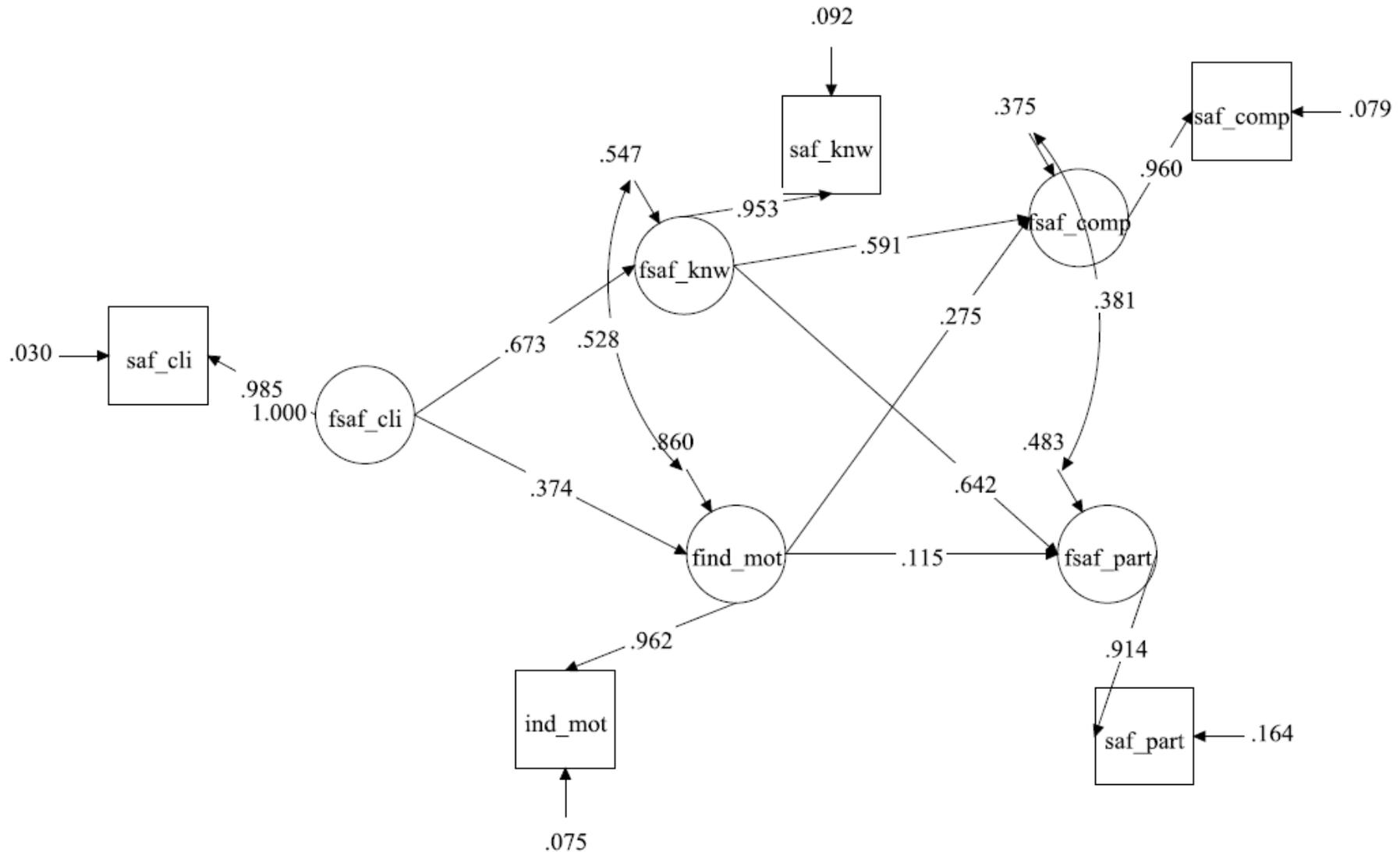
# MODELLO "RELIABILITY CORRECTION"

**MODEL:**

[neal 2011 item rel corr](#)

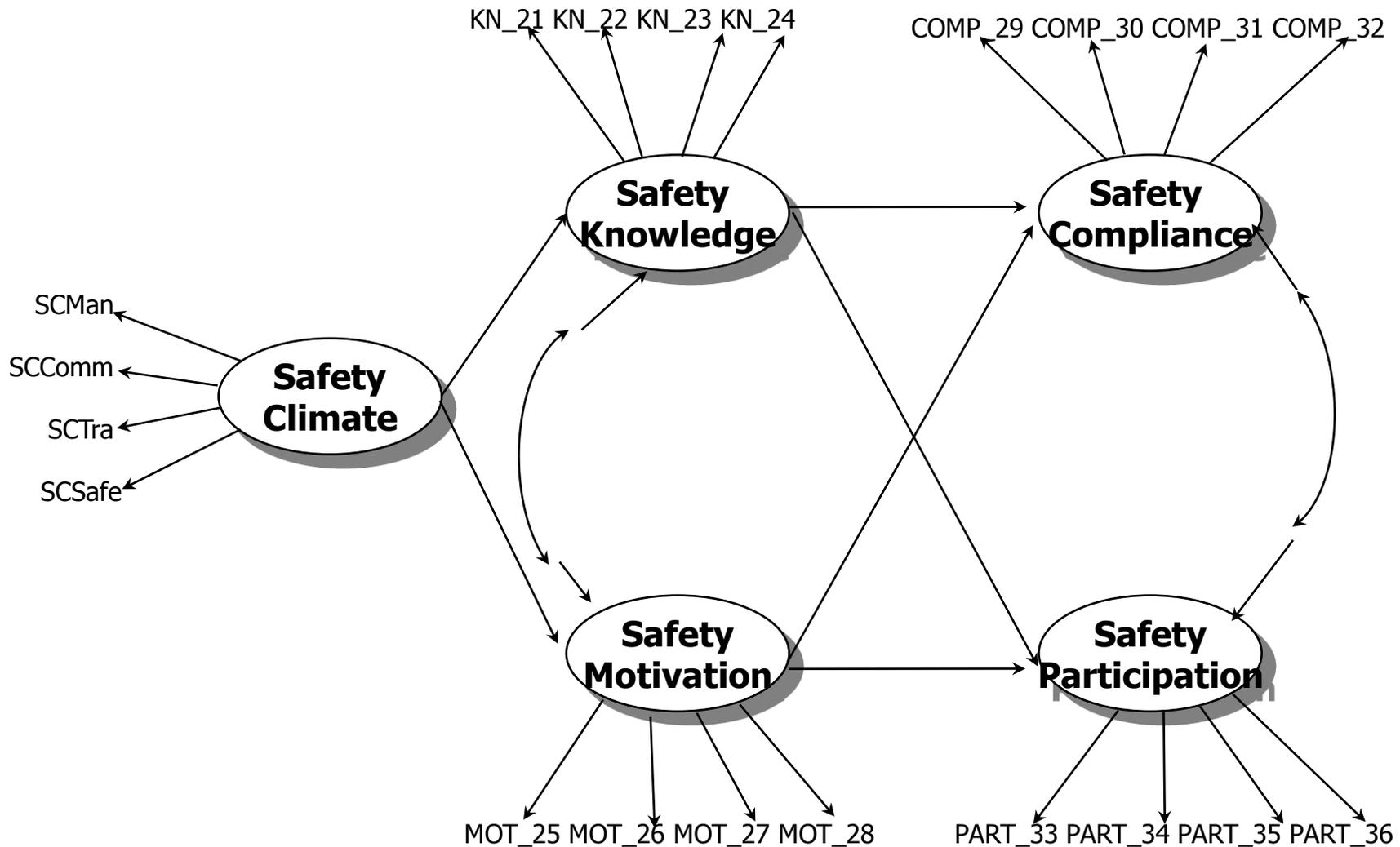
```
Fsaf_cli BY saf_cli;  
saf_cli @13.745;  
Fsaf_knw BY saf_knw ;  
saf_knw @2.104;  
Find_mot BY ind_mot ;  
ind_mot @1.608;  
FSAF_COMP BY SAF_COMP ;  
SAF_COMP @1.987;  
FSAF_PART BY SAF_PART ;  
SAF_PART @4.082;  
  
Fsaf_knw on Fsaf_cli;  
Find_mot on Fsaf_cli;  
FSAF_COMP ON Fsaf_knw Find_mot;  
FSAF_PART ON Fsaf_knw Find_mot;  
Fsaf_knw WITH Find_mot;  
FSAF_COMP WITH FSAF_PART;
```

# MODELLO "RELIABILITY CORRECTION"



[neal\\_2011\\_item\\_rel\\_corr](#)

# MODELLO "PARTIAL AGGREGATION"





# MODELLO "PARTIAL AGGREGATION"

usev are

KN\_21 KN\_22 KN\_23 KN\_24 MOT\_25 MOT\_26 MOT\_27 MOT\_28  
 COMP\_29 COMP\_30 COMP\_31 COMP\_32 PART\_33 PART\_34  
 PART\_35 PART\_36 SCMan SComm SCTra SCSafe;

define:

SCMan = SAFCLI1 +SAFCLI2 +SAFCLI3 +SAFCLI4;

SComm =SAFCLI5 +SAFCLI6 +SAFCLI7 +SAFCLI8 +SAFCLI9;

SCTra =SAFCLI10 +SAFCLI11 +SAFCLI12 +SAFCLI13;

SCSafe =SAFCLI14 +SAFCLI15 +SAFCLI16 ;

analysis: estimator=MLMV;

[neal 2011 part aggreg](#)

MODEL:

saf\_cli by SCMan SComm SCTra SCSafe;

saf\_knw by KN\_21 KN\_22 KN\_23 KN\_24 ;

ind\_mot by MOT\_25 MOT\_26 MOT\_27 MOT\_28;

SAF\_COMP by COMP\_29 COMP\_30 COMP\_31 COMP\_32;

SAF\_PART by PART\_33 PART\_34 PART\_35 PART\_36 ;

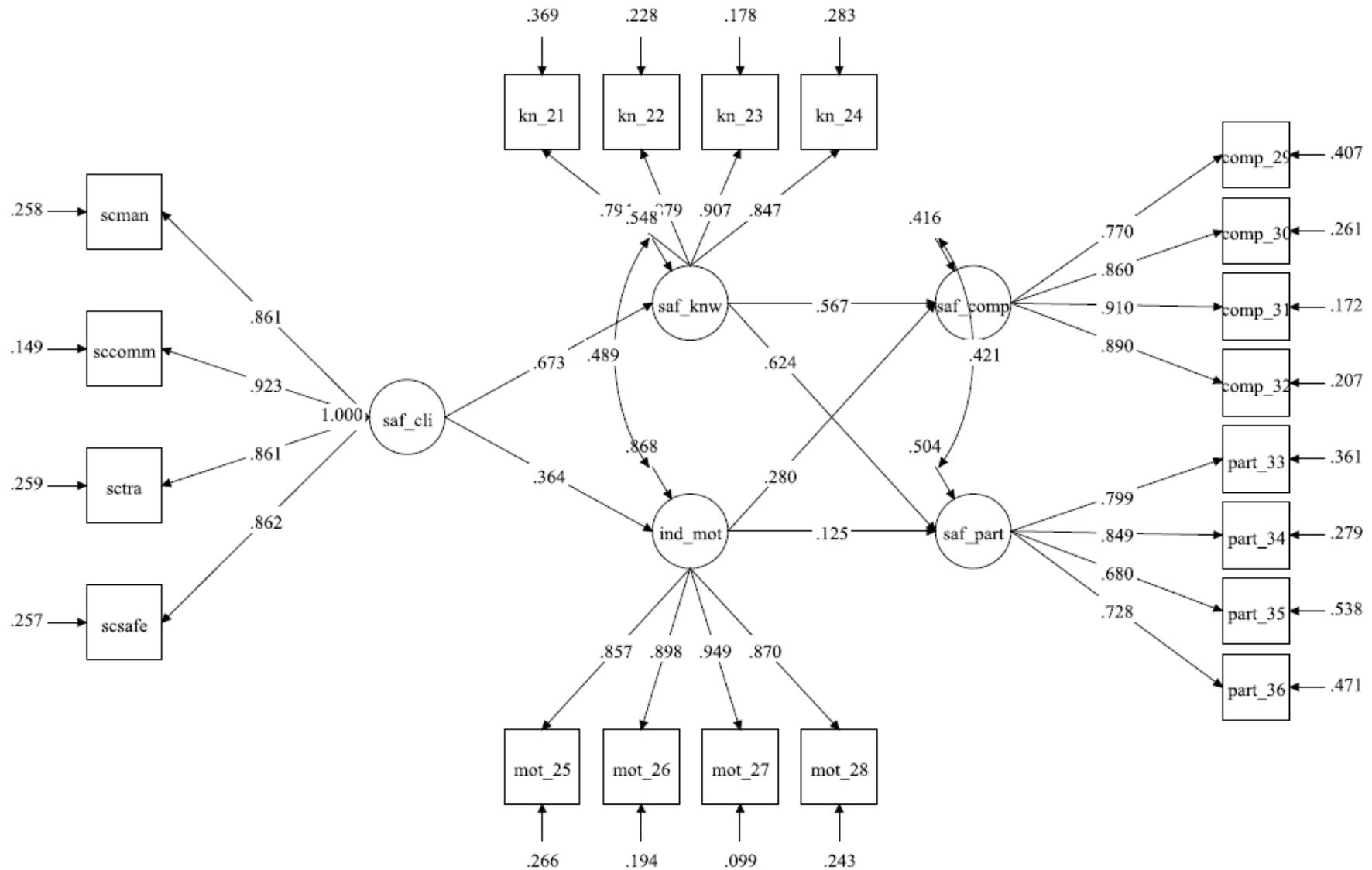
saf\_knw on saf\_cli; ind\_mot on saf\_cli;

SAF\_COMP ON saf\_knw ind\_mot;

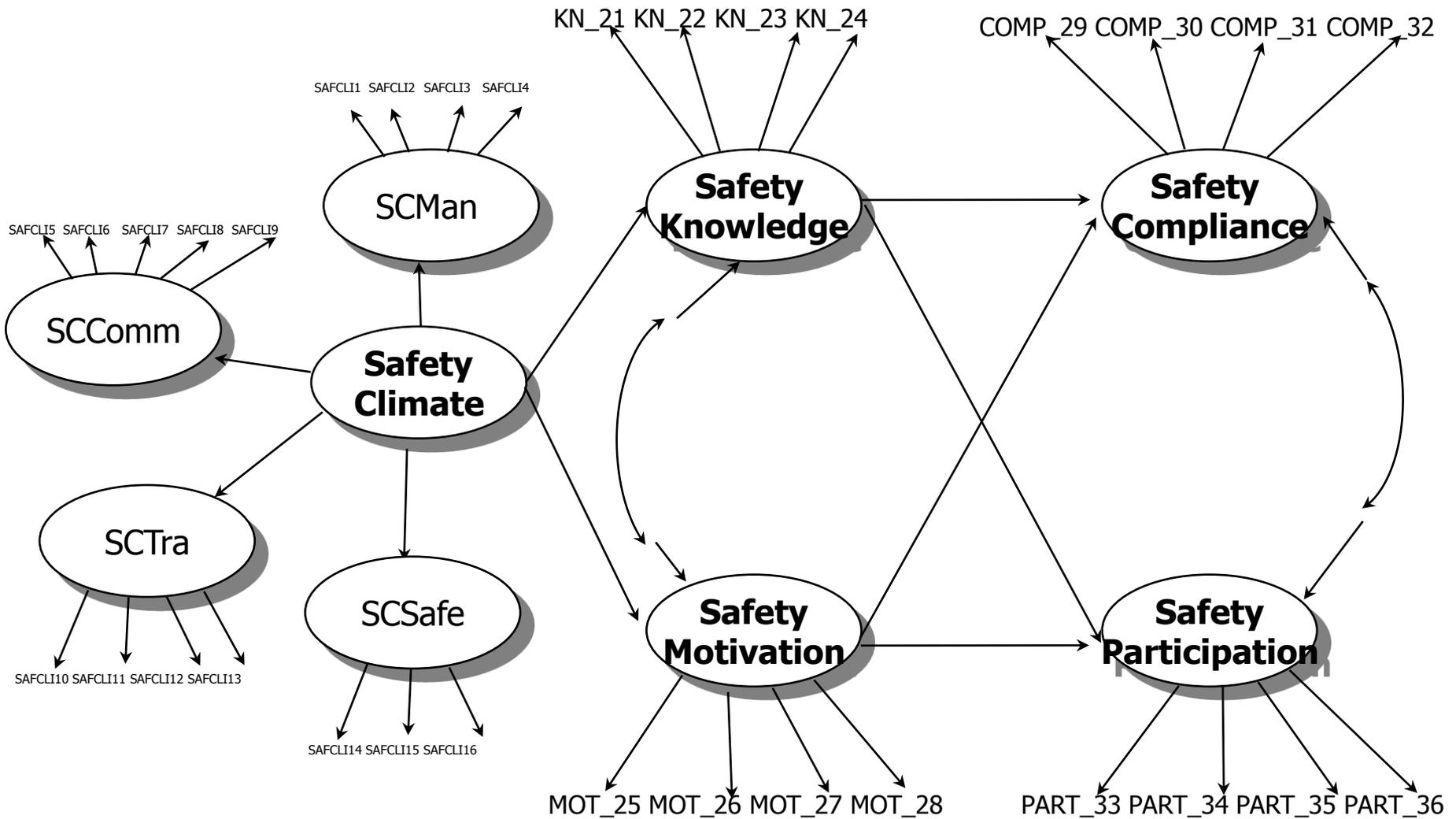
SAF\_PART ON saf\_knw ind\_mot;

saf\_knw WITH ind\_mot; SAF\_COMP WITH SAF\_PART;

# MODELLO "PARTIAL AGGREGATION"



# MODELLO "FULL"





# MODELLO FULL

## MODEL:

**SCMan by SAFCLI1 SAFCLI2 SAFCLI3 SAFCLI4;**  
**SCComm by SAFCLI5 SAFCLI6 SAFCLI7 SAFCLI8 SAFCLI9;**  
**SCTra by SAFCLI10 SAFCLI11 SAFCLI12 SAFCLI13;**  
**SCSafe by SAFCLI14 SAFCLI15 SAFCLI16 ;**  
**saf\_cli by SCMan SCComm SCTra SCSafe;**  
**saf\_knw by KN\_21 KN\_22 KN\_23 KN\_24 ;**  
**ind\_mot by MOT\_25 MOT\_26 MOT\_27 MOT\_28;**  
**SAF\_COMP by COMP\_29 COMP\_30 COMP\_31 COMP\_32;**  
**SAF\_PART by PART\_33 PART\_34 PART\_35 PART\_36 ;**

[neal 2011 item SEM.inp](#)

**saf\_knw on saf\_cli;**  
**ind\_mot on saf\_cli;**  
**SAF\_COMP ON saf\_knw ind\_mot;**  
**SAF\_PART ON saf\_knw ind\_mot;**  
**saf\_knw WITH ind\_mot;**  
**SAF\_COMP WITH SAF\_PART;**

**SAF\_COMP ON saf\_cli ;**  
**SAF\_PART ON saf\_cli ;**



## MODELLI A CONFRONTO

	chi_2	df	TLI	CFI	Rmse	p(RMSEA)	SRMR
Var. oss.	170	2	0,67	0,93	0,24	0,001	0,059
Rel. Corr.	117	2	0,78	0,96	0,2	0,001	0,041
Partial agg.	802	162	0,94	0,95	0,052	ns	0,06
Full SEM	1400	452	0,95	0,95	0,038	ns	0,065

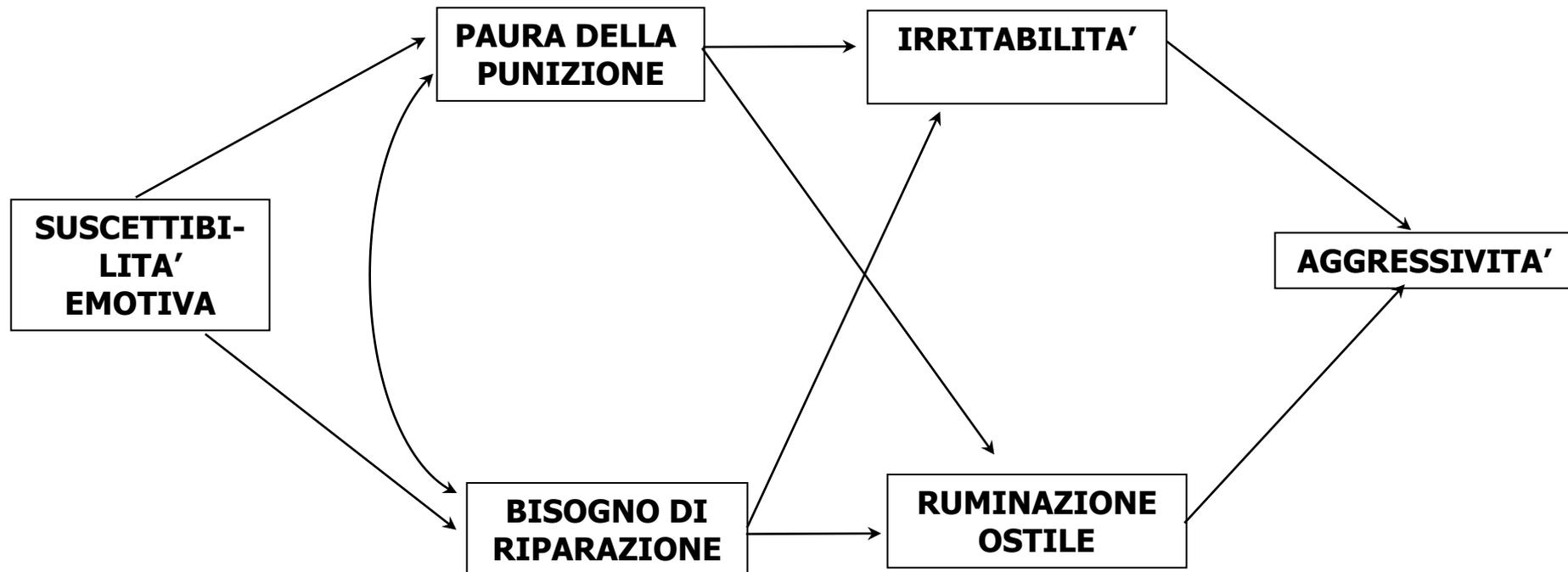
	Var. oss.	Rel. Corr.	Partial agg.	full SEM
Climate -> Knowledge	0.607	0.673	0.673	0.672
Climate -> Motivation	0.347	0.374	0.364	0.362
Knowledge-> Compl.	0.503	0.591	0.567	0.567
Motivation-> Compl.	0.307	0.275	0.280	0.280
Knowledge-> Particip.	0.519	0.642	0.624	0.624
Motivation-> Particip.	0.157	0.115	0.125	0.125
R <sup>2</sup> Knowledge	0.369	0.453	0.452	0.452
R <sup>2</sup> Motivation	0.120	0.140	0.132	0.131
R <sup>2</sup> Compliance	0.521	0.625	0.584	0.584
R <sup>2</sup> Participation	0.386	0.517	0.496	0.496



## MODELLI A CONFRONTO

- **I modelli di aggregazione totale hanno molti meno gradi di libertà. Se si utilizzano tali modelli, le stime fornite dall'approccio "reliability correction" sono meno distorte di quelle ottenute fissando a 0 la varianza residua**
- **I modelli di disaggregazione parziale e totale hanno molti più gradi di libertà. Essi correggono le stime dei parametri per l'errore di misurazione. Vanno sempre preferiti ai modelli di aggregazione totale**
- **Nel nostro esempio modelli di disaggregazione parziale e totale danno risultati sovrapponibili. Il modello "reliability correction" fornisce stime non troppo distanti da quelli "parcels" e "full". Il modello con variabili osservate non corrette è quello più critico**

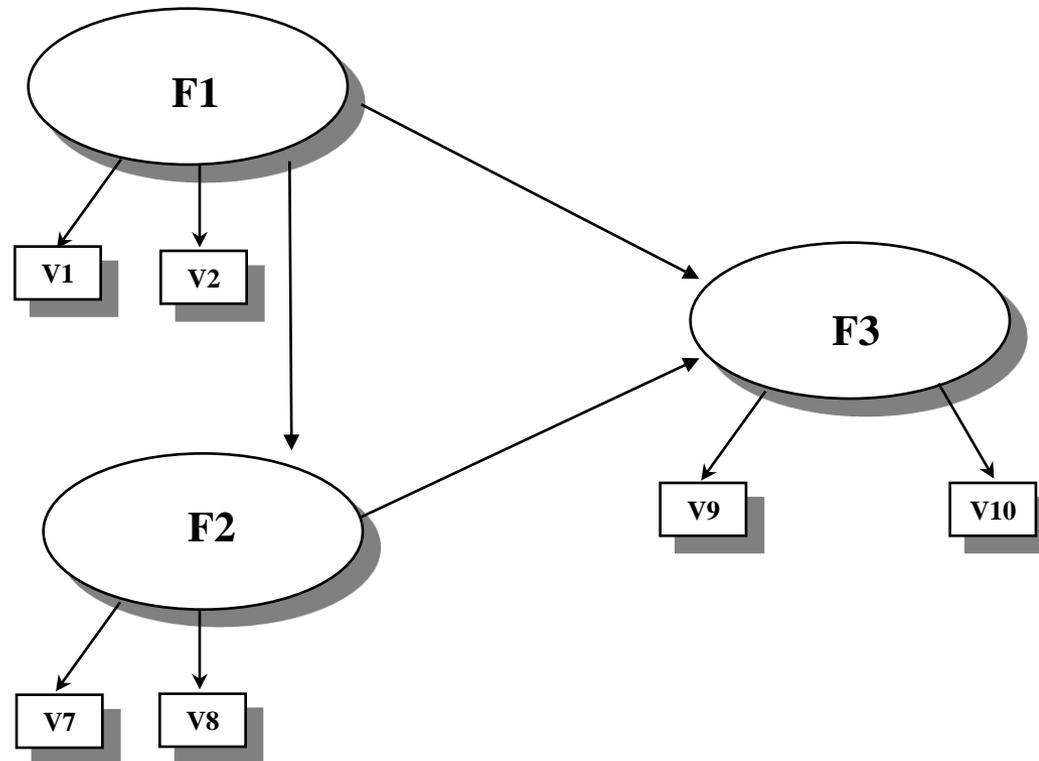
## ESERCIZIO: REALIZZAZIONE DI UN MODELLO DI EQUAZIONI STRUTTURALI CON VARIABILI OSSERVATE



**FILE IS OBSERVED\_PARIS\_2011.DAT;**

## ESERCIZIO: REALIZZAZIONE DI UN MODELLO DI EQUAZIONI STRUTTURALI COMPLETO

Analizzare il seguente modello di equazioni strutturali completo:





## **ESERCIZIO: REALIZZAZIONE DI UN MODELLO DI EQUAZIONI STRUTTURALI COMPLETO**

**V1=ISEE6; V2=ISEE2; V3=ISEE13; V4=ISEE11; V5=ISEE12;  
V6=ISEE3; V7=COMMUN; V8=MONIT; V9=SELF\_EST; V10=SODD;  
V11=OTTIM;**

**F1: autoefficacia nell'espressione delle emozioni positive**

**F2: relazione con i genitori**

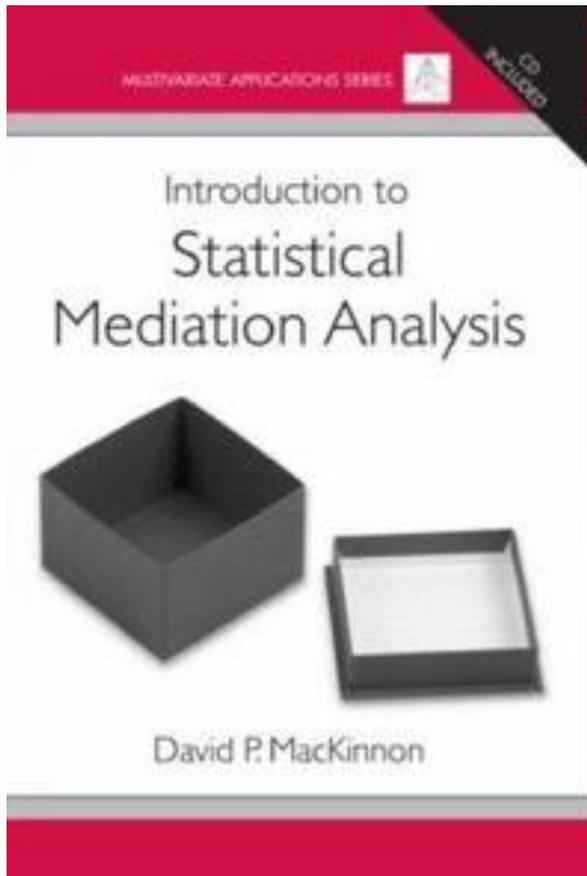
**F3: benessere**

**NUMERO DI SOGGETTI=412**

**La matrice di varianza/covarianza osservata si trova nel file ES2.dat**

# L'analisi della mediazione statistica

## Riferimenti concettuali e metodologici:



<http://www.public.asu.edu/~davidpm/ripl/mediate.htm>



## L'analisi della mediazione statistica

**Definizione di Mediatore:** Una variabile intermedia nel processo causale che lega una variabile indipendente a una dipendente

**Alcuni esempi dalla letteratura psicologica**

*Attitudes cause intentions which then cause behavior (Azjen & Fishbein, 1980)*

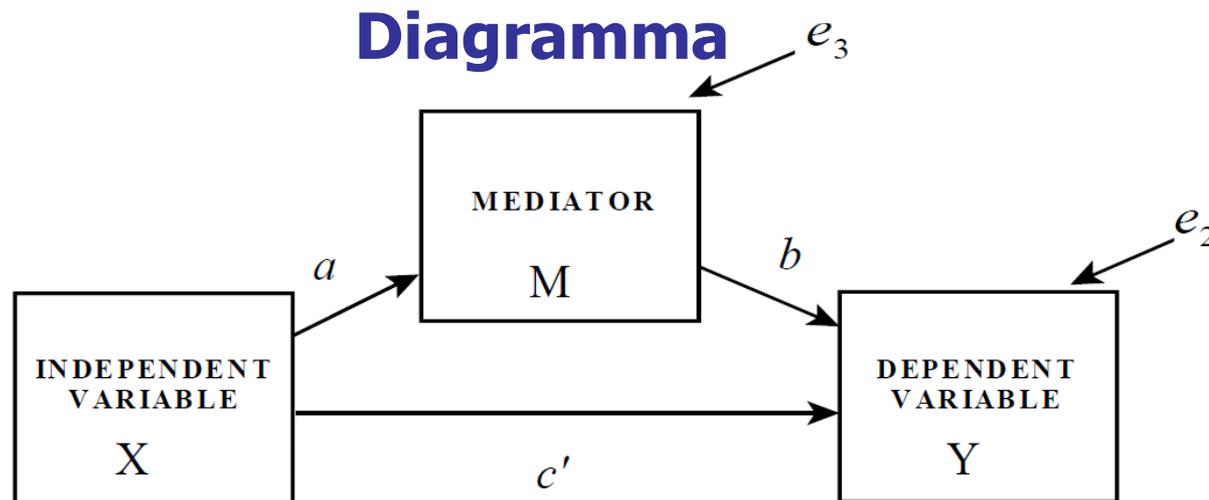
*Prevention programs change norms which promote healthy behavior (Judd & Kenny, 1981)*

*Increasing exercise skills increases self-efficacy which increases physical activity (Bandura, 1977)*

# L'analisi della mediazione statistica

## Tre modi per specificare un modello di mediazione

**Descrizione Verbale:** la variabile M è intermedia nella sequenza causale che lega X a Y.



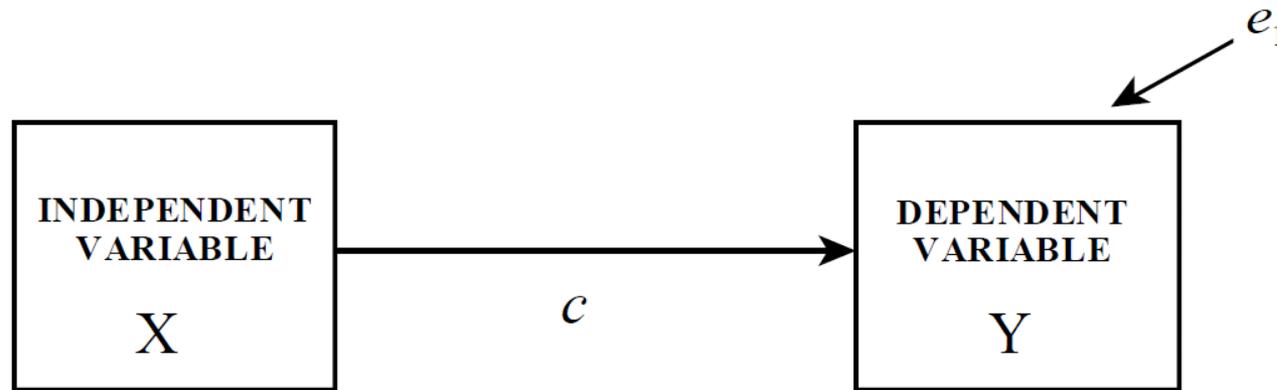
## Equazioni

$$Y = i_2 + c'X + bM + e_2$$

$$M = i_3 + aX + e_3$$

## L'analisi della mediazione statistica

Nella sua forma più semplice la mediazione rappresenta l'aggiunta di una terza variabile nella relazione tra una variabile dipendente  $Y$  e una variabile indipendente  $X$ , dove la  $X$  "causa" la terza variabile di mediazione "M".

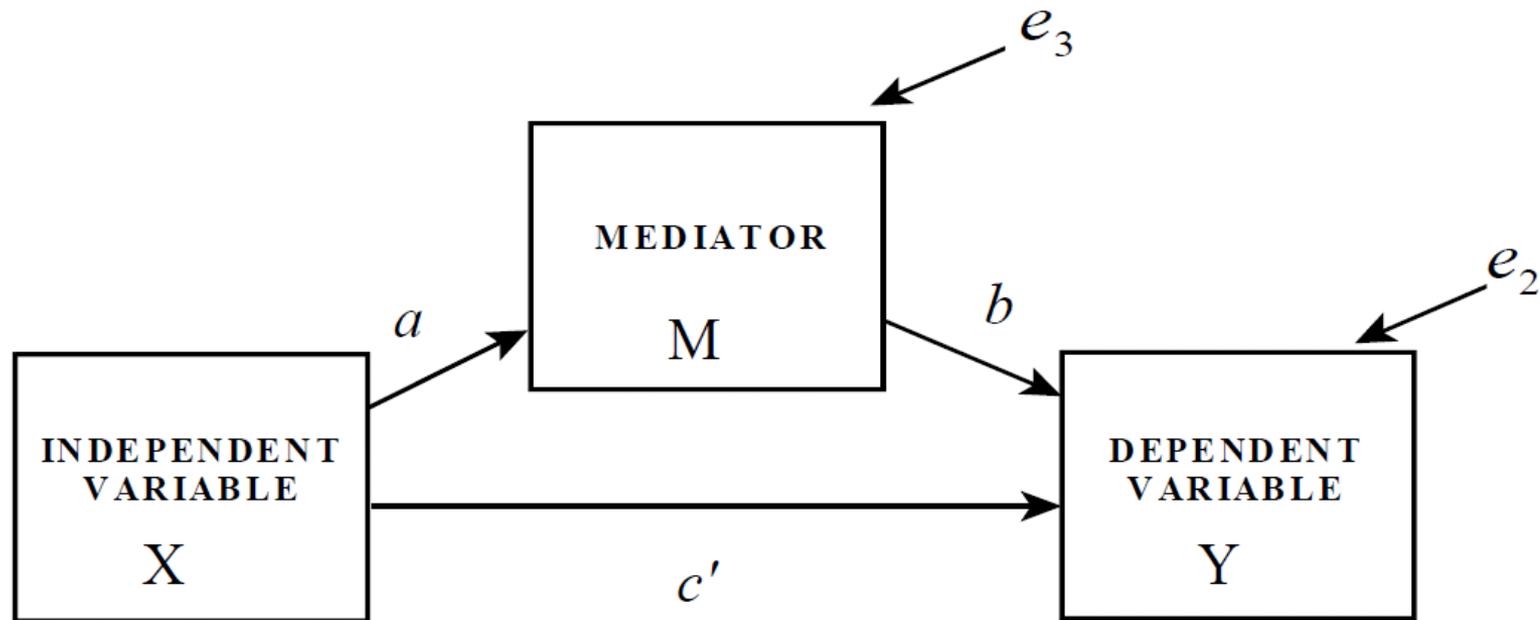


$$Y = i_1 + cX + e_1$$

**Modello senza mediatore "M"**

# L'analisi della mediazione statistica

## Modello con il mediatore "M"



$$Y = i_2 + c'X + bM + e_2$$

$$M = i_3 + aX + e_3$$

## L'analisi della mediazione statistica

Ci sono tre approcci principali all'analisi della mediazione statistica:

(a) l'approccio "causal steps"

(b) l'approccio della differenza dei coefficienti

(c) l'approccio del prodotto dei coefficienti

Tutti questi metodi utilizzano l'informazione proveniente dalle seguenti equazioni di base:

$$Y = i_1 + cX + e_1 \quad (\text{eq. 1})$$

$$Y = i_2 + c'X + bM + e_2 \quad (\text{eq. 2})$$

$$M = i_3 + aX + e_3 \quad (\text{eq. 3})$$



## L'analisi della mediazione statistica

L'approccio "causal steps" è quello proposto da Judd e Kenny (1981) e da Baron e Kenny (1986) e consiste nella verifica delle seguenti condizioni:

- a) È necessaria una relazione significativa tra X e Y nell'eq. 1
- b) È necessaria una relazione significativa tra X e M nell'eq. 3
- c) È necessaria una relazione significativa tra M e Y nell'eq. 2, al netto della relazione tra X e Y nella stessa equazione

## L'analisi della mediazione statistica

L'approccio "causal steps" è quello proposto da Judd e Kenny (1981) e da Baron e Kenny (1986) e consiste nella verifica delle seguenti condizioni:

d) Il coefficiente  $c$  che lega  $X$  a  $Y$  nell'eq. 1 deve essere maggiore del coefficiente  $c'$  che lega  $X$  a  $Y$  nell'eq. 2 (ovvero in presenza del mediatore  $M$ )

d') Nella formulazione originale di Judd e Kenny l'enfasi era sulla mediazione totale: per cui si è in presenza di un effetto di mediazione quando  $c$  è significativo e  $c'$  non lo è

*Baron and Kenny's (1986) article had been cited by 12,688 journal articles as of September 2009, according to Social Sciences Citation Index.*

## L'analisi della mediazione statistica

La condizione  $d'$ , **mediazione totale**, è irrealistica nelle scienze sociali (Baron & Kenny, 1986)

Si parla di **mediazione parziale**, invece, quando  $c'$  è minore di  $c$ , ma non è necessariamente non-significativo. E' una condizione più realistica e in quanto tale è diventata parte del "causal steps approach".

Una volta accertate le condizioni per la presenza della mediazione, di solito viene utilizzato il **prodotto dei coefficienti ( $ab$ )** per stimare l'effetto di mediazione.



## L'analisi della mediazione statistica

**Problemi nell'approccio "causal steps" di Judd e Kenny (1981) e Baron e Kenny (1986):**

- a) La potenza della verifica di questo approccio è molto bassa. Sono necessari tanti test di ipotesi nulle, e questo abbassa la potenza e aumenta l'errore di tipo I.**
- b) L'approccio mira a definire quali sono le condizioni per cui è presente una mediazione: il test esplicito dell'effetto indiretto di X su Y è accessorio.**
- c) E' difficile estendere questo approccio per incorporare mediatori multipli e valutarne l'effetto sulla variabile dipendente.**



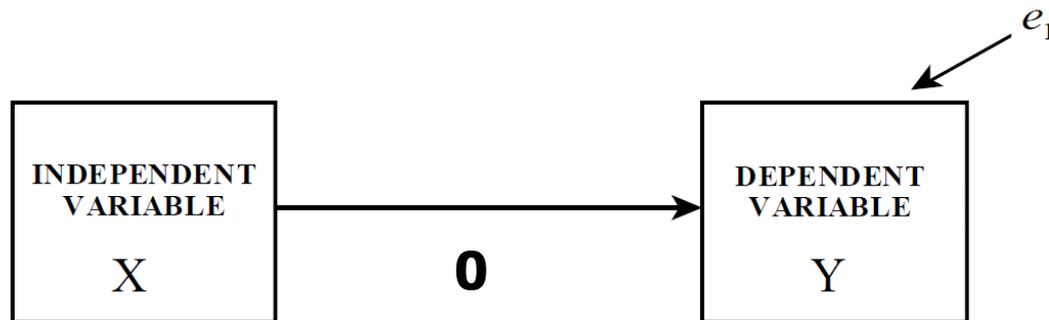
## L'analisi della mediazione statistica

**Problemi nell'approccio "causal steps" di Judd e Kenny (1981) e Baron e Kenny (1986):**

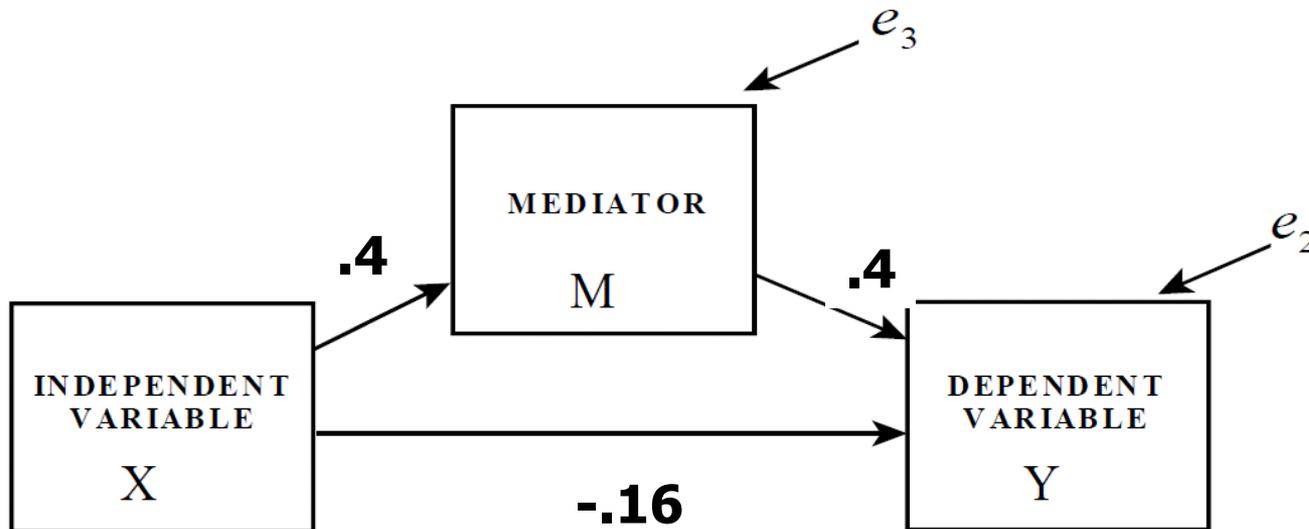
**d) Può esserci mediazione anche in presenza di un effetto non significativo di X su Y: si tratta dei cosiddetti modelli di **mediazione inconsistente**, ovvero modelli in cui la variabile di mediazione M agisce come soppressore della relazione tra X e Y per cui in presenza di M la relazione tra X e Y, prima non significativa, ora lo diventa.**

**In questi modelli solitamente l'effetto diretto e l'effetto indiretto sono di segno opposto e si cancellano, come è illustrato nella figura successiva.**

## L'analisi della mediazione statistica



### Modello senza il mediatore



### Modello con il mediatore

## L'analisi della mediazione statistica

**Calcolo dell'effetto di mediazione (o "indiretto"):  
differenza e prodotto dei coefficienti.**

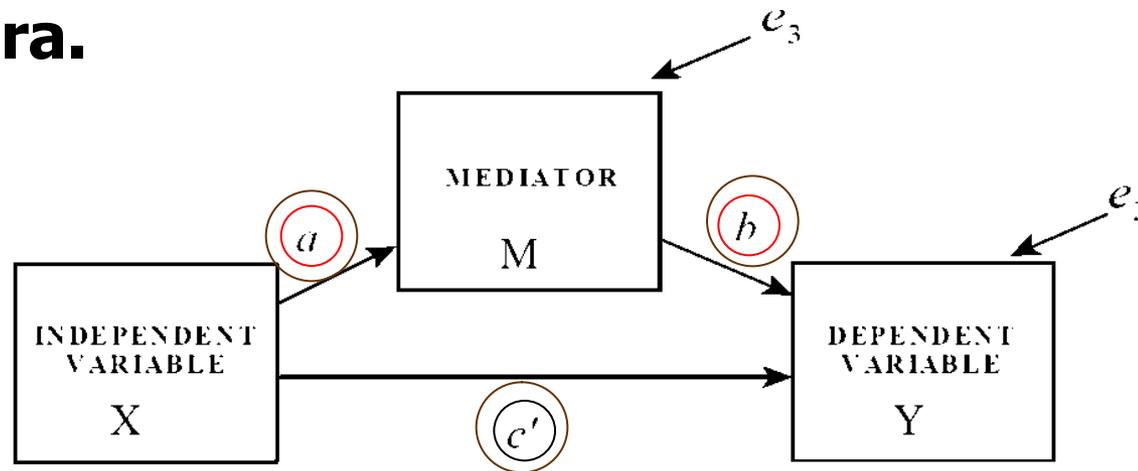
La **differenza** dei coefficienti ( $c-c'$ ) nelle eq. 1 e 2 corrisponde alla riduzione dell'effetto di X su Y dovuto all'aggiunta di M.

Il **prodotto** dei coefficienti ( $ab$ ) nelle eq. 2 e 3 corrisponde al razionale (Alwin & Hauser 1975) per cui la mediazione dipende da quanto la variabile indipendente X cambia il mediatore M (coefficiente a), e da quanto il mediatore M influenza la variabile dipendente Y al netto dell'impatto di X su Y (b).

MacKinnon et al. (1995) hanno dimostrato l'equivalenza algebrica di questi due differenti metodi di stima dell'effetto di mediazione

## L'analisi della mediazione statistica

L'approccio del **prodotto dei coefficienti** è migliore perché può essere facilmente applicato a modelli con: mediatori multipli, variabili indipendenti e dipendenti multiple, variabili osservate e latenti. In questo approccio la **scomposizione degli effetti** è molto chiara.



Effetto **Indiretto** di X su Y =  $ab$

Effetto **Diretto** di X su Y =  $c'$

Effetto **Totale** di X su Y =  $ab+c'$

# L'analisi della mediazione statistica

## Inferenza statistica ed effetti di mediazione

L'errore standard più comunemente utilizzato per  $ab$ ,  $s_{ab}$ , è dato dalla seguente formula derivata da Sobel (1982), che si basa sulle derivate prime ed utilizza il metodo multivariato delta (Folmer, 1981).

$$s_{ab} = \sqrt{a^2 s_b^2 + b^2 s_a^2 + 2ab cov_{ab}}$$

In questa formula il termine  $cov_{ab}$  è generalmente uguale a zero. Si tratta della formula utilizzata in molti programmi per i SEM, EQS (Bentler, 1997), Mplus (Muthén & Muthén, 2009) e LISREL (Jöreskog & Sörbom, 2001).

# L'analisi della mediazione statistica

## Inferenza statistica ed effetti di mediazione

La stima dell'effetto di mediazione e del suo errore standard può essere utilizzata per costruire intervalli di confidenza per l'effetto di mediazione. Gli intervalli di confidenza sono utilizzati perché incorporano l'errore nella stima del parametro, fornendo un insieme di possibili valori per un effetto piuttosto che un singolo valore stimato.

Limiti dell'intervallo di confidenza:  $ab \pm z_{\text{crit}} s_{ab}$

$LSup = ab + z_{\text{crit}} s_{ab}$ ,  $LInf = ab - z_{\text{crit}} s_{ab}$

$z_{\text{crit}}$  è il valore critico. Si può utilizzare anche la  $t$ .

Limiti per l'intervallo al 95% :  $ab \pm 1.96 s_{ab}$

$LSup = ab + 1.96 s_{ab}$ ,  $LInf = ab - 1.96 s_{ab}$



# L'analisi della mediazione statistica

## Inferenza statistica ed effetti di mediazione

**Se lo zero è incluso nell'intervallo di confidenza l'effetto di mediazione non è statisticamente significativo. Invece, se zero è fuori dell'intervallo, l'effetto di mediazione è statisticamente significativo.**

**Si può esaminare la significatività dell'effetto di mediazione dividendo la stima "ab" per il suo errore standard ( $s_{ab}$ ) e confrontando questo valore con quello critico nella distribuzione normale. Se il valore assoluto di questo rapporto è maggiore/uguale a 1.96 allora l'effetto di mediazione è significativamente diverso da zero al livello di significatività dello 0.05.**



## L'analisi della mediazione statistica

### Problemi nella verifica delle ipotesi sull'effetto di mediazione

**I metodi descritti sopra, basati sul coefficiente sviluppato da Sobel possono dare risultati inaccurati in diverse circostanze.**

**Il problema più ricorrente è quello della ridotta potenza della verifica derivante dal fatto che la distribuzione campionaria del prodotto  $ab$  è asimmetrica.**

**Mackinnon et al. (2006) hanno sviluppato un programma per il calcolo degli intervalli "asimmetrici" di confidenza, **PRODCLIN**:**  
**<http://www.public.asu.edu/~davidpm/ripl/Prodclin/>**

**Il metodo di Mackinnon et al. (2006) può però sottostimare l'incidenza dell'errore del I tipo quando  $a=0$  e  $b \neq 0$ , o  $a \neq 0$  e  $b=0$ . In questi casi è più prudente utilizzare il metodo di Sobel.**



## Inferenza Statistica: Il metodo Bootstrap

**Il Bootstrap è una tecnica non parametrica basata sul campionamento con re-immissione. Essa consente di generare una rappresentazione empirica della distribuzione campionaria dell'effetto indiretto trattando il campione ottenuto di ampiezza  $n$  come una rappresentazione della popolazione in miniatura.**

**Il campionamento viene condotto con re-immissione, per cui un nuovo campione di ampiezza  $n$  è costruito campionando i casi dalla popolazione originale ma consentendo che ogni caso una volta estratto possa essere re-immesso nel campione in modo che possa essere estratto di nuovo per costruire un nuovo campione di ampiezza  $n$ .**



# Inferenza Statistica: Il metodo Bootstrap

**Una volta che il campione è costruito, i coefficienti  $a$  e  $b$  sono stimati su questo data-set ri-campionato, e il prodotto dei coefficienti calcolato.**

**Questo processo viene ripetuto per un totale di  $k$  volte (almeno 5000 volte).**

**Alla fine, il ricercatore avrà a disposizione  $k$  stime diverse dell'effetto indiretto, e la distribuzione di tali stime rappresenta una approssimazione empirica della distribuzione campionaria dell'effetto indiretto.**



## Inferenza Statistica: Il metodo Bootstrap

**Questa procedura consente di ottenere un intervallo di confidenza per l'effetto indiretto basato sulle stime bootstrap. Se il valore zero non è compreso tra il limite inferiore e quello superiore dell'intervallo, allora l'effetto indiretto non è zero con una confidenza del  $ci\%$ .**

**Questo procedimento è concettualmente equivalente a rifiutare l'ipotesi nulla che l'effetto indiretto è zero al livello di significatività  $ci\ 100-ci\%$ .**

# Inferenza Statistica: Il metodo Bootstrap

**Limite Inferiore**



**( $k=1000$ ,  
25<sup>a</sup> posizione)**

**Limite Superiore**



**( $k=1000$ ,  
976<sup>a</sup> posizione)**

**Se  $k=1000$ , per un intervallo di fiducia al 95%, mettendo in ordine crescente i  $k$  valori di  $ab$ , il **limite inferiore** dell'intervallo è definito come il **valore di  $ab$  nella 25<sup>a</sup> posizione**, e il **limite superiore** è il **valore di  $ab$  nella 976<sup>a</sup> posizione**.**



# Inferenza Statistica: Il metodo Bootstrap

**Il bootstrap è l'approccio più sensibile per individuare gli effetti indiretti quando sono presenti.**

**Il bootstrap non fa assunzioni sulla distribuzione campionaria.**

**Non si basa sul calcolo di errori standard, e quindi è immune al problema che può verificarsi quando le assunzioni non sono rispettate.**



## Inferenza Statistica: Il metodo Bootstrap

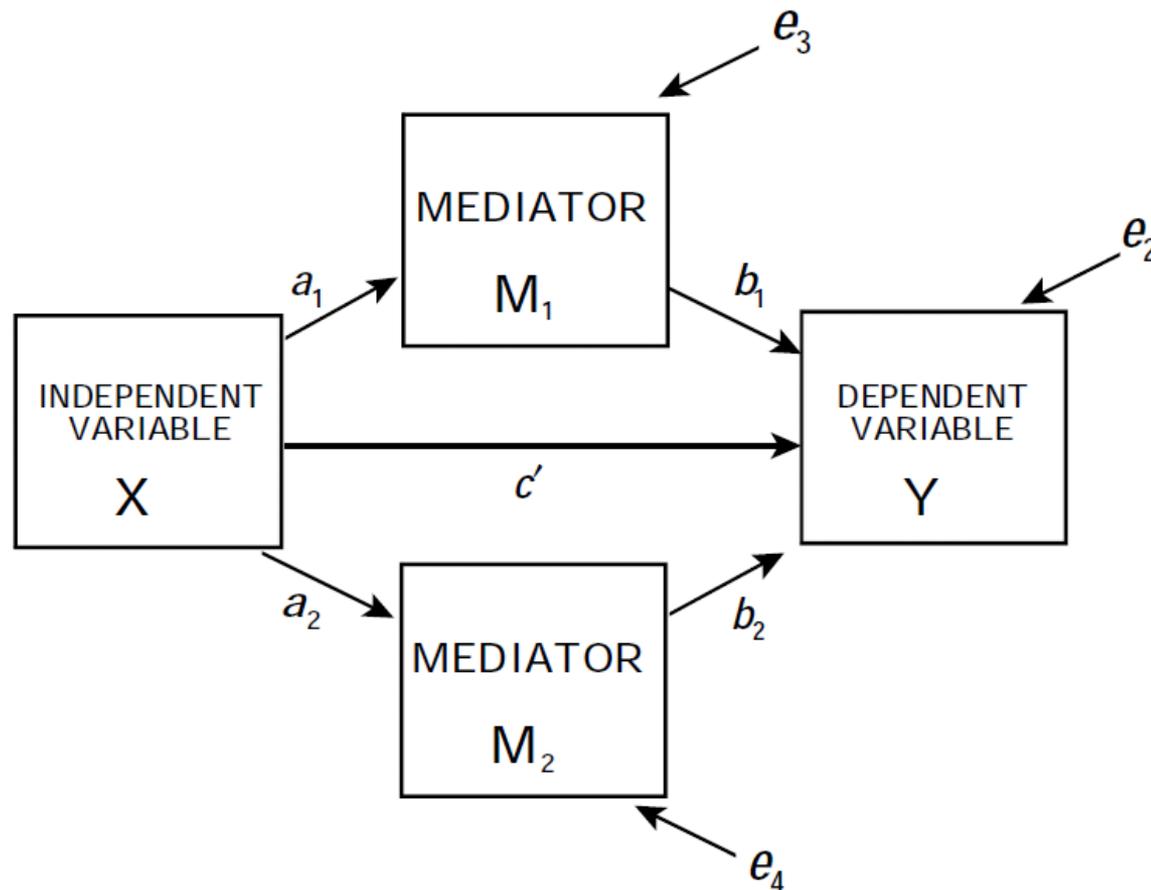
**Per queste proprietà il bootstrap ha maggiore potenza statistica di qualsiasi altra strategia di verifica di ipotesi utilizzabile per esaminare la significatività statistica degli effetti indiretti.**

**Si tratta di un approccio molto generale, che può essere utilizzato in ogni analisi di mediazione. Hayes e Preacher hanno sviluppato delle macro in SPSS e in SAS che possono essere scaricate per poter effettuare i test bootstrap degli effetti indiretti.**

**Mplus implementa la procedura bootstrap in modo estremamente semplice.**

# L'analisi della mediazione statistica

## Modelli con mediatori multipli



# L'analisi della mediazione statistica

## Modello con Mediatori Multipli

**Equazioni di regressione usate per stimare il modello di mediazione con 2 mediatori**

$$Y = i_2 + c'X + b_1M_1 + b_2M_2 + e_2$$

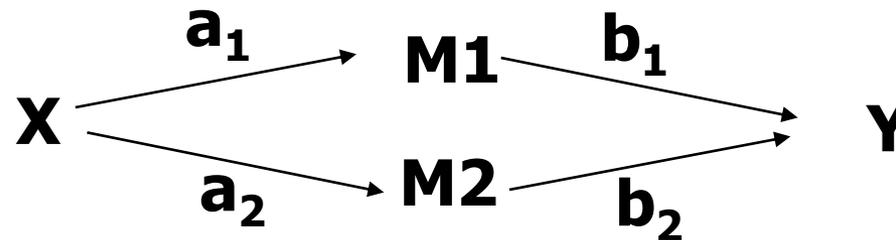
$$M_1 = i_3 + a_1X + e_3$$

$$M_2 = i_4 + a_2X + e_4$$

## L'analisi della mediazione statistica

### Scomposizione degli effetti nei modelli con mediatori multipli

Nei modelli con mediatori multipli è possibile definire due tipi di effetti di mediazione:



La mediazione specifica è l'effetto dovuto a uno specifico mediatore,  $a_1b_1$  (per  $M1$ ),  $a_2b_2$  (per  $M2$ ).

Mediazione totale: è l'effetto dovuto a tutti i possibili percorsi attraverso i quali  $X$  influenza indirettamente  $Y$ , ovvero  $a_1b_1 + a_2b_2$ .

# L'analisi della mediazione statistica

## Scomposizione degli effetti nei modelli con mediatori multipli

**L'effetto diretto** è definibile sempre come:

$$X \xrightarrow{c'} Y$$

**L'effetto totale:** è la somma di tutti gli effetti che agiscono su  $Y = c' + (a_1b_1 + a_2b_2)$

**Effetto diretto:**  $c'$

**Effetto indiretto (attraverso M1):**  $a_1b_1$

**Effetto indiretto (attraverso M2):**  $a_2b_2$



# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

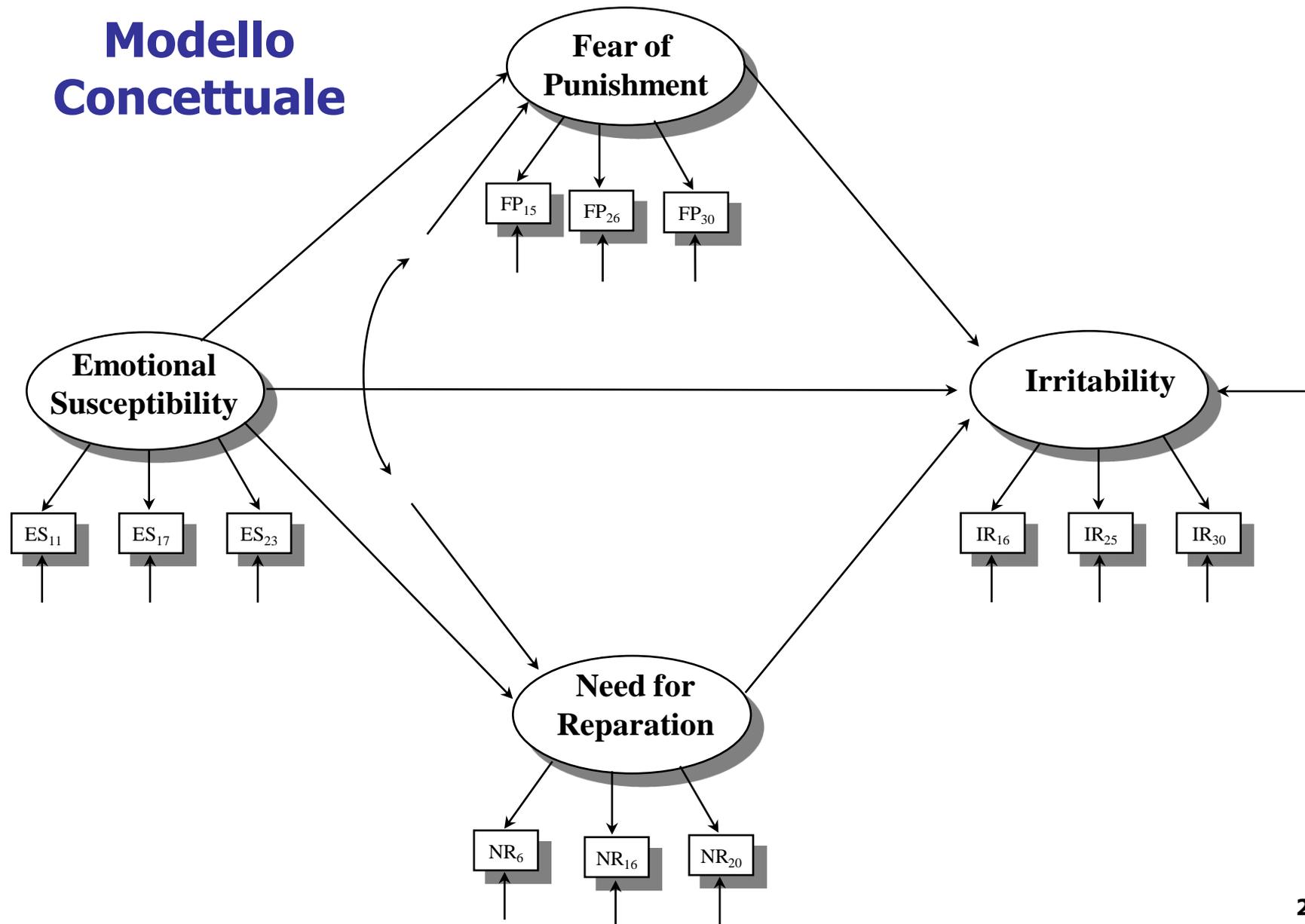
**I modelli di mediazione rappresentano un caso specifico dei SEM che può essere implementato in MPLUS.**

**Il modello generale consente di testare modelli più complicati con variabili indipendenti e dipendenti multiple, latenti e osservate.**

**Bisogna utilizzare i metodi di stima della massima verosimiglianza o dei minimi quadrati, perché i residui delle variabili dipendenti correlano.**

# L'analisi della mediazione statistica

## Modello Concettuale



# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

**Gli effetti indiretti possono essere ottenuti tramite i comandi IND e VIA nella sezione della sintassi MPLUS "MODEL INDIRECT".**

**IND è utilizzato per ottenere tutti gli effetti indiretti (totali e specifici).**

**MODEL INDIRECT:**

**Y IND X;**

**Y è la variabile dipendente, X la indipendente.**

# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

**VIA è utilizzata per ottenere solo l'effetto indiretto specifico definito dalle variabili nel comando.**

### MODEL INDIRECT:

**Y VIA M1 X;**

**Y è la variabile dipendente, X la indipendente, M1 il mediatore. Solo l'effetto indiretto specifico X->M1->Y viene calcolato.**



# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

TITLE: SEM

DATA:

**PARIS\_2011\_SEM**

FILE IS SEM\_PARIS.DAT;

VARIABLE:

NAMES ARE

es11 es17 es23 ir16 ir25 ir30  
FP15 FP26 FP30 nr6 nr16 nr20 ;

USEV ARE

es11 es17 es23  
ir16 ir25 ir30  
FP15 FP26 FP30  
nr6 nr16 nr20 ;

# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

MODEL:

EMSUS BY es11 es17 es23;

FEARPUN BY FP15 FP26 FP30 ;

NEEDREP BY NR6 NR16 NR20 ;

IRRIT BY ir16 ir25 ir30;

FEARPUN ON EMSUS;

NEEDREP ON EMSUS;

IRRIT ON EMSUS FEARPUN NEEDREP;

NEEDREP WITH FEARPUN;

MODEL INDIRECT:

IRRIT IND EMSUS;

OUTPUT: STANDARDIZED SAMPSTAT TECH1 TECH4 CINTERVAL

MODINDICES (3.84) ;

Calcola gli intervalli di  
confidenza

Calcola la scomposizione  
degli effetti

Calcola la covarianza  
tra le stime



# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from EMSUS to IRRIT				
Total	0.321	0.059	5.426	0.000
Total indirect	0.151	0.036	4.224	0.000
Specific indirect				
IRRIT FEARPUN EMSUS	0.180	0.037	4.855	0.000
IRRIT NEEDREP EMSUS	-0.029	0.016	-1.853	0.064
Direct				
IRRIT EMSUS	0.170	0.063	2.708	0.007



# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
<b>Effects from EMSUS to IRRIT</b>							
Total	0.169	0.205	0.224	0.321	0.418	0.437	0.473
Total indirect	0.059	0.081	0.092	0.151	0.210	0.221	0.243
<b>Specific indirect</b>							
<b>IRRIT</b>							
<b>FEARPUN</b>							
EMSUS	0.085	0.108	0.119	0.180	0.242	0.253	0.276
<b>IRRIT</b>							
<b>NEEDREP</b>							
EMSUS	-0.070	-0.060	-0.055	-0.029	-0.003	0.002	0.011
<b>Direct</b>							
<b>IRRIT</b>							
EMSUS	0.008	0.047	0.067	0.170	0.273	0.292	0.331

**Lower/Upper .5%: p=.01; Lower/Upper 2.5%: p=.05;  
Lower/Upper 5%: p=.10**



# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

### Stime Bootstrap

**ANALYSIS:**

**Bootstrap = 5000;**

**PARIS\_2011\_SEM**

**MODEL:**

EMSUS BY es11 es17 es23;

FEARPUN BY FP15 FP26 FP30 ;

NEEDREP BY NR6 NR16 NR20 ;

IRRIT BY ir16 ir25 ir30;

FEARPUN ON EMSUS;

NEEDREP ON EMSUS;

IRRIT ON EMSUS FEARPUN NEEDREP;

NEEDREP WITH FEARPUN;

**MODEL INDIRECT:**

**IRRIT IND EMSUS;**

**OUTPUT: STANDARDIZED SAMPSTAT TECH1 TECH4**

**CINTERVAL (BOOTSTRAP) MODINDICES (3.84) ;**

# L'analisi della mediazione statistica

## Modelli di mediazione in MPLUS

CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
<b>Effects from EMSUS to IRRIT</b>							
Total	0.170	0.204	0.223	0.321	0.418	0.438	0.480
Total indirect	0.061	0.081	0.091	0.151	0.220	0.238	0.260
<b>Specific indirect</b>							
IRRIT							
FEARPUN							
EMSUS	0.094	0.110	0.120	0.180	0.253	0.268	0.294
IRRIT							
NEEDREP							
EMSUS	-0.087	-0.069	-0.061	-0.029	-0.006	-0.002	0.005
<b>Direct</b>							
IRRIT							
EMSUS	-0.010	0.036	0.058	0.170	0.277	0.298	0.338

La stima bootstrap per i limiti dell'IC dell'effetto specifico di IRRIT attraverso NEEDREP ora è significativa ( $p < .05$ ).



## L'analisi della mediazione statistica

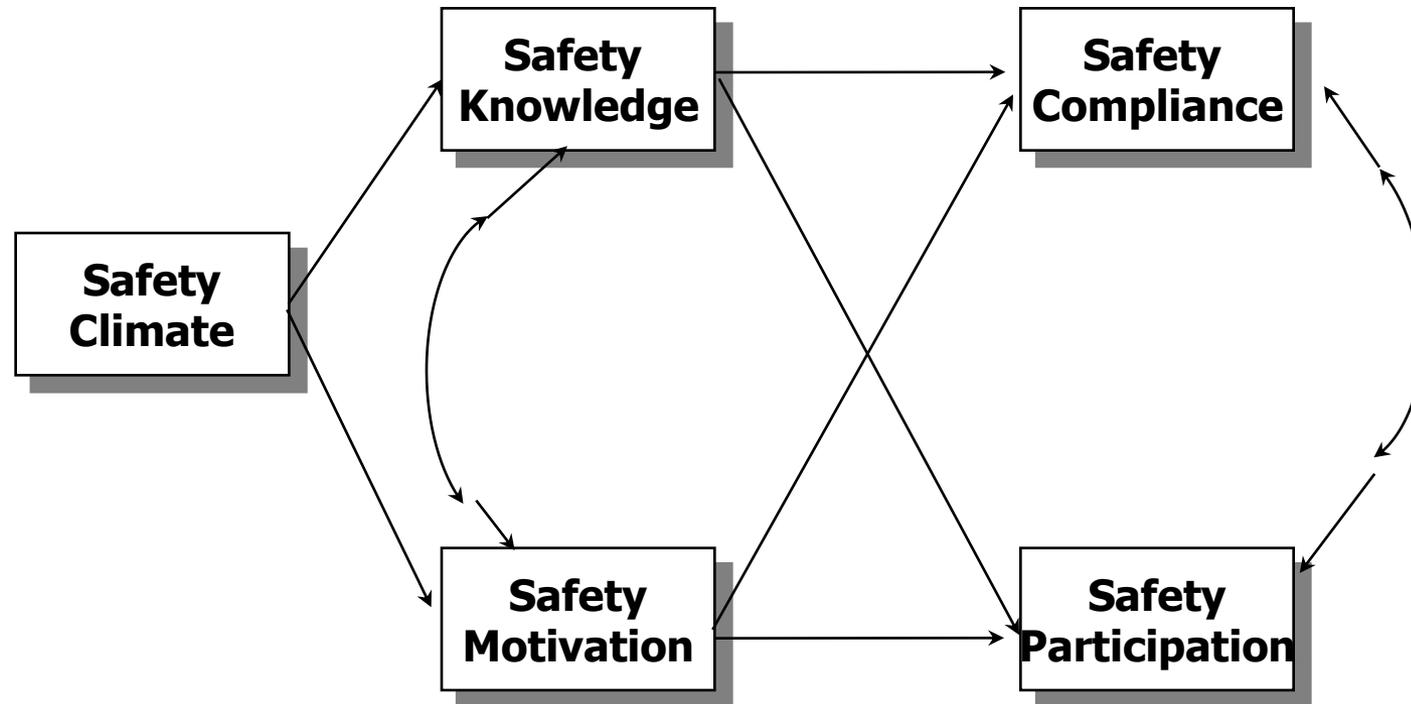
- L'approccio di Baron e Kenny è problematico e va sostituito con approcci più moderni, basati sulla stima dell'effetto indiretto
- Il metodo più utile per calcolare l'effetto di mediazione è quello che si basa sul prodotto dei coefficienti
- Soprattutto su piccoli campioni è importante utilizzare corrette procedure per la stima dei parametri (es, quelle basate sugli intervalli asimmetrici)

## I SEM per l'analisi di Mediazione

**I SEM offrono la massima flessibilità per esaminare modelli con: variabili indipendenti multiple, mediatori multipli (in parallelo o concatenati), variabili dipendenti multiple, relazioni ricorsive o non ricorsive, variabili osservate e/o latenti**

**I SEM garantiscono flessibilità nel calcolo dei parametri utilizzando stimatori **full information**. Le tecniche "limited information" (come i minimi quadrati ordinari) stimano i parametri di ciascuna equazione separatamente. Le tecniche "Full information" (FIML) stimano i parametri di un modello simultaneamente per tutte le equazioni.**

**I SEM esaminano sia i singoli parametri sia il modello nella sua globalità, grazie all'utilizzo di molteplici indici di fit. I SEM consentono di esaminare modelli molto complessi, come quelli multilivello o quelli su gruppi multipli.**

**ESERCIZIO 7: ANALISI DELLA MEDIAZIONE STATISTICA****Utilizzando il modello dell'esempio**

- **Esaminare gli effetti diretti, indiretti e totali**
- **Calcolare gli intervalli di fiducia al 95% per gli effetti indiretti specifici applicando la procedura basata sul metodo bootstrap**



## POTENZIALITÀ DEI SEM

- \* **Analisi simultanea delle relazioni tra costrutti (osservati e latenti) specificati da una teoria**
- \* **Incorporazione dell'errore di misurazione nel modello**
- \* **Esame probabilistico della consistenza tra teoria e dati**
- \* **Possibilità di considerare simultaneamente:**
  - **variabili osservate e variabili latenti**
  - **strutture di covarianza e medie strutturali**
  - **strutture di covarianza e/o medie su più campioni**
  - **variabili continue e categoriali (MPLUS)**
  - **data set complessi (modelli multilivello)**



## LIMITI DEI SEM

- \* **Necessità di ipotesi teoriche trasferibili in un modello**
- \* **Necessità di dati di qualità elevata (analisi preliminari)**
- \* **Condizioni stringenti di applicazione (es., ML)**
- \* **Condizioni stringenti per l'identificabilità dei modelli**
- \* **Uso "esplorativo" - non guidato dalla teoria (es., MI)**
- \* **Interpretazione causale dei nessi**
- \* **Direzionalità dei nessi di influenza**