LISREL per principianti*
Introduzione ai modelli di equazione strutturale
(versione ridotta)

Germano Rossi
germano.rossi@unimib.it†

5 dicembre 2007
vers. 0.3.4

*I titoli fra parentesi quadre indicano paragrafi vuoti, ancora da scrivere

†Università degli Studi di Milano-Bicocca, Dipartimento di Psicologia -

Indice

1 Dalla regressione alle equazioni strutturali

Questo capitolo ha lo scopo di introdurre il lettore al concetto di equazioni strutturali, soprattutto da un punto di vista intuitivo ovvero senza entrare negli aspetti più propriamente matematici, ma limitandosi a quelli logici e di significato. Presenterà anche alcune nozioni fondamentali per la comprensione dei capitoli seguenti.

Per la comprensione del testo, potrebbe essere necessario conoscere qualche nozione di algebra matriciale e di regressione lineare semplice e multipla. Se non avete queste conoscenze potete fare riferimento a queste altre dispense: *Appunti di algebra matriciale* (Rossi, 2002) e *Appunti sulla regressione lineare semplice e multipla* (?, ?), disponibili allo stesso sito in cui avete trovato questa dispensa.

1.1 Terminologia

Durante lo studio della statistica di base, si affrontano alcuni concetti relativi alle variabili, sia intese come scale di misura sia in base alla loro funzione.

La **variabile** va intesa come l'insieme delle misurazioni di una certa caratteristica di un certo aspetto della realtà, misurazioni che fanno riferimento a diversi casi statistici. Ad esempio, in psicologia, una persona che partecipa ad una ricerca come soggetto, è da considerarsi un "caso statistico" e le caratteristiche che vengono rilevate su di lui, come il sesso, l'età, l'altezza in cm, le risposte ad item di atteggiamenti o ad un questionario di conoscenze, sono tutte "variabili".

Una prima suddivisione, che solitamente si affronta all'inizio della statistica inferenziale, è quella fra variabile dipendente e indipendente.

La **variabile indipendente** è una variabile che misura un carattere della realtà che, presumibilmente, non viene influenzato da altri aspetti della realtà, una variabile che non dipende quindi da altre variabili. Esempi di variabili che vengono considerate indipendenti, in ambito psicologico, sono il sesso e l'età. Dovrebbe apparire chiaro che la maggior parte di queste variabili non sono *realmente* indipendenti, ma lo sono solo rispetto all'ambito di ricerca in cui vengono considerate. In psicologia, non possiamo certo dire che il sesso o l'età dipendono dall'intelligenza o dalla cultura..., mentre in ambito medico, il sesso può essere una variabile dipendente da altre.

Proprio per questo motivo, si considera **variabile dipendente** quella variabile che viene influenzata da altre caratteristiche e che cambia di valore al variare di quelle. Per esempio, la ricchezza del vocabolario di una persona può dipendere dall'educazione scolastica, dalla famiglia di origine e/o dal numero di libri letti.

Per comprendere la tecnica delle equazioni strutturali, dobbiamo introdurre altre due classificazioni dicotomiche delle variabili. Innanzitutto la dicotomia fra variabile osservata e variabile latente, successivamente quella fra variabile esogena e variabile endogena.

La **variabile osservata** è una variabile misurabile ed effettivamente misurata con determinati "strumenti" di misurazione o di rilevazione. Le variabili solitamente usate e studiate in statistica di base sono tutte variabili osservate: l'età di una persona, il reddito di una famiglia, il numero di errori in una prova...

La **variabile latente**, invece, è una variabile che non è direttamente misurabile e di cui noi osserviamo (ed eventualmente misuriamo) solo gli effetti. L'intelligenza non è misurabile direttamente, ma ci sono dei comportamenti che noi giudichiamo "manifestazioni di intelligenza" e che sono misurabili. Analogamente per la cultura, l'apprendimento e perfino l'inconscio. Quando Freud ipotizzò l'inconscio, lo fece sulla base di manifestazioni comportamentali (come i lapsus) che gli fecero pensare che questi comportamenti avessero alle spalle una causa unica, non direttamente osservabile, quindi latente.

Una variabile viene chiamata **esogena** (cioè esterna) in relazione ad un modello di nessi causali.

Nell'ambito di questi modelli, la variabile esogena è una variabile latente (oppure osservata) che svolge sempre e soltanto funzione di variabile indipendente, ovvero di variabile che causa un effetto.

Sempre nell'ambito dei modelli causali, si chiama **variabile endogena** (cioè interna) quella variabile che può svolgere sia funzione di variabile dipendente sia di variabile indipendente. Queste variabili possono cioè essere effetto di alcune variabili e contemporaneamente causa per altre. E' endogena anche una variabile che svolge sempre il ruolo di dipendente.

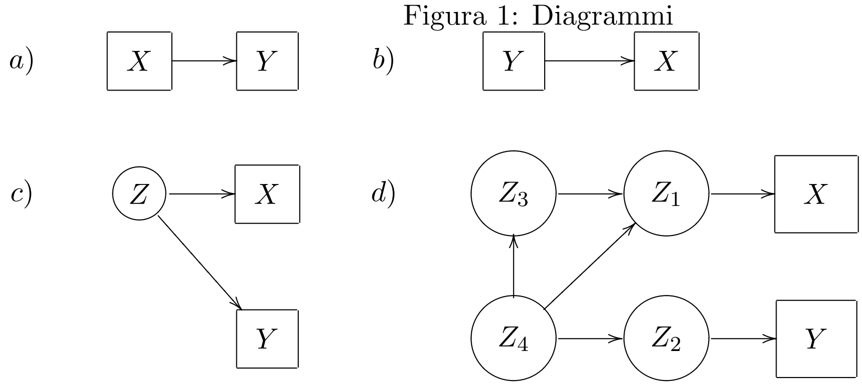
1.2 Modellazione grafica

Abbiamo parlato di nessi causali fra variabili. E' possibile rappresentare graficamente questi nessi causali, ovvero le relazioni fra le variabili (osservate/latenti, dipendenti/indipendenti) rispettando alcune semplici regole (riassunte in Tab. 1). All'interno della grafica dei modelli causali, le variabili osservate vengono distinte da quelle latenti tramite l'uso di rettangoli, mentre le relazioni causali fra le variabili viene espressa da una freccia che origina nella variabile-causa e termina sulla variabile-effetto. Una doppia freccia (generalmente curva) indica una relazione reciproca non causale (ad es. una correlazione o una covarianza). Usando queste regole si possono rappresentare diverse tecniche che corrispondono a casi semplificati dei modelli di equazione strutturale. Nei paragrafi che seguono vedremo alcune di queste tecniche e vedremo anche come si rappresentano graficamente.

Quando si studia la correlazione ci viene anche insegnato che quest'indice statistico non implica alcun effetto di causalità. E' un indice di associazione fra due variabili e ci informa soltanto sul loro andamento reciproco: al variare di una, varia anche l'altra variabile considerata. Se cercassimo però di trovare una spiegazione alla correlazione fra X e Y, ossia a questo andamento concomitante, avremmo diverse possibilità:

1. X è causa di Y (Fig. 1a);
2. Y è causa di X (Fig. 1b);
3. X e Y sono entrambi causati da una terza variabile Z (Fig. 1c) e questo fa in modo che, al variare di Z, anche X e Y cambino in modo concomitante, dandoci l'impressione che esista un qualche tipo di legame fra le due variabili;
4. X e Y sono entrambi causati da un blocco di variabili Z_1, Z_2 ... Z_n (Fig. 1d) con gli stessi effetti appena descritti.

Figura 1: Diagrammi



Ho usato il termine "causato" per indicare la relazione direzionata fra le due variabili. Possiamo anche dire che una variabile "influenza" l'altra oppure che la "modula" o la "dirige". A livello grafico

(Fig. 1) questa "influenza" è rappresentata da una freccia: la variabile che riceve la freccia è la variabile che dipende, che viene influenzata, modulata, causata. Inoltre, le variabili che abbiamo effettivamente misurato (quelle osservate) sono rappresentate da rettangoli, mentre quelle non misurate (ipotizzate, latenti, presunte) sono indicate da cerchi o da ellissi. La rappresentazione grafica di Fig. 1 si chiama anche "modello causale" o "modello relazionale".

Variabili osservate	$\longrightarrow$	quadrato, rettangolo
Variabili latenti	$\longrightarrow$	cerchio, ellisse
Influenza, spiegazione	$\longrightarrow$	freccia
Relazione reciproca (correlazione)	$\longrightarrow$	doppia freccia

Tabella 1: Regole della modellazione causale

2 Regressione lineare

Per capire pienamente i modelli di equazioni strutturale occorre avere alcune conoscenze di regressione lineare (cfr. dispensa) e di algebra matriciale (cfr. dispensa). Per ora ci limiteremo a cercare di capire come si passa da una regressione semplice ad un'equazione strutturale.

Tramite una funzione matematica (in particolare quella di una retta) si cerca di vedere se:

- una variabile osservata viene spiegata da un'altra variabile osservata (Regressione lineare semplice)
- o da altre variabili osservate (Regressione lineare multipla)
- più variabili osservate vengono spiegate da altre variabili osservate (Regressione lineare multivariata)

Tabella 2: Esempio di regressione semplice

Empatia	Soddisfazione
70	4
94	5
36	2
48	1

Esempio

Un ricercatore sta studiando la relazione fra il grado di empatia degli psicoterapeuti e la soddisfazione dei loro pazienti (Tab. 2). In uno studio pilota, utilizza 4 coppie paziente-terapeuta. Con un test misura l'empatia del terapeuta (maggiore il valore, maggiore l'empatia) e chiede al paziente di giudicare il proprio grado di soddisfazione su una scala da 1 (poco) a 5 (molto). All'aumentare del grado di empatia del terapeuta aumenta la soddisfazione del paziente. I dati possono essere facilmente interpretati tramite una retta (Fig. 2).

Sostanzialmente, quando diciamo che una variabile ne "spiega" un'altra, intendiamo dire che stiamo cercando di stabilire una relazione di causa-effetto. La variabile "spiegata" è l'effetto, quella che spiega (o influenza) è la causa.

La regressione lineare utilizza il principio matematico della retta per "cercare" questa relazione causale. Sulla base del numero e del tipo di variabili implicate, avremo differenti tipi (e complessità) di regressione.

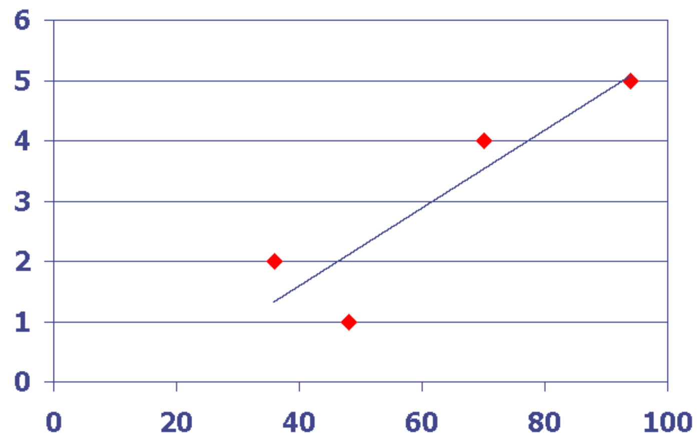


Figura 2: Rappresentazione grafica

2.1 Regressione lineare semplice

Quando abbiamo una sola variabile indipendente e una sola variabili dipendente, parliamo di regressione lineare semplice. La variabile indipendente è una variabile osservata (generalmente indicata con X) che spiega (o influenza) la variabile dipendente, anch'essa osservata (generalmente indicata con Y, Fig. 3). E' però possibile che la variabile Y non venga interamente spiegata da X, in tal caso le parti non utilizzate, ovvero non spiegabili, vengono chiamate *errori*. Per questo motivo, accanto al rettangolo che rappresenta la variabile Y, mettiamo una ε (o una e) con una freccia verso la Y stessa, ad indicare che una parte di questa variabile è "spiegabile" prendendo in considerazione un errore casuale.

$$\boxed{X} \longrightarrow \boxed{Y} \longleftarrow \varepsilon$$

Figura 3: modello relazionale di una regressione semplice

Osservando la Fig. 3 possiamo dire che la variabile Y è riproducibile, con certa approssimazione, sulla base della variabile X e di un errore casuale. Se proviamo a formulare questo concetto dobbiamo scrivere che $Y = X + e$. E' però possibile che la variabile X non contribuisca completamente a generare la Y e quindi possiamo considerare che la X venga "pesata": la metà, il doppio, i tre quarti... Per cui una formalizzazione più corretta sarebbe $Y = bX + e$, dove b è la parte di X (ovvero il "peso") che concorre a spiegare Y. Inoltre è possibile che per ottenere Y sia necessario aggiungere o togliere un valore costante. In questo caso avremmo:

$$Y = a + bX + e$$

A questo punto, se avete qualche ricordo di geometria, diventa evidente che questa è esattamente la formula generale di una retta.

2.2 Regressione lineare multipla

La relazione causale implicata dalla regressione semplice è abbastanza semplicistica perché ipotizza che una sola variabile possa essere la spiegazione di un'altra. E' più probabile che esistano più cause concomitanti che concorrono, in misura diversa, a spiegare la dipendente.

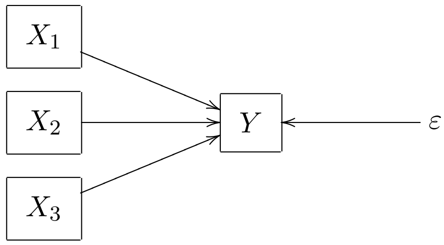


Figura 4: modello relazionale di una regressione multipla

Nella regressione lineare multipla, al posto di una sola variabile indipendente, ne abbiamo 2 o più. Anche in questo caso si tratta di usare una retta che può esistere in un iperpiano a 3 dimensioni (se ci sono due indipendenti) o a più dimensioni (con più di 2 indipendenti). Il tipico modello relazionale per rappresentare una regressione multipla è indicato in Fig. 4). Ovviamente l'importanza delle variabili X nello spiegare la variabile dipendente saranno diverse, alcune incideranno pesantemente, altre pochissimo.

Osservando il modello relazionale di Fig. 4 possiamo ampliare la formula della regressione semplice in questo modo:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + e$$

2.3 Regressione lineare multivariata

Se la regressione multipla è un ampliamento di quella semplice, in quanto presuppone più variabili indipendenti che vanno a spiegare una singola variabile dipendente, una situazione ancora più realistica è quella rappresentata da una regressione multivariata.

In questo caso, vi sono diverse variabili osservate X che influenzano diverse variabili osservate Y (Fig. 5). Con questo tipo di modello le variabili dipendenti possono essere pensate come l'effetto concomitante di diverse cause (le X), ciascuna con un peso e un'importanza diversa. Ovvero, osservando a partire dalle X, che ogni variabile indipendente può influenzare diverse variabili dipendenti.

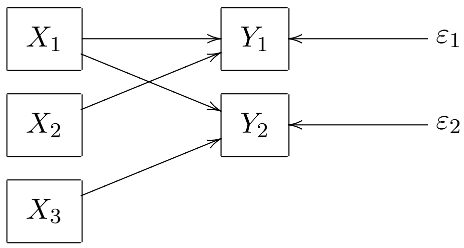


Figura 5: modello relazionale di una regressione multivariata

Sulla base del modello di Fig. 5 possiamo scrivere due equazioni differenti, una per ciascuna delle

Y:

$$
\begin{aligned}
Y_1 &= b_1 X_1 + b_2 X_2 & & +e \\
Y_2 &= b'_1 X_1 & +b'_3 X_3 & +e
\end{aligned}
$$

La variabile indipendente X_1 influisce su entrambe le dipendenti, mentre le altre ne influenzano una sola.

I tre modelli visti finora, utilizzavano tutti variabili osservate e usavano solo variabili osservate. Se introduciamo delle variabili latenti, la complessità dei modelli aumenta e contemporaneamente aumenta anche la loro capacità di adeguarsi alla realtà che voglio spiegare.

3 Analisi fattoriale

L'analisi fattoriale può essere genericamente pensata come una tecnica che cerca di individuare e/o associare una variabile latente a due o più variabili osservate. Ricordiamo che una variabile latente è una caratteristica che non possiamo misurare direttamente e che quindi ipotizziamo sulla base dei suoi effetti. Se questa variabile latente (causa) è una caratteristica particolare, tutte le sue manifestazioni (effetti) saranno "variazioni sul tema" di una stessa caratteristica. Ribaltando il problema, se più variabili osservate sembrano avere qualcosa in comune fra loro, quel "qualcosa in comune" potrebbe essere una caratteristica associabile ad una variabile latente. Storicamente, queste caratteristiche comuni a più variabili misurate furono chiamate "fattori" e solo successivamente si cominciò a considerarle delle variabili latenti. Sempre per ragioni storiche, le tecniche che individuano queste variabili sottostanti si chiamano "analisi fattoriale".

In questo tipo di analisi, il problema principale nasce dal fatto che, se la variabile è latente, noi possiamo solo ipotizzare la sua esistenza.

Esistono due diversi tipi di analisi fattoriale, una di tipo esplorativo e una di tipo confermativo. Come esprime bene il loro nome, il primo tipo di analisi si occupa di esplorare le variabili misurate alla ricerca di un andamento (fattore) comune, di isolarlo e di stimarne la potenza e il valore; è quindi un'analisi che si utilizza quando non si hanno le idee chiare, si ipotizza genericamente che le variabili misurate abbiano qualche concomitanza (ad esempio correlazioni alte), ma non si sa esattamente quali siano le variabili e quali e quanti siano i fattori. Il secondo tipo di analisi si utilizza invece quando si conoscono già le relazioni esistenti e si vogliono verificare matematicamente, ad esempio dopo aver effettuato un'analisi esplorativa, si può passare ad un'analisi confermativa per vedere "quanto reggono" effettivamente i risultati trovati.

3.1 Analisi fattoriale esplorativa

Serve per associare una o più variabili latenti (che non si conoscono) ad un gruppo di variabili osservate che si presuppone abbiano qualche cosa in comune, ma non si sa esattamente 'cosa'.

Esempio 1

In una prova (ad esempio un test) sono presenti 4 item relativi a:

1. Contare e ordinare sequenze numeriche;
2. Soluzione di problemi aritmetici;
3. Trovare errori di sintassi;
4. Trovare errori grammaticali.

E' possibile ipotizzare che tutti (o alcuni) di questi item abbiano qualcosa in comune? Se sì, cosa? e con cosa?

Posso ipotizzare la presenza di due variabili latenti, una che ha "qualcosa a che fare" con i numeri e l'altra con la scrittura. Nell'analisi fattoriale esplorativa, comunque, a causa del tipo di approccio matematico, tutte le variabili osservate presenteranno un legame con tutti i fattori (variabili latenti) anche se con "pesi" (importanza) diversi (Fig. 6). Se effettuo i calcoli per una analisi fattoriale esplorativa posso appunto calcolare la "relazione" che ciascun item ha con i fattori. Successivamente, terrò in considerazione le relazioni consistenti e ignorerò quelle talmente piccole da essere insignificanti.

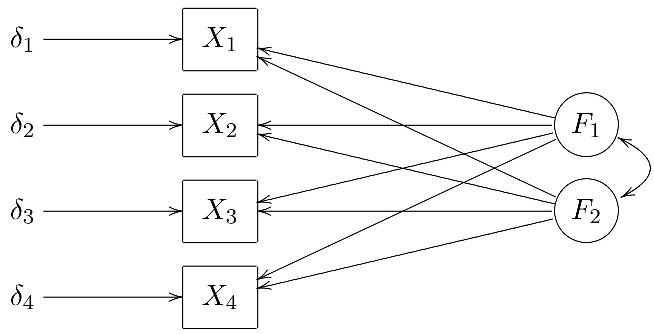


Figura 6: modello relazionale di una analisi fattoriale esplorativa

Nel modello relazionale di Fig. 6, le due variabili latenti sono connesse fra loro da una linea curva e ciò significa che le due variabili latenti hanno una associazione reciproca. Questo vuol dire, in genere, che i fattori sono fra loro correlati (si parla in questo caso di una "soluzione obliqua"). Se le variabili latenti non fossero fra loro correlate, avremmo una "soluzione ortogonale". L'origine di questi nomi (obliqua e ortogonale) nasce dai primi metodi di analisi fattoriale che erano di tipo grafico; in tal caso una soluzione ortogonale implicava lavorare in uno spazio formato da angoli retti (e quindi più semplice), mentre una soluzione obliqua, da uno spazio con angoli diversi da 90°. Va anche ricordato che una correlazione perfetta corrisponde ad un angolo retto, mentre una correlazione maggiore di 0 (in valore assoluto) ma inferiore a 1, corrisponde ad angoli compresi fra 0 e 90°.

Ovviamente, alcune variabili latenti possono essere correlate tra loro e altre no.

3.2 Analisi fattoriale confermativa

Come è già stato anticipato, l'analisi confermativa serve per verificare che una o più variabili latenti che abbiamo ipotizzato legate con determinate variabili osservate siano effettivamente legate a quelle variabili. A differenza dell'analisi esplorativa, solo alcune variabili osservate contribuiscono ai Fattori (variabili latenti) (Fig. 7).

Esempio

Nell'esempio del paragrafo precedente, potremmo essere noi stessi ad ipotizzare che gli item 1 e 2 sottostiano ad una variabile latente di *abilità numerica* e che gli item 3 e 4 indichino una variabile latente di *abilità del linguaggio scritto*.

L'analisi confermativa ci dirà quanto "buono" è il modello che abbiamo ipotizzato.

4 Equazioni strutturali

A questo punto la situazione può complicarsi ulteriormente (ma adeguandosi sempre più alla complessità della realtà studiata). Avendo misurato determinate variabili osservate e ipotizzando una serie di variabili latenti sottostanti, possiamo formulare una serie di relazioni di influenza (causa-effetto) fra le variabili latenti, che ovviamente spiegheranno gli andamenti delle variabili osservate misurate.

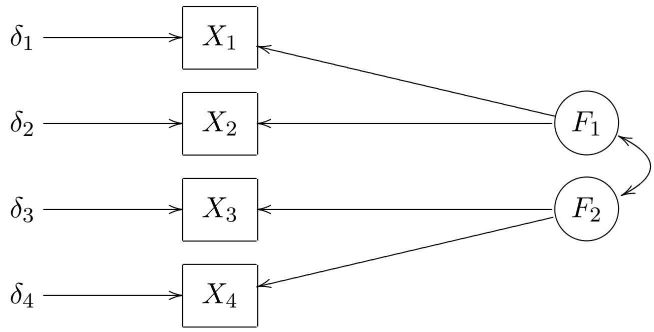


Figura 7: modello relazionale di un'analisi fattoriale confermativa

Esempio

Ipotizziamo (Fig. 8) di aver effettuato la misura dell'intelligenza (`Intellig.`) tramite due test (lo Stanford-Binet e il WISC-R) e di aver anche utilizzato due strumenti per la misura del successo (il California e il Metropolitan). Possiamo ipotizzare che l'intelligenza di un individuo abbia influenza sul suo successo (la freccia che va da `Intellig.` fino a `Successo1`). Possiamo anche immaginare di aver rimisurato il successo di un individuo a distanza di due anni (`Successo2`) sempre tramite gli stessi test. E' facile immaginare che il successo ottenuto precedentemente influenzi quello successivo.

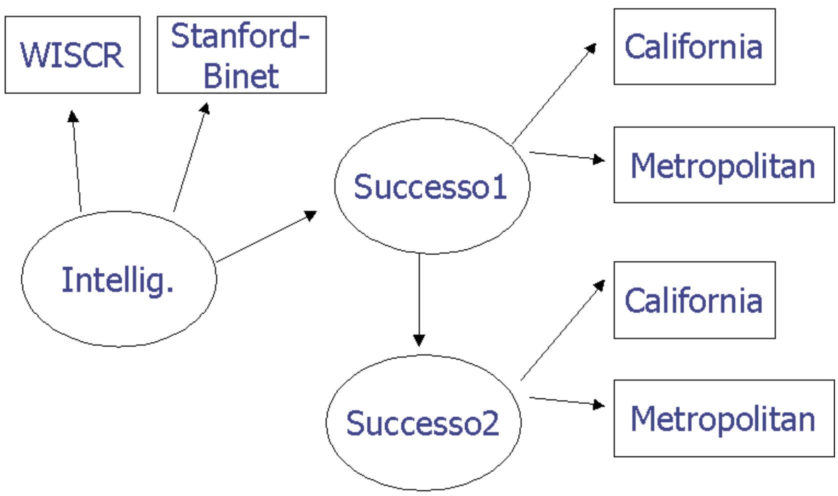


Figura 8: Modello relazionale complesso

In questo modello abbiamo 6 misurazioni effettive (variabili osservate) e 3 variabili latenti. Il modello propone alcune ipotesi sulle relazioni implicate dalle variabili latenti e cerca di spiegare tali relazioni. Se queste relazioni esistono, devono esistere anche fra le variabili osservate che sono direttamente influenzate dalle latenti e tali relazioni devono essere esprimibili tramite funzioni matematiche.

Le variabili osservate che noi usiamo in un modello esplicativo, possono essere organizzate in una tabella di dati; su questa tabella è possibile calcolare una matrice di correlazione di ogni variabile con ciascuna delle altre. I modelli di equazione strutturale, utilizzando relazioni matematiche lineari tipo quelle delle regressioni, cercano di riprodurre le relazioni esistenti fra le variabili osservate. Con queste relazioni, possiamo poi calcolare un'ipotetica matrice di correlazione fra le variabili osservate. A questo punto abbiamo due matrici di correlazioni, una calcolata con i dati osservati direttamente e una ottenuta come stima tramite il modello relazionale. Se confrontiamo le due matrici di correlazione (quella fra le osservate e quella stimata dal modello) possiamo vedere quanto differiscono fra loro. Se differiscono di poco, le relazioni ipotizzate fra le variabili (osservate e latenti) possono essere considerate una spiegazione per le relazioni esistenti fra le osservate.

In linea generale il processo che porta ad un'analisi di equazioni strutturali si svolge in 5 passaggi (Bollen & Lang, 1993, pp. 1-2):

1. Ipotizzare un modello;
2. Identificarlo;
3. Stimarlo;
4. Verificarlo;
5. Ri-formulare il modello.

5 LISREL passo a passo

5.1 Introduzione

In questo capitolo affronteremo `Lisrel`, il più famoso programma per equazioni strutturali, le sue istruzioni e la sua logica in modo graduale. L'obiettivo principale è quello di imparare come ottenere un modello teorico che abbiamo in mente, perciò inizieremo con un modello semplice di due variabili e procederemo con modelli sempre più complessi di tre o più variabili. Non ci focalizzeremo invece sulla scelta del modello migliore, che sarà oggetto di un'altro capitolo.

Nella prima parte del capitolo prenderemo in esame le regressioni lineari (semplici e multiple) (cfr. dispensa) e vedremo la loro riproducibilità tramite i modelli di equazione strutturale. Tramite questi esempi progressivi impareremo le sintassi e la notazione usata da Lisrel. Nella seconda parte del capitolo, amplieremo queste semplici modelli, affrontando le regressioni multivariate e la path analysis, sempre utilizzando le sole variabili osservate. Infine verranno introdotte le variabili latenti, dapprima nell'ottica dell'analisi fattoriale e successivamente nei veri e propri modelli di equazione strutturale più complessi.

Le sintassi riportate non saranno complete perché conterranno solo le parti che serviranno all'intento del capitolo stesso. Sintassi più ampie sono riportate in Appendice, mentre le sintassi complete possono essere trovate sui manuali di Lisrel o nell'help in linea del programma stesso. Per sperimentare concretamente i modelli presentati, potete usare la `versione studenti` del programma (esistono le versioni per Windows, Linux e OSX su Macintosh) che potete trovare all'indirizzo `http://www.ssicentral.com` [Al momento di compilazione di questa dispensa l'url esatto era `http://www.ssicentral.com/lisrel/student.html` e la versione disponibile la 8.8]. Dopo aver scaricato il programma (un file con estensione `exe` per Windows), eseguitelo per installare il programma.

5.2 Sintassi generale LISREL

In questa sezione, verrà presentata a grandi linee la sintassi di Lisrel che poi verrà usata in modo progressivo nella sezione successiva. In Lisrel è possibile eseguire un programma in 4 modi:

- usando i comandi Prelis;
- usando i comandi Simplis;
- usando i comandi Lisrel;
- creando un diagramma di percorso.

In questo libro useremo solo i comandi Lisrel e alcuni comandi tramite l'uso dei menù che, eventualmente, possono produrre i comandi in sintassi Prelis.

Se preferite o se l'approccio grammaticale vi confonde, potete andare direttamente alla sezione 6 e poi tornare a questa di volta in volta per verificare le sintassi e il senso dei comandi costruiti.

Possiamo suddividere la sintassi di Lisrel in 6 blocchi di cui solo 4 sono indispensabili (quelli in corsivo):

- *Descrizione dati*;
- *Inserimento dati*;
- Selezione dati;
- *Modello base*;
- Altri parametri, vincoli;
- *Output*.

Nelle sintassi che seguono, si useranno le seguenti regole:

- le parti in MAIUSCOLO devono essere trascritte esattamente, anche se non è necessario rispettare il maiuscolo;
- le parti in minuscolo indicano informazioni che devono essere fornite al programma sulla base delle nostre necessità;
- le parentesi graffe indicano delle opzioni alternative le une alle altre; almeno una dev'essere indicata;
- ⊔ indica uno spazio di separazione ed è usato qui per sottolineare le parti che devono essere separate dalle altre;
- ↩ significa "andare a capo"; però può essere sostituito da un punto e virgola (;);
- le parti fra parentesi quadre sono facoltative;
- il simbolo di paragrafo (§) indica un'opzione che viene usata di default, cioè automaticamente se non viene effettuata nessuna scelta o se non viene utilizzata quella particolare opzione.

Quando scrivete il programma Lisrel, i simboli ⊔, ↩, §, { e } non devono comparire nelle istruzioni.

Ogni istruzione di Lisrel inizia sulla prima colonna di una riga con almeno le prime due lettere di un *comando* (DA, LA, RA, CM, KM, SD, ME, SE, MO, LK, LE, FR, FI, ST, VA, PD, OU) e seguite da *sottocomandi* (NI, NO, NK, ...). I sottocomandi possono essere indicati in qualunque ordine. Se non è esplicitamente richiesto dalla sintassi del comando, una linea di istruzione è composta da una sola riga che può essere lunga fino a 255 caratteri. Per migliorare la leggibilità si può spezzarla su più righe, usando una `C` alla fine di ogni riga. Sempre per migliorare la leggibilità, le righe spezzate possono essere rientrate di uno o due spazi. I comandi che richiedono degli "a capi" non necessitano della `C` finale. Gli "a capo" possono essere sostituiti da un punto e virgola (;).

5.3 Descrizione dati (DATA)

La prima sezione è quella che descrive i dati su cui andremo a lavorare ed è obbligatoria. L'istruzione `DA` (abbreviazione di `DATA`) deve comparire all'inizio di una riga e tutto il testo che la precede viene ignorato. Serve a comunicare a Lisrel alcune informazioni su cui lavorare.

$$\text{DA} \sqcup \text{NI=ni} \sqcup \text{NO=} \begin{Bmatrix} 0 \\ no \end{Bmatrix} \sqcup MA = \begin{Bmatrix} CM^{\S} \\ KM \\ PM \end{Bmatrix}$$

`ni` è il numero delle variabili che verranno lette, mentre `no` è il numero delle osservazioni (o casi statistici o soggetti) che compongono i dati. Bisogna precisare che `ni` fa riferimento alle variabili fornite al programma nella sezione successiva e non alle variabili da utilizzare nell'analisi. `ni` indica quante variabili devono essere "lette" in fase di *Inserimento dati* (par. 5.4).

Il sottocomando `MA` (`matrix`) serve per specificare quale tipo di matrice vogliamo sottoporre ad analisi. **Attenzione:** non è la matrice dei dati che formiremo, ma quella su cui effettuare i calcoli. Lisrel può usare sia la matrice di covarianza (`CM`) sia quella di correlazione e in questo caso sia una matrice che contiene le correlazioni di Pearson (`KM`) sia le correlazioni policoriche (`PM`).

Esempi:

```
DA NI=6 NO=100 MA=CM
da ni=5 No=36 MA=km
```

5.4 Inserimento dati

$$\begin{Bmatrix} \text{RA} \\ \text{CM} \\ \text{KM} \\ \text{PM} \end{Bmatrix} \begin{Bmatrix} [\sqcup\text{FI}]\text{=nomefile} \\ \hookleftarrow [* \hookleftarrow]\text{dati} \end{Bmatrix}$$

$$\left[\text{SD} \begin{Bmatrix} [\sqcup\text{FI=nomefile} \\ \hookleftarrow [* \hookleftarrow]\text{dati} \end{Bmatrix} \right]$$

$$\left[\text{ME} \begin{Bmatrix} [\sqcup\text{FI=nomefile} \\ \hookleftarrow [* \hookleftarrow]\text{dati} \end{Bmatrix} \right]$$

La seconda sezione è quella dell'inserimento effettivo dei dati che possono essere di diverso tipo: dati grezzi (`RA`) cioè una matrice con le variabili in colonna e i casi statistici in riga oppure una matrice di correlazione (`KM`) o di varianza/covarianza (`CM`) o policorica (`PM`).

`SD` e `ME` indicano rispettivamente le deviazioni standard (`SD`) e le medie (`ME`); si usano generalmente con `KM/PM` o con `CM`. Con l'eccezione dei comandi `SD` e `ME`, uno solo degli altri può essere utilizzato. Dopo le istruzioni `RA`, `CM`, `KM` si deve indicare il file esterno da cui leggere i dati (sottocomando `FI=`) oppure andare a capo e indicare il formato da usare e quindi scrivere i dati direttamente.

Se i dati vengono inseriti direttamente nel programma è possibile usare un formato Fortran oppure il formato libero (indicato dall'asterisco e obsoleto nelle ultime versioni). In quest'ultimo caso, il valore da assegnare ad una variabile viene separato dal successivo con uno spazio. Sia che usiate un file di dati, sia che li scriviate direttamente nel programma, ricordate di usare il *punto decimale* al posto della *virgola*, quindi 1.08 e non 1,08.

Se si indica il nome di un file, si devono seguire le norme del sistema operativo utilizzato. Un semplice nome, indica generalmente un archivio dati nella stessa cartella/directory del programma, mentre se si vuole indicare un percorso assoluto o relativo, va messo fra virgolette o apici.

Il numero di variabili che vengono inserite in questa sezione (direttamente o tramite un file) deve necessariamente coincidere con il parametro `ni` del comando `DA`.

Il comando `PM` che legge una matrice di correlazioni policoriche può essere seguito dal comando `AM` che legge una matrice di varianze/covarianze asintotiche.

$$\mathrm{PM}\left\{\begin{array}{c}[\sqcup\mathrm{FI}=\mathrm{nomefile}\\ \hookleftarrow[*\hookleftarrow]\mathrm{dati}\end{array}\right\}$$

$$\mathrm{AM}\left\{\begin{array}{c}[\sqcup\mathrm{FI}=\mathrm{nomefile}\\ \hookleftarrow[*\hookleftarrow]\mathrm{dati}\end{array}\right\}$$

Esempi

```
RA
*
1 2 3 4 5

RA FI=esempio.dat

CM='C:/dati/lisrel/esempio.cov'
```

```
CM
2.7
1.08 1.03
0.52 0.36 .3

KM
1
.30 1
.17 .75 1
```

5.5 Selezione dati

5.5.1 Label

I dati letti possono contenere molte variabili. Lisrel, nel momento in cui le userà, tenderà ad assegnare loro nomi convenzionali come X1, X2, Y1, Y2 e così via. Se si vuole che le variabili abbiano un nome che abbia un senso per noi, bisogna usare l'istruzione `LA` (abbreviazione di `Label`, cioè etichetta).

$$\mathrm{LA}\left\{\begin{array}{c}\hookleftarrow\\ ;\end{array}\right\}\mathrm{elenco\_nomi}[/]$$

Dopo il comando LA, **si deve andare a capo** e scrivere i nomi da assegnare alle variabili secondo l'ordine con cui compaiono nella sezione precedente. Inoltre bisogna indicare tanti nomi quante sono le variabili definit con il sottocomando `ni`. Le etichette da attribuire vanno separate con uno spazio e se vogliamo includere uno spazio nel nome della variabile, dobbiamo scriverle fra virgolette. Ogni etichetta può essere lunga fino a 8 caratteri, quelli indicati in sovrappiù verranno ignorati.

Se si forniscono meno etichette del dovuto bisogna usare il segno `/` per terminare l'elenco e verranno nominate solo le variabili indicate mentre le altre assumeranno i nomi automatici. Non è possibile saltare alcune variabili nell'assegnare i nomi. Quindi se avete 10 variabili (`ni=10`) e dovete usare solo le ultime 3 tramite le loro etichette, dovete assegnare dei nomi a tutte e dieci le variabili.

Esempi

```
DA NI=7 ...
LA
AGG1 Agg2 LIVStud Voto /
```

```
DA NI=4 ...
LA
Base Terra "Tit. st." Maturita
```

Questa sezione non è obbligatoria. Se non volete assegnare nomi alle variabili osservate, il comando `LA` non è necessario. E' però molto utile attribuire un nome alle variabili osservate in modo che sia poi più facile usarle.

Questo comando dev'essere usato necessariamente dopo DA, ma non necessariamente prima della lettura dei dati o della loro selezione (`RA`, `CM`, `KM`...) o della loro selezione (`SE`).

5.5.2 Select

$$\text{SE}\begin{Bmatrix}\hookleftarrow\\ ;\end{Bmatrix}\begin{Bmatrix}\text{nomi\_var}\\ \text{posizione\_var}\end{Bmatrix}[/]$$

I dati che verranno letti all'inizio (e indicati con il sottocomando `ni`) possono contenere più variabili rispetto a quelle che servono effettivamente nel modello che si intende analizzare, per cui diventa necessario usare il comando `SE` (abbreviazione di `Select`) per selezionarle. Dopo il comando `SE` bisogna andare a capo e indicare la sequenza delle variabili da usare o con il numero della loro posizione oppure con il nome loro assegnato con il comando `LA`. La barra finale (/) termina il comando ed è obbligatoria se si seleziona un sottoinsieme di variabili (sarebbe meglio usarla sempre). Ad es. se il file dati contiene 10 variabili (`DA NI=10`) ma io uso solo le prime tre, userò il comando `select` in questo modo:

```
DA NI=10 ...
SE
1 2 3 /
```

Un secondo utilizzo del comando `SE` è quello di porre le variabili nell'ordine richiesto da Lisrel che si aspetta di trovare dapprima le variabili di tipo Y e poi quelle di tipo X (il senso di questa affermazione diverrà chiaro più avanti).

Anche questa sezione non è obbligatoria, infatti se vengono usate tutte le variabili inserite e se sono già nell'ordine previsto da Lisrel, `SE` non è necessaria. Tuttavia, se si utilizza, deve necessariamente essere usata primo di `MO`.

Esempi

```
SE
4 3 1 2 /

LA
voto livstud agg1 agg2 /
SE
agg1 agg2 voto livstud /
```

5.6 Modello base (MODEL)

Il comando `MOdel` serve a specificare le variabili che serviranno per il modello relazionale.

MO NX=nx [NY=ny] [NE=ne] [NK=nk] [specifica_matrici]

`NX` (*Number of X*) indica il numero di variabili osservate esogene di tipo `X`; `NY` il numero di variabili osservate endogene del tipo `Y`; `NE` (*Number of Eta*) il numero di latenti endogene (il tipo `ETA`) e `NK` (*Number of Ksi*) il numero di latenti esogene (il tipo `CSI`).

Al posto di `specifica matrici` bisogna indicare le matrici implicate nel modello relazionale, la loro forma e come devono essere considerate. Il modello completo di Lisrel prevede 8 matrici diverse: LX (lambda X), LY (lambda Y), BE (beta), GA (gamma), PH (phi), PS (psi), TD (theta delta), TE (theta epsilon). LX, LY, BE e GA indicano i legami fra le variabili e corrispondono ai parametri di regressione. PH è la matrice di varianza/covarianza delle latenti KSI, mentre PS, TD e TE rappresentano le varianze/covarianze degli errori. In tutti i casi la sintassi è del tipo:

matrice=dimensione,riempimento

`matrice` è una delle otto matrici di Lisrel, quindi LX, LY, BE, GA, PH, PS, TD o TE.

`dimensione` serve per indicare al programma come devono essere considerate le matrici e può essere `FU` (*full*=piena; tutte gli elementi possono essere diversi), `SY` (simmetrica; la metà superiore destra è speculare lungo la diagonale principale a quella inferiore sinistra), `DI` (diagonale; si considerano solo i valori lungo la diagonale) o `ZE` (zero).

`riempimento` può essere `FR` (*free*=libera; siamo interessati a tutti gli elementi) o `FI` (fissa; non siamo interessati a nessun elemento). Il parametro `FR` riempie la matrice corrispondente con il valore 1, mentre `FI` la riempie con degli zero.

Non tutti questi simboli si possono usare con tutte le matrici, ma lo vedremo quando ciascuna matrice verrà presentata. Il paragrafo 9 presenta, in un'unica tabella, tutti i simboli e le matrici utilizzate da Lisrel, con le dimensioni e i valori assunti automaticamente.

In linea di massima, le matrici che costituiscono i legami fra le variabili (LX, LY, BE e GA) sono da considerarsi piene (FU) ovvero tutti gli elementi della matrice possono contenere dei valori; la matrice che contengono le varianze/covarianze fra le latenti (PH) è da considerarsi simmetrica (SY) per il significato stesso di covarianza; infine le matrici che contengono le varianze/covarianze degli errori (PS, TD, TE) sono da considerarsi diagonali (DI) perché quasi sempre siamo interessati alle sole varianze, ma possono essere anche simmetriche (SY).

5.7 Altri parametri e vincoli

Tutti i comandi di questa sezione devono essere usati dopo `MOdel` e possono essere indicati in qualunque ordine. Se due linee di istruzioni richiedono operazioni opposte fra loro, avrà valore l'ordine della sequenza dei comandi.

Possono essere divisi in due blocchi, due istruzioni che assegnano etichette alle variabili latenti e una serie di istruzioni che agiscono sui parametri del modello. Per meglio comprendere questa seconda serie di istruzioni è importante ricordare che i parametri corrispondenti agli elementi di una matrice possono essere di tre tipi:

- **fissi**, vale a dire che non vogliamo vengano utilizzati perché non sono stati presi in considerazione nel modello;
- **liberi**, cioè vogliamo utilizzarli e chiediamo al programma di effettuare una stima del loro valore;
- **vincolati**, ovvero li vogliamo usare, però chiediamo al programma di non stimarli e di impostarli, invece, al valore che noi abbiamo deciso.

5.7.1 Etichette delle latenti

$$\text{LK} \begin{Bmatrix} \hookleftarrow \\ ; \end{Bmatrix} \text{etichette [/]}$$

$$\text{LE} \begin{Bmatrix} \hookleftarrow \\ ; \end{Bmatrix} \text{etichette [/]}$$

Questi due comandi servono (analogamente a `LA`) per assegnare dei nomi alle variabili latenti `KSI` (`LK`, label ksi) e a quelle di tipo `ETA` (`LE`, label eta). Ovviamente non è necessario assegnare delle etichette, se non vi servono (in questo caso Lisrel utilizzerà dei nomi automatici). E si possono assegnare solo le etichette delle latenti che si usano effettivamente, vale a dire che non potete assegnare etichette alle `Eta` se non le usate.

La barra finale si usa quando volete assegnare nomi solo ad una parte delle variabili. E' però buona cosa usarla sempre, in quanto evita che si generino errori imcomprensibili se si aggiungono variabili latenti e ci si dimentica di aggiungere un nome.

5.7.2 Fissare parametri

FI ⊔ matrice(riga,colonna)[⊔matrice(riga,colonna)]

Serve a fissare (`FIxed`) i parametri che non si vogliono stimare; Lisrel pone a 0 l'elemento indicato da `riga,colonna` della `matrice`.

Per ogni parametro teoricamente presente nel modello relazionale (gli elementi della matrice), Lisrel può stimare un valore. Con questo comando si inibisce la stima di alcuni parametri, dando ordine al programma di ignorare un certo elemento di una determinata matrice.

Si possono fissare contemporaneamente più elementi di matrici diverse. Inoltre non è strettamente necessario utilizzare la sintassi indicata, è sufficiente che le tre componenti (matrice, riga e colonna) siano presenti e separate l'una dall'altra. L'uso di `FI` dipende dal modo in cui è stata dichiarata la matrice nel comando `MO`: se una matrice è stata dichiarata `FR` (libera), si usa `FI` per bloccare i parametri da non stimare.

Esempi

```
FI LX(1,1)
FI GA 1,2 BE 2 1
```

5.7.3 Liberare parametri

FR ⊔ matrice(riga,colonna)[⊔matrice(riga,colonna)]

Svolge la funzione opposta a `FI` e serve a liberare (`FRee`) i parametri che vogliamo stimare, se la matrice (nel suo complesso) è stata dichiarata `FI` nel comando `MO`. Lisrel pone a 1 l'elemento indicato da `riga,colonna` della `matrice`.

Esempi

```
FR LX(1,1)
FR GA 1,2 BE 2 1
```

5.7.4 Fissare e liberare in blocco

PA ⊔ matrice ↩ [matrice]

Questa istruzione (`PAttern`) serve per sostituire i comandi `FI` e `FR`; `matrice` è una delle otto matrici di Lisrel (`LX`, `LY`, `PH`, `PS`, `TD`, `TE`, `BE`, `GA`), mentre `[matrice]` è il contenuto da assegnare a quella matrice. Gli elementi da usare sono solo lo zero (`0`) e l'uno (`1`). Zero corrisponde al comando `FI`, mentre 1 al comando `FR` e quindi i parametri corrispondenti all'elemento messo a 1 verranno stimati.

Esempi

```
MO ... NE=3 NK=2 ...
PA GA
1 0
0 1
0 0
```

```
MO ... NE=3 NK=2 ...
PA GA
1 0 0 1 /
```

Non è necessario che gli elementi siano disposti a matrice, possono anche essere scritti tutti su una stessa riga; vanno ovviamente indicati tanti elementi quanti sono quelli della matrice considerata. Se si vogliono indicare meno elementi del dovuto, si può usare una / per indicare che tutti gli elementi successivi sono da considerarsi uguali a 0.

5.7.5 Assegnare un valore ai parametri

VA ⊔ valore ⊔ matrice(riga,colonna)

ST ⊔ valore ⊔ matrice(riga,colonna)

Servono ad impostare un certo parametro ad un certo valore.

In teoria il comando `VA` (`Value`) serve per impostare un determinato parametro ad un certo valore, mentre `ST` (`Starting Value`) serve per suggerire un valore da cui partire per la stima del parametro. Però il loro comportamento dipende interamente dal fatto che il parametro sia stato precedentemente impostato come fisso (da non stimare) o come libero (da stimare). Se il parametro è fisso entrambi i comandi assegnano il valore ai parametri indicati e tale valore sarà la stima finale del parametro. Se invece il parametro è libero, entrambi i comandi suggeriscono a Lisrel di iniziare la stima a partire da un certo valore, ma il parametro potrà cambiare nel corso dell'elaborazione.

Esempi

```
ST .5 LX(1,1)
VA 1.0 GA 1,2 BE 2 1
```

5.7.6 Assegnare un blocco di valori

MA ⊔ matrice ←˒ [matrice]

Questa istruzione (`Matrix Assign`) costituisce una combinazione dei comandi `ST`, `VA` e `PA` in quanto serve per impostare i parametri di una matrice a particolari valori; `matrice` è una delle otto matrici di Lisrel (`LX`, `LY`, `PH`, `PS`, `TD`, `TE`, `BE`, `GA`), mentre `[matrice]` è il contenuto da assegnare a quella matrice. Gli elementi da usare sono i valori che si vogliono assegnare ai singoli parametri e possono essere diversi da 0 e da 1. Anzi, con questo comando, zero e uno corrispondono al loro reale valore numerico e quindi assegnano il valore 0 o il valore 1 al parametro. Per essere ancora più espliciti, 0 e 1 *non indicano i parametri da stimare o da fissare* Inoltre, i valori inseriti nella matrice, svolgono funzione di VA o ST in base al fatto che la matrice (o il singolo parametro) sia `FI` o `FR`.

Esempio

```
MO NX=6 NK=2 LX=FU,FI...
MA LX
.75 0
.68 0
.59 0
 0  .82
 0  .79
 0  .65
```

In questo esempio i valori impostati saranno i parametri effettivi che il programma userà, in quanto la matrice `LX` è stata definita come fissa.

Anche in questo caso, disporre i valori in forma di matrice ha solo lo scopo di facilitare la lettura perché i dati possono essere scritti in modo continuo, separati da uno spazio.

5.8 Output

5.8.1 Grafico

PD

Il comando PD (`path diagram`) disegna il diagramma del modello relazionale impostato ed è molto utile da usare per diversi motivi:

1. se il programma contiene un errore il diagramma non viene creato;
2. se il diagramma viene creato possiamo verificare velocemente se corrisponde con quello che avevamo intenzione di studiare;
3. usando il menù a tendina della finestra del grafico, possiamo vedere diverse informazioni: i parametri stimati, quelli standardizzati, i t di Student, gli indici di modifica. Quest'ultimi vengono visualizzati nel grafico anche se non sono stati esplicitamente richiesti fra i risultati.

5.8.2 Risultati

$$\text{OU} \sqcup \begin{Bmatrix} \text{ALL} \\ \text{ME, RS, FS, ...} \end{Bmatrix}$$

Il comando `OU` (`OUtput`) indica a Lisrel che cosa deve stampare nei risultati. In genere è sufficiente scrivere `OU` senza altri parametri.

I parametri più utili, sono:

- `ND=n`, stampa tutti i risultati usando n decimali (il default è 2);
- `NP=n`, numero di decimali da usare quando i numeri vengono scritti in un file esterno (il default è 3);
- `MI`, stampa gli indici di modifica;
- `EF`, stampa gli effetti indiretti e quelli totali;
- `FS`, stampa i punteggi fattoriali
- `SS`, le variabili latenti vengono standardizzate, ma non le osservate;
- `SC`, sia le variabili latenti che quelle osservate vengono standardizzate.

Altri parametri sono ormai obsoleti o si usano solo quando richiesti dal programma:

- `AD=n`, imposta il numero di interazioni massimo per arrivare ad una soluzione; se il numero di iterazioni supera quel valore (per default 20) il programma si ferma;
- `ME=metodo`, indica il tipo di algoritmo da usare per la stima dei parametri. Sono diponibili: ML (*maximum likelihood*, massima verosimiglianza), UL (*unweighted least squares*, minimi quadrati non pesati), GL, WL...

In ogni caso, tutti questi parametri si possono trovare (con relativa spiegazione) nell'help all'interno del programma.

6 Alcuni semplici programmi in Lisrel

Se avete scaricato la versione studenti di Lisrel e l'avete installata, potete provare immediatamente i programmi che seguono.

Per scriverli, avviate Lisrel, scegliete `File|New`, quindi selezionate `Syntax only` dal menù a discesa che compare. Nella nuova finestra (vuota) che si apre, scrivete le istruzioni e salvatele con un nome (preferibilmente con estensione `.ls8`). **Attenzione**: Lisrel non esegue il programma se non lo avete salvato almeno la prima volta. Per eseguirlo, è sufficiente fare clic sul pulsante dell'omino che corre su cui è sovrapposta una L.

6.1 Regressione semplice

Proviamo a scrivere le istruzioni Lisrel per una regressione lineare semplice (usando i dati di Tab.3). Applicando le formule della regressione (cfr. la dispensa corrispondente), i parametri che dovremmo ottenere sono: $b = 0.807$, $a = 0.362$ e $r^2 = 0.894$.

Tabella 3: Dati di esempio

Y	X
3	2
2	3
4	5
5	7
8	8
22	25

Iniziamo con la descrizione dei dati (comando `DA`): abbiamo due variabili in lettura (quindi `ni=2`) e 5 casi (`no=5`). Usiamo l'istruzione `RA` per inserire i dati e l'istruzione `LA` per assegnare un nome alle due variabili (anche se, per semplicità, usiamo i nomi classici). Nel comando `MO` dobbiamo indicare che abbiamo una variabile di tipo X e una di tipo Y. Infine chiediamo la visualizzazione del grafico e i risultati generici (ma con 3 decimali).

I dati vanno inseriti riga per riga, ovvero tutti i dati di un soggetto (un valore per ognuna delle variabili indicate, in questo caso due) e poi quelle del secondo soggetto e così via. Non è necessario andare a capo ad ogni soggetto, l'importante è che ogni valore sia separato dall'altro da almeno uno spazio (perché abbiamo scelto il formato "libero").

Qui di seguito le istruzioni complete (il punto esclamativo indica l'inizio di un commento):

```
Esempio di regressione lineare semplice

da ni=2 no=5 ma=cm
RA ! asterisco facoltativo
*
3 2 2 3 4 5 5 7 8 8
la
Y X
mo nx=1 ny=1
pd
OU nd=3
```

Notate come abbiamo scritto alcune istruzioni in minuscolo e altre in maiuscolo: per Lisrel non ha importanza.

Se salvate queste istruzioni nel file `esempio01.ls8`, quando lo eseguirete, nella stessa directory otterrete un file `esempio01.out` che contiene i risultati in formato testo e un file `esempio01.pth` che contiene il grafico (richiesto con il comando `PD`). Alla fine dell'esecuzione, il programma Lisrel apre automaticamente i due file dei risultati; dovreste quindi ritrovarvi con tre finestre aperte: il programmino che avete scritto, i risultati testuali, il grafico.

Quest'ultimo dovrebbe essere la finestra che compare al livello più alto e il grafico contenuto (ignorando i colori) dovrebbe corrispondere alla Fig.9. Dal grafico vediamo già che il parametro di regressione vale 0,81; gli altri valori indicati sono rispettivamente la varianza della X (6.50) e la varianza degli errori della Y (1.06).

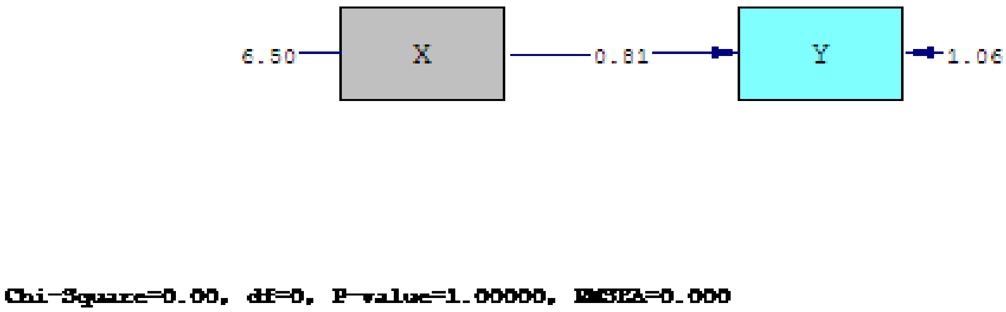


Figura 9: Esempio01 - Path Diagram

Analizzando i risultati (finestra con estensione `.out`), vediamo e commentiamo solo i più importanti (i risultati sono stati ri-impaginati). Dapprima Lisrel riepiloga le variabili:

```
Number of Input Variables  2
Number of Y - Variables    1
Number of X - Variables    1
Number of ETA - Variables  1
Number of KSI - Variables  1
Number of Observations     5
```

Possiamo notare che, oltre alle informazioni che abbiamo indicato noi nelle istruzioni (`NI`, `NO`, `NY`, `NX`), Lisrel aggiunge il numero di `ETA` e di `KSI` ponendole entrambe pari a 1. *Eta* e *ksi* sono i nomi che Lisrel usa per le variabili latenti. Quando però ci sono solo variabili osservate, Lisrel assume un'eguaglianza automatica fra `X` e `KSI` e fra `Y` e `ETA`, quindi sottintende due variabili latenti, una endogena e una esogena che coincidono perfettamente con le osservate (vedremo poi cosa significa).

```
    Covariance Matrix                        Means

              Y          X                  Y          X
       --------   --------           --------   --------
   Y      5.300                         4.400      5.000
   X      5.250      6.500
```

Segue la matrice di varianza/covarianza calcolata usando le stime (cioè con $N-1$ anziché N) e le medie di ciascuna delle variabili osservate. Quindi, Lisrel elenca i parametri che verranno stimati (`Parameter Specifications`) numerandoli in modo progressivo.

Infine i parametri che ci servono (`LISREL Estimates`). Gamma (γ) è il nome che Lisrel assegna al legame fra le ETA (η) e le KSI (ξ) (in questo contesto ricordiamo che `eta=y` e `ksi=x`) e quindi, nel nostro caso, corrisponde al parametro di regressione $b = .807$ (arrotondato diventa .81); Phi (ϕ) è la varianza delle KSI e corrisponde a quella già calcolata in precedenza (6.50); infine PSI (ψ) è la varianza degli errori. In tutti e tre i casi, Lisrel stampa il valore del parametro, la sua deviazione standard fra parentesi tonde e il valore di t di Student (t-test). Dividendo il parametro per la deviazione standard si ottiene il valore t. Tale valore è da considerarsi statisticamente significativo se superiore a 2 (in valore assoluto).

```
 LISREL Estimates (Maximum Likelihood)
```

```
       GAMMA        PHI         PSI

           X           X           Y
    --------    --------    --------
Y      0.808       6.500       1.060
     (0.233)     (5.307)     (0.865)
       3.465       1.225       1.225
```

Gli ultimi due parametri sono rispettivamente l'r^2 (`Squared Multiple Correlations for Structural Equations`) e la costante (`Alpha`) ovvero l'intercetta.

```
Squared Multiple Correlations          ALPHA
for Structural Equations
                                           Y
          Y                         --------
   --------                            0.362
      0.800                          (1.308)
                                       0.276
```

Infine le statistiche relative alla bontà del modello. In questo caso, il modello è saturo (cioè tutti i parametri stimabili sono stati stimati) e quindi la bontà del modello è massima, il χ^2 è 0 e anche i gradi di libertà (torneremo più avanti sul significato di questi indici).

```
                   Goodness of Fit Statistics

                      Degrees of Freedom = 0
        Minimum Fit Function Chi-Square = 0.00 (P = 1.00)
Normal Theory Weighted Least Squares Chi-Square = 0.00 (P = 1.00)

          The Model is Saturated, the Fit is Perfect !
```

[GRAFICO PD da confrontare con quello di fig. 1.2]

6.1.1 Esercizio

Tenendo in considerazione le istruzioni usate nell'esempio precedente e le sintassi presentate in precedenza (pag. 11), provate a scrivere le istruzioni per eseguire la regressione lineare semplice dei dati di Tab. 4.

Per facilitarvi, osservate attentamente il modello relazionale di Fig. 10 a pag. 22 e considerate che avete 2 variabili e 8 osservazioni.

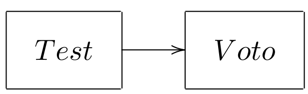


Figura 10: modello relazionale

Tabella 4: Dati fittizi

	Test	Voto
A	12	8
B	10	7
C	14	8
D	9	5
E	9	6
F	13	9
G	11	7
H	8	5

Provate poi a scrivere queste istruzioni direttamente nel programma Lisrel e cercate di identificare i parametri corrispondenti all'intercetta, alla pendenza e alla proporzione di varianza spiegata.

6.1.2 Soluzione

Si tratta di copiare le istruzioni precedenti con tre modifiche:

1. `no=8` perché abbiamo 8 osservazioni anziché 5;
2. usiamo l'istruzione `LA` per assegnare i nomi alle variabili;
3. usiamo l'istruzione `SE` per porre le variabili nell'ordine richiesto da Lisrel; nei dati, la variabile `test` (X) viene prima di `voto` (Y), ma Lisrel vuole che siano indicate prima le variabili dipendenti e poi quelle indipendenti.

Confrontate il programma scritto da voi con quello che segue:

```
da ni=2 no=8 ma=cm
ra
12 8 10 7 14 8 9 5 9 6 13 9 11 7 8 5
la
test voto
se
voto test
mo nx=1 ny=1
pd
ou
```

6.2 Regressione multipla

Estendiamo il concetto e proviamo a scrivere il programma LISREL per la regressione multipla indicata dai dati di Tab. 5.

Si tratta di ampliare e adattare i due programmi scritti finora, in quanto abbiamo 4 variabili indipendenti e una dipendente. Quindi `NI` diventerà 5, dobbiamo cambiare le etichette e usare SE per metterle nell'ordine corretto e quindi impostare `NX` al valore appropriato:

```
DA NI=5 NO=5 MA=CM
RA
```

X_1	X_2	X_3	X_4	Y
1	2	1	1	1
2	3	1	3	2
4	3	2	2	3
5	5	2	3	4
2	4	2	2	2

Tabella 5: Dati per la regressione multipla

```
1 2 1 1 1 2 3 1 3 2 4 3 2 2 3 5 5 2 3 4 2 4 2 2 2
LA
X1 X2 X3 X4 Y
SE
5 1 2 3 4 /
MO NY=1 NX=4
PD
OU
```

Dal momento che abbiamo attribuito dei nomi alle variabili, in alternativa, potevamo usare anche il comando:

```
SE
Y X1 X2 X3 X4 /
```

L'istruzione `LA`, così come l'abbiamo scritta qui (cioè per assegnare le etichette `X1`, `X2`...) è del tutto inutile perché Lisrel fa esattamente la stessa cosa se non usiamo il comando `LA`.

6.2.1 Uso di `select`

Torniamo un attimo sull'istruzione `SE`.

I dati della Tab. 5 mostrano che le variabili sono elencate nell'ordine X_1, X_2, X_3, X_4 e Y. Questo, in genere, significa che il file dei dati contiene le variabili esattamente in quest'ordine. Poiché l'ordine con cui abbiamo i dati non è lo stesso necessario per Lisrel, usiamo l'istruzione `SE` per metterle nell'ordine corretto. Analizziamo la relazione fra i comandi SE e MO.

```
DA NI=5 NO=5 MA=CM
SE
5 1 2 3 4 /
MO NY=1 NX=4
```

Il sottocomando NY ci informa che abbiamo una variabile Y. Siccome Lisrel si aspetta di trovare le Y indicate per prime, la prima variabile del comando SE sarà una Y (in questo caso la quinta).

Il sottocomando NX ci dice che abbiamo quattro X e siccome devono venire dopo le Y, le variabili dalla seconda alla quinta (4+1) saranno le X. Guardando il comando SE, vediamo che la prima variabile dei dati (X_1) sarà anche la prima X di Lisrel e così via.

Proviamo con un esempio un po' più complesso:

```
DA NI=7 NO=90 MA=CM
LA
ISMS ISSMS TP CERCA CONS REDD INDUS /
```

```
....
SE
 6 5 4 3 7 /
MO NX=1 NY=4 ...
```

In totale il modello usa 5 variabili su 7. Abbiamo 4 variabili di tipo Y, quindi le prime quattro variabili indicate dal comando SE sono le Y e la successiva è una X. Le prendiamo in questo ordine:

	variabile			
Ordine	numero	nome	tipo	latente
1	6	REDD	Y_1	η_1
2	5	CONS	Y_2	η_2
3	4	CERCA	Y_3	η_3
4	3	TP	Y_4	η_4
5	7	INDUS	X_1	ξ_1

6.2.2 Le matrici implicate

Nel programma precedente, avevamo 4 X e 1 Y, a livello di Lisrel queste diventano rispettivamente delle KSI (ξ) e delle ETA (η). Lisrel chiama i legami fra le Ksi e le eta con il nome di gamma (γ) e li pone nella matrice GA che, incrociando 1 eta e 4 ksi, avrà dimensione 1×4. Dal momento che nell'istruzione MO non abbiamo indicato né la forma né il contenuto di Gamma, Lisrel assume il default, cioè `GA=FU,FR` che corrisponde a GA=[γ_{11} γ_{12} γ_{13} γ_{14}] (v. Fig. 11) e stima tutti e quattro i parametri. Se cercate la matrice Gamma nei risultati dell'esempio precedente, la troverete indicata due volte, una prima volta nella zona `Parameter Specifications` e una seconda nella zona `Lisrel Estimates`. Nel primo caso, Lisrel numera progressivamente tutti i parametri della matrice che dovrà stimare; nel secondo caso, indica i parametri che ha effettivamente stimato. In Lisrel, quando si lavora con sole variabili osservate, sono previste anche le matrici BE, PH e PS che spieghiamo subito. Beta (BE) è la matrice che indica i legami fra le Eta. In questo esempio abbiamo una sola eta e quindi il legame con se stessa non va stimato. In effetti, non avendo indicato né la forma né il contenuto di BE, Lisrel ha usato il default, cioè `BE=ZE,FI`.

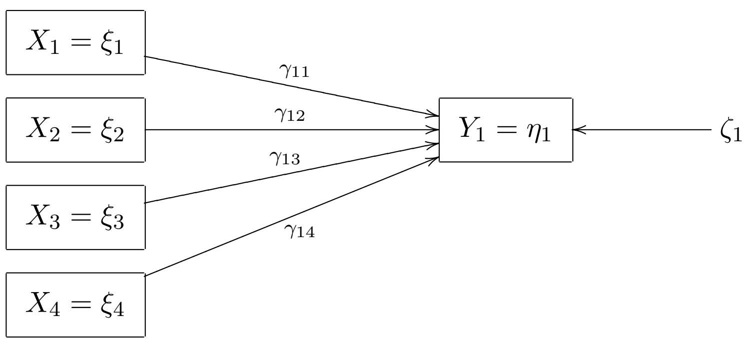


Figura 11: modello relazionale dei dati di Tab. 5

6.3 Regressione multivariata

Per estendere ulteriormente il concetto, ampliamo la possibilità del modello introducendo una seconda variabile indipendente.

Ipotizziamo di studiare le relazioni che legano le valutazioni dell'insegnante da parte dei suoi studenti e contemporaneamente la valutazione che l'insegnante fa dei suoi studenti. Abbiamo misurato 6 variabili, 3 riguardano l'insegnante e 3 riguardano lo studente (Tab. 6). Ipotizziamo che la chiarezza dell'esposizione (`CHIAR`), gli stimoli ricevuti (`STIMOLI`) e l'interesse suscitato (`INTERES`) spieghino la valutazione che gli alunni hanno dato all'insegnante (`VALINS`), mentre la valutazione che l'insegnante ha dato agli studenti (`VALSTUD`), cioè il voto, dipende dalla chiarezza dell'esposizione (`CHIAR`), dagli stimoli ricevuti e dalle conoscenze acquisite (`CONOS`).

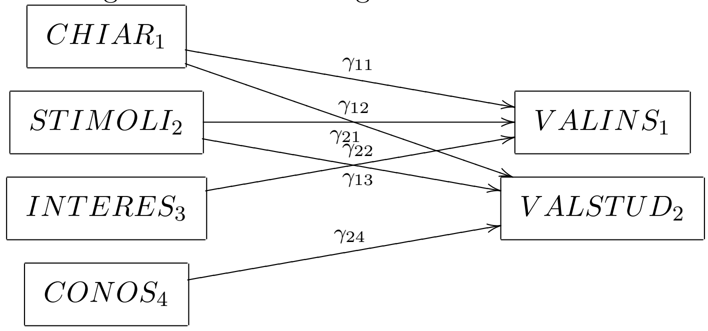
V1	VALINS	valutazione finale sull'insegnante	Ins.
C1	CHIAR	chiarezza dell'esposizione	Ins.
S2	STIMOLI	stimoli ricevuti durante il corso	Stu.
C4	CONOS	conoscenze acquisite	Stu.
I3	INTERES	interesse suscitato	Ins.
V2	VALSTUD	valutazione finale sullo studente	Stu.

	Matrice varianza/covarianza					
	V1	C1	S2	C4	I3	V2
V1	0.64					
C1	0.74	1.17				
S2	0.64	0.73	1.19			
C4	0.14	0.04	0.05	0.38		
I3	0.27	0.17	0.27	0.28	0.62	
V2	0.42	0.51	0.41	0.02	0.25	0.52

Tabella 6: Esempio di regressione multivariata

La Fig. 12 presenta queste relazione secondo le regole della modellazione causale; le variabili sulla sinistra sono tutte variabili indipendenti e quelle sulla destra dipendenti. Ai nomi delle variabili è stato aggiunto un numero progressivo per facilitare le indicazioni dei parametri da stimare.

Figura 12: Modello regressione multivariata



Essendo tutte variabili osservate, Lisrel farà coincidere le variabili indipendenti con le KSI e le dipendenti con le ETA. Notiamo come il modello di Fig. 12 è suddivisibile in due regressioni multiple:

1. VALINS = $b_1\times$ CHIAR + $b_2\times$ STIMOLI + $b_3\times$ INTERES
2. VALSTUD = $b'_1\times$ CHIAR + $b'_2\times$ STIMOLI + $b'_3\times$ CONOS

Il programma Lisrel è questo:

```
DA NI=6 NO=32 MA=CM
CM
0.64
0.74  1.17
0.64  0.73  1.19
0.14  0.04  0.05  0.38
0.27  0.17  0.27  0.28  0.62
0.42  0.51  0.41  0.02  0.25  0.52
LA
VALINS1 CHIAR1 STIMOLI2 CONOS4 INTERES3 VALSTUD2
SE
1 6 2 3 5 4 /
MO NY=2 NX=4 BE=FU,FI GA=FU,fi
FR GA 1 1 GA(1,2) ga(1,3) C
   GA 2 1 GA 2 2 GA 2 4
PD
OU
```

In alternativa:

```
MO ... GA=FU,fr
FI GA 1 4 GA(2,3)
```

La matrice gamma (dai risultati) mostra i parametri che abbiamo chiesto di stimare, la loro deviazione standard e la statistica t di Student associata ad ogni parametro. La funzione di questa statistica è quella di permetterci di sapere se l'apporto del parametro al modello è significativo o meno. L'ipotesi nulla che viene utilizzata è H_0 : stima = 0. Questa ipotesi viene verificata tramite un t di Student a livello $\alpha = .05$, il cui valore critico (ipotesi bidirezionale) è 1.96, generalmente arrotondato a 2 perché è più facile da identificare.

```
 LISREL Estimates (Maximum Likelihood)

         GAMMA

              CHIAR1   STIMOLI2   INTERES3     CONOS4
            --------   --------   --------   --------
  VALINS1       0.48       0.19       0.22        - -
              (0.07)     (0.07)     (0.08)
                7.03       2.77       2.85

 VALSTUD2       0.35       0.13        - -      -0.01
              (0.12)     (0.12)                (0.17)
                2.92       1.07                 -0.03
```

Valori inferiori a 2 (in valore assoluto) indicano parametri che possono essere eliminati; in questo caso i legami γ_{22} (1.07) e γ_{24} (-0.03), cioè non è vero che la valutazione dello studente pu= essere spiegata sulla base degli stimoli ricevuti e/o delle conoscenze acquisite. però non vanno eliminati entrambi immediatamente, ma è meglio eliminarli uno alla volta, in questo caso partendo da quello meno significativo in assoluto (γ_{24}).

Se affrontiamo i calcoli con la tecnica della regressione multipla, otteniamo gli stessi risultati:

$$VALINS1 \quad \begin{bmatrix} 1.17 & .73 & .17 \\ .73 & 1.19 & .27 \\ .17 & .27 & .62 \end{bmatrix}^{-1} \begin{bmatrix} .74 \\ .64 \\ .27 \end{bmatrix} = \begin{bmatrix} .48 \\ .19 \\ .22 \end{bmatrix}$$

$$VALSTUD2 \quad \begin{bmatrix} 1.17 & .73 & .04 \\ .73 & 1.19 & .05 \\ .04 & .05 & .38 \end{bmatrix}^{-1} \begin{bmatrix} .51 \\ .41 \\ .02 \end{bmatrix} = \begin{bmatrix} .35 \\ .13 \\ -.01 \end{bmatrix}$$

La matrice PHI riproduce la matrice di varianza/covarianza che abbiamo utilizzato in input, mentre PSI indica la varianza degli errori.

```
PSI
Note: This matrix is diagonal.

    VALINS1   VALSTUD2
   --------   --------
       0.09       0.28
     (0.02)     (0.08)
       3.67       3.67


Squared Multiple Correlations for Structural Equations

    VALINS1   VALSTUD2
   --------   --------
       0.86       0.45
```

Avendo posto due variabili dipendenti, la matrice `Alpha` ci mostra un valore di intercetta (o costante) per ciascuna.

```
ALPHA

    VALINS1   VALSTUD2
   --------   --------
       0.02       1.11
     (0.21)     (0.40)
       0.10       2.75
```

Quando stimiamo i parametri di un modello relazionale, stiamo cercando di vedere se, quel modello ridotto rispetto a quello saturo, è sufficiente per spiegare i dati di partenza. In base ai parametri stimati, viene stimata una matrice di varianza/covarinza che viene poi confrontata con quella di partenza. A questo punto abbiamo due matrici di varianza: una di dati reali, osservati ($\sum$) e una di valori attesi sulla base del modello ($\sum(\Theta)$). La differenza fra queste due matrici di varianze produce dei residui, cioè gli errori del modello. Se applichiamo una statistica di chi-quadro (cioè un confronto statistico fra valori osservati e valori teorici) a queste due matrici, otterremo un numero tanto più piccolo quanto più piccoli sono i residui. Usando la statistica del χ^2, confrontata con la relativa distribuzione e i relativi gradi di libertà, sappiamo che piccoli valori di χ^2 corrispondono ad alti valori di probabilità e viceversa. Quindi a noi servono piccoli valori di χ^2 che siano statisticamente non significativi (cioè alti valori di probabilità). La sezione `Goodness of Fit Statistics` di Lisrel raccoglie un insieme di indici statistici che si usano per stabilire se il modello è "buono" in se e per se (adeguamento assoluto), se un modello è migliore di un altro (indici di confronto).

```
                    Goodness of Fit Statistics

                       Degrees of Freedom = 3
          Minimum Fit Function Chi-Square = 18.44 (P = 0.00036)
   Normal Theory Weighted Least Squares Chi-Square = 12.87 (P = 0.0049)
             Estimated Non-centrality Parameter (NCP) = 9.87
          90 Percent Confidence Interval for NCP = (2.30 ; 24.94)

                  Minimum Fit Function Value = 0.59
            Population Discrepancy Function Value (F0) = 0.37
           90 Percent Confidence Interval for F0 = (0.085 ; 0.92)
           Root Mean Square Error of Approximation (RMSEA) = 0.35
          90 Percent Confidence Interval for RMSEA = (0.17 ; 0.55)
           P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0071

               Expected Cross-Validation Index (ECVI) = 2.25
          90 Percent Confidence Interval for ECVI = (1.75 ; 2.59)
                    ECVI for Saturated Model = 1.56
                  ECVI for Independence Model = 4.89

  Chi-Square for Independence Model with 15 Degrees of Freedom = 120.07
                      Independence AIC = 132.07
                          Model AIC = 60.87
                        Saturated AIC = 42.00
                      Independence CAIC = 146.86
                         Model CAIC = 120.04
                        Saturated CAIC = 93.78

                     Normed Fit Index (NFI) = 0.85
                   Non-Normed Fit Index (NNFI) = 0.27
                Parsimony Normed Fit Index (PNFI) = 0.17
                   Comparative Fit Index (CFI) = 0.85
                   Incremental Fit Index (IFI) = 0.87
                    Relative Fit Index (RFI) = 0.23

                        Critical N (CN) = 20.08

                Root Mean Square Residual (RMR) = 0.040
                        Standardized RMR = 0.072
                   Goodness of Fit Index (GFI) = 0.88
              Adjusted Goodness of Fit Index (AGFI) = 0.15
             Parsimony Goodness of Fit Index (PGFI) = 0.13
```

Tra tutti gli indici riportati da Lisrel i primi da considerare sono nelle due righe iniziali: un $\chi^2 = 18,44$ con 3 gradi di libertà corrisponde ad una probabilità di 0.00036, che confrontata con un livello α convenzionale del 5%, risulta significativa. Quindi dobbiamo concludere che il modello interpretativo dei dati che abbiamo utilizzato non si adegua (in inglese, *fit*) a sufficienza ai dati reali, indicando che i residui sono grandi rispetto ai valori attesi.

Gli indici di modifica funzionano al contrario rispetto ai t di Student. Il loro scopo è quello di evidenziare legami che non abbiamo preso in considerazione e che andrebbero invece aggiunti. Gli indici indicano di quanto diminuirebbe il chi-quadro generale se aggiungessimo al modello quel parametro. Si considerano, di solito, indici di modifica maggiori di 4. Nel nostro caso gli indici di modifica ci suggeriscono di aggiungere un legame γ_{23} e/o un legame β_{21}. Anche per gli indici di modifica si procede un passo alla volta.

```
Modification Indices and Expected Change
```

```
        Modification Indices for BETA

            VALINS1   VALSTUD2
           --------   --------
 VALINS1        - -       1.90
VALSTUD2       5.34        - -

        Modification Indices for GAMMA

             CHIAR1   STIMOLI2   INTERES3     CONOS4
           --------   --------   --------   --------
 VALINS1        - -        - -        - -       2.64
VALSTUD2        - -        - -       6.65        - -

No Non-Zero Modification Indices for PHI

        Modification Indices for PSI

            VALINS1   VALSTUD2
           --------   --------
 VALINS1        - -
VALSTUD2       1.92        - -
```

Il procedimento di aggiustamento del modello, chiamato "modellamento" ma anche *fitting*, ha lo scopo di arrivare ad un modello che si adegui il più possibile. Per far questo si devono cancellare i legami non significativi (t-test) o aggiungere i legami importanti (indici di modifica), sempre uno alla volta. Ma è un procedimento che affronteremo in dettaglio più avanti.

6.4 Un modello più complesso

La regressione lineare semplice pu= servire sia per "predire l'andamento" di una variabile a partire da un'altra, sia come "spiegazione causa-effetto" fra le due variabili. Nella realtà, mentre la **funzione predittiva** si applica ai soli dati osservati (anche se sono numeri generati a caso), quella esplicativa deve avere anche una sussistenza logica e teorica. Ne consegue che la regressione lineare semplice utilizzata con funzione esplicativa può essere considerata come un'eccessiva semplificazione della realtà, se l'ambito in cui si inserisce è sufficientemente complesso (come l'economia, la sociologia e la psicologia).

L'aggiunta di ulteriori variabili predittive (regressione lineare multipla) sebbene migliori entrambe le funzioni, è ancora una semplificazione pesante della realtà. Infatti una regressione lineare multipla ipotizza che le variabili predittrici siano fra loro pochissimo correlate (ma correlate), al punto che se le correlazioni fra le variabili indipendenti fossero nulle, la regressione multipla si ridurrebbe ad una serie di regressioni semplici.

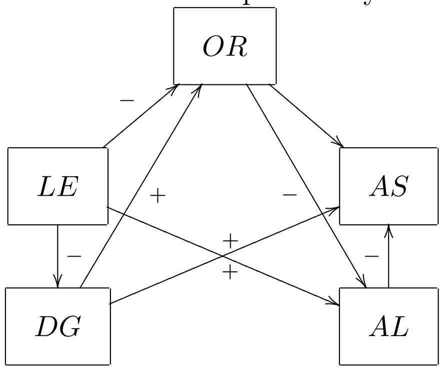
Tuttavia, l'effetto delle variabili predittrici sulle variabili dipendenti è diretto, o almeno, questo è quello che viene sottinteso dalla regressione lineare semplice e multipla.

E' invece possibile che le correlazioni fra le variabili predittive indichino una possibile relazione causale. Per questo motivo, storicamente, si è sviluppata la tecnica della *path analysis* che si è poi evoluta nei modelli di equazione strutturale.

Facciamo un esempio concreto considerando il modello relazionale rappresentato in Fig. 13 derivato da una ricerca di Hoge e Carroll (1975) e ripreso da Tacq (1997).

Studiando la relazione fra le credenze cristiane (in particolare l'ortodossia religiosa, OR) e l'atteggiamento anti-semita (AS), molti ricercatori trovano una correlazione positiva. Sappiamo già che questa correlazione potrebbe dipendere da un'altra variabile che potrebbe avere influenza su entrambe,

Figura 13: Modello di path analysis



sia il credere sia l'anti-semitismo. Cercando di capire questa relazione, possiamo raccogliere altre informazioni e trovare che anche l'atteggiamento dogmatico (DG) correla positivamente con l'ortodossia e con l'antisemitismo. Analogamente si potrebbe scoprire che l'atteggiamento libertario (AL) correla negativamente sia con l'ortodossia sia con l'antisemitismo; ed infine trovare che il livello educativo (LE), mentre correla positivamente con l'atteggiamento libertario, correla negativamente sia con il dogmatismo che con l'ortodossia religiosa.

Il modello presenta tre diverse regressioni lineare multiple e una regressione semplice che si potrebbero verificare singolarmente. Le loro equazioni sono:

$AS = b_0 + b_1 OR + b_2 DG + b_3 AL$
$AL = b_0 + b_1 OR + b_2 LE$
$OR = b_0 + b_1 LE + b_2 DG$
$DG = b_0 + b_1 LE$

In queste equazioni, i b_i sono i parametri della *path analysis.*

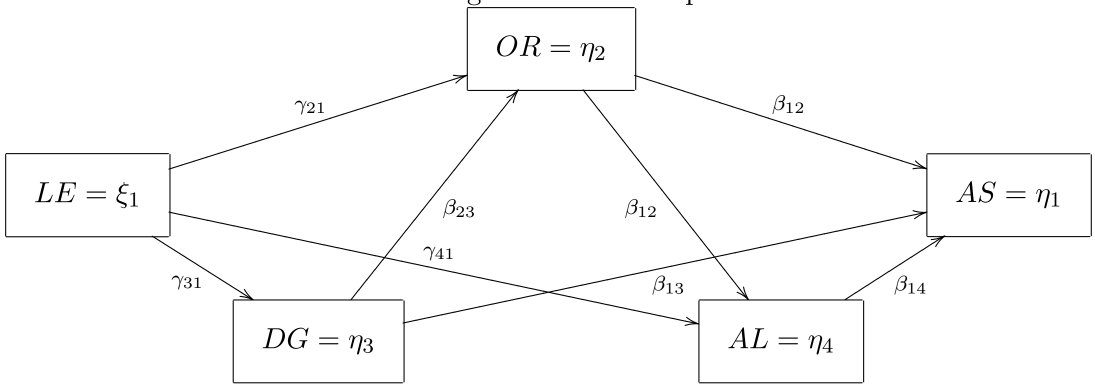
Per tradurre questo modello in LISREL, dobbiamo identificare la natura delle variabili. Innanzitutto abbiamo solo variabili osservate; quattro delle cinque variabili ricevono frecce (AS, AL, OR, DG) e quindi, in quel momento, fungono da variabili dipendenti; tre di loro (AL, OR e DG) servono per spiegare, in alcuni modelli (ovvero equazioni), una delle variabili dipendenti e quindi fungono anche da variabili indipendenti. Infine solo LE serve per spiegare altre variabili ma non funge mai da dipendente. Dobbiamo quindi concludere che mentre LE è una variabile esogena che coinciderà con un KSI, tutte le altre sono endogene e coincideranno con delle ETA.

Lisrel indica con *beta* (β) i legami fra le KSI (ξ) e le ETA (η) e con *gamma* (γ) quelli fra le ETA. Se applichiamo le regole della modellazione causale, dovremmo ottenere il grafico di Fig.14, dove è stato indicato anche il ruolo delle variabili (LE=ξ_1) in base all'ordine in cui le inseriremo (comando `RA`). Per interpretare il grafico, ricordiamo che gli indici di ogni legame fanno riferimento rispettivamente alla variabile a cui punta la freccia e alla variabile che la invia. Quindi γ_{41} indica una freccia che origina dalla prima Ksi e termina sulla quarta Eta.

Sulla base di questo grafico possiamo scrivere il programma. Riassumiamo: 5 variabili osservate, 10 misurazioni, dati grezzi.

```
DA ni=5 no=10 ma=cm
ra
3 2 1 8 9   3 1 2 8 6  2 5 3 6 4   6 2 3 5 7
4 3 4 7 6   5 6 5 4 4  6 5 6 4 5   4 8 7 6 5
9 7 8 2 1   8 3 9 5 6
```

Figura 14: Grafico per Lisrel




```
LA
AS1 OR2 DG3 AL4 LE1
```

Nel comando LA, dopo il nome simbolico di ogni variabile abbiamo inserito anche il numero sequenziale con cui Lisrel le identifica; questo ci aiuta a scrivere le istruzioni.

Nel comando MO dobbiamo indicare una variabile X e 4 variabili Y; essendo tutte osservate, Lisrel le farà coincidere automaticamente con le KSI e le ETA. Per semplicità, dichiariamo le matrici BE e GA come FU,FI (`full` e `fixed` ovvero piena e bloccate); questo significa che i legami 12 e 21 sono diversi fra di loro (`full`) e che, per ora, non vogliamo stimare nessun parametro (`fixed`). Non possiamo però eseguire il programma solo con l'istruzione MO, per cui dobbiamo liberare i parametri che vogliamo stimare (con l'istruzione FR).

```
mo nx=1 ny=4 BE=FU,FI GA=FU,FI
fr be 1 2 be 1 3 be 1 4 be 2 3 be 4 2
fr ga 2 1 ga 3 1 ga 4 1
pd
ou
```

Avremmo potuto agire in modo diverso.

Avremmo potuto dichiarare le matrici FU,FR (piene, libere cioè con tutti i parametri da stimare) e poi usare l'istruzione FI per bloccare i parametri che non interessano. Cioè:

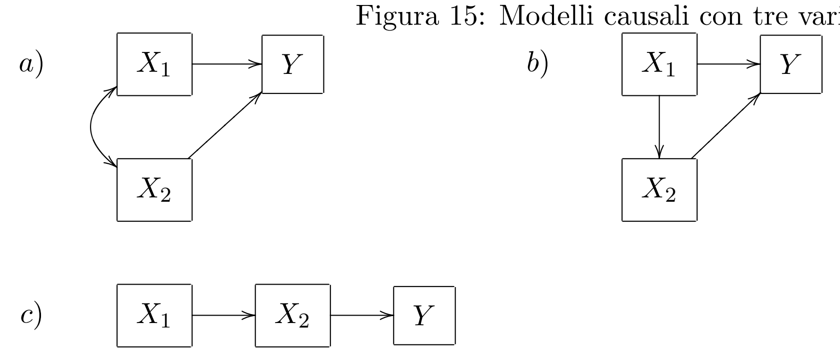
```
mo nx=1 ny=4 BE=FU,FR GA=FU,FR
fi be(1,1) be(2,1) be 2 2 be 2 4 be 3 1 be 3 2 be 3 3 be 3 4 be 4 1
fi be 4 3 be 4 4 ga 1 1
```

Come si può notare, questa soluzione è "parsimoniosa" (cioè usa meno istruzioni) solo per la matrice GA mentre non lo è affatto per la matrice BE. Per capire meglio il problema, costruiamo le due matrici.

$$\text{BE} = B = \begin{bmatrix} 0 & \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & 0 & \beta_{23} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_{42} & 0 & 0 \end{bmatrix} \qquad \text{GA} = \Gamma = \begin{bmatrix} 0 \\ \gamma_{21} \\ \gamma_{31} \\ \gamma_{41} \end{bmatrix}$$

$$NE \times NE = 4 \times 4 \qquad\qquad NE \times NK = 4 \times 1$$

Figura 15: Modelli causali con tre variabili



Possiamo osservare che la matrice BE implica 16 elementi associati a 16 possibili parametri e noi vogliamo stimarne solo 5, mentre la matrice GA implica 4 elementi e ne stimiamo 3. Quindi la soluzione ottimizzata sarà:

```
mo nx=1 ny=4 BE=FU,FI GA=FU,FR
fr be 1 2 be 1 3 be 1 4 be 2 3 be 4 2
fi ga(1,1)
```

Un'altra possibilità è quella di usare il comando `PA`.

```
PA BE
0 1 1 1
0 0 1 0
0 0 0 0
0 1 0 0
PA GA
0
1
1
1
```

In questa istruzione, lo 0 indica che non vogliamo stimare il parametro e 1 che lo vogliamo stimare.

6.5 Variazioni sul tema

Se prendiamo in considerazione 2 variabili predittrici e una variabile predetta, il modello relazionale relativo ad una regressione lineare multipla è quello rappresentato dalla Fig. 15a, ma con le stesse variabili sono possibili due altri modelli che non coincidono con una regressione multipla.

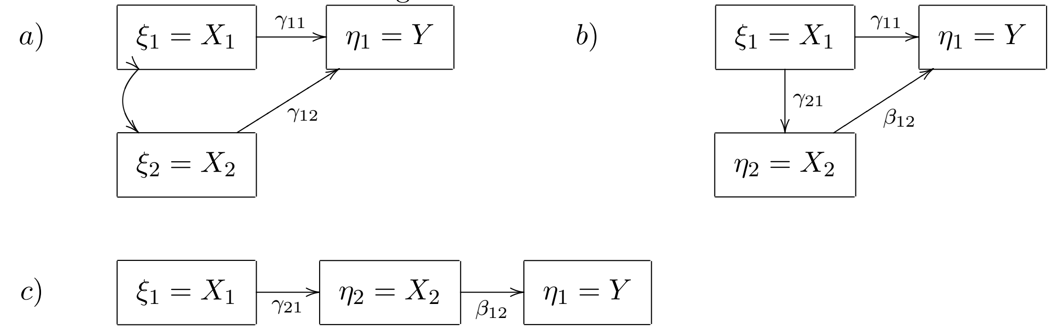
Nel caso della Fig 15b la correlazione fra le due variabili predittrici viene ipotizzata come una relazione causale. X_2, nella sua influenza su Y non agisce indipendentemente da X_1 e, d'altra parte, X_1 agisce su Y sia in modo diretto sia in modo indiretto attraverso la sua influenza su X_2.

Nella situazione rappresentata in Fig 15c, invece, X_1 non agisce direttamente su Y ma solo tramite X_2.

Trasformiamo i tre grafici in notazione Lisrel (Fig.16) e confrontiamoli fra loro:

La differenza sostanziale fra a e b e c sta nel fatto che la regressione lineare multipla implica che tutte le variabili predittive siano esogene (mandano frecce ma non ne ricevono) mentre i modelli di

Figura 16: Modelli lisrel con tre variabili



equazione strutturale permettono che una variabile predittiva sia endogena (cioè contemporaneamente mandi e riceva frecce). Le matrici dei legami implicati sono:

$$a) \quad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} \end{bmatrix}$$

$$b) \quad \Gamma = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \qquad B = \begin{bmatrix} 0 & \beta_{12} \\ 0 & 0 \end{bmatrix}$$

$$c) \quad \Gamma = \begin{bmatrix} 0 \\ \gamma_{21} \end{bmatrix} \qquad B = \begin{bmatrix} 0 & \beta_{12} \\ 0 & 0 \end{bmatrix}$$

Presumendo che le variabili siano chiamate `Y`, `X1` e `X2`, i tre programmi Lisrel diventano:

a)
```
SE Y X1 X2 /
MO NY=1 NX=2 GA=FU,FR
```
b)
```
SE Y X2 X1 /
MO NY=2 NX=1 GA=FU,FR BE=FU,FI
FR BE(1,2)
```
c)
```
SE Y X2 X1 /
MO NY=2 NX=1 GA=FU,FR BE=FU,FI
FR BE(1,2)
FI GA(2,1)
```

6.6 Variabili latenti

Per introdurre le variabili latenti, ripartiamo dalla regressione lineare semplice. A suo tempo avevamo visto che Lisrel fa coincidere automaticamente la variabile X con una ξ e la variabile Y con una η. Che cosa succede in questa coincidenza? E' come se Lisrel "inventasse" due latenti i cui parametri con le osservate sono uguali a 1. Osserviamo il grafico di Fig. 6.6b) rispetto alla a).

Se costruiamo le equazioni di regressioni avremmo (scarti dalla media):

$$\begin{array}{llll} a) & Y = & b_1 X & +e \\ b) & X = & 1\xi & +0 = \xi \\ & Y = & 1\eta & +0 = \eta \\ & \eta = & b_1 \xi & +e \end{array}$$

e dal momento che $X = \xi$ e che $Y = \eta$, l'ultima equazione di regressione risulta uguale alla prima, cioè $Y = b_1 X + e$.

Usando i dati di 2 il programma in Lisrel sarebbe:

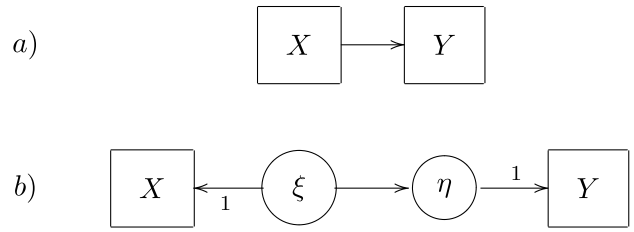


Figura 17: Regressioni semplice con latenti esplicitate

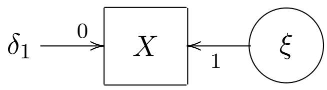
```
Esempio di regressione lineare semplice con latenti

da ni=2 no=5 ma=cm
RA
3 2 2 3 4 5 5 7 8 8
la
Y X
mo nx=1 ny=1 nk=1 ne=1
pd
OU nd=3
```

Se lo eseguite, otterrete gli stessi parametri.

7 I "modelli" Lisrel

Osservando la figura 6.6, possiamo vedere rappresentati (in modo sintetico) i tre modelli di Lisrel: i due modelli di misura e il modello strutturale.

Le variabili ksi ed eta sono variabili latenti e quindi non sono state misurate, le dobbiamo perciò stimare sulla base delle osservate che loro influenzano. Ecco quindi i due modelli di misura delle latenti. Il modello di misura delle ksi è costituito dalla latente ksi, dal suo legame con l'osservata di tipo x (chiamato *lambda x*), dall'osservata stessa e dall'errore delta; corrisponde a questo pezzo di modello:

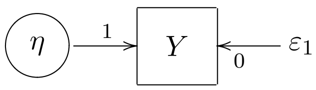


Che equivale all'equazione

$$\mathbf{X} = \Lambda_x \xi + \delta$$

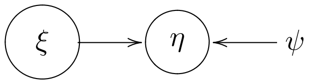
La eta, il suo legame con l'osservata di tipo y (chiamato *lambda y*) e l'osservata e l'errore epsilon, costituiscono il modello di misura delle eta:

$$\mathbf{Y} = \Lambda_y \eta + \varepsilon$$



Se invece consideriamo le latenti e i loro legami, abbiamo il Modello strutturale:

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta$$



In realtà, non tutti i tre modelli possono sussistere da soli. Il modello di misura delle ksi ha una sua indipendenza dagli altri perché corrisponde all'analisi fattoriale. Il modello strutturale può funzionare da solo se ogni latente è associata ad una e una sola osservata; si tratta di tutti gli esempi fatti finora e generalmente chiamati modelli persole variabili osservata. La coincidenza di una osservata con una latente (espressa tramite i lambda posti a 1 e gli errori a 0) permette di trattare questi disegni relazionali come fossero modelli strutturali. Al contrario il modello di misura delle eta non ha una ragione teorica per esistere da solo, in quanto dovrebbe essere trattato come una misura delle ksi.

7.1 Analisi fattoriale

Finora abbiamo utilizzato solo modelli che usavano variabili osservate e in particolare questi modelli assumevano che, ogni variabile osservata, coincidesse con una variabile latente. Esiste la possibilità che due o più variabili osservate siano legate ad una stessa variabile latente (Fig.18).

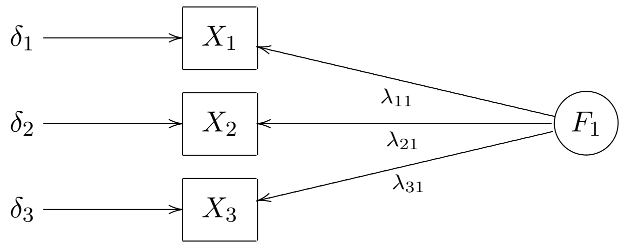


Figura 18: Modello fattoriale semplice

L'equazione matriciale che corrisponde a questo tipo di modello è:

$$\mathbf{X} = \Lambda_x \xi + \delta$$

Possiamo considerarla in forma più esplicita per ogni X:

$$\begin{aligned} X_1 &= \lambda^x_{11} F_1 + \delta_1 \\ X_2 &= \lambda^x_{21} F_1 + \delta_2 \\ X_3 &= \lambda^x_{31} F_1 + \delta_3 \end{aligned}$$

Ricordiamo che ogni variabile che riceve una freccia, cioè un legame, è oggetto di un'equazione di regressione); che Λ è la forma maiuscola di λ; che F_1 è il nome che abbiamo usato nel grafico al posto di ξ.

7.1.1 Esempio di AF con 1 fattore, programma, grafico e commento

[Corrisponde ad una delle esercitazione; versione provvisoria]

La seguente matrice di correlazioni riguarda 6 item del test di Brown per ADD (attention deficit disorder), 206 rispondenti adulti non patologici. Le variabili sono: *smemorato* - spesso altri dicono che sono smemorato; *dimentica* - spesso dimentica cose; *cancella* - molte cancellazioni in scittura; *omette* - cose dette in scrittura; *nontr* - non trattiene dettagli; *nonme* - non memorizza bene.

I dati provengono da una tesi di laurea: tesi E. Realini psicologia Bicocca 2003.

Secondo Brown i dati formano un singolo fattore.

```
DA ni=6 no=206 ma=km
LA
smemorato dimentica  cancella omette  nontr nonme
KM
1
.263     1
.104    .159      1
.123    .240     .364     1
.150    .410     .209    .213     1
.381    .138     .117    .207    .108     1
mo nx=6 nk=1 lx=fu,fr td=di,fr ph=sy,fr
lk; DEFICIT
pd; ou mi
```

Commento: Abbiamo 6 variabili e le usiamo tutte, quindi non serve usare SE e impostiamo NX=6. Essendo un'analisi fattoriale con 1 solo fattore, NK=1 e LX=FU,FR perché tutte le X devono avere un lambda che li collega all'unico fattore. Le covarianze degli errori (TD) sono posti DI,FR perché vanno stimati solo quelli di ogni singolo item (ovvero le varianze), mentre accettiamo l'assunto fondamentale che gli errori non correlino (e quindi non co-varino). La varianza della latente ksi è posta SY,FR perché dev'essere stimata. Lisrel la porrà automaticamente a 1 perché stiamo: 1) lavorando con una matrice di correlazione e abbiamo chiesto di analizzare una matrice di correlazione; 2) perché stiamo facendo un'analisi fattoriale e Lisrel automaticamente ipotizza che i fattori siano standardizzati (ricordate: storicamente l'AF si faceva a partire da una matrice di correlazione). Il comando LK serve solo a dare un nome alla latente. PD permette di vedere il grafico. Se fate eseguire il programmino sopra, notate cje la latente DEFICIT (ovale) manda frecce (lambda x) ad ogni osservata (quadrati) e ad ogni osservata è associato la varianza degli errori.

Commento avanzato: Se guardiamo il χ^2, vediamo che non è buono (è alto e significativo). Se poi consideriamo gli indici di modifica (o guardando nel file di output sotto `Modification Indices` oppure selezionando `Modification Indices` dal pannello `Estimates:` del grafico), vediamo che ci sono degli errori che correlano (nel grafico sono le frecce curve in grigio chiaro; nell'output la matrice `theta-delta`). Questo indica che una parte della covarianza degli errori (che dovrebbe essere casuale e non correlare) è invece spiegabile come influenza comune. Si tratta allora di aggiungere un'altro fattore che colleghi le variabili correlati. In questo caso specifico, dovremmo spostare le due X con l'indice di modifica più elevato su un secondo fattore: si tratta di `smemorat` e di `nonme`. Successivamente vedremo che ci saranno ancora errori correlati. Il modello finale prevede che la prima e l'ultima variabile variabile dipendano da un fattore, la seconda e la quinta da un secondo fattore e, infine, che la tera e la quarta formino un terzo fattore.

7.1.2 Esempio di AF con 2 fattori correlati, programma, grafico e commento

[Corrisponde ad una delle esercitazione; versione provvisoria]

LISREL7 manuale p 84 e seguenti

proveniente da Calsyn e Kenny (1977) 'Self concept of ability and perceived evaluation of others: cause or effect of academic achievement?'. Journal of educational psychology, 69, 139-145.

Le variabili sono:

x1	autocap	self concept of ability (autovalutazione di capacita)
x2	pgenval	perceived parental evaluation (percezione valutazione genitori)
x3	pinsval	perceived teacher evaluation (percezione valutazione insegnante)
x4	pamival	perceived friends evaluation (percezione valutazione amico/a)
x5	educasp	educational aspiration (apsirazioni educative)
x6	colplan	college plans (piano per college = educazione terziario)

556 studenti (bianchi) stati uniti

Si prevedono due fattori, uno sottostante le variabili x1, x2, x3, x4 (lo chiamiamo ABILITA, sia percepito da se stesso, sia percepito da altri) e un secondo fattore sottostante x5 e x6 (lo chiamiamo ASPIRAZIONE)

```
DA ni=6 no=556 ma=km
KM
1.00                         ! autocap
 .73 1.0                     ! pgenval
 .70  .68 1.0                ! pinsval
 .58  .61  .57 1.0           ! pamival
 .46  .43  .40  .37 1.0      ! educasp
 .56  .52  .48  .41  .72 1.0 ! colplan
LA; autocap pgenval pinsval pamival educasp colplan
MO nx=6 nk=2 ph=sy,fr lx=fu,fi td=di,fr
LK; ABILITA ASPIRAZIONE
pa lx
1 0
1 0
1 0
1 0
0 1
0 1
pd; ou mi
```

Commento: Avendo 2 fattori (latenti) NK va posto a 2 e, se si usa LK per assegnare etichette alle ksi, si danno nomi ad entrambe. Per associare le X alle latenti usiamo il comando PA impaginato in modo da rendere facile capire le associazioni (1 indica che quella X è associata a quel fattore; o che non lo è). La matrice PH delle correlazioni (stiamo facendo un'analisi fattoriale) la lasciamo completamente libera e in questo modo otterremo sia le varianze delle due latenti sia la covarianza (cioè la correlazione) fra le due. Quando non abbiamo motivi teorici per imporre che i fattori non correlino è sempre preferibile ipotizzare o lavorare su un modello correlati. Eventualmente, guardando il risultato potremo poi decidere di passare ad una soluzione ortogonale o non correlata.

Commento avanzato: Osservando il χ^2 vediamo che il modello è buono (valore basso e non significativo). Vediamo anche che i due fattori correlano a .67, una correlazione troppo alta per essere ignorata; inoltre, se guardiamo i `T-values` vediamo che è molto significativa e quindi non va azzerata. In questo modello, non è possibile una soluzione ortogonale.

8 Indici di adattamento

$$AIC = \chi^2 - 2gl \text{ (Joreskog)}$$
$$AIC = \chi^2 + 2t \text{ (Tanaka)}$$
$$CAIC = \chi^2 - (1 + lnN)gl \text{ (Bozdogan)}$$

9 Notazione Lisrel

		indicate da	matrice	ordine	default
Variabili	latenti endogene	η (eta)			
	latenti esogene	ξ (ksi)			
	osservate endogene	Y			
	osservate esogene	X			
Errori	delle η	ζ (zeta)			
	delle Y	ϵ (epsilon)			
	delle X	δ (delta)			
Parametri	tra η e Y	λ^y (lambda y)	Λ_y, LY	NYxNE	FU,FI
	tra ξ e X	λ^x (lambda x)	Λ_x, LX	NXxNK	FU,FI
	tra le η	β (beta)	B, BE	NExNE	ZE,FI
	tra ξ e η	γ (gamma)	Γ, GA	NExNK	FU,FR
		α (alpha)	A, NExNE		
Varianze/ covarianze	fra le ξ	ϕ (phi)	Φ, PH	NKxNK	SY,FR
	fra le ζ	ψ (psi)	Ψ, PS	NExNE	SY,FR
	fra le ε	θ^{ε} (theta-epsilon)	Θ_{ε}, TE	NYxNY	DI,FR
	fra le δ	θ^{δ} (theta-delta)	Θ_{δ}, TD	NXxNX	DI,FR

[mettere un grafico con tutte le variabili, gli errori e le frecce possibili]

10 Come disegnare un grafico di modello relazionale

Dare per scontato che il grafico andrà disegnato più volte fino a raggiungere un aspetto che renda chiare le relazioni implicate.

Ricordare che:

- le variabili osservate si indicano con un quadrato o con un rettangolo;
- le variabili latenti si indicano con un ellisse o un cerchio;
- le relazioni causa-effetto (A influenza B, B è influenzata da A; C spiega D, D è spiegata da C) si indicano con una freccia che parte dalla variabile-causa e termina sulla variabile-effetto;
- quando una variabile latente è misurata da una o più osservate, si presume che la latente *spieghi* le osservate;
- se la latente è spiegata da una sola osservata (si dice "misurata senza errore"), il loro λ sarà uguale a uno e il relativo errore (δ o ε) sarà uguale a zero;
- il primo indice di una relazione causa-effetto, indica la variabile effetto (che riceve la freccia) e il seconda la variabile causa (che invia la freccia);
- le correlazioni o le covarianze fra due variabili si indicano con una linea (generalmente curva) che termina con una freccia ad entrambe le estremità.

Predisporre un foglio su cui disegnare il grafico, dividendolo idealmente in due parti (sinistra e destra) e dividendo ulteriormente ciascuna parte in due parti.

Potete cominciare a disegnare ogni relazione ipotizzata, come se fosse un grafico a se stante, in modo da vedere quali variabili esistono. Per ogni variabile che ne influenza (o che spiega) un'altra, ci sarà una freccia che parte dalla prima e termina sull'altra; per ogni variabile spiegata (o influenzata), ci sarà almeno una freccia che arriva.

A) Identificare le variabili osservate;

B) identificare le variabili latenti.

Se ci sono solo variabili osservate:

a) identificare le variabili che svolgono sempre e solo il ruolo di variabili indipendenti, cioè quelle che servono a spiegare altre variabili osservate ma non sono spiegate da nulla; sono tutte variabili osservate esogene (le X) e andranno a coincidere con le KSI;

b) porre queste variabili nell'area sinistra del grafico racchiuse fra quadrati, nominandole come $X_i = \xi_i$;

c) le variabili osservate restanti dovrebbero essere tutte variabili che almeno una volta svolgono la funzione di variabile dipendente ovvero dovrebbero essere variabili che subiscono l'influenza di quelle già indicate nel grafico o di quelle considerate adesso (se così non fosse, avete sbagliato qualcosa); sono variabili endogene (le Y) e andranno a coincidere con le ETA;

d) porre queste variabili nell'area destra del grafico racchiuse fra quadrati e nominandole come $Y_i = \eta_i$;

e) identificare le relazioni esistente fra le $Y = \eta$ (sono i legami beta, β, BE), scrivendo accanto ad ogni β il numero corrispondente all'η di arrivo e quindi quello dell'η di partenza;

f) identificare le relazioni fra le $X = \xi$ e le $Y = \eta$ (sono i legami gamma, γ, GA), scrivendo accanto ad ogni γ il numero corrispondente all'η di arrivo e quindi quello della ξ di partenza;

g) per ogni η aggiungete l'errore ζ;

h) se le X hanno covarianze da tenere in considerazione, tracciate una linea curva e indicate a fianco l'appropriato simbolo ϕ.

Se ci sono sia variabili osservate che variabili latenti:

1) selezionare le variabili osservate;

2) dividere le variabili latenti in esogene (mandano frecce ma non ne ricevono) ed endogene (ricevono frecce e possono anche mandarne);

3) disegnare (su una *brutta*) le relazioni fra le osservate e le latenti, ricordando che se una latente è misurata da una o più osservate, la latente manda delle frecce alle osservate;

4) isolare le osservate che servono solo a misurare una variabile latente e che non vengono spiegate né da altre variabili osservate né da altre variabili latenti;

5) suddividere queste osservate in base alla latente; se la latente è esogene (ovvero manda frecce ma non ne riceve nessuna), porre queste variabili sull'estrema sinistra del grafico racchiuse in quadrati (o rettangoli). Sono le X.

6) se la latente è endogena (ossia riceve anche solo una freccia), porre queste osservate alla destra del grafico racchiuse in quadrati (o rettangoli), perché sono variabili Y;

7) porre nell'area centro-sinistra del grafico, le variabili latenti che sono misurate dalle X e che costituiscono i fattori. Sono le ksi (ξ), variabili latenti esogene;

8) tracciare le frecce fra le X e le ksi (sono i λ_x);

9) porre nell'area centro-destra del grafico, le variabili latenti che sono misurate dalle Y e che costituiscono le η (eta);

10) tracciare le frecce fra le Y e le eta (sono i λ_y);

11) se avanzano delle variabili osservate, devono coincidere con delle variabili latenti (eta o ksi) e vanno poste nell'area centro-sinistra o centro-destra del grafico racchiuse fra ellissi con le diciture $X_i = \xi_i$ oppure $Y_i = \eta_i$ (avranno dei λ posti a 1 ed errore uguale a 0;

12) disporre nell'area centrale le altre variabili latenti;

13) identificare le relazioni esistente fra le eta (sono i legami beta, β, BE);

14) identificare le relazioni fra le ksi e le eta (sono i legami gamma, γ, GA);

15) se le latenti esogene (le ksi) sono tra loro correlate, tracciare una doppia freccia curva e segnare la ϕ corrispondente;

16) segnare gli errori ζ delle η;

17) per ogni X, segnare il corrispondente errore δ, che andranno a formare la matrice theta-delta (θ_δ, TD);

18) per ogni Y, segnare il corrispondente errore ε, che andranno a formare la matrice theta-epsilon (θ_ϵ, TE).

10.1 Esercizi

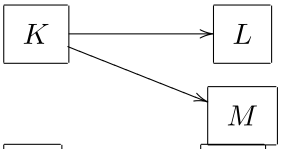
1. Disponiamo di 5 variabili osservate (K, L, M, N, e O) su un campione di N=200. K ha influenza su L e su M; L influenza N; O è influenzata da M; e N è influenzata da L e M.

2. In una ricerca con 78 genitori che hanno avuto da poco il loro primo figlio, è stato usato il questionario SPPQ (Social Project Perspective Questionnaire) di 22 items. In una successiva analisi, i primi 3 item vengono usati per misurare la variabile latente *Socialità*, i successivi 3 item per la variabile latente *Didattica* ed infine altri 3 per la latente *Controllo*. Socialità e Controllo correlano fra loro, mentre Didattica non correla con nessuno degli altri.

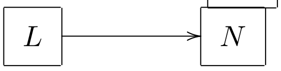
10.2 Soluzioni

1. Cominciamo col notare che il modello utilizza solo e soltanto variabili osservate. Disegniamo allora le relazioni fra le variabili così come ci vengono indicate:

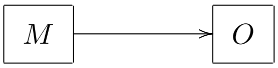
K ha influenza su L e su M:



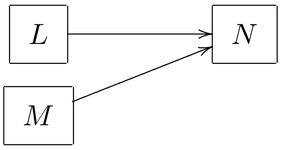
L influenza N:



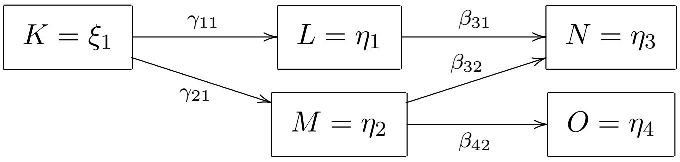
O è influenzata da M:



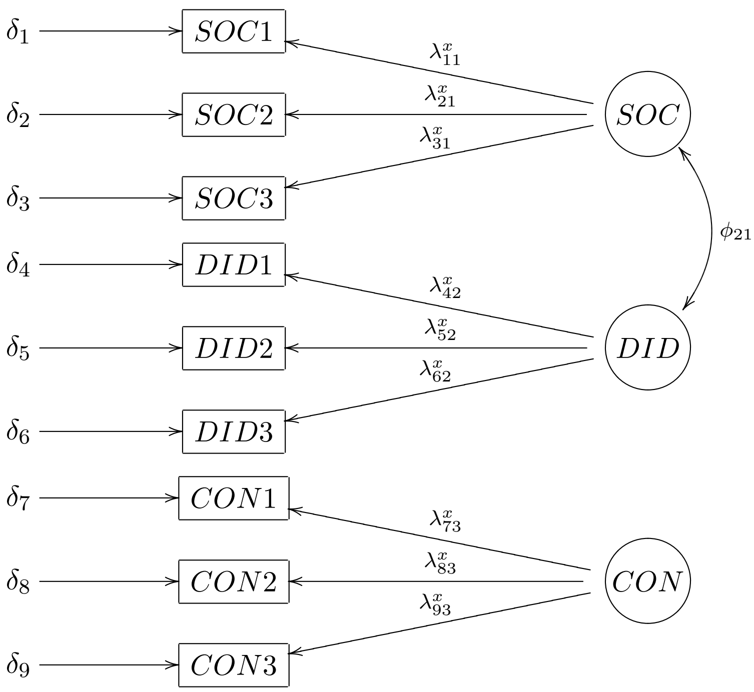
N è influenzata da L e M:



Quindi L, M, N e O ricevono almeno una freccia e saranno delle Eta, mentre K è l'unica che manda frecce ma non ne riceve mai, quindi sarà una Ksi. Mettendo tutto assieme e indicando anche i parametri in notazione Lisrel:



2. In questo caso abbiamo 9 variabili osservate e 3 variabili latenti. Ogni blocco di 3 osservate, serve per misurare una latente.



Riferimenti bibliografici

Hoge, D., & Carroll, J. (1975). Christian beliefs, nonreligious factors and semi-semitism. *Social Forces*, *53*(4), 581-594.

Rossi, G. (2002). *Appunti di algebra matriciale.* Available from http://web.newsguy.com/germano/mi/algebra.pdf ([Online http://web.newsguy.com/germano/mi/algebra.pdf])

Tacq, J. (1997). *Multivariate analysis technique in social science research. From problem to analysis.* London: Sage Publications.