QUADERNI DI RICERCA
Dipartimento di Scienze Sociali

12

A14
400/12

Roberto Albano
Daniela Molino

Analisi Fattoriale per le scienze sociali

Dipartimento di Scienze Sociali
Università degli Studi di Torino

www.dss.unito.it
dss@unito.it

via Sant'Ottavio 50
10124 Torino
011 6702606

ISBN 978–88–548–4399–8

I edizione: dicembre 2011

Indice

Introduzione

Tra le numerose tecniche di analisi multivariata dei dati, una delle più anziane è l'Analisi Fattoriale (d'ora in poi nel testo *AF*). La sua 'doppia' invenzione si può collocare all'inizio del Novecento:

— in ambito statistico, un punto di riferimento importante per l'*AF*, e per altre tecniche di analisi multivariata, è un articolo del 1901 di Karl Pearson, in cui si fa uso di strumenti di analisi matematica a quel tempo già consolidati, come la distribuzione normale multivariata di Bravais e la teoria degli autovalori e autovettori delle trasformazioni lineari;
— parallelamente, tra il 1904 e il decennio successivo, vengono gettate le sue fondamenta in ambito psicometrico con l'opera di Charles Spearman e di alcuni suoi collaboratori per misurare l'intelligenza negli esseri umani.

Si deve proprio a Spearman la prima concettualizzazione della caratteristica essenziale della tecnica: evidenziare quegli elementi comuni a più indicatori. Questa tecnica ha avuto successo in diversi campi del sapere scientifico; certamente in notevole misura nelle scienze sociali.

Le ragioni di questo successo risiedono nel fatto che l'*AF* permette di misurare proprietà che non hanno una definizione semplice e netta sul piano teorico e, conseguentemente, non sono rilevabili sul piano empirico mediante una singola operazione di misurazione come avviene per esempio per l'altezza fisica di una persona, la sua età, il suo stato civile. Le discipline sociali e psicologiche, come è noto, fanno ampio uso di concetti che non sono direttamente osservabili: si pensi a concetti complessi e senza un correlato empirico diretto e univoco come l'autoritarismo, il clima organizzativo, il *coping*, la secolarizzazione, il capitale umano, il capitale sociale, la partecipazione politica e

così via. Proprietà come queste possono essere rilevate soltanto mediante molteplici **indicatori**, sulla cui validità gli studiosi si trovino sostanzialmente d'accordo.

Il **rapporto di indicazione** (Marradi, 1980:40), che lega gli indicatori empirici al concetto teorico sottostante, è di natura semantica prima che di tipo matematico–statistico.

I singoli indicatori occupano uno spazio semantico solo parzialmente sovrapposto a quello del concetto teorico; l'insieme degli indicatori individuati dovrebbe ricoprire, se non l'intero suo spazio semantico, almeno una considerevole porzione[1]. Si può dire che l'insieme di indicatori di un concetto è aperto, potendosene formulare in linea di principio sempre di nuovi. Ciò determina tra l'altro che si possano formulare indicatori caratterizzati da **intercambiabilità**: seppur non totalmente coincidenti per lo spazio semantico comune, due indicatori possono essere in pratica considerati fungibili.

L'uso dell'*AF* è in genere mirato a ricondurre un insieme di variabili a una o più dimensioni comuni; talvolta, meno frequentemente, può anche essere usata per ricondurre più dimensioni latenti a una o più meta–proprietà (***AF* di secondo livello**). A titolo di esempio: un sottogruppo di item di una batteria può essere ricondotto alla dimensione 'superstizione e stereotipia', un altro gruppo alla dimensione 'distruttività e cinismo', un altro ancora al 'convenzionalismo'; queste tre dimensioni possono essere poi ricondotte a un concetto più generale quale è quello di 'autoritarismo', nel senso previsto dalla celebre 'teoria della personalità autoritaria' di Adorno–Horkheimer[2].

L'impiego di indicatori molto simili nel contenuto, al limite variabili solo nella forma linguistica, produce invece fattori di scarso interesse sostantivo; tuttavia, l'*AF* condotta su variabili molto simili può avere finalità di tipo metodologico: ad esempio per valutare in che misura

1. Cfr. de Lillo et al., 2011:31–5. Qui come in molte altre parti del testo facciamo riferimento a temi basilari della metodologia della ricerca, trattati in molti buoni manuali; tra questi scegliamo — salvo esigenze specifiche — di fare riferimento a uno dei più recenti in lingua italiana.

2. Che l'*AF* applicata alla scala F di Adorno dia o meno conferma dell'unidimensionalità sostenuta dagli studiosi della scuola di Francoforte è questione che qui può essere trascurata; per un approfondimento si veda ad es. Madge, 1962, trad. it. 1966, cap. X.

diverse formulazioni linguistiche siano intercambiabili, oppure per valutare preliminarmente la coerenza interna di scale di cui si vuole misurare l'**attendibilità**.

L'*AF* parte dagli indicatori (i *significans* del termine non osservativo) e dalle loro interrelazioni, per individuare, mediante opportune operazioni matematiche, le dimensioni ad essi sottostanti. Ciò non significa che con tale tecnica identifichiamo sempre a posteriori i fattori latenti. La loro individuazione *ex post* caratterizza uno stile di ricerca che definiamo 'esplorativo', contrapposto a uno stile 'confermativo', in cui il ricercatore deduce a priori, sulla base della riflessione teorica, la struttura dei legami tra le componenti del modello. In questo saggio ci concentreremo sull'*AF* Esplorativa, dedicando a quella Confermativa soltanto brevi cenni nel capitolo 6.

Sull'*AF* Esplorativa sono state sollevate importanti critiche circa la sua fondatezza sul versante matematico–statistico e sulla validità scientifica dei risultati che è in grado di produrre. Non sarà qui ripercorso il dibattito sugli aspetti epistemologici e ontologici attinenti ai fattori latenti, se essi cioè siano da intendersi come oggetti dotati di autonoma realtà piuttosto che mere astrazioni di comodo del ricercatore. Scorrendo, anche in modo non sistematico, gli articoli di ricerca in cui si fa uso dell'*AF*, è facile prendere atto che le variabili latenti sono considerate strumenti utili per la ricerca sociale da ricercatori che pure partono da concezioni di realtà sociale assai diverse, ora più orientate al 'realismo' ora più orientate al 'costruttivismo'. Sarà invece presentato un quadro generale degli aspetti più tecnici, vale a dire dei parametri da controllare e modificare nel modo più opportuno per individuare un numero di dimensioni latenti di molto inferiore a quello di un insieme di **variabili manifeste**[3], capaci di rendere conto delle relazioni intercorrenti tra queste ultime. Un'altra limitazione di campo consiste nel considerare esclusivamente il modello lineare nelle variabili e nei parametri. Per semplicità indicheremo d'ora in avanti l'*AF* come oggetto della presente trattazione sottintendendo, salvo indicazione contraria, 'Lineare' ed 'Esplorativa'.

3. Le variabili manifeste sono il risultato dell'applicazione di una **definizione operativa** a ciascuno degli indicatori; (cfr. de Lillo et al., 2011:29).

Capitolo I

Presentazione informale della tecnica

1.1. L'Analisi Fattoriale unidimensionale

Obiettivo dell'*AF* è quello di interpretare le covariazioni tra un numero elevato di variabili osservate empiricamente, le variabili manifeste, *come se* fossero dovute all'effetto di variabili non direttamente osservabili definite **fattori latenti comuni**. Il caso più semplice, da cui conviene iniziare, è quello unidimensionale, cioè a un solo fattore latente. Una rappresentazione grafica e un esempio serviranno a chiarire quanto detto.

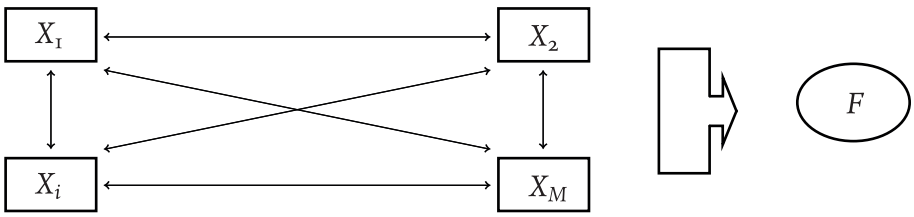

Figura 1.

Le variabili X_1, X_2, $X_i \ldots, X_M$, racchiuse in riquadri, sono quelle rilevate originariamente, ad esempio mediante la somministrazione di un questionario. La variabile F, racchiusa in un'ellisse, rappresenta invece il fattore latente comune alle variabili manifeste. Si parte dalle relazioni tra le coppie di variabili manifeste, rappresentate dalle frecce bidirezionali che collegano a coppie le variabili manifeste, per inferire l'esistenza di un fattore comune sottostante che renda conto, almeno per una parte rilevante, del comportamento di ognuna delle

variabili osservate, ma soprattutto che sia in grado di rendere conto, in massima parte, dell'interrelazione tra le variabili manifeste.

La misurazione del fattore latente comune non è dunque un'operazione empirica indipendente dalla misurazione delle variabili manifeste, anche se sul piano analitico si tratta di entità distinte.

Fatta questa distinzione, va ora richiesto al lettore un ulteriore sforzo concettuale: per comprendere il senso dell'*AF* occorre rappresentare il rapporto tra fattore latente comune e variabili manifeste come un rapporto di dipendenza in cui le ultime sono le variabili da spiegare.

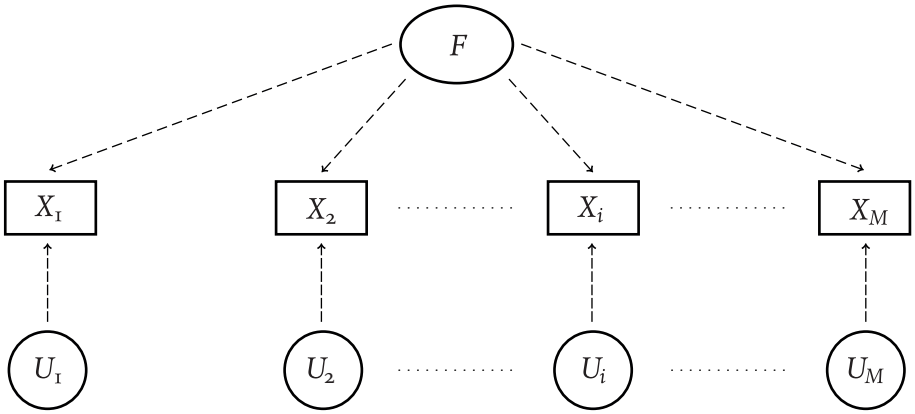

Figura 2.

Nella figura 2, oltre alle entità di cui abbiamo già detto, il fattore latente comune e le variabili manifeste, troviamo altre variabili inserite in cerchi: si tratta dei **fattori latenti unici**, variabili anch'esse non osservabili, ognuna delle quali influenza una sola variabile manifesta X_i.

Come dobbiamo interpretare le frecce che vanno dai fattori latenti alle variabili manifeste? In questo passaggio critico si colloca il senso dell'*AF*: interpretare le relazioni tra le variabili manifeste come covariazioni in assenza di causazioni dirette. Si tratta di un'applicazione del ben noto paradigma della **relazione spuria** (cfr. ad es. Albano, Testa 2002:140): l'associazione tra due variabili *X* e *Y* scompare quando si tie-

ne sotto controllo una terza variabile Z che agisce su X e Y. Rispetto al classico paradigma dell'analisi trivariata, va evidenziata un'importante differenza: le frecce della figura 2 non rappresentano una dipendenza genuina, perché le entità in gioco non sono semanticamente autonome; per questo motivo, al fine di marcare anche graficamente l'inseparabilità sul piano empirico di *explanans* e *explananda*, abbiamo tratteggiato le frecce.

Per riassumere: attraverso un input costituito da indici di associazione tra variabili (correlazioni lineari o altro) e particolari procedure matematiche e statistiche, stabiliamo con una procedura abduttiva (cioè induttiva e deduttiva) un rapporto di dipendenza *sui generis*, che interessa entità rilevate empiricamente (*explananda*) e entità solo ipotetiche (*explanans*). Al fattore latente comune (F) è attribuito il compito di interpretare le relazioni tra le variabili: questo è l'obiettivo prioritario di ogni *AF*. Ognuno dei legami tra fattore latente comune e variabili manifeste è rappresentato da un numero a_m, indicante il peso che il primo ha sulle seconde[1]. Ai Fattori latenti unici va il compito di interpretare la variabilità residua di ogni variabile manifesta. È utile a questo punto procedere con un esempio. Supponiamo di volere misurare nei paesi europei il livello di capitale umano (Becker 1964).

Il concetto di capitale umano può essere inteso come l'insieme di conoscenze, abilità, competenze e attributi individuali che facilitano la creazione di benessere personale sociale ed economico (OECD 2001). È chiaro che si tratta di un concetto multidimensionale e per sua natura composto da un mix eterogeneo di difficile misurazione diretta. L'uso delle credenziali educative può essere un indicatore per le abilità e le competenze degli individui, ma allo stesso tempo va messo in evidenza che queste possono essere acquisite anche al di fuori dei canali istituzionalizzati di formazione e istruzione. Conviene dunque, se possibile, ricorrere a un set di indicatori ecologici selezionati che ci permettano di individuare un fattore comune che esprima il concetto di capitale umano in modo articolato.

1. Detto con altri termini, questi coefficienti *a* rappresentano la **capacità predittiva** del fattore latente comune; successivamente, quando parleremo di calcolo dei punteggi fattoriali vedremo che si può anche costruire un peso che rappresenta un legame inverso, che va dalle variabili manifeste al fattore latente comune.

Possiamo ad esempio utilizzare alcuni indicatori rilevati attraverso le indagini internazionali sulle competenze degli studenti, in particolare quelli del *Programme for International Student Assessment* (Pisa) e del *Trend in International Mathematics and Science Study* (Timss), nonché attraverso indicatori strutturali:

— $V1$. competenze medie in matematica (Timss)
— $V2$. competenze medie linguistiche (Pisa)
— $V3$. competenze medie in scienze (Timss)
— $V4$. percentuale di diplomati nella fascia di età compresa tra i 18 e 29 anni
— $V5$. percentuale di laureati nella fascia di età compresa tra i 24 e 29 anni
— $V6$. percentuale di popolazione in età compresa tra i 24 e i 65 anni che partecipa a programmi di formazione continua (*Long life learning*)
— $V7$. percentuale di giovani che lasciano la scuola con un titolo di studio inferiore alla scuola secondaria di secondo grado nella fascia di età compresa tra i 18 e i 24 anni (*Early school leavers*)
— $V8$. spesa interna lorda per Ricerca e Sviluppo.

Queste variabili risultano legate tra loro, cioè al variare dell'una corrisponde una covariazione di un'altra, talvolta in senso positivo (per esempio $V5$ e $V6$), tal altra in senso negativo (per esempio $V1$ e $V7$). Già una semplice ispezione visiva della matrice di correlazioni **R** può portarci all'individuazione di una dimensione generale sottostante, con la quale riassumere le relazioni osservate.

Tuttavia, è facile intuire che quanto maggiore è il numero di variabili osservate, tanto più difficile risulterà considerare simultaneamente le correlazioni affidandoci a una mera esplorazione informale della matrice di correlazioni. Con sole 8 variabili, ad esempio, abbiamo 28 correlazioni diverse da prendere in considerazione. Più in generale, se le variabili sono M, le correlazioni non ridondanti tra tutte le possibili coppie di variabili sono:

$$\frac{M \cdot (M-1)}{2}$$

L'*AF* è utile proprio per semplificare l'osservazione di processi articolati e complessi (come può essere appunto il 'capitale umano') mediante la rilevazione di un certo numero di indicatori. Se il fattore individuato sia poi un'entità concretamente distinta dalle variabili manifeste, o se si tratti di un fattore solo analiticamente distinto ma operativamente definito da quelle, una mera astrazione matematica (un costrutto dotato di capacità euristica), è questione che dipende dallo sviluppo della teoria in quel campo e dalle scelte epistemologiche del ricercatore.

1.2. L'Analisi Fattoriale multidimensionale

Nella figura 2 abbiamo rappresentato una struttura latente unidimensionale[2]. Si può ora generalizzare il modello per studiare strutture multidimensionali; facciamo l'esempio di due fattori latenti (figura 3):

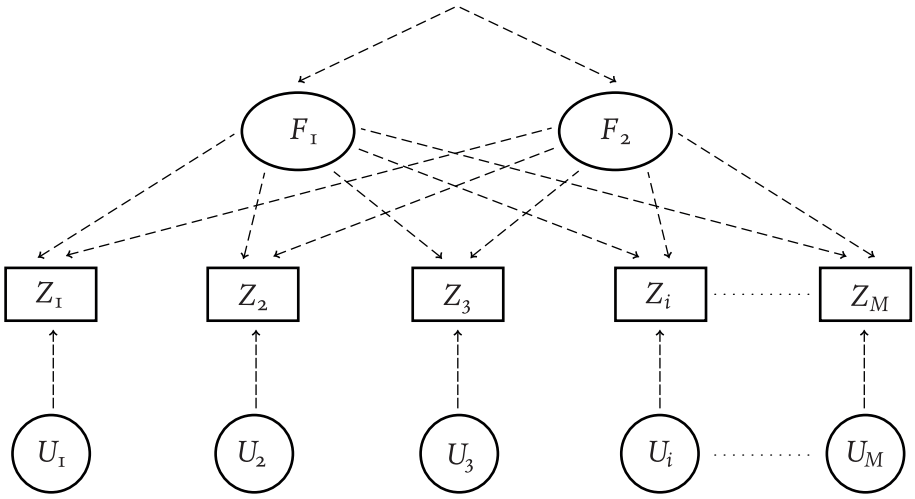

Figura 3.

2. In psicometria, e precisamente nella **teoria classica dei test**, si parla di **modello congenerico** (Jöreskog, Sörbom, 1988:76 e ss).

dove per rendere conto delle interrelazioni tra le variabili sono necessari due fattori comuni (che possono essere correlati tra loro).

La generalizzazione alle strutture multidimensionali, avviata negli anni Trenta del Novecento dagli studi di Thurstone sugli atteggiamenti, rende la tecnica flessibile ma apre un problema fondamentale: come si stabilisce il numero dei fattori comuni sottostanti? In un ambito genuinamente esplorativo non conosciamo a priori il numero dei fattori da estrarre, o al massimo abbiamo alcune ipotesi vaghe; tanto meno sappiamo dei legami che intercorrono tra i fattori e le variabili osservate. Come vedremo (par. 3.5), esistono dei criteri per farsi un'idea di massima del numero dei fattori da estrarre e per confrontare i risultati di diverse prove, sia dal punto di vista sintattico (valutazione a mezzo di indici statistici), sia dal punto di vista semantico (valutazione sostantiva dei fattori). Tuttavia, in ultima istanza spetta al ricercatore, con un giudizio complessivo, in parte anche guidato da intuizione, individuare la soluzione ottimale.

1.3. I passi caratteristici della tecnica e quelli supplementari

Prima di passare a considerare gli aspetti formali del modello, descriviamo a grandi linee i passi procedurali che caratterizzano l'esame di un insieme di variabili manifeste mediante l'*AF*.

1) Il primo passo consiste nella selezione dell'insieme di variabili che reputiamo possano fungere da indicatori, ossia per le quali si ipotizza l'esistenza di uno o più fattori latenti comuni.
   I criteri di selezione variano in funzione degli obiettivi della ricerca e delle risorse disponibili. Si possono individuare due modalità estremamente diverse di selezione delle variabili manifeste:
   - una prima modalità consiste nel condurre l'*AF* su una selezione di variabili rilevate in ricerche precedenti; questa modalità rientra nella cosiddetta **analisi secondaria dei dati** (de Lillo et al., 2011:139); accanto all'evidente vantag-

gio dal punto di vista dell'impiego di risorse, va evidenziato che le variabili disponibili potrebbero essere inadeguate allo scopo di individuare i fattori latenti comuni ipotizzati dal ricercatore;
— una seconda modalità prevede la costruzione di nuovi indicatori, o la selezione di item da repertori (nazionali o internazionali), comunque da rilevare in una nuova ricerca; in questo caso, in genere si hanno già a priori alcune ipotesi sulle dimensioni latenti che si intendono misurare (perlomeno sul numero).

In pratica poi ci si muove combinando analisi secondaria dei dati e nuove rilevazioni, tra formulazione di nuovi indicatori e selezione dai repertori.

2) Il secondo passo è la costruzione di una matrice contenente misure di concordanza tra tutte le coppie di variabili manifeste. Nell'*AF* Esplorativa, si tratta più spesso della **matrice di correlazioni lineari**, **R**, ma l'input può variare in funzione del livello di scala delle variabili manifeste (par. 2.1), nonché della finalità della ricerca: ad esempio, nel confronto tra campioni distinti, l'impiego della **matrice varianze–covarianze, S**, (una per ciascun campione a confronto), permette di calcolare e confrontare le medie dei fattori nei gruppi.
3) Il terzo passo consiste nel determinare in via induttiva, a partire dai dati contenuti nell'input minimo, il numero K ottimale di fattori latenti comuni, in grado di riprodurre adeguatamente (a meno di uno scarto minimo) le correlazioni tra le M variabili manifeste ($K \ll M$). La determinazione a priori del numero dei fattori avviene invece in un quadro che è già, almeno in parte, di analisi confermativa.
4) Il quarto passo consiste nella stima dei parametri, ossia dei coefficienti di impatto dei fattori latenti sulle variabili manifeste. Questo passo è detto comunemente **estrazione dei fattori**[3].

3. Si parla anche di *costruzione* dei fattori, data la loro natura non osservativa. Estrazione è comunque il termine più diffuso nella letteratura specialistica e tecnica.

Per estrarre i fattori sono stati sviluppati diversi algoritmi, anche molto diversi tra loro. I diversi metodi statistici hanno in comune il fatto di basarsi sullo stesso modello (equazione fondamentale e assunti; alcuni metodi richiedono assunti aggiuntivi). Attraverso le stime dei parametri si riproduce al meglio (secondo criteri prefissati) l'input minimo con una **matrice delle correlazioni riprodotte**. Quest'ultima è talvolta impiegata per individuare il numero dei fattori da estrarre; è inoltre sempre utile nella valutazione della soluzione fattoriale. I coefficienti stimati vengono raggruppati in una ***pattern matrix***, che rappresenta il nucleo della soluzione di una *AF*. Mediante questi coefficienti si cerca di fornire un'interpretazione semantica dei fattori estratti: si dà, cioè, ai fattori latenti un'etichetta che sintetizzi il contenuto delle variabili manifeste, guardando soprattutto a quelle che presentano i coefficienti più elevati. In un'indagine sui consumi, ad esempio, se i coefficienti più elevati per un certo fattore riguardano comportamenti come 'andare a teatro', 'ascoltare musica classica', 'leggere saggistica' ecc., potremo definire il fattore con un'etichetta del tipo 'consumi colti'.

5) Fin qui abbiamo considerato passi essenziali della tecnica. Il quinto passo, detto della **rotazione dei fattori**, si presenta solo quando i fattori comuni sono almeno due. La rotazione è in tal caso utile (non obbligatoria) per semplificare l'operazione di interpretazione dei fattori. L'operazione di rotazione si chiama così perché le variabili manifeste possono essere viste come punti–vettore in uno spazio a K dimensioni, dove K è il numero dei fattori. Ciò che è ruotato sono dunque gli assi di riferimento, cioè proprio i fattori. La rotazione non fa altro che ridefinire in modo più opportuno le coordinate dei vettori che rappresentano le variabili, lasciando inalterata la posizione relativa di tali vettori. Tale operazione lascia perciò inalterata la soluzione da un punto di vista globale; essa però ha un'utilità di carattere semantico, se i fattori estratti sono due o più. Nella *pattern matrix* non ruotata, di solito, ogni variabile ha legami diversi

da zero con più fattori; ciò rende difficile distinguere i fattori e interpretarli. Con la rotazione, si cerca, in linea di massima, di far passare gli assi di riferimento, i fattori, tra addensamenti di punti–vettore (variabili) in modo che risultino, il più possibile, distinti da altri addensamenti che saranno attraversati da altri fattori.
Anche per le rotazioni sono disponibili metodi diversi; esse sono classificabili in **rotazioni ortogonali**, dove la rotazione degli assi è soggetta al vincolo della perpendicolarità tra gli assi, e **rotazioni oblique**, dove tale vincolo è rilasciato del tutto o parzialmente. Nel caso di rotazioni oblique, il passo che stiamo esaminando si articola nei tre seguenti sottopassi:

— costruzione della ***pattern matrix***, contenente i coefficienti di regressione standardizzati delle variabili manifeste sui fattori latenti comuni;
— costruzione della ***structure matrix***, contenente le correlazioni tra le variabili manifeste e i fattori latenti comuni (nella rotazione ortogonale coincide con la *pattern matrix*);
— costruzione della **matrice di correlazione tra i fattori latenti**.

6) L'ultimo passo dell'*AF* come tecnica di analisi multivariata è la valutazione del modello. Si cerca, con l'ausilio di apposite statistiche, descrittive e in alcuni casi inferenziali, di valutare quanto il modello teorico si adatta alla struttura empirica dei dati. Inoltre, facendo ricorso ad alcuni parametri standard rintracciabili nella letteratura, si può valutare la significatività sostanziale dei fattori estratti.

7) Spesso, chi fa uso dell'*AF* prosegue con un passo supplementare: la stima dei **punteggi fattoriali** (*factor scores*). Esso consiste nell'attribuire a ogni caso della matrice $C \times V$ un punteggio, che rappresenta la posizione su ognuno dei fattori comuni estratti. L'input minimo in questo caso è rappresentato dalla matrice di correlazione e dalla matrice *Casi per Variabili*.

Si parla di *stima* dei punteggi fattoriali, perché nell'*AF* non è possibile calcolarli esattamente, trattandosi di un modello probabilistico[4]. In tal modo l'*AF* viene ad assumere una posizione a cavallo tra il mondo delle tecniche di analisi multivariata, che studiano la relazioni tra variabili, e quello delle **tecniche di assegnazione** (Ricolfi, 2002:4 e ss.) che producono nuove variabili. I punteggi fattoriali possono poi essere inseriti in altre analisi statistiche; ad esempio, in semplici incroci bivariati (come facciamo più avanti, nel capitolo 5) o come variabili, dipendenti o indipendenti, in modelli di regressione multipla.

4. Il calcolo esatto dei punteggi individuali è invece possibile in una tecnica simile, ma formalmente e sostantivamente diversa, che è l'Analisi in Componenti Principali (si veda il par. 3.3)

Capitolo II

Pre–condizioni

2.1. L'input minimo della tecnica

Nell'*AF*, l'**input minimo**, vale a dire la matrice dati contenente l'informazione sufficiente e necessaria per applicare la tecnica di analisi (Ricolfi, 2002:14), è tipicamente una matrice **R** di correlazioni lineari di Pearson (*product–moment correlation matrix*), calcolata a partire da una **matrice Casi per Variabili** $(C \times V)$[1]. Le variabili dovrebbero essere, a rigore, almeno a livello di scala di intervalli; tuttavia, poiché tale condizione spesso non è soddisfatta nelle scienze sociali, ci si accontenta di **variabili quasi–cardinali** (de Lillo et al., 2011:50 e 216-25), ossia variabili in cui non esiste la possibilità di verificare empiricamente l'effettiva uguaglianza degli intervalli tra le varie modalità; considereremo successivamente i correttivi e le cautele necessari per utilizzare questa sofisticata tecnica in presenza di variabili di livello inferiore.

Indichiamo innanzitutto come condizione veramente imprescindibile, che la matrice $C \times V$, perlomeno nella parte che ci serve per calcolare l'input minimo, sia una matrice ***column conditional*** nell'accezione di Carroll e Arabie (1980), ossia permetta confronti tra le celle per ogni singola colonna. È stato mostrato, anche empiricamente, che le misure di atteggiamenti (ad esempio mediante le tanto utilizzate scale Likert o i termometri), non rispettano sempre questa condizione, in quanto i soggetti intervistati utilizzano lo strumento di misura in modo non univoco[2]. In tal caso, se si vuole applicare l'*AF*, si dovrebbe-

1. Talvolta, raramente nell'*AF* Esplorativa, l'input minimo è costituito dalla matrice varianze–covarianze, **S** (Comrey, Lee, 1992, trad. it. 1995:399).

2. Questo **response style**, ossia l'uso idiosincratico delle scale di risposta, rientra in una più ampia famiglia di 'distorsioni' involontarie del dato reale che nella letteratura

ro compiere innanzitutto delle manipolazioni dei punteggi grezzi tali da rendere la matrice $C \times V$ *column conditional* (Albano, Testa 2002:37). A tale scopo, in alcuni casi è possibile ricorrere alla cosiddetta **deflazione dei dati**, in pratica una standardizzazione operata sui casi invece che sulle variabili (Albano, Testa 2002:90). Questa procedura richiede però batterie di item piuttosto numerose e fortemente eterogenee, cioè attinenti a temi diversi; pertanto non sempre è possibile applicarla ai dati in esame. In ogni caso, è importante che in sede di rilevazione dei dati vengano presi tutti gli accorgimenti che prevengano al massimo questo tipo di distorsione (come altri del resto).

In secondo luogo, una rigorosa applicazione dell'*AF* richiede che le coppie di variabili abbiano, perlomeno approssimativamente, una **distribuzione normale bivariata**. Da una distribuzione normale bivariata deriva una relazione lineare tra *X* e *Y* (mentre non è necessariamente vero il contrario). Se *X* e *Y* hanno una distribuzione normale bivariata, esse sono da considerarsi indipendenti allorquando la loro correlazione sia prossima a zero (Piccolo, 2000:425); in caso contrario, una correlazione pari a zero non indica necessariamente indipendenza: le variabili potrebbero infatti essere correlate in modo non lineare. L'assenza di distribuzioni normali bivariate non comporta necessariamente l'abbandono del modello fattoriale lineare, a patto che le relazioni intercorrenti tra le variabili manifeste e le relazioni tra le variabili manifeste e i fattori siano approssimativamente di tipo lineare.

Un altro assunto, quello di **multinormalità**, è invece indispensabile solo se desideriamo compiere valutazioni circa la significatività statistica del numero di fattori (Harman, 1976[3]:24 e capp. 9–10).

La normalità è una caratteristica di cui possono godere pienamente solo variabili almeno a livello di scala di intervalli; tuttavia, come abbiamo già detto, ciò spesso non avviene. Se si dispone di variabili con una distribuzione univariata quasi–normale, possiamo prendere anche in considerazione l'idea di sottoporle alla procedura di **standardizzazione normalizzata**, che consiste nel convertire i punteggi grezzi in punteggi standardizzati forzandoli al contempo ad assumere

psicometrica sono indicate come **response bias** (cfr. Roccato, 2003).

la forma della curva normale (Comrey, Lee, 1992, trad. it. 1995:47).

Che cosa fare invece quando si dispone di variabili quasi–cardinali, la cui distribuzione è decisamente lontana dalla forma normale? La letteratura metodologica offre a tal proposito indicazioni e valutazioni discordanti. Alcuni sostengono che, in pratica, se il numero di categorie della variabile è sufficientemente ampio (diciamo a partire da sei o sette modalità), si possono calcolare le correlazioni lineari di Pearson con qualsiasi variabile a modalità ordinate, senza ottenere eccessive distorsioni: ciò in ragione della **robustezza** mostrata dalle correlazioni prodotto–momento in studi di simulazione[3]. Altri, per mantenere un più stretto rigore metodologico, propongono delle varianti all'input minimo, miranti a ridurre le distorsioni che potrebbero insorgere nel calcolo di correlazioni prodotto–momento su questo tipo di variabili (soprattutto effetti di attenuazione della forza della relazione).

Semplificando al massimo, vediamo alcune tra le più importanti varianti proposte per variabili ordinali[4]. Conviene distinguere tre situazioni sostanzialmente diverse che possono portare a ottenere una misurazione ordinale; da ognuna di esse deriva un diverso input minimo 'ideale' per l'*AF*.

a) I dati sono originariamente rilevati su scale continue; tuttavia la presenza di un ampio errore di misurazione può rendere opportuno sostituire i valori originari con un numero minore di modalità solo ordinate. Come esempio, immaginiamo di aver intervistato un gruppo di individui e di aver chiesto di indicare quale sia il loro gradimento per un numero elevato di partiti. È presumibile che diversi elementi alimentino l'errore di valutazione dato dal soggetto (incertezza intrinseca, livello di attenzione ecc.): il passaggio da una misura a livello di rapporti a una ordinale ha il vantaggio di attenuare l'errore (Il sig. Rossi ha risposto che su una scala di gradimento da zero a 100 valuta '20'

3. Questa posizione non è sempre accuratamente argomentata dai suoi sostenitori; esistono comunque anche delle basi teoriche che sono state sviluppate a partire dai primi studi di simulazione compiuti da Labovitz già alla fine degli anni Sessanta (cfr. Labovitz, 1970; O'Brien, 1979).

4. Per una trattazione più approfondita si rinvia a Basilevsky, 1994, cap. 8.

il partito *A*, '50' il partito *B* e così via; al partito *Q*, il quindicesimo della lista dà un voto pari a '15'. I partiti *A* e *Q* hanno quindi ricevuto valutazioni diverse, ma in realtà è probabile che per il sig. Rossi in pratica essi si equivalgano).
Un altro caso in cui può essere utile questo abbassamento del livello di scala è quello in cui le variabili presentano relazioni non lineari: la riduzione a un numero limitato di modalità costituenti una variabile ordinale può avere come effetto quello di linearizzare le relazioni. In casi come questi, dove cioè le variabili sono originariamente continue ma trasformate per opportunità in ordinali, la misura di concordanza più appropriata è il coefficiente di **correlazione per ranghi di Spearman**.

b) Le modalità di ciascuna variabile sono genuinamente ordinali: le relazioni tra modalità sono rappresentabili esclusivamente con gli operatori 'maggiore di' e 'minore di'. Esempi sono: il titolo di studio; il livello di carriera aziendale; alcuni indici di status sociale; le graduatorie di un certo numero di soggetti osservati, in base ad alcune competenze di tipo psico–sociale mostrate in una serie di prove; l'ordine di importanza dato a un insieme di oggetti, per esempio istituzioni sociali.
Un approccio consigliato nella letteratura specialistica consiste nel ricorrere a operatori di concordanza non parametrici, come il **tau di Kendall**; un'altra strada, più pragmatica, è quella di ignorare la natura non metrica dei dati, ricorrendo a misure euclidee come il coefficiente di correlazione di Spearman (Basilevsky, 1994:512 e ss.).

c) I dati sono all'origine rilevati su scale ordinali; tuttavia si può assumere ragionevolmente che la proprietà rilevata sia per sua natura continua. Questo approccio, noto come ***underlying variable approach*** (cfr. Kampen, Swyngendouw, 2000), può essere seguito da chi analizza scale quasi–cardinali, come capita per esempio nella misura degli atteggiamenti. Il coefficiente adatto può essere anche in questo caso la correlazione per ranghi di Spearman; tuttavia, se è plausibile assumere che il continuum sottostante a ogni rilevazione ordinale sia distribuito normalmente, la misura di concordanza più appropriata è il coefficiente

di **correlazione policorica**, se entrambe le variabili osservate sono ordinali, oppure la **correlazione poliseriale**, nel caso che una di esse sia una variabile continua (Jöreskog, Sörbom, 1986).

È decisamente sconsigliabile ricorrere all'*AF* Lineare se le relazioni tra le variabili non sono monotoniche (condizione controllabile previamente mediante una ispezione dei **diagrammi di dispersione** di tutte le distribuzioni congiunte tra coppie di variabili). In tal caso infatti, non solo le stime dei parametri saranno distorte, ma è anche elevato il rischio che la soluzione ottenuta sia caratterizzata da eccesso di dimensionalità[5]. Nel caso di relazioni monotoniche non lineari sarebbe meglio linearizzare le relazioni tra variabili mediante una opportuna trasformazione matematica (ad esempio logaritmica).

Cautela è richiesta anche per l'applicazione, molto diffusa, dell'*AF* a batterie di item dicotomici. Una prima obiezione potrebbe essere che essa non è lecita in quanto le variabili non sono continue; tale obiezione può però essere facilmente superata applicando per esempio l'*underlying variable approach* e calcolando in questo caso le **correlazioni tetracoriche** (o le **correlazioni biseriali**, se una delle due variabili osservate è continua). Il vero problema però è un altro. Occorre infatti tenere presente che la relazione tra un fattore e un item dicotomico *non può* essere lineare, perlomeno non su tutto il continuum del fattore, ma al massimo in un certo intervallo centrale, escluse cioè le code della distribuzione: discorso che non suonerà nuovo a chi ha esperienza del cosiddetto **modello di regressione lineare binomiale** (Pisati, 2003) , in cui la variabile dipendente è dicotomica. A rigore sarebbe perciò meglio passare ai modelli della ***Item Response Theory***, in cui il modello per lo studio dei dati non è la retta ma la **curva logistica** (o altre curve sigmoidali). In pratica, questo non è necessario se le variabili osservate hanno un basso **parametro di discriminazione** (o

5. Sul tema, noto anche come 'problema di Coombs' si veda la trattazione approfondita in Ricolfi, 1999:254. Va inoltre segnalato che sono stati sviluppati modelli di analisi fattoriale non lineare, che però esulano dalla presente trattazione. Per chi fosse interessato all'argomento, il riferimento teorico principale è l'opera dello psicometrico Roderick P. McDonald che ha elaborato tali modelli a partire dall'inizio degli anni Sessanta. La sua teoria è stata implementata nel software NOHARM da Colin Fraser.

'sensibilità'): in tal caso le curve che rappresentano il legame tra variabili manifeste e fattori latenti sono meno 'ripide' e il modello fattoriale lineare rappresenta una buona approssimazione (cfr. McDonald, 1982), per lo meno a scopo esplorativo.

Ai nostri fini, è sufficiente aver fatto cenno all'esistenza delle varianti dell'input minimo dell'*AF* Esplorativa; d'ora in avanti ci atterremo a una posizione per così dire 'tradizionale', considerando come input minimo una matrice di correlazioni prodotto–momento calcolate su variabili cardinali (o assimilabili) legate tra loro e con il fattore latente in modo lineare (o quasi).

In chiusura di questo paragrafo sulle pre–condizioni relative all'input minimo, spendiamo ancora alcune parole circa il livello che le correlazioni devono assumere. Un buon input per l'*AF* Esplorativa è una matrice contenente correlazioni consistenti; in genere si ottengono scarsi risultati con correlazioni inferiori a $|0,3|$; del resto, la presenza di singole correlazioni superiori a tale soglia non garantisce che la matrice sia adeguata per il modello. Sono stati messi a punto degli **indici diagnostici di fattorializzabilità** della matrice, che ci possono suggerire se vale la pena applicare questo modello ai nostri dati.

Se le variabili manifeste sono indipendenti, la matrice di correlazioni assomiglierà a una **matrice identità** (vd. appendice I): pertanto, un primo test può consistere nel controllare l'ipotesi che la matrice delle correlazioni nella popolazione sia una matrice identità ($H_0 : \mathbf{R} = \mathbf{I}$). Questo è ciò che fa il **test di sfericità di Bartlett**. Avrà senso applicare l'*AF* solo se il valore della statistica–test è elevato e basso il *p–value* osservato, perché in tal modo si potrà respingere l'ipotesi nulla secondo le consuete soglie di significatività[6].

Poiché si tratta di un test inferenziale la sua corretta interpretazione si ha quando il campione è probabilistico; inoltre, deve valere l'assunto che le variabili nella popolazione abbiano una distribuzione normale multivariata.

Un paio di indici descrittivi, che non richiedono la casualità del campione, si basano sul **coefficiente di correlazione parziale** (cfr.

6. Per un'illustrazione generale dei test di ipotesi statistica rimandiamo a Albano, Testa, 2002, par. 6.5.

Barbaranelli, 2003:63–4). Assumendo che le variabili siano influenzate da fattori comuni e che i fattori unici siano incorrelati, i coefficienti di correlazione tra ciascuna coppia di variabili manifeste devono essere prossimi a zero, una volta tenute sotto controllo le altre variabili. Se vi sono molte correlazioni parziali grandi (in modulo), occorre rivedere la specificazione del modello (il set di variabili manifeste) o adottarne un altro.

Sulla correlazione parziale si basano l'*MSA* e il *KMO*.

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} p_{ij}^2} \qquad KMO = \frac{\sum\sum_{j \neq i} r_{ij}^2}{\sum\sum_{j \neq i} r_{ij}^2 + \sum\sum_{j \neq i} p_{ij}^2}$$

In entrambe le formule r_{ij} e p_{ij} rappresentano la correlazione osservata e quella parziale tra due variabili qualsiasi.

Il test **MSA**–Measure of Sample Adequacy si applica per una diagnosi di ciascuna singola variabile. In presenza di valori di MSA di scarsa entità si dovrebbero escludere dall'analisi le relative variabili, con conseguente rispecificazione del modello.

Un indice analogo, ma complessivo, è il **KMO–Kaiser–Meyer–Olkin**, (detto anche *overall MSA*).

Il livello minimo accettabile di *MSA* e *KMO* è 0,50, ma sono considerati valori buoni quelli che superano la soglia di 0,80.

2.2. Costruzione delle variabili e del campione

L'applicazione standard dell'*AF* richiede che la costruzione delle variabili manifeste avvenga secondo alcune regole. Un primo importante pre–requisito è che le variabili manifeste siano per costruzione indipendenti tra loro. Con ciò non si intende dire che non devono presentare concordanza (correlazioni o covarianze), condizione che invece è auspicabile. L'indipendenza in questione è di carattere logico: lo stato di un soggetto sulla variabile *Y* deve risultare da una rilevazione empirica autonoma, non da una derivazione logica a partire dalla conoscenza

dello stato del soggetto sulla variabile X. Un esempio in cui non si dà indipendenza logica è costituito dai **bilanci–tempo** delle persone: le variabili che rappresentano la quantità di ore giornaliere dedicata a certe attività, sono soggette a un vincolo di riga[7]. In casi come questi si dà **multicollinearità** tra K variabili: i valori di una variabile sono perfettamente prevedibili conoscendo gli stati sulle altre $K - 1$. Nel caso di **perfetta collinearità** la tecnica non funziona, in quanto la matrice di correlazioni non è invertibile (operazione che è alla base di molte procedure di analisi multivariata dei dati). Ci sono altre situazioni in cui le variabili non sono indipendenti per costruzione. È il caso per esempio degli **item a scelta multipla** (o 'punteggi ipsativi'), in cui si chiede al soggetto di scegliere tra M elementi un sottoinsieme per lui più importanti o preferiti. Spesso, per ognuno degli elementi si costruisce una variabile che ha come valore il rango attribuitogli da ciascun soggetto. In questi casi non si dà collinearità, ma è chiaro che le variabili costruite a partire da questa domanda non sono logicamente indipendenti; il problema principale è che esse potrebbero non risultare correlate anche quando una relazione in realtà esiste[8].

Altra condizione importante è che ogni variabile manifesta abbia un campo di variazione sufficientemente esteso. Le variabili con scarsa dispersione intorno al valore centrale risultano in genere poco

7. Consideriamo il caso più semplice, con due sole variabili che descrivono il bilancio–tempo di un campione di soggetti: $X1$–"tempo dedicato al lavoro" e $X2$–"tempo extralavorativo". Se un soggetto dichiara di destinare mediamente 10 delle sue ore giornaliere all'attività lavorativa, possiamo dedurre che il suo tempo medio di attività extra–lavorativa è pari a 14 ore giornaliere: le due variabili non sono per costruzione indipendenti, perché sottoposte a un vincolo di riga (la somma deve dare necessariamente 24).

8. Questo per vari motivi. Immaginiamo di sottoporre una lista di 15 oggetti agli intervistati e che sia chiesto loro di indicare quali oggetti pongono nei primi 5 posti. Per ogni oggetto verrà costruita una variabile che avrà come campo di variazione potenziale i valori interi da 1 a 6, dove 6 significa che l'oggetto non rientra nei primi cinque. Se per ipotesi un oggetto della lista non viene mai scelto tra i primi 5, la variabile corrispondente sarà una costante (una sfilza di 6) e risulterà necessariamente incorrelata con altre variabili. Un altro esempio: gli oggetti A e B vengono collocati sempre al primo e secondo posto, ora l'uno ora l'altro. Se avessimo lasciato liberi gli intervistati di dare a ciascun oggetto un voto indipendente, avremmo potuto anche osservare molte situazioni in cui i due oggetti ricevono lo stesso voto. Con il metodo dei punteggi ipsativi è facile, cioè, ottenere un correlazione più attenuata o tendenzialmente negativa rispetto al metodo delle valutazioni indipendenti.

correlate con altre variabili (per rendersene conto basta pensare alla formula dell'operatore statistico covarianza). È importante che le variabili osservate siano adatte a discriminare posizioni diverse sulla proprietà latente che si vuole misurare; inoltre, nessuna restrizione nel campo di variazione delle variabili deve essersi verificata in seguito a peculiarità nel metodo di selezione del campione (ad esempio: misurare l'interesse per la politica in un campione di lettori di una rivista politica fornirebbe molto probabilmente risposte del tutto o quasi del tutto omogenee).

In merito alle modalità di scelta dei casi e alla numerosità campionaria sono opportune le seguenti considerazioni.

a) Il campione non deve essere necessariamente probabilistico; tale requisito è indispensabile soltanto se si vogliono stimare i parametri ignoti di una **popolazione**. La casualità del campione non è invece necessaria in altri ambiti altrettanto importanti e frequenti della ricerca scientifica come ad esempio:
   - — nell'ambito di una **ricerca pilota**, che precede la ricerca vera e propria, in cui l'obiettivo è unicamente di testare gli strumenti (questionario, scale ecc.);
   - — nei **censimenti**, quando cioè il campione coincide con la popolazione obiettivo;
   - — quando la ricerca non ha un taglio **idiografico** ma **nomologico**, mirante cioè a individuare regolarità universalmente valide; si noti di passaggio che questo secondo caso è poco frequente al di fuori di situazioni sperimentali[9].

b) Con campioni piccoli, l'errore casuale dei coefficienti di correlazione meno attendibili determina un aumento del valore

9. Anche nella ricerca quasi–sperimentale o negli studi di covariazione accade di generalizzare i risultati ottenuti a una popolazione non specifica, soprattutto quando si studiano relazioni tra le variabili (il ricercatore non è interessato ai casi, se non in quanto 'supporto' attraverso il quale le variabili possono manifestarsi); devono però esserci delle buone ragioni per non considerare i risultati come idiosincratici alle unità selezionate. Questi temi, riguardanti la generalizzabilità dei risultati di una ricerca empirica, sono rubricati dai metodologi sotto la voce **validità della ricerca** (cfr. Pedon, Gnisci, 2004, cap. V).

assoluto delle correlazioni stesse (Comrey, Lee 1992, trad. it. 1995:284). All'aumentare della numerosità campionaria, aumenta di solito l'**attendibilità**[10] delle correlazioni statisticamente diverse da zero, ma esse tendono a diminuire in modulo. Secondo Comrey e Lee, 50 casi sono un numero decisamente scarso; 200 casi sono ritenuti un numero adeguato; campioni particolarmente numerosi ($N = 1000$) sono richiesti se si usano correlazioni diverse da quelle prodotto–momento di Pearson, per ottenere lo stesso livello di stabilità dei coefficienti di correlazione (Comrey, Lee, cit.:284). Si tratta comunque di indicazioni molto generali che non vanno prese in modo assoluto: ogni ricerca presenta delle specificità, su cui il ricercatore è chiamato a una attenta riflessione, anche per quanto riguarda la numerosità campionaria.

10. Ossia, in termini molto generali, la riproducibilità del risultato in prove ripetute ceteris paribus; per un approfondimento si veda de Lillo et al., 2011:37 e ss.

Capitolo III

Il modello

3.1. Aspetti terminologici

Consideriamo l'equazione di base dell'*AF*:

$$Z_{in} = a_{i1}F_{1n} + a_{i2}F_{2n} + \ldots + a_{ik}F_{kn} + \ldots + a_{iK}F_{Kn} + U_{in}; \quad \sum_{k=1}^{K} a_{mk}F_{kn} + U_{mn}$$

— Z_{in} è il punteggio standardizzato della variabile i–esima per l'individuo n–esimo
— a_{ik} è detto **saturazione fattoriale** o *factor loading* della variabile manifesta i–esima sul fattore k–esimo, con $k < m$
— F_{kn} è il punteggio standardizzato del k–esimo fattore comune per l'individuo n–esimo (*factor score*)
— U_{in} è il punteggio standardizzato dell'i–esimo fattore unico per l'individuo i–esimo.

Ognuno dei fattori unici si può idealmente scomporre così[1]:

$$U = S + E$$

dove S rappresenta il **fattore specifico** che influisce sulla corrispondente variabile manifesta Z, mentre E è l'**errore accidentale**.

È definita **comunalità** di una variabile manifesta la quantità:

$$\sum_{k=1}^{K} a_{ik}^2$$

1. Per semplificare, ometteremo le indicizzazioni delle variabili ogni volta che ciò non sia causa di fraintendimenti; preferiamo sacrificare il rigore notazionale per metterci dalla parte di chi affronta la lettura delle formule da non specialista.

La comunalità di una variabile viene talvolta indicata con h^2 e rappresenta cioè la parte di varianza di una variabile manifesta spiegata dai fattori comuni.

Il fattore U elevato al quadrato determina la cosiddetta **unicità**; essa è data quindi dalla somma di S^2 e di E^2, che sono definite rispettivamente **specificità** e **varianza dell'errore stocastico**.

Consideriamo ora il seguente sistema di equazioni:

$$Z_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1K}F_K + U_1$$

$$Z_2 = a_{21}F_1 + a_{22}F_2 + \ldots + a_{2K}F_K + U_2$$

$$\vdots$$

$$Z_M = a_{M1}F_1 + a_{M2}F_2 + \ldots + a_{MK}F_K + U_M$$

Esso prende il nome di ***factor pattern***; raccogliendo gli elementi a in una matrice **A**:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MK} \end{bmatrix}$$

otteniamo la cosiddetta *pattern matrix*.

Consideriamo ora un altro sistema di equazioni:

$$r_{z_i F_1} = a_{i1} + a_{i2}r_{F_1F_2} + \ldots + a_{ik}r_{F_1F_k} + \ldots + a_{ik}r_{F_1F_k}$$

$$\vdots$$

$$r_{z_i F_k} = a_{i1}r_{F_kF_1} + a_{i2}r_{F_kF_2} + \ldots + a_{ik}r_{F_kF_k} + \ldots + a_{ik}$$

che definiamo ***factor structure***.

Raccogliendo in modo opportuno i termini posti a sinistra nel sistema di equazioni in una matrice **S**, ovvero come segue:

$$\begin{bmatrix} r_{z_1F_1} & r_{z_1F_2} & \cdots & r_{z_1F_K} \\ \vdots & \vdots & \vdots & \vdots \\ r_{z_MF_1} & r_{z_MF_2} & \cdots & r_{z_MF_K} \end{bmatrix}$$

otteniamo la cosiddetta *structure matrix*. Ricordiamo che quest'ultima matrice è distinguibile dalla *pattern matrix* solo quando si effettua una rotazione obliqua.

Infine, una puntualizzazione circa l'uso del termine saturazione. Parliamo di saturazioni, senza ulteriori specificazioni, per indicare i coefficienti da stimare quando ci riferiamo alla soluzione non ruotata oppure ruotata in modo ortogonale. Nel caso di rotazioni oblique invece, poiché le matrici di parametri fornite dalla tecnica per interpretare i fattori sono due, *A* e *S*, chiameremo ***regression loading*** gli elementi della *pattern matrix* e ***correlation loading*** gli elementi della *structure matrix*.

3.2. Gli assunti sulla struttura fattoriale

Si sarà notata una certa somiglianza del modello dell'*AF* con quello ben più noto della regressione multipla; tuttavia, a differenza di quanto avviene in quest'ultima, i fattori, cioè le variabili indipendenti nel modello fattoriale, sono variabili inosservabili e quindi incognite. Senza porre ulteriori restrizioni, la stima del modello non sarebbe possibile: infatti si potrebbero individuare infiniti valori per i parametri *a* e per i fattori *F*, tutti in grado di soddisfare le equazioni del modello. Il modello della *AF* Esplorativa si basa perciò sui seguenti assunti:

— $E(F_k) = 0$ (il valore medio di ciascun fattore comune è pari a zero);
— $Var(F_k) = 1$ (la varianza di ciascun fattore comune è pari a uno);
— Cov $(F_k, U_i) = 0$ (i fattori comuni sono indipendenti, o ortogonali, rispetto ai fattori unici);
— Cov $(U_i, U_j) = 0$ (i fattori unici sono ortogonali tra loro).

Inoltre, senza perdita di generalità, si assume che:

— Cov $(F_k, F_j) = 0$ (vedremo però che tale condizione viene spesso rilasciata in fase di rotazione dei fattori e precisamente nelle rotazioni oblique).

3.3. Analisi Fattoriale e Analisi in Componenti Principali

Alla base dell'*AF* riposa l'idea che ognuno dei punteggi *Z* che concorrono a produrre la matrice di correlazioni possa essere espresso come somma ponderata del punteggio nei fattori comuni e di quello del fattore unico (a sua volta composto da un fattore specifico e un fattore di errore).

I punteggi *Z*, *F*, *U* sono tutti standardizzati, pertanto aventi media pari a zero e varianza pari a 1.

Le saturazioni (a_{ij}), che ricordiamo sono oggetto di stima nell'*AF*, hanno di norma valore compreso tra −1 e +1 (escludendo per ora il discorso delle rotazioni oblique).

L'**Analisi in Componenti Principali** (*ACP*), tecnica sviluppata da Hotelling negli anni Trenta — ma le cui basi sono già contenute nei lavori degli statistici Bravais (cit. in Basilevsky, 1994:99) e Pearson (1901) — non va confusa con l'*AF*.

Osserviamo l'equazione dell'*ACP* (La notazione cambia volutamente, per marcare la differenza con l'*AF*):

$$Z_{in} = p_{i1}C_{1n} + p_{i2}C_{2n} + \ldots + p_{iM}C_{Mn}$$

Ognuna delle *M* variabili osservate è descritta linearmente nei termini di *M* nuove componenti incorrelate: $C_1 C_2 \ldots C_M$.

Le componenti a loro volta non sono altro che combinazioni lineari delle variabili *Z*. Le componenti sono costruite in modo tale che la prima spieghi la % massima di varianza del sistema, la seconda la % massima di varianza residua e così via; analiticamente abbiamo:

$$C_1 = c_{11}Z_1 + c_{12}Z_2 + \ldots + c_{1M}Z_M = \sum_{m=1}^{M} c_{1m}Z_m$$

$$\mathrm{VAR}(C_1) = \max$$

sotto il vincolo che:

$$\sum_{i=1}^{M} {c_{1i}}^2 = 1$$

Nella *AF* ognuna delle *M* variabili osservate è ridescritta linearmente nei termini di *K* fattori comuni (con $K \ll M$), di un fattore specifico e di una componente stocastica. Nella *ACP* si vuole spiegare un numero *M* di variabili manifeste con un numero *M* di componenti che variamente contribuiscono alle varianze dei test. Le differenze con l'*AF* si attenuano per quanto concerne i risultati ottenuti con l'**Analisi in Componenti Principali Troncata**, quando cioè si trattiene un numero di *K* di componenti molto inferiore al numero *M* di variabili manifeste.

L'*ACP* Troncata fornisce la seguente scissione di una matrice di Correlazioni (o Var Cov):

$$\mathbf{R}_{(M \cdot M)} = {}_K\mathbf{P}_{(M \cdot K)} \cdot {}_K\mathbf{P}'_{(K \cdot M)} + \mathbf{D}_{(M \cdot M)}$$

dove ${}_K\mathbf{P}$ rappresenta una matrice di pesi di *K* componenti principali, ${}_K\mathbf{P}'$ è la sua trasposta e **D** è la matrice contenente i residui dovuti al fatto che non abbiamo considerato le ultime $M - K$ componenti (in pedice tra parentesi è indicato l'ordine di ciascuna matrice).

Questa variante dell'*ACP* è spesso utilizzata al posto della *AF* vera e propria. Ciò trova le sue origini nella maggior semplicità concettuale di questa tecnica e in una posizione epistemologica improntata all'empirismo: quest'ultima rifiuta il concetto di variabile latente e l'idea stessa che il dato empirico si possa considerare composto da 'segnale puro' (quello determinato dai fattori comuni) e da 'rumore' (determinato dal fattore specifico e dalla componente erratica).

Le due tecniche non sono comunque in alcun modo equivalenti da un punto di vista logico, anche se spesso (ma non necessariamente) possono portare a risultati molto simili.

L'*ACP*, nella sua versione completa o in quella troncata, è una tecnica orientata alla varianza di un sistema di equazioni lineari: essa ha infatti come funzione obiettivo quella di riscrivere un sistema di variabili, sostituendo a queste delle componenti ordinabili per quantità di varianza spiegata da ognuna del sistema complessivo. Un'*ACP* si ritiene riuscita se un numero ridotto di componenti, *K*, riesce a riprodurre gran parte (indicativamente, l'80–90%) della varianza del sistema originario.

L'*AF* si 'nutre' invece di covariazioni tra le variabili: la funzione obiettivo, comunque venga specificata, mira a annullare tali covariazioni imputandole all'azione di fattori latenti comuni. La varianza spiegata, come vedremo è un *by–product*, non è quasi mai l'obiettivo principale.

Per la misurazione di costrutti teorici a mezzo di indicatori è più importante l'*AF*: questa tecnica è in grado di filtrare la parte indicante degli indicatori dal 'rumore', composto dall'errore di misurazione e dai fattori specifici. Con l'*ACP* troncata invece rumore e segnale puro sono codificati insieme.

Se il modello generatore di dati è:

$$\mathbf{R}_{(M\cdot M)} = {}_{K}\mathbf{A}_{(M\cdot K)} \cdot {}_{K}\mathbf{A}'_{(K\cdot M)} + \mathbf{U}^2_{(M\cdot M)}$$

dove ${}_{K}\mathbf{A}$ è un matrice di parametri da stimare e $\mathbf{U}^2$ una matrice di disturbi, si può essere tentati di usare per tale stima la scissione di **R** fornita dalla *ACP* Troncata (vedi sopra): le strutture del processo generatore e della suddetta scissione sembrerebbero essere isomorfe. Ma così non è: la matrice $\mathbf{U}^2$ è diagonale (i fattori unici si assumono mutuamente incorrelati), mentre non lo è la matrice **D**.

Quindi ${}_{K}\mathbf{P}$ non approssima di norma ${}_{K}\mathbf{A}$.

I risultati di un'*AF* e di un'*ACP* Troncata possono essere molto simili (nella seconda le saturazioni sono inflazionate rispetto alla prima, ma presentano lo stesso *pattern*) (Harman, 1976[3]:135); in particolare i risultati non differiscono sostanzialmente quando:

— le comunalità sono molto alte (cioè se il rumore è basso: cosa che avviene molto di rado nelle scienze sociali e del comportamento);
— il numero di variabili osservate è molto ampio; al loro aumentare, infatti, il peso dei valori in diagonale diminuisce rispetto a quello delle covarianze.

Quando le risorse di calcolo erano scarse si ricorreva all'*ACP* Troncata in luogo dell'*AF* per semplificare i calcoli. Oggi, comunque, questa giustificazione è venuta meno; va segnalato tuttavia che alcuni importanti software statistici trattano, a nostro parere in modo fuorviante,

l'*ACP* come uno dei diversi metodi di estrazione di fattori latenti comuni.

L'*ACP* Troncata è una tecnica utile per altri scopi: primo fra tutti quello di costruire indici sintetici a partire da batterie più o meno ampie di item in cui non è rintracciabile una precisa struttura latente[2].

Sino qui abbiamo considerato il rapporto tra variabili osservate e costrutto teorico come un rapporto di indicazione. Sono definiti ***reflective indicators*** (Bagozzi, 1994:331) quegli indicatori che 'riflettono' alcune proprietà della dimensione sottostante: per usare una metafora, noi guardiamo il costrutto latente attraverso le lenti delle variabili manifeste. Se un insieme di indicatori riflessivi è riferito alla stessa dimensione è lecito attendersi che essi covarino (stiamo misurando la stessa proprietà con strumenti diversi). Inoltre, poiché la variabile latente ha una sua realtà parzialmente autonoma dagli indicatori, si può applicare il concetto di **intercambiabilità** degli indicatori empirici (Lazarsfeld, 1958:105 e ss): possiamo osservare la stessa proprietà latente mediante una diversa batteria di indicatori.

Diversi sono i cosiddetti ***causal indicators*** o *formative indicators* (Bagozzi, 1994:331; Bollen, 2001:7283): questi sono indicatori che presi singolarmente misurano certi aspetti di una proprietà più complessa, la quale è data, per definizione, dalla somma dei suoi indicatori. In questo caso, non è richiesto che gli indicatori mostrino una qualsiasi struttura di covarianza, perché non stiamo misurando una variabile latente responsabile delle covariazioni delle variabili manifeste, bensì 'costituendo' una proprietà complessa, combinando proprietà più semplici che possono anche essere in competizione tra loro (per es: consumo alcolico). Inoltre, non ha senso parlare di intercambiabilità degli indicatori: un set diverso di indicatori costituisce una proprietà necessariamente diversa.

2. Per usi più specialistici dell'Analisi in Componenti Principali rinviamo a Dunteman, 1989. Per un ulteriore approfondimento sulle differenze tra Analisi Fattoriale e Analisi in Componenti Principali, rinviamo a Widaman, 2007.

3.4. Dall'input alla matrice riprodotta

A partire dai punteggi *Z* (variabili manifeste standardizzate) vogliamo risalire alla matrice delle saturazioni, passando attraverso la matrice di correlazioni (quest'ultima, ricordiamo, è l'input minimo). Per comprendere tale processo, si faccia in un primo momento il ragionamento inverso: cioè si immagini che sia nota la parte destra dell'equazione fondamentale e che oggetto di calcolo siano le variabili *Z*. Si tratta naturalmente di un puro esperimento mentale, con cui fingiamo di conoscere il punteggio degli individui sui fattori latenti (comuni, specifici e d'errore), nonché i coefficienti (saturazioni) con cui tali punteggi si combinano in modo additivo. Per ricavare i valori delle variabili *Z* ricorreremmo allora alla seguente operazione matriciale:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} & 1 & & \\ a_{21} & a_{22} & \cdots & a_{2K} & & 1 & \\ \vdots & \vdots & \vdots & \vdots & & & \ddots \\ a_{M1} & a_{M2} & \cdots & a_{MK} & & & 1 \end{bmatrix} \times \begin{bmatrix} F_{11} & \cdots & F_{1N} \\ \vdots & \vdots & \vdots \\ F_{K1} & \cdots & F_{KN} \\ U_{11} & \cdots & U_{1N} \\ \vdots & \vdots & \vdots \\ F_{M1} & \cdots & F_{MN} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & Z_{22} & \cdots & \\ \vdots & & \cdots & \\ Z_{M1} & Z_{M2} & \cdots & Z_{MN} \end{bmatrix}$$

dove *K* è il numero dei fattori comuni, *M* è il numero delle variabili manifeste, *N* il numero dei casi.

In termini più sintetici possiamo scrivere:

$$\underset{(M\cdot(K+M))}{\mathbf{A}_u} \cdot \underset{((K+M)\cdot N)}{\mathbf{F}_u} = \underset{(M\cdot N)}{\mathbf{Z}}$$

Nell'*AF* compiamo il percorso inverso a quello appena ipotizzato: a partire dalla conoscenza della matrice $\mathbf{Z}$ cerchiamo di risalire agli elementi contenuti nelle celle della matrice $\mathbf{A_u}$, o meglio quelli contenuti nelle sue prime *K* colonne; chiamiamo **A** questa porzione di matrice, che come abbiamo già visto è la *pattern matrix*.

Analogamente, chiamiamo **F** la parte superiore di $\mathbf{F_u}$, cioè le prime *K* righe di quest'ultima.

I valori a_{mk} contenuti nella pattern matrix possono essere interpretati come i pesi beta delle regressioni delle variabili osservate sui fattori

latenti nonché, se escludiamo il caso delle rotazioni oblique, come coefficienti di correlazione tra variabili manifeste e fattori comuni. Vediamo ora come, conoscendo la matrice $\mathbf{Z}$ e calcolando le correlazioni tra le variabili si può giungere a stimare gli elementi della matrice $\mathbf{A}$.

Se nella formula della correlazione prodotto–momento:

$$r_{ij} = \text{Cov}\,(Z_i, Z_j) = \frac{1}{N}\sum_{n=1}^{N} Z_{in} Z_{jn}$$

sostituiamo il membro di destra dell'equazione di base (vedi par. 3.1) otteniamo:

$$r_{ij} = \frac{1}{N}\sum_{n=1}^{N} \begin{array}{l}(a_{i1}F_{1n} + a_{i2} + F_{2n} + \ldots + a_{ik}F_{kn} + U_{in}) \\ \cdot(a_{J1}F_{1n} + a_{J2} + F_{2n} + \ldots + a_{Jk}F_{kn} + U_{jn})\end{array}$$

Assumendo che i fattori siano tutti incorrelati fra loro (condizione che può essere rilasciata dopo aver stimato il modello, in sede di rotazione), e svolgendo il prodotto precedente otteniamo:

$$r_{mj} = \frac{1}{N}\sum_{n=1}^{N} (a_{m1}a_{j1}F_{1n}{}^2 + a_{m2}a_{j2}F_{2n}{}^2 + \ \ldots\ + a_{mk}a_{jk}F_{kn}{}^2)$$

inoltre, essendo i fattori espressi anch'essi in unità standardizzate, abbiamo che:

$$\frac{1}{N}\sum_{n=1}^{N} F_n{}^2 = \sigma_F{}^2 = 1$$

Pertanto possiamo semplificare e scrivere:

$$r_{ij} = a_{i1} \cdot a_{j1} + a_{i2} \cdot a_{j2} + \ldots + a_{ik} \cdot a_{jk}$$

in altri termini, la correlazione tra la variabile i–esima e la variabile j–esima corrisponde alla sommatoria dei prodotti delle loro saturazioni fattoriali nei fattori comuni.

Considerando simultaneamente tutte le r_{ij} otteniamo la matrice di correlazione; in notazione matriciale:

$$\mathbf{R_u} = 1/N \cdot (\mathbf{Z} \cdot \mathbf{Z}')$$

Poiché:

$$\mathbf{Z} = \mathbf{A_u} \cdot \mathbf{F_u}$$

possiamo scrivere:

$$\mathbf{R_u} = \mathbf{A_u} \cdot \left[\frac{\mathbf{F_u} \cdot \mathbf{F_u}'}{N} \right] \cdot \mathbf{A_u}'$$

e semplificando, avendo assunto che i fattori siano tra loro incorrelati, si ottiene:

$$\mathbf{R_u} = \mathbf{A_u} \cdot \mathbf{A_u}'$$

Infine, ammesso che i nostri assunti siano validi e senza riguardo per ora a ciò che compare nelle celle della diagonale principale, possiamo scrivere in notazione matriciale[3]:

$$\mathbf{R_{com}} = \mathbf{A} \cdot \mathbf{A}'$$

ricordando che con **A** si indica la parte di matrice $\mathbf{A_u}$ contenente solo i coefficienti dei fattori comuni. Quest'ultima equazione è chiamata ***fundamental factor theorem*** (Thurstone, 1935:70).

$\mathbf{R_u}$ e $\mathbf{R_{com}}$ sono identiche tranne che nella diagonale principale: $\mathbf{R_u}$ contiene degli uno, $\mathbf{R_{com}}$ contiene le comunalità (h^2). Ciò equivale a dire: per spiegare le correlazioni tra variabili, sono sufficienti i fattori comuni. Per spiegare la varianza totale sono necessari i fattori comuni e quelli unici. Il processo di estrazione dei fattori può seguire diversi metodi; inizia nell'approccio tradizionale, che possiamo associare al nome di Thurstone, con una matrice $\mathbf{R_{com}}$, dopo che sono state stimate in qualche modo le comunalità e inserite nella diagonale principale[4].

3. «*The product of the reduced factor matrix by its transpose is the reduced correlation matrix*», (Thurstone, 1947:81)

4. L'idea di mettere le comunalità nella diagonale principale è stata legittimata da Thurstone. Il rango di una matrice di correlazioni raramente è minore del numero di variabili

Le stime iniziali per le comunalità di solito usate sono o le correlazioni multiple al quadrato di ogni variabile con le restanti variabili, o la correlazione più alta in una riga della matrice di correlazione. Il processo prosegue, mediante iterazioni, sino a che per esempio, secondo certi parametri prefissati, le stime delle comunalità non si stabilizzano, oppure sino a che non si ottiene una matrice $\mathbf{R}_{\mathbf{repr}}$ sufficientemente simile a $\mathbf{R}_{\mathbf{com}}$ negli elementi extradiagonali.

Mentre $\mathbf{R}_{\mathbf{com}}$, a eccezione della diagonale principale, contiene valori empirici, $\mathbf{R}_{\mathbf{repr}}$ contiene valori teorici, ricavati dal modello. Sono stati sviluppati molti metodi per identificare una matrice di saturazioni tale che:

$$\mathbf{A} \cdot \mathbf{A}' = \mathbf{R}_{\mathbf{repr}} \cong \mathbf{R}_{\mathbf{com}}$$

Due di questi metodi saranno tra poco illustrati. Per ottenere

$$\mathbf{A} \cdot \mathbf{A}' = \mathbf{R}_{\mathbf{u}}$$

si utilizza l'*ACP* (dove, ricordiamo, non ci sono fattori specifici e di errore). Ciò richiede una matrice $\mathbf{A}$ della stessa grandezza di $\mathbf{R}_{\mathbf{u}}$. Scopo invece dell'*AF* è quello di spiegare una matrice $\mathbf{R}$ di ordine M con una matrice $\mathbf{A}$ di ordine $M \cdot K$, con $K \ll M$.

Ricapitolando, le matrici di correlazione coinvolte nell'*AF* sono: $\mathbf{R}_{\mathbf{u}}$ (correlazioni osservate), $\mathbf{R}_{\mathbf{com}}$ (correlazioni osservate e comunalità nella diagonale principale), $\mathbf{R}_{\mathbf{repr}}$ (correlazioni riprodotte dal modello). Si può per completezza considerare $\mathbf{R}_{\mathbf{res}}$, la matrice dei residui, data dall'operazione $\mathbf{R}_{\mathbf{com}} - \mathbf{R}_{\mathbf{repr}}$; sebbene essa derivi da questa semplice operazione algebrica, l'abbiamo messa in risalto perché essa risulta importante per la valutazione del modello, o in alcuni metodi di estrazione è addirittura l'argomento della funzione obiettivo dell'algoritmo.

Considereremo tra breve due dei metodi di estrazione dei fattori oggi più diffusi. È però opportuno soffermarsi dapprima sul modo

correlate; quindi se il numero dei fattori deve essere uguale al rango si mancherebbe l'obiettivo principale dell'analisi. Ripetute prove empiriche convinsero Thurstone che la tecnica di sostituire le autocorrelazioni con le comunalità aveva l'effetto di ridurre il rango della matrice (Thurstone, 1947:484).

con cui l'*ACP* individua le componenti; in tal modo avremo ulteriori elementi di differenziazione delle due tecniche.

Per semplicità consideriamo due sole variabili osservate, X e Y, e assumiamo che la loro distribuzione sia normale bivariata. Graficamente otteniamo una figura simile a un copricapo di forma tondeggiante se la correlazione tra X e Y è assente (figura 4).

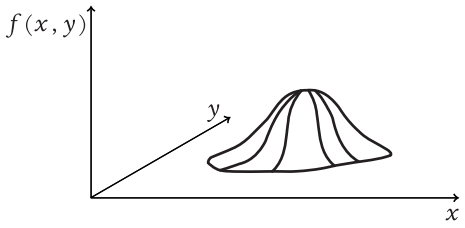

Figura 4.

Per rappresentare graficamente la presenza di una correlazione, si supponga di 'tirare' il cappello (elastico) lungo due direzioni opposte: più si tira (più la correlazione è alta), più il cappello assomiglierà alla parte centrale di un cappello da carabiniere (figura 5).

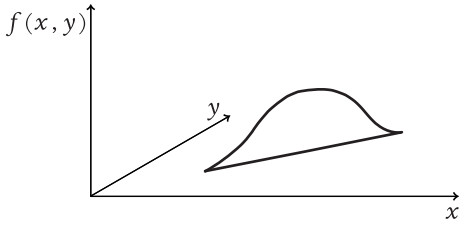

Figura 5.

Guardiamo ora la nostra relazione bivariata dall'alto, immaginando una correlazione positiva moderata. Possiamo usare come metodo di rappresentazione le curve di livello (figura 6: per semplificare considereremo ora le due corrispondenti variabili standardizzate).

Elevati valori di X tendono a associarsi prevalentemente a valori alti di Y (e viceversa): pertanto i punti si addensano nei quadranti I e

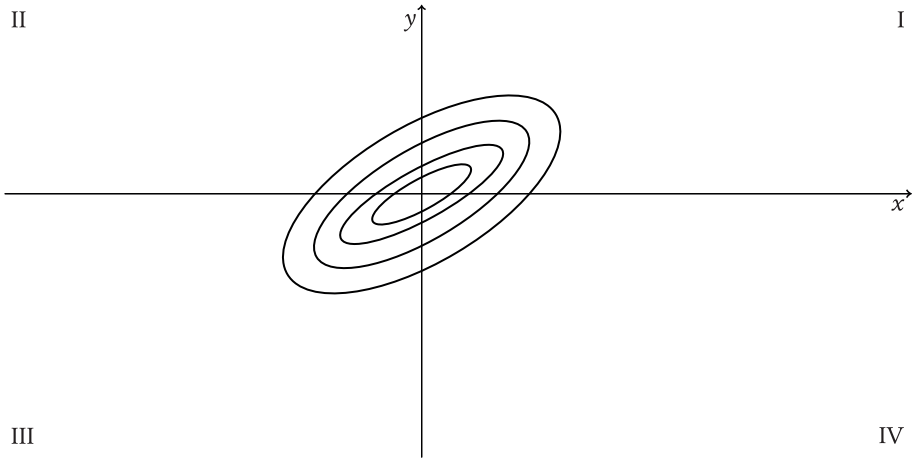

Figura 6.

III. Le **curve di livello** (isoipse[5]) formano delle ellissi concentriche (in assenza di relazione avremmo avuto dei cerchi concentrici come in figura 7). Queste ellissi sono attraversate longitudinalmente dall'asse P_1 (asse maggiore); un secondo asse, P_2 (asse minore), viene tracciato perpendicolarmente a P_1 (figura 8).

Supponiamo di voler rappresentare la posizione relativa di ogni punto in riferimento a uno solo dei due assi. La scelta cadrà naturalmente sull'asse P_1, perché esso è più vicino all'insieme dei punti delle ellissi di quanto non lo sia P_2 (e ogni altro possibile asse).

È importante sottolineare come non faccia differenza localizzare i punti rispetto agli assi P_1 e P_2 piuttosto che in riferimento agli assi X e Y; infatti anche se la rotazione muta le coordinate dei punti, non c'è alcuna perdita o aggiunta di informazione su questi ultimi. Se invece usiamo un solo asse per localizzare i punti, c'è perdita di informazione: possiamo localizzare ogni singolo punto solo approssimativamente; il margine di incertezza (o errore), è tanto più alto quanto meno stretta è la correlazione tra le variabili. Con l'asse principale, la perdita di informazione è minore. Nel caso (puramente ipotetico) di correlazione

5. L'isoipsa è una curva che congiunge i punti che stanno a una stessa altezza: ricordiamo che stiamo osservando una distribuzione in tre dimensioni.

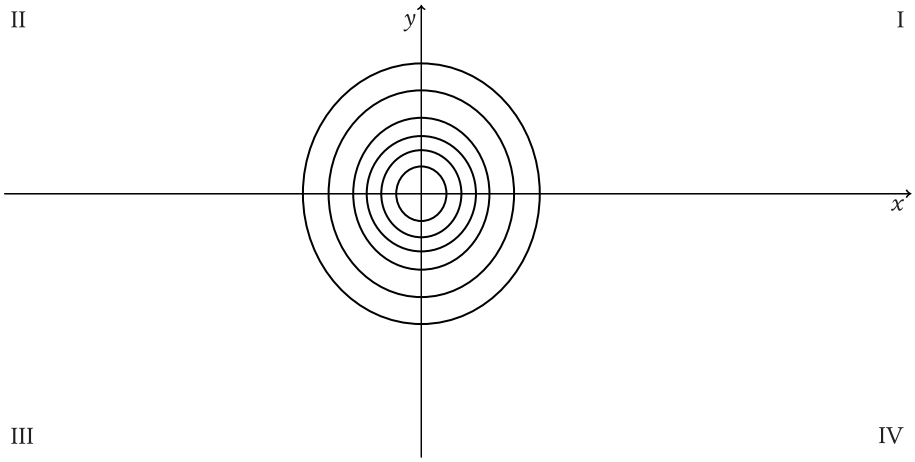

Figura 7.

perfetta, P_I conterrebbe tutta l'informazione utile per descrivere la posizione dei punti nel piano.

Ovviamente il discorso si può generalizzare a relazioni multivariate e, con riferimento agli assi principali, può essere esteso anche a relazioni bivariate e multivariate con distribuzioni non normali. In generale, la retta che individua l'asse principale è quella per la quale è minima la somma dei quadrati delle distanze dei punti dalla stessa tracciate perpendicolarmente (analogamente a quanto accade nella cosiddetta regressione ortogonale (cfr. Ricolfi, 2000:37).

Un altro modo di dire che la prima componente è quella che contiene il maggiore ammontare di informazione, è che essa spiega una quantità di varianza dei dati maggiore di quanto possa spiegare ogni altra componente. L'*ACP*, in effetti, procede proprio estraendo le componenti in ordine decrescente della quantità di varianza spiegata da ognuna.

Compito dell'*ACP* è quello di riparametrizzare le variabili in modo da ricavare una gerarchia di componenti basata sulla loro capacità di riassumere informazione.

Lo strumento fondamentale per arrivare a tale scomposizione e ge-

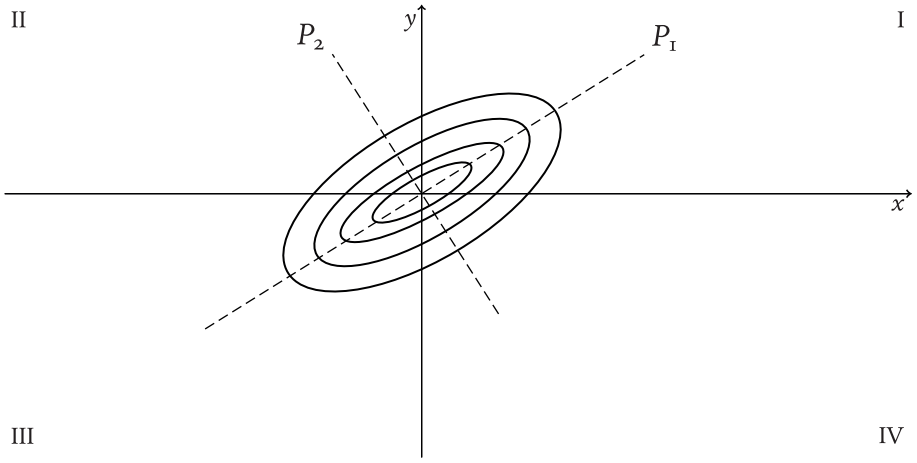

Figura 8.

rarchizzazione dell'informazione è dato dall'equazione caratteristica, che in forma matriciale sintetica si esprime così:

$$\mathbf{R} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$$

dove $\mathbf{v}$ e λ sono rispettivamente un **autovettore** e un **autovalore** associati alla matrice $\mathbf{R}$ (si veda l'appendice II). L'autovalore rappresenta la quantità di varianza spiegata da una componente. Una matrice di correlazioni empiriche ha di solito più autovalori non nulli: nell'*ACP* questi di norma sono tanti quanti le variabili osservate (M) (a meno che non vi siano casi di collinearità che però vanno evitati come si è già detto); gli M autovalori hanno M autovettori associati. La formula dell'equazione caratteristica può perciò essere scritta come segue:

$$\mathbf{R}(\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_M) = (\lambda_1\mathbf{v}_1, \lambda_2\mathbf{v}_2, \ldots \lambda_M\mathbf{v}_M)$$

o meglio

$$\mathbf{R} \cdot \mathbf{V} = \mathbf{V} \cdot \Lambda$$

dove λ è una matrice diagonale con valori $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_M$, posti in ordine decrescente.

Gli elementi di ogni autovettore, dopo che questo sia stato 'normalizzato' (cioè dopo che ogni suo elemento sia stato diviso per la lunghezza dell'autovettore, e poi moltiplicato per la radice dell'autovalore associato), costituiscono le saturazioni che soddisfano la seguente proprietà:

$$max\ h_i^2 = a_{1i}^2 + a_{2i}^2 + \cdots + a_{Mi}^2$$

Occorre in altri termini massimizzare la varianza spiegata dal fattore i–esimo, sotto la seguente condizione:

$$r_{jk} = \sum_{i=1}^{M} a_{ji} \cdot a_{ki}$$

Ovviamente, quando si opera su matrici di correlazioni il valore massimo di a_i è pari a 1.

Nell'*ACP* Troncata si individuano le prime K componenti ($K \ll M$) che insieme spiegano una buona percentuale di varianza, ignorando quelle che ci fanno perdere poca informazione sui punti.

Se l'*ACP* è orientata a spiegare la varianza, l'*AF* è invece mirata a dare conto delle correlazioni (o delle covarianze) tra le variabili manifeste come fenomeno interpretabile con la presenza di fattori sottostanti. In pratica, come abbiamo già visto ciò equivale a regredire le variabili osservate su uno o più fattori ipotetici, non direttamente osservati, su cui si fanno alcuni assunti.

3.5. Scelta del numero di fattori e loro estrazione

Il criterio di determinazione del numero dei fattori da estrarre va individuato prima di iniziare le stime / calcoli dei parametri. I criteri più adottati sono i seguenti:

— il numero dei fattori può essere previsto dal ricercatore sulla base di un ragionamento teorico–deduttivo, o sulla base di altre

pre-cognizioni (in questo caso però è da valutare il passaggio all'approccio confermativo);
— si può iniziare stimando la soluzione per $K = 1$ e poi procedere a successive elaborazioni aumentando K di una unità per volta, sino a quando si trova una soluzione soddisfacente;
— si può adottare uno dei criteri basati sul rendimento marginale di ogni componente estratta: molti manuali (per es. Comrey, Lee, 1992, trad. 1995:146 e ss.) propongono il criterio di Kaiser–Guttman dell'*autovalore* >1 (criterio di *default* in alcuni software)[6], oppure lo *Scree test* di Cattell[7], che opera sul diagramma di caduta degli autovalori.

Fissato il numero dei fattori comuni, l'algoritmo prescelto calcola — o stima — i valori dei loading.

Il primo metodo di estrazione di Fattori che prendiamo in considerazione è il **Metodo dei Minimi Quadrati Ordinari**, o 'non ponderati', d'ora in avanti **ULS** (*Unweighted Least Squares*). La logica di tale metodo è quella di determinare il valore dei parametri in modo da minimizzare i residui (contenuti in $\mathbf{R_{res}}$) derivanti dalla differenza tra la matrice di correlazioni originaria e quella riprodotta.

L'algoritmo calcola i parametri in modo che questi permettano la ricostruzione della matrice originaria nel modo più congruente possibile per quanto riguarda gli elementi extra–diagonali (per un numero di fattori che rispetti il criterio scelto in precedenza: è chiaro che maggiore è il numero di fattori estratti e migliore è l'adattamento)[8].

6. L'autovalore rappresenta la quota di varianza riprodotta dal fattore (o, nella *ACP*, dalla Componente); ogni fattore in più estratto spiega una quota progressivamente decrescente di varianza; la regola proposta da Kaiser è giustificata dal fatto che autovalori pari o inferiori a uno portano a estrarre fattori che riproducono meno varianza di una qualunque variabile manifesta (si ricorda che le variabili manifeste sono standardizzate e perciò hanno varianza pari a uno).

7. Illustriamo lo *scree test* nell'Appendice I.

8. Nella diagonale troviamo le comunalità di ogni variabile; a differenza di altre procedure di estrazione dei fattori, qui le comunalità sono un *by–product* dell'algoritmo e non un punto di inizio. Ciò da un lato costituisce un vantaggio, poiché evita il problema, controverso, della corretta stima a priori delle comunalità; d'altro canto può portare a soluzioni improprie, definite in letteratura come *Heywood Case*, caratterizzate da valori di

Il nucleo della soluzione consiste nella minimizzazione della seguente funzione:

$$F_{ULS} = 1/2[Tr(\mathbf{R_u} - \mathbf{R_{repr}})^2]$$

dove $\mathbf{R_{repr}}$ è la matrice di correlazioni riprodotte, contenente degli uno in diagonale principale, proprio come $\mathbf{R_u}$. In altre parole, la funzione da minimizzare è data dalla somma dei quadrati delle differenze tra correlazioni osservate e correlazioni riprodotte mediante le saturazioni[9].

Veniamo ora brevemente al metodo di stima della Massima Verosimiglianza, o *ML* (*Maximum Likelihood*). L'applicazione di questo metodo è particolarmente indicata quando:

a) si è in possesso di un campione probabilistico;
b) si può assumere che le variabili siano distribuite nella popolazione in modo normale multivariato.

3.6. Metodi di rotazione ortogonale e obliqua

Per comprendere il problema della rotazione dei fattori è utile immaginare uno spazio cartesiano che abbia come *assi di riferimento* i fattori estratti. Per semplicità, rappresentiamo graficamente solo due fattori latenti (F_1 e F_2) e due variabili manifeste (V_A e V_B). Sul piano cartesiano identificato dai due fattori (assi) possiamo rappresentare geometricamente ognuna delle due variabili come un vettore individuato da una coppia di coordinate che corrispondono alle saturazioni delle variabili sui fattori. Le saturazioni delle variabili sui fattori rappresentano il sistema di *coordinate (ascissa e ordinata)*.

comunalità maggiori di uno e dunque assurdi.

9. Un altro metodo della famiglia dei minimi quadrati, molto utilizzato, è quello dei **Minimi Quadrati Generalizzati**, o **GLS** (*Generalized Least Squares*). Il criterio di minimizzazione è lo stesso, ma le correlazioni sono pesate inversamente per l'unicità delle variabili: in tal modo si assegna maggior peso alle variabili con elevate comunalità (McDonald 1985), cioè si dà più peso alle stime più stabili.

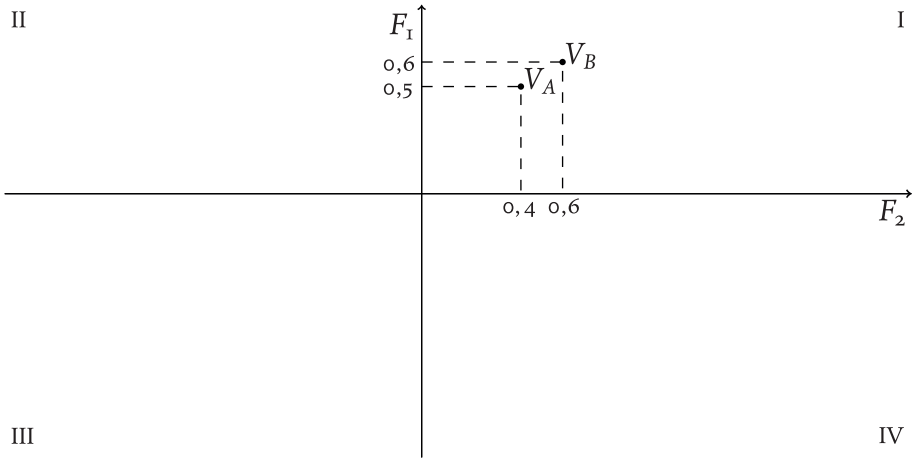

Figura 9.

Nel nostro esempio (figura 9), la variabile V_A ha saturazione $a_{1A} = 0,5$ sul fattore F_1 e $a_{2A} = 0,4$ sul fattore F_2.

La variabile V_B ha saturazione 0,6 su entrambi i fattori. Ruotando gli assi di riferimento è possibile cambiare il sistema di coordinate dei punti–vettore, lasciando inalterata la posizione relativa (correlazione) di questi ultimi. La somma dei quadrati delle saturazioni per ogni variabile resta invariata. Ruotando per esempio di un angolo $\theta = 30°$ entrambi gli assi (rotazione ortogonale), otteniamo le nuove coordinate dei punti–vettore (e dunque le nuove saturazioni) in un sistema di riferimento nuovo (con assi fattoriali F_1' e F_2') che rispetto al precedente conserva la perpendicolarità tra i fattori.

— La variabile V_A ha ora saturazione $a_{1A}' = 0,63$ sul fattore F_1' e $a_{2A}' = 0,1$ sul fattore F_2'.
— La variabile V_B ha saturazione $a_{1B}' = 0,82$ sul fattore F_1' e $a_{2B}' = 0,22$ sul fattore F_2'.

I valori delle saturazioni ruotate sono stati ottenuti applicando le seguenti formule, ricavate con una serie di passaggi di trigonometria:

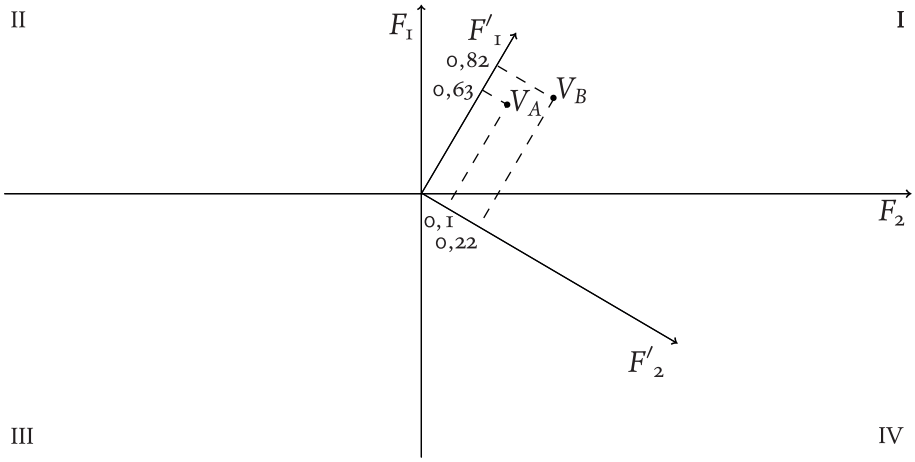

Figura 10.

$$a'_{1A} = a_{1A}\cos\theta + a_{2A}\sin\theta \quad e \quad a'_{2A} = a_{2A}\cos\theta - a_{1A}\sin\theta$$

$$a'_{1b} = a_{1b}\cos\theta + a_{2b}\sin\theta \quad e \quad a'_{2b} = a_{2b}\cos\theta - a_{1B}\sin\theta$$

Pertanto, è evidente che esiste un numero praticamente infinito di rotazioni possibili già solo per due fattori; il numero di soluzioni cresce ulteriormente all'aumentare del numero dei fattori.

Inoltre, si ricorderà che sino ad ora abbiamo vincolato, per semplicità, i fattori a essere ortogonali tra loro. Ma una volta estratti i fattori, ottenute la matrice $\mathbf{R}_{\mathbf{repr}}$ e le comunalità, non è più necessario vincolare i fattori all'ortogonalità, se ciò non appare utile a fini interpretativi.

Occorrono dunque dei criteri per scegliere una rotazione tra le tante possibili.

Un primo approccio consiste nella scelta della soluzione che a un esame grafico soddisfa maggiormente il ricercatore. La scelta può essere attuata sulla base di una valutazione sintetica *by intuition*, oppure basata su qualche criterio analitico di interpretazione, ma in ogni caso

è ampiamente guidata dalla capacità soggettiva di riconoscimento di pattern (addensamenti di punti–vettore). Le cosiddette 'rotazioni a mano', richiedono da parte del ricercatore una certa esperienza, soprattutto quando ci sono molti fattori da esaminare e le variabili non formano cluster facilmente distinguibili.

Una seconda strada, la più praticata, è quella di affidarsi ai criteri analitico–matematici di rotazione implementati nei metodi di rotazione presenti nei package statistici. Si tratta di regole formalizzate, che non richiedono valutazioni del ricercatore nell'applicazione al caso singolo[10].

I metodi analitici che andiamo ora a considerare si basano in gran parte su un principio generale, proposto da Thurstone nel 1947: il **principio della struttura semplice**. Esso non fa altro che riprendere un canone più generale della spiegazione scientifica, quello della **parsimonia**. Nella rotazione fattoriale, per raggiungere la struttura semplice si devono rispettare alcune regole; ci limitiamo a sintetizzarle in due principi (per una maggior precisione si veda Harman, 1976[3]:98):

— nella matrice fattoriale ruotata, ogni variabile deve avere almeno un *loading* nullo, ma possibilmente anche di più;
— ogni fattore deve avere almeno *K loading* nulli (*K*: numero dei fattori comuni) e questi devono essere diversi tra i vari fattori.

Secondo Cattell (1978, cit. in Kline, 1994, trad. it. 1997:69), i fattori a struttura semplice oltre alla facilità di interpretazione godono delle seguenti proprietà:

— sono più facilmente replicabili; la replicabilità della struttura semplice si estende anche ai casi in cui le variabili rilevate non siano del tutto identiche a quelle rilevate nella ricerca precedente;
— negli studi su matrici artificiali, in cui i fattori sottostanti, per così dire 'genotipici', sono noti, è stato dimostrato che le rota-

10. Il fatto che si tratti di regole a–priori non significa che siano 'migliori' di quelle che richiedono il giudizio del ricercatore: per una critica serrata dei metodi analitici si veda Comrey, Lee, 1992, trad. it. 1995:228 e ss.

> zioni verso una struttura semplice producono fattori 'fenotipici' che approssimano quelli genotipici[11].

Prima di considerare alcuni dei metodi di rotazione analitici più diffusi nei *package* statistici, riprendiamo la distinzione, già accennata, tra rotazioni ortogonali e rotazioni oblique.

Se le prime sono caratterizzate dal vincolo della perpendicolarità degli assi, nel caso delle rotazioni oblique invece si permette una certa libertà, di grado più o meno ampio, nello stabilire l'ampiezza dell'angolo formato dalle coppie di fattori: detto in altro modo, si permette ai fattori stessi di essere tra loro correlati. Ci si può chiedere il perché di questa ulteriore complicazione. Come osservava Cattell per la ricerca psicologica, e estendendo tale osservazione alla ricerca sociale, è quanto mai improbabile pensare che i concetti misurati in questi campi di indagine siano totalmente distinti e incorrelati. D'altro canto, se i fattori sono effettivamente incorrelati ciò emergerà anche con una rotazione obliqua[12].

Ricordiamo che se la rotazione è obliqua, otteniamo due matrici distinte di parametri: la *pattern matrix* e la *structure matrix*. La scelta di ricorrere alla seconda piuttosto che alla *pattern matrix* per interpretare il significato sostantivo dei fattori estratti è piuttosto soggettiva. Se i fattori sono poco correlati, le due matrici non sono molto diverse; se sono molto correlati, la *structure matrix* contiene in genere valori più elevati e ciò rende difficile distinguere i diversi fattori tra loro. In pratica, conviene ricorrere alla *structure matrix* in sede di etichettamento dei fattori quando già dall'esame della *pattern matrix* emergono dubbi sulla effettiva distinguibilità di due fattori: se le variabili manifeste risultano correlate allo stesso modo con entrambi i fattori, si può rafforzare l'ipotesi che il fattore sostantivo in realtà sia uno solo.

11. Può essere utile talvolta quantificare la semplicità fattoriale di una variabile o di un fattore, o della matrice dei loading nel suo complesso; alcune semplici misure sono riportate in Barbaranelli, 2003:151–2.

12. Ciò non significa che la rotazione obliqua sia sempre preferibile. Spesso nell'analisi esplorativa si forzano i fattori all'ortogonalità per renderli chiaramente distinguibili e per decidere quanti sono in effetti dotati di un significato sostantivo autonomo. In un passo successivo, una volta chiarito il numero di fattori da estrarre, sarà opportuno, a meno di ragioni specifiche, ricorrere a una rotazione obliqua.

La *structure matrix* infine può essere usata, dopo che l'operazione di etichettamento è già avvenuta, per individuare variabili che, pur non risultano validi indicatori di un fattore latente, siano comunque delle **proxy** di questo[13].

Nel caso di rotazioni ortogonali, per costruzione i fattori sono incorrelati; la *pattern* e la *structure matrix* coincidono e contengono le correlazioni tra variabili manifeste e fattori latenti comuni.

Varimax

Varimax massimizza la somma delle varianze delle saturazioni al quadrato calcolate in ogni colonna della pattern matrix (diciamo: max V). Ciò ha come effetto, in linea di principio, quello di ottenere che parte delle saturazioni di ogni colonna siano molto prossime a 1, parte molto prossime a zero e poche saturazioni di grandezza intermedia; in tal modo i fattori tendono a essere molto distinti tra loro (cosicché l'operazione di etichettamento dovrebbe essere agevolata).

Non sempre è possibile ottenere una struttura semplice mantenendo l'ortogonalità dei fattori; se però ciò è possibile, allora Varimax è la procedura più efficace.

Quartimax

Tende a rendere massima la somma delle varianze delle saturazioni al quadrato di ogni riga (diciamo max Q): ciò avviene quando una variabile è caricata su un solo fattore. Essa enfatizza pertanto la semplicità di interpretazione delle variabili, al contrario di Varimax che enfatizza la semplicità di interpretazione dei fattori. Spesso dà origine a un fattore generale, avente saturazioni medio–elevate su tutto il set

13. Una variabile proxy rinvia anch'essa a qualcosa di celato: vedo del fumo perché c'è del fuoco (che solo momentaneamente è nascosto alla mia vista). La proxy quindi sta in un rapporto di 'segnalazione' con il concetto da misurare, non di 'indicazione'. Un esempio per la ricerca sociale è l'uso che fa Inglehart (1990, trad. it 1993:46) della variabile 'grado di soddisfazione per la propria vita' per rilevare il 'civismo' degli intervistati; in questo caso è molto chiaro che l'Autore non sfrutta un rapporto di indicazione semantica, bensì una mera regolarità empirica verificata in precedenti ricerche.

di variabili. Per tale motivo non viene usato, a meno che non ci siano buone ragioni per ipotizzare l'esistenza di un fattore generale.

Equamax (o Equimax o Orthomax)

Se Quartimax semplifica l'interpretabilità delle righe (variabili) della Pattern Matrix, e Varimax quella delle colonne (fattori), Equamax applica entrambi i criteri con un peso appropriato (diciamo: max $\alpha Q + \beta V$).

Direct Oblimin

Le soluzioni analitiche delle rotazioni oblique alla struttura semplice sono numerose (Gorsuch nel 1983 ne illustrava 19, considerandole solo un campione!). Rotazioni oblique analoghe a Varimax sono Covarimin e Biquartimin; analoga a Quartimax è Quartimin. Direct Oblimin è analoga a Equamax.

Il grado di massima correlazione tra i fattori permessa è governato da un parametro δ; di solito è fissato a un valore tale da non permettere una correlazione troppo elevata tra i fattori.

Promax

Questo metodo inizia con una soluzione ortogonale, quale potrebbe essere Varimax. Le saturazioni ottenute vengono poi elevate a potenza: al crescere dell'esponente le grandezze delle saturazioni diminuiranno e tale diminuzione sarà tanto più rapida quanto più piccoli sono i valori di partenza. La prima soluzione ortogonale è poi ruotata con un metodo obliquo in modo tale da approssimare al meglio la matrice delle saturazioni elevate a potenza[14]. I fattori risulteranno tanto più correlati tra loro, quanto più alte sono le potenze a cui sono elevate le saturazioni iniziali.

14. La rotazione di una matrice di loading verso una matrice obiettivo, seguendo alcune regole poste a priori (analogamente a quanto avviene nella rotazione verso una struttura semplice) è detta soluzione di *Procruste*. In questo caso la matrice obiettivo è quella dei loading della prima soluzione elevati a potenza.

3.7. La stima dei punteggi fattoriali

Una volta estratto un adeguato numero di fattori e fornita una loro interpretazione plausibile, l'*AF* potrebbe ritenersi conclusa.

A questo punto però il ricercatore può procedere oltre con un passo supplementare, rappresentato dalla stima dei punteggi fattoriali. L'utilità di tali punteggi consiste nel fatto di sostituire dei profili di risposta composti da un numero di stati su variabili manifeste solitamente elevato con dei profili composti da pochi stati su variabili latenti. Inoltre, attraverso una rappresentazione grafica dei punteggi individuali sui piani individuati da coppie di fattori, è possibile riconoscere profili di risposta anomali (*outliers*).

A differenza dell'*ACP*, dove i punteggi individuali vengono *calcolati* esattamente perché manca la componente aleatoria, nel modello di *AF* essi possono essere soltanto stimati.

Alcuni autori ritengono adeguati i coefficienti contenuti nella Pattern Matrix, cioè i *regression loading*, per la stima dei *factor score* (Kline, 1994, trad. it. 1997). In realtà, come abbiamo già avuto modo di osservare nel capitolo I, le saturazioni non rappresentano il contributo delle variabili manifeste alla individuazione dei fattori, ma la dipendenza delle prime dai secondi (per una critica più dettagliata si veda Marradi, 1980). È più corretto dunque calcolare dei 'pesi' appositi, detti **factor score coefficients**.

Esistono vari metodi per stimare i punteggi fattoriali; uno dei più diffusi è il metodo della regressione, di seguito descritto. Il punteggio individuale sul fattore comune può essere stimato come combinazione lineare delle variabili originarie ricorrendo alla seguente equazione:

$$F_{kn} = \pi_{1k}Z_{1n} + \pi_{2k}\,Z_{2n} + \cdots + \pi_{ik}Z_{in} + \cdots + \pi_{Mk}Z_{Mn}$$

dove:

- F_{kn} è il punteggio standardizzato del fattore k–esimo per l'individuo n–esimo
- π_{ik} è il coefficiente di regressione standardizzato della variabile i–esima sul fattore k–esimo

— Z_{in} è il punteggio standardizzato della variabile *i*–esima per l'individuo *n*–esimo.

L'equazione differisce da quella di una normale regressione multipla, in quanto non solo i pesi fattoriali π sono ignoti, ma anche la variabile dipendente *F*.

Quindi il problema diventa quello di stimare i pesi π, senza conoscere i punteggi fattoriali, perché anzi questi andranno calcolati successivamente, una volta che i π saranno noti.

Fortunatamente, per ottenere i pesi π è irrilevante conoscere i punteggi fattoriali; si può sfruttare invece l'informazione contenuta nelle correlazioni tra le variabili manifeste (ricavabili dalla matrice di input) e nelle correlazioni tra le variabili manifeste e i fattori comuni (ricavabili dalla *structure matrix* o, se la rotazione è ortogonale, dalla *pattern matrix*).

L'equazione in forma matriciale che lega questi tre tipi di dati è la seguente (per semplicità consideriamo un singolo fattore latente comune, *F*):

$$\mathbf{R} \cdot \pi = r_{iF}$$

dove **R** è la matrice delle correlazioni tra le variabili manifeste, π è il vettore colonna che contiene i pesi fattoriali da stimare, e r_{iF} è un vettore colonna delle correlazioni tra le variabili e il fattore comune F. In altri termini, ognuna delle correlazioni tra le singole variabili manifeste e il fattore comune è data dalla somma ponderata di tutte le correlazioni tra la *i*–esima variabile manifesta e tutte le altre (compresa l'autocorrelazione).

Per trovare i pesi fattoriali dobbiamo perciò risolvere la seguente equazione:

$$\pi = \mathbf{R}^{-1} \cdot r_{iF}$$

ammesso che **R** sia invertibile.

Capitolo IV

La valutazione della soluzione

Nella valutazione del risultato complessivo di un'applicazione dell'*AF* Esplorativa occorre tenere distinti due piani, consistenti nel:

— valutare l'adeguatezza del modello specificato rispetto ai dati nel campione considerato (vedi par. 4.1);
— valutare la generalizzabilità del modello, ossia la sua applicabilità a casi diversi da quelli esaminati (par. 4.2).

La ricerca di misure complessive non esaurisce la problematica della valutazione: è anche necessario considerare criteri di valutazione locali, soprattutto riferiti a ogni singolo fattore estratto (par. 4.3).

4.1. Adeguatezza del modello ai dati

Sono quattro i criteri principali per valutare globalmente il successo di un'*AF* da un punto di vista descrittivo (Ricolfi, 1987; 2002): della varianza spiegata, della parsimonia, dell'adattamento, del rendimento.

Per **varianza spiegata dai fattori** s'intende la capacità dei fattori estratti di rendere conto della variabilità presente nell'insieme delle variabili manifeste. Naturalmente al crescere della dimensionalità del modello fattoriale, ossia all'aumentare del numero di fattori latenti, aumenta la quota di varianza spiegata. In altre parole, all'aumentare dei *benefici* (in termini di comunalità totale) richiesti al modello teorico aumentano anche i *costi* da affrontare, soprattutto in termini di un maggior numero di fattori estratti.

In sede di valutazione dell'esito di un'*AF* il criterio della varianza spiegata non può mai essere disgiunto da quello della parsimonia

della soluzione. Per esempio: un modello che riproduce il 60% della varianza totale di un insieme di otto item, è un buon risultato se il numero di fattori utilizzati è pari a 2, scadente se i fattori sono quattro.

Un semplice indice assoluto di parsimonia, P, mette a confronto la soluzione scelta con la soluzione meno parsimoniosa possibile, o detto in altri termini, caratterizzato da maggiore **complessità**:

$$P = K_{max} - K$$

dove K è il numero dei fattori latenti comuni.

Se vogliamo confrontare soluzioni in cui il valore massimo di K cambia, dobbiamo relativizzare l'indice:

$$\mathbf{P}_{rel} = \mathbf{P}/\mathbf{P}_{max}$$

Il criterio della varianza spiegata e quello della parsimonia si possono combinare in un unico criterio di **rendimento** che tenga conto al contempo dei costi e dei benefici dell'analisi. Il rendimento globale di una soluzione fattoriale si calcola nel seguente modo:

$$rend = \frac{\lambda_1 + \lambda_2 + \ldots \lambda_K}{K}$$

dove λ sono gli autovalori corrispondenti ai fattori estratti e K è il numero dei fattori estratti

Sulla scorta di diverse ricerche empiriche, si può adottare come regola del pollice che quest'indice non sia inferiore a 2 (Ricolfi, 1987:97): questo risultato indica che i benefici ottenuti sono doppi rispetto ai costi sostenuti. Un tale rendimento è auspicabile indipendentemente dalla percentuale di varianza spiegata e dal numero di fattori estratti.

L'idea di un indice di **adattamento** per la valutazione del modello si può trovare già in Thurstone (1947), anche se è soprattutto grazie ai lavori di Harman e Jones (1966) della metà degli anni sessanta che questo criterio di valutazione si è affermato. Con il termine adattamento si intende la capacità dei fattori estratti di riprodurre le correlazioni tra le variabili originarie.

Secondo Thurstone, dopo aver estratto il primo fattore, si deve verificare la significatività della matrice dei residui: dunque, per i

residui diversi da zero, valutare la probabilità che tale differenza sia casuale. Se tale matrice contiene residui significativamente diversi da zero, occorre estrarre un secondo fattore, e così via (in realtà per Thurstone questo era un criterio di scelta del numero dei fattori, e non un criterio di valutazione del *fit* del modello).

Presa alla lettera, questa indicazione ha però lo svantaggio di portare a una rappresentazione il più delle volte poco parsimoniosa. Ci chiediamo allora: per giudicare soddisfacente una soluzione di un'*AF*, tutti i residui devono essere non significativamente diversi da zero o la maggior parte di essi? E nel caso la risposta fosse la seconda: in che percentuale? Occorre dire che non esistono a tal proposito risposte precise. Una regola di buon senso è quella di lasciare al ricercatore la scelta di tale percentuale, possibilmente fatta prima di iniziare l'*AF*.

Presentiamo due modi di formalizzare il criterio dell'adattamento: un primo indice di valutazione, che chiamiamo **repr**, può essere ricavato dalla matrice $\mathbf{R_{res}}$: si conta il numero di residui maggiori di |0,05| e se ne calcola la percentuale sul totale dei residui. In linea di massima, una percentuale di residui minore del 10% è considerata soddisfacente; con numerosità campionarie basse (diciamo al di sotto di 1000 casi) si possono considerare soddisfacenti anche valori un po' più elevati.

Un altro indice di adattamento si può utilizzare se viene applicato ULS, ricorrendo allo stesso valore della funzione di adattamento usata come criterio di estrazione dei fattori. Poiché il valore di tale funzione dipende dal numero di variabili osservate, il suo valore minimo va relativizzato.

Per rendere la misura indipendente dall'ordine della matrice di correlazione (pari al numero delle variabili) si utilizza come criterio l'indice **RMR**-***Root Mean square Residuals*** (Jöreskog, Sörbom, 1988:43), che ha la seguente formula:

$$RMR = \left[\frac{1}{M(M-1)/2} \cdot \sum d_{mj}{}^2 \right]^{1/2}$$

dove d indica lo scarto tra correlazioni effettive e riprodotte, *M* il

numero delle variabili[1]. Come regola del pollice si può ritenere soddisfacente un $RMR \leq 0,05$ tenendo conto che si tratta di una regola molto lasca.

In un'*AF*, a differenza di quanto avviene nell'*ACP*, il primato non dovrebbe andare alla coppia parsimonia e varianza, bensì alla coppia parsimonia e adattamento. Ciò è abbastanza facile da comprendere, se si ricorda che l'*AF* mira principalmente a spiegare un'ampia serie di correlazioni con pochi fattori. Ciò non significa che la varianza spiegata sia irrilevante, ma solo che è in genere secondaria rispetto all'adattamento.

4.2. Generalizzabilità e replicabilità del modello

Un valido modello Fattoriale dovrebbe auspicabilmente avere la caratteristica di essere applicabile anche ad altri casi, distinti da quelli costituenti il campione su cui si è condotta l'analisi. Possiamo distinguere due aspetti:

a) quello della generalizzazione mediante la statistica inferenziale
b) quello della replicabilità dei fattori.

a) Quando si utilizza come metodo di estrazione *ML* (o anche *GLS*), nell'output viene fornito anche un indice di **bontà di adattamento del modello** ai dati, con il quale si valuta se il numero dei fattori estratti è sufficiente a rendere conto delle relazioni tra variabili manifeste. Per campioni probabilistici estratti da popolazioni normali multivariate, la statistica–test tende, al crescere della numerosità campionaria, a distribuirsi secondo la **variabile aleatoria chi quadrato** (cfr. Albano, Testa, 2002:181–2).

1. Nella formula originale il numeratore della formula è $M(M+1)/2$, perché i due autori considerano come input la matrice varianze covarianze, in cui la stessa diagonale principale è informativa. Anche con una matrice di correlazioni può andare bene la formula originaria: specificamente nel caso che il metodo di stima usato sia ML o P *AF* o GLS; in questi casi, i dati non ridondanti includono anche quelli in diagonale principale: le stime delle comunalità iniziali nella matrice di input e delle comunalità finali nella matrice riprodotta dal modello.

Occorre sottolineare una importante differenza circa l'interpretazione di questa statistica per la valutazione della bontà di adattamento del modello rispetto a quella, probabilmente familiare a molti tra i lettori, adottata nel test di associazione tra variabili categoriali (tavole di contingenza).

In questo secondo caso, lo ricordiamo, i valori alti della statistica–test permettono di respingere l'ipotesi nulla che afferma l'assenza di relazione tra le variabili (Albano, Testa, 2002:234 e ss.). Nel caso in esame invece, sono desiderabili valori bassi della statistica–test, che ci permettano di *non* respingere l'ipotesi nulla, per la quale i dati riprodotti dal modello non sono significativamente diversi da quelli originari (e che pertanto il numero di fattori estratti è sufficiente a interpretare la matrice di correlazioni).

Supponiamo di avere sottoposto all'*AF* una batteria di 20 item e di essere incerti sul numero di fattori da estrarre. Supponiamo inoltre che ci siano le condizioni per applicare *ML* e che i risultati di due prove, rispettivamente a 6 e a 7 fattori, siano i seguenti:

n. di fattori estratti	statistica–test	gradi di libertà*	significatività
6	121,87	85	,00
7	71,18	71	,26

* Il numero dei gradi di libertà è dato sottraendo il numero di parametri liberi dal numero di vincoli, e precisamente da: N° correlazioni tra v. manifeste – loading da stimare + N° coppie distinte di fattori latenti comuni (questi ultimi sono vincoli, perché le correlazioni tra fattori sono fissate a zero).

Tabella 1.

Usando il criterio del chi quadrato nell'esempio precedente, occorrono almeno 7 fattori per rappresentare adeguatamente le relazioni tra le 20 variabili.

Nell'*AF* Esplorativa si tende maggiormente a utilizzare gli indici descrittivi, anche in considerazione del fatto che le condizioni necessarie per l'applicazione dell'inferenza raramente sono soddisfatte. Inoltre, occorre sempre tenere conto che il chi quadrato è una statistica il cui valore dipende direttamente dalla grandezza del campione: pertanto, con campioni molto grandi, questo criterio porta spesso a una sovrastima del numero di fattori da estrarre.

b) È importante nell'applicazione dei modelli di *AF*, come in altri ambiti dell'analisi multivariata, evitare soluzioni ad hoc, valide solo a descrivere un campione di casi. Questo è ancora più richiesto se siamo in un ambito di *AF* Esplorativa, in cui è relativamente semplice ottenere una soluzione e, con un po' di immaginazione, trovare interpretazioni convincenti dei fattori latenti comuni.

Una volta validato su un dato campione un modello di cui siamo convinti, occorre procedere appena possibile a una sua riapplicazione su altri casi, preferibilmente con una nuova ricerca sul campo o anche, eventualmente, cercando fonti per l'analisi secondaria che presentino un set simile di variabili manifeste (il che non è semplice). Quando ciò è stato fatto, si potrà verificare la congruenza tra le soluzioni ottenute nei diversi campioni: in pratica, si tratterà di valutare se i pattern dei loading dei fattori latenti si assomigliano o se differiscono nei diversi campioni. Un **indice di congruenza fattoriale** è quello proposto da Tucker (1951); consideriamo, per semplicità e senza perdita di generalità, due soli campioni, A e B, su cui abbiamo misurato un fattore latente comune:

$$\phi_{{}^{A}a{}^{B}a} = \frac{\sum_j {}^{A}a_j {}^{B}a_j}{\sqrt{\sum_j {}^{A}a_j{}^{2} \sum_j {}^{B}a_j{}^{2}}}$$

dove ${}^{A}a_j$ e ${}^{B}a_j$ sono i vettori dei loading nei due campioni.

Solitamente si ritengono stabili i fattori se l'indice supera la soglia di 0,90.

Il phi di Tucker può essere impiegato anche per valutare la ripetibilità della soluzione ricavata da un campione all'interno di partizioni di quest'ultimo: in questo caso più che la replicabilità valutiamo la consistenza interna della soluzione adottata.

4.3. Valutazione della rilevanza sostantiva dei fattori

Accanto ai criteri di valutazione statistica in senso proprio, vengono comunemente considerate anche delle soglie convenzionali, o standard, con cui confrontare i legami tra fattori latenti e variabili

manifeste. Questi standard hanno lo scopo di aiutare il ricercatore a decidere se l'applicazione della tecnica alle variabili in esame ha prodotto dei fattori semanticamente ben individuabili o viceversa delle entità assolutamente effimere e dotato di scarso significato sostanziale.

Poiché l'identificazione di un fattore e la sua interpretazione dipendono dalle variabili manifeste a cui esso è legato, è logico che tali standard valutativi riguardino innanzitutto proprio la forza e il numero di legami significativi tra variabili osservate e fattore latente.

Innanzitutto occorre decidere quando un legame tra fattore comune e variabile manifesta è da ritenersi saliente. Si tenga conto che una correlazione pari a 0,3 indica che il 9% della varianza della seconda è riprodotta dal primo. Questo valore è considerato generalmente come la soglia minima di salienza di una correlazione tra variabile manifesta e fattore latente. Pertanto, le variabili manifeste da utilizzare per identificare il fattore dovrebbero avere una correlazione con quest'ultimo uguale o superiore a tale soglia. Questo limite inferiore non è però da interpretare in modo troppo rigido: si possono accettare anche valori leggermente inferiori se il modello è parsimonioso. Ma nella scelta di tale soglia intervengono anche valutazioni di carattere disciplinare. Questo livello infatti può anche essere abbassato in una ricerca dove abbiamo a che fare con costrutti molto sfaccettati (perché irriducibilmente complessi, come capita spesso in sociologia, o perché ancora poco raffinati da un punto di vista concettuale).[2]

Comrey e Lee (1992, trad. it. 1995:317) propongono il seguente standard per decidere dell'utilità potenziale di una variabile manifesta nell'interpretazione del fattore (tabella 2).

Fattori con correlazioni solo sufficienti o scarse richiedono cautela nell'interpretazione. Con molte valutazioni buone si può essere un po' più definitivi nel dare un nome al fattore[3].

2. In psicometria, il livello minimo accettato dei *loading* è di solito più elevato, in quanto i fattori latenti oggetto di misura sono costrutti più circoscritti nella definizione. Una pratica è quella di eliminare gli item con loading più bassi e condurre nuovamente l'Analisi Fattoriale (pratica ripresa anche nella ricerca sociale: cfr. Marradi in Borgatta, Jackson, 1981, Gangemi, 1982).

3. A volte emergono fattori che sono idiosincratici agli specifici casi scelti (cioè non emergerebbero se avessimo selezionato altre unità). Per evitare questo problema, Nunnally

correlazione*	% di varianza spiegata	valutazione
,71	50%	eccellente
,63	40%	molto buona
,55	30%	buona
,45	20%	sufficiente
,32	10%	scarsa

(*) Si ricorda che esse coincidono con le saturazioni quando si è attuata una rotazione ortogonale; nel caso di rotazione obliqua le correlazioni sono contenute nella structure.

Tabella 2.

Per valutare la significatività sostantiva dei fattori estratti, vanno considerati tre aspetti strettamente interconnessi:

a) la sovradeterminazione dei fattori;
b) la **significazione**, ossia l'attribuzione di un significato, soprattutto attraverso i marker;
c) la gerarchia dei fattori.

a) Per **sovradeterminazione dei fattori** si intende l'esistenza di un buon numero di variabili con saturazioni significativamente diverse da zero nel fattore (quando la rotazione è obliqua è bene considerare i *regression loading* per decidere se un fattore è sovradeterminato o meno).
b) Alcune delle saturazioni significativamente diverse da zero, diciamo non meno di tre, devono essere elevate solo su un fattore; le variabili con saturazioni elevate su un solo fattore sono chiamate ***marker*** nella letteratura psicometrica. Se ci sono troppe variabili complesse i fattori saranno difficilmente distinguibili e interpretabili.
c) D'altro canto, fattori identificati solo da variabili molto simili fra loro da un punto di vista di contenuto semantico sono di scarso interesse; si parla in tal caso di fattori situati a un livello molto basso nella gerarchia dei fattori. Conviene eliminare i

propone una regola del pollice: è necessario che i casi siano almeno 10 volte più numerosi delle variabili manifeste (Nunnally, 1978).

'doppioni' e vedere se emerge ugualmente il fattore. Oppure si può procedere a una sotto–estrazione, cioè estrarre un numero inferiore di fattori anche se ciò peggiora il fit del modello.

Capitolo V

Un'applicazione empirica

Di seguito presentiamo un'analisi empirica a partire dai dati secondari tratti dalla *European Values Survey* (EVS), un'indagine campionaria condotta a livello europeo. Si tratta di un'importante rilevazione periodica di atteggiamenti e opinioni che ha tra i suoi principali obiettivi l'individuazione delle dimensioni fondamentali su cui si articola la sfera dei valori morali civili, "intesi come giudizi su ciò che è bene e ciò che è male, giustificabile o ingiustificabile rispetto a un insieme di atti nei confronti dei beni pubblici e dei diritti della persona" (Sciolla, 1999:282).

I dati qui utilizzati per l'applicazione empirica dell'*AF* Esplorativa riguardano la quarta rilevazione (*wave* del 2008) e tre paesi: Italia, Francia e Spagna. La scelta dei paesi è avvenuta in ragione del fatto che sugli stessi abbiamo un confronto per le tre *wave* precedenti della EVS, sulle quali è stato condotto uno studio che ha evidenziato la presenza di due dimensioni latenti concernenti la morale civile (Albano, Loera, 2004b:227): il civismo e il libertarismo. La batteria sottoposta ad *AF* è composta, come allora, da 9 item che riguardano giudizi di giustificabilità di determinate azioni, rilevati con una scala di valutazione che va da 1 a 10, a seconda che l'intervistato ritenga il comportamento in esame rispettivamente mai o sempre giustificabile[1].

Tra gli argomenti sottoposti al giudizio degli intervistati riportiamo di seguito (tabella 3) quelli che sono stati presi in considerazione nell'analisi.

Analizzando i dati delle tre precedenti *wave* relative ai paesi citati, erano state evidenziate due dimensioni sottostanti alla morale civica:

1. Per un esame più approfondito delle dimensioni della morale civile rinviamo a Albano, Loera, 2004a.

Dimensione di civismo	Dimensione di libertarismo
Non pagare il biglietto sui mezzi pubblici	Aborto
Non pagare le tasse (o pagarle meno del dovuto)	Divorzio
Cercare di ottenere dallo stato benefici a cui non si ha diritto	Omosessualità
Accettare denaro non dovuto (bustarelle)	Eutanasia
	Suicidio

Tabella 3.

— 'civismo vs antinormativismo' che raggruppa giudizi sulla giustificabilità di comportamenti lesivi dell'interesse pubblico. Tende al civismo chi risponde che tali comportamenti non sono giustificabili e all'antinormativismo chi risponde che lo sono;
— 'integrismo vs libertarismo' che riunisce giudizi sulla giustificabilità di comportamenti relativi alla sfera privata che sono oggetto di disputa nell'opinione pubblica dal punto di visto religioso, morale e giuridico. Tende all'integrismo chi risponde che tali comportamenti non sono giustificabili e al libertarismo chi risponde che lo sono.

Possiamo verificare se, a partire dal medesimo set di variabili manifeste, l'*AF* Esplorativa ci permette di estrarre gli stessi fattori anche per l'ultima wave, relativa al 2008, in tutti i tre paesi. Le scelte tecniche operate nella ricerca del 2004 sono state riapplicate a tutte le analisi di seguito riportate[2].

La matrice di input è una matrice di correlazioni lineari di Pearson. Il metodo di estrazione dei fattori utilizzato è quello dei *minimi quadrati non pesati*, ULS. I motivi di tale scelta sono essenzialmente tre:

— questo metodo non richiede una distribuzione normale multivariata delle variabili osservate ed è perciò preferibile quando non si dispone di variabili metriche;
— la funzione obiettivo del metodo minimizza lo scarto tra corre-

2. Le analisi sono state realizzate con il package *PASW* Statistics 18; si rimanda all'appendice I per le istruzioni.

lazioni riprodotte e correlazioni osservate, obiettivo primario di un'*AF*;
— non è richiesta la stima previa delle comunalità, operazione carica di arbitrarietà e in genere criticabile da un punto di vista statistico.

Per la rotazione è stata utilizzato il metodo *direct oblimin*[3], in quanto era nelle nostre aspettative che questi fattori avessero aree di sovrapposizione semantica e che quindi fossero correlati.

Per una corretta interpretazione delle correlazioni tra i fattori, occorre ricordare l'orientamento degli indicatori: questi sono tutti costituiti da scale che variano da 1 = 'comportamento mai giustificabile' a 10 = 'comportamento sempre giustificabile'. Di conseguenza i due fattori latenti, ancorati alle variabili manifeste, costituiscono due *continuum* dove l'estremo sinistro individua una posizione di totale 'rigorismo morale', nel senso di assoluto divieto dei comportamenti implicati; l'estremo destro, viceversa, individua una posizione di totale 'permissivismo'; infine, il punto centrale del *continuum* individua coloro che hanno una concezione condizionale ('relativistica') dei divieti. Consideriamo, per iniziare, i risultati sul campione complessivo dei tre paesi (tabella 4).

I risultati dell'analisi e i parametri di valutazione permettono di affermare che i due fattori comuni estratti interpretano adeguatamente le correlazioni tra i nove item; essi inoltre sono una sostanziale replica dei fattori individuati nella ricerca del 2004.

La stessa soluzione appare adeguata anche per ciascun singolo paese (tabella 5).

Volendo valutare in modo più preciso la stabilità dei modelli nei diversi contesti nazionali possiamo calcolare il coefficiente phi di Tucker (tabella 6).

Si è proceduto anche all'estrazione di un solo fattore per verificare

3. Si tenga conto che il parametro δ, che governa il grado di obliquità massima, è stato lasciato a zero, cioè il valore di default del programma, perché così si ottiene una matrice di *loading* che si avvicina alla struttura semplice (nota come *direct quartimin*) (cfr. Comrey, Lee, 1992², trad. it. 1995, p. 488). È possibile che questi valori aumentino se si permette un grado di obliquità massimo ($\delta = 0,8$).

Totale (3 paesi)		
	Fattori latenti	
Variabili manifeste	Libertarismo	Civismo
Aborto	,85	
Divorzio	,84	
Omosessualità	,73	
Eutanasia	,65	
Suicidio	,48	
Non pagare il biglietto		,66
Abuso benefici		,60
Evadere le tasse		,57
Accettae bustarelle		,45
Correlazione tra i fattori	0,33	
N. casi	5373	
Varianza spiegata	43,92%	
Rendimento	1,98	
Adattamento (repr)*	2%	

L'adattamento è stato valutato mediante la percentuale di residui superiori a |0,05| tra correlazioni osservate e correlazioni riprodotte dal modello. Più questa percentuale tende a zero, più le correlazioni riprodotte sono considerabili sostanzialmente uguali a quelle osservate. Come valore soglia è ragionevole una percentuale del 15% di residui >|0,05|.

Tabella 4.

	Italia		Francia		Spagna	
	Fattori latenti		Fattori latenti		Fattori latenti	
Variabili manifeste	Libertarismo	Civismo	Libertarismo	Civismo	Libertarismo	Civismo
Aborto	,85		,83		,83	
Divorzio	,84		,83		,80	
Omosessualità	,68		,69		,76	
Eutanasia	,68		,56		,69	
Suicidio	,55		,47		,38	
Non pagare il biglietto		,73		,62		,77
Abuso benefici		,63		,62		,57
Evadere le tasse		,59		,60		,55
Accettare bustarelle		,47		,35		,52
N° casi	1233		2933		1217	
Varianza spiegata	46,41%		40,63%		46,24%	
Rendimento	2,09		1,83		2,08	
Adattamento (Repr)	5%		2%		5%	

Tabella 5.

	phi di Tucker medio
Italia–Francia	0,95
Italia–Spagna	0,96
Francia–Spagna	0,95
phi medio totale	0,95

Tabella 6.

	Adattamento	Varianza spiegata	Rendimento
Totale (3 Paesi)	61%	30,95%	2,79
Italia	52%	30,65%	2,76
Francia	47%	27,42%	2,47
Spagna	83%	31,56%	2,84

Tabella 7.

l'adeguatezza di una soluzione più parsimoniosa. La soluzione a un fattore è però notevolmente peggiore dal punto di vista delle capacità del modello di riprodurre adeguatamente le correlazioni tra le variabili, come si può vedere dai valori dell'indice di adattamento riportati nella tabella 7.

Nella soluzione a due fattori si verifica, in tutti Paesi, una netta distinzione tra due dimensioni latenti, semanticamente distinguibili; tutte le variabili manifeste hanno una saturazione almeno sufficiente. Nella soluzione a un fattore, per contro, solo sei delle nove variabili manifeste sono ancorate al fattore latente. Tra le sei variabili che saturano sul fattore latente, cinque fanno riferimento alla dimensione del libertarismo e presentano *loading* buoni o eccellenti; una variabile, invece, si riferisce alla dimensione del civismo e presenta una saturazione piuttosto debole sul fattore latente.

Nel complesso, i risultati offrono degli indizi solidi in favore di un modello fattoriale a due fattori comuni: per i paesi considerati si conferma, anche per la *wave* 2008 della *EVS*, che il set di variabili manifeste attinenti ai giudizi di giustificabilità si strutturano lungo due dimensioni della morale, il civismo e il libertarismo.

Un passo ulteriore è quello di usare l'*AF* per creare due indici sintetici di civismo e libertarismo. Una volta stimati, i punteggi fattoriali

vengono poi dicotomizzati, al fine di creare una tipologia sui profili morali degli intervistati nei tre paesi (Albano, Loera 2004b:228); come punto di taglio per la dicotomizzazione si è scelta la media della distribuzione (equivalente a zero).

		CIVISMO	
		Sotto la media	*Sopra la media*
LIBERTARISMO	*Sotto la media*	Integristi antinormativi	Integristi civici
	Sopra la media	Libertari antinormativi	Libertari civici

Tabella 8.

Nella tabella 9 sono riportate le percentuali dei quattro tipi morali così ricavati, ripetute per ogni paese; la stima è avvenuta nel campione complessivo dei tre paesi: in tal modo, le colonne della tabella sono perfettamente confrontabili. Queste percentuali non sono invece confrontabili, se non indicativamente, con quelle della ricerca pubblicata nel 2004: una piena comparabilità è ottenibile riapplicando lo stesso modello di *AF* al campione complessivo dei tre paesi per le 4 wave (compito che però esula dagli obiettivi di questo saggio).

	Italia	Francia	Spagna
Integristi antinormativi	15,5%	12,9%	9,1%
Integristi civici	56,8%	28,6%	38,0%
Libertari antinormativi	11,8%	32,2%	25,1%
Libertari civici	15,9%	26,2%	27,8%

Tabella 9.

Capitolo VI

Brevi cenni all'Analisi Fattoriale Confermativa

Una trattazione introduttiva dell'*AF* richiede in conclusione alcuni brevi cenni al modello confermativo; questo è stato proposto e sviluppato dallo statistico–psicometrico svedese Karl Jöreskog (1935 –) a partire dai primi anni Settanta del secolo scorso.

Il modello confermativo (*AF*) rappresenta certamente un approccio evoluto perché supera le indeterminatezze statistiche del modello esplorativo[1]. Molteplici sono le funzionalità e applicazioni dell'*AF*. L'applicazione tipica consiste nel mettere a prova alcune ipotesi dedotte da un quadro teorico, riguardanti le fonti di variabilità latenti responsabili delle covariazioni in un set di variabili osservate. Abbiamo visto nelle pagine precedenti che il modello esplorativo, non impone alcun vincolo a priori sul numero dei fattori da estrarre, sui *loading* sui legami tra i fattori (comuni e unici). Il modello confermativo permette di porre dei vincoli su ognuno di questi parametri, ma più spesso solo su alcuni di essi. Come minimo il ricercatore specifica a priori il numero di fattori latenti.

Spesso si desidera ruotare la matrice originariamente calcolata dalla tecnica verso una *matrice bersaglio*. Ciò comporta specificare quali *loading* sono liberi, devono cioè essere stimati dal modello, e quali sono fissati a zero, cioè sono ritenuti nulli a priori. In alcuni casi si desidera 'ancorare' i fattori a determinate variabili manifeste, cosa che si ottiene fissando a 1 il relativo *loading*. In tal modo si fornisce ai fattori latenti l'unità di misura (Corbetta, 2002:155). Talvolta si ritiene che due indicatori siano intercambiabili, e quindi si dichiara l'uguaglianza di due *loading*, che poi vengono stimati dal modello con valori eguali.

1. Per chi è interessato ad approfondire l'argomento, tra i vari testi in circolazione segnaliamo Hoyle, 2000 e Corbetta, 2002.

Un'altra applicazione dell'*AF* consiste nel mettere a confronto modelli alternativi derivati da teorie rivali. Un caso particolarmente interessante di confronto è quello tra **nested models**: questi si hanno, in generale, quando un **modello pieno** contiene tutti i termini di un **modello ridotto** più altri termini. Nel caso dell'*AF*, potremmo ad esempio provare su un set costante di variabili costituenti test di personalità un modello derivato da una teoria pentafattoriale della personalità vs. un modello derivato da una teoria quadrifattoriale, in cui due fattori del modello pieno sono considerati un solo fattore comune nel modello ridotto.

Si potrebbe proseguire con altri esempi anche più specifici: questi, tuttavia, ci sembrano esaurienti a fornire una prima idea di cosa significhi spostarsi in un ambito confermativo. Può essere utile, a questo punto, fornire una rappresentazione grafica che aiuta a cogliere la differenza tra questo approccio e quello esplorativo (si confronti con la figura 3):

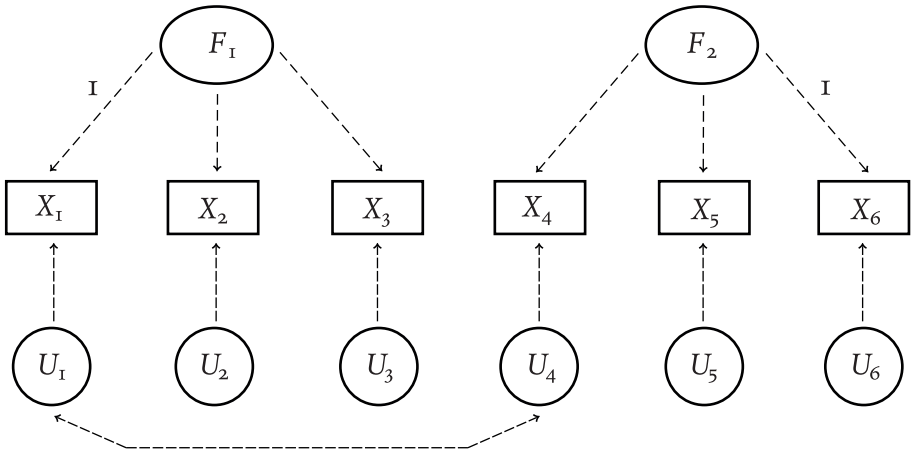

Figura 11.

Nell'esempio raffigurato si è imposto che:

a) i fattori comuni siano due;
b) i fattori comuni non siano correlati;
c) il primo fattore abbia la stessa unità di misura della variabile X_1;
d) il secondo fattore abbia la stessa unità di misura della variabile X_6;
e) il primo fattore sia saturato dalle prime tre variabili manifeste e l'altro dalla seconda terna;
f) due fattori unici siano correlati[2].

Nell'*AF* Confermativa la matrice di input è spesso una matrice **S** – varianze – covarianze; l'informazione in essa contenuta, di stimare oltre ai *loading* anche la media e la varianza di ciascun fattore, cosa che può essere utile soprattutto nell'analisi comparativa (tornando all'esempio del capitolo 5, potremmo chiederci se il fattore del civismo ha una media superiore in Italia o altrove).

Anche nell'*AF* Confermativa si possono impiegare più metodi di stima dei parametri; il più utilizzato è in genere *Maximum Likelihood* a meno di necessità specifiche.

Sono disponibili molti software per l'*AF* Confermativa: *LISREL* (il più noto), *EQS*, *MPLUS*, *AMOS* e altri ancora.

L'approccio confermativo dell'*AF* è in via di diffusione nelle scienze sociali. In questo ambito tuttavia, l'*AF* Confermativa è in genere intesa come componente di un modello più generale, dove essa costituisce il **modello di misurazione**, concernente le relazioni tra indicatori e costrutti latenti, che affianca il **modello strutturale**, concernente le relazioni di dipendenza tra costrutti latenti.

Questo più ampio apparato comprendente i modelli di misurazione e strutturale, è noto come **SEM – Structural Equation Model**[3], e

2. Può capitare quando ci sono cosiddetti effetti di "metodo": il tema non può qui essere sviluppato, ma un esempio semplice può dare un' idea approssimativa. Immaginiamo due domande di un questionario che si riferiscano a due fattori sostantivi ben distinti ma che abbiano un termine in comune (ad es. la parola 'politici'), a cui molti intervistati reagiscono, indipendentemente dal contenuto delle domande; questo 'rumore' linguistico può produrre una certa quantità di correlazione tra le due variabili che non ci sarebbe con due domande di contenuto semantico identico ma con un diverso *wording*.

3. In letteratura sono diffuse anche altre etichette: analisi delle strutture di covarianza,

costituisce uno degli strumenti più avanzati, ma anche più complessi da gestire con piena consapevolezza, per la ricerca quantitativa.

Modelli JKW (dagli autori, Jöreskog, Keesing, Wiley), Modelli Lisrel (dal nome del software inventato da Jöreskog e altri). Jöreskog e van Thillo (1973) parlano di "modelli di equazioni strutturali con variabili latenti". Tuttavia oggi, a livello internazionale, tende a prevalere l'acronimo SEM.

Conclusioni

Gli sviluppi impressi ai modelli di *AF* avutisi a partire dalla fine degli anni Sessanta hanno permesso a questa tecnica di superare le diffidenze da parte degli statistici più esigenti e rigorosi. L'*AF* Confermativa ha superato gli elementi di indeterminazione che caratterizzavano l'*AF* Esplorativa; inoltre, l'immagine di questa tecnica psicometrica si è irrobustita grazie alla sintesi con tecniche tipiche di altre tradizioni, in special modo quella sociometrica e quella econometrica.

Sarebbe errato però concludere che l'*AF* Esplorativa rappresenti un modello superato. Se l'*AF* Confermativa è la via più adeguata nel **contesto della giustificazione**, quindi nei momenti maturi della riflessione teorica in cui si controllano ipotesi sufficientemente dettagliate, quella esplorativa resta uno strumento indispensabile nel **contesto della scoperta**[4], in cui 'le domande di ricerca' prevalgono sui 'problemi', l'intuizione e le conoscenze tacite maturate nell'esperienza prevalgono su protocolli e procedure. Più che di opposizione o superamento, il rapporto tra *AF* Confermativa e *AF* Esplorativa si presenta come complementare: anche nell'ambito della stessa ricerca le due modalità possono essere combinate in una **tandem analysis**, in cui l'*AF* Esplorativa viene usata per generare un modello e l'*AF* Confermativa per sottoporlo a criteri più severi di prova statistica. In ogni caso, i fattori latenti estratti dovranno essere oggetto di ulteriore validazione su altri campioni; questa operazione, a sua volta, potrà richiedere l'impiego dell'*AF* Esplorativa per raffinare i modelli già usati in passato o per generare ulteriori modelli ritenuti più adeguati, in corrispondenza con la più recente letteratura teorica e con gli indicatori disponibili.

4. L'origine di questa distinzione diffusa nella filosofia della scienza può essere rintracciata molto in là nel tempo e in vari autori tra cui Popper, Reichenbach, Frege; per come qui intesa si veda Bruschi, 2005:6–7.

Per questi motivi, crediamo che l'*AF* Esplorativa continuerà a giocare un ruolo importante nella ricerca sociale, psicologica e economica accanto ad altre tecniche di analisi multivariata.

Appendice I
Istruzioni software e output

Le procedure computazionali che stanno alla base dell'*AF* si trovano implementate nei package e nei software statistici più diffusi. Vedremo ora come si esegue un'*AF* con il package *PASW Statistics 18*; verranno inoltre mostrati i principali oggetti che costituiscono l'output ottenuto con questo software. Prima di passare all'esame di alcuni esempi realizzati con *PASW Statistics 18* è utile riepilogare i principali elementi che sono coinvolti nella tecnica:

a) La matrice di input (tipicamente la matrice di correlazioni direttamente immessa dal ricercatore o calcolata dalla procedura a partire dalla $C \times V$);
b) l'elenco delle variabili di cui andrà specificato il nome e il modo di trattare gli eventuali valori mancanti (*missing values*);
c) il criterio con cui stabilire il numero di fattori da estrarre;
d) il metodo di estrazione dei fattori;
e) il metodo dell'eventuale rotazione dei fattori;
f) il metodo dell'eventuale stima dei punteggi fattoriali.

PASW Statistics 18 offre un'ampia scelta di opzioni per ciascuno di questi punti che devono essere oggetto di attente valutazioni da parte dell'analista. Il package, come per ogni altra tecnica, offre dei valori di default, cioè predefiniti; sconsigliamo però un loro uso acritico: infatti è proprio in questo caso che si corre il rischio di incappare in uno dei maggiori errori metodologici che si annidano nella procedura FACTOR così come è implementata. *PASW Statistics 18*, come altri package del resto, considera l'*ACP* come metodo di default per un'*AF*. Da quanto abbiamo detto in precedenza, dovrebbe essere a questo punto chiaro che invece le due tecniche non vanno confuse.

È utile riassumere in un quadro sinottico i valori di *default* (perlomeno per quanto concerne gli aspetti principali delle sintassi).

Matrice dati in input	$C \times V$
Trasformazione in input minimo	Correlazioni
Statistiche	Soluzione iniziale
Metodo di estrazione	Componenti Principali
Numero di fattori da estrarre	Numero di autovalori ≥ 1
Metodo di rotazione	Nessuno
Numero max di iterazioni per la convergenza nell'estrazione	25
Soglia di convergenza nell'estrazione*	p<0,001
Criterio di convergenza per la rotazione	Nessuno
Valori mancanti	Listwise

*. Questo valore rappresenta la soglia sotto cui l'algoritmo converge, interrompendo le iterazioni. Ciò accade quando l'incremento del valore di comunalità è minore di 0,001.

Tabella 10.

Servendoci dei dati EVS del 2008 utilizzati per l'analisi empirica del capitolo 5 comprendente un numero di casi corrispondente a 5373 (campione totale), mostriamo i principali passaggi. Ricordiamo che l'obiettivo dell'analisi era di individuare i fattori latenti in grado di interpretare le correlazioni tra i 9 item di giustificabilità di alcuni comportamenti. Di seguito[5] sono riportate le istruzioni di sintassi utilizzate per l'analisi empirica con *PASW Statistics 18*.

```
FACTOR
  /VARIABLES v1 to v9
  /MISSING LISTWISE
  /PRINT UNIVARIATE INITIAL REPR EXTRACTION ROTATION KMO
EXTRACTION
  /FORMAT SORT
  /PLOT EIGEN ROTATION
  /CRITERIA FACTORS (2) ITERATE (25)
  /EXTRACTION ULS /EXTRACTION ULS
  /CRITERIA ITERATE (25) DELTA (0)
  /ROTATION OBLIMIN
```

5. Si noti che nell'esempio presentato abbiamo rinunciato ad adottare i valori di default della tecnica, per operare precise scelte.

```
/SAVE = REG(ALL)
/METHOD CORRELATION.
```

Nell'ordine, le istruzioni di questo programma indicano:

— quali sono le variabili coinvolte, in questo caso dalla *v*1 alla *v*9 nella $C \times V$ `[VARIABLES]`[6];
— il criterio adottato per trattare i *missing values* `[MISSING] [LISTWISE]`[7];
— la richiesta di fornire nell'output: le statistiche descrittive monovariate (media e deviazione standard) per ogni variabile manifesta `[PRINT UNIVARIATE]`; la soluzione iniziale cioè le statistiche iniziali con tanti fattori quante sono le variabili manifeste e i relativi *eigenvalue* `[PRINT INITIAL]`; la matrice delle correlazioni riprodotte `[PRINT REPR]` dove sulla diagonale principale sono presenti le comunalità stimate, nel triangolo sinistro inferiore le correlazioni riprodotte, nel triangolo destro superiore i residui tra le correlazioni osservate e quelle riprodotte e in calce è riportata la percentuale di residui con valore assoluto maggiore di 0,05; la soluzione a 2 fattori non ruotata `[PRINT EXTRACTION]`; la soluzione a 2 fattori ruotata `[PRINT ROTATION]`, cioè la *pattern matrix*, la *structure matrix* e la matrice di correlazione tra i fattori; il test di sfericità di Bartlett `[PRINT EXTRACTION KMO]`;
— la richiesta di ordinare le variabili manifeste della *pattern matrix* e della *structure matrix* secondo valori decrescenti delle saturazioni sui rispettivi fattori latenti `[FORMAT SORT]`;

6. La matrice di input potrebbe essere sostituita da una matrice delle correlazioni tra le variabili manifeste con il seguente comando `[MATRIX=IN (COR=*)]`.

7. *Listwise* significa esclusione dall'analisi di tutti i casi che hanno almeno un valore mancante su una qualunque delle variabili manifeste introdotte nel modello. In alternativa, può essere chiesto il trattamento dei dati mancanti *Pairwise*, ovvero la selezione di casi per coppie di variabili, a fronte di quella generalizzata rappresentata da *Listwise*. Ciò significa, nel nostro caso, che dovendo calcolare la correlazione tra le variabili manifeste X_1 e X_2 si considerano validi tutti quei casi che non hanno *missing* su una delle due variabili, indipendentemente dal fatto che abbiano *missing* su altre coppie di variabili. L'utilizzo di questo criterio ha il vantaggio di trattare più informazione rispetto a *listwise*, ma può dare origine a matrici non positive semidefinite e quindi non processabili dalla tecnica. Il comando da utilizzare è [MISSING PAIRWISE].

— il grafico di caduta degli autovalori (*scree plot*) `[PLOT EIGEN]` e il grafico con la rappresentazione delle variabili manifeste nello spazio fattoriale ruotato avente come assi i fattori estratti `[PLOT ROTATION]`;
— il criterio di scelta del numero di fattori da estrarre; in questo caso viene fissato a 2 `[CRITERIA FACTORS (2)]`; il numero massimo di iterazioni dell'algoritmo di stima delle saturazioni `[ITERATE (25)]`;
— il metodo di estrazione prescelto `[EXTRACTION ULS]`;
— il metodo di rotazione, in questo caso *direct oblimin* `[ROTATION OBLIMIN]`[8], con parametro di obliquità δ fissato a zero, come di default `[DELTA(0)]`; il numero massimo di iterazioni dell'algoritmo di rotazione `[CRITERIA ITERATE (25)]`;
— l'input minimo è rappresentato dalla matrice di correlazioni `[METHOD=CORRELATION]`.
— la stima dei punteggi fattoriali con il metodo della regressione, che vengono salvati come nuove variabili nella matrice dei dati `[SAVE = REG(ALL)]`.

Riepilogando i principali elementi informativi che compaiono nell'output della procedura `FACTOR` sono i seguenti:

a) Statistiche descrittive
b) Il test di sfericità di Bartlett
c) Comunalità (iniziali e finali)
d) Varianza totale spiegata (iniziale e finale)
e) Grafico decrescente degli autovalori (eigenvalues)
f) Matrice fattoriale (*Factor matrix*)
g) Matrice delle correlazioni riprodotte — Comunalità riprodotte sulla diagonale principale
— Residui nel triangolo inferiore
h) Matrice dei modelli (*Pattern matrix*)

8. Qualora il ricercatore decida di non ruotare gli assi il comando da utilizzare è `[ROTATION NOROTATE]`.

i) Matrice di struttura (*Structure matrix*)
j) Matrice di correlazione tra i fattori

Di seguito riportiamo alcuni elementi dell'output:

Test KMO e di Bartlett		
Misura di adeguatezza campionaria KMO (Kaiser Meyer Olkin)		,830
Test di sfericità di Bartlett	Chi-quadrato appross.	13746,618
	df	36
	Sig.	,000

Tabella 11.

Il test KMO registra il valore di 0,830 che è da considerare buono perché superiore al valore soglia di 0,80. La statistica–test di Bartlett è significativa ($p - value < 0,01$).

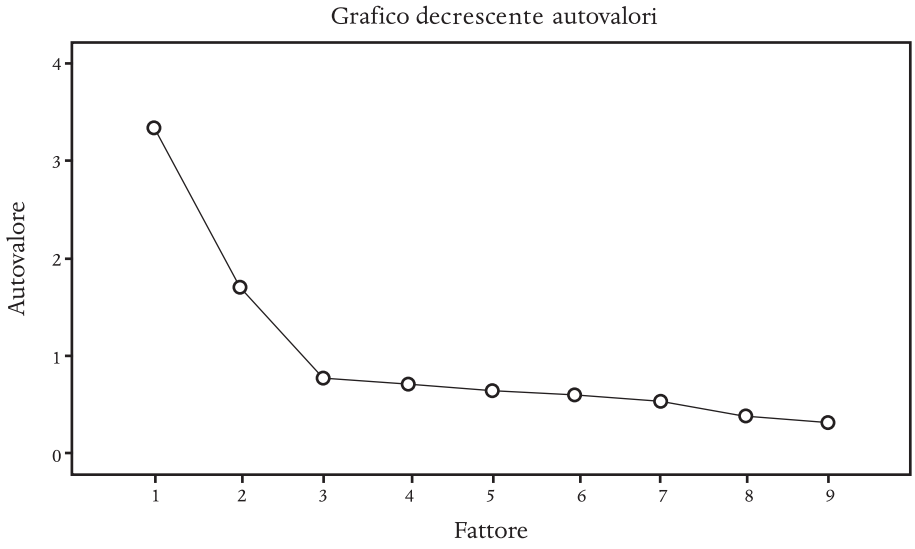

Figura 12.

Nel diagramma di caduta degli autovalori (figura 12) si individua un punto di gomito in corrispondenza del terzo fattore; il criterio

dello *scree test* proposto da Cattell suggerisce di estrarre i fattori che precedono il gomito, in questo caso due fattori[9].

Varianza totale spiegata

Fattore	Autovalori iniziali			Pesi dei fattori non ruotati			Pesi dei fattori ruotati[a]
	Totale	% di varianza	% cumulata	Totale	% di varianza	% cumulata	Totale
1	3,329	36,987	36,987	2,851	31,677	31,677	2,756
2	1,709	18,993	55,980	1,102	12,248	43,925	1,604
3	,785	8,720	64,701				
4	,714	7,928	72,629				
5	,631	7,009	79,638				
6	,599	6,661	86,299				
7	,535	5,941	92,240				
8	,395	4,384	96,624				
9	,304	3,376	100,000				

Metodo di estrazione: Minimi quadrati non pesati.
Due fattori presentano autovalori superiori a uno; i due fattori estratti riproducono complessivamente il 43,9% della varianza.

Tabella 12.

9. Se il criterio di Kaiser si basa su un'analogia con la teoria economica dell'utilità marginale, quello di Catell si fonda su una similitudine con la geologia, precisamente con la misurazione dei burroni. *Scree* è, infatti, la 'falda detritica' che si adagia ai piedi di un burrone; come i detriti creano disturbo nella misurazione della vera altezza di un burrone, e quindi non vanno considerati, così nello *scree test* occorre scartare i fattori che stanno a destra del 'gomito', cioè in una sorta di falda detritica della misurazione dei fattori.

Correlazioni riprodotte

		Abuso benefici	Evadere le tasse	Accettare busterelle	Omosessualità	Aborto	Divorzio	Eutanasia	Suicidio	Non pagare il biglietto
Correlazione riprodotta	Abuso benefici	,212[a]	,294	,264	,112	,140	,115	,130	,149	,279
	Evadere le tasse	,294	,416[a]	,374	,094	,123	,088	,125	,166	,382
	Accettare busterelle	,264	,374	,337[a]	,079	,104	,073	,107	,146	,342
	Omosessualità	,112	,094	,079	,512[a]	,601	,583	,473	,371	,191
	Aborto	,140	,123	,104	,601	,706[a]	,684	,556	,438	,235
	Divorzio	,115	,088	,073	,583	,684	,665[a]	,537	,417	,201
	Eutanasia	,130	,125	,107	,473	,556	,537	,440[a]	,352	,209
	Suicidio	,149	,166	,146	,371	,438	,417	,352	,292[a]	,224
	Non pagare il biglietto	,279	,382	,342	,191	,235	,201	,209	,224	,372[a]
Residui[b]	Abuso benefici		,033	-,009	,003	,008	-,007	-,001	,005	-,031
	Evadere le tasse	,033		-,021	-,013	,010	,015	-,009	-,016	,001
	Accettare busterelle	-,009	-,021		-,016	,001	-,004	,008	,010	,027
	Omosessualità	,003	-,013	-,016		,003	,022	-,034	-,008	,037
	Aborto	,008	,010	,001	,003		,007	-,007	-,001	-,017
	Divorzio	-,007	,015	-,004	,022	,007		,000	-,043	,004
	Eutanasia	-,001	-,009	,008	-,034	-,007	,000		,066	-,015
	Suicidio	,005	-,016	,010	-,008	-,001	-,043	,066		-,006
	Non pagare il biglietto	-,031	,001	,027	,037	-,017	,004	-,015	-,006	

Metodo di estrazione: Minimi quadrati non pesati.

a. Comunalità riprodotte

b. I residui vengono calcolati tra le correlazioni osservate e riprodotte. Ci sono dei residui non ridondanti 1 (2,0%) con valori assoluti maggiori di 0,05.

La matrice riporta nel blocco superiore le correlazioni riprodotte tra le 9 variabili e in quello inferiore i residui tra le correlazioni osservate e quelle riprodotte. La percentuale dei residui maggiori di |0,05| sul totale dei residui non ridondanti (2,0%) indica un buon adattamento del modello ai dati.

Tabella 13.

Matrice dei modelli[a]

	Fattore	
	1	2
Aborto	,848	-,024
Divorzio	,837	-,075
Omosesualità	,727	-,039
Eutanasia	,654	,028
Suicidio	,481	,137
Evadere le tasse	-,051	,660
Accettare bustarelle	-,054	,595
Non pagare il biglietto	,110	,565
Abuso benefici	,033	,448

Metodo estrazione: minimi quadrati non pesati.
Metodo rotazione: Oblimin con normalizzazione di Kaiser.
a. La rotazione ha raggiunto i criteri di convergenza in 3 iterazioni.

Tabella 14.

Matrice di struttura

	Fattore	
	1	2
Aborto	,840	,251
Divorzio	,813	,196
Omossesualità	,715	,198
Eutanasia	,663	,241
Suicidio	,525	,293
Evadere le tasse	,163	,643
Accettare bustarelle	,294	,601
Non pagare il biglietto	,139	,578
Abuso benefici	,179	,459

Metodo estrazione: minimi quadrati non pesati.
Metodo rotazione: Oblimin con normalizzazione di Kaiser.

Tabella 15.

Matrice di correlazione dei fattori

Fattore	1	2
1	1,000	,325
2	,325	1,000

Metodo di estrazione: minimi quadrati non pesati.
Metodo di rotazione: Oblimin con normalizzazione di Kaiser.

Tabella 16.

Appendice II
Elementi di algebra matriciale

Scalari, vettori, matrici

Chiamiamo *scalare* un qualsiasi numero reale.

L'*algebra matriciale* si distingue da quella scalare (o elementare) in quanto consiste di operazioni che comprendono oggetti più complessi degli scalari.

Questi oggetti sono rappresentati dai *vettori* e dalle *matrici*.

Si dice vettore un insieme di scalari ordinato su una riga o su una colonna. Se la disposizione degli scalari è orizzontale si parla di *vettore–riga*, se è verticale di *vettore–colonna*.

Due esempi, rispettivamente di un vettore–riga e di un vettore-colonna, sono i seguenti:

$$\begin{bmatrix} +1 & -7 & +0,3 \end{bmatrix}; \qquad \begin{bmatrix} +4 \\ -2 \\ +1; \end{bmatrix}$$

Il segno '+' può essere sottinteso e quindi omesso.

Si noti che l'ordine è importante; i due seguenti vettori sono distinti in quanto hanno gli stessi scalari ma ordinati in modo diverso:

$$\begin{bmatrix} +1 & -7 & +0,3 \end{bmatrix}; \qquad \begin{bmatrix} -7 & 1 & 0,3 \end{bmatrix}$$

Un vettore può talvolta essere riscritto come combinazione lineare di altri vettori. Si considerino ad esempio i tre seguenti vettori:

$$\mathbf{a} = \begin{bmatrix} 1 & -2 & 5 \end{bmatrix}; \qquad \mathbf{b} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}; \qquad \mathbf{c} = \begin{bmatrix} 1 & 0 & 5 \end{bmatrix}$$

c può essere riscritto come combinazione lineare di **a** e **b**:

$$\mathbf{c} = \mathbf{a} + 2 \cdot \mathbf{b}$$

Da un punto di vista geometrico un vettore è un segmento orientato, immerso in uno spazio *K*–dimensionale, così individuato: $\mathbf{v} = (v_1, v_2, v_3, \ldots, v_k)$; i valori nella parentesi sono dette *componenti* e si interpretano come coordinate.

Si definisce *lunghezza* o *norma* di un vettore v e si denota con $||\mathbf{v}||$ il risultato della seguente espressione:

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2 + \ldots v_k^2}$$

Un insieme di scalari ordinato su righe e colonne è detto *matrice*.

$$\begin{bmatrix} 1 & -7 & 0,3 \\ 4 & 5 & -2 \end{bmatrix}$$

Come è facile vedere, una matrice si può considerare alternativamente come un insieme ordinato di vettori-colonna (affiancati) o di vettori–riga (impilati)[10]. D'altro canto i vettori possono essere visti come matrici particolari (al limite anche uno scalare).

La dimensione di una matrice, più precisamente detta *ordine della matrice*, si esprime nel modo seguente:

$$(R \cdot C)$$

dove *R* è il numero delle righe che moltiplica *C*, il numero delle colonne.

Se $R = C$, cioè se il numero delle righe coincide con quello delle colonne, siamo in presenza di una matrice *quadrata*; in caso contrario la matrice è detta rettangolare.

In una matrice quadrata possiamo individuare la *diagonale principale* come quell'insieme di valori posti sulla diagonale che va dall'angolo in alto a sinistra a quello in basso a destra. Quando gli elementi, posti sopra e sotto la diagonale principale, individuati dalla stessa coppia di indici (ma invertiti), sono uguali allora la matrice è detta *simmetrica*.

10. Consideriamo qui solo matrici a due entrate cioè rappresentabili sul piano; l'algebra matriciale comunque opera anche su matrici cubiche e ipercubiche.

$$\begin{bmatrix} 1 & 3 & -0,5 \\ 3 & 4 & 2 \\ -0,5 & 2 & 0 \end{bmatrix}$$

Espressione di un sistema di equazioni lineari mediante matrici

Un sistema di M equazioni lineari in K incognite si può scrivere nella seguente forma:

$$\begin{aligned} W_1 &= a_{11}F_1 + a_{12}F_2 + \ldots + a_{1k}F_k + \ldots + a_{1K}F_K \\ &\vdots \\ W_m &= a_{m1}F_1 + a_{m2}F_2 + \ldots + a_{mk}F_k + \ldots + a_{mK}F_K \\ &\vdots \\ W_M &= a_{M1}F_1 + a_{M2}F_2 + \ldots + a_{Mk}F_k + \ldots + a_{MK}F_K \end{aligned}$$

in alternativa possiamo esprimere lo stesso sistema come segue:

$$\begin{bmatrix} W_1 & a_{11} & a_{12} & \cdots & a_{1k} & \cdots & a_{1K} \\ \vdots & & & & & & \\ W_m & a_{m1} & a_{m2} & \cdots & a_{mk} & \cdots & a_{mK} \\ \vdots & & & & & & \\ W_m & a_{M1} & a_{M2} & \cdots & a_{Mk} & \cdots & a_{MK} \end{bmatrix}$$

Questa scrittura più compatta è detta *matrice completa del sistema*.

Matrici speciali: diagonale, scalare, identità, triangolare, unitaria, nulla

Una matrice quadrata è detta *diagonale* quando tutti gli elementi esterni alla diagonale principale hanno valore zero.

Se solo gli elementi sotto la diagonale principale sono tutti uguali a zero, la matrice è detta *triangolare superiore*; viceversa, se solo gli elementi sopra la diagonale sono tutti uguali a zero è detta *triangolare inferiore*.

Una matrice diagonale **K**, in cui gli elementi posti sulla diagonale principale sono uguali a un valore costante k, è detta *matrice scalare*. È facile verificare che:

$$\mathbf{K} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{K} = k \cdot \mathbf{A} = \mathbf{A} \cdot k$$

Una matrice scalare con $k = 1$, è detta *matrice identità*, ed è indicata con **I**; essa presenta analogie con il numero 1 nell'algebra scalare: per esempio il prodotto di una matrice **A** con una matrice identità è uguale alla matrice **A** stessa; più in generale si dice che **I** è l'elemento neutro rispetto al prodotto.

Una matrice quadrata o rettangolare (o un vettore) contenente solo valori 1 è una *matrice–unitaria* (*vettore–unitario*).

Analoga al numero zero dell'algebra scalare è la *matrice nulla* (o il vettore nullo), i cui elementi sono tutti pari a zero: essa rappresenta l'elemento neutro rispetto all'addizione.

Operazioni tra matrici e tra matrici e scalari: addizione, prodotto, potenze

Due matrici, **A** e **B** possono essere sommate o sottratte se sono *compatibili rispetto alla somma*; perché lo siano devono essere dello stesso ordine. Il risultato sarà una matrice **C** dello stesso ordine, in cui ogni elemento c_{ij} sarà dato dalla somma (o sottrazione) di a_{ij} e b_{ij}.

Esempio.

$$\begin{bmatrix} 1 & 3 & -0,5 \\ 3 & 4 & 2 \\ -0,5 & 2 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 4 & -0,5 \\ 3 & -1 & 2 \\ 3,6 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 7 & -1 \\ 6 & 3 & 4 \\ 3,1 & 4 & 0 \end{bmatrix}$$

Una matrice **A**, qualunque sia il suo ordine, è moltiplicabile per uno scalare k. Il risultato dell'operazione è una matrice dello stesso ordine in cui gli elementi si ottengono moltiplicando quelli di **A** per k.

Esempio.

$$\begin{bmatrix} 1 & 3 & -0,5 \\ 3 & 4 & 2 \\ -0,5 & 2 & 0 \end{bmatrix} \cdot 4 = \begin{bmatrix} 4 & 12 & -2 \\ 12 & 16 & 8 \\ -2 & 8 & 0 \end{bmatrix}$$

Due matrici, **A** e **B** possono essere moltiplicate se sono *compatibili rispetto al prodotto*; perché lo siano, è necessario che il numero di colonne della prima sia uguale al numero di righe della seconda. Detto altrimenti, se la prima è di ordine $(N \cdot K)$ la seconda deve essere di ordine $(K \cdot M)$ (dove M e N possono essere uguali o diversi).

Il risultato sarà una matrice di ordine $(N \cdot M)$ i cui elementi si ottengono secondo la seguente formula:

$$c_{nm} = \sum_{k=1}^{K} a_{nk} \cdot b_{km}$$

Esempio.

$$\begin{bmatrix} 1 & -2 \\ -0,5 & 0 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 & -3 \\ 2 & 0,5 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & -5 \\ -2 & 0 & 1,5 \\ 16 & 1 & -7 \end{bmatrix}$$

Si noti che la moltiplicazione tra matrici non gode della proprietà commutativa: i due prodotti $\mathbf{A} \cdot \mathbf{B}$ e $\mathbf{B} \cdot \mathbf{A}$, ammesso che siano possibili entrambi, non danno di norma lo stesso risultato.

Poiché l'ordine dei fattori è importante, occorre usare un linguaggio più preciso di quello usato nell'algebra scalare, dove si direbbe che '*a* moltiplica *b*' o viceversa.

Consideriamo per esempio il prodotto $\mathbf{B} \cdot \mathbf{A}$: si dice che '**B** pre–moltiplica **A**', oppure che '**A** post–moltiplica **B**'.

In un prodotto tra una qualsiasi matrice di ordine $M \cdot K$ e una qualsiasi matrice di ordine $K \cdot M$, se $M < K$ si parla di *prodotto interno*; se $M \geq K$ si parla di *prodotto esterno*. Il prodotto interno di due vettori (necessariamente con lo stesso numero di elementi) ha come risultato uno scalare, e pertanto è detto *prodotto scalare*; il prodotto esterno di due vettori invece ha come risultato una matrice quadrata con un numero di righe e di colonne pari al numero di elementi dei vettori originari.

L'operazione di elevazione a potenza è applicabile solo a matrici quadrate: $\mathbf{A}^m = \mathbf{A} \cdot \mathbf{A} \cdot \ \ldots \ \cdot \mathbf{A}$ (m volte $\mathbf{A}$).

Come nell'algebra scalare un qualsiasi numero elevato a zero è pari per convenzione a 1, cioè l'elemento neutro rispetto al prodotto, anche nell'algebra matriciale una qualsiasi matrice (quadrata) elevata a zero è uguale all'elemento neutro rispetto al prodotto che qui è rappresentato dalla matrice **I**.

L'operazione di divisione di una matrice per un'altra non è definita nell'algebra matriciale; tuttavia, come nell'algebra scalare, si può vedere la divisione come la moltiplicazione di una matrice per l'*inversa* di un'altra (vedi oltre).

Valori caratteristici associati a una matrice: rango, traccia, determinante, autovalori e autovettori

Si definisce *rango* di una matrice il numero di *vettori linearmente indipendenti* contenuti in una matrice **A**, che può essere quadrata o rettangolare.

Una matrice quadrata che ha rango uguale al numero delle righe e delle colonne, cioè all'ordine, si dice di *rango pieno*.

Di una matrice quadrata qualsiasi si può calcolare la *traccia*, cioè quello scalare ottenuto sommando tutti gli elementi posti sulla diagonale principale:

$$\mathrm{Tr}\{\mathbf{A}\} = \sum_{i}^{n} a_{ii}$$

Abbiamo visto che in genere $\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A}$; notiamo ora che le tracce dei due prodotti, se sono entrambi definiti, coincidono, ovvero:

$$\mathrm{Tr}\{\mathbf{A} \cdot \mathbf{B}\} = \mathrm{Tr}\{\mathbf{B} \cdot \mathbf{A}\}$$

Il *determinante* è un altro valore caratteristico di una qualsiasi matrice quadrata, che ha molti utilizzi nell'algebra matriciale. Il calcolo del determinante è particolarmente laborioso per matrici di ordine superiore a 3.

Ci limiteremo qui a fornire l'esempio di calcolo più semplice, quello del calcolo del determinante di una matrice di ordine $(2 \cdot 2)$ e successivamente una definizione generale con un esempio per una matrice di ordine 3. Data una generica matrice **A** $(2 \cdot 2)$:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

il determinante è dato dall'operazione:

$$a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

Indichiamo ora $a_{11} \cdot a_{22}$ e $a_{12} \cdot a_{21}$ con la locuzione '*prodotti elementari*' di **A**. In generale un prodotto elementare ha N fattori, con N che dipende dall'ordine della matrice $(N \cdot N)$; questi fattori sono elementi provenienti dalla matrice secondo la regola che da ogni riga e ogni colonna deve essere estratto un solo elemento. Una matrice di ordine $(N \cdot N)$ ha $N!$ prodotti elementari. Si noti che per individuare facilmente gli elementi dei prodotti elementari si può seguire la seguente regola: costruire tutte le permutazioni semplici dei valori che può assumere ognuno dei due pedici di ciascun elemento a_{ij}.

In una matrice di ordine $(2 \cdot 2)$ i valori assumibili da ognuno dei due pedici sono 1 e 2; le permutazioni semplici sono $(1, 2)$ e $(2, 1)$; in una matrice di ordine $(23 \cdot 3)$ le permutazioni semplici sono $(1, 2, 3)$, $(2, 3, 1)$ $(3, 1, 2)$ $(1, 3, 2)$ $(2, 1, 3)$ $(3, 2, 1)$; e così via.

A questo punto è facile individuare i prodotti elementari con una tabellina come la seguente (il primo pedice assume sempre i valori in ordine crescente, in questo caso 1, 2, 3; il secondo pedice assume i valori della colonna delle permutazioni):

Prodotti elementari	Permutazione associata
$a_{11}a_{22}a_{33}$	(1,2,3)
$a_{12}a_{23}a_{32}$	(2,3,1)
$a_{13}a_{21}a_{32}$	(3,1,2)
$a_{11}a_{23}a_{32}$	(1,3,2)
$a_{12}a_{21}a_{33}$	(2,1,3)
$a_{13}a_{22}a_{31}$	(3,2,1)

Tabella 17.

Possiamo ora dare una definizione generale di determinante: quello scalare ottenuto dalla somma di tutti i prodotti elementari dotati di segno.

Resta quindi da definire il segno del prodotto elementare.

Per farlo torniamo alla tabella precedente, seconda colonna. Definiamo pari la prima serie $(1, 2, 3)$ (per definizione) e pari tutte quelle permutazioni in cui si conta un numero pari di spostamenti di valori rispetto alla prima serie per ottenere la permutazione; si definiscono dispari le altre permutazioni. Quando la permutazione è pari, il prodotto elementare è di segno positivo, altrimenti è di segno negativo.

Aggiungiamo ora alla tabella precedente altre due colonne:

Prodotti elementari	Permutazione associata	Pari o dispari	Prodotti con segno
$a_{11}a_{22}a_{33}$	(1,2,3)	pari	$+a_{11}a_{22}a_{33}$
$a_{12}a_{23}a_{32}$	(2,3,1)	pari	$+a_{12}a_{23}a_{31}$
$a_{13}a_{22}a_{31}$	(3,1,2)	pari	$+a_{13}a_{21}a_{32}$
$a_{11}a_{23}a_{32}$	(1,3,2)	dispari	$-a_{11}a_{23}a_{32}$
$a_{12}a_{21}a_{33}$	(2,1,3)	dispari	$-a_{12}a_{21}a_{33}$
$a_{13}a_{22}a_{31}$	(3,2,1)	dispari	$-a_{13}a_{22}a_{31}$

Tabella 18.

Il determinante gode di una serie di proprietà, tra cui le due seguenti (che non dimostriamo):

— $\text{Det}\{\mathbf{A} \cdot \mathbf{B}\} = \text{Det}\{\mathbf{A}\} \cdot \text{Det}\{\mathbf{B}\}$
— Se $\mathbf{X}$ è una matrice triangolare, superiore o inferiore, allora, il suo determinante è uguale al prodotto degli elementi della diagonale principale.

Quest'ultima proprietà è sfruttata per calcolare il determinante di matrici di ordine superiore a 3, dopo averle opportunamente trasformate in matrici triangolari (argomento per il quale rimandiamo a testi specialistici).

Dati una matrice quadrata $\mathbf{A}$, un vettore $\mathbf{v}$ compatibile per la postmoltiplicazione e uno scalare λ, tali da soddisfare la seguente relazione:

$$\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$$

allora λ viene detto *autovalore* (o *eigenvalue*) e $\mathbf{v}$ *autovettore*.

Una matrice può avere più autovalori e autovettori e questi ultimi hanno la caratteristica di essere linearmente indipendenti.

Ulteriori operazioni su singole matrici: trasposizione, inversione, estrazione di diagonale, partizione.

Data una matrice $\mathbf{A}$ di ordine $(M \cdot K)$, si definisce *trasposta* di $\mathbf{A}$ e si denota con $\mathbf{A}'$, quella matrice di ordine $(K \cdot M)$ la cui prima colonna è la prima riga di $\mathbf{A}$, la cui seconda colonna è la seconda riga di $\mathbf{A}$ e così via sino alla K–esima colonna (che è la K–esima riga di $\mathbf{A}$).

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -0,5 \\ 4 & 4 & 7 \\ -2 & 8 & 0 \end{bmatrix} \quad \mathbf{A}' = \begin{bmatrix} 1 & 4 & -2 \\ 3 & 4 & 8 \\ -0,5 & 7 & 0 \end{bmatrix}$$

Si noti che una matrice simmetrica è uguale alla sua trasposta.

Una matrice può essere moltiplicata per la sua trasposta e il risultato è una matrice quadrata simmetrica detta *prodotto–momento*. Se

$$\mathbf{A} \cdot \mathbf{A}' = \mathbf{A}' \cdot \mathbf{A} = \mathbf{I}$$

la matrice $\mathbf{A}$ è detta *ortogonale*.

Ricordando quanto detto circa la traccia di un prodotto tra matrici, abbiamo che:

$$\mathrm{Tr}\{\mathbf{A} \cdot \mathbf{A}'\} = \mathrm{Tr}\{\mathbf{A}' \cdot \mathbf{A}\} = \sum_{i}^{n} \sum_{j}^{m} a_{ij}{}^{2}$$

Il valore assunto da $\mathrm{Tr}\{\mathbf{A} \cdot \mathbf{A}'\}$ è una misura di quanto $\mathbf{A}$ differisce dalla matrice nulla.

Più in generale, la differenza tra due matrici $\mathbf{A}$ e $\mathbf{B}$ di ordine $(M \cdot N)$, può essere riassunta nella quantità:

$$d^2 = \mathrm{Tr}\left\{(\mathbf{A}-\mathbf{B})\cdot(\mathbf{B}-\mathbf{A})'\right\} = \sum_i^n \sum_j^m (a_{ij}b_{ji})^2$$

Si noti che $d^2 = 0$ solo se $\mathbf{A} = \mathbf{B}$.

Una matrice $\mathbf{A}(N \cdot N)$ post-moltiplicata da un vettore $(N \cdot 1)$ e contemporaneamente pre-moltiplicata dallo stesso vettore trasposto fornisce un valore scalare δ detto *forma quadratica*; in formula:

$$\delta = \mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x}$$

Il legame con le equazioni quadratiche è facilmente osservabile con un esempio come il seguente.

Sia:

$$ax^2 + 2bxy + cy^2 = \delta$$

un'equazione quadratica (i termini sono variabili al quadrato o prodotti di variabili).

In forma matriciale la precedente equazione diventa:

$$\begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} a & b \\ b & c \end{bmatrix} \cdot \begin{bmatrix} x & y \end{bmatrix} = \delta$$

Le forme quadratiche non sono limitate a due variabili.

Una matrice quadrata con forma quadratica positiva o nulla ($\delta \geq 0$) è detta *positiva semidefinita*.

Una matrice simmetrica e positiva semidefinita è detta *gramiana*; la gramianità di una matrice è una condizione importante nell'analisi dei dati: per esempio è un pre–requisito per un modello di *AF*, in quanto una matrice non gramiana produce degli autovalori negativi (cioè delle varianze negative).

L'operazione di inversione è definita solo per le matrici quadrate; se **A** è una matrice quadrata, si definisce inversa di **A** e si indica solitamente con la notazione $\mathbf{A}^{-1}$ quella matrice che post–moltiplicata o pre–moltiplicata da **A** fornisce (in entrambi i casi) la matrice **I**.

Non tutte le matrici quadrate sono invertibili: se non lo sono si dicono *singolari*.

Per valutare a priori se la matrice è non singolare, e quindi invertibile, si può calcolarne il rango oppure il determinante: se essa non è di rango pieno o, il che è equivalente, ha determinante uguale a zero, non è invertibile.

Inoltre, come si può dimostrare a partire dal fatto che

$$\mathrm{Det}\{\mathbf{A}\cdot\mathbf{B}\} = \mathrm{Det}\{\mathbf{A}\}\cdot\mathrm{Det}\{\mathbf{B}\}:$$

$$\mathrm{Det}\left\{\mathbf{A}^{-1}\right\} = \frac{1}{\mathrm{Det}\{\mathbf{A}\}}.$$

L'*estrazione della diagonale*, indicata dall'operatore Diag, è un'operazione definita solo per le matrici quadrate: essa consiste nella costruzione di un vettore–colonna, i cui elementi coincidono, nello stesso ordine dall'alto verso il basso con quelli della matrice in argomento.

Talvolta è utile esprimere un insieme di sottomatrici distinte combinandole in una sola matrice, detta *a blocchi*; viceversa talvolta è utile vedere un'unica grande matrice come un insieme concatenato di sottomatrici.

Per esempio la seguente matrice:

$$C = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & b_{11} & b_{12} \\ 0 & 0 & b_{21} & b_{22} \end{bmatrix}$$

in cui gli indici sono stati appositamente scelti per mostrarne la partizione, si compone di quattro sottomatrici $\mathbf{A}(2\cdot 2)$, $\mathbf{0}(2\cdot 2)$, $\mathbf{0}(2\cdot 2)$, $\mathbf{B}(2\cdot 2)$ concatenate in modo opportuno.

In forma sintetica si può scrivere:

$$C = \left\langle \begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{B} \end{array} \right\rangle$$

Bibliografia

Albano R., Loera B. (2004a), *La struttura dei valori di cittadinanza. L'analisi fattoriale per lo studio delle configurazioni valoriali*, Working Paper n. 6, Torino.

Albano R., Loera B. (2004b), *Note metodologico–statistiche*, in L. Sciolla, *La sfida dei valori*, il Mulino, Bologna.

Albano R., Testa S. (2002), *Introduzione alla statistica per la ricerca sociale*, Carocci, Roma.

Bagozzi R.P. (ed.) (1994), *Principles of Marketing Research*, Blackwell, Oxford [UK].

Barbaranelli C. (2003), *Analisi dei dati. Tecniche multivariate per la ricerca psicologica e sociale*, LED, Milano.

Basilevsky A. (1994), *Statistical Factor Analysis and Related Methods. Theory and Application*, Wiley, New York.

Becker G.S. (1964), *Capitale umano: un'analisi teorica ed empirica, con particolare riferimento alla Pubblica Istruzione*, University of Chicago Press, Chicago.

Bollen K.A. (2001), *Indicator: Methodology*, in N.J. Smelser, P.B. Baltes (editors in chief), *International Encyclopedia of the Social and Behavioral Sciences*, Elsevier Science, Oxford [UK], pp. 7282–7287.

Bruschi A. (2005), *Metodologia della ricerca sociale*, Laterza, Bari.

Carroll J.D., Arabie P. (1980), *Multidimensional Scaling*, in «Annual Review of Psychology», 31.

Cattell R. (1978), *The Scientific Use of Factor Analysis in the Behavioral and Life Sciences*, Plenum Press, New York.

Comrey A.L., Lee H.B. (1992), *A First Course in Factor Analysis*, Erlbaum Hillsdale [NJ], 2nd ed. (trad. it. *Introduzione all'analisi fattoriale*, LED, Milano 1995).

Corbetta P. (2002), *Metodi di analisi multivariata per le scienze sociali*, il Mulino, Bologna, II ediz.

De Lillo A., Arosio L., Sarti S., Terraneo M., Zoboli S. (2011), *Metodi e tecniche della ricerca sociale. Manuale d'uso per l'indagine quantitativa*, Pearson, Milano–Torino.

Dunteman G.H. (1989), *Principal Component Analysis*, Sage Univ. Paper, 69, Beverly Hills.

Gangemi G. (1982), *Criteri guida nell'analisi fattoriale. Dalla Simple Structure alla Two–Stage Component Analysis*, «Quaderni di Sociologia», 1, pp. 73–92.

Gorsuch (1983), *Factor Analysis*, L. Erlbaum Assoc., Hillsdale (NJ), 2nd ed.

Harman H. (1967), *Modern Factor Analysis*, Univ. Of Chicago Press, Chicago, 1976[3]

Harman H., Jones W.H. (1966), *Factor Analysis by Minimizing* «Psychometrika», 31, pp. 351–68.

Hotelling H. (1933), *Analysis of a Complex of Statistical Variables into Principal Components*, «Journal of Educational Psychology», 24, pp.417–41 e pp. 498-520.

Hoyle R.H. (2000), *Confirmatory Factor Analysis*, in: Tinsley H.E.A., Brown S.D. (eds), *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, «Academic Press», pp. 465-97.

Inglehart R. (1990), *Culture Shift in Advanced Industrial Society*, Princeton, Princeton Univ. Press, trad. it. *Valori e cultura politica nella società industriale avanzata*, Padova, Liviana 1993.

Jöreskog K.G., Sörbom D. (1988), *Lisrel 7. A Guide to the Program and Applications*, SPSS Inc, Chicago (Ill.).

Jöreskog K.G., Sörbom D. (1986), *Prelis. A Program for Multivariate Data Screening and Data Summarization. A Preprocessor for Lisrel*, Scientific Software Inc., Mooresville (Indiana).

Kampen J., Swyngedouw M. (2000), *The Ordinal Controversy* «Quality and Quantity», 34, pp. 87-102

Kline P. (1994), *An Easy Guide to Factor Analysis*, Routledge, London, (trad. it. *Guida facile all'analisi fattoriale*, Astrolabio, Roma 1997).

Labovitz S. (1970), *The Assignment of Numbers to Rank Order Categories*, «American Sociological Review», 35, pp. 515-24.

Lawley D. N., Maxwell A. E. (1963), *Factor Analysis as a Statistical Method*, Butterworth, London.

Lazarsfeld P.F. (1958), *Evidence and Inference in Social Research*, 'Dedalus', vol. 87, n. 4, pp. 99-130

Madge J. (1962), *The Origins of Scientific Sociology*, The Free Press of Glencoe, New York (trad. it. *Lo sviluppo dei metodi di ricerca empirica in sociologia*, il Mulino, Bologna 1966).

Marradi A. (1980), *Concetti e metodi per la ricerca sociale*, La Giuntina, Firenze.

Marradi A. (1981), *Factor Analysis as an Aid in the Formation and Refinement of Empirically Useful Concepts*, in: Borgatta E.F., Jackson D.J. (eds), *Factor Analysis and Measurement in Sociological Research. A Multidimensional Perspecitve*, Sage, London-Beverly Hills

McDonald R.P. (1982), *Linear Versus Nonlinear Models in Item Response Theory* «Applied Psychological Measurement», 4, Fall, pp. 379-96.

Nunnally J.C. *Psychometric Theory*, McGraw Hill, New York.

Oecd (2001), *The well-being of Nations. The role of human and social capital,* Paris.

O'Brien R.M. (1979), *The Use of Pearson's R with Ordinal Data*, «American Sociological Review», vol. 44, October, pp. 851-57.

Pearson K. (1901), *On Lines and Planes of Closest Fit to Systems of Points in Space*, «Philosophical Magazine», ser. 2, 6, pp. 559-72.

Pedon A. Gnisci A. (2004), *Metodologia della ricerca psicologica*, il Mulino, Bologna.

Piccolo D. (2000), *Statistica*, il Mulino, Bologna.

Pisati M. (2003), *L'analisi dei dati*, il Mulino, Bologna.

Ricolfi L. (1987), *Sull'ambiguità dei risultati delle analisi fattoriali*, 'Quaderni di Sociologia', 8, pp. 95-129.

Ricolfi L. (1999), *Destra e sinistra? Studi sulla geometria dello spazio elettorale*, Omega Ed., Torino.

Ricolfi L. (2000), *Tre variabili*, Franco Angeli, Milano (prima edizione: 1993).

Ricolfi L. (2000), *Manuale di analisi dei dati. Fondamenti*, Laterza, Bari.

Roccato M. (2003), *Desiderabilità sociale e acquiescenza. Alcune trappole delle inchieste e dei sondaggi*, Led, Milano.

Sciolla L. (2004), *La sfida dei valori*, il Mulino, Bologna.

Thurstone L.L. (1935), *The Vectors of Mind*, University Press, Chicago.

Thurstone L.L. (1947), *Multiple Factor Analysis*, University Press, Chicago.

Tucker L.R. (1951), *A Method for Synthesis of Factor Analysis Studies*, Dept. of the Army, Washington D.C.

Widaman K.F. (2007), *Common Factors Versus Component: Principals and Principles, Errors and Misconception*, in Cudeck R., Mac Callum R.C. (eds), *Factor Analysis at 100*, Lawrence Erlbaum Associates, Mahwah (NJ).

Indice analitico

QUADERNI DI RICERCA

Dipartimento di Scienze Sociali

1. Fiorenzo Girotti
Da burocrazia ad azienda.
Il ruolo della dirigenza nella trasformazione organizzativa del Comune di Torino
ISBN 88-88057-39-0

2. Massimiliano Vaira
La riforma universitaria:
strategie e leadership
ISBN 88-88057-39-4

3. Massimo Follis
Apprendimento e flessibilità del lavoro:
la logica delle carriere organizzative nel post-fordismo
ISBN 88-88057-49-8

4. Roberto Albano
Introduzione all'analisi fattoriale per la ricerca sociale
ISBN 88-88057-50-1

5. Lorenzo Venturini
Out of Danger?
An analysis of the trends in child poverty rates, in European Union Member States, between 1993 and 1998
ISBN 978-88-88057-51-4

6. Roberto Albano, Barbara Loera
La struttura dei valori di cittadinanza.
L'analisi fattoriale per lo studio delle configurazioni valoriali
ISBN 88-88057-52-8

7. Roberta Ricucci , Paola Maria Torrioni
Le regole della vita familiare:

differenze di classe, di background culturale e di genere
ISBN 88-88057-53-6

8. Barbara Loera , Raffaella Ferrero Camoletto
Capitale sociale e partecipazione politica dei giovani
ISBN 88-88057-54-4

9. Giuseppe Tipaldo
L'analisi del contenuto nella ricerca sociale.
Spunti per una riflessione multidisciplinare
ISBN 978-88-88057-80-4

10. Rosalba Altopiedi
Il doping nello sport.
Discorsi e pratiche delle organizzazioni sportive
ISBN 978-88-88057-93-4

11. Lorenzo Todesco
Il fenomeno dell'instabilità coniugale nei paesi occidentali.
Uno sguardo d'insieme
ISBN 978-88-88057-97-2

Compilato il 18 gennaio 2012, ore 11:04
con il sistema tipografico LaTeX 2_ϵ

Finito di stampare nel mese di dicembre del 2011
dalla «ERMES. Servizi Editoriali Integrati S.r.l.»
00040 Ariccia (RM) – via Quarto Negroni, 15
per conto della «Aracne editrice S.r.l.» di Roma