Rilevazione e analisi statistica del dato soggettivo

Parte II

I modelli di *scaling*

1. IL MODELLO ADDITIVO

1.1 GLI OBIETTIVI E GLI ASSUNTI

Il modello *additivo*[1] ha l'obiettivo di misurare, classificare e ordinare gli oggetti rispetto all'attributo al fine di individuare le differenze individuali rispetto alla caratteristica (Spector, 1992); alla base del modello vi sono due importanti assunti riguardanti la natura degli indicatori:

a. *l'insieme degli indicatori misura solamente un attributo* ovvero, facendo riferimento ad un modello fattoriale, gli indicatori linearmente combinati dovrebbero essere correlati solamente con un singolo fattore comune[2];
b. *ciascun indicatore è monotonamente correlato al continuum dell'attributo latente*: ciò vuol dire che nel caso si misuri, per esempio, un atteggiamento, più favorevole (sfavorevole) è l'atteggiamento del soggetto, maggiore (minore) è il suo punteggio atteso.

Dati tali obiettivi, questo modello di *scaling* richiede:

- <u>indicatori</u> dello stesso tipo selezionati sulla base della loro capacità nel discriminare tra casi e non sulla base delle loro posizioni relative sul continuum,
- <u>livelli di classificazione</u> dello stesso tipo associate a tutti gli indicatori; le diverse risposte di ciascun caso devono essere combinate in modo tale che sia possibile rappresentare differenze valide e affidabili tra i casi.

Tale modello può essere visto secondo due prospettive:

- <u>Prospettiva geometrica</u>: l'approccio additivo assume che uno degli insiemi di punti varia sistematicamente rispetto alla dimensione, mentre l'altro fluttua in modo casuale. In altre parole:
  - le variazioni sistematiche osservate nelle reazioni dei casi agli indicatori[3] è attribuita alle differenze individuali,
  - gli indicatori vengono considerati *misure ripetute*.

  Per ciascun oggetto nel primo insieme si sommano i valori di tutti gli oggetti nel secondo insieme in modo tale che le fluttuazioni si annullano a vicenda fornendo così una stima accurata della posizione del punto del primo insieme lungo la dimensione sottostante. Notare che i punti

---

[1] Tale modello è molto utilizzato nella rilevazione del soggettivo (atteggiamenti, disposizioni o opinioni personali, attitudini, ecc.). In genere la serie di item definiti è rappresentata da affermazioni rispetto alle quali i soggetti riferiscono il proprio grado di accordo su una scala ordinale composta, in genere, da cinque livelli: "molto d'accordo", "d'accordo", "indeciso", "in disaccordo", "molto in disaccordo" (scala Likert); a ciascun livello è associato un valore, nell'ordine da 1 a 5 o da 0 a 4. Assumendo che le risposte a ciascuna affermazione siano tra loro equivalenti è possibile calcolare un punteggio totale individuale sommando i valori associati a tutte le affermazioni.

[2] Il modello additivo presenta alcuni tratti comuni all'approccio fattoriale; in particolare gli assunti comuni sono:
- gli indicatori sono tra loro indipendenti e legati al tratto latente (punteggio vero),
- gli errori di misurazione non sono correlati tra loro, né con la dimensione latente,
- gli errori di misurazione sono dovuti al caso.

Infine possiamo aggiungere che il modello fattoriale consente di superare la condizione di parallelismo, che assume che tutti gli indicatori siano tra loro equivalenti, vista nel caso del modello additivo; attraverso il *factor loading* è infatti possibile sapere quanto ogni indicatore contribuisce alla formazione e composizione dei fattori.
I due approcci si differenziano rispetto alla definizione di "errore": nel modello fattoriale l'errore è formato da due componenti, una casuale, come per il modello additivo, l'altra attribuita alla variazione caratteristica di ciascun indicatore, non presente nel modello additivo; occorre però ricordare che il modello fattoriale consente di stimare solo la somma di tali due componenti (*unicità*).

[3] In questo caso gli indicatori sono rappresentati da variabili o item.

che rappresentano gli oggetti nell'insieme non scalato non sono fissati lungo la dimensione; essi variano in modo marcato da una riga all'altra della matrice dei dati.

- Prospettiva di misurazione: gli item individuali che vengono sommati per produrre la scala rappresentano funzioni di livello ordinale della dimensione latente. In altre parole i valori numerici assegnati alle righe della matrice dei dati sono correlati in modo monotono alla caratteristica sottostante. Sommare gli item comporta sommare le funzioni. Le funzioni monotone sommate dovrebbero essere lineari, in quanto le peculiarità delle specifiche funzioni monotone degli item dovrebbero compensarsi. Come già sappiamo se vi è una funzione specifica (lineare in questo caso) tra la caratteristica sottostante e un insieme empirico di assegnazioni numeriche, è stato raggiunto un livello di misurazione a intervalli.

Esistono diverse tecniche di *scaling* che, pur essendo trattate come differenti, rispondono al modello additivo ovvero rappresentano manifestazioni differenti dello stesso modello di base[4].

Dati e matrici. Le misure costruite sulla base di questo modello semplificano la rappresentazione delle osservazioni empiriche sommando i livelli in almeno uno dei modi di una matrice *multi-way multi-mode*; nella seguente figura una matrice $n*k$ viene ridotta sommando le colonne all'interno delle righe della matrice; il risultato è una matrice $n*1$ contenente punteggi di scala per gli oggetti rappresentati nelle righe; il modello di *scaling* riguarda i punti per gli n oggetti posizionati lungo la dimensione sommando i k oggetti all'interno di ciascuna riga:

		Matrice di input *V: two-way, two-mode* v_{ij} : presenta il grado in cui l'oggetto *i* domina l'oggetto *j*					
		Oggetti-colonna (variabili)					
		1	2	…	*j*	…	*k*
Oggetti-riga (unità)	1	v_{11}	v_{12}	…	v_{1j}	…	v_{1k}
	2	v_{21}	v_{22}	…	v_{2j}	…	v_{2k}
	…	…	…	…	…	…	…
	i	v_{i1}	v_{i2}	…	v_{ij}	…	v_{ik}
	…	…	…	…	…	…	…
	n	v_{n1}	v_{n2}	…	v_{nj}	…	v_{nk}

		Matrice di output X x_i : fornisce la stima della posizione dell'oggetto *i* lungo la dimensione
		Punteggi di scala
Oggetti-riga (unità)	1	x_1
	2	x_2
	…	…
	i	x_i
	…	…
	n	x_n

[4] La più conosciuta è quella che utilizza la tecnica Likert.

1.2 LA VERIFICA DEL MODELLO

Nel caso del modello additivo l'obbiettivo dell'analisi è principalmente quello di verificare che l'insieme degli indicatori misuri la stessa caratteristica ovvero condivida una comune varianza (*omogeneità* e *solidità*) (Carmines, 1992; Spector, 1992; Traub, 1994). Per fare questo è possibile distinguere principalmente tra due approcci:

- due componenti (parallele o non parallele); in questo caso lo schema sperimentale prevede la stima dell'affidabilità attraverso la verifica dell'*equivalenza tra componenti*; tra le tecniche per individuare le componenti la più utilizzata è lo *split-half*; le misure di equivalenza sono il coefficiente *Spearman-Brown*, se le componenti sono parallele, e il coefficiente *Rulon*, se le componenti risultano *tau-equivalenti* o *tau-essenzialmente-equivalenti*;
- più componenti ciascuna delle quali viene considerata come una misurazione separata (item); in questo caso lo schema sperimentale di stima dell'affidabilità è indicato come *consistenza interna*; l'affidabilità è stimata verificando l'*equivalenza tra* n *componenti*; le misure di equivalenza sono i coefficienti *KR-20*, *KR-21*, *alfa*, L_2; se le *n* componenti sono parallele, *tau-equivalenti* o *tau-essenzialmente-equivalenti*.

APPROCCIO		STIMA DELL'AFFIDABILITA'			PROBLEMI
		TIPO DI VERIFICA	METODI	TECNICHE E STRUMENTI	
Componenti	Parallele	equivalenza tra le due componenti (split-half)	confronto tra componenti	▪ coefficiente Spearman-Brown ▪ correlazione tra componenti ▪ coefficiente Rulon	identificazione delle componenti parallele
	Non parallele				identificazione delle componenti
Analisi della consistenza interna		confronto tra le n componenti	confronto tra item e tra ciascun item e lo strumento intero	▪ correlazione tra item ▪ correlazione item-totale ▪ coefficienti alfa, KR-20, KR-21, L_1, L_2	individuazione della omogeneità degli item

1.2.1 Affidabilità

1.2.1.1 Componenti parallele

Sappiamo che nel metodo *test-retest* i punteggi sono ottenuti tenendo costanti 'strumento' e 'modalità di somministrazione' mentre con il metodo delle forme parallele si tengono costanti momento e modalità di somministrazione e si varia lo strumento. Con le componenti parallele è possibile stimare l'affidabilità dell'intero strumento correlando due parti confrontabili dello strumento; tale approccio, detto anche *split-half*, consente di evitare la maggior parte dei problemi incontrati nell'applicazione dei due precedenti metodi.
Il limite del metodo è rappresentato dal fatto che la correlazione può variare anche in modo considerevole a seconda della tecnica di suddivisione dello strumento; a tale proposito è possibile identificare principalmente due tecniche:

- *suddivisione degli item secondo l'ordine di somministrazione* (prima metà e seconda metà degli item); tale tecnica non è consigliabile in quanto ciascun raggruppamento di item può

risultare influenzato, compromettendo il livello di affidabilità, in modo differenziato da particolari fattori e condizioni soggettive e/o ambientali non sempre controllabili o eliminabili (stanchezza, fatica, confidenza, noia, ecc.);
- *separazione degli item numerati con cifre dispari da quelli numerati con cifre pari* (*odd-even method*); è la tecnica più utilizzata in quanto consente di superare i problemi osservati con la precedente tecnica: non potendo escludere i fattori e le condizioni non controllabili si fa in modo che influenzino le due componenti in misura equivalente.

Ottenere la stima dell'affidabilità secondo questo metodo non è semplice, in quanto non sempre la suddivisione produce due componenti perfettamente parallele; è per tale motivo che la stima dell'affidabilità prevede due approcci diversi:

- Stima dell'affidabilità per componenti parallele: si procede correlando le due serie di punteggi ottenuti con i due sottogruppi di item (coefficiente di equivalenza); in realtà tale correlazione sottostima il valore di *rho* in quanto l'affidabilità è direttamente correlata con la numerosità del campione di item; questo vuol dire che il risultato di tale correlazione rappresenta l'affidabilità di due gruppi di item (strumenti paralleli) e non l'affidabilità dell'intero strumento; per questo il coefficiente di equivalenza viene sottoposto ad un correttivo che corrisponde ad una delle applicazioni più frequenti dell'equazione *Spearman-Brown* (di cui si parlerà più avanti) indicata come *formula Spearman-Brown per componenti parallele*:

$$rho_x = \frac{2r_{12}}{1+r_{12}}$$

dove
r_{12} correlazione tra le due componenti.
che produce una stima dell'affidabilità dell'intero strumento.

- Stima dell'affidabilità per componenti non parallele: le componenti si dicono "non parallele" quando i punteggi di ciascuna di esse sono uguali o diversi in quantità uguale per ciascun soggetto e quando gli errori standard di misurazione delle diverse parti non sono correlati tra loro. In questo caso le componenti sono dette, analogamente a quanto succedeva per gli strumenti, *tau-equivalenti* o *tau-essenzialmente-equivalenti*. Se si soddisfano i requisiti specificati, l'affidabilità può essere stimata utilizzando una formula riportata in letteratura ad opera di P.Rulon (nel 1939) ma attribuita a J.Flanagan (tutti studiosi di psicometria) che, utilizzando i punteggi osservati, produce una buona stima dell'affidabilità per l'intero strumento:

$$rho_x = 2\left(1 - \frac{\sigma_{y_1}^2 + \sigma_{y_2}^2}{\sigma_x^2}\right)$$

dove
$\sigma_{y_1}^2$ e $\sigma_{y_1}^2$ varianza dei punteggi osservati delle due parti dello strumento
σ_x^2 varianza dell'intero strumento.

Al contrario della stima dell'affidabilità per componenti parallele, qui non si richiede alcuna correzione che tenga conto della lunghezza dello strumento.

E' stato dimostrato che la formula di Rulon fornisce il limite inferiore dell'affidabilità dello strumento nel caso in cui le parti definite ricadono in una delle tre tipologie (componenti parallele, componenti *tau-equivalenti*, componenti *tau-essenzialmente-equivalenti*) ma anche nel caso in cui le componenti definite non ricadono in tali tipologie. Quindi più il risultato della formula di Rulon si avvicina ad 1, minore è l'incertezza riguardo all'affidabilità dello strumento anche quando le componenti dello strumento non soddisfano i requisiti di *parallelismo*.

1.2.1.2 Consistenza interna

L'analisi della *consistenza interna* (*internal consistency reliability*) mira a verificare l'entità della

componente comune a tutti gli item (omogeneità degli item); tale componente comune è attribuibile non solo alla capacità di un gruppo di item di misurare insieme qualcosa ma riflette la presenza di un sottostante costrutto comune. Gli item che riflettono una componente comune, dovrebbero condividere una comune varianza e, quindi, registrare alte correlazioni[5].
Le correlazioni tra gli item rappresentano il reciproco dell'errore; in altre parole la componente degli item non correlata con gli altri item è considerata "errore". Maggiore è la correlazione tra gli item, minore è la componente "errore" (DeVellis, 1991).
Stimare l'affidabilità di uno strumento attraverso la verifica dell'omogeneità degli item richiede l'utilizzo di tutte le informazioni contenute nei punteggi dei singoli item; il procedimento è di tipo iterativo e richiede l'analisi delle correlazioni tra item e delle correlazioni tra ogni item e tutti gli altri; tali informazioni trovano la finale informazione nei coefficienti di affidabilità, i cui valori (compresi tra 0 e 1) sono funzione del numero di item e del livello di intercorrelazione. Vediamo tali coefficienti (Carmines, 1992; Giampaglia, 1990; Spector, 1992).

Kuder-Richardson Formula 20 e 21

Kuder e Richardson hanno puntato la loro attenzione su strumenti composti da item dicotomici-binari. Delle molte equazioni da loro studiate, la 20[a] (1937) è divenuta quella più utilizzata:

$$KR-20 = \left(\frac{k}{k-1}\right)\left(1-\frac{\sum p_i q_i}{\sigma_x^2}\right)$$

dove

k	numero di item
p_i	proporzione di risposte positive di ciascun item i
q_i	$1-p_i$
$p_i q_i$	varianza dell'item i
σ_x^2	varianza dell'intero strumento (punteggio totale).

Il risultato prodotto dal coefficiente *KR-20* può essere più correttamente interpretato come il limite inferiore del livello di affidabilità dello strumento.
Esiste un'altra versione del *KR-20* che impone una condizione molto restrittiva in quanto assume che per tutti gli item vi sia la stessa proporzione di risposte positive (media per le variabili dicotomiche):

$$KR-21 = \left(\frac{k}{k-1}\right)\left(1-\frac{\mu_x - \mu_x^2/k}{\sigma_x^2}\right)$$

dove

k	numero di item
μ_x	media dei punteggi osservati dell'intero strumento
σ_x^2	varianza dell'intero strumento (punteggio totale).

Coefficiente *alfa*

Per stimare l'affidabilità di uno strumento nel 1951 Cronbach ha proposto l'utilizzo del coefficiente *alfa*, che confronta la varianza del punteggio totale con le varianze di tutti gli item (McIver, 1979).
In presenza di una totale mancanza di correlazione tra gli item, la varianza dello strumento corrisponde alla somma delle varianze di tutti gli item che lo compongono; un tale risultato indica che gli item osservati misurano aspetti specifici e non sono tra loro omogenei. *alfa* consente di confrontare tali varianze al fine di stimare il livello di affidabilità:

$$alfa = \left(\frac{k}{k-1}\right)\left(1-\frac{\sum \sigma_i^2}{\sigma_x^2}\right)$$

dove

k	numero di item
σ_i^2	varianza dell'item i

---

[5] E' questo un principio che è alla base anche dell'analisi fattoriale che, come vedremo, costituisce uno degli approcci alla misurazione multidimensionale.

σ_x^2 varianza dell'intero strumento (punteggio totale).

L'evidente analogia della formula di *alfa* con quella proposta da Kuder e Richardson consente di dire che il coefficiente *alfa* rappresenta la generalizzazione del *KR-20* al caso di strumenti composti da item non dicotomici.

Se si utilizza la matrice di correlazione tra gli item, il calcolo di *alfa* diviene:

$$alfa = \frac{k * r_m}{1 + r_m (k-1)}$$

dove

k numero di item

r_m media delle correlazioni tra tutti item.

Il coefficiente *alfa* può essere calcolato su dati sia in forma originale che standardizzata; in quest'ultimo caso si utilizza la matrice di varianza-covarianza[6] che, se calcolata su dati standardizzati, corrisponde alla matrice di correlazione con in diagonale valori unitari; in tal caso la formula per calcolare *alfa* diviene:

$$alfa = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum diag}{\sum tot}\right)$$

dove

k numero di item

diag elementi nella diagonale della matrice di varianza-covarianza (varianze dei punteggi dei singoli item)

Σtot somma di tutti gli elementi della matrice di varianza-covarianza.

Siccome nella matrice di correlazione in diagonale vi sono valori unitari, *Σdiag* corrisponde al numero di item (k).

Non è facile decidere se utilizzare dati in forma originale o in forma standardizzata; indicativamente se è possibile assumere per gli item uguaglianza tra le varianze della popolazione e quelle campionarie allora è possibile assumere che il calcolo di *alfa* non produca valori diversi utilizzando dati standardizzati o in forma originale. In caso contrario è più corretto utilizzare la prima versione della formula. Sia *alfa* che *KR-20* stimano l'affidabilità di uno strumento se le parti che lo compongono (gli item) sono parallele, *tau-paralleli* o *tau-essenzialmente-paralleli*, altrimenti fornisce un valore che può essere interpretato come limite inferiore dell'affidabilità.

Altri coefficienti

Prima del lavoro di Cronbach sul coefficiente *alfa*, L.Guttman nel 1945 aveva già provato a definire un coefficiente di affidabilità. In particolare Guttman era giunto a definire tre coefficienti L_1, L_2 e L_3 quest'ultimo formalmente corrispondente ad *alfa*. Guttman ha dimostrato che questi coefficienti stimano limiti inferiori dell'affidabilità anche quando non è possibile fare alcun assunto sulla natura della relazione tra le parti dello strumento.

Il più semplice tra i tre coefficienti, L_1, è definito nel seguente modo:

$$L_1 = 1 - \frac{\sum \sigma_i^2}{\sigma_x^2}$$

dove

σ_i^2 varianza dell'item i

σ_x^2 varianza dell'intero strumento (punteggio totale).

Il migliore coefficiente di stima del limite inferiore di affidabilità è considerato L_2, definito da Guttman nel seguente modo:

[6] Ricordiamo che in tale matrice i valori in diagonale rappresentano la varianza di ciascun item mentre il resto dei valori è rappresentato dalle covarianza tra tutte le possibili coppie di item.

$$L_2 = 1 - \frac{\sum \sigma_i^2}{\sigma_x^2} + \frac{\sqrt{\frac{k}{k-1}\sum_i \sum_j \sigma_{ij}^2}}{\sigma_x^2}$$

dove

k numero di item

σ_{ij}^2 covarianza tra i punteggi osservati per tutte le coppie di item ($i, j = 1,2,...,n$ con $i \neq j$)

σ_x^2 varianza dell'intero strumento (punteggio totale).

E' stato osservato che in genere L_2 produce un valore uguale o maggiore a quello ottenuto con il coefficiente *alfa*. Nella pratica il coefficiente più utilizzato è comunque *alfa* anche per la sua maggiore semplicità di calcolo. E' probabile che l'inserimento degli altri coefficienti nei più diffusi *package*, consentirà una loro maggiore diffusione. Infatti esistono *package* che consentono di calcolare alcuni o tutti tali coefficienti.

Occorre ricordare che i coefficienti L_1, L_2, *alfa* e *KR-20* non necessariamente producono un valore uguale o superiore a 0. Se la somma delle covarianze delle componenti è negativa, i valori dei coefficienti possono risultare negativi. In tal caso, naturalmente, i coefficienti forniscono praticamente un'informazione inutilizzabile sull'affidabilità dello strumento ma indicativa di una presenza di più dimensioni al di sotto delle diverse componenti (item). La seguente tabella riporta l'*output* per alcuni coefficienti relativi ad una scala.

INTERNAL CONSISTENCY DATA	
• SPLIT-HALF CORRELATION	.573
• SPEARMAN-BROWN COEFFICIENTI	.729
• GUTTMAN (RULON) COEFFICIENTI	.728
• COEFFICIENT ALPHA - ALL ITEMS	.715
• COEFFICIENT ALPHA - ODD ITEMS	.579
• COEFFICIENT ALPHA - EVEN ITEMS	.518

Come sappiamo in genere ci si attende un valore piuttosto alto di tali coefficienti per poter affermare che la scala è affidabile anche se non vi è concordanza tra gli autori sul valore minimo richiesto per ciascuno di tali coefficienti che può andare da .70 a .90.

Nel caso in cui sia stato registrato un valore non soddisfacente, si procede ad un'analisi approfondita dei singoli item per una loro selezione; tale selezione avviene sulla base di particolari caratteristiche.

1.2.2 Standard di affidabilità

Stabilire qual è il livello di affidabilità soddisfacente è un problema che dipende dal tipo di utilizzazione che deve essere fatto dello strumento sul quale è misurata l'affidabilità.

Tenendo presente che la messa a punto di uno strumento richiede la realizzazione di più esperimenti per la verifica dell'affidabilità, si può dire che per strumenti con obbiettivi prevalentemente descrittivi è possibile accontentarsi di un valore di affidabilità anche modesto; cercare in questi casi un valore di affidabilità maggiore può essere costoso in termini di tempo e di risorse non proporzionati agli obbiettivi dello strumento. Secondo alcuni autori (Nunnally, 1978) in questi casi il livello minimo accettabile di consistenza interna è 0.70. Quando però la misura dell'affidabilità riguarda strumenti le cui misurazioni prodotte dalla sua applicazione devono essere utilizzate per assumer importanti decisioni (si pensi al campo clinico o al campo sociale), il livello minimo di

affidabilità deve essere sufficientemente alto (0.90 o 0.95)[7].
Il mancato raggiungimento del livello di affidabilità auspicato può essere attributo

- alla scorretta formulazione del modello e della definizione del costrutto (per esempio errata individuazione della dimensionalità),
- alla presenza di item non corretti ovvero affetti da errore sistematico (item *biased*).

Mentre nel primo caso è necessario procedere ad una revisione del modello ed, eventualmente, ad una sua riformulazione nel secondo caso è necessario identificare gli item *biased* utilizzando strategia iterativa.

1.2.2.1 Selezione degli item

La strategia utilizzata per identificare gli item affetti da errore sistematico assume che gli item biased sono quelli che non condividono con gli altri la misurazione di una dimensione[8].
Per poter verificare tale condizione è necessario eseguire un procedimento iterativo i cui momenti sono:

a. calcolo del punteggio totale e osservazione della sua distribuzione di frequenza,
b. calcolo di indici relativi a ciascun item (indice di difficoltà, coefficiente di discriminazione),
c. calcolo e analisi delle correlazione tra tutti gli item e tra ciascun item e il punteggio totale (correlazione *item-totale*),

Al termine del procedimento si calcola il coefficiente di affidabilità; se i risultati del procedimento hanno individuato un item biased si prosegue ripetendo il procedimento stesso dopo aver eliminato tale item.[9]

A. <u>Calcolo del punteggio totale</u>
   In genere si calcola sommando semplicemente i valori che gli item hanno registrato per ciascun individuo. Il trattamento statistico del punteggio totale rappresenta un problema statistico risolto nell'ambito del più generale trattamento delle combinazioni lineari.
B. <u>Calcolo degli indici individuali</u>
   - *Proporzione di risposte corrette*:

$$P_i = \frac{y_i}{N}$$

---

[7] Una scala utilizzata per la stima dell’autosufficienza fisica di un gruppo a rischio (per esempio anziani) e i cui risultati devono consentire la programmazione di determinati servizi territoriali, deve necessariamente presentare un livello di affidabilità piuttosto alto.

[8] Date le sue caratteristiche, tale approccio si presenta particolarmente adatto nell'ambito dell'applicazione del modello additivo.

[9] Un altro approccio per l’identificazione degli item biased è quello basato sull’*Analisi della Varianza* (*ANalysis Of VAriance*, *ANOVA*) (Osterlind, 1983), che, come sappiamo, rappresenta una tecnica statistica molto potente e utile nell'ambito di molti ambiti ma piuttosto delicata e fragile quando utilizzata nell'ambito dell'individuazione del *bias* (altre strategie sono più convenienti e appropriate a questo fine). D'altra parte in questo ambito l'ANOVA è ancora molto utilizzata a causa della sua popolarità e della sua familiarità e riveste un'importanza storica (i primi studi di individuazione del *bias* ponevano al centro tale tipo di analisi).
Secondo tale approccio, dopo aver sottoposto lo stesso gruppo di item a due o più campioni estratti dalla stessa popolazione, si cerca di indagare il significato delle differenze riscontrate nei punteggi individuali. In particolare l'attenzione è concentrata sull'*interazione gruppi*item* anziché sugli effetti principali.
Come sappiamo gli item che misurano tratti o attributi diversi per sotto-gruppi estratti dalla stessa popolazione violano l'assunto richiesto di unidimensionalità. Parlando, per esempio, in termini di difficoltà, la perdità dell'unidimensionalità attraverso i sotto-gruppi molto probabilmente indica variazioni nei diversi livelli di difficoltà indipendentemente dalla differenza dei livelli di difficoltà totale dei gruppi. Un item è considerato non affetto da *bias* quando il livello di difficoltà ("probabilità di successo") su un item è lo stesso per i soggetti di uguale capacità della stessa popolazione indipendentemente della loro appartenenza ad un gruppo. Quindi è possibile inferire la presenza di *bias* ognivolta che si osserva un’interazione significativa *gruppi*item*. Per una trattazione sistematica dell’analisi della varianza si può ricorrere ad un manuale di statistica metodologica.

dove
y_i numero di risposte considerate "corrette" all'item *i*
N numero soggetti
Negli item dicotomici, corrisponde alla media.

- *Coefficiente di affidabilità* dell'item: prodotto tra deviazione standard dell'item e correlazione item-totale:

$$s_i = r_i \sqrt{P_i(1-P_i)}$$

- *Coefficiente di discriminazione*:

$$D_i = (P_a - P_b)_i$$

dove
P_a proporzione di risposte corrette all'item *i* di soggetti che sono nel 27% superiore dei punteggi totali
P_b proporzione di risposte corrette all'item *i* di soggetti che sono nel 27% inferiore dei punteggi totali.
Un'altra versione del coefficiente di discriminazione di un item è derivata da quella definita da Ferguson per un insieme di item (Guilford, 1954):

$$\delta = \frac{(n+1)\cdot\left(N^2 - \sum_{i=1}^{n}\sum_{j=1}^{k} f_{ij}^2\right)}{nN^2}$$

dove
δ coefficiente di discriminazione
N dimensione del campione
n numero di item del gruppo
k numero di punteggi definiti per ciascun item
f_{ij} frequenza del *j*-esimo punteggio per l'*i*-esimo item
Per un item singolo diventa

$$\delta_i = \frac{N^2 - \sum_{j=1}^{k} f_j^2}{N^2}$$

dove
δ_i coefficiente di discriminazione per l'item *i*
f_j frequenza del *j*-esimo punteggio

C. <u>Calcolo e analisi delle correlazione tra tutti gli item e tra ciascun item e il punteggio totale (correlazione *item-totale*),</u>
Un primo indice della capacità dei singoli item di condividere la misurazione di una dimensione comune è rappresentato dal livello di intercorrelazione presente tra i singoli item. La media di tali correlazioni può indicare la dimensione della componente comune a tutti gli item. La dispersione di tali correlazioni rispetto alla media indica quanto gli item tendono a variare nel contribuire alla componente comune[10].
Un altro indice è rappresentato dal livello di correlazione tra ciascun item e l'insieme degli altri item. La media di tali correlazioni può rappresentare un indice del livello di omogeneità di contenuto: se gli item sono tra loro omogenei dovremmo attenderci alte correlazioni tra il punteggio totale e i singoli item. L'osservazione di bassi valori di correlazione *item-totale* rilevano la presenza di item che, dimostrando di non condividere alcuna dimensione con gli altri item, possono essere definiti biased.
E' possibile stabilire in anticipo il numero di item che devono andare a comporre lo strumento finale e quindi selezionare un numero corrispondente di item tra quelli che registrano le correlazioni più alte. E' possibile anche

[10] Occorre a tale proposito ricordare che se da una parte la registrazione di un alto livello di correlazione può indicare e rilevare anche una sovrapposizione tra gli item e rivelare una ridondanza di informazioni (item che misurano la stessa cosa) dall'altra la mancanza assoluta di sovrapposizione può far pensare ad item che non appartengono alla stessa dimensione.

selezionare solo quegli item che registrano almeno un dato livello di correlazione (in genere 0.40).

Se un item registra una correlazione *item-totale* negativa vuol dire che non è classificato nella stessa direzione degli altri, ovvero che il contenuto dell'item è orientato in senso positivo quando gli altri lo sono in senso negativo, o viceversa; in questi casi l'item deve essere sottoposto a *riflessione*, invertendo il punteggio attribuito alle singole risposte dell'item in questione per uniformare la direzione delle risposte. Dopo la riflessione si procede nuovamente al calcolo del punteggio totale e delle correlazioni; in questi casi il numero di correlazioni positive aumenta come pure la dimensione media delle correlazioni.

Nella seguente tabella sono riportati i risultati di una delle iterazioni e relativi ad un gruppo di 10 item:

Item	Label	Item-total	Item-reliability	Excl. item	
		(*R*)		*R*	*alpha*
1	VAR83	0.695	0.467	0.623	0.837
2	VAR84	0.701	0.671	0.594	0.837
3	VAR85	0.646	0.454	0.561	0.841
4	VAR86	0.711	0.502	0.637	0.835
5	VAR87	0.523	0.425	0.402	0.853
6	VAR88	0.770	0.729	0.682	0.828
7	VAR89	0.542	0.414	0.432	0.850
8	VAR90	0.591	0.602	0.448	0.853
9	VAR91	0.727	0.599	0.642	0.833
10	VAR92	0.724	0.592	0.639	0.833
alpha	0.854				

Nell'esempio il gruppo di item risulta essere piuttosto omogeneo; l'esclusione a turno degli item dal calcolo del coefficiente alfa non ne fa aumentare il valore.

Una particolare attenzione va dedicata al coefficiente che si intende adottare per il calcolo della correlazione *item-totale*; in particolare se

- l'item è dicotomico puro allora di utilizza il coefficiente *punto-biseriale*,
- l'item e dicotomizzato allora si utilizza il coefficiente *biseriale* (Maggino, 2005).

Dopo l'identificazione e l'eliminazione dell'item (o degli item) biased si procede al calcolo del nuovo punteggio totale, delle correlazioni *item-totale*. Il procedimento iterativo termina quando i risultati osservati risultano soddisfacenti.

Come si è visto, gli item identificati come *biased*, in genere, vengono eliminati; in determinati casi è però anche possibile procedere in altro modo:

- ridefinendo e/o riconsiderando le categorie di risposta,
- riformulando la struttura verbale degli item e delle istruzioni,
- sostituendo l'item (l'item presenta problemi di validità).

In quest'ultimo caso può essere utile stimare il numero di item che dovrebbe essere aggiunto per raggiungere un livello accettabile di consistenza interna. A tale scopo è possibile utilizzare il coefficiente *Spearman-Brown (S-B)*

1.2.2.2 Coefficiente Spearman-Brown

Tale coefficiente consente di stimare:

a. gli effetti sul livello di affidabilità di un aumento o di una diminuzione del numero di item, per esempio stimare l'affidabilità di uno strumento con un numero maggiore di item a partire dall'affidabilità determinata per lo strumento originale,

b. il numero di item necessari per raggiungere un determinato livello di affidabilità.

Dato il valore di *alfa* per uno specifico numero di item, tale formula indica gli effetti su *alfa* di aumento o di una diminuzione del numero di item. Tale procedimento si basa sull'assunto che gli item aggiunti o eliminati presentano la qualità (affidabilità individuale) degli item iniziali; in caso contrario la formula può sovra-stimare o sotto-stimare il numero di item (McIver, 1979).

Stima dell'affidabilità di uno strumento con un numero diverso di item a partire dall'affidabilita dello strumento originale

Sapendo che l'affidabilità di uno strumento aumenta con il numero degli item, si può concludere che il principale modo per costruire strumenti più affidabili è quello di renderli più lunghi. In altre parole sappiamo che nei casi in cui è ragionevole aumentare la lunghezza dello strumento - tenendo presente che l'incremento non dovrebbe essere così grande da rendere inutilizzabile lo strumento - è possibile ottenere misurazioni maggiormente affidabili.

Il problema può essere quello di sapere quanto aumenta l'affidabilità con l'aumento del numero di item. Ciò ha condotto alla definizione di una particolare formula di Spearman e Brown detta *prophecy formula* (McIver, 1979). Proposta nello stesso momento dai due autori separatamente nel 1910, la formula consente di calcolare le variazioni del coefficiente di affidabilità di uno strumento in funzione dell'aumento (o della riduzione) del numero di item:

$$\textit{formula generalizzata di Spearman-Brown} = rho_x = \frac{k * rho_y}{1+(k-1)rho_y}$$

dove

rho_y affidabilità nota dello strumento

rho_x affidabilità da determinare

k a / b

dove

a numero di item dello strumento su cui calcolare l'affidabilità

b numero di item dello strumento con l'affidabilità nota.

Non è necessario che il fattore della lunghezza k nell'equazione sia un numero intero.

Disponendo di uno strumento di affidabilità rho_y=0.60, il raddoppio del numero di item produce un'affidabilità di

$$rho_x = (2*0.60)/(1+0.60) = 1.20/1.60 = 0.75$$

mentre un quadruplicamento

$$rho_x = (4*0.60)/(1+3*0.60) = 2.40/2.80 = 0.857.$$

Il livello di incremento del valore del coefficiente diminuisce gradualmente mano a mano che il valore di affidabilità si avvicina a 1.00. Se il livello di affidabilità previsto con la formula non aumenta in misura interessante si può concludere che lo strumento si presenta sufficientemente solido. La formula può essere utilizzata anche per stimare gli effetti sull'affidabilità della *riduzione del numero di item*[11].

E' importante sottolineare che, in relazione al numero di item maggiore o minore, la precisione della stima ottenuta con la formula dipende principalmente dalla differenza del numero di item tra uno strumento e la sua versione allungata o ridotta. Ciò vuol dire che non ci si dovrebbe aspettare una stima molto precisa se a partire dalla conoscenza dell'affidabilità di un'area di 5 item pretendiamo di stimare l'affidabilità dell'area con 40 item o viceversa. Nel caso in cui si voglia stimare l'affidabilità di uno strumento cui s'ipotizzi di raddoppiare il numero di item, la formula diviene:

$$rho_x = \frac{2rho_y}{1+rho_y}$$

dove

rho_x affidabilità dello strumento con lunghezza raddoppiata

rho_y affidabilità dello strumento originario.

Tale caso particolare ha condotto alla più nota applicazione della formula *S-B* ovvero quella che consente di stimare l'affidabilità di uno strumento attraverso il modello delle componenti parallele (*split-half*):

---

[11] Supponiamo di avere uno strumento con un livello di affidabilità di 0.75; se il valore di k è uguale a 0.5, la nuova affidabilità sarà:

$$(0.5*0.75)/[1+(0.5-1)0.60] = (0.375)/(1-0.375) = (0.375) / (-0.625) = 0.60$$

$$rho_x = \frac{2r_{12}}{1 + r_{12}}$$

dove

rho_x affidabilità dell'intero strumento

r_{12} correlazione tra le due componenti parallele.

Per poter utilizzare correttamente la formula è necessario soddisfare i seguenti assunti:
- gli item aggiunti (o eliminati) hanno le stesse caratteristiche statistiche (affidabilità individuale) degli item iniziali, conseguentemente gli item hanno uguale effetto sull'affidabilità,
- la correlazione media dello strumento di partenza deve essere la stessa dello strumento ipotizzato.

Tali assunti non vengono soddisfatti quando gli item nuovi e quelli vecchi differiscono sistematicamente e in modo significativo
- nel contenuto (ovvero provengono da campi di contenuto diversi),
- nell'affidabilità (ovvero se la correlazione media in un gruppo è più alta di quella dell'altro).

Il nome, non privo di ironia, attribuito alla formula suggerisce quanto sia difficile soddisfare i requisiti richiesti e conseguentemente quanto il risultato ottenuto sia da considerare con le dovute cautele. Ciò vale soprattutto per la seguente applicazione.

Numero di item necessari per raggiungere un determinato livello di affidabilità

La formula *Spearman-Brown* è stata generalizzata anche per stimare il numero di item necessari per raggiungere un desiderato livello di affidabilità. Per determinare quanto dovrebbe aumentare il numero di item per raggiungere il livello di affidabilità desiderato la formula *S-B* viene trasformata nel modo seguente:

$$k = \frac{rho_x\left(1 - rho_y\right)}{rho_y\left(1 - rho_x\right)}$$

dove

rho_x affidabilità desiderata

rho_y affidabilità nota

k di quanto deve aumentare il numero di item per ottenere l'affidabilità desiderata

L'errore standard di k è dato da

$$\sigma_k = \frac{k\left(1 - rho_y^2\right)}{rho_y^2\sqrt{N}}$$

L'applicazione di tale formula conduce alla conclusione che per raggiungere un'affidabilità moderatamente alta si dovrebbe aumentare il numero di item spesso anche in modo considerevole (e, il più delle volte, impraticabile)[12].

1.2.2.3 Un altro approccio alla selezione degli item: Transformed Item Difficulties

L'approccio noto come *Transformed Item Difficulties* (*TID*) assume (Osterlind, 1983) che un item è *biased* quando risulta più "difficile" per un gruppo rispetto all'altro: l'errore è indicato da una

[12] Supponiamo di avere uno strumento composto da 20 item con una affidabilità di 0.50; secondo il modello proposto, volendo raggiungere un livello di affidabilità di 0.80, dovremmo aumentare il numero di item di:

k = [0.80 (1-0.50)]/[0.50 (1-0.80)] = 0.40/0.10 = 4

ovvero è necessario aumentare il numero di item di quattro volte (20*4=80 item!!).

Nel caso in cui lo strumento disponga di 40 item e presenti un'affidabilità di solamente 0.20 e si desideri un'affidabilità di almeno 0.80:

k = [0.80 (1-0.20)]/[0.20 (1-0.80)] = 0.64/0.04 = 4

ovvero sarebbe necessario arrivare a 40*4=160 item!!. Ma sappiamo anche che con uno strumento con un così basso livello di affidabilità è principalmente necessario mettere in discussione l'intera sua formulazione.

differenza significativa tra gruppi rispetto alla difficoltà relativa dell'item. Tale approccio è concettualmente lineare e chiaro[13] anche grazie alla possibilità che offre di utilizzare tecniche grafiche per la rappresentazione delle difficoltà degli item; inoltre consente di rappresentare in forma di equazione sia i punteggi individuali (capacità) che quelli degli item (difficoltà).
L'ipotesi da verificare attraverso tale strategia non è solamente la presenza o l'assenza dell'interazione *gruppi*item*, anche se ciò non è fatto in senso formale di accettazione o di rifiuto dell'ipotesi nulla attraverso un criterio statistico, ma anche il livello relativo in cui determinati item possono variare tra gruppi.
Il procedimento è piuttosto semplice e si sviluppa nei seguenti momenti:

1. Determinazione dell'indice di difficoltà dell'item (per esempio *p*) per ciascuno dei gruppi da confrontare.
2. Conversione o trasformazione di tale indice in un punteggio standardizzato.
   Come sappiamo, l'indice di difficoltà più semplice da utilizzare è *p* che per poter essere utilizzato deve essere standardizzato. Nell'ambito dell'approccio *TID* la trasformazione di *p* in scala standardizzata è detta *delta*; la scala così ottenuta assume particolari valori; in particolare per l'item *i* e il gruppo *j*:
   $$delta_{ij} = 4z_{ij} + 13$$
   dove *z* rappresenta il valore *p* standardizzato.
   I valori *delta* così ottenuti compongono una scala ad intervalli, con punti tra loro equidistanti consentendo trasformazioni lineari, con media 13 e deviazione standard 4.
   Uno dei vantaggi dell'utilizzo dei valori *delta* sta nel fatto che consentono di evitare i valori negativi; infatti essi vanno da 0 a 26 che corrispondono, rispettivamente, ai valori *p* di .999 e .001; quindi valori alti di *delta* indicano item difficili mentre valori bassi indicano item facili[14].
   E' comunque possibile adottare altre trasformazioni in cui *z* rappresenta il *(1-p)-esimo* percentile della distribuzione normale standardizzata sarà sufficiente.
3. Creazione di un diagramma a punti in cui
   a. in ascissa vi sono i valori di difficoltà per il I gruppo,
   b. in ordinata vi sono i valori di difficoltà per il II gruppo,
   c. ciascun punto rappresenta un item.

   Il livello di dispersione dei punti nel grafico costruito è considerato una misura dell'interazione *gruppo*item*, una specie di coefficiente di correlazione inverso.
4. Interpolazione di una retta, detta asse maggiore dell'ellisse descritta dai punti (*major axis line*), interpretata come indice della relazione bivariata dei valori *delta* dei due gruppi.
   A partire da tale retta è possibile individuare la presenza di item *biased*. L'asse maggiore minimizza la distanza tra i diversi *delta* ed è individuato attraverso una procedura matematica che non coincide con quella dei minimi quadrati, in quanto, diversamente dall'analisi di regressione, qui non è possibile individuare una "variabile indipendente" e una "variabile dipendente" ovvero non esiste alcun criterio che consenta di decidere quale gruppo di punteggi deve essere regredito a partire dall'altro: per tale motivo la retta migliore è considerata quella che presenta la minima distanza perpendicolare dai punti[15]. A parte questa differenza l'approccio

---

[13] Questo approccio assume che gli item siano unidimensionali.

[14] Inoltre l'errore standard associato a *delta* rimane costante per tutti i livelli di difficoltà dell'item L'errore standard di $delta_{ij}$ è calcolato nel modo seguente:

$$ES_{delta_{ij}} = \frac{4}{N_j - 1}$$

L'errore standard di una singola porzione dovrebbe essere massimo .01; se risulta essere maggiore può sorgere il sospetto che la dispersione dei punti possa essere attribuita a fluttuazioni campionarie.

[15] La distanza perpendicolare di ciascun punto dall'asse maggiore è considerata una funzione dell'interazione di *gruppo*item* ed è data da:

appare molto simile a quello che consente di individuare la retta di regressione sulla base della seguente equazione:

$$y = a + bx$$

In questo caso le due costanti vengono calcolate nel modo seguente:

$$b = \frac{\left(\sigma_y^2 - \sigma_x^2\right) \pm \sqrt{\left(\sigma_y^2 - \sigma_x^2\right)^2 + 4r_{xy}^2\sigma_{xy}^2\sigma_y^2}}{2r_{xy}\sigma_x\sigma_y}$$

$$a = m_x - bm_y$$

dove

x e y simboli che si riferiscono ai due gruppi

σ_x e σ_y deviazioni standard rispettivamente del gruppo x e del gruppo y

r_{xy} coefficiente di correlazione tra i due gruppi di valori di *delta*.

5. Definizione di una funzione di distanza che consenta di verificare la distanza minima di ciascun item dalla retta.
   La presenza del *bias* è rilevata per quegli item che sono relativamente distanti dalla retta. Il problema a questo punto è quello di capire qual è il limite di tolleranza accettabile oltre il quale la distanza dall'asse maggiore rivela la presenza di un item *biased*. Uno degli approcci più comuni è quello che determina gli intervalli di confidenza dell'asse maggiore. Gli item che nel grafico appaiono al di fuori di tali intervalli possono essere giudicati "deviati". Spesso per stabilire i confini accettabili di grandezza del *bias* si usa il limite delle unità di *punti-z* a ±.75; in alcune occasioni tale limite è ritenuto troppo rigoroso e si preferisce fissare il limite delle unità di *punti-z* a ±1.5.

Nell'applicazione di questo approccio occorre prestare particolare attenzione al possibile insorgere di alcuni problemi.

- Abbiamo visto come la strategia basata sul *TID* di identificazione degli item *biased* è basata sull'assunto secondo il quale l'interazione *gruppi*item* rappresenta una valida indicazione per identificare gli item deviati; il livello di deviazione può essere considerato misura del *bias*.
  Nell'applicazione pratica di tale approccio occorre però tenere presente che in determinate circostanze tale assunto non può essere sostenuto; se per esempio gli item sono di varia difficoltà, l'interazione *gruppi*item* può esistere anche in un item perfettamente corretto (*unbiased*); quindi l'associazione tra item *bias* e interazione *gruppi*item* può risultare inappropriata.
- La proporzione di risposte corrette non rappresenta necessariamente una reale misura della difficoltà degli item, in quanto non ci si può aspettare che i componenti di due diversi gruppi reagiscano agli item con la stessa proporzione (i punti non necessariamente si posizionano esattamente sull'asse maggiore) anche nel caso di item *unbiased*. Se esistono due diversi livelli di forza discriminante per gli item (uno per ogni gruppo) il grafico dei valori di *delta* potrebbero ricadere su curve diverse, causando confronti ineguali. In questi casi è opportuno ricorrere all'approccio basato sull'*Item Response Theory*, di cui parleremo più avanti.
- La proporzione di casi che risponde correttamente ad un item così come è indicata dal valore di *delta* descrive nell'approccio *TID* non solamente l'item, in termini di difficoltà, ma anche il gruppo osservato, in termini di capacità; ciò rappresenta un grave limite all'identificazione di item affetti da errori sistematici soprattutto se basata sul valore *p*.

$$D_i = \frac{bX_i + a - Y_i}{\sqrt{\left(b^2 + 1\right)}}$$

dove

b pendenza dell'asse maggiore

a intercetta

X_i e Y_i punteggi *delta* per ciascun gruppo rispetto all'item i.

Per superare alcuni dei problemi visti in molte applicazioni pratiche la metodologia *TID* (Osterlind, 1983) presenta alcune interessanti varianti:

a. Per evitare i difetti attribuiti teoricamente alle trasformazioni viste in precedenza, la trasformazione i valori *p* in *punti-z* può utilizzare le medie e le deviazioni standard per gruppo; in questo caso la funzione di distanza applicata è quella vista in precedenza, anche se la pendenza dell'asse maggiore viene convenzionalmente definita come unitaria (b=1); concettualmente tale modifica rimane molto vicino alla strategia tradizionale.

b. Si assume una retta fissa con un'inclinazione di 45°. Dopo aver calcolato i valori *delta* per ciascun gruppo nel solito modo, anziché confrontare i valori *delta* tra loro, si verifica la normalità della distribuzione $delta_{i1}$-$delta_{i2}$ considerando la media e la varianza delle differenze come parametri. Tali parametri sono considerati limiti di confidenza rispetto ai quali è possibile fare inferenze per identificare gli item *bias* e *unbiased*; tale procedura è essenzialmente un test di bontà di adattamento (del tipo Kolmogorov-Smirnov).

c. Per l'identificazione degli item *biased*, l'interazione *gruppi*item* viene definita in maniera diversa:
- *effetti non-ordinali*: conseguenza di item che hanno diversi ranghi di difficoltà tra i due gruppi; tali effetti, considerati un potente indicatore della presenza di item *biased*, possono essere osservati calcolando la cograduazione (in genere *rho* di Spearman) tra i valori dei due gruppi; il valore $1\text{-}rho^2$ sta ad indicare l'aspetto puramente non-ordinale dell'interazione *gruppi*item*;
- *effetti ordinali*: conseguenza delle differenze relative nelle difficoltà degli item indipendentemente dai ranghi di entrambi i gruppi; per stimare gli effetti ordinali può essere utilizzato il coefficiente di correlazione *r*, consentendo in questo modo il confronto tra effetti ordinali e disordinali; per il calcolo di *r* si propone di non utilizzare i valori *delta* semplici ma quelli modificati calcolando la distanza tra valori di *delta* ordinati ($delta_1$-$delta_2$, $delta_2$-$delta_3$, ecc.); tali nuovi valori sono detti *decrementi di delta* e sono ottenuti all'interno di ciascun gruppo; *r* è quindi calcolato su tali valori.

Il valore $rho^2(1\text{-}r^2)$ è considerato una stima degli effetti puramente ordinali dell'interazione *gruppi*item*. La porzione residua (contributo congiunto alla varianza dell'interazione degli effetti ordinali e disordinali) dell'interazione *gruppi*item* è uguale a $rho^2\ r^2$.

d. Per superare uno dei principali limiti dell'approccio *TID*, ovvero la sua incapacità a definire la difficoltà degli item indipendentemente dai gruppi osservati, è possibile calcolare la correlazione parziale tra successo rispetto ad un determinato item e l'appartenenza di gruppo (o punteggi individuali attesi). In pratica questo approccio, controllando la capacità, supera uno dei suoi principali limiti mantenendone la sua semplicità.

1.3 I FATTORI CHE INCIDONO SULLA VERIFICA DEL MODELLO

In realtà gli errori di misurazione non sono dovuti tutti al campionamento degli item. Sono molti i fattori che influenzano la verifica del modello e, conseguentemente, del livello di affidabilità; la loro quantità e la loro tipologia dipendono dalla natura della caratteristica da misurare e dalla utilizzazione e applicazione che si fa dello strumento. Tra i fattori che possono incidere sul livello di affidabilità occorre considerare anche l'approccio sperimentale adottato; infatti ciascuno degli approcci è sensibile a fonti diverse di variazione nei punteggi individuali, ognuna delle quali produce un errore di misurazione in misura non sempre valutabile. Vediamo quali sono le fonti di variazione per ciascun metodo.
L'errore di stima nel caso dell'approccio delle *componenti parallele* in questo caso è attribuibile sia

all'interazione *soggetto-contenuto dello strumento* che all'interazione *soggetto-condizioni di somministrazione*. In quest'ultimo caso si fa riferimento anche fattori che possono incidere in maniera diversa sugli item, quali la progressiva stanchezza del soggetto durante la somministrazione. Come abbiamo già visto tale problema può in parte essere gestito adottando, per la suddivisione dello strumento, il metodo *odd-even* che consente di attribuire l'influenza di tale fattore in misura uguale alle due parti.

Accanto ai problemi attribuibili all'approccio sperimentale è possibile identificare altri fattori disturbanti tra i quali ricordiamo:

a. limiti di tempo, considerato come *tempo impiegato per sottoporre lo strumento* e inteso anche come accuratezza di esecuzione: maggiore è il tempo, maggiore è l'affidabilità. Come si può intuire tale fattore è comunque molto legato al numero di item impiegati. In ogni caso, superato un certo limite di tempo ottimale, l'affidabilità tende comunque a diminuire.
b. Caratteristiche degli item ovvero
    - *omogeneità degli item*: maggiore è l'*omogeneità*, maggiore è l'affidabilità;
    - *dipendenza tra item*: item troppo interdipendenti tendono a ridurre l'affidabilità in quanto rispetto a tali item i soggetti tendono a comportarsi nello stesso modo; tale situazione presenta lo stesso effetto della diminuzione del numero di item;
    - *capacità discriminante degli item*: minore è l'estensione del continuum misurato da ciascun item, minore è l'affidabilità; questo fattore è collegato con il successivo.
c. Qualità dello *scoring* inteso come
    - *oggettività e accuratezza di attribuzione dei punteggi agli item*: i giudizi soggettivi introducono maggiore variabilità; da ciò si può dedurre che maggiore è l'oggettività, maggiore è l'affidabilità; la mancanza di accuratezza è comunque difficilmente controllabile;
    - *livello di casualità delle risposte agli item*: si tratta della classica posizione di chi *tira a indovinare* ovvero di chi nel rispondere agli item sceglie la risposta a caso. Ciò causa alcune variazioni nel risultato da item a item ed abbassa l'affidabilità di tutto lo strumento; maggiore è la probabilità di rispondere bene (o comunque nella direzione della dimensione misurata) anche casualmente, minore è l'affidabilità; ricordiamo che la casualità è maggiore negli item dicotomici;
    - *response set*, ovvero tendenza di un soggetto a rispondere nello stesso modo a tutti gli item.
d. Eterogeneità/omogeneità del campione della sperimentazione: nella stima dell'affidabilità è importante considerare le caratteristiche sia della popolazione da cui è stato estratto il campione sul quale viene effettuato l'esperimento di affidabilità che della popolazione sulla quale lo strumento verrà utilizzato in pratica, tenendo presente che maggiore è l'omogeneità dei soggetti che compongono il campione, maggiore risulta essere l'affidabilità dello strumento e minore sarà la possibilità di estendere l'utilizzo dello strumento.
e. Livello di comprensione e interpretazione degli item: la riduzione dell'affidabilità può essere dovuta anche a cattive interpretazioni dell'item soprattutto quando:
    - le domande presentano particolari valenze emozionali;
    - le istruzioni sono inadeguate o difettose;
    - le domande presentano tranelli, trucchetti o inganni linguistici.

    Ecco perché occorre prestare molta attenzione nella scelta delle parole e delle espressioni; per esempio item espressi in una forma verbale sintetica e con costruzioni sintattiche semplici risultano di solito i migliori.
f. Altri fattori, così sintetizzati:
    - velocità di somministrazione;
    - velocità individuale nel rispondere;
    - distribuzione del tempo per rispondere agli item;
    - predisposizione soggettiva alla precisione;

- reazione ad elementi e stimoli disturbanti, per esempio alla stanchezza;
- persistenza di determinati atteggiamenti mentali o emozionali (positivi o negativi) durante tutto il tempo di somministrazione;
- malattie, preoccupazioni, eccitamenti, ecc.
- livello di apprendimento del soggetto.

Per esempio un soggetto potrebbe avvertire improvvisamente mal di testa a metà della somministrazione, potrebbe non accorgersi di non aver risposto a determinati item, potrebbe accorgersi a metà di aver compreso male le istruzioni per rispondere, ecc.

1.3.1 Numero ottimale di item

E' difficile conoscere in anticipo quanti item devono essere utilizzati per la costruzione di uno strumento. Per determinare il numero di item da impiegare esistono alcune regole empiriche come quella che ritiene 30 item dicotomici in grado di raggiungere un buon livello di consistenza interna. Non sempre però è possibile giungere a definire 30 item, a causa di problemi di applicabilità dello strumento ma anche per problemi di definizione dell'area stessa. In genere se gli item presentano una scala di risposta a più livelli per ottenere lo stesso livello di affidabilità è sufficiente un numero inferiore di item (a volte 10 item con una scala di risposta a sette punti possono raggiungere un'affidabilità di .80).
In fase di messa a punto dello strumento, se si conosce molto poco sull'omogeneità degli item, è possibile seguire una delle seguenti strategie:

- cominciare con un numero di item più grande per eventualmente eliminarli in fase di verifica del modello di scaling (*item analysis*),
- utilizzare un numero di item minore di quello ritenuto adeguato e di aggiungere via via nuovi item fino a quando si osserva un aumento dell'affidabilità della scala.

A tale proposito occorre tenere presente che l'aggiunta o l'eliminazione di un item comporta non solo la verifica dello strumento di misurazione ma anche della sua struttura concettuale; può essere interessante confrontare gli item scartati con quelli conservati: se gli item eliminati sono differenti nel contenuto è necessario riconsiderare la definizione della caratteristica misurata dai rimanenti item.
Contemporaneamente è necessario valutare anche la dimensione del campione di soggetti su cui verificare il modello; una regola empirica è quella di disporre di un numero di osservazioni uguale a cinque/dieci volte il numero di item.

1.3.2 Ulteriori verifiche, valutazioni e controlli

L'osservazione di un valore d'affidabilità soddisfacente non garantisce la reale verifica del modello di *scaling* e in primo luogo dell'unidimensionalità della scala; se per esempio si combinano due gruppi di item che misurano due diversi costrutti, tra loro correlati e con alti valori di affidabilità, è possibile ottenere ugualmente un buon livello di consistenza interna pur essendo il modello diventato, nel frattempo, bidimensionale.
Tutti i passaggi dell'analisi devono consentire di rivalutare tutti i momenti che hanno condotto ai risultati, dalla formulazione degli item al campione utilizzato per la validazione, alla definizione del costrutto.
Una volta terminato il procedimento di validazione della scala è possibile passare ad un ulteriore stadio di sviluppo della scala. E' possibile ripetere l'analisi su dati rilevati su un nuovo campione, per verificare nuovamente il livello di affidabilità e di validità della scala. La disponibilità di più stime di affidabilità, ottenute su diversi campioni, consente di generalizzare l'affidabilità della scala

a settori più ampi di soggetti. Un'altra buona abitudine è quella di calcolare sempre il coefficiente di affidabilità ad ogni applicazione, anche nel caso di scale già validate.

1.4 I LIMITI DEL MODELLO ADDITIVO

Il modello additivo ha trovato e trova larga applicazione nella ricerca sociale; ciò è dovuto sia alla sua base logica chiara che alle sue semplici procedure di applicazione e di verifica. Esso però è stato oggetto di critiche per diversi motivi.

Esso dipende interamente dall'assunto di fluttuazioni casuali tra gli item che vengono sommati per creare la scala. I criteri stabiliti per selezionare gli item (*item analysis*) utilizzati nelle tecniche *Likert* e *Thurstone* possono essere visti come strategie per cercare di assicurare la casualità delle differenze interitem. Nel caso in cui sia possibile soddisfare tale assunto l'approccio additivo risulta essere una tecnica di *scaling* molto potente. D'altra parte il modello additivo presenta due limiti prodotti dallo stesso assunto:

a. il metodo assume che in un insieme di punti tutti gli errori sono attribuibili a fluttuazioni casuali. Tale fluttuazioni potrebbero però verificarsi per altre ragioni come l'influenza simultanea di più dimensioni sottostanti. Il modello additivo però non considera a priori la multidimensionalità come presenza possibile; infatti una delle critiche rivolte a questo modello riguarda proprio il metodo di valutazione della reale <u>unidimensionalità</u> del gruppo di item in quanto basato su assunti troppo deboli[16]; in termini strettamente pratici, ciò significa che la verifica del modello di scala potrebbe produrre un buon adattamento anche quando le vere fonti di variazioni sono più di una. Ciò significa che tale approccio è *molto utile come tecnica di scaling* ma *molto carente come criterio di scaling*.
b. con l'approccio additivo è possibile scalare solo un unico insieme di punti tra i due che di solito costituiscono i dati del tipo *stimolo-unico*. Ciò è dovuta al fatto che, assumendo che il secondo insieme di punti varia in modo casuale, le stime precise per le posizioni dei punti appartenenti al secondo insieme non avrebbero significato.

Un altro appunto rivolto a tale metodo riguarda l'approccio che il modello additivo utilizza per la valutazione dell'adeguatezza (in termini di affidabilità e validità) dello strumento in quanto troppo legata alle informazioni ricavate dai dati empirici.

Un altra caratteristica criticata riguarda la definizione delle <u>scale di risposta</u> troppo rigidamente vincolata per tutti gli item (si richiede infatti lo stesso numero di livelli di risposta per tutti gli item).

Ma la critica più severa rivolta all'approccio prende spunto dal fatto che secondo il modello additivo, richiamandosi al modello classico di misurazione, la dimensione da misurare è espressa come <u>punteggio vero</u> definito come *il valore atteso di una caratteristica presente in un certo soggetto*. Tale dimensione misurata in un soggetto è definita solamente in funzione di un particolare strumento di misurazione. Se per esempio la dimensione misurata è rappresentata da una capacità, lo strumento sarà considerato:

- *difficile* quando i soggetti risultano avere bassa capacità;
- *facile* quando molti soggetti risulteranno avere alta capacità.

Analogamente definendo la *difficoltà* come la *proporzione di soggetti che risponde correttamente ad un item*, ne consegue che:

- la *facilità/difficoltà* di un item dipende dalla <u>capacità</u> dei soggetti misurati,
- la *capacità* di un soggetto dipende dalla <u>facilità</u> o <u>difficoltà</u> degli item.

[16] Vedremo infatti come una verifica indiretta dell'unidimensionalità è l'analisi della consistenza interna, basata sulle correlazioni tra ciascun item e il punteggio totale (correlazione *item-totale*); tale criterio è considerato insufficiente come prova di unidimensionalità in quanto non è in grado di registrare l'eventuale presenza di due o più sottoinsiemi di item corrispondenti a sottodimensioni della caratteristica misurata.

Ciò vuol dire che la verifica dell'affidabilità (e quella della validità) dello strumento è troppo legata al campione utilizzato per la sperimentazione con la conseguenza che

- le caratteristiche di uno strumento cambiano in funzione del gruppo di soggetti (*group-dependent*),
- le caratteristiche di un soggetto cambiano in funzione degli strumenti di misurazione (*test-dependent*).

Conseguentemente è molto difficile confrontare

- soggetti misurati con strumenti diversi,
- item per la cui validazione sono stati utilizzati campioni per la sperimentazione composti da soggetti con caratteristiche diverse.

Proviamo a questo punto a riassumere i limiti riscontrati che hanno condotto ad una valutazione insoddisfacente dell'approccio additivo:

- *Group-dependent*: la dipendenza delle caratteristiche degli item dai campioni di soggetti utilizzati per la sperimentazione rende gli indici di uso limitato nella pratica, soprattutto nei casi in cui le caratteristiche di tali campioni sono diverse rispetto a quelle della popolazione sulla quale verrà utilizzato lo strumento costruito e validato.
- *Test-dependent*: è difficile confrontare direttamente tra punteggi individuali ottenuti con strumenti diversi non essendo sempre possibile osservare e identificare relazioni funzionali tra due strumenti[17]. Se i soggetti presentano diversi livelli, per esempio, di capacità (lo strumento è più difficile per un gruppo che per un altro), i punteggi osservati contengono una diversa quantità di errore, ovvero sono affidabili in diversa misura[18]. Per ottenere informazioni più precise sulla reale capacità di un soggetto sarebbe quindi necessario osservare il risultato ottenuto in ciascun item. La difficoltà di confrontare punteggi individuali ottenuti con strumenti diversi è dovuta anche alla diversa precisione nel misurare la dimensione. Per superare il problema della presenza di livelli di versi di errori di misurazione sarebbe necessario stabilire diversi livelli rispetto alla dimensione, per esempio livelli diversi di capacità e di difficoltà.
- *Definizione degli strumenti paralleli*: l'affidabilità è valutata sulla base di correlazioni tra punteggi ottenuti con forme parallele che sono difficili, se non impossibili, da definire;
- *Stime di affidabilità con significatività ignota*: la valutazione dell'affidabilità è effettuate con coefficienti che forniscono stime che non possono essere sottoposte a verifica della significatività statistica;
- *Errore di misurazione uguale*: il modello classico assume che l'errore di misurazione sia funzione dell'affidabilità e della varianza della distribuzione dei punteggi e che, conseguentemente, sia uguale per tutti i soggetti; tale assunto però non è accettabile in quanto, come abbiamo visto, ciascuno strumento misura in modo impreciso e in modo disuguale soggetti di capacità differenti.

Un ultimo limite è dato dal fatto che l'approccio additivo è *test-oriented* piuttosto che *item-oriented*; infatti l'adozione del concetto di punteggio totale non consente alcuna valutazione delle risposte individuali ai singoli item; ciò vuol dire che a partire dal punteggio totale individuale non è possibile stimare le risposte ai singoli item.

---

[17] Tale limite è uguale a quello riscontrato per il modello sperimentale degli *strumenti paralleli*.

[18] Il punteggio zero ottenuto da un soggetto indica che il soggetto presenta un basso livello nella capacità misurata ma non fornisce alcuna informazione su quanto esattamente basso soprattutto se confrontato con lo stesso punteggio ottenuto da un altro soggetto.

1.5 UN ESEMPIO DI MODELLO ADDITIVO MULTIDIMENSIONALE: IL DIFFERENZIALE SEMANTICO[19]

1.5.1 Obiettivo del modello: la misura del significato

Nella definizione del significato di un concetto è possibile distinguere due aspetti:
- l'aspetto denotativo, che è uguale per tutti in quanto oggettivo,
- l'aspetto connotativo che invece cambia da soggetto a soggetto, in quanto legato alle reazione emotive ed affettive che ciascun oggetto evoca a livello individuale.

In molte situazioni di ricerca l'interesse è orientato a conoscere il secondo aspetto per poter confrontare tra loro gli individui.
Esistono più modi per rilevare il significato attribuito ad uno stimolo. Quello più semplice consiste nel chiedere direttamente al soggetto il significato. Un'altra modalità è quella richiedere parole in associazione a quella indicata come stimolo. In entrambi i casi, però, ci si trova ad affrontare i classici problemi legati alla rilevazione di dimensioni soggettive (Maggino, 2005); in particolare quelli legati al *respondent* (livello culturale, di personalità e di esperienza) e alla necessità di quantificare l'osservazione e di confrontare i risultati ottenuti.
Ciò si ottiene ponendo particolare attenzione nel costruire uno strumento che consente di misurare il significato attribuito ad uno stimolo attraverso una procedura di misurazione standardizzata.
I primi a porsi il problema di come misurare il significato connotativo di uno stimolo furono Osgood e i suoi collaboratori dell'Università dell'Illinois alla fine degli anni 50.
Secondo Osgood il significato affettivo[20] di un concetto è multidimensionale[21]. Tali dimensioni definiscono il cosiddetto *spazio semantico* la cui struttura è stabile mentre la collocazione del concetto in tale spazio varia tra soggetti. Per poter confrontare i soggetti, è necessario uno strumento che consenta di misurare il significato affettivo individuale attribuito al concetto. Tale strumento è il *Differenziale Semantico*, che consente di quantificare l'aspetto connotativo del significato attribuito ad uno stimolo senza porre domande dirette.
Il Differenziale Semantico è costituito da una serie di scale, ciascuna delle quali è composta da una coppia di aggettivi bipolari tra i quali è collocata una scala di *rating* (a 5 o 7 posizioni).[22]
Ogni scala rappresenta una componente del significato (dimensione dello spazio semantico) e può essere raffigurata come una retta passante per l'origine di questo spazio. Il numero di dimensioni che definisce tale spazio è teoricamente finito, ma sconosciuto. Il concetto è collocato soggettivamente in tale spazio da un numero finito di coordinate ciascuna delle quali corrisponde

---

[19] I rimandi al lavoro di Osgood fatti in questo capitolo si riferiscono al classico testo Osgood et al., 1957.

[20] Per significato «affettivo» Osgood intende il grado di disposizione favorevole o sfavorevole e in generale tutte le connotazioni di tipo emotivo che costituiscono una parte notevole del significato di un concetto al di là delle sue caratteristiche denotative (Arcuri, 1974).

[21] Osgood definisce il significato come un "processo di mediazione di rappresentazione": "ogni volta che lo stimolo, oltre che il significato, è contiguo con il significato, s'incrementa l'associazione con alcune parti del comportamento totale ricavato dal significato come un processo di mediazione di rappresentazione". Secondo Osgood:
1. i significati producono nell'individuo certi comportamenti;
2. i segni dei significati evocano nell'organismo dei processi di mediazione, che sono una parte del comportamento totale provocato dal significato;
3. il processo di mediazione e di rappresentazione è il significato, che è la condizione iniziale per produrre una risposta al segno, vale a dire un comportamento.

[22] Come abbiamo visto, date le sue caratteristiche, tale scala può essere classificata tra le scale non-comparative discrete.

alla posizione che ciascun soggetto attribuisce al concetto su ciascuna scala. Tale *punto* rappresenta il *significato affettivo dell'oggetto*. Dato che soggetti diversi possono attribuire significati affettivi diversi, ciascun oggetto può trovare collocazioni diverse nello spazio semantico.
La valutazione che il soggetto fa dell'oggetto per ciascuna scala/dimensione è individuata dalla posizione dell'oggetto rispetto all'origine e viene espressa in termini di

- qualità (*direzione* positiva o negativa)
- intensità (*distanza* dall'origine).

La complessità dello spazio semantico definito da un numero di dimensioni equivalente al numero delle scale, può essere ridotta condensando in *fattori*, ovvero dimensioni tra loro ortogonali, più scale che condividono lo stesso significato.
Secondo Osgood, nello spazio semantico è possibile collocare sia i significati semantici di più soggetti relativamente a un concetto che i significati semantici di più concetti per un soggetto[23].

Osgood ha cercato di verificare empiricamente tale struttura concettuale. In particolare con il suo gruppo ha cercato di verificare la natura e il numero delle dimensioni dello spazio semantico.
Per fare ciò ha condotto una serie di studi disegnati in modo tale da ridurre al minimo le fonti di variabilità derivanti dai soggetti, dalle scale e dai concetti (stimoli).

- *I soggetti*. Il campione dei soggetti, in teoria, sarebbe dovuto essere rappresentativo in maniera trasversale di tutta la popolazione in senso demografico. In pratica, ciò risulta sempre molto difficile e costoso da realizzare, per tanto Osgood è ricorso ad un campione composto da studenti universitari. L'utilizzo di un campione così composto se da una parte presentava lo svantaggio di non poter essere realmente rappresentativo della variabilità della popolazione teorica di riferimento, dall'altra presentava il vantaggio di essere composto da soggetti che avevano dimestichezza nel trattare concetti astratti e nell'eseguire il compito loro richiesto (risposte ad un questionario).
- *Gli aggettivi*. Inizialmente, il differenziale semantico somministrato nella pratica era composto da scale definite da coppie di aggettivi selezionati tra termini considerati di uso frequente tra gli studenti universitari. In studi successivi la definizione delle coppie ha fatto riferimento al Roget's Thesaurus, un dizionario dei sinonimi e contrari.
- *I concetti*. I concetti sottoposti ai soggetti erano scelti in modo tale che risultassero il più possibile diversi tra di loro ma allo stesso tempo familiari ai soggetti.
- *La rilevazione*. L'ordine di presentazione dei concetti e delle scale per ciascuno di essi non era casuale ma era "ruotato" in modo da controllare l'eventuale presenza di una variabilità causata dalla sequenza di somministrazione.
- *L'analisi dei dati*. I dati sono stati elaborati utilizzando tecniche di analisi fattoriale diverse (*Centroid factorization/Thurstone's Centroid Factor Method*; *D-Factorization*; *Centroid Factor Analysis*, *Quartimax Rotation of the centroid*, *Square Root Factorization*).

I risultati delle analisi confermarono la multidimensionalità dello spazio semantico e consentirono di attribuire un significato a ciascuna dimensione individuata. Confrontando i risultato prodotti dai diversi studi, tra le dimensioni individuate, tre risultarono quelle ricorrenti, denominate *evaluation*,

[23] Osgood osserva che nel caso del confronto tra concetti occorre tenere presente il "Principio di Congruenza": "ogni volta che due segni sono collegati da un'asserzione, la reazione di mediazione caratteristica di uno cambia verso la congruenza con la reazione di mediazione dell'altro; la grandezza del cambiamento è inversamente proporzionale alle intensità delle reazioni dell'interazione". In altri termini, quando due concetti interagiscono simultaneamente si modificano reciprocamente in proporzione alla loro intensità, apportando dei cambiamenti nel significato o creando nuove risoluzioni combinando i significati. Le interazioni sono tali che i processi di mediazione dei segni in questione sono modificati fino ad essere reciprocamente congruenti.

potency e *activity*.[24] La relazione tra tali dimensioni risultò essere sempre la stessa. In particolare, la dimensione *evaluation* risultava essere sempre la prima componente in termini di proporzione di varianza spiegata (generalmente i tre quarti), mentre successivamente risultavano le dimensione *potency* e *activity* con proporzioni di varianza spiegata simili. La presenza di una porzione non rilevante di varianza residua non spiegata ha consentito di concludere che tali dimensioni dominanti non esauriscono la composizione dello spazio semantico. Tali dimensioni sono risultate tra loro indipendenti (le valutazioni dei soggetti su una dimensione non erano correlate con quelle fatte per le altre dimensioni).

1.5.2 Costruzione del differenziale semantico

Gli studi svolti dal gruppo di Osgood consentirono di verificare non solo la bontà della struttura concettuale ipotizzata, ma anche di mettere a punto il differenziale semantico come strumento di misura in termini di obbiettività, affidabilità e validità (Maggino, 2004).
D'altra parte il Differenziale Semantico non si presenta come uno strumento di misura standardizzato ma come un approccio alla misurazione. In questo senso, ogni volta che si intende utilizzare tale approccio è necessario procedere alla costruzione dello strumento in funzione del contesto e degli obbiettivi conoscitivi specifici (Camozza, 1977;).
La costruzione avviene secondo le seguenti fasi:

- la selezione dei concetti/stimoli,
- la formulazione dell'ipotesi dimensionale,
- la definizione delle scale (coppie di aggettivi e scala di *rating*)
- la scelta della modalità di presentazione,
- la definizione della procedura di somministrazione,
- la definizione della procedura di analisi dei dati.

1.5.2.1 Selezione dei concetti/stimoli

L'obiettivo del differenziale semantico - nell'ottica di Osgood - era quello di riuscire a misurare il significato di concetti. L'approccio di misura da lui messo appunto può essere utilizzato anche per misurare il significato di oggetti, parole, personaggi, eventi, ciascuno dei quali può essere operativamente detto "stimolo". La selezione degli stimoli deve tener conto, naturalmente, della loro *rappresentatività* e *pertinenza* rispetto agli obiettivi della ricerca. Nel caso in cui gli obbiettivi conoscitivi richiedono il confronto delle valutazioni individuali tra più oggetti, nella selezione degli stimoli deve tenere conto dei seguenti requisiti:

- gli stimoli selezionati devono essere *univocamente* identificati da tutti i soggetti coinvolti nella rilevazione;
- gli stimoli selezionati devono risultare *familiari* a tutti i soggetti che devono eseguire le valutazioni; la presenza di stimoli non familiari può introdurre un errore di misurazione difficilmente individuabile (per esempio un *response set*);
- gli stimoli selezionati devono essere tra loro chiaramente distinti;
- l'insieme degli stimoli selezionati deve essere rappresentativo del fenomeno che si intende studiare.

---

[24] Ciascuna di tali dimensioni risultava caratterizzata da specifiche coppie di aggettivi, per esempio la dimensione *evaluation* è descritta dalle coppie "buono-cattivo", "bello-brutto", la dimensione *potency* da "potente-impotente", "forte-debole", la dimensione *activity* da "veloce-lento", "attivo-passivo".

1.5.2.2 Formulazione dell'ipotesi dimensionale

Come abbiamo visto, i risultati degli studi di Osgood consentirono di identificare un unico spazio semantico stabile, definito indipendentemente dallo stimolo. Le tre dimensioni dello spazio semantico identificato attraverso gli studi di Osgood sono però intimamente legate alla sua ipotesi di lavoro e ai concetti da lui studiati. Ciò vuol dire che ogni volta che si vuole utilizzare il differenziale semantico come approccio di misura occorre riformulare e rivalutare la dimensionalità dello spazio semantico in relazione agli obiettivi e conseguentemente agli stimoli. Per questo motivo appare ragionevole, prima di procedere alla definizione delle scale, formulare una ipotesi relativa alle dimensioni che definiscono lo spazio semantico e attraverso le quali i soggetti valuteranno lo stimolo.

1.5.2.3 Definizione delle scale

Come abbiamo detto, ciascuna scala che andrà a definire lo strumento di misura è individuata da una coppia di aggettivi e da una scala di *rating* (Heise, 1970). Mentre per questa il ricercatore deve stabilire unicamente la dimensione (5 o 7 posizioni), nella selezione delle coppie di aggettivi è importante considerare alcuni criteri; in particolare la scelta deve tenere conto:

- della *attinenza degli aggettivi agli stimoli selezionati* per ottenere misurazioni più sensibili;
- della *familiarità degli aggettivi ai soggetti che compongono il campione*;
- della *effettiva bipolarità della coppia di aggettivi* che devono essere realmente opposti nel loro significato;
- della *neutralità degli aggettivi selezionati*: i soggetti del campione non devono avere la percezione che gli aggettivi siano orientati verso particolari giudizi di valore.

Un'altra questione che si pone riguarda il numero di scale (coppie di aggettivi) da definire. Al riguardo non esistono regole ma è possibile individuare alcuni suggerimenti che fanno dipendere il numero di scale:

- dal numero di concetti/stimoli da valutare,
- dal numero di dimensioni ipotizzate (per ciascuna dimensione è auspicabile identificare un minimo di 4-5 scale),
- dal contesto in cui è inserito lo strumento (se il differenziale semantico è inserito all'interno di un questionario lungo e complesso oppure se è somministrato congiuntamente ad una scheda di rilevazione sintetica e semplice),
- dalle caratteristiche dei soggetti che devono rispondere (relativamente all'età, al livello di istruzione, ecc.),
- dalle modalità di somministrazione (rispetto alla presenza o meno di un rilevatore).

1.5.2.4 Scelta delle modalità di presentazione

Il differenziale semantico può essere presentato in diversi formati.
Osgood identifica diversi formati ottenuti dalla combinazione dell'ordine di presentazione dei concetti/stimoli e delle coppie di aggettivi; ritiene inoltre che le diverse modalità di presentazione non apportino alcuna alterazione del risultato. Di seguito si riportano i formati da lui proposti.

- Ciascuna coppia di aggettivi è somministrata relativamente a tutti gli stimoli selezionati.

Concetto / Stimolo **A**	Aggettivo "a" |___|___|___|___|___|___|___| Aggettivo "b"
Concetto / Stimolo **B**	Aggettivo "a" |___|___|___|___|___|___|___| Aggettivo "b"
Concetto / Stimolo **C**	Aggettivo "a" |___|___|___|___|___|___|___| Aggettivo "b"

Tale modalità dovrebbe consentire di evitare forme di *response set*.

- La serie delle coppie di aggettivi è somministrata per ciascuno stimolo:

Concetto / Stimolo **A**
Aggettivo "a" |___|___|___|___|___|___|___| Aggettivo "b"
Aggettivo "c" |___|___|___|___|___|___|___| Aggettivo "d"
Concetto / Stimolo **B**
Aggettivo "a" |___|___|___|___|___|___|___| Aggettivo "b"
Aggettivo "c" |___|___|___|___|___|___|___| Aggettivo "d"

L'ordine di presentazione dei concetti non è rilevante; al contrario, la sequenza e la polarità delle coppie di aggettivi rimangono le stesse per tutti gli stimoli.
Questo formato rende più semplice il compito dei soggetti che devono valutare e le successive fasi di codifica, trattamento e analisi dei dati.

- Ogni oggetto si trova sulla stessa linea della scala su cui deve essere valutato; i concetti e le scale sono combinati a caso e, altrettanto casualmente, sono disposti uno dopo l'altro.

Concetto / Stimolo "**C**"	Aggettivo "x" |___|___|___|___|___|___|___| Aggettivo "y"
Concetto / Stimolo "**F**"	Aggettivo "k" |___|___|___|___|___|___|___| Aggettivo "w"

Questo formato ha il vantaggio di minimizzare l'effetto alone in quanto, essendo sia gli stimoli che le coppie di aggettivi di volta in volta diversi, si richiede un livello di attenzione sempre alto.

In realtà le modalità di presentazione possibili sono moltissime; quella proposta da Osgood è solo una semplice classificazione. Si ritiene importante precisare che la modalità di presentazione del differenziale semantico non è il risultato di una scelta "finale" del ricercatore, ma è l'esito di un percorso: viene, infatti, identificata durante le fasi di costruzione del differenziale semantico. In corrispondenza di ciascuno dei passaggi necessari per la messa a punto dello strumento è necessario effettuare una breve riflessione anche sul modo in cui si ritiene che ciascun elemento costitutivo possa essere meglio rappresentato, valutando sempre anche gli obbiettivi conoscitivi e il disegno dell'indagine.

1. Selezione dei concetti/stimoli
   Dopo la selezione dei concetti/stimoli rispetto al formato di presentazione è necessario decidere:
   - la modalità di presentazione: riprendendo i suggerimenti di Osgood è necessario valutare l'opportunità di presentare tutte le scale da valutare in relazione ad un concetto oppure quella di presentare una scala da valutare per tutti i concetti;
   - l'ordine di presentazione: presenza o meno di un ordine ed eventualmente quale (in base a specifici criteri o seguendo convenzioni p.e. ordine alfabetico o crescente/decrescente).
2. Formulazione dell'ipotesi dimensionale
   In seguito alla identificazione delle dimensioni da valutare rispetto al formato di presentazione è necessario decidere:
   - l'ordine di presentazione: presenza o meno di un ordine ed eventualmente quale (in base a specifici criteri o seguendo convenzioni p.e. ordine alfabetico o crescente/decrescente);
   - la sequenza delle dimensioni: può rimanere fissa per tutti i concetti oppure essere modificata.

3. Definizione delle scale

   La definizione delle scale comprende sia l'identificazione delle coppie di aggettivi che quella della scala di *rating*.

   Rispetto alla coppie di aggettivi è possibile scegliere:
   - l'ordine di presentazione: presenza o meno di un ordine ed eventualmente quale (in base a specifici criteri o seguendo convenzioni p.e. ordine alfabetico);
   - la sequenza delle coppie di aggettivi: può rimanere fissa per tutti i concetti e per tutte le dimensioni oppure essere modificata.
   - la polarità della coppia di aggettivi: può rimanere fissa oppure essere invertita.

   Rispetto alla scala di *rating* è necessario determinare:
   - la dimensione: generalmente si sceglie una scala da 5 o 7 posizioni;
   - la modalità di rappresentazione. Vediamo di seguito l'elenco di tutto le possibili modalità di rappresentazione:

     Scala di rating semplice rappresentata con un continuum tra la coppia di aggettivi bipolari senza alcuna indicazione specifica.

Aggettivo "a"	|___|___|___|___|___|___|___|	Aggettivo "b"
Aggettivo "c"	|___|___|___|___|___|___|___|	Aggettivo "d"
Aggettivo "e"	|___|___|___|___|___|___|___|	Aggettivo "f"

Scala di rating con attribuzione di un simbolo alfabetico alle diverse posizioni.

	A	B	C	D	E	F	G	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
Aggettivo "c"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "d"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "f"

Scala di rating con attribuzione di etichette verbali alle diverse posizioni per agevolare la valutazione in termini di intensità.

	Molto	Abba-stanza	Poco	Neutro	Poco	Abba-stanza	Molto	
Aggettivo "a"								Aggettivo "b"
Aggettivo "c"								Aggettivo "d"
Aggettivo "e"								Aggettivo "f"

Scala di rating con indicazione numerica delle posizioni dal valore minimo in corrispondenza dell'aggettivo a destra al valore massimo in corrispondenza dell'aggettivo a sinistra.

	1	2	3	4	5	6	7	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
Aggettivo "c"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "d"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "f"

Aggettivo "a"	①	②	③	④	⑤	⑥	⑦	Aggettivo "b"
Aggettivo "c"	①	②	③	④	⑤	⑥	⑦	Aggettivo "d"
Aggettivo "e"	①	②	③	④	⑤	⑥	⑦	Aggettivo "f"

Scala di rating con indicazione numerica delle posizioni in relazione alla distanza dalla posizione neutrale centrale a cui corrisponde il valore 0: le posizioni più estreme rispetto a quella centrale (più vicine agli aggettivi) sono identificate dal valore più alto, mentre quelle più vicine alla posizione neutrale centrale dal valore più basso. Ai valori che identificano la *distanza dal centro viene associato un segno (positivo o negativo) per indicare la direzione di questa distanza: per convenzione positiva a destra e negativa a sinistra.*

	-3	-2	-1	0	+1	+2	+3	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
Aggettivo "c"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "d"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "f"

Aggettivo "a"	❸	❷	❶	○	①	②	③	Aggettivo "b"
Aggettivo "c"	❸	❷	❶	○	①	②	③	Aggettivo "d"
Aggettivo "e"	❸	❷	❶	○	①	②	③	Aggettivo "f"

Scala di rating con identificazione delle diverse posizioni in base alla graduazione di colore: dalla graduazione più tenue in corrispondenza dell'aggettivo a sinistra alla graduazione più forte in corrispondenza dell'aggettivo a destra.

Aggettivo "a"								Aggettivo "b"
Aggettivo "c"								Aggettivo "d"
Aggettivo "e"								Aggettivo "f"

Scala di rating con identificazione delle diverse posizioni in base alla graduazione di colore in relazione alla distanza dalla posizione neutrale centrale (bianca): le posizioni più estreme rispetto a quella centrale (più vicine agli aggettivi) sono identificate dalla graduazione di colore più forte, mentre quelle più vicine alla posizione neutrale centrale dalla graduazione più tenue.

Aggettivo "a"								Aggettivo "b"
Aggettivo "c"								Aggettivo "d"
Aggettivo "e"								Aggettivo "f"

Nella presentazione del differenziale semantico può essere necessario o semplicemente utile fare ricorso ad alcune accortezze grafiche che facilitino la corretta lettura delle diverse combinazioni di aggettivi. Il differenziale semantico è infatti uno strumento di misurazione particolarmente soggetto ad errori di compilazione proprio per la caratteristica struttura e complessità: si consideri, infatti, che a seconda dell'ipotesi dimensionale formulata il differenziale semantico può comprendere un numero molto elevato di coppie di aggettivi. Adottare alcune accortezze grafiche consente di contenere gli errori dovuti alla difficoltà di combinare correttamente le coppie di aggettivi.

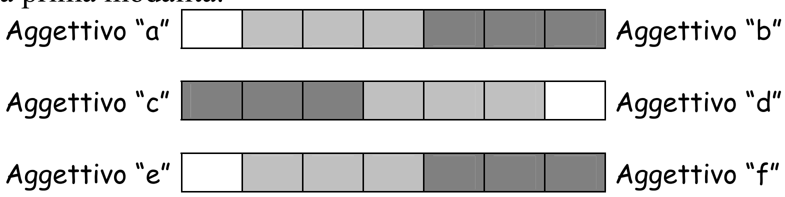
Quando la rappresentazione della scala non prevede alcun colore, questi possono essere utilizzati per identificare diverse righe e quindi diverse coppie di aggettivi. Di seguito si riportano alcuni esempi.

Aggettivo "a"	|___|___|___|___|___|___|___|	Aggettivo "b"
Aggettivo "c"	|___|___|___|___|___|___|___|	Aggettivo "d"
Aggettivo "e"	|___|___|___|___|___|___|___|	Aggettivo "f"

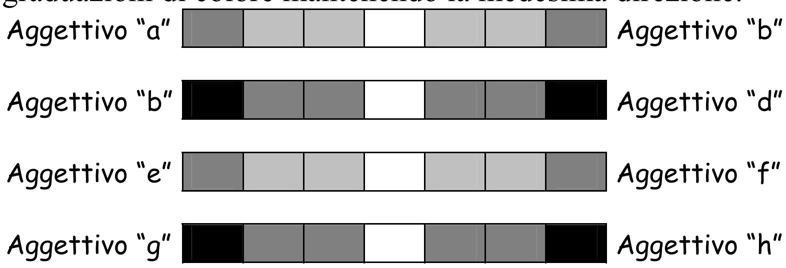
	1	2	3	4	5	6	7	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
Aggettivo "c"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "d"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "f"

	-3	-2	-1	0	+1	+2	+3	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
Aggettivo "c"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "d"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "f"

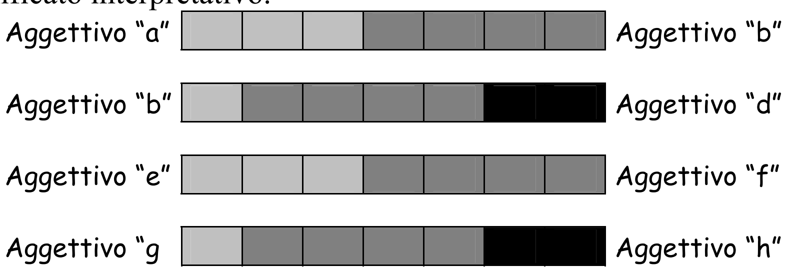
Anche quando il colore costituisce un elemento costitutivo della scala può comunque essere utilizzato per facilitare la lettura delle diverse coppie di aggettivi. Per esempio si può decidere di alternare la graduazione di colore: per la prima coppia di aggettivi la direzione graduazione più tenue a sinistra - graduazione più forte a destra, per la seconda coppia la direzione graduazione più forte a sinistra - graduazione più tenue a destra, per la terza coppia si riprende la prima modalità.



Un'altra accortezza che costituisce un supporto per la lettura delle combinazioni di aggettivi senza intervenire sulla direzione della graduazione di colore consiste nell'alternanza delle due diverse graduazioni di colore mantenendo la medesima direzione.



Questa strategia è funzionale anche nei casi in cui alla direzione sia stato attribuito un preciso significato interpretativo.



Indipendentemente dall'uso del colore un'altra strategia consiste nell'utilizzare diversi formati del carattere utilizzato per scrivere la coppia di aggettivi. Si può decidere di utilizzare per esempio:

- diversi stili (normale, corsivo, grassetto, ecc.) di un dato tipo di carattere (Comic Sans Ms)

Aggettivo "a"	___	___	___	___	___	___	___	Aggettivo "b"
Aggettivo "c"	___	___	___	___	___	___	___	*Aggettivo "d"*
Aggettivo "e"	___	___	___	___	___	___	___	**Aggettivo "f"**

- lo stesso stile (grassetto) di diversi tipi di carattere (Comic Sans Ms, Trebuchet MS, Berlin Sans FB);

	1	2	3	4	5	6	7	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
Aggettivo "c"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "d"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	**Aggettivo "f"**

- diversi effetti (normale, maiuscoletto, ombreggiatura e contorno) di un dato tipo di carattere (Comic Sans Ms);

	-3	-2	-1	0	+1	+2	+3	
Aggettivo "a"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "b"
AGGETTIVO "C"	☐	☐	☐	☐	☐	☐	☐	AGGETTIVO "D"
Aggettivo "e"	☐	☐	☐	☐	☐	☐	☐	Aggettivo "f"

- diverse altre combinazioni dei precedenti (carattere, stile, effetto).

Si tratta di accortezze il cui obiettivo principale, come è stato ampiamente detto, consiste nel contenimento dell'errore di rilevazione dovuto alla difficoltà di combinare correttamente le coppie di aggettivi, pertanto il loro uso deve risultare di supporto alla lettura e non affaticarla ulteriormente.

Di seguito si propone uno schema che sintetizza le riflessioni e le scelte connesse a ciascuna fase di costruzione del differenziale semantico.

Fase di costruzione/Elementi costitutivi del differenziale		Variabili e relative opzioni che intervengono nella modalità di presentazione	
Concetti		Modalità	Uno per volta
			Tutti insieme
		Ordine	Presenza di un ordine
			Casuale
Dimensioni		Ordine	Presenza di un ordine
			Casuale
		Sequenza	Fissa
			Variabile
Scale	Coppia di aggettivi	Ordine	Presenza di un ordine
			Casuale
		Sequenza	Fissa
			Variabile
		Polarità	Fissa
			Variabile
	Scala di *rating*	Dimensione	5 posizioni
			7 posizioni
		Rappresentazione	*Continuum* semplice
			Simbolo
			Etichetta verbale
			Valore numerico
			Grafica (colore)

1.5.2.5 Definizione della procedura di somministrazione

La somministrazione del differenziale semantico richiede che il soggetto indichi la posizione più o meno vicina all'aggettivo di ciascuna coppia che ritiene più adatto a descrivere lo stimolo in questione, pensando allo stimolo da valutare senza meditare troppo.

La definizione della procedura di somministrazione deve prendere in considerazione innanzitutto la presenza o meno di un rilevatore. Indipendentemente da ciò, la presentazione del differenziale semantico deve essere corredata da semplici e precise *istruzioni* che devono riguardare la richiesta di valutazione dello/degli stimolo/stimoli, la descrizione tecnica dello strumento, le modalità di indicazione della valutazione espressa. Una indicazione utile da specificare a ciascun soggetto riguarda il modo di porsi nella espressione della propria valutazione (il soggetto dovrebbe valutare senza soffermarsi troppo, velocemente, sulla base della prima impressione). Può essere utile anche inserire un esempio che consenta di illustrare le modalità di compilazione.

Se attraverso il differenziale semantico si vuole, per esempio, rilevare valutazioni individuali riguardanti la città in cui il soggetto risiede, è possibile far precedere la presentazione delle scale dalle seguenti indicazioni: *Di seguito è riportata una scheda che contiene una serie di coppie di aggettivi bipolari. Pensando alla sua città, segni una crocetta nella posizione più o meno vicina all'aggettivo che pensa sia più adeguato a descrivere la sua idea della città. Risponda senza soffermarsi troppo.*

1.5.2.6 Definizione della procedura di analisi dei dati

L'analisi dei dati rilevati attraverso il differenziale semantico ha due obbiettivi:

A. verificare la consistenza dello strumento,
B. confrontare i soggetti relativamente alle valutazioni individuali.

(A)

In generale la verifica della consistenza dello strumento prevede una valutazione di:

- la **consistenza dello strumento in termini di struttura dimensionale, la stabilità della struttura rispetto agli stimoli e la stabilità semantica delle coppie di aggettivi rispetto alle dimensioni**;
- l'**effettiva bipolarità delle coppie di aggettivi** utilizzati,
- la **stabilità dello strumento** rispetto alla presenza di *response-set*.

Il principale strumento statistico che consente di realizzare tali verifiche è la analisi fattoriale.

L'analisi fattoriale (Maggino, 2005) ha avuto origine in psicologia sperimentale come insieme di concetti che definiscono un particolare *modello di misurazione*, detto *fattoriale*. Alla base di tale modello vi è l'assunto secondo la quale esistono concetti ipotetici, quali l'intelligenza, la qualità della vita, ecc., non osservabili e misurabili direttamente, e che quindi rappresentano *fattori* o *dimensioni latenti*, misurati attraverso una o più variabili rilevate a loro volta tramite misure multiple. L'applicazione del modello, riducendo la complessità, consente di chiarire e verificare definizioni teoriche costituendo in questo senso uno strumento di verifica del significato di una o più variabili (valore euristico)[25].

Successivamente, il modello fattoriale non solo è stato generalizzato e applicato nei campi più disparati, ma ha prodotto un insieme di procedimenti che, studiati per verificare il modello, sono utilizzati come strumento generale di analisi statistica conosciuto con il termine di *analisi fattoriale*: questa rappresenta un insieme di metodi e tecniche statistiche e di procedure matematiche il cui comune obiettivo è quello di semplificare modelli interpretativi (per esempio dare risposte a domande del tipo *quali sono le cause dell'insuccesso scolastico*) attraverso l'analisi e la rappresentazione contemporanea di molte variabili per verificarne la struttura sottostante (anche in termini di dimensionalità) attraverso l'esame delle interrelazioni. Dati gli obiettivi, l'analisi fattoriale consente anche di sintetizzare le variabili originarie in modo significativo.

[25] Spearman, uno dei primi a formulare un modello di misurazione multifattoriale all'inizio del '900, ha definito e postulato i concetti di "fattore generale" e "fattore specifico" per poter misurare l'intelligenza; in particolare egli ha teorizzato un modello che mirava a descrivere l'intelligenza di un individuo mediante il minor numero possibile di caratteri risultati maggiormente significativi; secondo tale teoria l'attività mentale può essere considerata come l'effetto dell'azione di un fattore generale, rappresentabile lungo un continuum lineare, e di un insieme di altri fattori particolari (memoria, conoscenza, ecc.).

Gli assunti alla base di tale modello possono essere così riassunti:

a. le relazioni tra le variabili sono lineari;
b. la varianza totale nelle variabili è una funzione:
    - dei fattori (*comunanza*),
    - dei *disturbi* caratteristici di ciascuna variabile e degli errori di misurazione (*unicità*);
c. gli errori e i disturbi non sono correlati tra loro né con i fattori;
d. le variabili non sono casualmente correlate tra loro se non attraverso la reciproca relazione con i fattori;
e. la struttura dei fattori osservati riproduce fedelmente la struttura delle dimensioni sottostanti.

Nel verificare la struttura dimensionale spesso viene utilizzata l'analisi delle componenti principali. A tale proposito occorre ammettere che il modello delle componenti principali è simile a quello dei fattori comuni, in quanto entrambi consentono di sintetizzare e riassumere le variabili e cercano di spiegare parte della variazione di un insieme di variabili osservate sulla base di poche dimensioni sottostanti. Tuttavia pur essendo simili, i due approcci non si identificano completamente. Di seguito sono riassunte le principali differenze.

- Componenti Principali: non verifica statisticamente un modello ma semplicemente decompone la matrice di correlazione; l'obiettivo è quello di ridurre la dimensionalità dell'insieme di variabili che vengono sintetizzate attraverso una loro combinazione lineare: la *componente principale è funzione lineare delle variabili*. L'obiettivo non è quindi quello di spiegare le correlazioni tra le variabili, ma di spiegare più varianza possibile.
- Fattori Comuni: ha un modello sottostante che suddivide la varianza totale in varianza comune e varianza unica; l'attenzione è posta principalmente sulla spiegazione della varianza comune, piuttosto che sulla varianza totale; il principale assunto di tale modello definisce *ciascuna variabile come combinazione lineare di fattori comuni non osservabili e di fattori unici*. Non fornisce un'unica soluzione per la trasformazione delle variabili in fattori.

L'approccio dei fattori comuni è più conveniente quando le variabili contengono una quantità di errore di misurazione in quanto il modello prevede la considerazione della varianza unica; l'applicazione del modello delle componenti principali applicato in questi casi produce combinazioni lineari, affette da errori di misurazione non valutabili.

Quando l'obiettivo dell'analisi non è solo quello di verificare la struttura dimensionale ma anche quello di valutare i processi di aggregazione delle coppie di aggettivi, può essere utile analizzare i processi di aggregazione delle coppie di aggettivi attraverso la *Cluster Analysis gerarchica* (Maggino, 2005) che presenta l'ulteriore vantaggio di consentire una interpretazione semplice e immediata dei risultati. L'applicazione di tale analisi risulta appropriata soprattutto quando ciascuna scala viene riferita ad un'unica dimensione.

(B)

Quando la struttura dimensionale risulta stabile e interpretabile, può essere utile confrontare le posizioni che i soggetti assumono rispetto alle dimensioni. Per fare questo è necessario effettuare il calcolo del numero dei punteggi individuali equivalente al numero delle dimensioni. Ciò richiede di sintetizzare per ogni individuo le singole valutazioni riferite a ciascuna scala di ciascuna dimensione. Tali punteggi possono essere definiti in termini di semplici somme dei valori rilevati per ciascuna scala (approccio additivo).

Tale sintesi può però tenere conto anche del diverso peso (*score*) che ciascuna scala ha nel misurare la dimensione. L'analisi fattoriale non consente la determinazione degli score in quanto non presenta un numero sufficiente di stimatori (troppe incognite) nonostante i diversi tentativi fatti negli ultimi decenni per risolvere tale "indeterminatezza". Per questo motivo in questi casi si ricorre all'analisi delle componenti principali (Maggino, 2005) il cui obbiettivo è proprio quello di sintetizzare variabili tra loro molto correlate in nuove variabili sintetiche dette *componenti*.

L'insieme dei punteggi individuali per ciascuna dimensione (sia che siano stati calcolati con approccio additivo o come punteggi di componenti) rappresentano nuove variabili che il ricercatore può utilizzare all'interno di modelli più complessi di analisi (per esempio, analisi tipologica).

L'affidabilità dei punteggi multidimensionali

Nel caso in cui i punteggi individuali vengano calcolati come punteggi componenti può essere importante determinare il livello di affidabilità. Il procedimento non è semplice in quanto occorre tenere presenti alcuni elementi.

In presenza di *loading* differenti, e non uniformi, il punteggio individuale per ciascuna componente deve essere calcolato sommando i punteggi degli indicatori, ciascuno dei quali deve essere pesato per mezzo del *component score* che rappresenta il contributo originale di ciascun indicatore nella definizione di ciascun fattore. Tale peso presenta valori diversi dal *loading*. In questi casi l'affidabilità sarà data dalla seguente equazione

$$\alpha = \frac{\sigma_F^2 - \sum \left(1 - h_i^2\right) w_i^2}{\sigma_{Fs}^2} = \textit{affidabilità generalizzata}$$

dove

w_i *score* relativo all'indicatore *i*

h_i^2 comunanza dell'indicatore *i*

σ_F^2 varianza totale dei punteggi costruiti ovvero $\sigma_F^2 = \sum \sum w_i w_j r_{ij}$

che è equivalente alla somma degli elementi nella matrice di correlazione corretta in cui ciascun elemento è moltiplicato per i rispettivi due pesi w_i e w_j.

Appendice.
ANCORA SULLA STIMA DELL'AFFIDABILITÀ

A.1 L'APPLICAZIONI DEL COEFFICIENTE DI AFFIDABILITÀ

Date le sue caratteristiche concettuali e statistiche, il coefficiente *rho* trova particolari applicazioni in statistica; in particolare

1. nella individuazione degli intervalli di confidenza dei i punteggi veri (stima dell'errore standard di misurazione),
2. nella correzione dei coefficienti di correlazione dovuta alla non affidabilità nella misurazione delle variabili correlate (correzione dell'attenuazione) (Carmines, 1992).

A.1.1 Stima dell'errore standard di misurazione

Come abbiamo visto un'importante proprietà di una misurazione è data dalla dimensione dell'errore ad essa associata; minore è l'errore, più precisa è la misurazione.
Non è però possibile conoscere la dimensione della componente di errore di ciascuna particolare misurazione ma è possibile stimare la dimensione della varianza dell'errore.
Dalla statistica induttiva sappiamo che è possibile stimare l'errore standard utilizzando il coefficiente *r*

$$\textit{errore standard di stima} = \sigma_e^2 = \sigma_y^2\left(1 - r_{yx}^2\right)$$

da cui

$$\sigma_e = \sigma_y\sqrt{\left(1 - r_{yx}^2\right)}$$

Nel caso in cui le variabili sono espresse da punteggi standardizzati le precedenti espressioni divengono

$$\sigma_e^2 = 1 - r_{yx}^2 \text{ da cui } \sigma_e = \sqrt{1 - r_{yx}^2}$$

in quanto per le variabili standardizzate la deviazione standard è uguale a 1.
Occorre osservare che σ_e:

a. ha una relazione inversa con il quadrato della correlazione,
b. consente di determinare gli intervalli di confidenza per stimare i punteggi di una variabile a partire dai punteggi di un'altra variabile,
c. offre un approccio allo sviluppo di molti indici di relazioni tra variabili, per cui

$$\sigma_e^2 = \sigma_y^2\left(1 - r_{xy}^2\right) \qquad\qquad r_{xy}^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2}$$

Così la correlazione è inversamente correlata al rapporto tra σ_e^2 e σ_y^2. Quando questi due hanno lo stesso valore la correlazione è uguale a zero. Quando σ_e^2 è relativamente molto piccola rispetto a σ_y^2 la correlazione è molto alta. Tale rapporto è molto importante e viene indicato come grandezza relativa dell'errore (*RE*) nella stima di una variabile dipendente. Esso è una misura diretta della forza della relazione. Tale concetto può essere esteso a qualsiasi forma di relazione.

Tornando al problema iniziale, un indice tipico della dimensione dell'errore di misurazione può essere rappresentato dalla deviazione standard della distribuzione dei punteggi osservati per ciascun soggetto; tale indice è detto *errore standard di misurazione* (*Standard Error of Measurement, SEM*). Maggiore è l'*errore standard di misurazione* minore è l'affidabilità della rilevazione.

Sappiamo che

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 \text{ da cui } \sigma_e^2 = \sigma_x^2 - \sigma_t^2$$

e che

$$rho_x = \frac{\sigma_t^2}{\sigma_x^2} \text{ da cui } \sigma_t^2 = \sigma_x^2 rho_x$$

Se sostituiamo il termine σ_t^2 della precedente equazione $\sigma_e^2 = \sigma_x^2 - \sigma_t^2$ con quest'ultima $\sigma_t^2 = \sigma_x^2 rho_x$ si ottiene la varianza degli errori di misurazione espressa in termini di varianza e coefficiente di affidabilità dei punteggi osservati:

$$\sigma_e^2 = \sigma_x^2 - \sigma_x^2 rho_x$$

ovvero

$$\sigma_e^2 = \sigma_x^2 (1 - rho_x)$$

L'errore standard di misurazione è quindi

$$\sigma_e = \sigma_x \sqrt{(1 - rho_x)}$$

Rispetto al *coefficiente di affidabilità*, l'*errore standard* di misurazione, presenta il vantaggio di non dipendere dal campo di variazione del punteggio osservato. Maggiore è l'affidabilità minore è il valore di σ_e.

Poniamo di avere $rho_x = .81$ e $\sigma_e = 4$

allora $\sigma_e = 4\sqrt{1 - .6561} = 2.35$

Con un livello di significatività del 5% (e la distribuzione dei punteggi presenta caratteristiche di *normalità*) il vero valore sarà compreso tra $\pm 1.96 \cdot \sigma_e \pm$. Quindi possiamo dire che, in presenza di un punteggio *osservato* di 50, il punteggio *vero* ha il 95% di probabilità di ricadere tra 47.65 e 52.35.

In realtà la stima così calcolata rappresenta una *tipizzazione* della dimensione dell'errore atteso; infatti soggetti diversi presentano errori standard diversi; la stima presentata è valida se la deviazione standard è pressoché la stessa per tutti i soggetti.
Tale specifico e individuale errore standard di misurazione (*PSEM*) può essere diverso da soggetto a soggetto. A tale proposito si raccomanda di riportare i valori degli errori standard di misurazione almeno per i valori critici identificati nei punteggi.

A.1.2 Correzione dell'attenuazione

Tra le conseguenze della presenza dell'errore di misurazione (definito come differenza tra punteggio *vero* e punteggio osservato) vi è la riduzione (attenuazione) del livello di correlazione tra le variabili osservate. L'errore di misurazione di ciascuna variabile costituisce la parte non comune all'altra.
Il coefficiente di affidabilità delle due variabili consente di *correggere* le correlazioni *attenuate* dall'errore di misurazione. La correzione dell'attenuazione può essere data dal seguente rapporto:

$$r_{t_1 t_2} = \frac{r_{12}}{\sqrt{rho_1 rho_2}}$$

dove

$r_{t_1 t_2}$ correlazione attesa nel caso di perfetta affidabilità delle variabili 1 e 2

rho_1 affidabilità della variabile 1

rho_2 affidabilità della variabile 2

r_{12} correlazione tra variabile 1 e variabile 2.

Sapendo, per esempio, che

$r_{12} = .55$ $\quad rho_1 = .94$ $\quad rho_2 = .85$

allora $r_{t_1t_2} = 0.55 \big/ \sqrt{0.94 * 0.85} = 0.62$

Se la correzione riguarda solo una delle due variabili, sotto radice al denominatore comparirà solo il coefficiente di affidabilità di tale variabile.
In pratica la *correzione dell'attenuazione* rappresenta la stima della correlazione tra due variabili nel caso in cui queste fossero perfettamente affidabili. Vi sono però alcune controversie sull'applicazione di tale correzione. Osservando la questione superficialmente sembrerebbe di aver trovato un facile *espediente* per determinare una correlazione migliore di quella osservata nei dati. In realtà essa produce risultati che sono sempre ipotetici. Una corretta utilizzazione della *correzione dell'attenuazione* prevede l'uso di buone stime di affidabilità. Infatti se le variabili coinvolte presentano affidabilità molto basse tale correzione fornisce una stima molto debole o comunque inutilizzabile (a volte correlazioni poco affidabili ottengono, dopo la correzione, valori anche superiori a 1!!).

Stima dell'aumento della correlazione

L'applicazione della precedente formula consente di stimare anche l'incremento nella correlazione tra due variabili quando l'affidabilità viene incrementata di un particolare valore.

$$r_{t_1t_2} = \frac{r_{12}\sqrt{rho'_1\, rho'_2}}{\sqrt{rho_1 rho_2}} = r_{12}\sqrt{\frac{rho'_1\, rho'_2}{rho_1 rho_2}}$$

dove

$r_{t_1t_2}$ correlazione stimata tra le variabili 1 e 2 con le due corrispondenti affidabilità modificate

rho'_1 coefficiente di affidabilità modificato per la variabile 1

rho'_2 coefficiente di affidabilità modificato per la variabile 2

rho_1 coefficiente di affidabilità osservato della variabile 1

rho_2 coefficiente di affidabilità osservato della variabile 2

r_{12} correlazione tra variabile 1 e variabile 2.

Poniamo di avere una correlazione di .30 tra due variabili, ciascuna delle quali presenta un'affidabilità di .60. Se l'affidabilità di ciascuno strumento viene incrementata fino a .90, la correlazione che si otterrebbe se i due punteggi avessero tale affidabilità sarà:

$$r_{t_1t_2} = \frac{.30\sqrt{.90 * .90}}{\sqrt{.60 * .60}} = .45$$

Nel caso in cui venga modificata l'affidabilità di una sola variabile (per esempio la variabile 1) la formula diviene:

$$r_{t_1t_2} = r_{12}\sqrt{\frac{rho'_1}{rho_1}}$$

Occorre comunque dire che dopo tale stima le correlazioni raramente risultano essere molto diverse dalle correlazioni realmente osservate: con un aumento dell'affidabilità da .60 a .80 si ottiene un incremento della correlazione da .30 a .35. Tale differenza, pure importante, è comunque sempre minore di quanto ci si sarebbe potuto aspettare. Lo stesso procedimento può essere utilizzato per stimare la correlazione nel caso in cui i valori di affidabilità diminuiscano; ciò può risultare particolarmente utile nel caso in cui è necessario impiegare versioni più brevi di determinati strumenti. Se si conosce

- l'affidabilità degli strumenti lunghi,
- l'affidabilità degli strumenti corti,
- la correlazione tra gli strumenti lunghi

le affidabilità degli strumenti accorciati possono essere poste nella seconda versione della formula di correzione dell'attenuazione come rho'_1 e rho'_2 e le affidabilità degli strumenti più lunghi come

rho_1 e rho_2 .

A.2 L'INDICE DI AFFIDABILITÀ

Il coefficiente di affidabilità (rho_x) può essere interpretato come

- autocorrelazione dello strumento con se stesso,
- correlazione tra gli elementi dello strumento,
- percentuale del punteggio osservato che corrisponde al punteggio vero (se il coefficiente di affidabilità è uguale a .95 si può concludere che il 95% del punteggio osservato è quello del punteggio reale e che il 5% rappresenta la proporzione del punteggio osservato attribuibile ad errore di misurazione).

Sappiamo che qualsiasi *coefficiente di correlazione* non è altro che la radice quadrata del *coefficiente di determinazione* $\left(r = \sqrt{r^2}\right)$ e che il *coefficiente di determinazione* è dato dal rapporto tra la *devianza spiegata* e la *devianza totale*

$$r^2_{yy'} = \frac{\sum\left(y' - m_y\right)^2}{\sum\left(y - m_y\right)^2} = \frac{\sigma^2_{y'}}{\sigma^2_y}$$

Ciò consente di affermare che r^2 non è altro che il rapporto tra la varianza dei valori previsti (spiegati) e quella dei valori osservati (totale), in altre parole il quadrato del coefficiente di correlazione non è altro che la parte di variabilità di una variabile che è spiegata dalla variabilità dell'altra. Sappiamo che il rapporto tra la varianza vera e quella osservata rappresenta un'altra definizione di affidabilità

$$r^2_{tx} = rho_x = \frac{\sigma^2_t}{\sigma^2_x}$$

In altre parole, il coefficiente di affidabilità è il coefficiente di determinazione della correlazione tra il punteggio osservato e quello vero ovvero rappresenta la stima del quadrato della correlazione tra un gruppo di punteggi osservati e i corrispondenti punteggi veri (r_{tx}), ovvero r_{tx} (*indice di affidabilità*) è uguale alla radice quadrata del coefficiente di affidabilità:

$$r_{tx} = \sqrt{rho_x} = \textit{indice di affidabilità}$$

(la formula inversa è ovviamente $rho_x = r^2_{tx}$).

A.3 LA PRECISIONE DELLA STIMA DELL'AFFIDABILITÀ

Un problema che si può porre è quello di stabilire il livello di precisione della stima di affidabilità effettuata. La precisione con cui viene stimata l'affidabilità di qualsiasi strumento è una funzione diretta

- della dimensione del campione di item (maggior è il numero, tendenzialmente maggiore è il livello di affidabilità),
- della precisione con cui la correlazione media degli item in uno strumento stima la correlazione media di tutti gli item dell'area di contenuto (maggiore è tale correlazione media tra gli item, tendenzialmente maggiore è il livello di affidabilità).

Se tali correlazioni sono distribuite normalmente, un approssimato errore standard della stima dell'affidabilità vera è ottenuto nel modo seguente:

$$\sigma_r = \frac{\sigma_{r_{ij}}}{\sqrt{½k(k-1)-1}}$$

dove

σ_r errore standard di stima dell'affidabilità nell'intera area di contenuto

$\sigma_{r_{ij}}$ deviazione standard della distribuzione delle correlazioni osservate all'interno di un'area

k numero di item

Tale equazione rappresenta un semplice adattamento della consueta formula dell'errore standard della media campionaria[26]. La deviazione standard delle correlazioni all'interno dell'area viene considerata come stima della deviazione standard delle correlazioni nell'intera area di contenuto. Il denominatore rappresenta la radice quadrata del numero di tutte le possibili correlazioni tra i *k* item meno 1 (*gradi di libertà*).

Quindi rispetto all'affidabilità totale, l'errore di stima è direttamente correlato con la deviazione standard delle correlazioni tra item e la precisione della stima è direttamente correlata con il numero di item. Da ciò deduciamo che gli strumenti con un numero di item maggiore non solo sono i più affidabili ma anche i più precisi.

Se, per esempio, la correlazione media tra 10 item è .20 e la deviazione standard delle correlazioni è .10, allora l'errore standard delle correlazioni medie ottenute in un'area di 10 item è:

$$\sigma_r = \frac{.10}{\sqrt{(5*9)-1}} = \frac{.10}{\sqrt{44}} = \frac{.10}{6.63} = 0.015$$

Quindi ci si aspetta che il 95% dei valori campionari ricadano approssimativamente tra .17 e .23. Se nell'esempio ci fossero stati 40 item l'errore standard sarebbe stato solamente .0036. Comunque anche un'area con pochi item può registrare un'affidabilità piuttosto precisa.

Infine occorre osservare che i principi utilizzati per verificare la precisione della stima dell'affidabilità sono basati su un modello approssimato i cui assunti sono "indipendenza statistica" e "distribuzione normale delle correlazioni tra gli item".

[26] L'errore standard della media è dato dalla deviazione standard della distribuzione diviso il numero di soggetti meno 1. Nel nostro caso ciascuna correlazione è considerata come un caso.

2. I MODELLI CUMULATIVI

Come abbiamo visto, secondo il modello additivo l'insieme combinato degli item misura una sola dimensione, tutti item contribuiscono allo stesso modo alla descrizione della dimensione misurata e ciascun item è monotonamente correlato al continuum della caratteristica misurata.
Ciò vuol dire che nel caso si misuri, per esempio, un atteggiamento, più favorevole (o sfavorevole) è l'atteggiamento del soggetto, maggiore (o minore) è il punteggio totale atteso; conseguentemente gli item sono selezionati sulla base della loro capacità nel discriminare tra unità e non tra livelli diversi del continuum che rappresenta la caratteristica da misurare.
Secondo i modelli cumulativi, gli item non contribuiscono allo stesso modo alla descrizione della dimensione misurata. Ciascuno di essi discrimina i casi in punti diversi del continuum della caratteristica misurata; in altre parole gli item devono essere selezionati sulla base della loro capacità di discriminare non solo tra unità ma anche tra livelli diversi del continuum che rappresenta la caratteristica.Poniamo di voler collocare lungo il continuum metrico di una determinata caratteristica (ad intensità crescente), cinque casi.
Se noi conoscessimo l'unità di misura per poter collocare i casi lungo il continuum dopo aver "confrontato" ciascun caso con la scala di misura potremmo ottenere i seguenti valori (veri):

caso 1: 164 caso 2: 172 caso 3: 178 caso 4: 179 caso 5: 184 caso 6: 194

Lungo il continuum i soggetti sarebbero così posizionati:

```
             160                  170                  180                  190                  200
             ┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼
continuum    ═══════*═══════════════*═══════════*═*═══════════*═══════════════════*══════════════
                    ↑               ↑           ↑ ↑           ↑                   ↑
casi -->            1               2           3 4           5                   6
```

Non disponendo di un tale strumento di misura, per poter procedere alla collocazione dei casi lungo il continuum è possibile definire come indicatore un item che può assumere la seguente caratteristica:

valore maggiore di 168 ⊔ *si* ⊔ *no*

Dei cinque casi, il primo risponderebbe "no", mentre gli altri risponderebbero "sì".
Tali risposte consentono di ordinare i soggetti lungo il continuum rispetto al punto discriminante individuato dall'item. Ciò è reso possibile dal fatto che ciascuna risposta è interpretabile come indicazione della direzione della relazione tra soggetto e item: infatti una risposta positiva indica che il soggetto è al di sopra del limite indicato dall'item mentre una risposta negativa indica che il soggetto si posiziona sul limite o al di sotto.
Vediamo se e in che modo tale item riesce a collocare i soggetti lungo il continuum.

```
                     ─── no ──>|<──────── si ───────........
                               |
             ┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼─┼
continuum     ═════════════════╪══════════════════════════════════════════════════════════════
             160               |  170                  180                  190                  200
casi                         1 |  2,3,4,5,6
```

In seguito a tale misurazione i casi risultano essere discriminati in un punto e quindi divisi in due gruppi. Noi sappiamo però che in questo modo non siamo riusciti a discriminare in maniera precisa tutti i casi in quanto l'utilizzazione di un solo item di questo tipo non consente l'ordinamento preciso dei casi lungo il continuum relativo all'attributo scalabile; per fare ciò sono necessari più item (*scaling*) in grado di discriminare i casi in corrispondenza di altri punti del continuum; nel nostro caso potrebbero essere definiti i seguenti item:

a. valore maggiore di 160	□ *si*	□ *no*
b. valore maggiore di 168	□ *si*	□ *no*
c. valore maggiore di 176	□ *si*	□ *no*
d. valore maggiore di 181	□ *si*	□ *no*
e. valore maggiore di 187	□ *si*	□ *no*

Tali item individuano lungo il continuum punti diversi di discriminazione; a seconda delle risposte date ai cinque item, ciascun soggetto può così essere classificato rispetto al continuum (in pratica la posizione di ciascun caso corrisponde all'ultima risposta positiva - classificazione ordinale).

```
              |<------- a ---------------------------------------------------------...
              |                |<------- b ----------------------------------------...
 segmenti     |                |                |<---- c --------------------------...
              |                |                |          |<----- d --------------...
              |                |                |          |            |<--------- e ------...
              |                |                |          |            |
             -+-+-+-+-+-+-+-+--+-+-+-+-+-+-+-+--+-+-+-+-+--+-+-+-+-+-+--+-+-+-+-+-+-+-+-+-+-+-+-+-
 continuum   ===========================================================================================
            160               170                     180               190                     200
 casi   --> |       1          |        2       |  3 4     |     5      |          6
```

Appare evidente a questo punto che la collocazione di ciascun soggetto si avvicinerà sempre di più a quella "vera" (ovvero la misurazione sarà tanto più precisa) quanti più item utilizziamo e quanto più tali item sono capaci di discriminare i casi in punti diversi del continuum. Tali item, dovendo discriminare in punti diversi e ordinati del continuum, devono essere non solo unidimensionali ma anche *cumulativi*.

L'obiettivo dei modelli cumulativi è quello di individuare tali item e conseguentemente di classificare i casi; in altre parole secondo i modelli cumulativi l'*assegnazione dei punteggi riguarda sia gli item che compongono lo strumento che i casi cui è sottoposto*.

Tali modelli si basano sui seguenti assunti:

- *unidimensionalità*: l'insieme degli item esprimono la misura di una sola dimensione, misurano una sola caratteristica;
- *relazione differenziata* di ciascun item con la dimensione con conseguente
- *assenza di compensabilità* tra item rilevabile e definibile in termini di *gradualità/scalabilità*: gli item devono essere scelti in modo che risultino essere discriminanti a livelli diversi della stessa dimensione; in altre parole deve essere possibile ordinare gli item secondo un livello crescente di intensità (capacità, disposizioni, difficoltà, ecc.); gli item così selezionati presentano solo una parziale sovrapposizione di significato, consentendo di ottenere una *gradualità* della valutazione;*omogeneità*: tutti gli item sono rilevati con lo stesso tipo di livelli di classificazione;
- *esaustività*: l'insieme degli item deve rappresentare un inventario completo del dominio reale di una "dimensione" ovvero gli item devono ricoprire tutta la variabilità osservabile in modo da consentire una valutazione globale.Su questa base è possibile distinguere due distinti modelli cumulativi che si differenziano rispetto al diverso trattamento del problema dell'errore di misurazione, inteso come variazione non sistematica nelle risposte, o della varianza non sistematica (<u>varianza dell'errore</u>) e conseguentemente alla diversa definizione dei modelli di risposta (*trace line*); tali modelli sono:

- modello deterministico, secondo il quale non è possibile prevedere la definizione esplicita di errore o di variazione non sistematica nella misurazione; tutta la variazione nelle risposte è interamente attribuita alla posizione dei soggetti e alla posizione dell'item lungo il continuum che rappresenta la dimensione studiata; all'interno di questo modello non si fa alcuna previsione di varianza dell'errore; conseguentemente la probabilità di dare una determinata risposta ad un item può essere solo 0 (*beta*) o 1 (*alfa*) in qualsiasi punto del continuum sottostante l'attributo misurato; la versione più comune dei modelli cumulativi-deterministici è quella conosciuta con il nome *Guttman*, che ha trovato applicazione nello *scaling* di atteggiamenti;
- modello probabilistico, che prevede la definizione di errore casuale; secondo questo modello la probabilità di dare una determinata risposta ad un item può variare da 0 a 1 e non è ristretta a tali posizioni estreme.

Di solito tali modelli sono applicati ad una matrice di dati rettangolare con *n* risposte soggettive e *k* item (di solito dicotomici). Si assume che sia gli item che i soggetti occupano posizioni lungo la dimensione sottostante. Tale assunto, combinato con la relazione di dominanza dei dati *stimolo-unico*, conduce ad un ordinamento dei punti-soggetto e dei punti-item basato sull'accumulazione delle risposte.
Storicamente però il primo ad aver affrontato il problema di creare un continuum per una caratteristica a intensità crescente è stato Louis Thurstone, psicologo dell'università di Chicago, attraverso importanti lavori (Arcuri, 1974; McIver; 1979; Thurstone, 1927, 1959; Torgerson, 1958).

2.1 ALLE ORIGINI DEI MODELLI CUMULATIVI: LO SCALING DI THURSTONE

Secondo Thurstone per poter procedere alla collocazione di una serie di stimoli lungo un continuum è possibile utilizzare i giudizi espressi da un gruppo di "giudici"[1]. Il modello di *scaling* proposto è basato su alcuni assunti fondamentali.

La legge del giudizio per confronto

Ciascun oggetto sottoposto al giudizio di un soggetto suscita una risposta che è prodotta da un *processo discriminante* relativo all'attributo considerato. Tale processo discriminante è un costrutto teorico e rappresenta la valutazione che un individuo fa, relativamente all'attributo preso in considerazione, nel mettere a confronto uno stimolo con un altro.
Per ciascuno stimolo e ciascun attributo è possibile assumere l'esistenza di più *processi discriminanti* e quindi che il valore del processo discriminante di fronte a ripetute presentazioni (a più soggetti o allo stesso soggetto) di ciascuno stimolo subisca delle variazioni, legate alla presenza dell'errore di misurazione.[2] Tale variabilità conduce all'assunto secondo il quale esiste una distribuzione dei processi discriminanti e che per ciascuno stimolo esiste una risposta ricorrente o più frequente che rappresenta il cosiddetto *processo discriminante modale* (*modal discriminant process*). Si assume che la distribuzione di tutti i processi discriminanti suscitati da ciascuno stimolo rispetto al processo modale discriminante sia normale. Tale distribuzione normale può essere descritta da due parametri, media (che, come sappiamo, in una distribuzione normale è

---

[1] Facciamo un esempio concreto: consideriamo la situazione in cui si voglia costruire un continuum di prestigio delle professioni. Si individuano per questo delle professioni (stimoli) che si presume possano essere ordinate lungo il continuum psicologico rispetto al grado di prestigio che esprimono. Si chiede quindi ad un gruppo di "giudici" di valutare tali stimoli rispetto al livello di prestigio utilizzando una tecnica di *scaling* comparativo, confronti accoppiati (*paired comparisons*) oppure ordinamento per ranghi (*rank order*).

[2] Riprendendo l'esempio precedente, ciascun giudice fa una discriminazione esprimendo un giudizio che è legato al livello di prestigio associato a ciascuna occupazione.

uguale alla mediana e alla moda) e deviazione standard. Per un determinato stimolo, il processo modale discriminante definisce il suo valore di scala mentre la deviazione standard definisce la sua dispersione discriminante. Ciascuno stimolo potrà così essere posizionato lungo il continuum psicologico calcolando il corrispondente valore di scala[3].
L'assunto di base sottostante la legge del giudizio di confronto è che il grado in cui ciascuno stimolo può essere discriminato è funzione diretta della differenza dei loro status rispetto all'attributo in questione. Se la maggior parte dei "giudici" valuta lo stimolo A diverso da quello B relativamente al continuum considerato, le posizioni sulla scala di A e B dovrebbero riflettere tale differenza. Maggiore sarà la proporzione di "giudici" che ha considerato A diverso da B (e viceversa), maggiore sarà la distanza tra gli stimoli A e B sul continuum. Se ad A e B viene attribuito lo stesso valore di scala[4] si presume che essi posseggano la stessa quantità della proprietà considerata (McIver; 1979; Thurstone, 1927, 1959; Torgerson, 1958).
Vediamo come si procede nella pratica con dati prodotti da giudizi per confronto e dati prodotti per ordinamento (McIver; 1979; Thurstone, 1927, 1959; Torgerson, 1958).

2.1.1 Metodo per dati prodotti da giudizi per confronto

In questo caso, a ciascun soggetto vengono sono sottoposti tutti gli stimoli in tutte le possibili combinazioni a coppie (per *k* stimoli, il numero di confronti è $k(k-1)/2$)[5].
Dopo aver proceduto nella stessa maniera per tutti i soggetti, si costruisce una matrice quadrata in cui ciascuna riga e ciascuna colonna rappresenta uno stimolo e ciascuna cella presenta la frequenza con cui lo stimolo di colonna è stato preferito allo stimolo di riga.

Matrice delle frequenze		Stimoli						
		1	2	3	…	*i*	…	k
Stimoli	1		$f_{2>1}$	$f_{3>1}$	…	$f_{i>1}$	…	$f_{k>1}$
	2	$f_{1>2}$		$f_{3>2}$	…	$f_{i>2}$	…	$f_{k>2}$
	3	$f_{1>3}$	$f_{2>1}$		…	$f_{i>3}$	…	$f_{k>3}$
	….	…	…	…		…	…	…
	j	$f_{1>j}$	$f_{2>j}$	$f_{3>j}$	…	$f_{i>j}$	…	$f_{k>j}$
	….	…	…	…	…	….		…
	k	$f_{1>k}$	$f_{2>k}$	$f_{3>k}$	…	$f_{i>k}$	…	

Quindi $f_{2>1}$ indica la frequenza con cui lo stimolo 2 è stato giudicato superiore allo stimolo 1.
Tali frequenze possono essere trasformate in valori che sono proporzionali al numero di soggetti (*N*). Quindi, a partire da tale matrice è possibile costruirne un'altra che contiene in ciascuna cella la

[3] Riprendendo l'esempio precedente, al termine del procedimento tutte le professioni risulteranno essere ordinate lungo il continuum, dalla meno alla più prestigiosa.

[4] Ciò si verifica quando metà dei "giudici" considera lo stimolo A maggiore di B e l'altra metà considera B maggiore di A rispetto al continuum considerato.

[5] La sequenza con cui vengono presentate le coppie di stimoli deve essere predeterminata e dovrebbero essere seguite alcune accortezze:
- ciascuno stimolo dovrebbe apparire con uguale frequenza a destra e a sinistra dell'elenco (per controllare l'errore dovuto alla posizione);
- le posizioni di ciascuno stimolo sulla destra e sulla sinistra dovrebbero essere alternate;
- nessuno stimolo dovrebbe trovarsi in due coppie successive (la distanza tra le successive presentazioni dello stesso stimolo dovrebbe essere il più ampia possibile).

Queste indicazioni sono più facilmente applicabili con un numero *n* di stimoli dispari.

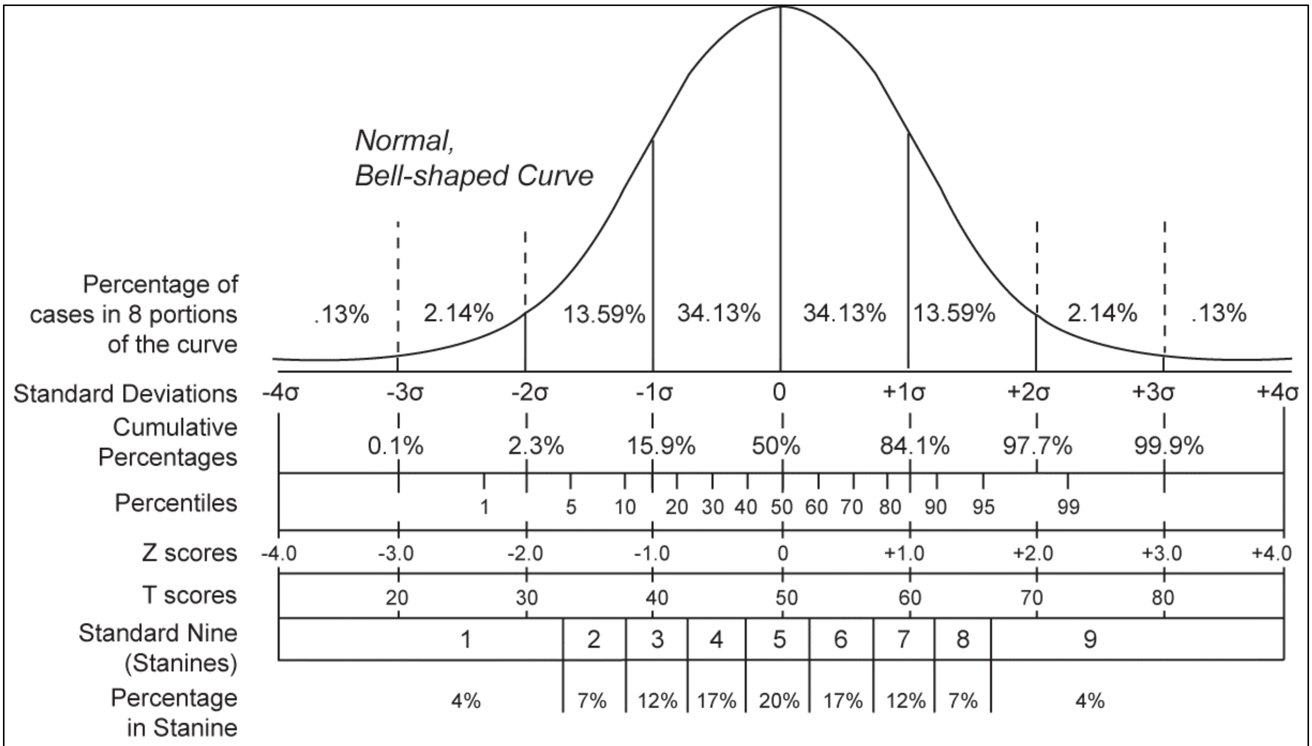
proporzione delle volte in cui ciascuno stimolo di colonna è stato preferito – rispetto all'attributo – a ciascuno stimolo di riga.

Matrice delle proporzioni		Stimoli						
		1	2	3	...	*i*	...	k
Stimoli	1		$p_{2>1}$	$p_{3>1}$	...	$p_{i>1}$	...	$p_{k>1}$
	2	$p_{1>2}$		$p_{3>2}$	...	$p_{i>2}$	...	$p_{k>2}$
	3	$p_{1>3}$	$p_{2>3}$		...	$p_{i>3}$	...	$p_{k>3}$
	....	...	...	...		....	...	...
	j	$p_{1>j}$	$p_{2>j}$	$p_{3>j}$	...	$p_{i>j}$	...	$p_{k>j}$
	....	...	...	...	...	...		...
	k	$p_{1>k}$	$p_{2>k}$	$p_{3>k}$	...	$p_{i>k}$	...	

Come si può facilmente osservare nella matrice vale la relazione:

$$p_{i>j} + p_{j>i} = 1$$

Secondo la legge del giudizio per confronto di Thurstone la distanza scalare tra due stimoli *i* e *j* lungo il continuum è data dal valore standard z_{ij} corrispondente alla proporzione di volte in cui uno stimolo è giudicato superiore ad un altro $\left(p_{i>j}\right)$. Il valore z_{ij} si ottiene facendo riferimento alla tavola della distribuzione normale. In particolare, è necessario identificare tra i valori della distribuzione cumulata (*cumulative percentages*) la proporzione in questione e successivamente individuare il corrispondente valore standardizzato (*Z scores*), come evidenziato nella seguente figura.
E' comunque possibile utilizzare anche le funzioni relative alla curva normale presenti in tutti i package statistici.



da *Wikipedia*

La matrice di proporzioni viene così trasformata in una nuova matrice nelle cui celle sono riportate le distanze scalari tra gli stimoli in unità standard normali.

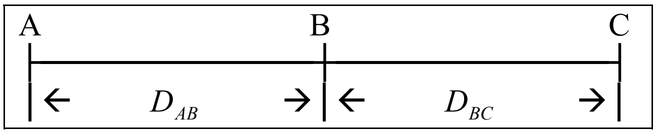
Matrice dei valori standard		Stimoli						
		1	2	3	…	*i*	…	k
Stimoli	1		z_{21}	z_{31}	…	z_{i1}	…	z_{k1}
	2	z_{12}		z_{32}	…	z_{i2}	…	z_{k2}
	3	z_{13}	z_{23}		…	z_{i3}	…	z_{k3}
	….	…	…	…		….	…	…
	j	z_{1j}	z_{2j}	z_{3j}	…	z_{ij}	…	z_{kj}
	….	…	…	…	…	….		…
	k	z_{1k}	z_{2k}	z_{3k}	…	z_{ik}	…	

Le distanze riportate in questa matrice consentono di disporre gli stimoli sul continuum metrico se il continuum è unidimensionale e in assenza di errore ovvero quando tali distanze soddisfano la *proprietà additiva*.

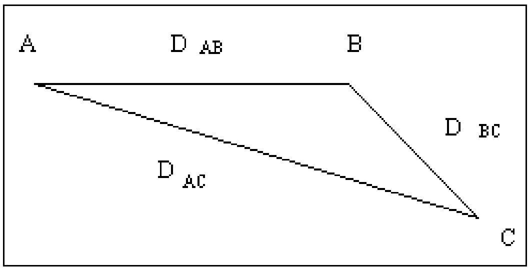
Secondo tale proprietà, dati tre punti *A*, *B* e *C* la distanza tra *A* e *C* deve essere uguale alla somma delle distanze tra *A* e *B* più quella tra *B* e *C*

$$D_{AC} = D_{AB} + D_{BC}$$

Geometricamente questa condizione è rappresentata nella seguente figura:



Se il continuum non è unidimensionale la proprietà additiva delle distanze non è soddisfatta e geometricamente può essere rappresentata come nella figura di seguito.[6]



In realtà le distanze determinate nel modo precedentemente presentato difficilmente potranno risultare perfettamente additive, pertanto per ridurre la quantità di errore è possibile determinare la distanza di ciascuno stimolo dalla media di tutti gli altri valori scalari. Sapendo che i valori di ciascuna colonna *i* rappresentano le distanze dello stimolo *i* da tutti gli altri stimoli, sommiamo tutti i *k* valori della colonna *i* e otteniamo

$$\sum_{j=1}^{k} z_{ij} = \sum_{j=1}^{k} S_i - S_j$$

dove

z_{ij} distanza dello stimolo *i* dallo stimolo *j*

S_i valore scalare dello stimolo *i*

S_j valore scalare dello stimolo *j*

[6] Questa situazione si può verificare quando gli stimoli sono classificati secondo criteri diversi. In altre parole, l'attributo su cui i soggetti devono confrontare gli stimoli assume diversi significati a seconda dei particolari simboli che vengono proposti nella valutazione. In questi casi è più utile applicare modelli di scaling multidimensionale.

che può anche essere scritta come

$$\sum_{j=1}^{k} z_{ij} = kS_i - \sum_{j=1}^{k} S_j$$

A questo punto dividiamo entrambi i termini per il numero dei valori della colonna (k) ed otteniamo $\bar{z}_i$ che rappresenta il valore scalare di i attraverso la sua deviazione dalla media di tutti gli altri valori scalari

$$\bar{z}_i = S_i - \bar{S}$$

dove

$\bar{z}_i$ media dei valori della colonna i

S_i valore scalare dello stimolo i

$\bar{S}$ media aritmetica dei k valori scalari

2.1.2 Metodo per dati prodotti da ordinamenti per ranghi

In questo caso, i k stimoli vengono disposti da ciascun caso in ordine secondo un certo criterio. Dopo aver proceduto nella stessa maniera per tutti i casi, si costruisce una matrice rettangolare in cui ciascuna riga rappresenta un soggetto e ciascuna colonna uno stimolo; in ciascuna cella è riportato il rango che ciascun caso ha attribuito a ciascuno stimolo.

		Stimoli						
		1	2	3	…	*i*	…	k
Casi	1	R_{11}	R_{21}	R_{31}	…	…	…	R_{k1}
	2	R_{12}	R_{22}	R_{32}	…	…	…	R_{k2}
	3	R_{13}	R_{23}	R_{33}	…	…	…	R_{k3}
	….	…	…	…	…	…	…	…
	j	R_{1i}	R_{2i}	R_{3j}	…	…	…	R_{kj}
	….	…	…	…	…	…	…	…
	n	R_{1n}	R_{2n}	R_{3n}	…	…	…	R_{kn}

A questo punto si procede al calcolo della somma dei ranghi di ciascuna colonna della matrice dei dati

$$\sum_{j=1}^{n} R_{ij}$$

Successivamente si calcola il rango medio di ciascuno stimolo i, dividendo la somma per il numero di casi

$$\bar{R}_i = \frac{\sum_{j=1}^{n} R_{ij}}{n}$$

E' possibile a questo punto determinare i ranghi percentili corrispondenti ai ranghi medi ottenuti[7]:

$$P_i = \frac{k - \bar{R}_i + 0.5}{k} * 100$$

dove

P_i rango percentile del rango medio registrato per lo stimolo i

[7] La moltiplicazione per 100 consente di individuare il percentile, altrimenti dalla formula si ottiene la proporzione di area sotto la curva normale corrispondente alle posizioni percentili di $\bar{R}_i$.

k numero degli stimoli
$\overline{R}_i$ rango medio dello stimolo *i*

A partire dai ranghi percentili è possibile determinare i valori scalari degli stimoli (*z*) servendosi della tavola della distribuzione normale[8].

Procedimenti alternativi

(A)

E' possibile che la matrice di input sia una tabella di frequenza a doppia entrata con in riga i ranghi e in colonna gli stimoli; in ciascuna cella è indicata la frequenza con cui lo stimolo corrispondente alla colonna che individua la cella è stato posizionato nel rango della riga corrispondente alla cella considerata. Poniamo di avere somministrato ad un gruppo di soggetti 10 stimoli; la tabella di frequenze avrà la seguente forma:

		Stimoli									
		1	2	3	4	5	6	7	8	9	10
Ranghi	1	$f_{1.1}$	$f_{2.1}$	$f_{3.1}$	$f_{4.1}$	$f_{5.1}$	$f_{6.1}$	$f_{7.1}$	$f_{8.1}$	$f_{9.1}$	$f_{10.1}$
	2	$f_{1.2}$	$f_{2.2}$	$f_{3.2}$	$f_{4.2}$	$f_{5.2}$	$f_{6.2}$	$f_{7.2}$	$f_{8.2}$	$f_{9.2}$	$f_{10.2}$
	3	$f_{1.3}$	$f_{2.3}$	$f_{3.3}$	$f_{4.3}$	$f_{5.3}$	$f_{6.3}$	$f_{7.3}$	$f_{8.3}$	$f_{9.3}$	$f_{10.3}$
	4	$f_{1.4}$	$f_{2.4}$	$f_{3.4}$	$f_{4.4}$	$f_{5.4}$	$f_{6.4}$	$f_{7.4}$	$f_{8.4}$	$f_{9.4}$	$f_{10.4}$
	5	$f_{1.5}$	$f_{2.5}$	$f_{3.5}$	$f_{4.5}$	$f_{5.5}$	$f_{6.5}$	$f_{7.5}$	$f_{8.5}$	$f_{9.5}$	$f_{10.5}$
	6	$f_{1.6}$	$f_{2.6}$	$f_{3.6}$	$f_{4.6}$	$f_{5.6}$	$f_{6.6}$	$f_{7.6}$	$f_{8.6}$	$f_{9.6}$	$f_{10.6}$
	7	$f_{1.7}$	$f_{2.7}$	$f_{3.7}$	$f_{4.7}$	$f_{5.7}$	$f_{6.7}$	$f_{7.7}$	$f_{8.7}$	$f_{9.7}$	$f_{10.7}$
	8	$f_{1.8}$	$f_{2.8}$	$f_{3.8}$	$f_{4.8}$	$f_{5.8}$	$f_{6.8}$	$f_{7.8}$	$f_{8.8}$	$f_{9.8}$	$f_{10.8}$
	9	$f_{1.9}$	$f_{2.9}$	$f_{3.9}$	$f_{4.9}$	$f_{5.9}$	$f_{6.9}$	$f_{7.9}$	$f_{8.9}$	$f_{9.9}$	$f_{10.9}$
	10	$f_{1.10}$	$f_{2.10}$	$f_{3.10}$	$f_{4.10}$	$f_{5.10}$	$f_{6.10}$	$f_{7.10}$	$f_{8.10}$	$f_{9.10}$	$f_{10.10}$

A partire da questa tabella di frequenza il procedimento è in sostanza equivalente a quello appena presentato.

(B)

E' possibile determinare, a partire dai ranghi attribuiti da ciascun soggetto a ciascuno stimolo, le valutazioni comparative del soggetto tra ciascuna coppia di stimoli e predisporre una matrice quadrata in cui righe e colonne rappresentano gli stimoli e in ciascuna cella è riportata la frequenza con cui lo stimolo di colonna è stato preferito all'altro rispetto al criterio adottato.
Poniamo di avere quattro stimoli – *a*, *b*, *c*, *d* – che sono stati posizionati da un soggetto nell'ordine *b*, *d*, *a*, *c* da tali relazioni di ordine è possibile dedurre i seguenti giudizi comparativi:

$b>d \quad b>a \quad b>c \quad d>a \quad d>c \quad a>c$

Osservando la stessa cosa anche per gli altri casi, è possibile determinare la proporzione di volte in cui il ciascuno stimolo è stato giudicato maggiore di ciascun altro stimolo.
Anche se semplice, il procedimento può rivelarsi un po' laborioso, specie in presenza di un numero elevato di casi e di stimoli. Ottenuta questa tabella si procede esattamente come per il metodo dei confronti a coppie.

[8] A tale proposito, confrontare la precedente figura relativa alla curva normale. E' inoltre possibile eliminare i valori negativi della scala effettuando una trasformazione lineare del tipo

$$z' = az + b$$

dove
z' nuovo valore scalare nella nuova unità di misura
z valore standard originario
a e *b* costanti, rispettivamente deviazione standard e media.

2.1.3 Una applicazione

Per esemplificare il procedimento fin qui esposto, verranno utilizzati i dati rilevati attraverso un'indagine effettivamente realizzata e finalizzata alla messa a punto di un questionario per la rilevazione della qualità della vita percepita attraverso l'uso del modello congiunto[9]. Il questionario prevedeva anche due particolari dimensioni per la rilevazione delle quali sono state applicate le tecniche del confronto a coppie e dell'ordinamento per ranghi. La rilevazione, dati gli obiettivi dell'indagine, ha coinvolto un campione ristretto (ca. 50) di studenti universitari.

2.1.3.1 Preferenze per i mezzi di trasporto[10]

Il problema dello spostamento nel centro delle città, è una questione che riguarda tutti quelli che hanno a che fare con la città sia per motivi di studio (come nel caso dei soggetti che compongono il nostro campione) o lavoro oppure perché ci abitano. Quindi è sembrato importante valutare questo aspetto che è stato rilevato attraverso la tecnica dei confronti accoppiati: ciascun soggetto ha indicato, tra due stimoli presentati contemporaneamente, a quale dei due va la sua preferenza.
La domanda appariva nel questionario nel modo seguente:

Per ogni coppia di mezzi di trasporto, metti una crocetta sul quadratino più vicino al mezzo che preferisci usare per spostarti in città			
Autobus	☐	☐	Automobile
Automobile	☐	☐	A piedi
A piedi	☐	☐	Bicicletta
Bicicletta	☐	☐	Moto-scooter
Moto-scooter	☐	☐	A piedi
Autobus	☐	☐	Bicicletta
Automobile	☐	☐	Moto-scooter
A piedi	☐	☐	Autobus
Bicicletta	☐	☐	Automobile
Moto-scooter	☐	☐	Autobus

Vediamo di seguito, le frequenze registrate per ciascun confronto effettuato:

CONFRONTO AUTOBUS - AUTOMOBILE	N	%
Autobus preferito a automobile	10	21.3
Automobile preferita a autobus	37	78.7
Totale	47	100

CONFRONTO AUTOMOBILE - A PIEDI	N	%
Automobile preferita a piedi	21	44.7
A piedi preferito a automobile	26	55.3
Totale	47	100

CONFRONTO A PIEDI - BICICLETTA	N	%
A piedi preferito a bicicletta	23	48.9
Bicicletta preferita a piedi	24	51.1
Totale	47	100

CONFRONTO BICICLETTA - MOTO	N	%
Bicicletta preferita a scooter	24	51.1
Scooter preferito a bicicletta	23	48.9
Totale	47	100

[9] Maggino F. (2005) *Importanza delle dimensioni di qualità della vita nelle preferenze dei cittadini: un'applicazione sperimentale dell'analisi congiunta.* Firenze University Press, Archivio E-Prints.

[10] L'applicazione tradizionale del modello Thurstone è fatta su dati prodotti da giudizi individuali; l'applicazione qui presentata è una estensione del procedimento al caso di dati espressi in termini di preferenza.

CONFRONTO MOTO – A PIEDI	N	%
Scooter preferito a piedi	25	53.2
A piedi preferito a scooter	22	46.8
Totale	47	100

CONFRONTO AUTOBUS - BICICLETTA	N	%
Autobus preferito a bicicletta	10	21.3
Bicicletta preferita a autobus	37	78.7
	47	100

CONFRONTO AUTOMOBILE – MOTO	N	%
Automobile preferita a scooter	23	48.9
Scooter preferito a automobile	24	51.1
Totale	47	100

CONFRONTO A PIEDI - AUTOBUS	N	%
A piedi preferito a autobus	35	74.5
Autobus preferito a piedi	12	25.5
Totale	47	100

CONFRONTO BICICLETTA - AUTOMOBILE	N	%
Bicicletta preferita a automobile	19	40.4
Automobile preferita a bicicletta	28	59.6
Totale	47	100

CONFRONTO MOTO – AUTOBUS	N	%
Scooter preferito a autobus	36	76.6
Autobus preferito a scooter	11	23.4
Totale	47	100

Di seguito vediamo le due matrici riassuntive, rispettivamente, delle frequenze e delle proporzioni:

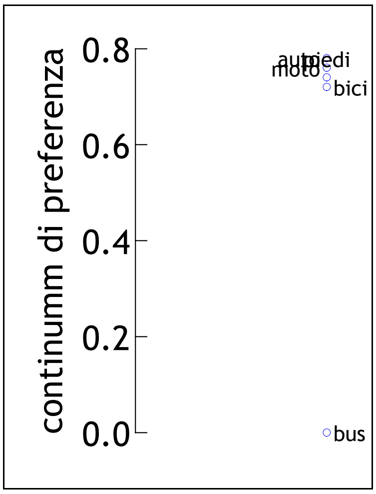
Matrice delle frequenze	BUS	AUTO	MOTO	BICI	PIEDI
BUS		37	36	37	35
AUTO	10		24	19	26
MOTO	11	23		24	22
BICI	10	28	23		23
PIEDI	12	21	25	24	

Matrice delle proporzioni	BUS	AUTO	MOTO	BICI	PIEDI
BUS	0.500	0.787	0.766	0.787	0.745
AUTO	0.213	0.500	0.511	0.404	0.553
MOTO	0.234	0.489	0.500	0.511	0.468
BICI	0.213	0.596	0.489	0.500	0.489
PIEDI	0.255	0.447	0.532	0.511	0.500

Nella seguente matrice sono invece riportati i valori standard normali corrispondenti alle proporzioni precedentemente osservate. Inoltre, per ciascuna colonna (ovvero per ciascuno stimolo) sono riportate la somma, la media e il valore scalare finale, ricalcolati in modo da non avere valori negativi:

Matrice dei valori standard normali	BUS	AUTO	MOTO	BICI	PIEDI
BUS	0	0.796	0.726	0.796	0.659
AUTO	-0.796	0	0.028	-0.243	0.133
MOTO	-0.726	-0.028	0	0.028	-0.080
BICI	-0.796	0.243	-0.028	0	-0.028
PIEDI	-0.659	-0.133	0.080	0.028	0
Somma	-2.980	0.880	0.810	0.610	0.680
Media	-0.600	0.180	0.160	0.120	0.140
+0.60	**0.00**	**0.78**	**0.76**	**0.72**	**0.74**

Mettendo in ordine i mezzi sul continuum di preferenza si può dedurre come il mezzo più preferito sia l'automobile insieme alla moto seguito dallo spostamento a piedi e dalla bici. Meno preferito in assoluto è l'autobus.



2.1.3.2 *Importanza attribuita agli ambiti di vita individuale*

Per valutare l'importanza attribuita a ciascun ambito proposto, dalla famiglia alla carriera, dalle amicizie al partner, ecc., è stata utilizzata una scala di ordinamento o *rank-order*. Ai soggetti sono stati presentati tutti gli stimoli contemporaneamente e ad essi doveva essere assegnato un numero di graduatoria secondo il criterio di importanza (1 per il più importante, 2 per il secondo più importante, fino a 10 il meno importante). Nel questionario la domanda appariva nella seguente forma:

Quale ambito ritieni più importante nella vita di una persona? Metti in ordine di importanza (1 il più importante, 10 il meno importante) i seguenti ambiti. (Due ambiti non possono avere la stessa posizione)	*Ordine*
il guadagno	
la carriera	
la famiglia	
il vicinato	
le amicizie	
l'aspetto fisico	
l'indipendenza economica	
gli ideali	
la salute	
il partner	

Vediamo nella seguente matrice i valori ordinali attribuiti ai dieci ambiti di vita (guadagno, carriera, famiglia, vicini, amici, aspetto fisico, indipendenza economica, ideali, salute, partner) da ciascuno dei casi osservati (come si potrà osservare, alcuni casi non hanno riferito alcun valore). Le ultime righe della matrice riportano i valori scalari ottenuti secondo la procedura presentata in precedenza.

	guadagno	carriera	famiglia	vicini	amici	fisico	indipend	ideali	salute	partner
	V36	V37	V38	V39	V40	V41	V42	V43	V44	V45
Q1	8	9	6	10	1	7	5	2	3	4
Q2										
Q3	10	7	4	8	3	9	6	5	1	2
Q4	10	7	5	9	1	8	6	3	2	4
Q5	6	2	3	10	7	8	4	9	1	5
Q6	10	6	2	7	4	8	9	5	1	3
Q7	6	8	2	9	4	5	7	10	1	3
Q8	7	8	4	10	3	6	2	9	5	1
Q9	7	4	1	10	6	9	8	5	3	2
Q10	8	9	2	10	3	7	6	5	4	1
Q11	8	7	4	10	5	9	6	1	2	3
Q12	7	5	1	4	3	10	9	8	2	6
Q13	9	8	4	10	1	7	5	2	3	6
Q14	6	7	2	10	3	9	5	8	1	4
Q15	7	9	3	10	4	8	5	6	1	2
Q16	9	8	1	10	5	7	6	4	2	3
Q17	9	7	3	10	1	5	8	2	6	4
Q18	5	8	2	10	3	9	4	7	6	1
Q19	4	3	2	10	5	9	8	7	1	6
Q20	5	6	1	10	4	9	7	8	2	3
Q21	8	7	1	10	5	9	6	4	3	2
Q22	8	7	4	10	3	6	5	9	1	2
Q23	8	4	2	10	5	9	6	7	1	3
Q24	9	7	1	10	5	6	8	3	2	4
Q25										
Q26	8	7	5	10	4	9	3	6	1	2
Q27	8	7	1	10	3	9	6	4	5	2
Q28	7	8	3	10	4	9	6	2	1	5
Q29										
Q30	7	9	2	10	3	8	5	4	6	1
Q31	9	4	2	10	3	8	5	6	7	1
Q32	7	4	2	10	6	9	5	8	1	3
Q33	5	4	1	10	7	9	6	8	3	2
Q34	7	4	1	10	5	9	8	6	2	3
Q35	8	7	2	10	5	9	4	6	1	3
Q36	9	7	2	10	6	8	5	4	1	3
Q37	5	3	1	10	2	9	4	8	6	7
Q38	10	4	2	7	3	9	6	8	1	5
Q39	10	8	3	9	2	7	6	5	4	1
Q40	3	4	8	10	7	5	2	9	1	6
Q41	9	8	1	10	3	4	7	5	6	2
Q42	10	9	1	6	3	8	7	5	4	2
Q43	9	4	2	10	5	7	6	8	1	3
Q44	8	6	1	9	3	10	7	4	5	2
Q45										
Q46	9	8	1	10	4	7	5	3	6	2
Q47	7	8	5	10	3	9	6	4	1	2
Q48										
Q49	8	10	1	9	4	7	5	6	3	2
Somma	337.00	286.00	107.00	417.00	169.00	348.00	255.00	248.00	120.00	133.00
Media	7.66	6.50	2.43	9.48	3.84	7.91	5.80	5.64	2.73	3.02
p=(k-R+.5)/k	0.284	0.4	0.707	0.102	0.666	0.259	0.47	0.486	0.777	0.748
z	-0.571	-0.253	0.545	-1.270	0.429	-0.646	-0.075	-0.035	0.762	0.668
z+1.270	**0.699**	**1.017**	**1.815**	**0.000**	**1.699**	**0.624**	**1.195**	**1.235**	**2.032**	**1.938**

Proviamo ora a vedere come, a partire dalle valutazioni ordinali, sia possibile ricavare le valutazioni comparative. Prendiamo in considerazione tutte le possibile coppie *ij* di stimoli e procediamo contando quante volte *i* è stato valutato con un rango maggiore di *j* e quante volte *i* è stato valutato con un rango minore di *j*. I confronti effettuati sono stati 45:

CONF1	N	%	CONF2	N	%	CONF3	N	%
V36>V37	13.	29.5	V36>V38	1.	2.3	V36>V39	37.	84.1
V36<V37	31.	70.5	V36<V38	43.	97.7	V36<V39	7.	15.9
Totale	44	100	Totale	44	100	Totale	44	100
CONF4	N	%	CONF5	N	%	CONF6	N	%
V36>V40	4.	9.1	V36>V41	23.	52.3	V36>V42	7.	15.9
V36<V40	40.	90.9	V36<V41	21.	47.7	V36<V42	37.	84.1
Totale	44	100	Totale	44	100	Totale	44	100
CONF7	N	%	CONF8	N	%	CONF9	N	%
V36>V43	13.	29.5	V36>V44	2.	4.5	V36>V45	3.	6.8
V36<V43	31.	70.5	V36<V44	42.	95.5	V36<V46	41.	93.2
Totale	44	100	Totale	44	100	Totale	44	100
CONF10	N	%	CONF11	N	%	CONF12	N	%
V37>V38	2.	4.5	V37>V39	41.	93.2	V37>V40	9.	20.5
V37<V38	42.	95.5	V37<V39	3.	6.8	V37<V40	35.	79.5
Totale	44	100	Totale	44	100	Totale	44	100
CONF13	N	%	CONF14	N	%	CONF15	N	%
V37>V41	28.	63.6	V37>V42	17.	38.6	V37>V43	18.	40.9
V37<V41	16.	36.4	V37<V42	27.	61.4	V37<V43	26.	59.1
Totale	44	100	Totale	44	100	Totale	44	100
CONF16	N	%	CONF17	N	%	CONF18	N	%
V37>V44	2.	4.5	V37>V45	6.	13.6	V38>V39	41.	93.2
V37<V44	42.	95.5	V37<V45	38.	86.4	V38<V39	3.	6.8
Totale	44	100	Totale	44	100	Totale	44	100
CONF19	N	%	CONF20	N	%	CONF21	N	%
V38>V40	33.	75.0	V38>V41	43.	97.7	V38>V42	40.	90.9
V38<V40	11.	25.0	V38<V41	1.	2.3	V38<V42	4.	9.1
Totale	44	100	Totale	44	100	Totale	44	100
CONF22	N	%	CONF23	N	%	CONF24	N	%
V38>V43	37.	84.1	V38>V44	22.	50.0	V38>V45	29.	65.9
V38<V43	7.	15.9	V38<V44	22.	50.0	V38<V45	15.	34.1
Totale	44	100	Totale	44	100	Totale	44	100
CONF25	N	%	CONF26	N	%	CONF27	N	%
V39<V40	44.	100.0	V39>V41	6.	13.6	V39>V42	3.	6.8
Totale	44	100	V39<V41	38.	86.4	V39<V42	41.	93.2
			Totale	44	100	Totale	44	100
CONF28	N	%	CONF29	N	%	CONF30	N	%
V39>V43	3.	6.8	V39>V44	44.	100.0	V39>V45	1.	2.3
V39<V43	41.	93.2	Totale	44	100	V39<V45	43.	97.7
Totale	44	100				Totale	44	100
CONF31	N	%	CONF32	N	%	CONF33	N	%
V40>V41	43.	97.7	V40>V42	36.	81.8	V40>V43	36.	81.8
V40<V41	1.	2.3	V40<V42	8.	18.2	V40<V43	8.	18.2
Totale	44	100	Totale	44	100	Totale	44	100
CONF34	N	%	CONF35	N	%	CONF36	N	%
V40>V44	16.	36.4	V40>V45	10.	22.7	V41>V42	5.	11.4
V40<V44	28.	63.6	V40<V45	34.	77.3	V41<V42	39.	88.6
Totale	44	100	Totale	44	100	Totale	44	100
CONF37	N	%	CONF38	N	%	CONF39	N	%
V41>V43	7.	15.9	V41>V44	2.	4.5	V41>V45	1.	2.3
V41<V43	37.	84.1	V41<V44	42.	95.5	V41<V45	43.	97.7
Totale	44	100	Totale	44	100	Totale	44	100
CONF40	N	%	CONF41	N	%	CONF42	N	%
V42>V43	19.	43.2	V42>V44	6.	13.6	V42>V45	4.	9.1
V42<V43	25.	56.8	V42<V44	38.	86.4	V42<V45	40.	90.9
Totale	44	100	Totale	44	100	Totale	44	100
CONF43	N	%	CONF44	N	%	CONF45	N	%
V43>V44	10.	22.7	V43>V45	7.	15.9	V44>V45	28.	63.6
V43<V44	34.	77.3	V43<V45	37.	84.1	V44<V45	16.	36.4
Totale	44	100	Totale	44	100	Totale	44	100

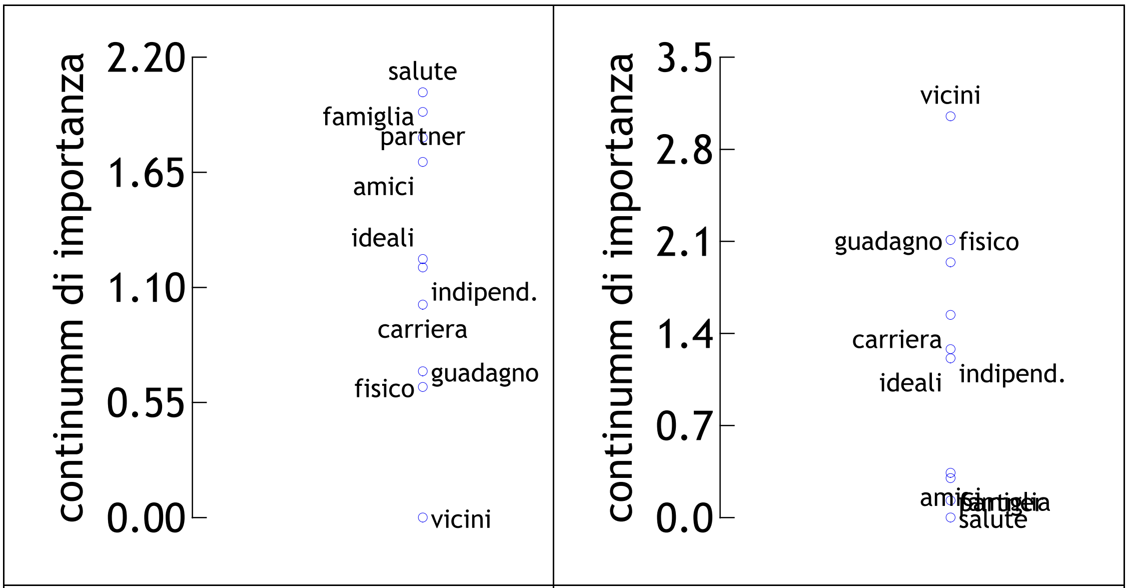
Si procede a questo punto nello stesso modo visto con l'esempio riguardante i mezzi di trasporto:

Matrice delle frequenze		V36	V37	V38	V39	V40	V41	V42	V43	V44	V45
guadagno	V36	-	13	1	37	4	23	7	13	2	3
carriera	V37	31	-	2	41	9	28	17	18	2	6
famiglia	V38	43	42	-	41	33	43	40	37	22	29
vicini	V39	7	3	3	-	0	6	3	3	0	1
amici	V40	40	35	11	44	-	43	36	36	16	10
fisico	V41	21	16	1	38	1	-	5	7	2	1
Indipendenza	V42	37	27	4	41	8	39	-	19	6	4
ideali	V43	31	26	7	41	8	37	25	-	10	7
salute	V44	42	42	22	44	28	42	38	34	-	28
partner	V45	41	38	15	43	34	43	40	37	16	-

Matrice delle proporzioni		V36	V37	V38	V39	V40	V41	V42	V43	V44	V45
guadagno	V36	0.500	0.295	0.023	0.841	0.091	0.523	0.159	0.295	0.045	0.068
carriera	V37	0.705	0.500	0.045	0.932	0.205	0.636	0.386	0.409	0.045	0.136
famiglia	V38	0.977	0.955	0.500	0.932	0.750	0.977	0.909	0.841	0.500	0.659
vicini	V39	0.159	0.068	0.068	0.500	0.001	0.136	0.068	0.068	0.001	0.023
amici	V40	0.909	0.795	0.250	0.999	0.500	0.977	0.818	0.818	0.364	0.227
fisico	V41	0.477	0.364	0.023	0.864	0.023	0.500	0.114	0.159	0.045	0.023
Indipendenza	V42	0.841	0.614	0.091	0.932	0.182	0.886	0.500	0.432	0.136	0.091
ideali	V43	0.705	0.591	0.159	0.932	0.182	0.841	0.568	0.500	0.227	0.159
salute	V44	0.955	0.955	0.500	0.999	0.636	0.955	0.864	0.773	0.500	0.636
partner	V45	0.932	0.864	0.341	0.977	0.773	0.977	0.909	0.841	0.364	0.500

Matrice dei valori standard normali		V36	V37	V38	V39	V40	V41	V42	V43	V44	V45
guadagno	V36	0	-0.5388	-1.9954	0.9986	-1.3346	0.0577	-0.9986	-0.5388	-1.6954	-1.4909
carriera	V37	0.5388	0	-1.6954	1.4909	-0.8239	0.3478	-0.2898	-0.2301	-1.6954	-1.0985
famiglia	V38	1.9954	1.6954	0	1.4909	0.6745	1.9954	1.3346	0.9986	0	0.4097
vicini	V39	-0.9986	-1.4909	-1.4909	0	-4.2649	-1.0985	-1.4909	-1.4909	-4.2649	-1.9954
amici	V40	1.3346	0.8239	-0.6745	4.2649	0	1.9954	0.9078	0.9078	-0.3478	-0.7488
fisico	V41	-0.0577	-0.3478	-1.9954	1.0985	-1.9954	0	-1.2055	-0.9986	-1.6954	-1.9954
Indipendenza	V42	0.9986	0.2898	-1.3346	1.4909	-0.9078	1.2055	0	-0.1713	-1.0985	-1.3346
ideali	V43	0.5388	0.2301	-0.9986	1.4909	-0.9078	0.9986	0.1713	0	-0.7488	-0.9986
salute	V44	1.6954	1.6954	0	4.2649	0.3478	1.6954	1.0985	0.7488	0	0.3478
partner	V45	1.4909	1.0985	-0.4097	1.9954	0.7488	1.9954	1.3346	0.9986	-0.3478	0
Somma		7.54	3.46	-10.59	18.59	-8.46	9.19	0.86	0.22	-11.89	-8.90
Media		0.75	0.35	-1.06	1.86	-0.85	0.92	0.09	0.02	-1.19	-0.89
+1.19		**1.94**	**1.54**	**0.13**	**3.05**	**.34**	**2.11**	**1.28**	**1.21**	**0**	**0.3**

Confrontiamo a questo punto i risultati ottenuti con i due procedimenti:



RANK ORDER	CONFRONTI ACCOPPIATI A PARTIRE DA *RANK ORDER*

Come si può osservare i due continuum risultano speculari (in quanto diverso è il significato attribuito ai dati utilizzati, rispettivamente dati ordinali e frequenze) e con metrica diversa (il primo continuum va da 0 a 2.20 ca., il secondo da 0 a 3.5 ca.). Conseguentemente risultano essere diverse anche le distanze metriche tra i diversi ambiti (da notare come paradossalmente il continuum più "lungo" le distanze risultano minori ovvero la discriminazione tra stimoli risulta peggiore). Ciò che risulta essere comunque rispettato in entrambe le soluzioni è l'ordine di importanza tra i diversi ambiti.

2.1.4 Osservazioni

Le scale così create vengono dette *scale Thurstone* o *scale differenziali*. In letteratura esistono molte versioni analitiche a seconda del numero di casi e di stimoli coinvolti e a seconda del modello sperimentale adottato (assunti).
Si tratta di strumenti raramente utilizzati in quanto non di semplice costruzione, rilevazione e analisi per la presenza di numerosi problemi.
Uno dei principali problemi che si pongono a livello teorico riguarda la reale possibilità di soddisfare alcuni assunti fondamentali di tale modello, quale quello della unidimensionalità del continuum psicologico (McIver; 1979).
Tra i problemi applicativi che tale modello pone vi è la scelta degli stimoli, che devono essere differenziati rispetto al continuum ipotizzato senza essere troppo "differenziati" e la tecnica di somministrazione degli stimoli, soprattutto quando si utilizza la tecnica dei confronti accoppiati. La riuscita dell'applicazione di tale tecnica è legata al numero di confronti che si richiedono; se il numero è troppo grande, il compito può rivelarsi gravoso sia in termini di tempo che di attenzione richiesti ai soggetti (Arcuri, 1974). Sono stati studiati alcuni accorgimenti che consentono di alleggerire il compito sia dei giudici che dei ricercatori (per esempio, in presenza di un numero elevato si casi, è possibile evitare di sottoporre tutte le coppie a tutti; ciò pone però problemi di altra natura).
Vedremo nei successivi capitoli come l'idea di misurare stimoli che risultano discriminanti in punti diversi del continuum sia stata realizzati attraverso modelli diversi.

2.2 *LA METODOLOGIA Q*

In maniera molto simile a quanto abbiamo visto con la scala Thurstone, l'obiettivo della metodologia *Q* è quello di posizionare lungo un continuum una serie di stimoli utilizzando le valutazioni (giudizi o preferenze) espresse in termini ordinali da un gruppo di individui.
Il procedimento richiede che ciascun soggetto (appartenente ad un gruppo detto *P-sample*) ordini sistematicamente (*Q-sort*) un insieme di stimoli (selezionati secondo un determinato criterio e che compongono un *Q-sample*) utilizzando la sua personale esperienza (struttura di riferimento interna del soggetto) (McKeown, 1988).

Definizione del *Q-Sample*

Come si è detto l'insieme degli stimoli da sottoporre ai soggetti è detto *Q-Sample*. L'individuazione di tali stimoli – come succede anche per gli altri modelli di *scaling* – rappresenta un passaggio delicato in quanto devono soddisfare l'assunto di unidimensionalità. In altre parole gli stimoli che devono essere confrontati tra loro devono provenire dallo stesso specifico dominio di contenuto.
Nell'ambito della metodologia *Q* la individuazione degli stimoli può avvenire secondo una definizione:

- *naturale*, quando sono individuati direttamente dai soggetti coinvolti nella rilevazione attraverso interviste, racconti scritti o altre fonti,
- *costruita*, quando sono individuati a partire da altre fonti.

Il principale vantaggio dei *Q-Sample* naturali è dato dal fatto che il significato è attribuito agli item dai soggetti stessi facilitando il compito loro richiesto.
I *Q-Sample* costruiti presentano il vantaggio di poter essere utilizzati in più contesti. Ciò ha consentito la definizione di *Q-Sort* standardizzati. E' possibile costruire *Q-Sample* a partire da item utilizzati per altri tipi di *scaling* che utilizzano, per esempio, scale di *rating*.
Tenuto conto che gli stimoli possono essere rappresentati non solo da affermazioni ma anche da foto, figure, ecc., si comprende come non sia sempre facile verificare che gli stimoli provengano dallo stesso universo di contenuto ovvero condividano una struttura di riferimento comune (il problema è analogo a quello della definizione della popolazione di individui che il campione dovrebbe rappresentare); ciò vale ancora di più quando si misurano dimensioni quali sentimenti, preferenze, valori, opinioni e tratti di personalità (McKeown, 1988).
E' possibile comunque attraverso l'analisi statistica verificare l'omogeneità degli stimoli.
Uno dei principali problemi da affrontare nella individuazione del *Q-Sample* è quello della definizione del numero di stimoli che dovranno comporlo. Tale definizione deve anche tenere conto della disponibilità e capacità dei soggetti che comporranno il *P-Sample*.

Definizione del *P-Sample*

Dati gli obiettivi della metodologia, l'individuazione degli individui cui sottoporre gli stimoli rappresenta un altro passaggio importante. A tale proposito si distingue tra

- campioni *estensivi*, che mirano ad essere rappresentativi della popolazione di cui si intendono esplorare gli atteggiamenti; i fattori che emergono possono essere considerati generalizzazioni di atteggiamenti presentati dai soggetti che definiscono un dato fattore; il rischio è quello di non vedere rappresentati tutti i fattori di atteggiamento esistenti;
- campioni *intensivi*, rappresentati da singoli casi selezionati, detti *spicemen*, per lo studio dei quali il metodo *Q* è particolarmente adatto; uno *spicemen* rappresenta un individuo ritenuto, per le sue caratteristiche, rappresentativo del tipo di soggettività che si vuole esaminare; l'analisi dei dati prodotti da tali campioni ha l'obbiettivo di esplorare le dinamiche della soggettività intraindividuale, precedentemente individuata nell'analisi estensiva. La validità di tali campioni è legata al criterio con il quale sono stati individuati.

I due approcci al campionamento non sono tra loro alternativi ma possono essere complementari.

2.2.1 *Q-Sort*

Successivamente alla definizione dei due campioni si procede alla rilevazione dei *Q-Sort*. Il procedimento di rilevazione richiede che il soggetto ordini gli stimoli secondo il criterio definito (per esempio il livello di accordo[11]) e seguendo le istruzioni indicate (*Condition of Instruction*, *CI*).
In alcuni casi è possibile richiedere al soggetto di ordinare lo stesso gruppo di stimoli più volte secondo più criteri (preferenze, valutazioni, ecc.).

La forma della distribuzione

Si chiede a ciascun soggetto di procedere all'ordinamento degli stimoli seguendo istruzioni che riguardano:

- il *numero* di gruppi/categorie entro i quali deve essere ordinati gli stimoli; in genere si raccomanda di utilizzare un numero piuttosto grande di raggruppamenti che consente una migliore discriminazione tra gli stimoli (l'effetto è quello di aumentare l'affidabilità del *sort*)[12];
- la *forma* della distribuzione dell'ordinamento degli stimoli; nella maggior parte dei casi si preferisce adottare la forma *normale* in quanto si tratta di una distribuzione che trova un adattamento a molti fenomeni osservati.

La critica rivolta a questo approccio riguarda il fatto che al termine della rilevazione tutti gli ordinamenti di tutti i soggetti presentano la stessa media e deviazione standard; in questa ottica, tale approccio appare come una violazione della soggettività in quanto non consente di conoscere il reale livello di risposta individuale. D'altra parte l'obiettivo dell'applicazione della metodologia Q non è quello di misurare gli individui ma di utilizzare le valutazioni individuali per collocare gli stimoli lungo un continuum. In ogni caso la violazione, da parte del soggetto, della forma della distribuzione non compromette né l'affidabilità della tecnica né la qualità dei dati.
Facciamo un esempio. Poniamo di voler studiare le preferenze rispetto a 100 statue con origini e stili molto diversi tra loro (Nunnally, 1978). A ciascun soggetto si chiede di esprimere le preferenze confrontando le fotografie delle statue e mettendole in ordine decrescente di preferenza. L'ordinamento deve rispettare la distribuzione normale. A tal fine, prima di presentare le 100 fotografie, a ciascun soggetto viene presentato il seguente schema che dovrà guidare il soggetto nella indicazione delle preferenze in modo che la distribuzione di queste risultino nella forma voluta:

	Numero di fotografie da inserire in ciascun gruppo											
	2	4	8	12	14	20	14	12	8	4	2	
Minore preferenza	-5	-4	-3	-2	-1	0[13]	1	2	3	4	5	Maggiore preferenza
	0	1	2	3	4	5	6	7	8	9	10	
	Numero di riferimento di ciascun raggruppamento											

Dopo aver mescolato le foto, per ridurre l'effetto che potrebbe avere l'ordine di presentazione, il ricercatore chiede al soggetto, prima di procedere all'ordinamento, di osservare tutte le raffigurazioni una dopo l'altra e di stenderle successivamente sul tavolo per poterle meglio confrontare.

[11] Per esempio: *ordina gli item da quello con cui sei più in accordo (+5) a quello con cui sei più in disaccordo (-5).*

[12] In un certo senso, l'adozione di una particolare forma di distribuzione può essere considerata come la definizione anticipata del numero di parimerito tra gli stimoli.

[13] Il punto "zero" non corrisponde ad un punto medio, ma ad un punto "neutrale" nel significato e senza significato psicologico.

Si può suggerire al soggetto di procedere all'ordinamento delle fotografie a partire dai due estremi del continuum verso il centro, in quanto gli estremi dei primi due raggruppamenti di solito vengono distinti più velocemente e facilmente. In particolare si chiede al soggetto di partire dall'estremo della "maggiore preferenza" verso il centro e di identificare le due statue maggiormente preferite e di metterle nel gruppo 10. A partire dalle 98 rimanenti fotografie il soggetto preleva le quattro successive statue preferite che devono essere posizionate nel gruppo 9. A partire dalle rimanenti 94 foto il soggetto individua le otto successive statue preferite da sistemare nel gruppo 8. Il soggetto continua così seguendo lo schema sopra presentato. Alla fine al soggetto si chiede di controllare l'intero *rating* per essere sicuro che le fotografie non siano fuori posto.
Al termine le fotografie risulteranno essere distribuite secondo il seguente schema:

Lontano dalla mia preferenza ←						→ Vicino alla mia preferenza				
-5	-4	-3	-2	-1	0	1	2	3	4	5
(3)	(4)	(4)	(7)	(7)	(10)	(7)	(7)	(4)	(4)	(3)
foto	foto	foto	foto	foto	foto	foto	foto	foto	foto	foto
foto	foto	foto	foto	foto	foto	foto	foto	foto	foto	foto
foto	foto	foto	foto	foto	foto	foto	foto	foto	foto	foto
	foto	foto	foto	foto	foto	foto	foto	foto	foto	
			foto	foto	foto	foto	foto			
			foto	foto	foto	foto	foto			
			foto	foto	foto	foto	foto			
					foto					
					foto					
					foto					

Se si chiede al soggetto si eseguire un altro *Q-Sort*, utilizzando lo stesso *Q-Sample* secondo diverse criteri, è consigliabile mescolare le foto prima di rieseguire l'operazione. L'esecuzione aggiuntiva di altri *Q-Sort* (purché non numerosi) sullo stesso *Q-Sample* può essere utile in quanto il soggetto ha già familiarità con il gruppo di stimoli e con il procedimento di ordinamento.

2.2.2 Analisi dei dati

I dati ottenuti con la metodologia *Q* vengono sottoposti ad un'analisi, molto criticata, che richiede l'esecuzione sequenziale di tre procedimenti statistici (McKeown, 1988):

a. <u>analisi della correlazione</u> tra gli N *Q-Sort* (le variabili sono rappresentate dai casi);
b. <u>analisi fattoriale</u> della matrice di correlazione $N * N$. I fattori prodotti rappresentano i punti di vista e l'associazione tra ciascun soggetto e ciascuno dei punti di vista è indicato dalla dimensione del valore del *factor loading*. Si assume che gli individui associati in modo significativo con un dato fattore condividano una prospettiva comune. Quindi il valore del *factor loading* indica il grado di associazione (positiva o negativa) tra ciascun *Q-Sort* individuale e l'atteggiamento composito o la prospettiva riferita dal fattore. I diversi fattori ortogonali evidenziano diversi punti di vista presenti nel campione di soggetti. L'interpretazione dei fattori non assume alcuna definizione a priori ma è dedotta dai *Q-Sort* individuali. L'analisi fattoriale è fondamentale per la metodologia *Q* in quanto rappresenta lo strumento statistico attraverso il quale raggruppare i soggetti sulla base dei *Q-Sort*. Dati i suoi obiettivi, l'approccio fattoriale qui utilizzato si presenta in modo diverso da quello classico (Maggino, 2005).[14]

[14] A tale proposito ricordiamo la distinzione tra analisi fattoriale *Q* e analisi fattoriale *R* (classica). Tecnicamente ciò che caratterizza l'analisi fattoriale *Q*, rispetto all'analisi *R*, è l'inversione della matrice dei dati sulla quale viene effettuata. Tale distinzione, seppure corretta, è insufficiente in quanto continuano ad esserci equivoci sul suo reale significato. L'inversione richiede che la matrice di correlazione sulla quale viene effettuata l'analisi fattoriale sia calcolata su dati non omogenei. A tale proposito si ricorda che dal punto di vista logico non si può parlare di correlazione tra casi ma di prossimità tra casi. La determinazione delle prossimità tra i casi richiede l'applicazione di tecniche particolari

c. calcolo dei punteggi fattoriali ovvero si attribuisce a ciascuno stimolo del *Q-Sample* un punteggio per ciascun fattore. Tali punteggi fattoriali consentono di comprendere e interpretare i significati dei fattori. Per il calcolo dei punteggi fattoriali dei particolari pesi (*score*) che tengano conto delle differenze tra i *Q-Sort*. L'espressione indicata da Spearman per il calcolo degli *score* è la seguente:

$$w = \frac{f}{1 - f^e}$$

dove
f *factor loading*
w peso fattoriale

I *factor score* sono calcolati come *punti-z* ma per convenienza vengono convertiti in numeri interi (da +5 a -5) per facilitare i confronti tra fattori.

Nel caso in cui si utilizzino *Q-Sample* sia estensivi che intensivi, il procedimento di analisi prevedere tre diverse fasi:

a. *fase estensiva*: un insieme di soggetti ordina uno stesso *Q-Sample* secondo le stesse condizioni; dopo aver correlato i *Q-Sample* e averli sottoposti ad analisi fattoriale, si identificano e si interpretano i fattori ottenuti; tale fase consente di verificare la corretta comprensione delle strutture di riferimento dei soggetti e l'affidabilità e la validità degli strumenti;
b. *fase di selezione*: si selezionano i casi che meglio rappresentano i punti di vista (fattori) emersi con la precedente analisi; tali casi sono quelli la cui *comunanza* è maggiormente spiegata da un determinato fattore;
c. *fase intensiva*: dopo aver selezionato uno o più soggetti, si procede alla somministrazione di un nuovo il *Q-Sort*. Occorre ricordare che nella metodologia *Q* pochi soggetti sono psicometricamente accettabili in quanto la prospettiva osservativa è rappresentata dal soggetto stesso. Qualsiasi spiegazione avanzata dai ricercatori è sottomessa alla struttura di riferimento del soggetto resa operativa nel *Q-Sort*. E' per questa ragione che le verifiche di affidabilità e di validità, centrali negli altri approcci sono completamente inutili all'interno dello schema psicometrico della metodologia *Q*.

(Maggino, 2005). Una volta costruita la matrice di correlazione, l'approccio matematico per estrarre i fattori è in teoria lo stesso utilizzato nell'analisi *R*.

3. I MODELLI CUMULATIVI. L'APPROCCIO DETERMINISTICO

Il modello deterministico ha l'obbiettivo di classificare sia i soggetti che gli item lungo un continuum. Esso è detto *deterministico* (Torgerson, 1958) in quanto nella sua formulazione è prevista la definizione di un modello ideale di riferimento del quale deve essere verificato l'adattamento ai dati osservati. In realtà tale approccio non consiste tanto nel determinare se il modello si adatta o meno ai dati quanto piuttosto nell'utilizzare il modello come approssimazione adeguata ai dati; in altre parole, essendoci sempre, tranne casi eccezionali, uno scarto tra previsione e realtà, il problema non è quello di controllare se c'è differenza, ma se questa rientra in certi margini di tolleranza; quindi nei casi di mancata corrispondenza perfetta tra modello e dati, il modello può servire come approssimazione ai dati reali. Ciò fa sorgere due importanti questioni:

- come calcolare lo scarto tra dati ideali e dati osservati,
- come definire i margini di tolleranza.

A tale proposito sono state sviluppate diverse tecniche che servono a indicare la bontà dell'approssimazione. Dato che i modelli ideali possono servire per rappresentare i dati empirici all'interno di un certo errore, espresso in termini di differenza tra il risultato atteso e il risultato ottenuto, la definizione di questo consente di valutare, attraverso un indice, il grado di approssimazione all'ideale.

Secondo il modello deterministico (Torgerson, 1958) per poter valutare se l'insieme di item consente di misurare sia gli item che i soggetti è necessario assumere che l'attributo rispetto al quale tutti gli item e i soggetti vengono misurati sia *scalabile*. Individuato un tale attributo (status socio-economico, pregiudizio, ecc.) è necessario fare due assunti:

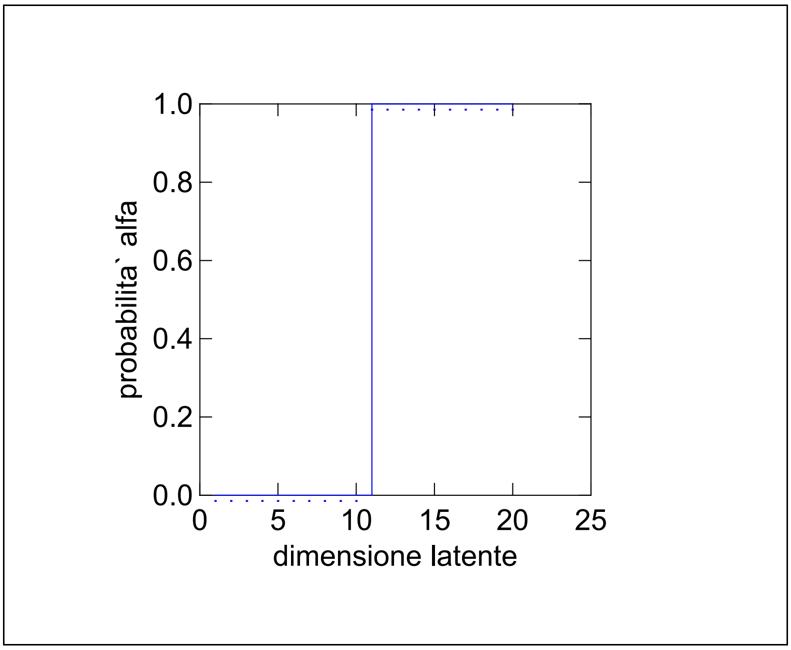
- i soggetti possono essere classificati rispetto a tale attributo su un singolo continuum;
- esiste una relazione tra gli item e il continuum postulato.

A questo punto è possibile definire e selezionare gli item che insieme devono rispondere a determinate caratteristiche. A tale proposito è possibile distinguere due diversi approcci: *Guttman* e alternativo, cui corrispondono due tipi di item.

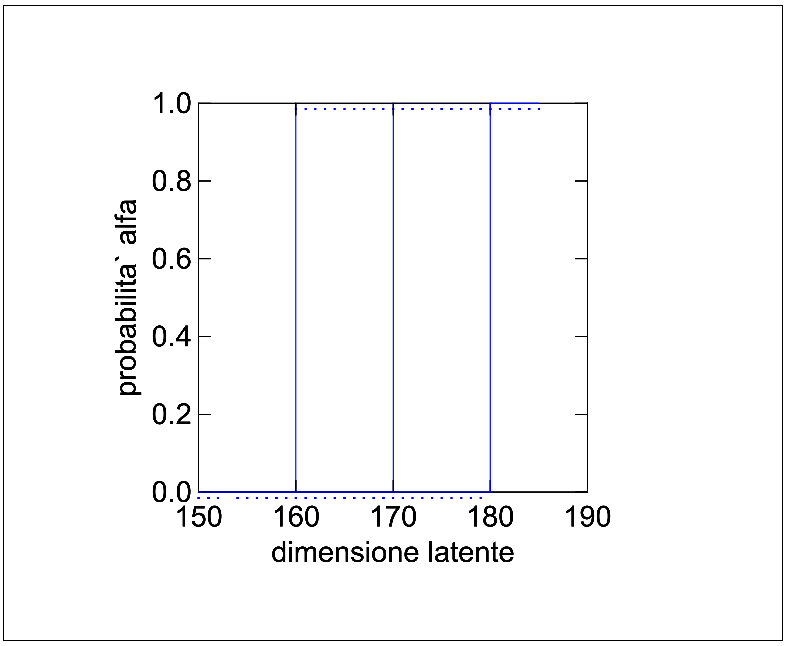
In termini geometrici ciascun item rappresenta, come per il modello additivo, una ripetizione della dimensione sottostante anche se, come vedremo, in punti diversi del continuum.

3.1 L'IPOTESI DI SCALOGRAMMA

Gli item utilizzati dal modello deterministico sono detti *monotoni* e sono definiti da una *trace line* in cui fino ad un certo punto del continuum dell'attributo la probabilità di risposta *alfa* è 0 (mentre *beta* è 1); oltre tale punto la probabilità di risposta *alfa* è 1; tale tipo di *trace line* è basata su una funzione detta *step function* e ha la seguente forma:



Concettualmente gli item che soddisfano il modello di *scaling Guttman* (Guttman, 1945, 1947;) possono essere visti in successione su una sola dimensione. In questa ottica ogni item rappresenta, alla luce dell'ipotesi, un punto intermedio fra l'item precedente e il successivo ovvero deve discriminare perfettamente in un particolare punto del continuum relativo all'attributo, diverso dagli altri.
Di seguito sono raffigurate le tre *trace line* corrispondenti a tre item che discriminano in tre punti diversi e che quindi soddisfano i requisiti del modello deterministico monotono:



Il modello può quindi essere applicato quando è necessario misurare caratteristiche e disposizioni per le quali è possibile identificare livelli crescenti di intensità come capacità, impegno, difficoltà, ecc.; quindi il gruppo di item che soddisfà il modello cumulativo *Guttman* deve essere in grado di misurare e ricoprire la variabilità di caratteristiche, disposizioni, difficoltà, capacità crescenti; questo vuol dire che essi devono essere ordinabili ovvero devono rispondere al concetto di *scalabilità*. Nel misurare capacità, per esempio, in pratica dovremmo ottenere che gli item considerati facili devono poter essere superati da tutti mentre gli item considerati difficili devono essere superati da alcuni (ricordiamo che

maggiore è la proporzione degli individui che superano un item, più facile e meno discriminante esso si presenta).
Quando
- gli item misurano livelli crescenti di intensità (gli item discriminano in punti diversi e ordinati sul continuum),
- i soggetti presentano livelli crescenti di intensità (i soggetti che rispondono positivamente ad un item rispondono positivamente anche ai precedenti e i soggetti che rispondono negativamente ad un item rispondono negativamente anche ai successivi),

è stato identificato il modello ideale che assume la forma dello *scalogramma* e che riflette la caratteristica di *perfetta scalabilità* (Arcuri, 1974). Lo scalogramma perfetto presenta una matrice con due triangoli, uno formato da tutti segni "+" (risposte positive) e uno formato da tutti segni "-" (risposte negative); con i dati del nostro esempio (tre item e sei soggetti) è stato raggiunto lo scalogramma perfetto:

Soggetti	Item		
	a	b	c
1	+	-	-
2	+	+	-
3	+	+	-
4	+	+	-
5	+	+	+
6	+	+	+
"+": risposta positiva			
"-": risposta negativa			

Quando viene identificato il modello ideale triangolare è possibile misurare ovvero ordinare lungo il continuum
- ciascun item, sulla base della proporzione di risposte positive registrata,
- ciascun soggetto, sulla base del numero di risposte positive rilevate (punteggio totale individuale).

L'identificazione del modello triangolare, secondo i criteri visti, consente di stabilire con esattezza a quali item ciascun soggetto ha risposto positivamente e a quali negativamente; la possibilità di prevedere per ciascun soggetto le risposte date a tutti gli item sulla base del punteggio totale rappresenta la prova che un item appartiene ad una singola dimensione sottostante.

Una scala deterministica, ma in generale le scale cumulative, semplifica l'informazione contenuta nei dati riducendo la singola matrice *n***k* a due vettori separati:
- *n**1 per le posizioni dei soggetti,
- *k**1 per le posizioni degli item.

		Matrice di input V: *two-way, two-mode* v_{ij} : presenta "1" se *i* domina *j*; "0" se *j* domina *i*					
		Oggetti-colonna (item)					
		1	2	…	*j*	…	*k*
Oggetti-riga (soggetti)	1	v_{11}	v_{12}	…	v_{1j}	…	v_{1k}
	2	v_{21}	v_{22}	…	v_{2j}	…	v_{2k}
	…	…	…	…	…	…	…
	i	v_{i1}	v_{i2}	…	v_{ij}	…	v_{ik}
	…	…	…	…	…	…	…
	n	v_{n1}	v_{n2}	…	v_{nj}	…	v_{nk}

Matrici di output					
X: punteggi dei soggetti			Y: punteggi degli item		
x_i : numero di punti-item posizionati a sinistra del soggetto i lungo la dimensione			y_i : numero di punti-soggetto posizionati a sinistra del punto-item lungo la dimensione		
soggetti	1	x_1	item	1	y_1
	2	x_2		2	y_2
	…	…		…	…
	i	x_i		*i*	y_i
	…	…		…	…
	n	x_n		*k*	y_k

Il <u>modello di *scaling*</u> riguarda

- i punti degli *n* soggetti ordinati lungo la dimensione secondo il numero di stimoli che dominano,
- i *k* punti-item ordinati lungo la dimensione secondo il numero di soggetti che dominano ciascuno di essi.

In termini geometrici il valore di ciascuna cella nella matrice dei dati fornisce un'informazione sulla coppia di punti soggetto-item. Tale valore è:

- "1" se il punto-soggetto è posizionato a destra del (domina il) punto-item lungo la dimensione sottostante,
- "0" se il punto-item è posizionato a destra del (domina il) punto-soggetto.

Ciascun punto è posizionato lungo la dimensione sommando il numero di "1" contenuti all'interno della propria riga o colonna della matrice dei dati.
La posizione del punto-soggetto sarà tanto più a destra quanto più grande è il numero di risposte positive date dal soggetto, in quanto domina tanti item. La posizione del punto-item sarà tanto più a sinistra quanto più grande è il numero di risposte positive ottenute dall'item, in quanto è dominato da tanti più soggetti.
Una *scala Guttman* composta da dati dicotomici rappresenta l'operazionalizzazione più semplice del modello di *scaling* cumulativo-deterministico. Il modello può comunque essere generalizzato anche ad item non dicotomici. L'interpretazione dei punti-item risulta in questo caso però un pò diversa. Per quanto si è detto il punto-item può essere interpretato come un *cut-point* lungo la dimensione in quanto serve come confine tra risposte positive e risposte negative per quell'item. In altre parole se tutti i soggetti ad un determinato item danno risposte

- negative, i loro punti saranno posizionati a sinistra del *cut-point*,
- positive, i loro punti saranno posizionati a destra del *cut-point*.

Nel caso di item dicotomici il punto-item e il corrispondente *cut-point* coincidono. Nel caso di item non dicotomici si ottengono più *cut-point*; il numero di *cut-point* è uguale a *q*-1, dove *q* corrisponde al numero di categorie. Gli item non dicotomici suddividono la dimensione in *q* segmenti. Ciò in realtà non complica lo *scalogramma* che viene costruito nello stesso modo. La differenza sta nel fatto che la matrice dei dati rappresentano in modo esplicito i *cut-point* invece che gli item; questo vuol dire che la matrice degli item dovrà presentare più colonne.

3.2 LA VERIFICA DEL MODELLO: LO SCALOGRAM ANALYSIS

Tale modello è basato sui seguenti assunti:

- *omogeneità*: tutte le grandezze rilevate esprimono una quantità della stessa natura in modo da legittimare la riunione di più item in un unico punteggio[1]; in altre parole il criterio è soddisfatto quando tutte le grandezze rilevate (risposte agli item) sono quantità della stessa natura e possono essere riunite in un punteggio generale che rappresenta la misura di un solo fattore;

- *esuastività*: l'insieme degli item deve rappresentare un inventario completo del dominio reale di una "dimensione" ovvero gli item devono ricoprire tutta la variabilità osservabile in modo da consentire una valutazione globale;
- *unidimensionalità*: l'insieme degli item è determinato da un insieme di attitudini strettamente connesse e/o dipendenti da una sola dimensione;
- *gradualità/scalabilità*: gli item devono essere scelti in modo tale che risultino essere superabili con livelli diversi della stessa attitudine; in altre parole deve essere possibile ordinare gli item secondo un livello crescente di intensità (capacità, disposizioni, ecc.); gli item così selezionati presentano una parziale sovrapposizione di significato; ciò consente di ottenere una *gradualità* della valutazione.

Se tali assunti vengono soddisfatti ne consegue la giustificazione teorica di un punteggio globale.
Il modello teorico si realizza perfettamente quando, per superare un item, sono indispensabili tutte le attitudini utilizzate per l'item precedente più un'attitudine aggiuntiva. Per poter verificare l'adattamento del modello ai dati, ovvero per poter determinare se un campione di item e un campione di soggetti sono conformi allo specifico insieme di criteri considerati requisiti dello *scaling* Guttman, si applica un procedimento detto *scalogram analysis*. In particolare l'analisi dello scalogramma consente di verificare se e in che misura la distribuzione reale delle risposte si discosta dalla distribuzione ideale o dalla combinazione ideale (scalogramma) ovvero se le risposte riproducono il modello triangolare teorizzato; quindi l'analisi dello scalogramma consente il confronto di una distribuzione concreta di soggetti rispetto a un modello teorico di perfetta scalabilità. Si registra una scala perfetta quando tutti i modelli individuali seguono questo andamento, ossia nessun individuo supererà un item se non ha superato un item più facile per l'intero gruppo.
Se gli item e i soggetti coprono la variabilità della caratteristica ordinale misurata, la distribuzione delle frequenze registrate dovrebbe presentare un numero decrescente di risposte dello stesso tipo. Se per esempio si dispone di 5 item dicotomici con difficoltà crescente e 10 soggetti con diverse attitudini si dovrebbero ottenere le seguenti distribuzioni di frequenza:

item	Numero di "si"	Numero di "no"
1	10	0
2	7	3
3	5	5
4	3	7
5	0	10

Per illustrare la procedura vediamo un semplice esempio in cui vi è un perfetto adattamento al modello deterministico; naturalmente la procedura può essere estesa ai casi più complessi. Poniamo di avere ottenuto, con 5 item dicotomici a 6 soggetti, i seguenti risultati:

[1] Nel caso di alta consistenza interna tra item, si può non solo affermare che la scala è omogenea, ma anche che le risposte degli intervistati sono coerenti. Questa deduzione è la diretta conseguenza del fatto che gli intervistati, nella grande maggioranza dei casi, hanno compreso le domande, ossia che la scala, così come è stata applicata, è idonea per lo studio.

Soggetti	Item				
	1	2	3	4	5
a	-	+	+	+	+
b	-	-	+	-	+
c	-	-	-	-	+
d	+	+	+	+	+
e	-	-	-	-	-
f	-	+	+	-	+
P	0.17	0.50	0.67	0.33	0.83

dove
- "+" indica *superamento dell'item/della prova* o *risposta affermativa*,
- "-" indica *non superamento dell'item/fallimento della prova* o *risposta negativa*.

A questo punto assegniamo alla risposta
- "+" (superamento dell'item/della prova o risposta affermativa) punteggio *1*,
- "-" (non superamento dell'item/fallimento della prova, risposta negativa) punteggio *0*.

Alla matrice così modificata aggiungiamo:
- il punteggio totale di ogni soggetto (PT),
- la proporzione di risposte positive per ogni item (P).

Soggetti	Item					Punteggio (PT)
	1	2	3	4	5	
a	0	1	1	1	1	4
b	0	0	1	0	1	2
c	0	0	0	0	1	1
d	1	1	1	1	1	5
e	0	0	0	0	0	0
f	0	1	1	0	1	3
P	0.17	0.50	0.67	0.33	0.83	

Avendo selezionato gli item sulla base dell'assunto di *scalabilità* si ricostruisce e si verifica tale ordine in sede d'analisi in funzione della proporzione di soggetti che hanno superato la prova. A tal fine si modifica

A. la posizione degli item (colonne della matrice) in modo tale che risultino ordinati secondo la proporzione di risposte positive (dalla più alta alla più bassa);
B. la posizione dei casi (righe della matrice) in modo tale che i soggetti siano ordinati secondo il punteggio totale (dal più alto al più basso).

A								B						
SOGGETTI	ITEM					PT		SOGGETTI	ITEM					PT
	5	3	2	4	1				5	3	2	4	1	
a	1	1	1	1	0	4		d	1	1	1	1	1	5
b	1	1	0	0	0	2		a	1	1	1	1	0	4
c	1	0	0	0	0	1		f	1	1	1	0	0	3
d	1	1	1	1	1	5		b	1	1	0	0	0	2
e	0	0	0	0	0	0		c	1	0	0	0	0	1
f	1	1	1	0	0	3		e	0	0	0	0	0	0
P	0.83	0.67	0.50	0.33	0.17			P	0.83	0.67	0.50	0.33	0.17	

Riattribuendo alle risposte *1* il segno + e alle risposte *0* il segno - potremo osservare come tali dati riproducano in modo preciso il modello triangolare (*perfetta scalabilità*): in ogni profilo, dopo un insuccesso in una prova, si incontrano solo insuccessi nelle prove più difficili; un punteggio di "4" indica che il soggetto ha superato i primi quattro item ma non l'ultimo:

B						
SOGGETTI	ITEM					PT
	5	3	2	4	1	
d	+	+	+	+	+	5
a	+	+	+	+	-	4
f	+	+	+	-	-	3
b	+	+	-	-	-	2
c	+	-	-	-	-	1
e	-	-	-	-	-	0

In questa breve analisi il modello ideale di riferimento è stato perfettamente osservato; nella pratica è però molto difficile poter osservare una perfetta riproduzione del modello triangolare; per questo motivo, per stabilire se il modello costituisce un'adeguata rappresentazione dei dati empirici, occorre definire il livello di deviazione tollerabile (valutazione della bontà di adattamento). Per fare ciò non basta la semplice osservazione delle distribuzioni di frequenza, occorre effettuare una validazione più approfondita, basata sui concetti di *riproducibilità*, *scalabilità* e *predicibilità*.

- <u>*Riproducibilità*</u>: *possibilità di riprodurre per ciascun soggetto, a partire dal punteggio totale, le risposte date a ciascun item*. Per tutti i soggetti e tutti gli item, è possibile calcolare la percentuale di riproducibilità.
- <u>*Scalabilità*</u>: *osservazione di item a difficoltà crescente*; l'osservazione della scalabilità consente di introdurre il concetto di predicibilità.
- <u>*Predicibilità*</u>: *possibilità di inferire dalla risposta data da un soggetto ad un item posto ad una certa soglia di difficoltà la risposta data alla domanda posta al di sotto di tale soglia*.

La predicibilità fra le risposte di item contigui è ricavata da due diversi punti di vista:

1. dall'item più facile al più difficile: chi non ha superato un item non supera l'item successivo più difficile: P_i;
2. dal più difficile al più facile: chi ha superato un item deve aver superato l'item precedente più facile: P_{i+1}.

Nel caso di predicibilità perfetta se un soggetto supera il terzo item deve aver superato necessariamente anche il secondo e il primo. A ciascun item si attribuisce un punteggio nel modo seguente:

- all'item *predetto* dall'item precedente, ovvero quando i soggetti che hanno superato l'item ha superato anche l'item precedente più facile, si attribuisce punteggio "1";
- all'item non *predetto* dall'item precedente (*indifferenza predittiva*), ovvero quando tra i soggetti che hanno superato l'item solo il 50% supera l'item precedente più facile, si attribuisce punteggio "0".

La perfetta *scalabilità* e la perfetta *predicibilità* possono verificarsi solo nei casi in cui esistano item più "facili" essenziali per la definizione di quelli più "difficili". Per questo lo *scaling* Guttman è particolarmente adatto a misurare attitudini e qualità crescenti.

3.2.1 Deviazione dal modello: l'errore

La procedura deterministica per stabilire se item e soggetti rispondono al modello (*scalogram analysis*) è basata sull'analisi dei modelli di risposta dei singoli soggetti, *response pattern*, all'insieme di item. Un <u>modello di risposta</u> indica l'insieme di risposte date da un soggetto agli item. Con *n* item dicotomici vi sono 2^n modelli di risposta possibili; se gli item formano davvero uno *scaling* cumulativo, si devono verificare solamente *n*+1 di tali modelli.
La presenza in un profilo individuale di una risposta positiva preceduta da una negativa costituisce

un errore elementare; quindi i profili individuali che registrano errori di scalabilità sono quelli che presentano passaggi diversi da insuccesso a successo, e seguono andamenti del tipo:

−+−−− +−+−+− ++−++ −−+−+ −−++−

E' proprio l'analisi di tali errori che conduce ad una valutazione globale dell'insieme di item. La mancata occorrenza dei modelli devianti consente alla procedura dello scalogramma di ordinare gli individui, gli item e gli intervalli categorici sul continuum sottostante a partire dai dati osservati (*scalogramma perfetto*).

Le deviazioni osservate dal modello perfetto e dalla forma ideale di scalogramma richiesta sono definiti *errori*. Si assume che la quantità di deviazione, o errore, osservata sia una funzione del fallimento degli item e dei soggetti nel conformarsi alle procedure di ordinamento.

La determinazione della quantità di errore non è però univoca: è possibile infatti identificare principalmente due forme di conteggio dell'errore cui corrispondono due approcci diversi alla valutazione dello *scaling*.

3.2.2 Tecniche di valutazione dell'errore

Tecnica Cornell: ***Minimizzazione dell'errore*** (Guttman, 1947)

La tecnica Cornell è basata sul criterio di minimizzazione dell'errore di una serie di modelli di risposte. L'assunto base di questa tecnica è che nessun item può possedere più errore che *non-errore*. L'ordinamento degli item sul continuum sottostante è una funzione di minimizzazione dell'errore tra le risposte osservate. Tale tecnica presenta il vantaggio di consentire inversioni nell'ordinamento degli item basati sul decremento delle probabilità marginali per raggiungere la più alta riproducibilità possibile. L'ordinamento degli item e i punteggi possono essere una funzione degli errori casuali anziché rappresentare un costrutto di risposte sottostanti agli item. *Il numero di errori rappresenta il numero minimo di risposte positive che devono essere cambiate in negative o di negative che devono essere cambiate in positive per trasformare le risposte osservate nel modello ideale di risposta*. Vediamo un esempio in cui per ciascun profilo di risposta (su quattro item) è riportato il corrispondente conteggio degli errori presenti:

Profili di risposta ⇒	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-
	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Assegnazione dell'errore ⇒	0	0	1	0	1	1	1	0	1	1	2	1	2	1	1	0

Il criterio di minimizzazione dell'errore (suggerito da Guttman) appare debole se confrontato con la teoria alla base del modello *Guttman*, in quanto, focalizzandosi solo sugli errori di risposte positive o negative, sottovaluta gli errori dovuti alla posizione. Vediamo un esempio; secondo la tecnica della minimizzazione dell'errore descritta, il modello di risposta -++- osservato contiene un solo errore rispetto al modello ideale più vicino che è +++-. Visto da un altro punto di vista però il profilo di risposta osservato potrebbe riflettere in realtà due errori: infatti se consideriamo che il modello ideale più vicino è ++-- (due risposte positive e quindi stesso punteggio), per trasformare il profilo osservato (-++-) in quello ideale il numero minimo di segni che devono essere cambiati è due. In questo senso con l'applicazione di questo criterio lo *scaling Guttman* perde una certa componente d'interpretabilità in quanto conduce ad un indebolimento dell'assunto cumulativo su cui è basata e ad una contraddizione della teoria alla base dello scalogramma.

Tecnica Goodenough-Edwards: ***Deviazione dalla Perfetta Riproducibilità*** (1957)

Una diversa procedura, nota con il nome di *GoodEnough-Edwards*, è basata su due principi tra loro

collegati:
1. il modello ideale di risposte di un soggetto è funzione diretta del numero di item cui il soggetto ha risposto in modo positivo;
2. gli item devono essere perfettamente riproducibili a partire dalle risposte dei soggetti.

Quindi l'errore deve essere
- misurato rispetto al numero di risposte che si allontanano dai modelli previsti,
- assegnato sulla base dell'assunto della perfetta riproducibilità,

assicurando il massimo rispetto della posizione degli item e dell'ordinamento dei soggetti.
Aldilà della maggiore plausibilità teorica, questa tecnica presenta anche alcuni vantaggi pratici in quanto l'analisi è basata sulle risposte osservate anziché sulle risposte derivate dalla procedura di minimizzazione dell'errore. Vediamo come vengono valutati gli errori se riprendiamo i profili con quattro item, precedentemente presentati:

Profili di risposta ⇒	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-
	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Assegnazione dell'errore ⇒	0	0	2	0	2	2	2	0	2	2	2	2	4	2	2	0

Anche se tale procedura conduce all'identificazione di un numero maggiore di errori rispetto al metodo precedente, rappresenta comunque una descrizione più accurata dei dati basati sulla teoria dello scalogramma.

3.2.3 Valutazione dell'adattamento del modello

Dato che è difficile raggiungere una perfetta corrispondenza tra modello teorico e modello osservato, è necessario stabilire, sulla base degli errori rilevati, se il modello osservato risponde alle caratteristiche dal modello teorico di *scaling* cumulativo ovvero è necessario verificare se i modelli di risposta osservati riflettono i modelli ideali di risposte. A tale fine sono stati definiti alcuni indici e coefficienti (McIver, 1979).

❖ COEFFICIENTE DI RIPRODUCIBILITA' PER CIASCUN ITEM (CR_i)

Sappiamo che la scalabilità è funzione di quanto i modelli di risposta osservati possano essere accuratamente riprodotti sulla base dei punteggi assegnati. Per questo per valutare il livello di scalabilità dei dati empirici ovvero per verificare complessivamente la predicibilità per ciascun item, Guttman propose un coefficiente che, confrontando gli errori osservati con tutti gli errori possibili, può essere interpretato *errore di riproducibilità*:

$$CR_i = 1 - \frac{n_{ie}}{n}$$

dove
n_{ie} errori di riproducibilità tra l'item i e l'item $i+1$ (ovvero numero totale di errori per l'item i)
n numero di risposte all'item (o numero di soggetti).

I valori di CR_i compresi tra 0.85 e 1 sono indici di bontà dello scalogramma. Secondo Torgerson (1958) gli item che presentano valori di CR_i minori di 0.85 dovrebbero essere scartati in quanto, non soddisfacendo il postulato della predicibilità, rivelano una loro appartenenza a una diversa dimensione.

❖ COEFFICIENTE DI RIPRODUCIBILITA' PER TUTTI GLI ITEM (CR)

Eliminati gli item con bassi valori CR_i, si calcola il coefficiente di riproducibilità per l'insieme di item, che dovrebbe assumere un valore centrale fra i valori dei singoli coefficienti degli item grazie alla compensazione degli errori. Per l'insieme di item il coefficiente di riproducibilità (CR) è interpretabile come la *proporzione di risposte agli item che possono essere correttamente riprodotte conoscendo il punteggio totale del soggetto* e rappresenta una misura della bontà di adattamento tra il modello di risposte osservato e quello ideale o previsto:

$$CR = 1 - \frac{\sum n_{ie}}{N}$$

dove

N numero di risposte (*nitem* * *nsogg*)

dove

nitem numero di item

nsogg numero di soggetti.

Il valore minimo accettabile del coefficiente di riproducibilità è considerato 0.90; tale valore indica che l'errore osservato nella riproduzione non supera il 10% delle risposte totali e consente

- o di interpretare il punteggio totale e
- o di considerare scalabili gli item e rappresentabili in uno *scaling* cumulativo,
- o di identificare un'unica dimensione sottostante.

Appare chiaro come il valore di CR dipenda dal metodo con cui si conteggiano gli errori. E' quindi importante, nel presentare i risultati, indicare il metodo utilizzato specificando CR_{error} per il metodo *Guttman* e CR_{ge} per il metodo *GoodEnough-Edwards*.

Esiste la possibilità di stimare l'*errore standard di riproducibilità*, ovvero la generalizzabilità (significatività statistica) della scala:

$$SE_{cr} = \sqrt{(1 - CR)\frac{CR}{N}}$$

❖ *PERCENT IMPROVEMENT* ($INP\%$)

L'item cui è associato il valore di CR_i più basso rappresenta il punto più debole del gruppo di item. In tal senso è possibile calcolare un indice del miglioramento che si può ottenere eliminando tale item (*percent improvement*):

$$INP\% = CR - \min(CR_i)$$

❖ *MINIMA RIPRODUCIBILITA' MARGINALE* (MMR)

Siccome esiste la possibilità che il coefficiente CR venga influenzato da distribuzioni marginali estreme, è consigliabile confrontare tale valore con la *Minima Riproducibilità Marginale (Minimal Marginal Reproducibility, MMR)*, equivalente alla *sommatoria delle sole frequenze modali di ciascun item*, ovvero il valore minimo di CR tra due item contigui per item dicotomici[2]; ciò richiama l'attenzione sul fatto che la riproducibilità di un item non può essere minore della proporzione delle risposte nella sua categoria modale; analogamente la riproducibilità totale non può essere minore della somma delle proporzioni delle risposte nella categoria modale per ciascun item, diviso per il numero di item.

Il calcolo di *MMR* può avvenire nel seguente modo:

$$MMR = \frac{\sum nm_i}{N}$$

dove

nm_i numero di risposte nella categoria modale (la più *scelta*) dell'item *i*

N numero di risposte (*nitem* * *nsogg*).

Il valore di tale coefficiente riflette la riproducibilità di una serie di item basata solo sulla conoscenza delle distribuzioni marginali degli item[3].

---

[2] Edwards A.L., "Modal Categories and Minimal Marginal Reproducibility", pp. 191-198, *Techniques of Attitude Scale Construction*, Appleton, Century-Crofts, Inc., New York, 1957 e *package* BMD.

[3] Di seguito vediamo alcuni esempi in cui si considerano i marginali di risposte positive per quattro item. Prendiamo il caso in cui le probabilità marginali dei quattro item siano .8, .6, .4 e .2 (quinta riga della tabella riportata di seguito) il valore della minima riproducibilità marginale sarà uguale a

La riproducibilità totale non può essere minore di MMR, ovvero la differenza tra CR e MMR deve essere di tale grandezza da poter attribuire un miglioramento nella previsione dei modelli di risposta allo *scalogram analysis*. Sappiamo che CR e MMR sono nella seguente relazione:

$$CR = \frac{(N - se)}{N} \qquad MMR = \frac{N - me}{N}$$

dove
N numero di risposte (*nitem* * *nsogg*)
se errori di *scaling*
me errori marginali, somma di tutte le frequenze non-modali[4].

I due coefficienti sono nella seguente relazione:

$$CR - MMR = \frac{N - se}{N} - \frac{N - me}{N} \qquad CR - MMR = \frac{me - se}{N}$$

Quindi la differenza tra i due coefficienti è funzione del miglioramento nella previsione fornito dall'insieme di item rispetto alle frequenze marginali degli item individuali.
Il valore di tale differenza va da 0 (nessun miglioramento nella previsione) a me/N (i dati si adattano in maniera perfetta allo *scaling Guttman*). Poiché l'errore massimo marginale che può verificarsi è 50%, tale differenza ha un massimo teorico di .50. Interpretare la differenza tra CR e MMR su una scala da 0 a me/N può essere difficile; per questo motivo sono stati suggeriti altri indici alternativi.

- COEFFICIENTE DI SCALABILITA' (CS)

Per meglio interpretare i precedenti coefficienti è importante applicare anche il *coefficiente di scalabilità* (CS) che consente di misurare la *capacità di un insieme di item di prevedere le risposte rispetto alle previsioni basate sulle frequenze marginali*:

$$CS = 1 - \frac{\sum n_{ie}}{me}$$

dove
n_{ie} numero di errori dell'item i
me errori marginali, somma di tutte le frequenze non-modali.

Tale coefficiente può essere espresso anche nel modo seguente:

$$CS = \frac{INP\%}{1 - \min(CR_i)}$$

Esso può assumere valori tra 0 e 1. Se le previsioni sono perfette, ovvero non vi sono errori di *scaling*, il valore di CS è 1. Se il gruppo di item non fornisce alcun miglioramento nella previsione (gli errori di *scaling* sono uguali agli errori marginali) il valore è 0. L'insieme di item presenta comunque una buona scalabilità se registra un valore di CS di almeno .60.

- COEFFICIENTE DI PREDICIBILITA' (CP_j)

Per ciascun soggetto è possibile calcolare un *coefficiente di predicibilità* (CP_j) basato sulle

---

$$MMR = (.8 + .6 + .6 + .8) / 4 = .7$$

Osserviamo inoltre come nel caso in cui i marginali osservati siano .95, .85, .15 e .05, MMR risulta essere uguale a .90; da ciò si può concludere che il valore di MMR è funzione dei marginali estremi.

item	*1*	.50	.60	.70	.70	.80	.80	.90	.85	.90	.95	.95
	2	.50	.60	.50	.70	.60	.80	.70	.85	.90	.85	.95
	3	.50	.40	.50	.30	.40	.20	.30	.15	.10	.15	.04
	4	.50	.40	.30	.30	.20	.20	.10	.15	.10	.05	.05
MMR		.50	.60	.60	.70	.70	.80	.80	.85	.90	.90	.95

[4] La formula generica per calcolare l'errore marginale dell'item i secondo l'ipotesi casuale è

$$me_i = c(n - nm)$$

dove
c probabilità di avere una risposta giusta per caso (con item dicotomici = .50)
n numero di risposte
nm numero di risposte nella categoria modale (la più *scelta*) dell'item.

previsioni realizzate e le previsioni possibili:

$$CP_j = \sum \frac{pr_j}{pp_i}$$

dove
pr previsioni realizzate
pp previsioni possibili.

❖ VERIFICA DELL'UNIDIMENSIONALITA'

Come sappiamo per poter adottare un item all'interno di un certo modello occorrono delle ragioni teoriche e delle ipotesi iniziali; in altre parole l'appartenenza degli item ad un unico ambito di contenuto è condizione essenziale per poter assumere un modello triangolare. Come abbiamo visto la verifica dell'unidimensionalità è fatta essenzialmente osservando il livello di riproducibilità, il livello di scalabilità e la casualità degli errori di risposta (la frequente osservazione di un particolare modello di errore fa ipotizzare più dimensioni sottostanti il gruppo di item[5]). L'osservazione di un soddisfacente livello di scalabilità tra item non consente di verificare l'unidimensionalità: infatti un item può risultare scalabile ma non avere alcun contenuto in comune con gli altri item e con la dimensione che si vuole misurare; esistono dei casi in cui item pur molto distanti e diversi tra loro anche nel contenuto consentono comunque di soddisfare il modello triangolare, come nel seguente esempio con quattro item:

a. *risolvi rispetto a* x *la seguente equazione:* $x^2+2x+9=16$
b. *qual è il significato della parola* severo*?*
c. *quanto fa "10*38"*
d. *quando usi l'ombrello?*

E' possibile che tali item somministrati a ragazzi tra i 10 e i 16 anni possano riprodurre in modo eccellente il modello triangolare dello *scalogramma*, in quanto qualsiasi soggetto che esegue il primo item correttamente probabilmente esegue correttamente anche gli altri, pur non appartenendo e non misurando lo stesso attributo.[6]

Vediamo schematicamente i diversi momenti di verifica del modello:

---

[5] In alcuni casi l'osservazione di errori sistematici può risultare utile all'identificazione di particolari gruppi di soggetti; nel caso per esempio di scale per la misurazione dell'autosufficienza fisica negli anziani è possibile individuare soggetti portatori di disabilità legate ad aspetti specifici, e quindi sul piano epidemiologico meno frequenti o imputabili ad aspetti non connessi con i processi di invecchiamento, che richiedono interventi individuali e particolari (ad esempio, una mutilazione ma anche fattori ambientali, barriere architettoniche ecc.).

[6] Secondo alcuni autori l'unidimensionalità di una scala Guttman può essere verificata indirettamente tramite il coefficiente di dipendenza di Yule: prendiamo in considerazione due item dicotomici *i* e *j* e costruiamo la loro distribuzione incrociata:

		i		
		si	no	
j	si	a	b	*(a+b)*
	no	c	d	*(c+d)*
		(a+c)	*(b+d)*	*n=(a+b+c+d)*

se *i* esprime un intensità maggiore rispetto a *j*, la frequenza *c* tenderà a 0 (compatibilmente con la presenza di errori casuali); il coefficiente *Q* di Yule consente di verificare l'ipotesi che *c* sia significativamente uguale a 0:

$$Q = \frac{(ad - bc)}{(ad + bc)}$$

Q, che non è altro che il coefficiente *gamma* (Maggino, 2005) nel caso particolare di dati dicotomici, assume valore 0 quando le variabili sono indipendenti mentre raggiunge l'unità tutte le volte che una qualsiasi delle quattro celle si presenta vuota. Secondo gli autori un valore di $Q \geq 0.80$ dovrebbe essere considerato soddisfacente.

VERIFICA DEL MODELLO DETERMINISTICO

⇓

A. ANALISI DELLA RIPRODUCIBILITA'

1. SELEZIONE DEGLI ITEM

a. *Coefficiente di riproducibilità* per ciascun item CR_i

$$CR_i = 1 - \frac{n_{ie}}{n}$$

bontà di adattamento $\Rightarrow CR_i \geq 0.85$

↓

b. *Percent Improvement INP%*

$$INP\% = CR - \min(CR_i)$$

Indice del miglioramento che si può ottenere eliminando l'item con il valore CR_i più basso

↓

c. Eliminazione degli item con $CR_i > 0.85$

⇓

2. RIPRODUCIBILITA' DELLA SCALA

a. *Coefficiente di riproducibilità CR*

$$CR = 1 - \frac{\sum n_{ie}}{N}$$

proporzione di risposte agli item che possono essere correttamente riprodotte conoscendo il punteggio totale del soggetto

↓

misura della bontà di adattamento tra
- il modello di risposte osservato
- il modello di risposte ideale o previsto

$CR > 0.90 \rightarrow$
- la scala riflette il 10% di errori
- consente di interpretare la scala
- consente di considerare scalabili gli item

b. *Minima Riproducibilità Marginale (Marginal Reproducibility, MMR)*

$$MMR = \frac{\sum nm_i}{N}$$

Sommatoria delle sole frequenze modali di ciascun item.
La riproducibilità totale non può essere minore di *MMR*

B. SCALABILITA'

1. Coefficiente di scalabilità (CS)

$$CS = 1 - \frac{\sum n_{ie}}{me} \quad \text{oppure} \quad CS = \frac{INP\%}{1 - \min(CR_i)}$$

Capacità di una scala di prevedere le risposte rispetto alle previsioni basate sulle frequenze marginali.

$CS \geq 0.60 \rightarrow$ scala con buona scalabilità
$CS = 1 \rightarrow$ previsioni perfette $\rightarrow$ non ci sono errori di scala
$CS = 0 \rightarrow$ scala non fornisce alcun miglioramento nella previsione

C. ANALISI DELLE PREDICIBILITA'

1. Coefficiente di predicibilità per ciascun soggetto (CP_j)

$$CP_j = \sum \frac{pr_i}{pp_i}$$

dove

n_{ie}	numero totale di errori perl'item i = errori di riproducibilità
n	numero di risposte all'item (o numero di soggetti)
N	numero di risposte (*nitem*nsogg*)
$nitem$	numero di item
$nsogg$	numero di soggetti
nm	numero di risposte nella categoria modale dell'item
me	errori marginali, somma di tutte le frequenze non-modali
pr	previsioni realizzate
pp	previsioni possibili

L'adattamento è considerato soddisfacente quando sono soddisfatti almeno tre requisiti:

a. $CR \geq .90$;
b. valore di *MMR* non eccessivamente alto;
c. differenza tra *CR* e *MMR* che indichi che si sia realizzato un miglioramento nella scalabilità in funzione della conoscenza dei punteggi totali;
d. errori non sistematici nei modelli di risposta.

3.3 I LIMITI DEL MODELLO DETERMINISTICO

Il più serio limite dello *scaling Guttman* è dato proprio dall'approccio deterministico. Tale modello non fornisce alcuna spiegazione delle deviazioni dai modelli di risposte perfettamente scalabili, ovvero non presenta alcuna teoria riguardo agli errori di *scaling*. Questi si verificano quando un soggetto dà una risposta negativa ad un item il cui punto è posizionato a sinistra del punto-soggetto o quando un soggetto dà una risposta positiva ad un item il cui punto è posizionato a destra del punto-soggetto. In entrambi i casi le risposte contraddicono il modello, data la geometria della soluzione di *scaling*. Nello standard dell'analisi dello *scaling Guttman* gli errori rimangono inspiegati: a partire dalla loro conoscenza si cerca solo, come vedremo, di sviluppare metodi che consentano di assegnare punteggi di scala alle osservazioni che presentano tali errori. Essi rimangono però un impedimento alla spiegazione della variabilità dei dati (McIver, 1979).

3.3.1 *Problemi di applicazione dello scalogramma*

3.3.1.1 Calcolo del punteggio individuale

Considerando item a risposte dicotomiche, il calcolo dei punteggi individuali è effettuato sommando il numero di risposte positive; in questo senso è identico al corrispondente metodo utilizzato per il modello additivo (scale *Likert*). Relativamente al punteggio totale, tra i due modelli esistono però delle differenze:

- modello additivo (scala *Likert*):
  - *calcolo punteggi*: quando l'affidabilità dell'insieme degli item è elevata;
  - *interpretazione dei punteggi*: un punteggio di 2 indica che il soggetto ha risposto in modo positivo a 2 tra gli item.
- modello cumulativo (scala *Guttman*):
  - *calcolo punteggi*: quando il valore del CR è maggiore di .90;
  - *interpretazione dei punteggi*: un punteggio di 2 indica che il soggetto ha risposto in modo positivo a 2 specifici item, quelli più *facili* o più *accettabili*.

3.3.1.2 Calcolo del punteggio con item a risposte multiple

Sicuramente il modello deterministico-cumulativo è più semplice da comprendere e da applicare nel caso in cui si utilizzino item dicotomici (Flament, 1976). Per questo la maggior parte dei modelli di scalogramma sono costruiti con item a risposte dicotomiche o riconducibili a valutazioni del tipo *successo|fallimento* nella prova. D'altra parte esistono, come già Guttman aveva indicato, applicazioni delle tecniche di *scalogram analysis* che utilizzano valutazioni graduate delle attitudini; si tratta di item a risposte multiple (ovvero un numero di categorie maggiore di due) con valutazione decrescente (Tesi, 1993). Le risposte multiple *graduate* permettono una maggiore finezza nella rilevazione delle posizioni individuali, ma rendono più complesso il calcolo degli errori di scalabilità (per ogni soggetto e ogni item).
E' però difficile trovare item con molte categorie tra loro distinguibili in modo chiaro e non ambiguo per tutti; per tale motivo spesso si consiglia di combinare tra loro le categorie che non si presentano chiaramente distinte. Sono state per questo stabilite diverse strategie di combinazione delle categorie; tra queste ricordiamo il metodo delle approssimazioni successive; tale criterio è basato sull'osservazione delle *abitudini verbali* dei soggetti che condividono posizioni simili pur indicando categorie diverse. In tale ottica si possono combinare tra loro le categorie considerate *neutrali*, quelle considerate *positive*, quelle considerate *negative*.
Come si può immaginare la validità di tale metodo è stata messa in discussione a causa del suo carattere empirico e soggettivo. Con altre tecniche, le risposte vengono dicotomizzate con una soglia arbitraria, sotto la quale le risposte sono valutate negative, sopra positive. Quando non è possibile una chiara dicotomizzazione per tutti gli item o si ritiene sia più corretto prevedere un sistema di punteggi differenziato per ogni item, è necessario porre particolare attenzione al peso da attribuire alle singole risposte per ciascun item. Quando la scala delle risposte è di questo tipo il coefficiente MMR assume una particolare importanza per la validazione dello scalogramma.

3.3.1.3 Assegnazione dei punteggi nel caso di errori

Analizzata e verificata la rispondenza della scala al modello cumulativo, il passaggio successivo è

quello dell'assegnazione dei punteggi ai soggetti. Nel caso in cui la scala utilizzata è perfetta, l'assegnazione dei punteggi è molto semplice e chiara: il punteggio di ciascun soggetto sarà uguale al numero di risposte positive. Il problema sorge quando le risposte di un soggetto non sono scalate ovvero quando il modello di risposta non riflette un modello ideale.
Uno dei problemi che si pongono nell'analizzare tali tipi di scale riguarda infatti l'opportunità di considerare anche i soggetti il cui modello di risposta devia da quello ideale e quindi di calcolare il punteggio totale. Non è possibile dare una soluzione unica a tale problema; il ricercatore può sostenere che l'individuo che non rispetta l'ordine degli elementi è in errore solo se il modello teorico sostiene un perfetto schema ordinale degli item. E' necessario essere consapevoli del fatto che la decisione di accettare o respingere la scala è comunque arbitraria e che il ricercatore dovrà prima o poi risolvere il problema dell'*errore*. La violazione del modello triangolare di risposta può suggerire che

a. uno degli assunti sottostanti il modello *Guttman* non è valido rispetto ai dati;
b. lo spazio in cui gli stimoli e i soggetti sono posizionati non è unidimensionale[7].

La raccomandazione è quindi quella di adottare una scala Guttman solo se il numero di errori è abbastanza piccolo (alto valore del coefficiente di riproducibilità). In questi casi occorre comunque prestare molta attenzione a come si calcola il punteggio totale.
Assumiamo che, su uno scalogramma con 5 item, sia stato osservato il seguente modello di risposta -++--. La prima osservazione che si può fare è che non esiste alcun numero da 0 a 5 che consenta di definire in modo accurato il soggetto sul continuum sottostante. D'altra parte è necessario affrontare il problema di quale punteggio assegnare in modo che l'interpretazione della scala ne sia influenzata il meno possibile.
Come abbiamo visto in questi casi si procede all'assegnazione dell'errore rispetto alla quale sono state sviluppate due forme:

1. assegnazione basata sui modelli di risposta ideali più vicini (*tecnica Cornell*),
2. assegnazione basata sul numero di risposte positive (*tecnica GoodEnough-Edwards*).

Se applichiamo il primo criterio, il modello osservato viene riclassificato nel modello più vicino +++-- cui viene assegnato il punteggio di 3.
Se applichiamo il secondo criterio a tale *pattern* di risposta, nel quale sono presenti solamente due "+", viene assegnato il punteggio 2.
E' evidente quindi come la possibilità di gestione diversa dell'errore introduce una seria limitazione nella interpretazione dei risultati della scala per i soggetti che presentano risposte non scalari.
Poniamo di avere osservato un altro modello: -+-+-.
Se applichiamo il primo criterio tale modello di risposta può essere riferito a diversi modelli ideali: ++++-, ++--- oppure -----; conseguentemente il punteggio che può essere assegnato è 4 oppure 2 oppure 0; quindi l'applicazione del primo criterio induce una certa arbitrarietà di assegnazione sia dell'errore che del punteggio individuale.
L'applicazione del secondo criterio conduce comunque all'assegnazione del punteggio 2.
E' comunque vero però che il problema dell'assegnazione è uno pseudo-problema dal momento che per poter procedere al calcolo del punteggio individuale è necessario che la scala risponda al modello cumulativo ovvero deve registrare un elevato valore di riproducibilità ($CR \geq .90$). E' comunque consigliabile utilizzare sempre il secondo criterio ovvero quello che registra la deviazione dalla perfetta riproducibilità.

3.3.1.4 Calcolo degli errori con item a risposta multipla

L'adozione di item a risposta multipla pone un altro problema. Usando i modelli di validazione per item dicotomici, ogni singolo errore di predicibilità è valutato 1, qualunque sia l'entità della

[7] Una ulteriore validazione di queste scale è infatti quella che riguarda la verifica dell'omogeneità degli item (tutte le domande devono interessare o misurare una sola capacità); a tal fine è possibile utilizzare l'analisi fattoriale.

differenza fra i punteggi di due item contigui. Ma con una valutazione a più punti il significato della differenza può essere più o meno rilevante; è quindi necessario sviluppare indicatori in grado di pesare gli errori di predizione (un punteggio più basso in un item più difficile) sulla base della distanza tra le valutazioni contigue. In pratica, l'entità di questa differenza è ricuperabile calcolando per ogni item il numero di errori commessi e la sommatoria dell'entità degli scarti.
E' per questo che accanto all'indice classico (CR_i) è stato sviluppato un altro coefficiente di scalabilità (CR_{iw}) (Tesi G. e al., 1993) che viene ricavato

- dalla proporzione di soggetti che non hanno superato un item ma hanno superato un item successivo più difficile diviso il totale dei soggetti che hanno superato la prova i+1 (n_e),
- dalla sommatoria della *distanza* fra le risposte negli stessi soggetti con errori di scalabilità (n_s).

Sulla base di queste due valutazioni è possibile calcolare gli indici di riproducibilità, rapportando la sommatoria degli scarti al valore massimo della sommatoria:

$$CR_i = 1 - \frac{n_e}{n} \qquad\qquad CR_{iw} = 1 - \frac{n_s}{n * p_{mm}}$$

n_e errori di riproducibilità tra l'item i e l'item $i+1$ (ovvero numero totale di errori per l'item i)
n_s scarti negativi tra l'item i e item $i+1$
n numero risposte (o numero di soggetti)
p_{mm} punteggio_*max* - punteggio_*min*

Analogamente il *coefficiente globale di riproducibilità* dello scalogramma, che esprime quanto l'insieme di item soddisfi la condizione *cumulativa* delle attitudini richieste per svolgere il compito preso in esame, è calcolato a partire dal totale degli errori in tutti gli item:

$$CR_w = 1 - \frac{\sum n_s}{N * p_{mm}}$$

n_s scarti negativi tra item contigui (i e $i+1$)
N numero di risposte (*nitem* * *nsogg*)
p_{mm} punteggio_*max* - punteggio_*min*

3.3.1.5 Dimensione della matrice

Nello *scalogram analysis* la presenza di troppi dati (item e soggetti) rappresenta un problema di complessa gestione. La complessità di analisi e di controllo consigliano in genere di adottare scale Guttman composte da pochi item anche se ciò potrebbe non consentire una corretta discriminazione tra i soggetti. D'altra parte la disponibilità di molti dati è auspicabile nei momenti di messa a punto della scala. Per ridurre la matrice analizzata ad una dimensione gestibile, sono stati messi a punto diversi metodi; tali metodi possono rappresentare dei validi sostegni nello sviluppo di una scala.

- Tecnica *H* (o degli *item inventati*)
  Se gli item da scalare fanno tutti parte dello stesso teorico universo di contenuto, ma nello stesso momento non sono importanti individualmente, è possibile creare *item inventati* combinando le risposte in due o più item. Gli item combinati sono quelli che presentano alla luce dei risultati la stessa difficoltà ovvero occupano la stessa posizione sul continuum di misurazione. Ciascuna risposta individuale ricondotta in un item è determinata sulla base delle risposte del soggetto a tutti gli item che costituiscono la nuova variabile inventata. L'utilizzazione di *item inventati* consente di eliminare risposte non scalari all'interno del gruppo di item combinati; utilizzando le risposte è possibile individuare la posizione individuale più probabile rispetto al gruppo degli item combinati. Tale tecnica si presta a diverse applicazioni, ma anche critiche, ed è stata utilizzata in origine per incrementare la riproducibilità. Può essere comunque utilizzata per consolidare i dati prima che l'analisi cominci.
- Tecnica dei *soggetti inventati*
  Si tratta di un metodo studiato per incrementare l'adattamento tra dati e modello cumulativo (riproducibilità) senza modificare, eliminare o combinare gli item. Nel caso in cui l'interesse è più orientato verso una analisi globale del fenomeno (attitudini del gruppo più che dei singoli individui) non è necessario conservare i modelli di risposta individuali. Conseguentemente è possibile creare un soggetto inventato allo stesso modo degli item inventati; ovvero è possibile combinare gli individui che presentano lo stesso punteggio totale in un unico soggetto ipotetico assumendo come punteggi per gli item le risposte predominanti dei componenti del gruppo.

- <u>Eliminazione di item con marginali estremi</u>
  Gli item con i marginali estremi non aumentano in modo significativo il numero di errori di scala e forniscono poche informazioni sulla loro reale appartenenza ad una data scala.
- <u>Eliminazione di item con marginali simili agli item presenti nella scala</u>
- <u>Campionamento di item</u>
- <u>Mantenimento dei soli item significativi</u>
- <u>Analisi di sottoinsiemi di contenuto</u>.

3.3.1.6 Dati missing

Un altro importante problema che l'analisi dello scalogramma deve affrontare è rappresentato dal modo di trattare i dati *missing*. Tale problema è comunque comune alla maggior parte delle ricerche empiriche ed è difficilmente evitabile (è difficile ricontattare i soggetti per completare le informazioni mancanti). Per tale motivo molti ricercatori sono spesso obbligati a sviluppare qualche procedura statistica per evitare la perdita di dati in parte già raccolti. Tali procedure riguardano i particolari criteri di analisi delle matrici che presentano dati *missing*. Alcuni criteri prendono in considerazione i modelli di risposta del soggetto a tutti gli item della scala:

a. se il soggetto non ha risposto a più del 50% degli item, il soggetto risulterà *non classificato*;
b. se il soggetto non ha risposto a meno del 50% degli item, alle risposte *missing* si attribuisce un valore corrispondente al punteggio medio della totalità dei soggetti.

Esistono altri complessi schemi per prevedere e attribuire un punteggio al valore *missing* a partire dalle risposte osservate o da altre informazioni. In alcuni casi la complessità e l'arbitrarietà di tali schemi annullano qualsiasi beneficio derivato dal loro utilizzo.

3.3.1.7 Significatività statistica

A tale modello di *scaling* vengono mosse altre critiche riguardanti in particolare la sua *significatività statistica*; secondo tali critiche le scale Guttman presentano le seguenti caratteristiche:

- *<u>Instabilità</u>: tali scale, soprattutto se composte da pochi item, sono soggette a variazioni ed errori casuali*. In realtà la scala Guttman, anche se composta da pochi item, conserva tutti gli elementi per il controllo della consistenza interna attraverso gli aspetti della predicibilità.
- *<u>Poca significatività</u>: lo scalogramma rappresenta un modello poco generalizzabile.* In effetti la verifica della scalabilità di un gruppo di item su un particolare campione non consente di affermare che applicando la stessa su un altro campione si ottengano gli stessi risultati complessivi. Si comprende quindi l'importanza della rappresentatività del campione su cui viene effettuata la validazione della scala.
  Anche se, come abbiamo visto, sono stati fissati alcuni criteri di accettabilità della scala *Guttman* $(CR > 0.9, MMR < 0.9, CS > 0.6)$ è possibile che la loro soddisfazione avvenga per caso e che quindi non fornisca una prova definitiva della aderenza degli item al modello cumulativo. A tale riguardo sono stati sviluppati diversi test di significatività come supporto aggiuntivo alla scalabilità degli item. Per verificare ciò è possibile effettuare controlli attraverso il *chi-quadro*, come confronto tra distribuzione reale e distribuzione teorica (casuale) e l'errore standard di riproducibilità (Guttman:
  $$ES_{rep} \cong \sqrt{(1-CR)*CR/N}$$

In conclusione va comunque ricordato come particolari modelli, nonostante le difficoltà applicative, spesso risultano importanti nella ricerca per lo sviluppo di teorie e dei relativi modelli matematici.

3.4 ALTRI MODELLI DETERMINISTICI

Il modello di scalogramma che abbiamo appena analizzato richiede che vengano soddisfatti diversi assunti riguardanti la caratteristica misurata (unidimensionale) e gli item (ordinabili lungo il continuum della caratteristica). E' comunque possibile ipotizzare scalogrammi più complessi, in cui non necessariamente si ipotizza una caratteristica unidimensionale e gli item sono ordinabili in modi diversi (Borg, 1995; Shye, 1985). A tale proposito è possibile identificare due diversi approcci:

- *Multidimensional Scalogram Analysis* (*MSA*), il cui obiettivo è quello di rappresentare i profili attraverso punti in uno spazio in modo tale che tale spazio possa essere suddiviso da ciascun item secondo le sue categorie di misurazione;
- *Partial Ordered Scalogram Analysis* (*POSA*), rappresenta i profili individuali in uno spazio geometrico, preservando le relazioni d'ordine tra essi.

In generale tali approcci risultano utili come:

a. strumenti per la definizione di ipotesi strutturali riguardanti le osservazioni;
b. procedure per valutare fenomeni la cui struttura empirica richiede più di un punteggio.

3.4.1 Modelli alternativi di scalogramma

Non sempre gli item utilizzati consentono di descrivere un perfetto scalogramma cumulativo; tra di essi è possibile osservare, all'interno del concetto di scalogramma relazioni diverse da quella osservata (cumulativa). Infatti la mancata osservazione di uno scalogramma perfetto non indica necessariamente una scorretta definizione di un modello o la sbagliata costruzione di uno strumento. In questi casi è forse possibile definire un modello di scalogramma che descriva delle nuove relazioni tra item o, meglio come vedremo, tra profili. Quando i profili individuali ottenuti con le osservazioni empiriche risultano mutuamente non confrontabili può essere comunque possibile identificare un'ipotesi di scala sottostante che ci consenta di sintetizzare in un punteggio interpretabile le informazioni degli item. Come vedremo il punteggio può non essere unico.

Secondo un approccio è necessario perseguire due obiettivi:

a. individuazione per ciascun caso di più profili e assegnazione a ciascuno di essi di un punteggio;
b. analisi delle variabili sottostanti (diversamente dallo scalogramma Guttman che ne presenta solo una).

Osserviamo i seguenti profili individuali ottenuti con quattro item (il codice '1' indica "fallimento" e il codice '2' indica "superamento"):

		item				*n. di performance positive*
		I_1	I_2	I_3	I_4	
casi	1	1	1	1	1	0
	2	1	1	2	1	1
	3	1	1	1	2	1
	4	2	1	1	1	1
	5	1	2	1	1	1
	6	1	1	2	2	2
	7	2	1	1	2	2
	8	2	2	1	1	2
	9	2	1	2	2	3
	10	2	2	1	2	3
	11	2	2	2	2	4
n. di performance positive		*6*	*4*	*4*	*6*	

Come si può facilmente osservare l'ipotesi cumulativa dello scalogramma Guttman qui non è sostenibile. Infatti pur avendo ordinato i punteggi individuali, questi non consentono di *riprodurre* le *performance* individuali relative ai singoli item.
Se però si ritiene che i profili ottenuti siano comunque legittimi è necessario operare in modo diverso[8]. Con il *diagramma Hasse* (detto anche *partial order diagram*) è possibile rappresentare i profili osservati; in esso ciascun profilo è scritto solamente una volta. Vediamo la rappresentazione degli undici profili presentati (in questa rappresentazione è rispettata la sequenza degli item: 1, 2, 3, 4):

Sequenza degli item 1234							Numero di performance positive
			1111				0
1121		1112		2111		1211	1
	1122		2112		2211		2
		2122		2212			3
			2222				4

Nel tracciare tale diagramma può essere utile mettere sullo stesso livello orizzontale i profili che presentano lo stesso numero di *performance* positive. Tale diagramma, ottenuto per prove ed errori, descrive le relazioni di confrontabilità tra i profili e soddisfa le condizioni del diagramma di ordine parziale: solo i profili confrontabili sono collegati da linee discendenti; notare inoltre che non presenta linee intrecciate. Un gruppo di profili che risponde a tali caratteristiche è almeno bidimensionale.

3.4.1.1 Modello diamante

Riordinando gli item rispetto ai totali di colonna visti nella loro presentazione (la nuova sequenza è I_3, I_4, I_1 e I_2) è possibile osservare una certa interessante regolarità; il nuovo ordinamento ha tenuto conto di una nuova caratteristica osservabile in senso verticale: la *media dei ranghi delle posizioni dei successi* (*center of success*):

Sequenza degli item 3412								Numero di performance positive
			1111					0
2111		1211		1121		1112		1
	2211		1221		1122			2
		2221		1222				3
			2222					4
1	1.5	2	2.5	3	3.5	4	⇐	Media dei ranghi delle posizioni dei successi

Per riprodurre ciascun profilo è necessario sia il punteggio di riga che quello di colonna; per esempio i punteggi 2 (di riga) e 3.5 (di colonna) consente di identificare solo il profilo 1122.

[8] Un modo per esempio può essere quello di identificare e isolare all'interno del gruppo di profili più sottogruppi che descrivono uno scalogramma Guttman.

Tale configurazione è detta "configurazione diamante" o "diagramma diamante", la cui logica può essere estesa ad un numero superiore di variabili.
La generalizzazione può riguardare anche il numero di possibili categorie (1, 2, 3, ..); in questo caso è necessario fare specificazioni aggiuntive riguardanti i pesi relativi da attribuire alle categorie.
L'ipotesi *diamante*, quando confermata dai dati, presenta potenti implicazioni teoriche e pratiche. Essa determina:

- la struttura dei concetti che sono misurati,
- gli strumenti di misurazione necessari per valutare completamente i casi rispetto a tali concetti.

Tale scalogramma presenta aspetti radicalmente diversi da quelli molto rigidi dello scalogramma Guttman; quest'ultimo può essere considerato un caso particolare della configurazione *diamante*.
Quando la stessa configurazione di profili ricorre in più applicazioni empiriche, aumenta la fiducia negli aspetti strutturali manifestati dai dati.

La logica interna della struttura e dei processi

Come abbiamo visto, tra i pochi concetti applicabili allo *scaling* Guttman vi è quello di "difficoltà"; nello scalogramma tipo *diamante* abbiamo notato come nella configurazione è possibile riscontrare un ordine tra gli item in modo tale che ogni profilo presenti una singola sequenza di *performance* positive (ovvero serie di *performance* positive non interrotte).
Un altro assunto che consente di stabilire il criterio di appartenenza di un profilo ad una configurazione *diamante* è il seguente: dato un particolare ordine tra gli item, se un caso ha superato due item deve aver superato anche tutti quelli posti tra i due in questione.
A partire da tale assunto diviene evidente un'interessante caratteristica della configurazione *diamante*: ogni profilo presenta un *primo* item superato e un *ultimo* item superato. Nel profilo "112221" il primo item superato è il terzo, mentre l'ultimo è il quinto.
E' possibile a questo punto definire una configurazione diamante sulla base di due punteggi:

a. posizione nel profilo del *primo* item superato,
b. posizione nel profilo dell'*ultimo* item superato.

Osserviamo il seguente diagramma in cui è possibile realizzare contemporaneamente due forme di *scaling*.

1° punteggio di scala: posizione del primo item superato	6	211111	221111	222111	222211	222221	222222
	5		121111	122111	122211	222221	122222
	4			112111	112211	112221	112222
	3				111211	111221	111222
	2					111121	111122
	1	111111					111112
		1	2	3	4	5	6
		2° punteggio di scala: posizione dell'ultimo item superato					

Notare che nel 1° punteggio di scala, il punteggio più alto viene attribuito a quei profili in cui, lungo la sequenza di item determinata dalla configurazione *diamante*, il primo "successo" avviene *presto*. Analogamente nel 2° punteggio di scala, il punteggio più alto viene attribuito a quei profili in cui l'ultimo "successo" avviene *tardi*. Si fa questo in modo da poter associare punteggi alti in entrambe le scale con la situazione in cui il caso "supera" molti item. Quindi il significato generale delle due scale rimane lo stesso ("successo" o "tendenza al successo") mentre ciascuna di esse mantiene il proprio specifico e ben definito significato ("inizio" e "fine" dei "successi").
Utilizzando la rappresentazione cartesiana (anche in termini ordinali più che quantitativi) ha consentito di posizione (o riprodurre) i profili osservati; tali scale presentano una simmetria

semantica nei loro significati, cosa che non era stata riscontrata con le due precedenti scale (*numero di performance positive* e *media delle posizioni dei successi*). Tale caratteristica (condivisione di un generale significato comune) estende la nozione di *scaling* a complesse configurazioni di profili, compreso quelle di alta dimensionalità.

Scaling che preserva l'ordine

Come abbiamo visto nel caso della configurazione *diamante*, studiando la struttura dei profili è stato possibile assegnare due punteggi ad ogni profilo: il primo item superato - l'ultimo item superato; tali punteggi sono determinati con riferimento ad uno specifico ordine tra gli item. Dato un insieme di profili che risponde al criterio *diamante*, l'assegnazione dei due punteggi consentono di riprodurre un unico profilo.
Un modo per estendere tale metodo di *scoring* a altre configurazioni più complesse è quello di aderire alla nozione di ordine tra gli item e di classificare i modelli di successo che possono apparire relativi ad un determinato ordinamento fisso. Un ulteriore elemento di complessità potrebbe essere quello di considerare i profili che contengono due sequenze di "successi"; tale situazione si presenta tridimensionale in quanto per poter riprodurre ciascun profilo sono necessari tre punteggi di scala.
Non sempre però i ricercatori hanno una chiara idea della specifica struttura dei concetti utilizzati o dei meccanismi dei processi che studiano in modo da poter formulare un'ipotesi in termini di modelli attesi e di profili legittimi.
Prendiamo in considerazione i punteggi assegnati ai profili nel diagramma precedente. Risulta abbastanza chiaro come rispetto al primo punteggio, i profili che condividono lo stesso punteggio definiscono una scala Guttman. Tale *scalabilità* può essere chiamata scalabilità condizionale: il condizionamento riguarda i valori del 1° punteggio. In questo caso la scalabilità è indicata dai valori del 2° punteggio (ricordiamo che in questo caso i profili selezionati condividono lo stesso punteggio rispetto al primo). E' possibile procedere anche in senso inverso.
Quando due profili sono confrontabili (uno è maggiore dell'altro) anche i rispettivi punteggi di scala manterranno la stessa relazione, come si può osservare dalla seguente tabella:

		punteggi	
profili	111111	1	1
	111211	3	4
	112211	4	4
	112221	4	5
	122221	5	5
	222221	6	5
	222222	6	6
		⇓	⇓
		Punteggi di tipo Guttman	

A questo punto è possibile definire la proprietà del diagramma: *il diagramma converte ogni profilo (composto da* n *item) in due punteggi di scala in modo tale che ogni coppia di profili confrontabili diviene una coppia di due punteggi confrontabili che conservano tra loro la stessa relazione d'ordine. Solamente profili confrontabili sono convertibili in punteggi di scala confrontabili*. Tale proprietà è associata ad un'altra: in studi ben disegnati *le scale identificate condividono lo stesso significato di un concetto esteso del quale rappresentano aspetti diversi*.
L'obiettivo della rappresentazione di una configurazione *diamante* è non solo quello di economizzare l'informazione (obiettivo di *scaling*) ma anche quello di consentire l'identificazione di fattori sottostanti i concetti studiati. Un possibile criterio per definire molte scale può essere il seguente: dato un insieme di profili osservati, si potrebbe cercare di identificare uno spazio di coordinate cartesiane, con il numero più piccolo di assi, che consenta di rappresentare tutte le

possibili relazioni tra i profili, compresa la "non-confrontabilità".
Il compito di identificare scale che preservino l'ordine può essere complesso e arduo anche nel caso bidimensionale. Le procedure automatiche *POSAC* (*Partially Ordered Scalogram Analysis with Coordinates*) consentono di affrontare tale compito in modo più agevole (Shye, 1985).
Vedremo ora come sia possibile ottenere due punteggi di scala (valori di coordinate) per profili che sono strutturati da un meccanismo interno specifico.

3.4.1.2 Modello action system

Abbiamo visto che:

- lo *scaling* "Guttman" implica il concetto di ordine tra gli item e tra le categorie di risposta (per esempio in termini di "difficoltà");
- lo *scaling* "diamante" è sostenuto dal concetto di ordine (per esempio in termini di tempo o in termini di passaggi all'interno di un procedimento).

E' possibile però osservare un altro tipo di *scaling* ordinale a partire da altri tipi di configurazioni, i cui profili manifestano altri tipi meccanismi o di logiche interne. Tra i diversi tipi ne è stato identificato uno in grado di riflettere alcune situazioni reali piuttosto complesse. Come vedremo, in questo caso, pur aumentando la complessità strutturale, non aumenta necessariamente la dimensionalità dell'insieme dei profili, che rimane bidimensionale. Tale diversa configurazione è stata scoperta e formulata da Shye (1985) nell'ambito del lavoro sull'*action system* (noto anche come *living system* o *open system*) ed è per questo chiamata "configurazione *action system*". La caratteristica essenziale comune a tutte le definizioni di *action system* è quella seconda la quale in un tale tipo di sistema è possibile "agire" ed "essere agiti" (per esempio si può "dare" e "ricevere").
Sapendo ciò possiamo definire i seguenti item:

- *G*: indica se in un certo momento un determinato sistema <u>"dà"</u> (1) o meno (0);
- *R*: indica se in un certo momento un determinato sistema <u>"riceve"</u> (1) o meno (0).

E' possibile a questo punto identificare quattro stati sistemici:

G	R	il sistema
1	1	dà e riceve
1	0	dà e non riceve
0	1	non dà e riceve
0	0	non dà e non riceve

Seguendo la teoria dei sistemi, è necessario introdurre due nuovi item:

- *X*: indica se in un certo momento un determinato sistema vi è <u>interazione</u> (1) o meno (0) con l'ambiente esterno;
- *I*: indica se in un certo momento per un determinato sistema vi è <u>equilibrio interno</u> (1) o meno (0).

Vediamo a questo punto quali sono le relazioni esistenti tra i quattro concetti sistemici (detti anche modi sistemici).
1ª condizione: l'*interazione* può verificarsi sia con il *dare* che con il *ricevere* ed è sicuramente assente se sono assenti *dare* e *ricevere*;
2ª condizione: l'*equilibrio interno* può essere interpretato come una specie di equilibrio tra *dare* e *ricevere* e può essere valutato come "presente" se *dare* e *ricevere* sono presenti.

Se G=0 e R=0 allora X=0. Se G=1 e R=1 allora I=1.

Date queste premesse, è possibile identificare i profili assenti (ovvero non osservabili):

GRXI	
0010	esclusi dalla 1ª condizione
0011	
1100	esclusi dalla 2ª condizione
1110	

Quindi dei 16 profili teoricamente possibili, si assume che solo 12 sono empiricamente osservabili:

GRXI
0000
0001
1000
1001
1010
1011
0100
0101
0110
0111
1101
1111

A questo punto l'obiettivo è quello di definire uno spazio bidimensionale che preservi l'ordine. Cominciamo con il distinguere i profili con *G*=0 da quelli con *G*=1:

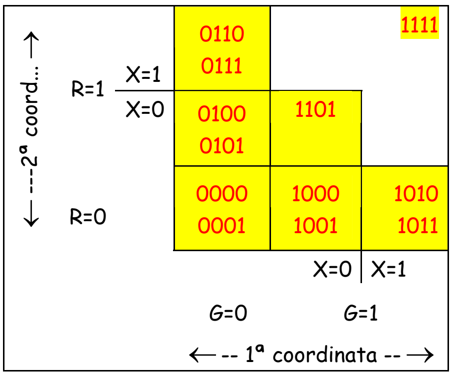
G=0	G=1
GRXI	GRXI
0000	1000
0001	1001
0100	1010
0101	1011
0110	1101
0111	1111

← 1ª coordinata →

A questo punto proviamo a distinguere i profili rispetto al valore *R* tracciando una retta orizzontale che divida ciascuna colonna in due.

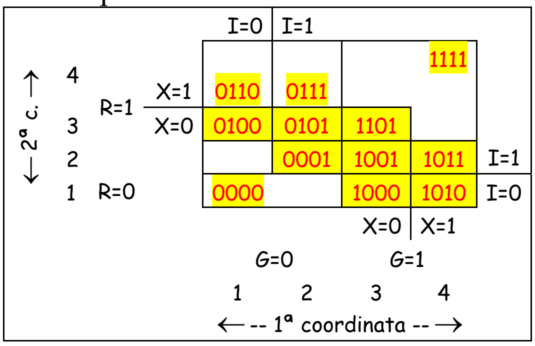
← 2ª coordinata →			
	R=1	0100 0101 0110 0111	1101 1111
	R=0	0000 0001	1000 1001 1010 1011
		G=0	G=1
		← 1ª coordinata →	

A questo punto in questo spazio bidimensionale è possibile separare i profili presenti secondo i valori di *X*. Per fare ciò è necessario individuare una linea di suddivisione appropriata che lasci tutti i profili con *X*=0 da una parte e i profili con *X*=1 dall'altra, senza alterare le classificazioni esistenti ottenute con i punteggi *G* e *R*.



Notare che la nuova linea lascia tutti i profili con *X*=0 al di sotto e a sinistra e i profili con *X*=1 al di sopra e a destra (verso i valori alti della due coordinate). La precedente suddivisione tra profili viene comunque mantenuta.

Dall'osservazione del diagramma risulta che *X*=1 è associato con valori estremi in almeno una coordinata. Per questo motivo un item che si comporta come *X* è detto *polarizzato* o *accentuato*.
A questo punto è possibile individuare una nuova linea che consenta di operare nello stesso modo ma con i valori *I*; tale linea dovrebbe lasciare tutti i profili con *I*=0 da una parte e quelli con *I*=1 dall'altra senza alterare le precedenti partizioni.



La nuova linea lascia al di sotto e a sinistra i profili con *I*=0 e al di sopra e a destra quelli con *I*=1. Notare che *I*=1 è associato con valori moderati in entrambe le coordinate. Per questa ragione un item che si comporta in questo modo è detto *moderante* o *attenuante*. Le coordinate sono ora ridefinite secondo 4 valori distinti (indicati con 1, 2, 3, 4). In questo modo a ciascun profilo vengono associati due nuovi punteggi, relativi alle due coordinate. Quindi a ciascun profilo vengono associati due punteggi. Un esame delle coppie di profili in tale diagramma consente di verificare se le relazioni d'ordine sono confermate. Per esempio il profilo 0111 (punteggio 24) è maggiore del profilo 0001 (punteggio 22) mentre il profilo 1010 (punteggio 41) e il profilo 0001 (punteggio 22) non sono tra loro confrontabili.
E' possibile a questo punto definire le regole strutturali per assegnare i valori delle coordinate (punteggi) a ciascun profilo della configurazione dell'*action system*.
1. Assegnazione del primo punteggio (prima coordinata) ad un profilo *GRXI*:
    - se *G*=0 e *I*=0 → assegnare al profilo il punteggio più basso (1)
    - se *G*=0 e *I*=1 → assegnare al profilo il punteggio successivo più basso (2)
    - se *G*=1 e *X*=0 → assegnare al profilo il punteggio successivo più alto (3)
    - se *G*=1 e *X*=1 → assegnare al profilo il punteggio più alto (4).
2. Assegnazione del secondo punteggio (seconda coordinata) ad un profilo *GRXI*:
    - se *R*=0 e *I*=0 → assegnare al profilo il punteggio più basso (1)
    - se *R*=0 e *I*=1 → assegnare al profilo il punteggio successivo più basso (2)
    - se *R*=1 e *X*=0 → assegnare al profilo il punteggio successivo più alto (3)
    - se *R*=1 e *X*=1 → assegnare al profilo il punteggio più alto (4).

Le configurazioni *action system* differiscono dalle altre per il preciso schema logico che governa le relazioni tra gli item in ogni profilo. Vi sono comunque molte caratteristiche essenziali che qualificano una configurazione come *action system*:

a. la configurazione presenta due item, detti *polari* che sono strutturalmente indipendenti tra loro ovvero: qualsiasi valore relativo ad uno, si può verificare con qualsiasi altro dell'altro item; i concetti basilari rappresentati dagli item polari sono di solito opposti e complementari nei loro significati (come *dare* e *ricevere*);
b. la configurazione presenta un item, detto *polarizzante* o *accentuante*, che è strutturalmente dipendente dagli item polari ovvero alcune combinazioni dei suoi valori con quelli degli item polari non sono osservabili nella configurazione; un tale tipo di item accentua le differenze delineate dagli item polari, nel determinare la configurazione; il concetto basilare rappresentato dall'item polarizzante varia con i contenuti del sistema studiato; il suo significato si caratterizza nella *incapacità di assemblare i poli*;
c. la configurazione presenta un item, detto *moderante* o *attenuante*, che è strutturalmente dipendente dai polari; il suo valore tende ad aumentare con l'aumento simultaneo in entrambi i polari; il suo ruolo nel determinare la struttura della configurazione è quello di attenuare o

ridurre la differenziazione delineata dagli item polari; il concetto basilare rappresentato dall'item moderante è tipicamente distinto nel suo significato dai due polari; esso può "de-enfatizzare" il significato di qualsiasi polo concorrente relativo a quello dell'altro polo.

Tali caratterizzazioni della configurazione *action system* sono sufficientemente generali e consentono di definire molti modelli concreti che sono definibili in termini di relazioni logiche specifiche tra diversi item sistemici.

Nei dati empirici, le regole viste possono essere soddisfatte solo parzialmente o in modo approssimato. Inoltre la generalizzazione di configurazioni sistemiche a dimensionalità maggiori è possibile ma poco indagata.

Può essere utile a questo punto presentare uno schema riassuntivo dei tre modelli di scalogramma presentati:

Configurazione dei profili	Dimensionalità (n. di scale)	Esistenza di ordine tra gli item	Una possibile interpretazione	N. di classificazioni di item	Dimensionalità dello spazio degli item secondo LSA[1]
Guttman	1	si	difficoltà	1 (0) difficoltà[2]	0[3]
Diamante	2	si	processi finiti continui	1 posizione nel tempo	1
Action System	2	no	comportamento tipo action system	3 (2) polare, accentuante, attenuante[4]	2

1. Versione particolare della più generale procedura di analisi *Smallest Space Analysis* (*Lattice Space Analysis*).
2. L'ordine in una scala Guttman è in realtà determinato dall'ordine delle categorie rispetto alla loro relativa difficoltà; quindi, in presenza di dicotomie, gli item non sono differenziati dalle classificazioni; si può dire che il numero di classificazioni è nullo.
3. Per definizione.
4. Gli item accentuanti e attenuanti possono essere formulati come elementi della stessa classificazione, riducendo il numero di classificazioni a 2.

3.4.2 Analisi di scalogrammi multidimensionali

Tale approccio (*Multidimensional Scalogram Analysis*, *MSA*) cerca di rappresentare profili come punti in uno spazio multidimensionale in modo tale che lo spazio possa essere suddiviso in regioni che contengano solo punti (profili) con la stessa struttura. In generale ciò è possibile solo se i dati presentano una particolare struttura.

Con k item osservati su n casi si registreranno n profili, ciascuno dei quali può essere considerato come una cella di una tabella di contingenza a "k vie" oppure come un punto in uno spazio a k dimensioni. I valori che compongono un profilo possono essere così considerati come coordinate in tale spazio. Se per ciascun item l'ordine delle categorie è arbitrario, la rappresentazione spaziale non è unica.

L'obbiettivo di tale approccio è quindi quello di verificare se è possibile definire una rappresentazione equivalente in uno spazio a bassa dimensionalità, soggetto alla restrizione chiamata *regionalità*; ciascun profilo apparirà come una cella o un punto in uno spazio a bassa dimensionalità in modo tale che tutti i profili che presentano la stessa struttura risulteranno essere contigui tra loro.

La rappresentazione dei profili come punti in uno spazio geometrico avviene in modo tale che

a. lo spazio possa essere suddiviso da ciascun item del profilo,
b. la soluzione sia ottenuta in uno spazio a bassa dimensionalità.

Esistono molti algoritmi per l'analisi di scalogrammi multidimensionali che si differenzianoscono tra loro per il livello di scala assegnato agli item e per la forma delle linee di partizione dello spazio:

- *MSA-I*: rappresenta l'algoritmo più generale; assume item nominali e lascia indefinita la forma delle linee di partizione;

- *MSA-II*: assume che tutti gli item hanno lo stesso *range* e che le linee di partizione siano circolari;
- *MSA-III*: assume linee di partizione, item per item, rette tra loro parallele.

Il primo algoritmo è basato sul concetto di contiguità e sulla definizione di *outer-point* e *inner-point*. Data una particolare distribuzione dei punti, osserviamo un item (X) ed un suo particolare valore individuale (x_i); è possibile determinare il limite, il confine della regione dei punti che appartengono a x_i attraverso l'individuazione degli *outer-point*. Tra tutti i punti che non appartengono a x_i ne esiste uno che risulta essere il più vicino a x_i; tale punto è detto *outer* rispetto a x_i. Tutti gli altri punti che non sono *outer* rispetto a x_i sono *inner*. L'insieme di tutti i punti appartenenti allo spazio di x_i si dice che occupano una regione contigua se ciascun *inner-point* di x_i è più vicino ad alcuni *outer-point* di x_i di quanto sia un *outer-point* di ciascun altro elemento di X.
Le deviazioni dalla perfetta contiguità sono prodotte da *inner-point* che sono più vicini ad alcuni *outer-point* di un'altra regione di un *outer-point* della propria regione. Per una data soluzione di *MSA-I* il *coefficiente di contiguità*, che può assumere valori da -1 (minima contiguità) a +1 (massima contiguità), tiene conto del numero dei punti devianti e della dimensione delle deviazioni calcolate per tutti gli item. La procedura è iterativa ed utilizza un algoritmo *steepest ascent*[9].
Nella pratica, gli spazi individuati attraverso questo tipo di analisi non sono facili da esplorare e interpretare, soprattutto nei casi in cui la dimensionalità è maggiore di 2 e quando la soluzione non risulta essere pienamente soddisfacente (coefficiente di contiguità basso). La metodologia si presenta comunque utile soprattutto nei casi in cui si ha l'obiettivo di suddividere gli item secondo diverse tipologie o mettere in rilievo particolari legami esistenti tra le diverse categorie.

3.4.3 Analisi di uno scalogramma parzialmente ordinato

Alla base dell'approccio *POSA* vi è l'assunto che secondo l'universo di contenuto indagato è rappresentato da tutti gli item prescelti, ciascuno dei quali presenta un *range* di possibili punteggi ordinati rispetto a tale universo di contenuto (Shye, 1985). L'analisi effettuata secondo tale approccio riguarda *profili*, ciascuno dei quali è definito come *insieme di soggetti con punteggi identici in tutti gli item*. Tale analisi consente una rappresentazione spaziale senza il bisogno di assunti e di statistiche aggiuntive; tale rappresentazione spaziale facilita l'interpretazione delle direzioni nello spazio; il significato generale del tratto viene comunque associato ad una direzione specificata in precedenza dello spazio geometrico derivato. Al termine dell'analisi sarà possibile assegnare a ciascuno profilo un numero di punteggi minore del numero di variabili originarie, in modo tale che una volta nota la configurazione spaziale, i profili originali possono essere riprodotti a partire dalle coordinate nello spazio geometrico.
Il *POSA* può essere considerato una estensione dello *scaling* tipo Guttman, caratterizzato come abbiamo visto dall'unidimensionalità e da un'unica soluzione di scalabilità tra i profili empirici. Occorre però dire che, dal punto di vista matematico, il passaggio dal caso speciale a quello generale non è risultato né semplice né ovvio; ciò non ha sicuramente favorito lo sviluppo di procedure analitiche d'analisi. Tra gli approcci analitici è possibile identificarne principalmente due, uno dei quali è stato definito da Coombs.

Approccio Coombs

Nel tentativo di proporre un procedimento di analisi di modelli di scalogramma parzialmente ordinali, Coombs, a metà degli anni '60, giunge a descriverne tre basati su un'analisi non-metrica che consente di produrre una configurazione multi-dimensionale. La selezione del modello più appropriato dipende dagli assunti adottati:

[9] A tale proposito si veda il *MultiDimensional Scaling*.

1. *Modello congiuntivo*, secondo il quale un individuo deve "superare" tutti gli item perché si possa affermare che possiede la caratteristica misurata. Per l'analisi si suggerisce di costruire molte scale. Sia gli item più difficili[10] che i profili con il maggior numero di errori/fallimenti individuano dimensioni distinte, ognuna delle quali definisce una scala. Con item dicotomici, ogni scala così ottenuta, definisce un ordine tra tali elementi. Tutte insieme le scale risultano in un ordine parziale tra gli item.
2. *Modello disgiuntivo*, secondo il quale un individuo deve "superare" almeno uno degli item perché si possa affermare che possiede la caratteristica misurata. Il tipo di analisi è molto simile al precedente, si invertono i ruoli dei valori "0" (fallimento) e "1" (superamento). L'analisi descritta può produrre due diverse dimensionalità per la stessa matrice dei dati a seconda del modello adottato.
3. *Modello compensatorio*, secondo il quale un individuo deve "superare" molti item perché si possa affermare che possiede la caratteristica misurata; una mancanza in determinati item può essere compensata dalla previsione in altri. La procedura può essere applicata solo a configurazioni bidimensionali. Essa prevede che si determini l'ordine tra gli item che compongono una tripletta, successivamente la combinazione di tali ordini consente di ottenere due ordini generali tra tutti gli item. Ciò limita l'applicazione ad una sottoclasse molto speciale di tali configurazioni. Tale sottoclasse è, inoltre, caratterizzata dal fatto che l'ordine degli item in una dimensione deve essere contrario di quello nell'altra dimensione.

Approccio Guttman

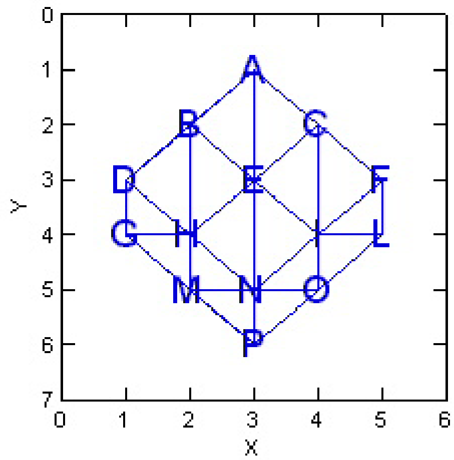
Questo approccio consente la rappresentazione dei profili e delle loro reciproche relazioni d'ordine parziale, attraverso un *diagramma Hasse*. In questo tipo di diagramma i profili vengono presentati come punti ed ogni coppia confrontabile di profili viene collegata da un segmento. Un profilo maggiore di un altro è posizionato nella parte alta del diagramma, rispetto ad una direzione pre-determinata. Con item dicotomici, l'insieme dei possibili profili è rappresentato in modo tale che le proiezioni perpendicolari sugli assi diagonali, che collegano il vertice "00..0" con il vertice "11..1", riflettano le relazioni d'ordine tra i profili confrontabili. Nei casi in cui, a causa dell'interdipendenza tra item, si ottengono solamente alcuni dei possibili profili, la costruzione del *diagramma Hasse* risulta particolarmente semplificata. Tale diagramma utilizza due dimensioni, una detta *joint* (congiunta) e l'altra *lateral* (laterale). Vedremo come in tale rappresentazione i profili che presentano lo stesso punteggio totale presentano la stessa posizione rispetto alla dimensione (o direzione) congiunta ovvero la stessa latitudine; i profili confrontabili ma con punteggio totale diverso presentano posizioni diverse rispetto alla direzione congiunta ma la stessa longitudine.

Di seguito osserviamo la configurazione ottenuta dall'analisi dei profili relativi alla partecipazione alle attività culturali di associazioni sindacali (esempio tratto da Shye, 1985); attraverso tale analisi la sequenza degli item è risultata la seguente:

1°: biblioteca	2=si 1=no
2°: dibattiti	
3°: attività sportive	
4°: cine-forum	
5°: escursioni	

La sequenza determina il contenuto della direzione laterale della configurazione (x); tale direzione va da *preferenza per programmi educativi* (valori bassi di x) a *preferenza per programmi di intrattenimento* (valori alti di x):

[10] Si assume che l'item più difficile sia quello che all'interno di un profilo risulta essere l'unico a non essere superato.



PROFILI	SIMBOLI	X	Y
11111	A	3.000	1.000
21111	B	2.000	2.000
11121	C	4.000	2.000
22111	D	1.000	3.000
21121	E	3.000	3.000
11122	F	5.000	3.000
22211	G	1.000	4.000
22121	H	2.000	4.000
21122	I	4.000	4.000
11222	L	5.000	4.000
22221	M	2.000	5.000
22122	N	3.000	5.000
21222	O	4.000	5.000
22222	P	3.000	6.000

x: direzione laterale ⇒ tipo di partecipazione

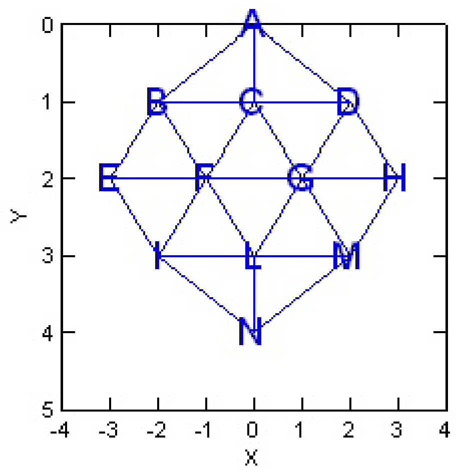
y: direzione congiunta ⇒ estensione della partecipazione totale

Di seguito vediamo i risultati ottenuti in una indagine sulle preoccupazioni rispetto alla perdita del lavoro in seguito alla computerizzazione in un gruppo di lavoratori (esempio tratto da Shye, 1985). In seguito all'analisi la sequenza degli item è risultata essere la seguente:

1°: lavoro	1=si 0=no
2°: stabilità	
3°: prospettive	
4°: interesse	

Gli item 1 e 4 definiscono la polarità del contenuto della direzione laterale: punteggi alti nell'item 1 sono associati con la parte sinistra del diagramma mentre i punteggi alti nell'item 4 sono associati con la parte destra. Gli altri due item, 2 e 3, risultano avere la funzione rispettivamente di polarizzatore e moderatore dell'asse laterale: punteggi alti per l'item 2 sono generalmente associati ai punteggi laterali estremi (molto alti o molto bassi) mentre i punteggi alti nell'item 3 sono associati con punteggi laterali mediani.

La sequenza determina il contenuto della direzione laterale della configurazione (x); tale direzione va da *enfasi sulla perdita di compensi estrinseci* (valori negativi di x) a *enfasi sulla perdita di compensi intrinseci* (valori positivi di x):



PROFILI	SIMBOLI	X	Y
0000	A	0.000	0.000
1000	B	-2.000	1.000
0010	C	0.000	1.000
0001	D	2.000	1.000
1100	E	-3.000	2.000
1010	F	-1.000	2.000
0011	G	1.000	2.000
0101	H	3.000	2.000
1110	I	-2.000	3.000
1011	L	0.000	3.000
0111	M	2.000	3.000
1111	N	0.000	4.000

x: direzione laterale ⇒

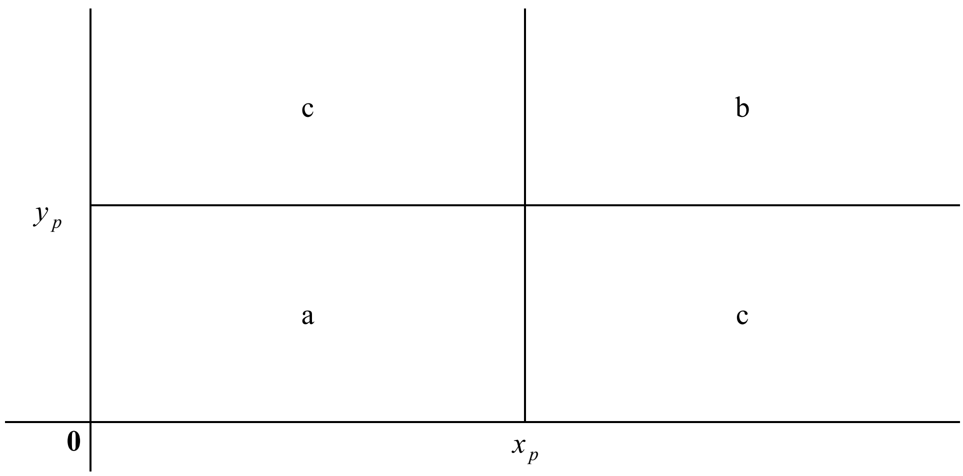
y: direzione congiunta ⇒

Entrambi gli esempi dimostrano gli sforzi effettuati per attribuire significati sostanziali alla seconda direzione nello spazio dei profili (*direzione laterale*). Il significato della *direzione congiunta* è invece determinato dal significato generale del tratto misurato, comune a tutti gli item.
Come risulta chiaro a questo punto, in tale configurazione bidimensionale ciascun profilo individuale (ovvero l'insieme dei punteggi originali degli item per ciascun caso) può essere riprodotto a partire da due punteggi, quello relativo alla direzione laterale e quello relativo alla direzione congiunta. In particolare, nel secondo esempio, un caso che registra un punteggio laterale di -1 e un punteggio congiunto di 2 (ovvero ha un livello intermedio di preoccupazione con una leggera enfasi sulle preoccupazioni riguardanti la perdita di ricompense esterne) ha un profilo di 1010 ovvero egli è preoccupato della perdita del lavoro e della perdita di prospettive ma non della stabilità e dell'interesse.
Per grandi matrici di dati, tale procedura può essere ardua da applicare. Inoltre l'assegnazione del significato alla direzione laterale diviene complessa se non impossibile. E' stata però definita una procedura analitica detta *POSAC*, (*Partially Ordered Scalogram Analysis with Coordinates*) che consente di calcolare e presentare graficamente una configurazione bidimensionale dello scalogrammi e di interpretare la direzione laterale di tale scalogramma.

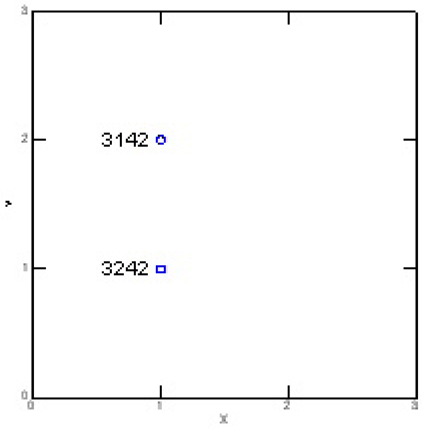
3.4.3.1 POSAC (Partially Ordered Scalogram Analysis with Coordinates)

Il *POSAC* rappresenta essenzialmente una procedura tecnica che consente di determinare l'adattamento dei profili osservati in uno spazio bidimensionale in modo tale da preservare la condizione di ordine ipotizzata. Il successo dell'applicazione di tale approccio ai dati empirici dipende dalla qualità del disegno sperimentale (lo schema concettuale, la scelta delle variabili, la popolazione osservate, ecc.).
L'obiettivo del *POSAC* è quello di, dato un insieme A' di profili osservati con n item, identificare il numero minimo di punteggi che consentano di collocare i profili osservati in modo tale che conservino le relazioni osservate di ordine e di confrontabilità, in altre parole, di verificare se è possibile assegnare due punteggi (corrispondenti a un punto in un piano cartesiano) a ciascun profilo di A', in modo tale che per qualsiasi coppia di profili, la loro relazioni possa essere rappresentata in modo corretto per mezzo dei loro corrispondenti profili di coordinate.
Si ottiene una rappresentazione perfetta quando i punteggi individuati descrivono perfettamente l'ordine e la confrontabilità dei profili originari.
Nel caso in cui non sia possibile ottenere per un certo scalogramma una perfetta rappresentazione in uno spazio bidimensionale l'approccio *POSAC* ricerca una soluzione ottimale attraverso l'osservazione dei valori ottenuti con il *coefficiente di corretta rappresentazione* (*CORREP*), che specifica la proporzione di coppie di profili, pesati attraverso le loro frequenze osservate, la cui relazione di confrontabilità sia perfettamente rappresentata. Il valore del coefficiente di corretta rappresentazione va da 0 a 1 (soluzione perfetta).
Risulta chiaro a questo punto come l'insieme delle coppie di punteggi per ciascun profilo può essere pensato come uno spazio bidimensionale a coordinate cartesiane in cui le coordinate X e Y rappresentano rispettivamente il primo e il secondo punteggio. Le due coordinate per ciascun punto nello spazio XY possono essere considerate un profilo che consente ai punti del piano di formare un insieme parzialmente ordinato. Per ciascun punto nello spazio, è possibile individuare nello spazio tre diverse regioni:

a. regione dei punti che sono minori di un dato punto P;
b. regione dei punti che sono maggiori di un dato punto P;
c. regione (composta da due diverse sezioni) dei punti che non sono confrontabili con il punto P (al centro della configurazione).

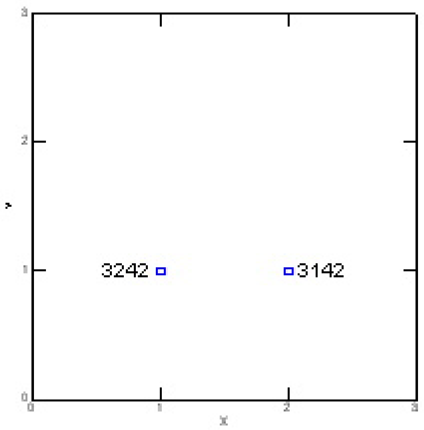


Poniamo di avere tre profili: "3142", "3242" e "1118": all'interno di un diagramma cartesiano esistono tre possibili diverse collocazioni, e tre possibili punteggi, per poter rappresentare le posizioni reciproche dei primi due profili in modo tale che conservino la loro reciproca relazione ("3242" > "3142"):



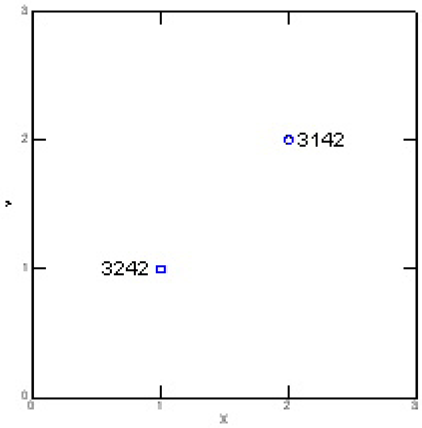
Punteggi

- 3142 → 11
- 3242 → 12



Punteggi

- 3142 → 11
- 3242 → 21



Punteggi
- 3142 → 11
- 3242 → 22

Per poter rappresentare nello stesso diagramma anche il terzo profilo occorre individuare la regione dei punti non confrontabili con le coordinate 1,1 e 2,2 (ovvero all'interno dell'intersezione delle regioni dei punti non confrontabili a 11 e dei punti non confrontabili a 22) inserendo per esempio la coordinata 3; in questo modo il profilo "1118" può assumere la posizione, e il punteggio, 03. Naturalmente non è possibile individuare una nuova posizione nel piano per qualsiasi nuovo profilo in modo tale che vengano preservate le sue relazioni con tutti gli altri; ciò vuol dire che non esiste sempre la possibilità di rappresentare correttamente in uno spazio bidimensionale tutte le relazioni d'ordine esistenti tra i profili di un certo gruppo.

POSAC: Bontà di adattamento

Dato un insieme di profili, è necessario definire un criterio di bontà di adattamento che consenta di stabilire qual è il tipo di collocazione bidimensionale che meglio descrive le relazioni d'ordine tra loro osservate. Un primo indice che consente tale valutazione è rappresentato da un coefficiente basato sulla proporzione di profili rappresentati in modo corretto; tale coefficiente tiene conto anche della frequenza registrata da ciascun profilo. L'algoritmo è basato sulla minimizzazione di una funzione proposta da Louis Guttman negli anni '80. La soluzione *POSAC* rappresenta una approssimazione allo spazio minimo definito. Tale algoritmo è molto sensibile alla *approssimazione iniziale* utilizzata; questa può essere basata su un'ipotesi riguardante, per esempio, la polarità tra item; in altri casi il procedimento per definire l'approssimazione iniziale prevede che vengano eseguiti in successione i seguenti momenti:

a. calcolo della matrice dei coefficienti di debole monotonicità (*weak monotonicity coefficients*, wm)[11];

[11] Il coefficiente di monotonicità tra due item con range ordinati, *A* e *B* è:

$$\textit{coefficiente di monotonicità} = \frac{\sum_q \sum_p (a_p - a_q)(b_p - b_q)}{\sum_q \sum_p |a_p - a_q||b_p - b_q|} \qquad q, p = 1,\ldots,N$$

dove

a_p punteggio del caso *p* per l'item *a*
a_q punteggio del caso *q* per l'item *a*
b_p punteggio del caso *p* per l'item *b*
b_q punteggio del caso *q* per l'item *b*

I valori di questo coefficiente vanno da +1 a -1 e indicano quanto un aumento nei punteggi di un item è accompagnato

b. identificazione dei due item (i_0 e j_0) che presentano la minore correlazione positiva (item estremi);
c. determinazione della posizione di ciascun profilo $a = a_{i_0} ... a_{j_0} ... a_n$:
   - calcolo del punteggio del profilo (somma dei punteggi registrati dagli item),
   - *livellamento* rispetto alla somma $x_a + y_a$ delle sue coordinate;
   - valutazione della sua prossimità relativa ad uno o all'altro degli item estremi attraverso $a_{i_0} - a_{j_0}$ che equalizzata alla differenza tra le coordinate $x_a - y_a$.

In questo modo è possibile ottenere i valori di x_a e y_a per tutti i profili *a*.
Riprendendo i dati di un esempio presentato da Shye (1985) e relativi a quattro item, osserviamo in pratica il procedimento:

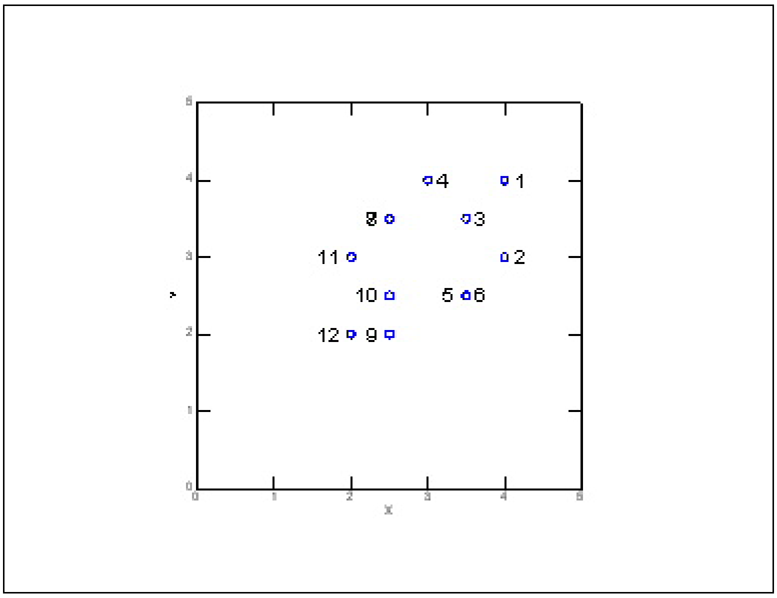
identificazione profilo	profilo	punteggio	item$_1$ - item$_4$	coordinata x dell'approssimazione iniziale	coordinata y dell'approssimazione iniziale
1	2222	8	0	4	4
2	2221	7	1	4	3
3	2212	7	0	3.5	3.5
4	1222	7	-1	3	4
5	2121	6	1	3.5	2.5
6	2211	6	1	3.5	2.5
7	1212	6	-1	2.5	3.5
8	1122	6	-1	2.5	3.5
9	2111	5	1	2.5	2
10	1211	5	0	2.5	2.5
11	1112	5	-1	2	3
12	1111	4	0	2	2

I coefficienti *wm* per i quattro item di tale configurazione, assumendo frequenze uguali per tutti i profili, sono date nella seguente matrice.

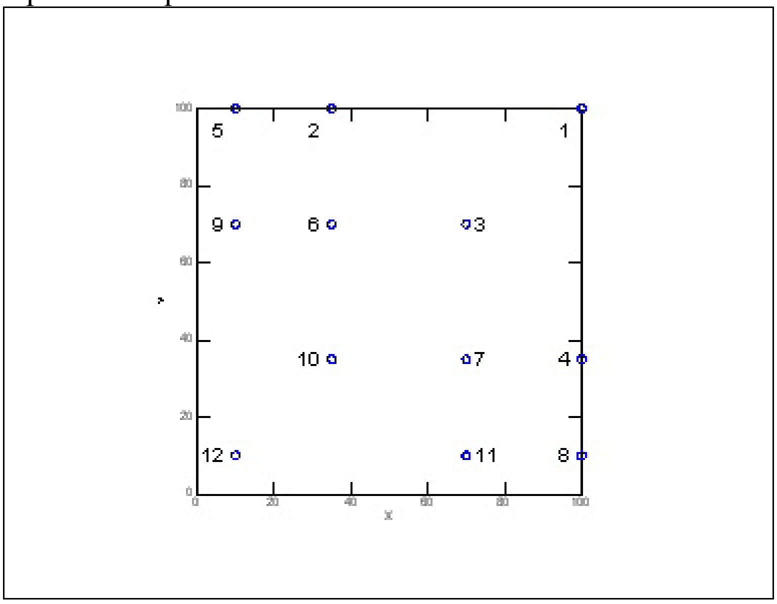
	I$_1$	I$_2$	I$_3$	I$_4$
I$_1$	1			
I$_2$	1/3	1		
I$_3$	1/3	1/17	1	
I$_4$	-3/5	1/3	1/3	1

In tale matrice osserviamo come gli item 1 e 4 sono quelli con la più bassa correlazione positiva. Nella terza e quarta colonna della tabella vengono presentati i due parametri iniziali per la configurazione dei profili. Quando questi due parametri iniziali sono equalizzati rispettivamente a $x_a + y_a$ e $x_a - y_a$, è possibile individuare i valori di x_a e y_a per ogni profilo. Tali coordinate sono presentate nelle colonne 5 e 6 della tabella e proiettate nel seguente diagramma in cui ciascun profilo dal numero di identificazione del profilo (prima colonna della tabella).

da un aumento (a da una diminuzione) nei punteggi dell'altro item.



L'approssimazione iniziale viene migliorata attraverso ripetute iterazioni, effettuate in due fasi, nel tentativo di trovare una approssimazione soddisfacente. L'output della soluzione finale presenta per ciascun profilo i valori *X*, *Y*, *J=X+Y*, *L=100+X-Y* e quindi la configurazione bidimensionale che, nel caso dell'esempio presentata potrebbe essere:



Uno dei pochi *package* che presentano al loro interno la procedura *POSAC* è il *Systat* il cui procedimento pratico, in sintesi, prevede i seguenti passaggi:

a. ordinamento degli item da sinistra a destra in modo tale che la dimensione orizzontale mostri i valori 1 che si spostano da sinistra a destra all'interno dei profili;
b. ordinamento dei profili in senso verticale rispetto al punteggio totale;
c. ordinamento dei profili da sinistra a destra;
d. individuazione di profili che non si adattano al modello e verifica che l'intera soluzione che venga influenzata da essi.

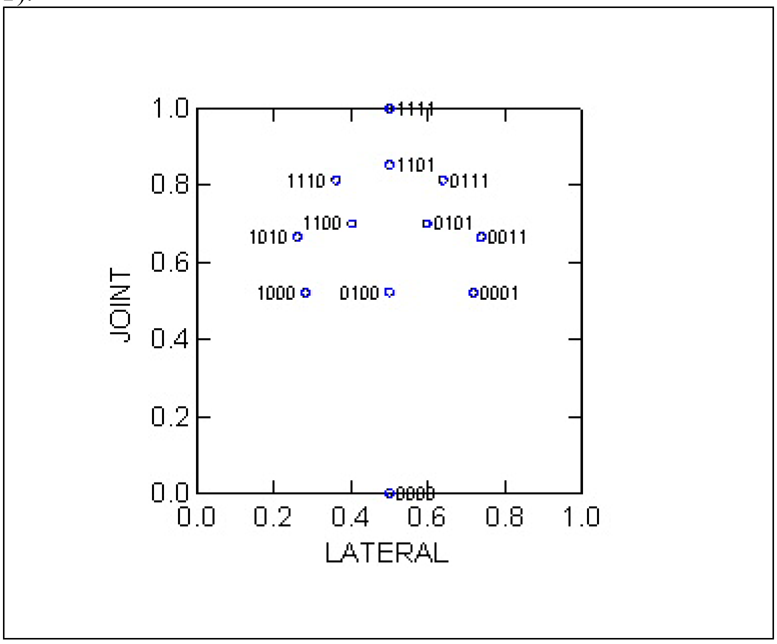
In un certo senso l'ultimo momento rappresenta un requisito ambiguo che comunque dipende dal primo momento; se per esempio avessimo i due profili "1010" e "0101", scambiando il secondo con il terzo item, avremmo i profili "0011" e "1100", corrispondenti a quelli estremi, che ci consentirebbero di ottenere una soluzione ben adattata.

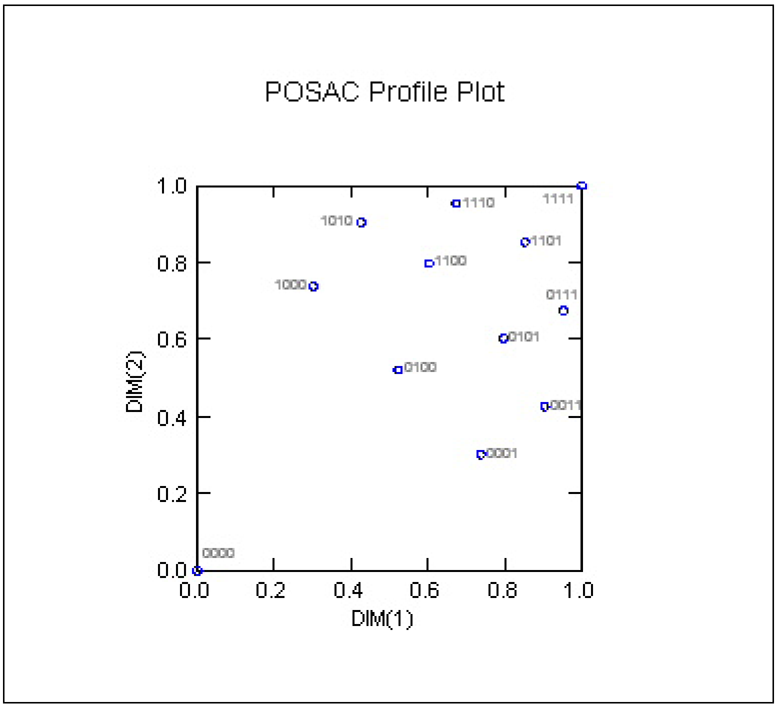
Per realizzare il primo passaggio il procedimento *POSAC* prevede la definizione della matrice di monotonicità attraverso il coefficiente di monotonicità e quindi l'ordinamento della matrice ottenuta per mezzo di un algoritmo di *multidimensional scaling*. Il procedimento è iterativo e produce al termine l'ordine degli item per la definizione del profilo, le coordinate per ciascun profilo e un valore che consente di valutare la rappresentazione ottenuta in termini di perdita (*final loss value*); minore è tale valore, migliore è la rappresentazione. Prima di predisporre la rappresentazione grafica, la procedura programmata all'interno del *package Systat* calcola la radice quadrata delle coordinate per rendere la direzione laterale lineare piuttosto che curvilinea; in questo modo il grafico risulta ruotato di 45°.

Riprendendo i dati dell'ultimo esempio si ottengono per ciascun profilo le coordinate per la direzione *joint* e la direzione *lateral* e le corrispondenti nuove coordinate (rispettivamente Dim_1 e Dim_2):

Profili	Dim 1	Dim 2	*Joint*	*Lateral*
1111	1.000	1.000	1.000	0.500
1101	0.853	0.853	0.853	0.500
1110	0.674	0.953	0.814	0.360
0111	0.953	0.674	0.814	0.640
1100	0.603	0.798	0.700	0.403
1010	0.426	0.905	0.665	0.261
0011	0.905	0.426	0.665	0.739
0101	0.798	0.603	0.700	0.597
1000	0.302	0.739	0.520	0.281
0001	0.739	0.302	0.520	0.719
0100	0.522	0.522	0.522	0.500
0000	0.000	0.000	0.000	0.500

Di seguito vediamo la rappresentazione spaziale con le coordinate originali (*lateral*joint*) e con le coordinate trasformate (*dim 1*dim 2*):





I punteggi finali ottenuti con la procedura *POSAC* possono essere interpretati osservando le relazioni tra i contenuti degli item e il ruolo che giocano nello strutturare lo spazio *POSAC*.

Appendice.
ALLE ORIGINI DEL MODELLO DETERMINISTICO: LA SCALA BOGARDUS

La scala Bogardus propone un modello che può essere considerato il precedente storico di quello deterministico. Tale scala è nata per poter misurare gli atteggiamenti verso gruppi etnici ma ha trovato applicazioni anche per misurare gli atteggiamenti verso classi sociali, gruppi religiosi, e così via. In questa scala gli item vengono definiti come gradini di una scala con *problematicità* crescente. La scala originariamente misurava la *distanza sociale* (Bogardus, Emory, "Racial Distance Changes in the United States During the Past Thirty Years", *Sociology and Social Research*, Vol. XLIII, pp. 127-130, November, 1958); tale concetto si riferisce al modo e all'intensità con cui le persone percepiscono e tendono ad avere rapporti con altre persone nelle diverse situazioni di interazione sociale; essa doveva individuare i diversi gradi in cui tale fenomeno si manifesta, in modo da fornire dei dati per una adeguata interpretazione (Arcuri, 1974); le affermazioni che definivano gli item indicano gerarchicamente diversi livelli di "intimità" o di contatto sociale, con un certo gruppo di individui: si passa infatti dallo stretto vincolo di parentela, attraverso i legami di amicizia, di lavoro, fino al rifiuto di un qualsiasi contatto. In particolare la scala originaria era basata su una serie di domande che richiedevano se si era disposti ad accettare i neri come residenti nella stessa città, come vicini di casa, come amici, come mariti della figlie. La scala assume risposte cumulative in quanto chi accetta un negro in casa lo accetta anche come vicino di casa. Di seguito vediamo una sequenza di insieme di item che definisce una tipica scala *Bogardus*:

1. Lo accetterei come parente stretto	5. Lo accetterei come concittadino
2. Lo accetterei al mio club, come amico personale	6. Lo accetterei nel mio paese come turista
3. Lo accetterai come vicino di casa	7. Lo escluderei dal mio paese
4. Lo accetterei come compagno di scuola	

L'ipotesi dell'autore è che quanto più è stretto il contatto che il soggetto è disposto ad ammettere con una persona ad un certo gruppo, tanto più favorevole è il suo atteggiamento nei confronti delle persone di tale gruppo; al contrario, quanto meno intimo è il contatto che il soggetto è disposto ad ammettere, tanto meno favorevole è il suo atteggiamento. Dal punto di vista dell'analisi il giudizio emesso dal soggetto può essere sintetizzato in due indici:

a. *ampiezza del contatto sociale* (*Social Contact Range*, *SCR*): indice rappresentato dal numero di item di contatto sociale che il soggetto è disposto ad ammettere.
b. *distanza del contatto sociale* (*Social Contact Distance*, *SCD*): indice rappresentato dal numero di item di più stretto contatto sociale che non vengono scelte dal soggetto.

Tra i due indici vi è una relazione inversa: quanto minore è il numero di item di contatto ammesse, tanto maggiore è la distanza sociale che il soggetto intende stabilire tra sé e la persona giudicata. Naturalmente nel caso di rifiuto totale della persona il *SCR* assume valore 0, dato che non è ammesso nessun tipo di contatto, mentre *SCD* diventa pari al massimo ottenibile. Nel caso in cui si sottoponga la scala ad un campione di soggetti, è possibile calcolare

- *SCR-medio*: indice medio del numero di item di contatto sociale scelte dai soggetti;
- *SCD-medio*: indice medio del numero di item di contatto sociale non accettate dai soggetti.

E' importante che i due indici siano utilizzati contemporaneamente in quanto il solo indice di ampiezza del contatto sociale non fornisce un'informazione completa. Uno dei problemi riscontrati in questo tipo di scala è dato dal fatto che non esiste alcun modo per determinare la reale distanza tra i vari punti sulla scala.

Tale scala risulta comunque utile nei casi in cui si vogliano studiare gli atteggiamenti degli individui verso particolari gruppi di persone ma anche nei casi in cui si vogliano misurare gli atteggiamenti verso classi sociali, occupazioni e gruppi religiosi. Infatti le categorie di accettazione o di rifiuto possono essere formulate in maniera diversa a seconda delle esigenze della ricerca condotta; è possibile per esempio affrontare ricerche riguardanti contesti culturali diversi o aree di problemi diversi (negli anni passati in Italia si sono visti esempi di applicazione nello studio del pregiudizio nei confronti dei meridionali o degli atteggiamenti nei confronti degli obiettori di coscienza mentre di recente si sono viste applicazioni che riguardavano i rapporti con gli extra-comunitari).

4. I MODELLI CUMULATIVI. L'APPROCCIO PROBABILISTICO

I modelli cumulativi-probabilistici si richiamano all'*Item Response Theory* (*IRT*), che a sua volta si rifà alla teoria del tratto latente, formalmente attribuisce la variazione nelle risposte a parametri riguardanti sia gli item che i casi; in questo senso le risposte individuali costituiscono un indicatore della relazione tra item e soggetto, in quanto ciascuna di esse è frutto della posizione rispetto all'attributo sia dell'item che del soggetto. Quindi la posizione di accettazione (superamento, accordo) o di rifiuto (fallimento, disaccordo) del soggetto rispetto ad un particolare item dipende sia dalle sue caratteristiche (capacità, atteggiamenti, attitudini, opinioni, ecc.) che da quelle riflesse dall'item (difficoltà, presentazione di un'opinione, ecc.).
In altre parole, secondo la *teoria del tratto latente*, la 'performance' di una data misura è funzione della posizione del caso lungo il continuum del tratto misurato e dell'errore casuale. L'obiettivo è quello di stimare tale funzione, assumendo l'indipendenza locale (la 'performance' di un caso rispetto ad un item è statisticamente indipendente ed è spiegabile solo in funzione delle caratteristiche individuali). La relazione è descritta dalla curva caratteristica dell'item (*Item Characteristic Curve*, *ICC*).
La definizione di modelli matematico-probabilistici ha consentito la definizione di criteri statistici per la valutazione della bontà di adattamento del modello ai dati e l'individuazione di procedure soddisfacenti di accettazione o rifiuto dell'ipotesi di misurazione.
I modelli di *Item Response Theory* si distinguono dagli altri in quanto non sono *test-dependent*, ovvero sono espressi a livello di item (*item oriented*) e non a livello di gruppo di item (*test-oriented*) e producono misurazioni precise (quantitative) sia dell'item (in genere in termini di difficoltà) che dei casi (in genere in termini di capacità) sulla stessa scala,
Inoltre secondo tali modelli le caratteristiche di ciascun item e di ciascun soggetto non cambiano in funzione dei campioni per la sperimentazione, ovvero

- gli item selezionati *non* sono *group-dependent*,
- i punteggi individuali *non* sono *test-dependent*.

Ciò consente

- la validazione sulla base di una sola applicazione, senza il bisogno di definire strumenti paralleli in senso stretto,
- la costruzione di banche di item validati su gruppi diversi (<u>economicità</u>),
- la scelta di item che coprano l'intera estensione del continuum (<u>ottimizzazione della selezione degli item</u>).

I metodi di analisi degli item inoltre consentono di ottenere precise informazioni sulle modalità interne di risposta utilizzabili con <u>funzioni *diagnostiche*</u>.[1]

4.1 GLI ASSUNTI

Gli assunti che definiscono tali modelli sono l'*unidimensionalità*, l'*indipendenza locale* e la funzione che lega risposta e item.

[1] Riferimenti per l'intero capitolo: Andersen, 1972; Andersen, 1973; Andrich, 1988; Boch, 1981; Hambleton, 1991; Koch, 1995; Lord, 1952, 1968, 1974, 1980, 1984; Ludlow, 1995; McDonald, 1989; Misley, 1984; Rasch, 1960; Sijtsma, 2002; Swaminathan, 1982, 1985, 1986; Torgerson, 1958; Urry, 1974.

- UNIDIMENSIONALITA'
  Si assume che la risposta di un soggetto ad un item sia determinata e possa essere spiegata da una sola componente, da un solo fattore unico dominante chiamato tratto latente; gli item individuati devono misurare solo tale caratteristica. In realtà l'assunto (comune ad altri modelli) è difficile da soddisfare in modo rigoroso.
  A tale proposito si pensi come la misurazione di certi attributi sia difficile
    - perché non necessariamente immodificabili per l'intervento di componenti (apprendimento, esperienza, memoria, ecc.) che possono modificare nel tempo il tratto misurato,
    - per l'influenza che in misura variabile e non nota possono esercitare altri fattori (cognitivi, di personalità, di motivazione, di ansietà, di capacità di lavorare velocemente, di tendenza a rispondere in maniera casuale nel caso di risposte dubbiose, ecc.).

  Nell'ambito dell'*IRT*, comunque, sono stati definiti anche modelli *multidimensionali* che possono essere adottati nei casi in cui si ipotizzano risposte spiegabili da più di una dimensione.

- INDIPENDENZA LOCALE
  Secondo l'assunto di indipendenza locale, tenuto costante il valore del tratto latente che influenza la risposta, non esiste alcuna relazione tra le risposte di ciascun soggetto ai diversi item; in altre parole le risposte di un soggetto agli item sono statisticamente indipendenti e sono spiegabili solo in funzione delle caratteristiche individuali. La dimensione misurata costituisce lo *spazio completo latente* (*complete latent space*). Quando l'assunto di unidimensionalità è soddisfatto, tale spazio riguarda una sola capacità. La proprietà d'indipendenza locale può essere matematicamente formalizzata nel modo seguente:

$$P(U_1,U_2,...,U_i,...,U_n|d) = P(U_1|d)P(U_2|d)...P(U_i|d)...P(U_n|d) = \prod_{i=1}^{n} P(U_i|d)$$

dove

n numero totale di item

d capacità che influenza la risposta del soggetto ad uno strumento

U_i risposta del soggetto all'item *i* (*i*=1, 2, ..., *n*)

$P(U_i|d)$ probabilità di risposta di un soggetto con capacità *d*

con

$P(U_i = 1|d)$ probabilità di risposta corretta

$P(U_i = 0|d)$ probabilità di risposta scorretta

ovvero la probabilità di risposta di un soggetto ad un gruppo di item è uguale al prodotto delle probabilità associate alle risposte del soggetto a ciascuno degli item[2].

Con tale definizione, l'assunto di indipendenza locale sembra complicato da soddisfare: è difficile pensare che le risposte di un soggetto a molti item non siano tra loro correlate ovvero che siano tra loro indipendenti; in realtà l'unico elemento che deve legare tali risposte è il valore del tratto latente: tenendo costante tale valore (analisi di correlazione parziale), le risposte devono risultare non correlate. Quindi se la relazione tra le diverse risposte di un soggetto ad un gruppo di item è spiegata da un unico aspetto (caratteristica misurata), rendendo costante tale aspetto, le risposte divengono indipendenti[3]. Per questa ragione l'assunto di indipendenza locale viene detto anche assunto di indipendenza *condizionale*.

Gli assunti di *unidimensionalità* e di *indipendenza locale* sono molto collegati tra loro, infatti se il primo risulta valido, lo sarà conseguentemente anche il secondo: identificato uno spazio latente completo, che influenza le risposte, è soddisfatto l'assunto di indipendenza locale.

---

[2] Per esempio, se il modello di risposta a tre item di un soggetto è *1,1,0* (ovvero $U_1=1$, $U_2=1$ e $U_3=0$) allora l'assunto di indipendenza locale implica che:

$$P(U_1 = 1, U_2 = 1, U_3 = 0|d) = P(U_1 = 1|d)P(U_2 = 1|d)P(U_3 = 0|d) = P_1 P_2 Q_3$$

[3] Si tratta di un principio che si ritrova anche in altri modelli come quello fattoriale.

Vediamo un esempio. Poniamo di avere un item che misura la capacità matematica ma che richiede anche un alto livello di padronanza linguistica; possiamo trovarci davanti a due situazioni:

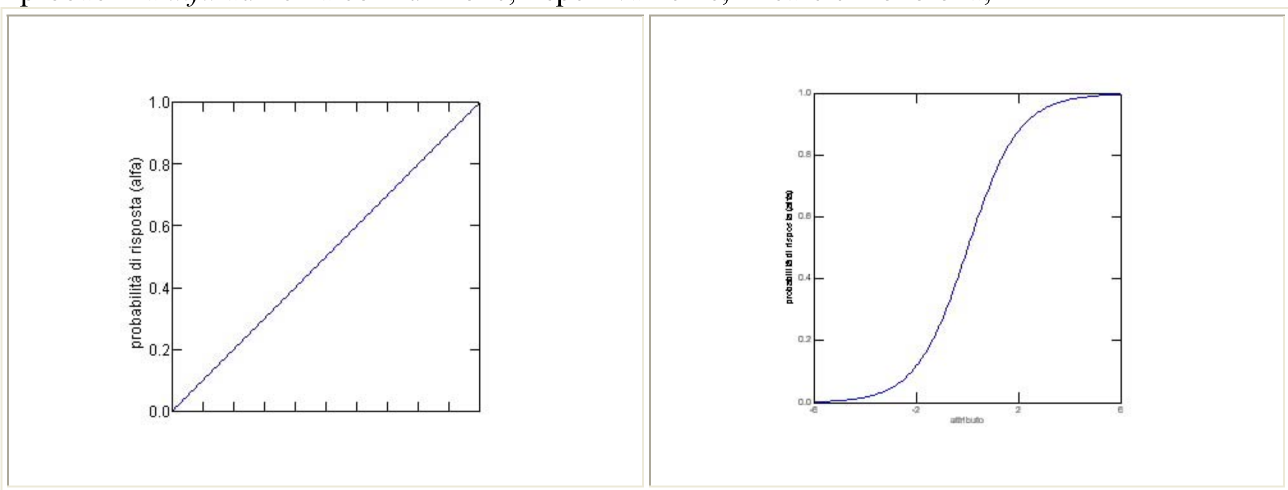
- i soggetti presentano tra loro diversi livelli di competenza linguistica; in questo caso si può ipotizzare che i soggetti con bassa capacità linguistica non rispondano correttamente all'item indipendentemente dalla loro competenza matematica; quindi una dimensione estranea alla capacità matematica influenza la risposta all'item; conseguentemente l'assunto di indipendenza locale non può venire soddisfatto;
- i soggetti presentano tra loro gli stessi livelli di competenza linguistica; in questo caso si può ipotizzare che le risposte ottenute possano essere attribuite solo alla capacità matematica: l'assunto di indipendenza locale è soddisfatto.

Altri casi in cui non è possibile sostenere l'indipendenza locale sono quelli in cui, per esempio, l'item contiene già elementi o informazioni tali da influenzare la risposta all'item in questione o ad altri item. E' possibile che solo alcuni soggetti percepiscano tali elementi e che quindi ne risultino influenzati; tale capacità di percezione risulta essere una dimensione al di fuori della capacità realmente misurata.

- FUNZIONE CARATTERISTICA DELL'ITEM

  Tale assunto riguarda la relazione tra la variabile non direttamente osservabile (tratto latente) e la variabile realmente osservabile (risposta all'item): maggiore è il valore del tratto latente maggiore è la probabilità di rispondere affermativamente (o, secondo i casi, correttamente). La relazione tra risposta e dimensione latente crescente può essere descritta da una funzione monotona, secondo la quale all'aumentare del livello di una data caratteristica (per esempio di capacità), aumenta la probabilità di rispondere in modo affermativo o nella direzione della dimensione misurata (per esempio in modo corretto) ad un item; è quindi possibile affermare che *la probabilità di un soggetto di rispondere correttamente ad un certo item dipende dalla disposizione del soggetto e dalle caratteristiche dell'item* ovvero dipende dalla correlazione tra caratteristica misurata e caratteristiche dell'item. La probabilità di dare una risposta affermativa/corretta in funzione del valore individuale del tratto latente per ciascun item è definita da una funzione matematica (*Item Characteristic Function*, *ICF*) e rappresentata da una curva (*Item Operating Characteristics* o *Item Characteristic Curve*, *ICC*).

  E' possibile identificare e definire diverse *ICC* come si può osservare dai due esempi rappresentati di seguito che indicano come all'aumentare dell'intensità dell'attributo la probabilità *alfa* aumenta con funzione, rispettivamente, lineare e monotona;



4.1.1 *Invarianza dei parametri*

La caratteristica che maggiormente distingue l'*IRT* dalla teoria classica della misurazione è la proprietà dell'invarianza della capacità del soggetto e dei parametri relativi all'item. Secondo tale proprietà i parametri che caratterizzano

- ciascun item, non dipendono dalla distribuzione della capacità dei soggetti,

- la capacità di un soggetto, non dipendono dal gruppo di item.

Se il modello di *IRT* si adatta ai dati, la *ICC* di un determinato item risulta la stessa, indipendentemente dalla distribuzione della capacità del campione di soggetti utilizzati per stimare i parametri degli item.
In altre parole i valori di capacità più alti presentano anche probabilità più alte di rispondere correttamente all'item rispetto ai soggetti con valori più bassi, indipendentemente dalla distribuzione del gruppo di appartenenza; quindi i soggetti che presentano lo stesso livello di capacità hanno la stessa probabilità di fornire una risposta corretta all'item, indipendentemente dal gruppo di appartenenza.
Sapendo che la probabilità di successo per un soggetto con una data capacità è determinata dai parametri dell'item, anche i parametri dell'item per i due gruppi devono essere uguali.
La proprietà di invarianza è importante non solo teoricamente in quanto consente importanti applicazioni come quella di effettuare confronti, di costruire banche di item, di analizzare l'errore sistematico di particolari item e di stimare gli errori standard delle stime delle capacità individuali, diversamente dal modello classico di misurazione in cui viene definito un solo errore uguale per tutti i casi[4].

4.2 I MODELLI

Esistono molte tecniche che consentono di stimare la versione probabilistica del modello di *scaling* cumulativo. Tutte sono in grado di gestire le deviazioni empiriche dalla forma perfetta di scala cumulativa. Con tali tecniche è inoltre possibile valutare l'adattamento del modello di *scaling* rispetto all'ipotesi nulla che le risposte ai singoli item sono tra loro statisticamente indipendenti (Swaminathan, 1982, 1985, 1986).
La distinzione tra i diversi modelli basati sull'*IRT* è fatta sulla base del numero e del tipo di parametri utilizzati nel definire la funzione adottata:

- Modello logistico con un parametro: si tratta di uno dei modelli probabilistici più noti, conosciuto spesso nella versione detta *Rasch* dal nome dello studioso che la ha sviluppata (1960); secondo questo modello le *ICC* sono rappresentate da curve logistiche parallele i cui valori variano tra 0 e 1; questo modello ha la proprietà dell'"oggettività specifica": i soggetti sono misurati su una scala a rapporti indipendente dagli item utilizzati e viceversa; rappresenta il miglior approccio alla individuazione di gruppi di item per i quali le risposte individuali ai primi item determinano la selezione, tra i successivi item, dei migliori per la misurazione del particolare soggetto.
- Modello con due parametri: Lord nel 1952 è stato il primo a sviluppare un modello con due parametri basato sull'assunto rappresentato dall'*ogiva normale* (distribuzione normale cumulata). Una *ICC* con tale forma descrive un item che è maggiormente discriminante nei punti in cui la curva è più ripida. Più ripida è tale sezione della curva, maggiore è la correlazione biseriale dell'item con l'attributo[5]. Al diminuire della correlazione tra item e attributo, la curva tende ad appiattirsi fino ad assumere la posizione di linea retta orizzontale. L'adozione del modello che fa riferimento all'ogiva normale consente di identificare per ciascun

---

[4] La caratteristica dell'invarianza non è nuova per la statistica. Essa riguarda infatti anche i modelli di regressione lineare: stabilita l'equazione e verificato il modello di regressione, la pendenza e l'intercetta della retta sarà la stessa in qualsiasi sottopopolazione della variabile *X*. Ricordiamo invece che un indice come il coefficiente di correlazione, parametro non caratterizzante direttamente la retta di regressione, cambia se calcolato su un sottogruppo. Quindi mentre il parametro di pendenza non dipende dalle caratteristiche della sottopopolazione, il coefficiente di correlazione sì.
L'applicazione dello stesso concetto anche ai modelli di *IRT* ci permette di considerare questi come modelli di regressione non lineare.

[5] Se tale sezione fosse verticale, le code scomparirebbero, l'item correlerebbe perfettamente con l'attributo; in questo caso l'item risponderebbe alle caratteristiche proprie del modello deterministico-cumulativo.

item una zona critica (individuata da un punto critico di discriminazione) corrispondente al livello di incertezza delle risposte dei soggetti; allontanandosi da tale zona in entrambe le direzioni, l'incertezza si riduce in modo evidente: la probabilità di dare una risposta positiva all'item è bassa al di sotto del valore che identifica la zona critica e alta al di sopra. Questo consente di definire, nel caso in cui si misurino capacità, la difficoltà dell'item. L'individuazione del punto critico per ogni item consente di selezionare gli item che consentono di discriminare in determinati punti.

Birnbaum successivamente sostituì la funzione normale con la funzione *logistica*, che presenta il vantaggio di poter essere più facilmente e convenientemente trattata dal punto di vista matematico e di godere di alcune importanti proprietà statistiche[6].

- Modello Birnbaum (tre parametri): le *ICC* sono rappresentate da curve logistiche che possono presentare diverse pendenze e asintoti che riflettono la possibilità di rispondere correttamente anche a caso e, in alcune varianti, prevedono item a scelta multipla; tale modello consente di selezionare il gruppo di item più "informativo" per un dato gruppo di soggetti.
- Modello Mokken: modello non parametrico che richiede curve (con valori che vanno da 0 e 1) che non si intersecano e che conducono ad un numero limitato di violazioni delle caratteristiche degli item di una scala Guttman (item ordinali); tale modello, oggetto di un vasto dibattito, presenta gli assunti meno restrittivi e un concetto diverso di qualità della misurazione.

Anche se i modelli probabilistici in genere fanno riferimento:
- alla caratteristica del soggetto in termini di *capacità*,
- alla caratteristica dell'item in termini soprattutto di *difficoltà*,

tali modelli possono trovare applicazione e quindi essere generalizzati anche ad altri casi.
Tra i più noti e diffusi modelli di *IRT* di seguito esamineremo i più diffusi modelli logistici.

4.2.1 I modelli logistici

4.2.1.1 Modello con un parametro

Il modello con un parametro assume che l'unica caratteristica che influenza la risposta del soggetto sia la difficoltà dell'item stesso (b_i). La *ICF* che descrive tale modello è la seguente:

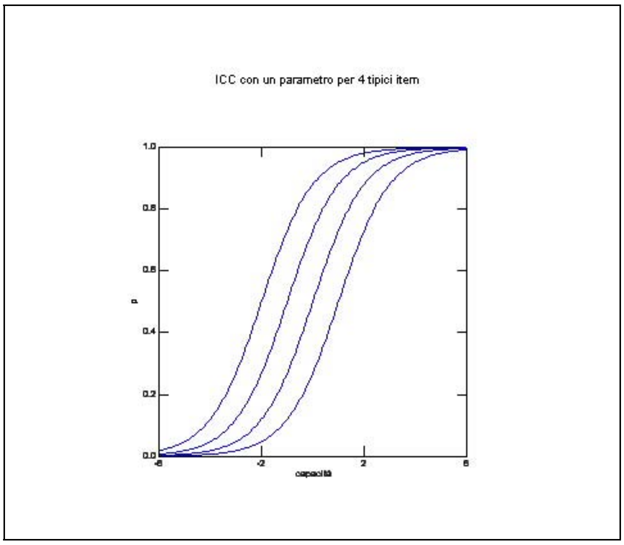
$$P_i(d) = \frac{e^{d-b_i}}{1+e^{d-b_i}} \qquad i = 1,2,\ldots,n$$

dove

$P_i(d)$	probabilità del soggetto *j* con capacità *d* di rispondere correttamente all'item *i*
e	numero trascendente (come il *p-greco,* π) il cui valore è 2.718 (corretto ai tre decimali)
d	parametro di capacità del soggetto (*j*)
b_i	parametro di difficoltà dell'item *i*
n	numero di item.

Il parametro b_i può essere interpretato come indicatore della posizione della *ICC* in relazione alla scala di capacità. Tale parametro infatti rappresenta il punto della scala di capacità dove la probabilità di dare una risposta corretta è 0.5; maggiore è il valore del parametro b_i, maggiore è il livello di capacità richiesto ad un soggetto per aver una possibilità del 50% di rispondere in modo corretto, maggiore è la difficoltà dell'item.
In genere, quando i valori di capacità dei soggetti sono standardizzati, i valori di difficoltà degli item (b_i) variano da -2.0 a +2.0. Di seguito sono presentati alcuni esempi di *ICC* per il modello con un parametro (lateralmente sono riportati i relativi valori del parametro).

[6] Per alcuni utili richiami sulle definizioni di logaritmi e *logit* si veda l'appendice a questo capitolo.



Le curve differiscono solo per la loro posizione rispetto alla scala di capacità; infatti da tale grafico si osserva che gli item più

- difficili sono localizzati all'estrema destra ovvero nella parte in cui i valori di capacità sono più alti (valori di b_i vicini a 2.0);
- facili sono posizionati all'estrema sinistra ovvero nella parte in cui i valori della scala di capacità sono più bassi (valori di b_i vicini a -2.0)

Secondo questo approccio i soggetti con capacità molto bassa hanno probabilità nulla di rispondere correttamente all'item. Questo modello logistico, matematicamente equivalente al modello *Rasch*, è basato su assunti restrittivi, la cui adeguatezza dipende dalla natura dei dati e dall'importanza dell'applicazione stessa.

4.2.1.2 Modello con due parametri

Il modello con due parametri può essere considerato l'estensione del precedente modello rispetto al quale presenta un nuovo parametro: <u>discriminazione dell'item</u>[7] (a_i); questo è proporzionale alla pendenza della *ICC* nel punto b_i della scala di capacità.
In teoria il parametro di discriminazione dell'item è definito su una scala che va da $-\infty$ a $+\infty$; nella pratica tale parametro difficilmente supera 2.0; i valori molto bassi definiscono *ICC* per le quali la capacità aumenta gradualmente mentre i valori molto alti di a_i identificano *ICC* con le maggiori pendenze corrispondenti quelli che sono gli item più utili a discriminare i soggetti in punti diversi del continuum di capacità. Accanto a tale parametro è stato introdotto anche un fattore di *scaling* che rende la funzione logistica più simile possibile alla funzione normale (D[8]).
L'equazione matematica alla base del modello con due parametri (Birnbaum, 1968) è la seguente:

$$P_i(d) = \frac{e^{Da_i(d-b_i)}}{1+e^{Da_i(d-b_i)}} \qquad\qquad i = 1,2,\dots,n$$

dove

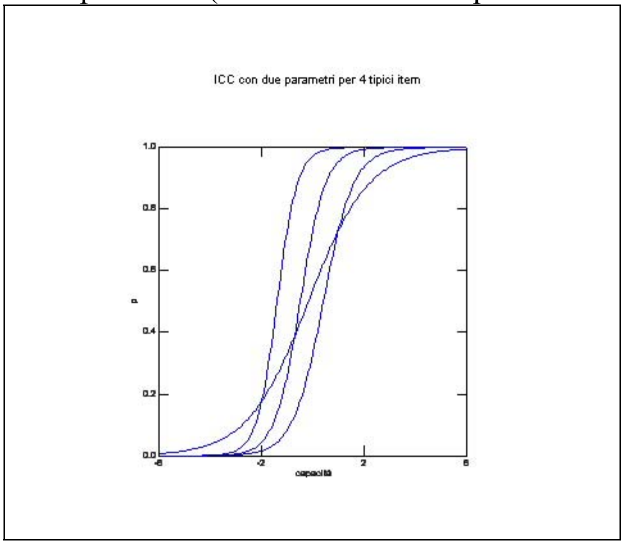
n	numero di item
$P_i\,(d)$	probabilità del soggetto *j* con capacità *d* di rispondere correttamente all'item *i*
e	numero trascendente (come il π) il cui valore è 2.718 (corretto ai tre decimali).
d	parametro di capacità del soggetto (*j*)

[7] A differenza del modello classico, che considera tutti gli item ugualmente discriminanti, il valore di tale parametro cambia da item a item.

[8] E' stato dimostrato che quando *D*=1.7, i valori di $P_i(d)$ per il modello dell'ogiva normale con due parametri e per il modello logistico con due parametri differiscono nel valore assoluto di meno di 0.01 per tutti i valori di *d*.

b_i parametro di difficoltà dell'item *i*
a_i parametro di discriminazione dell'item *i*
D fattore di *scaling* che avvicina la funzione logistica a quella normale.

Di seguito vediamo alcuni esempi di *ICC* (lateralmente sono riportati i relativi valori dei parametri).



Le curve non sono parallele come nel caso del precedente modello; ciò è dovuto ai diversi valori del parametro di discriminazione cui corrispondono pendenze diverse. Inoltre per ciascuna curva le asintoti sono uguali a zero; ciò indica che oltre i parametri considerati non esiste alcuna altro elemento che spieghi le risposte.

4.2.1.3 Modello con tre parametri

Il nuovo parametro introdotto con questo modello rappresenta la probabilità dei soggetti con bassa capacità di rispondere correttamente all'item (c_i). Tale parametro consente di prendere in considerazione l'esecuzione rispetto al limite inferiore della scala di capacità, dove *tirare a indovinare* può rappresentare un fattore che influenza la risposta. Tale parametro risulta particolarmente adatto ai casi in cui si sottopongono scale di risposta in cui appaiono scelte convincenti ma scorrette.

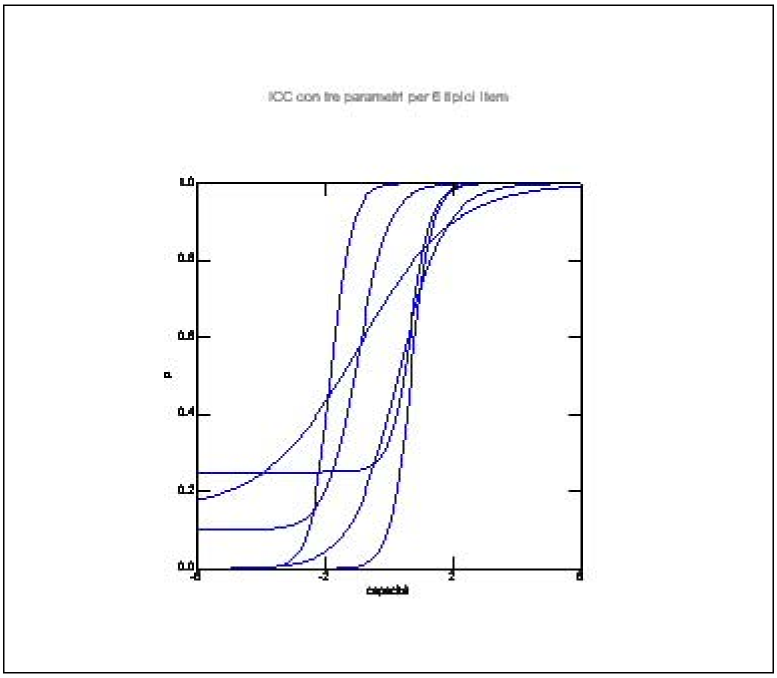
La funzione matematica alla base del modello logistico con tre parametri è la seguente:

$$P_i(d) = c_i + (1 - c_i)\frac{e^{Da_i(d-b_i)}}{1 + e^{Da_i(d-b_i)}} \qquad i = 1,2,\ldots,n$$

dove

$P_i(d)$ probabilità del soggetto *j* con capacità *d* di rispondere correttamente all'item *i*
e costante (come il π) il cui valore è 2.718 (corretto ai tre decimali)
c_i parametro del *livello pseudo-casuale*
d parametro di capacità del soggetto (*j*)
b_i parametro di difficoltà dell'item *i*
a_i parametro di discriminazione dell'item *i*
D fattore di *scaling* che avvicina la funzione logistica a quella normale
n numero di item.

Di seguito vediamo alcuni esempi tipici di *ICC* definiti da tale modello (lateralmente sono riportati i relativi valori dei parametri).



Il confronto tra le diverse *ICC* corrispondenti ai diversi item consente di evidenziare il ruolo dei diversi parametri:

- *parametro di difficoltà*: gli item più difficili (1, 2, 3) sono posizionati sull'estremo più alto della scala di capacità;
- *parametro di discriminazione*: confrontando gli item 1 e 2 (o gli item 1, 3 e 4 con gli item 2, 5 e 6) si evidenziano le diverse "ripidità";
- *parametro di casualità*: confrontando le asintoti appare evidente come per gli item 3, 5 e 6 vi sia, anche se a livelli diversi, la possibilità di rispondere correttamente tirando a indovinare.

4.2.2 Altri modelli

Riassumendo i modelli logistici visti, possiamo dire che i modelli di risposta ad un item di un soggetto con una determinata capacità sono definiti da:

- difficoltà dell'item (modello con un parametro);
- difficoltà e discriminazione dell'item (modello con due parametri);
- difficoltà, discriminazione dell'item e casualità di risposta (modello con tre parametri).

Nell'ambito dell'*Item Response Theory* sono stati sviluppati altri modelli, molti dei quali possono essere applicati ad item non dicotomici. E' il caso del modello con due parametri presentato da Bock nel 1972 (HAMBLETON & altri, 1991), sviluppato per essere applicato nei casi di item a risposta multipla. Esso viene detto <u>modello a risposte nominali</u> (*nominal response model*) in quanto considera le scale di risposta nominali ovvero non prevede alcun ordine a priori nelle modalità di risposta. Lo scopo di tale modello è quello di massimizzare la precisione delle stime di capacità utilizzando tutta l'informazione contenuta nelle risposte del soggetto, non solo nel caso di risposta corretta all'item. Secondo tale modello, la probabilità che un soggetto selezioni una particolare risposta k (su m risposte disponibili) all'item i è data da

$$P_{ik}(d) = \frac{e^{a_{ik}^*\left(d - b_{ik}^*\right)}}{\sum_{h=1}^{m} e^{a_{ih}^*\left(d - b_{ih}^*\right)}} \qquad i = 1,2,\ldots,n;1,2,\ldots,m$$

Per ciascuna d, la somma delle probabilità per le m opzioni è uguale a uno ovvero $\sum P_{ik} = 1$. Le quantità b_{ik}^* e a_{ik}^* rappresentano i parametri di item relativi alla *k-esima* opzione[9].

Esiste un altro modello detto *graded response* presentato da Samejima nel 1969 che può essere applicato ad item con scale di risposte con categorie ordinate. In questo modello, come nel precedente, si cerca di ottenere più informazioni relative alle risposte scorrette. Dato il maggiore utilizzo di scale di risposta ordinali nella ricerca sociale si comprende l'interesse suscitato da tale modello. Supponiamo che le categorie di risposta di un item siano ordinate dal valore più piccolo a quello più grande e che siano indicate con $x_i = 0,1,...,m_i$ dove $m_i + 1$ rappresenta il numero di categorie per l'*i-esimo* item. La probabilità di un soggetto di rispondere ad un item in una particolare categoria o ad una più alta può essere determinata attraverso una estensione del modello logistico con due parametri:

$$P_{xi}^*(d) = \frac{e^{Da_i(d-b_{xi})}}{1+e^{Da_i(d-b_{xi})}}$$

dove

b_{x_i} livello di difficoltà della categoria m_i.

Con $m_i + 1$ categorie, è necessario stimare m_i valori di difficoltà per ciascun item, più un parametro di discriminazione di item. La reale probabilità di un soggetto di ottenere un punteggio di x_i è data dalla seguente espressione

$$P_{xi}(d) = P_{xi}^*(d) - P_{xi+1}^*(d)$$

Per esempio con 50 item e 5 punti di scala per item, dovrebbero essere stimati un totale di (50*4)+50=250 valori di parametri di item!!

4.3 LA VERIFICA DEL MODELLO

La verifica dei modelli probabilistici prevede principalmente due passaggi:

a. STIMA DEI PARAMETRI che caratterizzano il modello prescelto. Essendo tali parametri ignoti, il primo e più importante momento nell'applicazione dei modelli di *IRT* è quello della loro stima. Tale stima, come in altri ambiti della statistica, avviene necessariamente a partire dalle osservazioni sperimentali (risposte dei soggetti). Il momento della stima dei parametri è particolarmente delicato in quanto la validità dell'applicazione dell'*IRT* dipende dalla possibilità di disporre di soddisfacenti procedure di stima; tali parametri sono la capacità e quelli che riguardano gli item; ricordiamo che nei modelli logistici che abbiamo visto questi ultimi sono, a seconda del modello,
    - difficoltà dell'item *i* (b_i),
    - discriminazione dell'item *i* (a_i),
    - casualità di risposta all'item *i* (c_i), corrispondente alla possibilità anche con bassa capacità di rispondere correttamente.

b. VERIFICA DELL'ADATTAMENTO DEL MODELLO AI DATI ovvero della possibilità di questo di prevedere e/o spiegare in modo adeguato i dati.

4.3.1 Stima dei parametri

Come sappiamo la probabilità di dare una risposta corretta è definita dai parametri dell'item e della capacità; tali parametri sono però sconosciuti: l'unico elemento noto è rappresentato dalle risposte

[9] Il simbolo "*" indica "fino a"; quindi $P_{x_i}^*$ indica "*P* fino a x_i".

dati dagli individui. Per tale motivo il primo è più importante momento nell'applicazione dell'*IRT* è quello della stima dei parametri che caratterizzano il modello di risposta degli item prescelti ovvero la capacità per ciascun soggetto e i parametri assunti per ciascun item. Alla possibilità di disporre di soddisfacenti procedure per la stima dei parametri del modello è legato il successo delle successive applicazioni dell'*IRT*.
La stima dei parametri può essere eseguita in diversi modi. La strategia adottata per la stima dei parametri, nel caso di dati campionari, è quella che identifica i valori dei parametri che producono la migliore curva di adattamento. Tale problema è simile a quello che viene affrontato nell'ambito dell'analisi di regressione: i parametri che caratterizzano il modello di regressione (coefficienti di regressione) devono essere stimati a partire dai dati osservati; in tale sede il criterio che consente di definire il migliore adattamento, come sappiamo, è quello dei minimi quadrati.
I modelli di *IRT*, a differenza di quelli di regressione, non consentono l'utilizzazione di tale criterio in quanto non sono lineari e non dispongono dei valori osservabili per la variabile indipendente (*d* non è infatti osservabile direttamente). Vedremo come per tali modelli si utilizza principalmente l'approccio di stima basato sulla *massima verosimiglianza*.[10]

4.3.1.1 Stima della capacità

Riprendendo l'assunto di indipendenza locale, il prodotto delle probabilità di osservare le risposte di ciascun item è uguale a:

$$P(U_1, U_2, ..., U_n|d) = P(U_1|d) * P(U_2|d) * ... * P(U_i|d) * ... * P(U_n|d) = \prod_{i=1}^{n} P(U_i|d)$$

dove

n	numero di item
d	capacità che influenza la risposta del soggetto
U_i	risposta del soggetto all'item *i* (*i*=1,2,...,*n*)
$P(U_i|d)$	probabilità di risposta di un soggetto con capacità *d*

Tenendo conto del fatto che

$P(U_i = 1|d) \rightarrow$ probabilità di risposta corretta

$P(U_i = 0|d) \rightarrow$ probabilità di dare una risposta scorretta

la funzione di probabilità può essere riscritta:

$$P(U_1, U_2, ...U_i, ..., U_n|d) = \prod_{i=1}^{n} P(U_i|d)^{U_i} \left[1 - P(U_i|d)\right]^{1-U_i} \text{ ovvero } \qquad \prod_{i=1}^{n} P_i^{U_i} Q_i^{1-U_i}$$

dove

P_i	$P(U_i|d)$
Q_i	$1 - P(U_i|d)$

Tale equazione rappresenta un'espressione della probabilità congiunta di un modello di risposta.
Quando il modello di risposta non è teorico ma osservato $(U_i = u_i)$, l'interpretazione probabilistica

[10] Il metodo di *stima di massima verosimiglianza* (in inglese *Maximum Likelihood Estimation*, *MLE*) si basa su una serie di approssimazioni successive ai valori incogniti dei veri parametri. A differenza del metodo dei minimi quadrati comuni, che valuta la *bontà di adattamento* ai dati del modello calcolando la somma delle differenze al quadrato tra i valori predetti e quelli osservati, il metodo di *massima verosimiglianza* calcola la probabilità di osservare ciascun possibile valore, assumendo che un dato parametro sia vero. Il parametro che risulta associato alla probabilità più alta costituisce la stima di *massima verosimiglianza*.
A tale metodo non sono associate formule simili a quelle utilizzate per la stima dei minimi quadrati ma algoritmi che consentono di esaminare in modo iterativo più parametri fino a quando non viene identificato quello migliore. Nella prima fase si stabiliscono delle stime iniziali dei parametri; una serie di iterazioni produce in successione nuove stime e le confronta con quelle precedenti; le iterazioni continuano fino a quando le stime ottenute nel ciclo appena concluso differiscono da quelle ottenute nel precedente di una quantità inferiore a uno certo valore stabilito. Le distribuzioni campionarie delle stime di massima verosimiglianza sono note solo nel caso di grandi campioni quando sono normali.

non è più appropriata. L'espressione per la probabilità congiunta è chiamata *funzione di verosimiglianza* (*likelihood*, *L*):

$$L(u_1, u_2, ..., u_i, u_n | d) = \prod_{i=1}^{n} P_i^{u_i} Q_i^{1-u_i}$$

Dato che P_i e Q_i sono funzioni di *d* e dei parametri degli item, anche la funzione di verosimiglianza rappresenta una funzione di tali parametri. L'applicazione di tale funzione può essere migliorata sottoponendola a trasformazione logaritmica[11]; utilizzando le proprietà dei logaritmi la funzione di verosimiglianza diviene funzione *log-verosimiglianza*:

$$\log_n L(u|d) = \sum_{i=1}^{n} [u_i \log_n P_i + (1 - u_i) \log_n (1 - P_i)]$$

dove **u** rappresenta il vettore delle risposte agli item. Il valore massimo di *d* prodotto dalla funzione per un soggetto è definito come *stima di massima verosimiglianza* di *d*.

Il problema dell'individuazione del valore massimo non è secondario. Il valore che massimizza la funzione può essere determinata utilizzando una procedura di ricerca automatica. Le procedure più efficienti sono quelle che, nel punto in cui la funzione raggiunge il massimo, la pendenza della funzione è zero. Così la stima di massima verosimiglianza può essere determinata utilizzando un metodo di approssimazione. L'utilità delle stime di massima verosimiglianza (*maximum likelihood estimes*, *MLE*) sta nel fatto che hanno proprietà asintotiche ben note; l'asintote si riferisce al numero di item all'aumentare del quale, la *MLE* di *d* $(\hat{d})$ è distribuita normalmente con media *d*. Ciò comporta che la distribuzione asintotica di $\hat{d}$ è centrata sul valore di *d*, quindi la *MLE* di $\hat{d}$ non è affetta da errore sistematico con un numero alto di item. E' possibile determinare per $\hat{d}$ l'errore standard e l'intervallo di confidenza. Accanto alla stima di massima verosimiglianza è utilizzabile anche la procedura Bayesiana che consente di risolvere i problemi presentati dall'approccio di massima verosimiglianza. Ricordiamo che le distribuzioni campionarie delle stime di massima verosimiglianza sono note solo nel caso di grandi campioni.

4.3.1.2 Stima dei parametri degli item

Nel descrivere le procedure per stimare *d* abbiamo assunto noti i parametri degli item. D'altra parte anche i parametri degli item devono essere stimati. Come abbiamo visto per stimare la capacità di un soggetto si assumono noti i parametri degli item e si applica la funzione di verosimiglianza alle risposte ottenute. Al contrario se si vogliono stimare i parametri degli item si assumono noti i valori individuali di *d* e si applica la funzione di verosimiglianza alle risposte ottenute:

$$L(u_1, u_2, ..., u_j, ..., u_N | d, a, b, c) = \prod_{j=1}^{N} P_j^{u_j} Q_j^{1-u_j}$$

dove

a, *b*, e *c* parametri degli item (modello con tre parametri)

L'applicazione della funzione di verosimiglianza per la stima dei parametri degli item, a differenza di quella per la stima della capacità, non richiede l'assunto d'indipendenza locale ma quello d'indipendenza delle risposte di *N* soggetti ad un item. Essendo noti i valori *d*, la stima dei parametri degli item è piuttosto semplice ed è confrontabile con la procedura descritta in precedenza; la differenza sta nel fatto che la funzione di verosimiglianza per i parametri degli item è multidimensionale.

4.3.1.3 Stima dei parametri di item e capacità

La situazione più difficile da risolvere, ma anche la più comune, è quella in cui devono essere

[11] Il vantaggio di tale trasformazione è quello di poter utilizzare, come abbiamo visto, le proprietà dei logaritmi.

stimati sia i parametri degli item che la capacità; in questo caso è necessario prendere in considerazione simultaneamente tutte le risposte a tutti gli item di tutti i soggetti.
La funzione di verosimiglianza con *N* soggetti che rispondono a *n* item, assumendo l'indipendenza locale, è

$$L(u_1, u_2, \ldots, u_N | d, a, b, c) = \prod_{i=1}^{n} \prod_{j=1}^{N} P_{ij}^{U} Q_{ij}^{1-U}$$

dove

U u_{ij}

u_j modello di risposta del soggetto *j* agli *n* item

d vettore degli *N* parametri di capacità

a, ***b***, ***c*** vettori dei parametri degli item per gli *n* item

Considerando che:

a. il numero dei parametri degli item è *n* nel modello con un parametro, 2*n* nel modello con due parametri e 3*n* in quello con tre,
b. il numero di parametri di capacità è *N*,

il numero massimo di parametri da stimare è 3*n*+*N*.
Prima di procedere con la stima è necessario affrontare il problema dell'indeterminatezza. Nella funzione di verosimiglianza vista in precedenza i parametri degli item e di capacità non sono determinati in maniera unica.
La funzione per l'item relativa al modello con tre parametri

$$P_i(d) = c_i + (1 - c_i) \frac{e^{Da_i(d-b_i)}}{1 + e^{Da_i(d-b_i)}} \qquad i = 1,2,\ldots,n$$

Se in tale equazione sostituiamo

$d \to d^* = ad + \beta$ $\qquad$ $b \to b^* = ab + \beta$ $\qquad$ $a \to a^* = a/\alpha$

la probabilità di una risposta corretta rimane invariata ovvero

$$P(d) = P(d^*)$$

Siccome i valori di α e β sono costanti e arbitrari, la funzione di verosimiglianza non avrà un massimo unico ovvero sarà indeterminata e quindi non consentirà di cercare il massimo valore di verosimiglianza. Tale problema non esiste nella stima di *d* quando i parametri degli item sono noti o nella situazione parallela in cui i parametri degli item sono stimati disponendo dei parametri di capacità.
Il problema dell'indeterminatezza può essere risolto scegliendo una scala arbitraria per i valori di capacità e per i valori di *b*, per esempio standardizzando gli *N* valori di capacità e gli *n* valori di difficoltà. Ciò consente di confrontare le stime dei parametri degli item effettuate su gruppi diversi.
Una volta eliminata l'indeterminatezza, è possibile determinare i valori dei parametri di capacità e degli item che massimizzano la funzione di verosimiglianza.
Una delle procedure che consente di fare ciò è la *stima congiunta della massima verosimiglianza* (*joint maximum likelihood estimation*) che viene eseguita in due fasi (*stage*):

1ª fase	○	scelta dei valori iniziali per il parametro di capacità per ciascun soggetto: *log_n (numero di risposte corrette / numero di risposte scorrette)*
	○	standardizzazione di tali valori per eliminare l'indeterminatezza;
	○	stima dei parametri dell'item, considerando noti i valori di capacità;
2ª fase	○	stima dei parametri di capacità, considerando noti i valori di dei parametri dell'item.

Tale procedura viene ripetuta secondo vari passaggi (*step*) fino a quando i valori delle stime non cambiano tra due successivi passaggi.
La procedura congiunta di massima verosimiglianza anche se concettualmente convincente presenta degli svantaggi:

a. non consente stime per capacità perfette o con punteggio zero;
b. non consente stime dei parametri per quegli item ai quali tutti i soggetti hanno risposto correttamente (o scorrettamente);

c. non produce stime consistenti[12] dei parametri di item e capacità per il modello con tre parametri.

E' possibile superare alcuni di tali problemi utilizzando un approccio alternativo che produce stime Bayesiane ottenute utilizzando distribuzioni a priori (Swaminathan, 1982, 1985, 1986).

Il problema dell'inconsistenza delle stime congiunte di massima verosimiglianza è dovuta alla simultanea stima dei parametri degli item e di capacità. Tale problema però scompare se i parametri degli item possono essere stimati senza fare alcun riferimento ai parametri di capacità ma a specifiche distribuzioni dei parametri di capacità, considerando i soggetti estratti causalmente da una popolazione. Ne risultano *stime di massima verosimiglianza marginale* (*marginal maximum likelihood estimes*) che presenta positive proprietà asintotiche: le stime dei parametri degli item sono consistenti ovvero si avvicinano al valore del parametro all'aumentare del numero di soggetti.

Per poter ottenere la funzione di verosimiglianza marginale dei parametri degli item, è necessario approssimare la distribuzione di capacità. Per una buona approssimazione della distribuzione di capacità è importante disporre di un grande numero di soggetti ovvero la procedura di massima verosimiglianza marginale dovrebbe essere applicata solo in presenza di un campione ampio di soggetti.

Una volta stimati i parametri degli item con tale procedura, le stime dei parametri degli item possono essere utilizzate per stimare le capacità utilizzando il metodo visto in precedenza.

Oltre alle procedure di stima *di massima verosimiglianza*, *di massima verosimiglianza marginale* e *Bayesiana* esistono altre procedure, così riassumibili (Swaminathan, 1982, 1985, 1986):

- Procedura di stima *congiunta di massima verosimiglianza* (Lord 1974, 1980), applicabile ai modelli con uno, due e tre parametri; i parametri di capacità e degli item sono stimati simultaneamente.
- Procedura di stima di *massima verosimiglianza marginale* (Bock & Aitkin 1981), applicabile ai modelli con uno, due e tre parametri; la stima della capacità avviene successivamente a quella degli altri parametri.
- Procedura di stima di *massima verosimiglianza condizionale* (Andersen 1972, 1973, Rasch 1960), applicabile solamente al modello ad un parametro; la funzione di verosimiglianza è condizionata sul numero di punteggi corretti.
- Procedure di stima *Bayesiana congiunta e marginale* (Mislevy 1984, Swaminathan 1982, 1985, 1986), applicabile ai modelli ad uno, due e tre parametri; si definiscono distribuzioni a priori attribuite ai parametri degli item e di capacità, eliminando così alcuni dei problemi, come l'impropria stima dei parametri e la mancanza di convergenza, che si hanno con le procedure di massima verosimiglianza marginale e congiunta.
- Procedura di *stima euristica* (Urry 1974, 1978), applicabile principalmente ai modelli con due e tre parametri.
- Procedure basate su *analisi fattoriale non-lineare* (McDonald 1967, 1989), applicabile al modello con due parametri e ad un modello modificato con tre parametri in cui i valori *c* sono fissi.

4.3.1.4 Stima dei parametri per un modello semplice

Come abbiamo visto per il modello logistico con un parametro (modello *Rasch*), la risposta corretta è più probabile quando la capacità del soggetto supera la difficoltà dell'item. Tale modello assume che la probabilità di un soggetto di rispondere correttamente ad un item sia dovuto a due parametri:

- *capacità* del soggetto (d_j, posizione del soggetto *j* sul continuum),
- *difficoltà* dell'item (b_i, posizione dell'item *i* sul continuum).

Entrambi i parametri possono essere misurati dall'item; vediamo come:

➢ capacità del soggetto: *rapporto tra numero di risposte corrette* (x_j) *e il numero di risposte errate* $(n - x_j)$ *fornite dal soggetto* j *a tutti gli item*

$$d_j = \frac{x_j}{n - x_j}$$

dove

[12] Ricordiamo che una stima è detta "consistente" quando all'aumentare della dimensione del campione aumenta la probabilità di avvicinarsi al valore corrispondente della popolazione.

x_j numero di risposte corrette fornite dal soggetto *j* a tutti gli item

$n - x_j$ numero di risposte errate del soggetto *j*

n numero di item.

- <u>difficoltà dell'item</u>: *rapporto tra numero di risposte errate* $(N - y_i)$ *e numero di risposte corrette* (y_i) *fornite all'item* i *da tutti i soggetti del campione*

$$b_i = \frac{N - y_i}{y_i}$$

dove

y_i numero di risposte corrette fornite all'item *i* da tutti i soggetti del campione

$(N - y_i)$ numero di risposte errate all'item *i*

N numero di soggetti del campione.

Tali calcoli vengono effettuati dopo aver escluso dall'analisi i soggetti che hanno ottenuto i punteggi massimi e minimi e gli item che hanno registrato risposte costanti.

Disponendo della misura della capacità dei soggetti e della difficoltà degli item, la probabilità di risposta può essere definita come la differenza tra i due valori; quindi se la capacità del soggetto *j* (d_j) è maggiore della difficoltà dell'item *i* (b_i) allora la probabilità di rispondere correttamente all'item è maggiore di 0.5, ovvero

$$\text{se } (d_j - b_i) > 0 \text{ allora } P_i(d) > 0.5$$

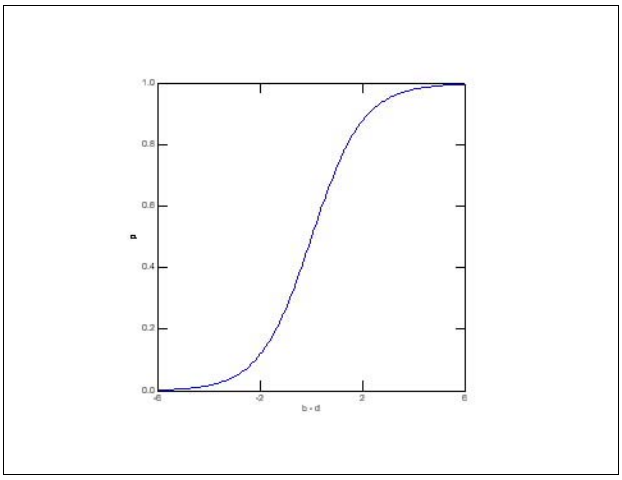
da cui

$$\text{se } (d_j - b_i) < 0 \text{ allora } P_i(d) < 0.5 \qquad \text{se } (d_j - b_i) = 0 \text{ allora } P_i(d) = 0.5$$

Ciò corrisponde al modello con un parametro definito dalla seguente equazione matematica

$$P_i(d) = \frac{e^{(d_j - b_i)}}{1 + e^{(d_j - b_i)}}$$

cui corrisponde la seguente curva:



Tali osservazioni sono sostenibili quando sia la *capacità* che la *difficoltà* sono misurate e sono esprimibili con la stessa unità di misura. Per poter ottenere una scala di misura con intervalli uguali la precedente equazione viene sottoposta a trasformazione logaritmica:

$$P_i(d) = \frac{e^{(\delta_j - \beta_i)}}{1 + e^{(\delta_j - \beta_i)}}$$

dove

δ_j $\log_n(d_j) \rightarrow$ *logit correct*

β_i $\log_n(b_i) \rightarrow$ *logit incorrect*

che corrispondono alle definizioni iniziali di capacità e di difficoltà. Quindi il *logit*

- della capacità del soggetto è il logaritmo naturale del rapporto tra il numero di risposte corrette e il numero di risposte errate fornite dal soggetto *j* per l'item che si trova all'origine

della scala ovvero che ha difficoltà zero;
- della difficoltà degli item è il logaritmo naturale del rapporto tra il numero di risposte errate e il numero di risposte corrette ottenute dall'item *i* per il soggetto che si trova all'origine della scala ovvero che ha capacità zero.

Le due quantità (d_j e b_i)[13] così calcolate e trasformate vengono riferite ad un'unica scala lineare che va da $-\infty$-a $+\infty$. Nella pratica i valori di difficoltà e quelli di capacità vanno da -4 a +4 *logit*:
- i valori negativi di capacità indicano soggetti con basse prestazioni,
- i valori negativi di difficoltà indicano item facili.

Per ciascuna stima è possibile calcolare i corrispondenti errori standard:
- Errore standard della capacità del soggetto:

$$E(d_j) = X\left[\frac{n}{x_j}(n - x_j)\right]^{1/2}$$

dove

x_j numero di risposte corrette fornite dal soggetto *j* a tutti gli item

$n - x_j$ numero di risposte errate del soggetto *j*

n numero di item.

- Errore standard della difficoltà dell'item:

$$E(b_i) = Y\left[\frac{N}{y_i}(N - y_i)\right]^{1/2}$$

dove

y_i numero di risposte corrette fornite all'item *i* da tutti i soggetti del campione

$N - y_i$ numero di risposte errate all'item *i*

---

[13] Sono stati definiti ed elaborati altri algoritmi di stima della capacità e della difficoltà:
- Capacità del soggetto:

$$d_j = X(\delta_j) \quad \rightarrow \quad d_j = X \log_n\left(\frac{x_j}{n - x_j}\right)$$

dove

n numero di item

X fattore di espansione di capacità, ovvero

$$\left[\frac{(1 + U/1.7^2)}{(1 - UV/1.7^4)}\right]^{1/2}$$

U varianza degli *n* valori β_i

V varianza degli *N* valori δ_j.

- Difficoltà dell'item:

$$b_i = Y\left(\beta_i - \frac{\sum \beta_i}{n}\right) \quad \rightarrow \quad b_i = Y\left[\log_n\left(\frac{N - y_i}{y_i}\right) - \frac{\sum \log_n\left(\frac{N - y_i}{y_i}\right)}{n}\right]$$

dove

N numero di soggetti del campione

Y fattore di espansione di difficoltà, ovvero

$$\left[\frac{(1 + V/1.7^2)}{(1 - UV/1.7^4)}\right]^{1/2}$$

U varianza degli *n* valori β_i

V varianza degli *N* valori δ_j

N numero di soggetti del campione.

Ciascun errore standard può essere interpretato come *indice di non-affidabilità*; il suo valore è maggiore per i valori che nella scala di misurazione stanno agli estremi in quanto in questi casi è minore l'informazione fornita. L'errore standard della misura della difficoltà $(E(b_i))$ dipende dalle capacità dei soggetti. Il valore soglia di $E(b_i)$ è 0.25; gli item che superano tale valore sono da considerare sospetti.

Se nel modello è stato compreso anche il parametro di discriminazione (la cui procedura di stima non è qui presentata) occorre tenere presente che gli item con valori di discriminazione negativi sono scartati dagli strumenti che misurano capacità in quanto indicano che all'aumentare della capacità del soggetto diminuisce la probabilità di rispondere correttamente all'item, situazione non accettabile.

Se i soggetti in un campione hanno una capacità che si situa all'inizio della scala, ovvero hanno una probabilità di successo di 0.5, la precisione della misura è massima.

Come per tutti i modelli di *IRT*, le stime di difficoltà degli item che si ottengono dopo la procedura di controllo dell'adattamento dei risultati del modello e della successiva eliminazione di soggetti e di item incoerenti, sono praticamente indipendenti dai soggetti particolari; ciò vuol dire che lo strumento può essere tarato anche con una sola applicazione.

Riassumiamo nel seguente schema quanto visto finora.

STIMA DEI PARAMETRI NELL'APPROCCIO LOGISTICO CON UN PARAMETRO

⇓

Stima dei valori dei parametri che producono la migliore curva che descrive la probabilità di un soggetto di rispondere correttamente ad un item:

⇓

Si escludono dall'analisi:

- soggetti che hanno ottenuto i punteggi massimi e minimi
- item che hanno registrato risposte costanti

1. capacità del soggetto (d_j, posizione del soggetto *j* sul continuum): rapporto tra:
   - *numero di risposte corrette* (x_j)
   - *numero di risposte errate* $(n - x_j)$

   fornite dal caso *j* a tutti gli item

$$d_j = \frac{x_j}{n - x_j}$$

2. difficoltà dell'item (b_i, posizione dell'item *i* sul continuum): rapporto tra:
   - numero di risposte errate $(N - y_i)$
   - numero di risposte corrette (y_i)

   fornite dall'item *i* da tutti i casi

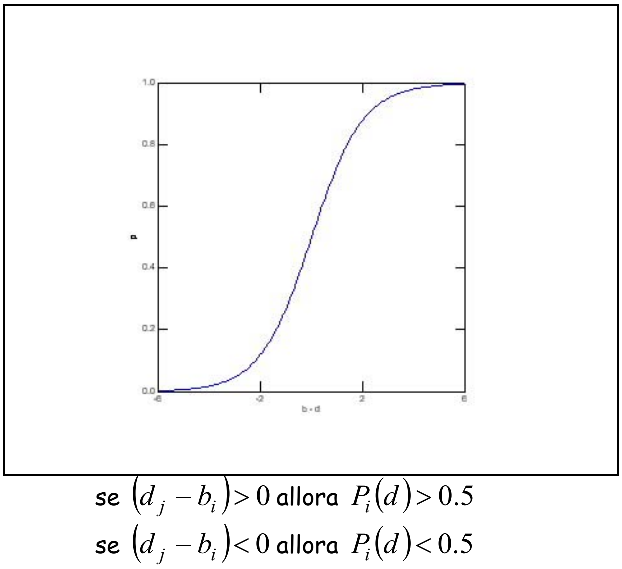
$$b_i = \frac{N - y_i}{y_i}$$

quanto maggiore è la differenza tra *capacità-soggetto* e *difficoltà-item* tanto maggiore è lo scarto quadrato che indica l'inaccettabilità di una risposta sbagliata (con $d_j > b_i$) o una risposta corretta (con $d_j < b_i$)

3. probabilità di risposta:

(capacità del soggetto *j*) – (difficoltà dell'item *i*)

Modello $\rightarrow$ $P_i(d) = \dfrac{e^{(d_j - b_i)}}{1 + e^{(d_j - b_i)}}$



se $(d_j - b_i) > 0$ allora $P_i(d) > 0.5$

se $(d_j - b_i) < 0$ allora $P_i(d) < 0.5$

se $(d_j - b_i) = 0$ allora $P_i(d) = 0.5$

4. <u>trasformazione di capacità e difficoltà</u> su stessa unità di misura:

unica scala lineare per d_j e b_i $\rightarrow$ trasformazione logaritmica

$\downarrow$

LOGIT

- della capacità del soggetto = *logit correct* = $\log_n(d_j) = \delta_j$
- della difficoltà dell'item = *logit incorrect* = $\log_n(b_i) = \beta_i$

$\downarrow$

Valori in teoria: da $-\infty$ a $+\infty$

in pratica: da –4 a +4

- i valori negativi di capacità indicano soggetti con basse prestazioni
- i valori negativi di difficoltà indicano item facili

$\downarrow$

Modello $\rightarrow$ $P_i(d) = \dfrac{e^{(\delta_j - \beta_i)}}{1 + e^{(\delta_j - \beta_i)}}$

5. <u>errori standard</u> $\rightarrow$ indici di *non-affidabilità*

a. <u>errore standard della capacità del soggetto</u>:

$$E(d_j) = X\left[\frac{n}{x_j}(n - x_j)\right]^{1/2}$$

b. <u>errore standard della difficoltà dell'item</u>:

$$E(b_i) = Y\left[\frac{N}{y_i}(N - y_i)\right]^{1/2}$$

$E(b_i)$:

- dipende dalle capacità dei soggetti
- valore soglia = 0.25
- se $E(b_i) > 0.25$ $\rightarrow$ item sospetto

dove

n	numero di item
N	numero di soggetti
x_j	numero di risposte corrette fornite dal soggetto *j* a tutti gli item
Y_i	numero di risposte corrette fornite all'item *i* da tutti i soggetti
$n - x_j$	numero di risposte errate del soggetto *j*
$N - Y_i$	numero di risposte errate all'item *i*
X	fattore di espansione di capacità
Y	fattore di espansione di difficoltà

4.3.2 Adattamento del modello

Perché i modelli di *IRT* possano essere considerati validi, è necessario verificare il loro livello di adattamento ai dati. L'adattamento di un modello *IRT* ai dati indica che sono state raggiunte le caratteristiche desiderabili. Un modesto e debole adattamento del modello non produce parametri invarianti di item e di capacità. Quindi la riuscita delle specifiche applicazione di *IRT* non è legata solo alla possibilità di disporre di stime consistenti dei parametri ma soprattutto dalla possibilità di

valutare il livello di adattamento[14] tra dati e modello.
Il modello può essere considerato appropriato, per un particolare gruppo di dati, quando è in grado di prevedere o spiegare in modo adeguato i dati stessi. Alcuni autori raccomandano di valutare e verificare l'adattamento del modello sulla base di tre tipi di evidenza:

1. verifica della validità degli assunti del modello, utile nel selezionare modelli di *IRT* da utilizzare nella verifica del secondo e terzo tipo di evidenza;
2. verifica dell'invarianza dei parametri (item e capacità), essenziale in quanto tutte le applicazioni di *IRT* dipendono da tale caratteristica;
3. verifica dell'accuratezza delle previsioni del modello nell'utilizzare i dati reali o simulati; ciò richiede una valutazione del livello di spiegazione dei risultati da parte del modello *IRT* e della possibilità di comprendere la natura delle discrepanze tra modello e dati e delle conseguenze.

Nella scelta del modello più appropriato può essere utile verificare il livello di adattamento di più modelli e di confrontare i risultati con quelli ottenuti con dati simulati.

4.3.2.1 Verifica degli assunti

L'analisi dei principali assunti che stanno alla base dei modelli di *Item Response* consente di selezionare il modello migliore. Tale analisi in genere riguarda:

- l'*unidimensionalità* e l'*indipendenza locale* (per tutti i modelli),
- la *discriminazione* (modello con due e tre parametri),
- la *possibilità di rispondere a caso* (modello con tre parametri).

Di seguito vediamo alcuni tra i metodi di verifica utilizzabili tra i quali si noteranno alcuni utilizzati anche dal modello classico di *item analysis* e considerati validi anche in questo caso.

➢ **Unidimensionalità**

Sono stati individuati moltissimi indici di valutazione dell'unidimensionalità (c'è chi ne ha contati ben 88!!); tra i pochi metodi giudicati più validi e soddisfacenti vi sono quelli basati sull'analisi fattoriale e l'analisi dei residui; di seguito vediamo alcuni tra i più interessanti.

- *Plot* degli *eigenvalue* (dai più grandi ai più piccoli) dell'analisi fattoriale applicata alla matrice di correlazione tra gli item per verificare se è presente un primo fattore determinante.
- Confronto dei *plot* degli *eigenvalue* di due matrici di correlazione tra item: una con dati reali e l'altra con dati casuali (con distribuzione normale, stessa dimensione campionaria e stesso numero di variabili dei dati osservati). Se nei dati osservati l'assunto di unidimensionalità è soddisfatto, i due grafici dovrebbero essere sostanzialmente simili.
- Verifica dell'assunto di indipendenza locale investigando le matrici di covarianza o di correlazione calcolate sui soggetti raggruppati secondo diversi valori di capacità. L'assunto di unidimensionalità sarà soddisfatto (anche approssimativamente) se i valori al di fuori della diagonale sono piccoli e vicini a zero.
- Adattamento al modello di analisi mono-fattoriale non-lineare della matrice di correlazione tra gli item e analisi dei residui.
- Uso di un metodo di analisi fattoriale basato direttamente sull'*IRT*. Per spiegare il vettore delle risposte si assume una versione multidimensionale del modello con tre parametri basato sull'ogiva normale. La stima dei parametri del modello è complicata e richiede molto tempo ma i risultati ottenuti possono essere promettenti. Di particolare interesse è l'adattamento di una soluzione unidimensionale ai dati.
- Verifica della presenza di particolari item che violano l'assunto per vedere se "funzionano" in maniera diversa. I valori *b* per tali item sono calibrati separatamente in sotto-gruppi e nel gruppo totale di item. Il contesto della taratura dell'item diviene trascurabile se gli assunti del modello vengono soddisfatti. Se il grafico dei valori di *b* tarati nei due contesti è lineare con una dispersione confrontabile con gli errori standard associati con le stime dei parametri dell'item, l'assunto di unidimensionalità è soddisfatto.

➢ **Indici di Equi-Discriminazione**

- Analisi delle correlazioni (biseriali e punto-biseriali) tra item e punteggi totali secondo il modello classico di *item analysis*. Quando le correlazioni sono ragionevolmente omogenee, è possibile procedere alla selezione di

[14] L'adattamento dei risultati al modello è detto *fit* mentre il mancato adattamento è detto *misfit*.

un modello che assume item equidiscriminanti.

- **Minima Possibilità di Rispondere A Caso**
  - Verifica delle risposte dei soggetti con bassa capacità agli item più difficili; se i livelli di esecuzione sono vicini a zero, l'assunto è soddisfatto.
  - Verifica attraverso il *plot* delle regressioni dei punteggi item-strumento. L'assunto sarà soddisfatto quando gli item mostreranno *performance* vicino a zero per soggetti con punteggi bassi.
  - Analisi del livello di difficoltà dello strumento, dei limiti di tempo e della struttura degli item per verificare l'eventuale presenza della possibilità di rispondere a caso.
- **Altre particolari verifiche**
  - Rapporto tra la varianza del gruppo di item omessi e la varianza del gruppo di item con risposte scorrette; l'assunto è soddisfatto quando il rapporto è vicino a zero.
  - Confronto tra i punteggi dei soggetti cui si sono dati limiti di tempo specificati e quelli dei soggetti senza limiti di tempo. Grosse sovrapposizioni nell'esecuzione indicano che l'assunto è soddisfatto.
  - Verifica della percentuale di soggetti che completano l'esecuzione, la percentuale di soggetti che completano il 75% e il numero di item completati dall'80% dei soggetti. Se quasi tutti i soggetti completano quasi tutti gli item, la velocità di esecuzione non rappresenta un fattore importante.

4.3.2.2 Verifica dell'invarianza

I metodi di verifica dell'invarianza dei parametri del modello sono molti e, in ogni caso, piuttosto semplici. L'invarianza può essere studiata somministrando ai soggetti due o più gruppi di item; gli item di ciascun gruppo variano molto in difficoltà relativamente alla dimensione misurata. Per ciascun soggetto e per ogni gruppo di item vengono determinate le stime di capacità. Le due serie di stime di capacità per tutti i soggetti vengono messe a confronto attraverso un grafico. Se l'assunto che il punteggio atteso di capacità di ciascun soggetto è indipendente dagli item scelti è soddisfatto, tale grafico dovrebbe definire una retta con pendenza 1. L'errore di misurazione può produrre qualche dispersione dei punti intorno a tale retta. Il modello di *IRT* può non essere considerato adatto ai dati a disposizione quando non è possibile osservare una relazione lineare (con pendenza=1 e intercetta=0) o la dispersione eccede quella attesa dalla conoscenza degli errori standard delle stime di capacità.

1. **Invarianza delle stime dei parametri di capacità**. Confronto delle stime di capacità per diversi campioni di item (per identificare item difficili e facili, item che riflettono diverse categorie di contenuto all'interno del gruppo di item definiti). L'invarianza è soddisfatta quando le stime non differiscono molto tra loro.
2. **Invarianza delle stime dei parametri di item**. Confronto tra i modelli di stima di parametri degli item (valori *b*, valori *a* e/o valori *c*) ottenuti in due o più sottogruppi della popolazione cui è rivolto lo strumento. Quando le stime sono invarianti, il *plot* dovrebbe essere lineare con scarti che riflettono solo errori di campionamento.

4.3.2.3 Verifica delle previsioni del modello

La successiva verifica riguarda l'adattamento (*fit*) dei risultati al modello; tale verifica si basa su alcune considerazioni: se i valori di capacità dei soggetti e i valori di difficoltà degli item si estendono approssimativamente sullo stesso intervallo (sufficientemente ampio) della variabile da misurare allora sarà possibile tracciare una curva simile a quella rappresentata dall'*ICC*. Lo scostamento dei punti osservati da questa curva consente di determinare il grado di incoerenza rispetto al modello. La procedura seguita per tale verifica è iterativa con affinamenti successivi che depurano i dati da quelli incoerenti ovvero da quelli che, non adattandosi al modello, presentano un alto valore di *misfit*.

La valutazione dell'adattamento viene fatta attraverso la verifica delle previsioni del modello applicato. Di seguito vediamo quali sono le tecniche più utilizzate.

- Analisi dei residui e dei residui standardizzati dell'adattamento del modello ai dati.
- Confronti delle distribuzioni dei punteggi osservati e predetti ottenute assumendo corretti tutte

le stime dei parametri del modello. L'analisi dei risultati è fatto con metodi grafici o applicando particolari statistiche come il *chi-quadro*.

- Analisi degli effetti di particolari condizioni sperimentali come la posizione degli item, la velocità, noia, istruzioni, ecc.
- Diagramma di dispersione delle stime di capacità e dei corrispondenti punteggi. L'adattamento è accettabile quando la relazione si presenta forte con una dispersione intorno alla *ICC* che riflette l'errore di misurazione.
- Applicazione di una miriade di test statistici per determinare l'adattamento totale del modello, dell'item e individuale.
- Confronti dei parametri veri e stimati degli item e delle capacità con l'ausilio di metodi di simulazione computerizzati.
- Analisi della "robustezza" del modello attraverso l'uso di metodi di simulazione; per esempio è possibile studiare le implicazioni dell'adattamento dei modelli *IRT* unidimensionali a dati multidimensionali.

Tra i diversi approcci il più utilizzato è sicuramento il metodo che verifica le previsioni del modello attraverso l'<u>analisi dei residui </u>(detti anche scarti/errori e a volte "residui *raw*") degli item; vediamo come lo studio dei residui e/o dei residui standardizzati possa fornire preziose informazioni nella scelta del modello di *IRT*; dopo aver:

- scelto il modello di *IRT*,
- stimato i parametri di item e di capacità,
- calcolato le previsioni riguardanti le esecuzioni di vari gruppi di capacità, assumendo la validità del modello scelto,

si confrontano i risultati previsti con quelli osservati. Un residuo r_{ij} rappresenta la differenza tra l'esecuzione osservata e l'esecuzione attesa dell'item per un determinato sottogruppo di soggetti:

$$r_{ij} = x_{ij} - P_{ij}$$

dove

i	item
j	categoria dell'item (sottogruppo)
x_{ij}	proporzione osservata di risposte corrette per l'item i e la *j-esima* categoria di capacità
P_{ij}	proporzione attesa di risposte corrette ottenuta utilizzando il modello *IRT* adottato

Se si fa riferimento ad un unico soggetto allora:

x_{ij}	risposta data dal soggetto j all'item i
P_{ij}	corrispondente probabilità di risposta come prevista dal modello.

Per determinare la *proporzione attesa di risposte corrette* (P_{ij}) in ciascuna categoria di capacità si utilizzano le stime dei parametri del modello ipotizzato. In particolare si determina il valore *d* utilizzando il punto centrale della categoria di capacità come valore di capacità rappresentativo per la categoria; si calcola quindi la probabilità di una risposta corretta utilizzando tale valore.
Per identificare le *categorie di capacità*, il continuum corrispondente viene di solito suddiviso in intervalli di uguale dimensione (da 10 a 15). Gli intervalli non dovrebbero essere troppo piccoli per evitare che le statistiche risentano dell'instabilità soprattutto in presenza di piccoli campioni. D'altra parte, gli intervalli dovrebbero essere definiti in modo tale che i soggetti in ciascuna categoria risultino omogenei in termini di capacità.
L'uso degli scarti è limitato dal fatto che non tiene conto dell'errore campionario associato con il punteggio atteso all'interno di ciascuna categoria di capacità. Per tenere conto di tale errore campionario viene calcolato il *residuo standardizzato* (o *scarto standard*, z_{ij}) ottenuto dividendo i residui per i loro errori standard.
In pratica lo *scarto standard* serve per definire il calcolo del coefficiente di *misfit* (t):

$$z_{ij} = \frac{x_{ij} - P_{ij}}{\sqrt{\dfrac{P_{ij}(1-P)}{N_j}}}$$

dove

N_j	numero di soggetti nella categoria di capacità *j*
$P_{ij}(1-P_{ij})$	varianza.

Quanto più i dati si approssimano al modello tanto più lo scarto standard sarà distribuito normalmente, con media 0 e varianza 1. Quindi per analizzare l'adattamento è possibile esaminare quanto tali scarti si approssimano ad una distribuzione normale oppure quanto i loro quadrati z_{ij}^2 si approssimano ad una distribuzione di *chi-quadro* con un grado di libertà[15].

Quanto maggiore è la differenza tra la capacità del soggetto e la difficoltà dell'item, tanto maggiore risulta lo scarto quadrato che indica l'inaccettabilità di una risposta sbagliata (con $d_j > b_i$) o di una risposta corretta (con $d_j < b_i$).

Di solito per verificare l'adattamento del modello vengono applicati anche test statistici, in genere il Q_1, statistica derivata dal *chi-quadro*. Il Q_1 per l'item *i*

$$Q_{1i} = \sum_{j=1}^{m} \frac{N_j\left(x_{ij} - P_{ij}\right)^2}{P_{ij}\left(1-P_{ij}\right)} \quad = \quad \sum_{j=1}^{m} z_{ij}^2$$

La statistica Q_1 è distribuita come il *chi-quadro* con $m-k$ gradi di libertà, dove *k* è il numero di parametri presenti nel modello *IRT* ed *m* è il numero delle categorie in cui è stato suddiviso il continuum. Se il valore osservato della statistica supera il valore critico (presentato nella tavola del *chi-quadro*), l'ipotesi nulla che la *ICC* si adatti ai dati viene rifiutata e deve essere trovato un nuovo modello di adattamento. E' possibile inoltre determinare:

- il grado di *misfit* di un soggetto sommando i quadrati degli scarti del soggetto *j* per tutti gli item *i* $\left(z_j^2\right)$,
- il grado di *misfit* di un item sommando i quadrati degli scarti dell'item *i* per tutti i soggetti *j* $\left(z_i^2\right)$.

Le medie normalizzate delle quantità z_j^2 e z_i^2 sono i due coefficienti *misfit* t_j e t_i.

Entrambi i coefficienti possono essere calcolati come di seguito:

- <u>*misfit* della capacità del soggetto</u>: $t_j = \left[\log_n\left(v_j\right) + v_j - 1\right] * \left[(n-1)/8\right]^{1/2}$

  dove

n	numero di item
v_j	$z_j^2/(n-1)$

- <u>*misfit* della difficoltà dell'item</u>: $t_i = \left[\log_n\left(v_i\right) + v_i - 1\right] * \left[(N-1)/8\right]^{1/2}$

  dove

N	numero di soggetti
v_i	$z_i^2/(N-1)$.

Non sempre nelle applicazioni si studia e si valuta in modo adeguato l'adattamento modello-dati e le conseguenze dei *misfit*. A causa delle poche applicazioni, non si conosce molto sul livello di adeguatezza dei modelli di *IRT*. Le poche applicazioni spesso utilizzano statistiche non appropriate allo studio della bontà di adattamento; conseguentemente le conclusioni risultano poco affidabili. In genere tali applicazioni hanno utilizzato il modello della *verifica dell'ipotesi* i cui test sono, come si sa, troppo sensibili alla dimensione del campione dei soggetti: con campioni di grosse dimensioni, è molto più probabile giungere alla conclusione di mancanza di adattamento modello-dati anche nei

[15] Nel caso in cui le risposte siano binarie è possibile ottenere per z_{ij}^2 una semplificazione dell'espressione: $z_{ij}^2 = \exp\left|d_j - b_i\right|$

casi di piccolo allontanamento dal modello. D'altra parte l'utilizzo di piccoli campioni non consente una significativa e utile stima di parametri a causa della presenza di grossi errori standard. A ciò va aggiunto il fatto che le distribuzioni campionarie di alcune statistiche di bontà di adattamento utilizzate non sono sempre note con precisione.
Vediamo di seguito la sintesi del procedimento descritto di verifica delle previsioni del modello.

VALUTAZIONE E VERIFICA DELL'ADATTAMENTO DEL MODELLO

⇓

VERIFICA DELL'ACCURATEZZA DELLE PREVISIONI DEL MODELLO APPLICATO

⇓

VALUTAZIONE DEL MISFIT

⇓

procedura iterativa:

1. <u>Analisi dei residui</u> (scarti/errori o residui raw) degli item

   Dopo aver:
   - scelto il modello di IRT,
   - stimato i parametri di item e di capacità,

   si confrontano i risultati previsti con quelli osservati.

   Residuo = (esecuzione osservata) – (esecuzione attesa)

   ↓

   $$r_{ij} = x_{ij} - P_{ij}$$

   Quanto più gli scarti approssimano la distribuzione normale tanto più i dati approssimano il modello

2. <u>Residuo standardizzato</u>

   scarto standard = residuo / errore standard

   $$z_{ij} = \frac{x_{ij} - P_{ij}}{\left[P_{ij}(1-P)\right]^{1/2}} \qquad \text{oppure} \qquad z_{ij} = \frac{x_{ij} - P_{ij}}{\sqrt{\dfrac{P_{ij}(1-P)}{N_j}}}$$

3. <u>Coefficiente di *misfit*</u>
   - <u>*Misfit* della capacità del soggetto</u>:

     z_j^2 = somma dei quadrati degli scarti del soggetto *j* per tutti gli item

     media normalizzata di z_j^2 = $t_j = \left[\log_n(v_j) + v_j - 1\right] * \left[(n-1)/8\right]^{1/2}$
   - <u>*Misfit* della difficoltà dell'item</u>:

     z_i^2 = somma dei quadrati degli scarti dell'item *i* per tutti i soggetti

     media normalizzata di z_i^2 = $t_i = \left[\log_n(v_i) + v_i - 1\right] * \left[(N-1)/8\right]^{1/2}$

4. <u>Test di verifica dell'adattamento del modello</u>

   Per l'item *i* $Q_{1i} = \sum_{j=1}^{m} \frac{N_j\left(x_{ij} - P_{ij}\right)^2}{P_{ij}\left(1 - P_{ij}\right)} \quad = \quad \sum_{j=1}^{m} z_{ij}^2$

   Distribuita come il chi-quadro con m-k gradi di libertà

   Se Q_{1i} > valore critico → adattamento mancato

dove

n	numero di item
N	numero di soggetti
N_j	numero di soggetti nella categoria di capacità *j*
P_{ij}	probabilità di risposta del soggetto *j* all'item *i* <u>oppure</u> proporzione attesa (secondo il modello) di risposte corrette per l'item i nella categoria di capacità *j*
$P_{ij}(1-P_{ij})$	varianza
k	numero di parametri presenti nel modello IRT
v_j	$z_j^2/(n-1)$
v_i	$z_i^2/(n-1)$
m	numero di categorie in cui è stato suddiviso il continuum
x_{ij}	risposta data dal soggetto j all'item i oppure proporzione osservata di risposte corrette per l'item *i* nella categoria di capacità *j*

4.3.2.4 Funzioni informative

Secondo il modello probabilistico, ciascun item può essere caratterizzato e definito oltre che dai parametri visti, anche da una particolare funzione detta *funzione informativa* (*item information function*):

$$I_i(d) = \frac{[P'_i(d)]^2}{P_i(d)Q_i(d)} \qquad i = 1, 2, ..., n$$

dove

n numero totale di item

d capacità che influenza la risposta del soggetto ad uno strumento

i item

$I_i(d)$ informazione fornita dall'item i alla capacità d

$P_i(d)$ *IRF*, *item response function*

$P'_i(d)$ derivata di $P_i(d)$ rispetto a d

$Q_i(d)$ $1 - P_i(d)$

Tale equazione è applicabile ai modelli logistici (con uno, due o tre parametri). Nel caso del modello logistico con tre parametri essa è modificata nel modo seguente:

$$I_i(d) = \frac{2.89 \cdot a_i^2(1-c_i)}{\left(c_i + e^{1.7a_i(d-b_i)}\right)\cdot\left(1+e^{-1.7a_i(d-b_i)}\right)^2}$$

dove a, b, c, d sono naturalmente, rispettivamente, i parametri di discriminazione, difficoltà e del livello *pseudo-casuale* dell'item e di capacità del soggetto.
Da tale equazione è possibile dedurre il ruolo che i parametri b, a e c hanno nella funzione informativa dell'item:

- l'informazione è maggiore quando il valore b è vicino a d di quando il valore b è lontano da d,
- l'informazione è generalmente maggiore quando il valore del parametro a è alto,
- l'informazione aumenta all'avvicinarsi a zero del parametro c.

Le funzioni informative dell'item, consentendo una valutazione dei singoli item, possono giocare un ruolo importante nello sviluppo di uno strumento con misure multiple, in quanto indicano il contributo che ogni item dà alla stima della capacità in punti diversi lungo il continuum di capacità. Tale contributo dipende in gran parte dalla forza di discriminazione di un item (pendenza di P_i); la posizione in cui tale contributo sarà realizzato è dipendente dalla difficoltà dell'item.
Dato che, in genere, il valore delle funzioni informative è minore quando c>0 di quando c=0, si potrebbe essere tentati di adottare i modelli con uno o due parametri. I valori delle funzioni informative che ne risultano saranno più grandi; comunque le curve relative alle funzioni informative per i modelli con uno o due parametri saranno utili solo quando le *ICC* da cui sono derivate si adattano ai dati. L'uso di *ICC* che non si adattano adeguatamente conducono a risultati scorretti.
In pratica, le funzioni informative consentono di selezionare quegli item che presentano il livello di informazione necessario per soddisfare le specifiche esigenze dello strumento che si intende costruire[16].

---

[16] Il modello probabilistico si presenta particolarmente adatto alla selezione di item per la misurazione di capacità; il metodo che consente di selezionare gli item è particolarmente potente grazie alle specifiche caratteristiche del modello quali

- l'invarianza dei parametri degli item, che consente di utilizzare campioni diversi per la messa a punto;
- la possibilità di misurare sulla stessa scala sia i parametri degli item che la capacità dei soggetti;

A tale proposito, Lord nel 1977 ha definito una procedura che consente di utilizzare le funzioni degli item come elementi informativi per costruire scale che soddisfino le specificazioni considerate. Tale procedura impiega una banca di item, per ciascuno dei quali sono disponibili sia le stime dei parametri degli item secondo il modello prescelto che le funzioni informative. I passaggi suggeriti per tale procedura sono i seguenti:

1. definizione della forma della funzione informativa desiderata (*target information function*, *TIF*); per costruire una

Le funzioni informative di più item presentano la caratteristica di essere additive; quindi è possibile determinare una funzione informativa relativa ad un gruppo di item nel modo seguente:

$$I(d) = \sum_{i=1}^{n} I_i(d)$$

In altre parole gli item contribuiscono in modo indipendente alla funzione informativa del gruppo. Come abbiamo visto questa caratteristica non è riscontrabile in altri modelli di *scaling* per i quali non è possibile determinare per ciascun item le caratteristiche e il contributo all'affidabilità totale in modo indipendente dalle caratteristiche di tutti gli altri item.
Infine ricordiamo che la quantità di informazione fornita è inversamente correlata alla precisione con cui la capacità è stimata:

$$SE(\hat{d}) = \frac{1}{\sqrt{I(d)}}$$

dove
$SE(\hat{d})$ errore standard di stima

All'interno dell'*IRT*, l'errore standard di $\hat{d}$, $SE(\hat{d})$, che rappresenta la deviazione standard della distribuzione normale asintotica della stima di massima verosimiglianza della capacità per un dato valore vero di capacità *d*, svolge lo stesso ruolo dell'errore standard di misurazione nella teoria classica di misurazione.
Per poter confrontare più funzioni informative relative al la stessa capacità è possibile calcolare l'indice di efficienza relativa:

$$RE(d) = \frac{I_A(d)}{I_B(d)}$$

dove
$RE(d)$ efficienza relativa
$I_A(d)$ funzione informativa dello strumento *A*
$I_B(d)$ funzione informativa della strumento *B*
Se per esempio
$I_A(d) = 25.0$
$I_B(d) = 20.0$
allora $RE(d) = 1.25$
ovvero a livello *d* lo strumento *A* funziona come se fosse 25% più lungo dello strumento *B*; in altre parole

---

scala che
- misura una capacità con un ampio *range*, la *TIF* dovrebbe essere abbastanza piatta; ciò consente di costruire una scala che presenta una precisione di stima di capacità uniforme per tutto il *range*;
- richiede la definizione di un *cut-off* sulla scala di capacità che consenta di separare i soggetti misurati in due gruppi, la *TIF* dovrebbe essere molto appuntita vicino al punteggio di *cut-off*.

2. selezione dalla banca degli item con le funzioni informative che interessano;
3. calcolo della funzione informativa dell'intera scala;
4. ripetizione del precedente passaggio, eliminando item o inserendone altri fino al raggiungimento della funzione informativa di scala che più si avvicina alla *TIF*.

La procedura suggerita da Lord consente di costruire una scala che discrimina bene in una particolare regione del continuum di capacità; questo vuol dire che gli item possono essere selezionati in modo da massimizzare l'informazione di scala nella regione della capacità per i soggetti che devono essere misurati. E' il caso della costruzione di una scala che misura un rendimento; una tale scala dovrebbe avere item più semplici, ovvero che misurano *performance* più basse, quando utilizzata nel *pretest* e item più difficili, ovvero che misurano *performance* più alte, nel *post-test*. In ciascuna occasione, la precisione di misurazione sarà massimizzata nella regione di capacità dove i soggetti più probabilmente dovrebbero posizionarsi.
Ricordiamo però che il solo uso di criteri statistici per la selezione di item non assicura una scala con validità di contenuto; la sopravvalutazione dei criteri statistici è facile e comoda ma produce la conseguenza di non tenere in giusto conto il ruolo che il contenuto degli item ha nello sviluppo di una scala.

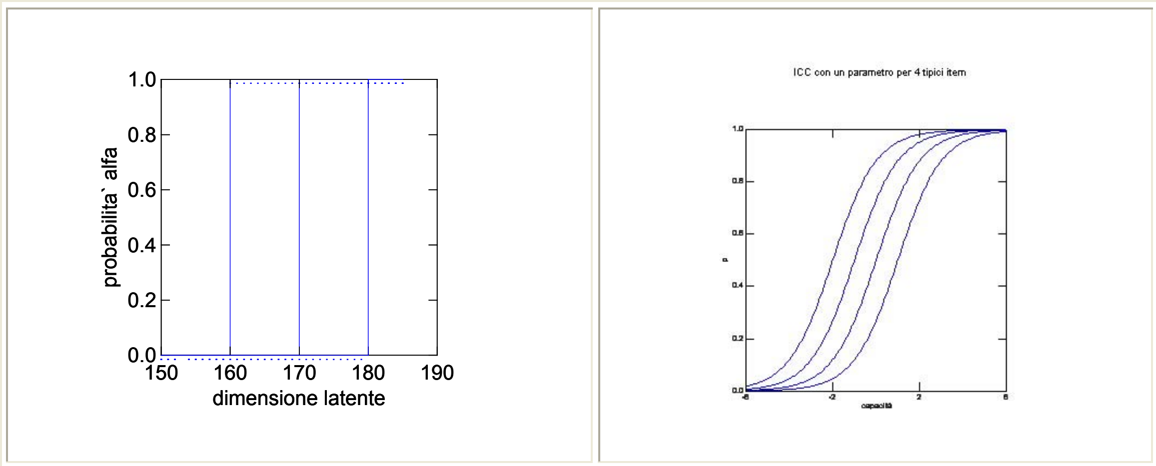
- *B* dovrebbe essere allungato del 25% per produrre la stessa precisione di misurazione di *A* a livello *d*;
- *A* potrebbe essere accorciato del 20% per continuare a produrre stime di capacità a livello *d* possedendo la stessa precisione di *B*.

Tali conclusioni, riguardati l'allungamento e la riduzione, sono basate sull'assunto che gli item aggiunti o eliminati siano confrontabili nella loro qualità statistica agli altri item. In altre parole, l’efficienza relativa svolge un ruolo molto simile a quello svolto dal coefficiente Spearman-Brown nel modello additivo.

4.4 ALCUNE CONSIDERAZIONI

4.4.1 Confronto tra modelli deterministici e probabilistici

Le due seguenti figure presentano abbastanza chiaramente la differenza tra le due versioni di una scala cumulativa e in particolare i limiti del modello deterministico (tre item) rispetto a quello probabilistico (quattro item), soprattutto se si pensa che praticamente mai i modelli di *scaling* adattano perfettamente i dati.



Come appare evidente mentre nel caso del modello deterministico non è prevista una risposta negativa ad un item che ricade a sinistra del punto-soggetto, nel caso del modello probabilistico la probabilità di una risposta positiva a ciascun item da parte di un soggetto aumenta via via che il punto-soggetto si sposta verso le posizioni più alte (a destra) lungo la dimensione.
Una delle riflessioni che si possono fare a questo punto riguarda la relazione tra modello e dati: quanto più il ricercatore è in grado di rendere sempre più espliciti gli assunti sulla natura degli errori quanto più diviene possibile accettare grandi discrepanze tra i modelli di *scaling* perfetto, definiti a livello di modello, e gli stessi dati empirici. Sembrerebbe quindi che le versioni probabilistiche del modello cumulativo siano non solo più realistiche dell'approccio deterministico, ma più adatte a costruire scale anche a partire da dati che presentano errori.

4.4.2 Modelli cumulativi e multidimensionalità

Uno dei problemi che pongono le procedure di *scaling* basate sul modello cumulativo è dato dal fatto che non sono in grado di rilevare la struttura dimensionale sottostante un insieme di dati, ovvero nel caso in cui vi siano più fonti di variabilità.
Le misure di bontà di adattamento utilizzate per verificare il modello cumulativo possono indicare quando una dimensione singola risulta essere inadeguata a rappresentare i dati. Comunque esse non possono essere utilizzate per distinguere tra le diverse ragioni di un tale risultato (perdita della

struttura, errore di misurazione, struttura dimensionale più complessa).
In generale gli approcci cumulativi funzionano meglio quando vengono applicati ai dati che si adattano un modello dimensionale nel quale entrambi gli insiemi di oggetti variano in modo sistematico rispetto ad un unico attributo.
Se il ricercatore ha il sospetto che:

- solamente uno dei due insiemi contiene una variazione sistematica,
- vi sono molte fonti di variabilità in un insieme di osservazioni,

può essere più proficuo orientarsi verso un altro approccio di *scaling*.
A tale proposito abbiamo visto, comunque, come sia i modelli deterministici che quelli probabilistici presentano vari tentativi di definire anche modelli multidimensionali.

Appendice.
LOGARITMI E LOGIT

I logaritmi

I *logaritmi* originariamente sono stati sviluppati dai matematici per potere trattare con maggiore facilità grandi numeri in problemi complessi, Successivamente si è riscontrato che potevano essere utili nelle descrizioni matematiche di molti fenomeni naturali e anche psicologici e sociali. L'utilizzo dei logaritmi consente di sostituire operazioni quali la moltiplicazione, la divisione e elevazione a potenza con operazioni più semplici ovvero, rispettivamente, l'addizione, la sottrazione e la moltiplicazione. Le regole basilari della matematica dei logaritmi sono le seguenti:

- La moltiplicazione è sostituita dall'addizione: $\log_n(x * y) = \log_n(x) + \log_n(y)$
  es. il logaritmo di 6(.778) più il logaritmo di 2(.301) = 1.08 ovvero il logaritmo di 12.
- La divisione è sostituita dalla sottrazione: $\log_n(x/y) = \log_n(x) - \log_n(y)$
  es. il logaritmo di 6(.778) meno il logaritmo di 2(.301) = .477 ovvero il logaritmo di 3.
- L'elevazione a potenza è sostituita dalla moltiplicazione: $\log_n(x^a) = a * \log_n(x)$
  es. il logaritmo di 6(.778) volte 2 = 1.556 ovvero il logaritmo di 36.

Un altro vantaggio nell'utilizzo dei logaritmi è dato dalla possibilità di utilizzo nella definizione di scale di misura e nella loro rappresentazione grafica: uguali distanze (cicli logaritmici) vengono trasformate automaticamente in uguali rapporti (i seguenti intervalli espressi in forma logaritmica risultano uguali: 1-10, 10-100, 100-1000).
Per questi motivi i logaritmi vengono utilizzati anche in statistica.
Il vantaggio di tale trasformazione è quello di poter utilizzare le proprietà dei logaritmi:

$$\log_n(xy) = \log_n(x) + \log_n(y) \qquad \text{e} \qquad \log_n(x^a) = a * \log_n(x)$$

che semplifica notevolmente i calcoli.

I *logit*

I *logit* sono unità matematiche o di probabilità logistica relativa a una data osservazione. In particolare un *logit* si ottiene calcolando il logaritmo naturale dell'*odds* di *p* che è dato dal rapporto tra *p* e il suo reciproco; nel nostro caso quindi il *logit* è dato dal logaritmo naturale del rapporto tra la probabilità di dare una risposta corretta (p_{ij}) e la probabilità di dare una risposta errata $(1 - p_{ij})$:

$$\log_n\left[\frac{p_{ij}}{(1 - p_{ij})}\right]$$

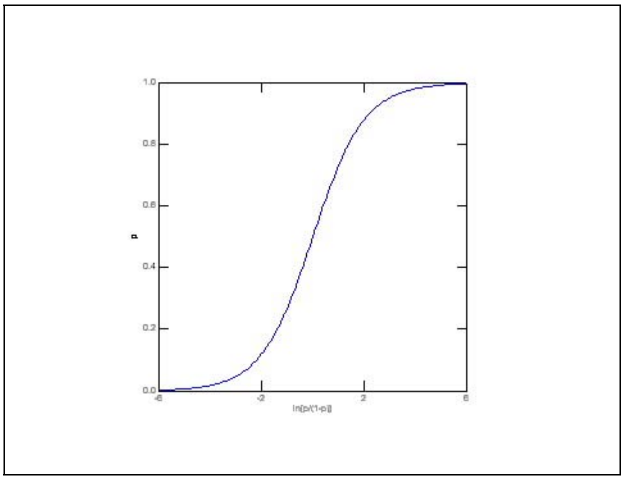
Il *logit* dà origine ad una scala ad intervalli e può essere sottoposto a qualsiasi trasformazione lineare. Si distribuisce simmetricamente intorno ad un valore centrale. Quando p=0.50, il *logit* sarà uguale a zero infatti:

$$\log_n\left[\frac{0.50}{0.50}\right] = \log_n(1) = 0$$

Via via che la dicotomia si sposta verso una delle due direzioni, ovvero approssimando 0 o 1, i valori di *logit* si allontanano da 0; nella seguente tabella osserviamo la corrispondenza tra p, $1 - p$ e *logit* per alcuni valori di p:

p	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1-p	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
logit	-2.20	-1.39	-0.85	-0.41	0.00	0.41	0.85	1.39	2.20

Da notare che mentre i valori di *p* nella tabella sono equidistanti (infatti distano sempre 0.10), la distanti tra i corrispondenti valori dei *logit* non è sempre la stessa ma aumenta allontanandosi da 0.50; inoltre all'aumentare del valore assoluto del *logit*, la probabilità corrispondente si avvicina ai valori estremi, senza raggiungerli mai.
Non esiste alcun limite superiore o inferiore del *logit* ma quando *p* è uguale esattamente a 1 o a 0 il *logit* risulta indefinito. La distribuzione continua dei valori di probabilità trasformati in *logit* è la seguente:



Come si potrà notare la curva logistica e la curva normale standardizzata cumulata sembrano molto simili e in generale è possibile fare per entrambe le stesse deduzioni. Come si può notare la trasformazione logistica è praticamente lineare per una porzione di valori di *p* (in particolare tra 0.20 e 0.75); questo vuol dire che all'interno di tale intervallo il modello lineare di probabilità produce risultati molto simili a quello del modello logistico. Quando la probabilità si avvicina ai due estremi (0 e 1) la trasformazione è in modo evidente non-lineare. A tale proposito è possibile fare alcune osservazioni. Tale trasformazione appiattisce le probabilità molto elevate e molto basse; questo vuol dire che i *logit* risultano molto utili quando è necessario effettuare confronti tra proporzioni a livelli diversi. Sia dalla tabella che dal grafico appare evidente come i valori di *logit* siano simmetrici; infatti la trasformazione logistica di un *odds* $\left(p/(1-p)\right)$ e del suo reciproco $\left((1-p)/p\right)$ produce un *logit* di valore uguale ma di segno opposto.

5. LE MAPPE PERCETTIVE

Sotto il termine *perceptual mapping* sono compresi diversi *modelli di scaling* che hanno il comune obiettivo di rappresentare e di identificare le dimensioni chiave sottostanti le valutazioni date da individui relative ad un insieme di oggetti.
I metodi per l'identificazione delle mappe percettive consentono di rappresentare le strutture mentali che gli individui utilizzano per percepire ed integrare informazioni. Tali metodi sono stati spesso applicati a dati prodotti da questionari utilizzati in indagini di *marketing* o a dati prodotti da sperimentazioni nell'ambito della linguistica e della psicologia cognitiva. Attualmente tali metodi possono essere applicati anche quando i dati non sono strettamente legati alla percezione.
Tali metodi richiedono che i dati siano rappresentati in matrici di prossimità (somiglianza o preferenza) al fine di

- verificare la presenza di dimensioni sottostanti le valutazioni individuali di prossimità tra oggetti,
- verificare l'importanza relativa di ciascuna dimensione,
- rappresentare in maniera adeguata gli oggetti nello spazio geometrico definito dalle dimensioni individuate conoscendo la prossimità tra loro.

Le differenze tra le posizioni reciproche degli oggetti possono essere attribuite a:

a. *variazioni nella dimensionalità*: pur potendo ipotizzare che gli individui giudicano utilizzando un numero limitato di dimensioni, non necessariamente gli individui percepiscono gli oggetti con la stessa dimensionalità;
b. *variazioni nell'importanza*: pur potendo ipotizzare che tutti gli individui valutino gli oggetti utilizzando le stesse dimensioni, non necessariamente gli individui attribuiscono la stessa importanza a ciascuna dimensione,
c. *variazioni nel tempo*: nel caso di indagini longitudinali è possibile anche ipotizzare che non necessariamente le valutazioni individuali rimangono stabili nel tempo.

I metodi di *mapping* possono essere suddivisi in due gruppi[1]:

- metodi per l'analisi e la rappresentazione di dati di somiglianza; a questo gruppo appartiene il *Multidimensional Scaling* che presenta diverse tecniche (metriche e non metriche, per dati individuali o aggregati); con dati metrici la procedura matematica si richiama a quella fattoriale;
- metodi per l'analisi e la rappresentazione di preferenze; le relazioni psicologiche e percepite tra gli stimoli sono rappresentate come relazioni geometriche tra i punti in uno spazio multidimensionale; il più conosciuto tra tali metodi è sicuramente l'*unfolding*, che consente di rappresentare in un unico spazio sia gli stimoli che gli individui.

---

[1] Data la definizione di *perceptual mapping*, a tali metodi è possibile affiancare anche l'analisi delle corrispondenze (Maggino, 2005). Infatti, i dati di input dell'analisi delle corrispondenze sono rappresentati in forma di tabella di contingenza che indica un'associazione qualitativa tra righe e colonne. L'analisi delle corrispondenze scala le righe e le colonne in unità corrispondenti in modo tale che possano essere rappresentate graficamente nello stesso spazio. Tali mappe spaziali forniscono degli spunti riguardo a:

- somiglianze e differenze tra le righe rispetto ad una determinata categoria di colonna,
- somiglianze e differenze tra le colonne rispetto ad una determinata categoria di riga,
- relazioni tra righe e colonne.

Anche se gli algoritmi alla base dell'analisi delle corrispondenze sono ripresi dall'analisi delle componenti principali, i raggruppamenti che risultano dall'analisi possono essere interpretati in termini di prossimità tra le righe e le colonne della tavola di contingenza. Le categorie che risultano essere più vicine possono essere considerate più simili rispetto alla struttura sottostante.
Il vantaggio dell'analisi delle corrispondenze, rispetto agli altri metodi di *mapping*, è rappresentato dal fatto che la raccolta dei dati è molto più semplice; anche se non sempre le distanze tra i due insiemi (righe e colonne) possono essere significativamente interpretate.

5.1 LA MAPPA DELLE SOMIGLIANZE: IL MULTIDIMENSIONAL SCALING

L'obiettivo del *Multidimensional Scaling* (*MDS*) (Cox, 1994; Kruskal, 1978; Torgerson, 1958) è quello di creare una rappresentazione geometrica multidimensionale detta *mappa*, che richiede valori espressi in termini di distanze, delle relazioni esistenti tra *K* oggetti utilizzando le prossimità note[2]. Ciò consente di rappresentare, attraverso grafici a due dimensioni, le relazioni anche complesse tra le variabili. L'*output* di tale procedura consiste proprio in una rappresentazione spaziale espressa da una *configurazione geometrica di punti*, come in una mappa, che dovrebbe rendere i dati di più semplice comprensione. Ciascun punto nella configurazione riflette la struttura non nota dei dati.
Per realizzare ciò occorre assumere tre modelli: delle prossimità, spaziale e delle distanze. Tali modelli possono in qualche modo essere considerati tra loro indipendenti; in altre parole, un particolare modello spaziale può essere combinato con uno dei diversi modelli delle distanze. Per questo motivo è possibile individuare diversi approcci al *MDS*, che si distinguono tra loro per:

- tipo di prossimità osservata (tipologia della matrice di input),
- tipo di modello di distanze e di misura del livello di adattamento tra prossimità e distanze;
- tipo di modello spaziale (per esempio, possibilità di rotazione degli assi della configurazione finale)
- l'*output* prodotto dall'analisi.

Per comprendere la logica alla base del *MDS* può essere utile presentare un esempio. In genere per rappresentare la posizione geografica delle città di un territorio si dispone di cartine geografiche che riproducono in scala la posizione reciproca tra le città. Il rapporto di scala consente di ricostruire le distanze chilometriche reali tra le città.

Si supponga di disporre solo delle distanze chilometriche (percorrenza stradale) tra 32 città europee.
L'obiettivo può essere di definire una configurazione spaziale dimensionale interpretabile (mappa) definita da due dimensioni interpretabili ("*nord/sud*" e "*est-ovest*") e all'interno della quale riprodurre le reali posizioni reciproche delle città; le distanze stradali non consentono di riprodurre direttamente la mappa in quanto sono determinate sulla base delle vie di comunicazione e, conseguentemente, possono essere interpretate in termini statistici in termini di prossimità (dissomiglianza). E' quindi necessario stimare le distanze reali reciproche (distanza in linea d'aria) tra le città a partire dalla matrice delle distanze stradali.

[2] Il *multidimensional scaling*, pur essendo nato in campo psicometrico, è un approccio che ha trovato spazio come metodo generale di analisi. Per ciò, nella trattazione tecnica di questo metodo, verranno utilizzati in modo intercambiabile sia il termine di variabile che quello di oggetto.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	BB	CC	DD	EE	FF
A	0																															
B	313	0																														
C	155	287	0																													
D	184	128	204	0																												
E	67	258	186	131	0																											
F	84	229	91	152	95	0																										
G	23	301	133	173	81	67	0																									
H	144	169	202	41	90	115	138	0																								
I	228	130	267	63	180	219	228	90	0																							
J	24	289	138	161	59	70	21	120	210	0																						
K	47	266	134	138	54	47	41	97	187	23	0																					
L	47	288	183	159	30	98	60	119	209	49	51	0																				
M	186	397	322	279	149	237	199	228	282	188	190	139	0																			
N	285	119	312	100	232	252	274	142	70	262	239	260	352	0																		
O	82	323	218	175	45	133	95	135	225	84	86	35	104	276	0																	
P	233	412	129	332	290	211	210	330	395	231	238	270	409	446	305	0																
Q	41	315	147	205	103	96	32	170	259	58	78	85	227	316	120	224	0															
R	178	347	63	267	234	153	155	274	330	176	183	215	354	381	250	65	169	0														
S	113	186	95	112	113	40	95	107	175	98	78	124	262	212	164	234	129	169	0													
T	123	236	52	156	155	58	106	157	219	107	106	156	295	256	192	179	125	114	58	0												
U	85	219	134	97	60	44	79	69	160	62	38	77	209	197	105	255	120	197	59	103	0											
V	129	370	265	238	107	180	142	201	287	131	133	82	97	339	62	352	167	297	206	238	159	0										
W	52	272	104	183	108	59	29	154	247	50	57	89	228	283	124	181	43	126	86	80	87	171	0									
X	92	225	174	96	35	84	92	56	149	75	51	63	184	196	80	295	129	237	99	143	40	144	105	0								
Y	175	146	141	140	156	103	158	125	203	161	141	170	305	240	201	272	192	204	63	93	96	252	150	136	0							
Z	226	91	250	41	172	192	214	82	39	202	179	201	320	60	216	373	250	322	153	197	138	279	224	137	181	0						
AA	146	387	282	239	109	197	159	199	289	148	150	99	40	341	64	369	182	314	223	256	168	57	188	145	265	280	0					
BB	71	243	127	120	63	37	55	93	183	48	24	71	212	220	106	248	88	190	55	96	23	153	64	48	118	161	171	0				
CC	143	177	128	83	113	70	125	79	146	115	91	130	262	184	158	254	159	191	30	80	53	212	116	84	58	124	221	76	0			
DD	123	236	240	108	56	151	137	67	121	115	112	86	161	209	94	362	159	304	160	210	107	159	164	67	187	149	160	115	131	0		
EE	119	194	178	66	65	90	113	25	115	95	72	91	214	167	111	301	150	243	92	142	44	174	131	30	120	107	176	67	63	68	0	
FF	84	217	103	140	83	13	67	102	205	70	47	93	232	244	130	224	95	165	31	72	31	177	58	71	94	181	195	24	61	138	77	0

Citta'	Simbolo	Citta'	Simbolo	Citta'	Simbolo	Citta'	Simbolo	Citta'	Simbolo
Amsterdam	A	Budapest	H	Kobenhavn	O	Oslo	V	Venezia	CC
Athinai	B	Bucaresti	I	Lisboa	P	Paris	W	Warszawa	DD
Barcelona	C	Düsseldorf	J	London	Q	Praha	X	Wien	EE
Beograd	D	Frankfurt	K	Madrid	R	Roma	Y	Zürich	FF
Berlin	E	Hamburg	L	Milano	S	Sofija	Z		
Bern	F	Helsinki	M	Marseille	T	Stockholm	AA		
Bruxelles	G	Istambul	N	München	U	Stuttgart	BB		

Per raggiungere tale obiettivo occorre (in sequenza):

a. disporre della matrice delle prossimità (distanze stradali tra le città),
b. definire la funzione che lega prossimità osservate e distanze da stimare (modello delle distanze) e il numero di dimensioni che spiegano le prossimità (modello spaziale),
c. stimare le distanze euclidee a partire dalle prossimità determinando una funzione che leghi le due misure (coordinate spaziali),
d. collocare, sulla base delle distanze calcolate, gli oggetti (le città) su una mappa (carta geografica),
e. valutare la bontà di adattamento del modello spaziale ai dati di partenza, tra prossimità iniziali e distanze stimate; tale valutazione mira a verificare se ciò che è stato ottenuto rappresenta una valida stima della reale posizione geografica esistente.

In questo esempio la mappa finale dovrebbe essere rappresentata da due dimensioni interpretabili in termini di dimensioni geografiche (*nord-sud* e *est-ovest*). L'orientamento della mappa sarà quello che risulterà più facilmente interpretabile; questo vuol dire che, tenendo fisse le distanze calcolate, è possibile ruotare la mappa in qualsiasi modo; nell'esempio risulterà più facilmente riconoscibile l'orientamento *est-ovest* come prima dimensione (*ascisse*) e *nord-sud* come seconda dimensione (*ordinata*). Il procedimento di stima della reale misura può complicarsi sia dalla presenza nei dati di errori di misurazione (detti *rumori*), che dalla mancanza di informazioni sul numero di dimensioni necessarie per riprodurre la *mappa*.

L'esempio aiuta a comprendere le potenzialità del MDS che in realtà consente di affrontare i classici problemi della misurazione del soggettivo.

5.1.1 Verifica del modello

Cercando di sintetizzare, il procedimento prevede i seguenti momenti:

a. costruzione della matrice di prossimità tra gli oggetti;
b. definizione degli assunti ovvero del modello delle distanze e del modello spaziale,
c. inizio del procedimento iterativo che prende il via dalla individuazione di una configurazione arbitraria iniziale:
    - stima delle distanze a partire dalle prossimità osservate e dello spazio utilizzando il modello delle distanze e quello spaziale adottati,
    - confronto tra la matrice delle prossimità e quella delle distanze ovvero valutazione del livello d'adattamento tra prossimità rilevate (δ) e distanze stimate (d) attraverso il calcolo di un indice di bontà della soluzione;
    - calcolo delle disparità tra prossimità e distanze;
    - il procedimento iterativo prosegue fino a quando il confronto tra prossimità e distanze non indica un alto adattamento;
d. determinazione delle coordinate per ciascun oggetto secondo il modello spaziale adottato e quindi costruzione della nuova matrice (***X***), di dimensione $N * K$, contenente le coordinate di ciascun oggetto *i* per ciascuna dimensione *r* $(x_{i1}, x_{i2}, ..., x_{iR})$ e rappresentazione grafica di tutti i punti x_{ir};
e. interpretazione delle relazioni tra gli oggetti e delle dimensioni ottenute (con eventuale rotazione degli assi per giungere ad una migliore interpretazione).

5.1.1.1 Definizione del modello delle distanze

Il *MDS* assume che la relazione tra gli oggetti, espressa in termini di prossimità, può essere riespressa in termini di distanze all'interno di uno spazio definito; per procedere a tale trasformazione è necessario definire la relazione tra prossimità[3] e distanze; tale relazione può essere genericamente definita da una funzione:

$$d_{ij} = f(\delta_{ij})$$

Esistono diverse funzioni che derivano da diversi modelli delle distanze; di seguito sono presentati i principali.

Modelli metrici

- *Modello classico*: assume l'uguaglianza tra dissomiglianze e distanze:

[3] Si ricordi che se le prossimità sono espresse da somiglianze, tra queste e le distanze esisterà una relazione inversa; per questo motivo spesso, prima di procedere alla conversione, si può procedere alla riflessione dei valori di prossimità; questo vuol dire che se le prossimità sono state misurate in termini di somiglianze, queste vengono trasformate in dissomiglianze. La riflessione dei valori non produce alcun problema in quanto la trasformazione conserva tutta l'informazione contenuta nella matrice originale dei dati.

$$d_{ij} = \delta_{ij} \qquad \text{per } i, j = 1,...,K$$

La rigidità dell'assunto di base rende questo modello molto potente; d'altra parte l'applicazione di questo modello anche quando l'assunto non può essere soddisfatto, rende spuri i risultati prodotti.

- *Modello lineare*: assume tra dissomiglianze e distanze una relazione lineare:

$$d_{ij} = a + b\delta_{ij}$$

I valori a e b devono essere stimati attraverso il metodo dei minimi quadrati. Il modello classico può essere considerato un caso particolare del modello lineare. Quando l'origine della retta stimata dai minimi quadrati (e che descrive la relazione tra prossimità e distanze) coincide con l'origine degli assi cartesiani allora $a = 0$ e $b = 1$ e quindi la relazione diviene la seguente:

$$d_{ij} = b\delta_{ij}$$

I punti quindi, si posizionano lungo la bisettrice. Quando i punti si dispongono lungo una retta non passante per l'origine, ma parallela alla bisettrice (a diverso da 0), è necessario stimare il valore di a, valore costante che indica, per tutte le dissomiglianze, la divergenza tra queste e le distanze. In questo caso il metodo prende il nome di *metodo della costante additiva*.

Una variante al modello metrico frequentemente utilizzata è quella detta *delle coordinate principali*.
I principali vantaggi dell'approccio metrico risiedono nel fatto di consentire un'agevole individuazione del numero (grazie alla possibilità di utilizzare i tipici strumenti dell'approccio fattoriale) e interpretazione delle dimensioni (grazie alla possibilità di rotare gli assi e la configurazione spaziale).

Modello non-metrico (*analisi delle prossimità*)

Questo modello può essere adottato quando non è possibile soddisfare i forti assunti dei modelli metrici; esso assume che tra prossimità e distanze vi sia una relazione *monotona* considerando le prossimità solo in termini ordinali. Tutte le funzioni monotone adottabili in questo ambito, e che soddisfano il requisito di monotonicità (lineare, esponenziale, logaritmica, ecc.), soddisfano le seguenti relazioni:

$$\delta_{ij} < \delta_{ls} \rightarrow f(\delta_{ij}) \leq f(\delta_{ls}) \qquad \text{per tutti } 1 \leq i, j, l, s \leq K$$

In questo senso il modello metrico può essere considerato un caso particolare del più generale modello non-metrico. Applicato a prossimità che presentano proprietà metriche, il modello non-metrico consente di superare il problema della presenza di errori nella valutazione delle prossimità.
Il primo modello messo a punto è stato quello metrico; gli altri modelli proposti sono generalizzazioni successive del primo modello metrico proposto[4].

5.1.1.2 Definizione del modello spaziale

La definizione del modello spaziale comporta la definizione delle caratteristiche formali dello spazio multidimensionale (Kruskal, 1978); per poter giungere a tale definizione occorre assumere che

- l'attributo rispetto al quale si misura l'oggetto è rappresentato da più aspetti/dimensioni;
- la dimensionalità dello spazio corrisponde alla dimensionalità dell'attributo;
- la posizione di un oggetto nello spazio corrisponde ai valori che l'oggetto registra rispetto all'attributo multidimensionale;

A questo punto è possibile definire il modello spaziale o geometrico (caratteristiche formali dello spazio ipotizzato) in termini di relazione tra *dimensionalità dello spazio* e *proiezioni*

[4] Il modello non-metrico è stato sviluppato nel 1962 da Shepard che ha focalizzato l'attenzione sulla relazione tra prossimità e distanze, dimostrando che era possibile derivare soluzioni metriche assumendo una relazione ordinale tra prossimità e distanze. Nel 1964 Kruskal, lavorando direttamente sull'*analisi di prossimità* di Shepard, ha inserito nel processo iterativo la funzione valutativa della bontà dell'adattamento delle soluzioni progressivamente ottenute.

dei punti sugli assi dello spazio con *le distanze tra i punti*. In particolare, il modello spaziale adottato consente di attribuire un significato ai numeri assegnati agli oggetti.

5.1.1.3 Determinazione delle distanze

Sulla base degli assunti, si procede alla stima delle distanze e alla determinazione della configurazione spaziale (posizionamento dei punti nello spazio); tale configurazione è individuata attraverso un procedimento iterativo per il quale è necessario definire una configurazione spaziale di partenza attraverso particolari algoritmi che presentano soluzioni comunque arbitrarie[5].
Ad ogni iterazione si procede alla valutazione della bontà di adattamento che consiste nel confronto tra distanze calcolate e prossimità iniziali. Se l'adattamento non risulta soddisfacente, si procede ad una nuova collocazione dei punti. Il procedimento prosegue fino a quando la valutazione dell'adattamento non produce un valore soddisfacente. L'obiettivo è quello di massimizzare la *bontà-di-adattamento* (o di minimizzare la *perdita-di-adattamento*). Nel caso di basso adattamento si procede alla determinazione delle disparità $(\hat{d})$ in base alle quali trasformare le distanze in modo tale che il loro ordinamento rispecchi quello delle prossimità e produca gli scarti minimi tra d e δ .

5.1.1.4 Valutazione dell'adattamento

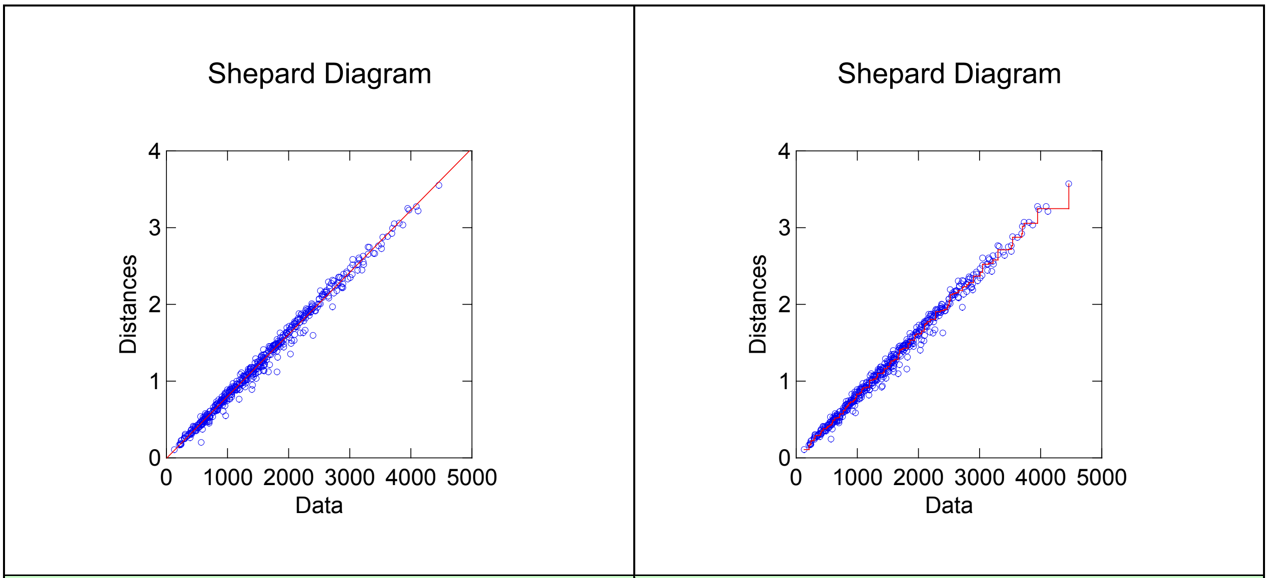
Il diagramma di *Shepard*

Una tecnica utile a verificare l'adattamento è quella che utilizza il diagramma di Shepard. In tale diagramma ciascun punto corrisponde ad una coppia di stimoli (i, j); la posizione di ciascun punto corrisponde alle coordinate corrispondenti alle

- prossimità osservate (δ_{ij}) in ascisse,
- distanze (d_{ij}) in ordinata.

Nel modello perfetto, quando tutte le distanze riprodotte sono identiche alle distanze osservate, tutti i punti ricadono su una linea retta e presentano un andamento ascendente, dal basso a sinistra (a piccole differenze corrispondono piccole distanze) verso l'alto a destra (a grandi differenze corrispondono grandi distanze).
Quando l'adattamento non è perfetto, il grafico consente di valutare le discrepanze tra la funzione teorica e l'andamento ottenuto. Tali discrepanze sono rappresentate da rette (parallele all'asse y) che esprimono la distanza tra i punti osservati e i punti stimati dalla funzione.
Quando si assume una funzione monotona, i punti risultano collocati su curve che presentano quasi sempre una caratteristica forma frastagliata.
Tale *diagramma* consente anche di individuare eventuali *outlier* oppure particolari deviazioni sistematiche dal modello, importanti da un punto di vista teorico in quanto possono suggerire l'applicazione di altri tipi di funzione.
Nella seguente figura sono rappresentati i diagrammi di Shepard delle stime delle distanze ottenute applicando, rispettivamente, il modello lineare e il modello monotono alle distanze stradali delle 32 città.

[5] E' consigliabile avere e provare sempre diverse configurazioni di partenza per evitare problemi come quello del minimo locale (di cui si parlerà più avanti).



MODELLO LINEARE	MODELLO MONOTONO

I coefficienti

La valutazione della bontà di adattamento può avvenire utilizzando anche alcuni coefficienti detti *objective function*[6]. Tali coefficienti, applicabili sia nell'ambito dell'approccio metrico che non-metrico, possono essere utilizzati non solo per la valutazione della configurazione finale ma anche per le configurazioni ottenute lungo il procedimento iterativo. L'interpretazione dei valori dei coefficienti deve tenere conto che tali valori sono funzione anche del numero di dimensioni (R) e del numero di oggetti (K). Di seguito sono presentati i coefficienti più comuni.

- **Il coefficiente di *stress***: i valori prodotti da tale coefficiente, definito nel 1964 da Kruskal per valutare la bontà di adattamento (Kruskal, 1978), possono essere interpretati in termini di varianza residua (non spiegata); in altre parole esprime quanta parte della variabilità delle prossimità non viene interpretata dalla configurazione ottenuta:

$$stress = \sqrt{\frac{\sum\left(\delta_{ij} - d_{ij}\right)^2}{\sum \delta_{ij}^2}}$$

Un'altra versione del coefficiente, proposta sempre da Kruskal, è la seguente:

$$stress = \sqrt{\frac{\sum\left(\delta_{ij} - d_{ij}\right)^2}{\sum\left(\delta_{ij} - \bar{\delta}\right)^2}}$$

dove $\bar{\delta}$ rappresenta la media delle prossimità.

Il valore dello *stress* varia da 0, che indica un adattamento perfetto, a 1, che indica una completa mancanza di adattamento, tra la matrice della configurazione ottenuta (distanze riprodotte) e la matrice delle prossimità osservate. Naturalmente, essendo in pratica molto difficile ottenere un adattamento perfetto, l'obiettivo è quello di ottenere il valore di *stress* più basso possibile. Kruskal ha cercato di dare delle indicazioni utili per la interpretazione dei valori dello *stress:*

[6] In letteratura tale criterio generale si trova con termini diversi quali *criterion function*, *error function*, *evaluation function*, *merit function*, *goodness-of-fit function*, *badness-of-fit function*.

valori di stress	livello di adattamento
0	Perfetto
0.001-0.025	Eccellente
0.026-0.05	Buono
0.06-0.10	Sufficiente
0.11-0.20	Scarso - carente
>0.20	Soluzione da ignorare.

- **Il coefficiente di *alienazione***: dal punto di vista concettuale la funzione di perdita definita dal coefficiente di *alienazione* equivale a quella definita per il coefficiente di *stress*, l'algoritmo, proposto da Guttman nel 1968 è piuttosto diverso:

  $$alienazione = \sqrt{\left(1-\mu^2\right)}$$

  dove μ rappresenta il *coefficiente di continuità* per la configurazione che è dato da

  $$\mu = \sqrt{\frac{\left(\sum d_i d_i^*\right)^2}{\left(\sum d_i^2\right)\left(\sum d_i^{*2}\right)}}$$

  dove d^* si riferisce al valore di partenza della distanza. E' stato dimostrato che la minimizzazione di tale coefficiente equivale a quella del coefficiente di *stress*.
- **L'indice *gamma* di Shepard**: tale indice consente di valutare quanto la configurazione geometrica si allontana da una relazione monotona con le prossimità iniziali:

  $$gamma = \frac{\sum_{i<j}\left[\delta_{ij} - K({}_2 d_{ij})\right]^2}{\frac{N(N-1)}{2}}$$

  dove

  $K\left({}_2 d_{ij}\right)$ rango della distanza finale tra i punti *i* e *j*.

  $N(N-1)/2$ numero di confronti che si effettuano.

  Gamma quindi non è altro che la media degli scarti quadratici tra gli elementi della matrice di partenza e quella della soluzione ottenuta.

- **L'indice di interpretazione** (α): tale indice, valido sia per l'analisi metrica che non-metrica, può essere interpretato come misura della strettezza del legame tra i dati originari trasformati e la soluzione finale:

  $$\alpha = 1/2 + \frac{1/2\left(\sum_{i,j}\delta_{ij} d_{ij}\right)}{\left(\sum_{i,j}\delta_{ij}^2\right)\left(\sum_{i,j} d_{ij}^2\right)}$$

  L'indice varia tra
  - 0, massima discrepanza tra i due tipi di dati,
  - 1, perfetta prevedibilità dei dati di partenza sulla base della configurazione ottenuta.

In generale, può essere utile applicare diverse *objective function* in modo da mettere a confronto i risultati, anche se l'esperienza ha comunque mostrato che l'applicazione di funzioni diverse non conduce a risultati molto differenti tra loro.
Non essendo nota la loro distribuzione campionaria, tali coefficienti non sono interpretabili in termini di significatività; conseguentemente non è possibile verificare la significatività statistica

della configurazione individuata[7]. Se, comunque, si applicano i coefficienti in maniera corretta, è ragionevole supporre che la dimensionalità ottenuta approssimi quella vera.
Nella seguente tabella sono riportati i valori di *stress* prodotti dalle iterazioni effettuate per stimare le distanze tra le 32 città secondo il modello lineare e monotono.

MODELLO DI DISTANZE: FUNZIONE LINEARE		MODELLO DI DISTANZE: FUNZIONE MONOTONA	
Iteration	STRESS	Iteration	STRESS
0	0.058877	0	0.051314
1	0.048413	1	0.041978
2	0.045902	2	0.038861
3	0.045113	3	0.037623
Stress of final configuration is: 0.04511		Stress of final configuration is: 0.03762	
Proportion of variance (RSQ) is: 0.99134		Proportion of variance (RSQ) is: 0.99399	

5.1.1.5 Problemi nell'interpretazione dell'adattamento: minimo globale e minimo locale

Osservare un soddisfacente valore per i coefficienti visti non può essere considerato necessariamente un indice di buon adattamento in quanto tale valore può essere ottenuto anche in presenza di una violazione degli assunti del modello di distanze (monotonicità della relazione tra prossimità e distanze). Questo vuol dire che è necessario assicurarsi che non si sia in presenza di una soluzione degenerata a causa di uno dei principali problemi del procedimento iterativo: la possibilità di ottenere un valore, per i coefficienti, minimo distorto.
Per comprende i concetti di minimo globale e locale può risultare utile un esempio. Si immagini un territorio molto irregolare con avvallamenti e rilievi. Ciascun punto del territorio corrisponde ad una delle configurazioni ottenute lungo il procedimento iterativo e può essere individuato da tre parametri:

- un'altezza, che rappresenta il valore della *objective function*,
- due orientamenti-dimensioni (nell'esempio più semplice).

Cercare la configurazione con il massimo adattamento corrisponde a cercare il punto più basso del territorio (altezza minima). Siccome è impossibile avere informazioni sui valori di adattamento per tutte le configurazioni, la ricerca di quella migliore corrisponde alla ricerca del punto più basso del terreno percorrendolo con gli occhi bendati (procedimento iterativo) a partire da un punto individuato a caso (un paracadutista lanciato sul territorio con gli occhi bendati in una notte buia da un aereo ad alta quota) definito come configurazione di partenza; a partire da tale punto lo stesso paracadutista bendato procede, tastando con un piede il territorio circostante, alla ricerca del punto più basso del terreno, spostandosi nella direzione in cui riesce ad individuare la pendenza maggiore. Dal nuovo punto il paracadutista procede nello stesso modo. Quando egli raggiunge un punto dal quale non riesce a individuare alcuna pendenza maggiore, si ferma.
Se il terreno non è molto irregolare e presenta un unico avvallamento, è molto probabile che il paracadutista alla fine giunga nel punto più basso del territorio. Se il terreno non presenta un unico avvallamento ma molte irregolarità, allora il paracadutista ha la possibilità di fermarsi al centro di un'irregolarità che non necessariamente rappresenta il massimo avvallamento. Se il punto rappresenta l'avvallamento massimo (adattamento massimo) il valore di adattamento è detto

[7] Per verificare la significatività statistica della discrepanza tra prossimità originali e quelle della configurazione finale possono comunque essere utilizzati i seguenti indici:

- *chi²* di Pearson, quando le prossimità iniziali sono rappresentate da frequenze;
- *rho* di Spearman, quando si tratta di ranghi variabili,
- *indice di concordanza C*, basato sul *chi²*, che consente di verificare la bontà del risultato dell'analisi non-metrica:

$$C = 1 - \frac{\sum_i \sum_r \left(x_{ir} - x_{ir}^{*2}\right)}{\sum_i \sum_r x_{ir}^{*2}}$$

dove

x_{ir} coordinata dell'oggetto *i* sull'asse *r* nella configurazione di partenza

x_{ir}^{*} coordinata dell'oggetto *i* sull'asse *r* nella configurazione ottenuta.

minimo globale, in caso contrario si parla di *minimo locale*. Come si può intuire, è più probabile trovare un minimo locale che un minimo globale. In genere la differenza tra minimo locale e minimo globale è poca; quando la configurazione risulta non interpretabile è molto probabile che la differenza tra i due minimi sia molto grande.
Nei casi in cui si pensa di avere ottenuto un minimo locale è consigliabile ripetere l'analisi con una configurazione di partenza diversa. Utilizzare molte e diverse configurazioni di partenza equivale a lanciare molti paracadutisti in punti diversi che proseguono la ricerca separatamente. E' improbabile che le diverse configurazioni di partenza convergano tutte ad un comune minimo locale a meno che esso non rappresenti il minimo globale. Quando si ottengono minimi locali molto diversi e valori di adattamento soddisfacenti, è probabile che le configurazioni ottenute non siano utilizzabili.
L'osservazione del diagramma *Shepard* consente di verificare la presenza di una soluzione degenerata che si verifica quando
- i punti della configurazione vanno a formare un denso raggruppamento,
- i punti, che invece di configurare una retta compongono una caratteristica forma a scala consistente di pochi e ampi scalini, si concentrano intorno a tali scalini.

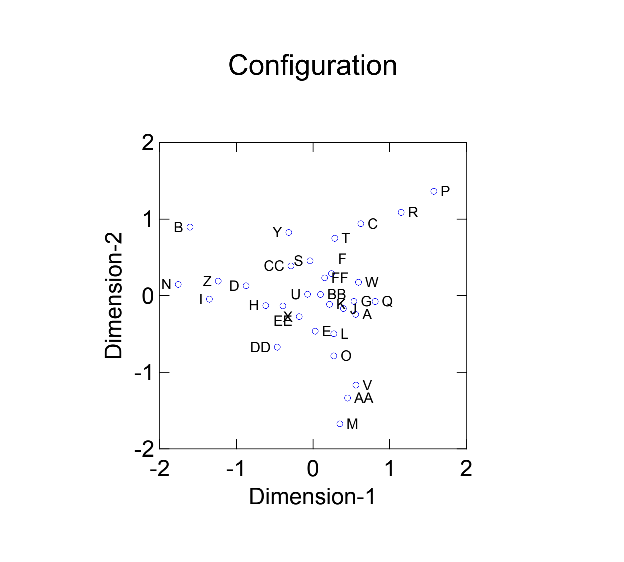
Il pericolo di un minimo locale è minore se la scelta della configurazione di partenza è fatta in maniera ragionata, per esempio traendola dalla configurazione ottenuta dall'applicazione dell'approccio metrico sugli stessi dati.

5.1.1.6 Interpretazione della configurazione

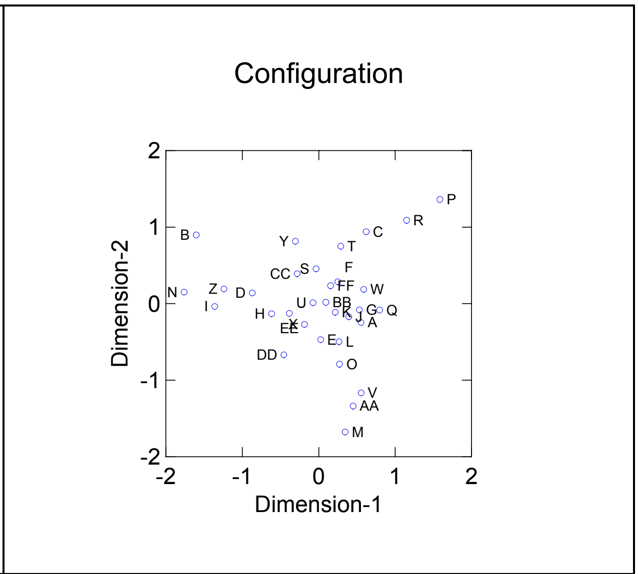
La configurazione ottenuta può essere interpretata (Kruskal, 1978) in due direzioni:
- *osservazione e interpretazione delle relazioni reciproche tra gli oggetti*: ciò significa essenzialmente interpretare le prossimità; per fare questo si può procedere, per esempio, individuando nella configurazione dei possibili raggruppamenti degli oggetti; in questa fase può essere utile applicare anche altre tecniche, come la *cluster analysis* gerarchica;
- *osservazione e interpretazione del significato degli assi*, analogamente a quanto si fa nell'analisi fattoriale[8]; a tal fine può essere utile ricordare che:
  - il numero di dimensioni interpretabili può anche essere minore della dimensionalità reale dello spazio (e quindi del costrutto);
  - la presenza di una dimensione non interpretabile non significa necessariamente che la dimensione non abbia realmente alcuna interpretazione; in genere la difficoltà di interpretazione può essere causata sia dall'arbitrario orientamento degli assi che dalla presenza di un'alta dimensionalità (difficile da valutare per la difficoltosa rappresentazione grafica).

Nella seguente figura sono rappresentate le mappe ottenute con i dati dell'esempio delle 32 città (le relative coordinate per ciascuna città sono presentate nella successiva tabelle).
Osservando la configurazione ottenuta con il modello lineare e quello monotono è possibile notare, confrontando le posizioni reciproche delle città, come le due dimensioni ottenute possono essere interpretate secondo l'ipotesi fatta (*nord/sud* e *est-ovest*).

[8] Tra *MDS* e analisi fattoriale esistono alcuni importanti elementi comuni. Entrambi i metodi di analisi utilizzano una matrice simmetrica di partenza (di prossimità per il *MDS* e di correlazione per l'analisi fattoriale). Come è noto la correlazione rappresenta uno degli approcci che consentono di definire una misura di prossimità. E' però importante ricordare che l'interpretazione dei coefficienti di correlazione in termini di misura di somiglianza non è la migliore in quanto le correlazioni distorcono le distanze tra i punti. Esiste un caso in cui le correlazioni sono correlate in modo monotono alle distanze tra punti: quando tutti i punti sono collocati alla stessa distanza dall'origine (ciò corrisponde alla standardizzazione degli assi).



FUNZIONE LINEARE



FUNZIONE MONOTONA

SIMBOLO	CITTA'	Coordinate			
		Funzione monotona		Funzione lineare	
		DIM(1)	DIM(2)	DIM(1)	DIM(2)
A	AMSTERDAM	0.554	-0.248	0.558	-0.246
B	ATHINAI	-1.601	0.895	-1.605	0.893
C	BARCELONA	0.622	0.939	0.625	0.938
D	BEOGRAD	-0.868	0.139	-0.873	0.129
E	BERLIN	0.026	-0.472	0.030	-0.465
F	BERN	0.251	0.285	0.242	0.286
G	BRUXELLES	0.533	-0.084	0.535	-0.076
H	BUDAPEST	-0.615	-0.134	-0.616	-0.131
I	BUCARESTI	-1.360	-0.039	-1.351	-0.046
J	DÜSSELDORF	0.394	-0.173	0.397	-0.169
K	FRANKFURT	0.217	-0.114	0.217	-0.114
L	HAMBURG	0.268	-0.498	0.273	-0.499
M	HELSINKI	0.347	-1.679	0.351	-1.674
N	ISTAMBUL	-1.761	0.150	-1.759	0.145
O	KOBENHAVN	0.274	-0.792	0.272	-0.788
P	LISBOA	1.587	1.360	1.578	1.361
Q	LONDON	0.798	-0.086	0.812	-0.077
R	MADRID	1.150	1.088	1.151	1.086
S	MILANO	-0.033	0.453	-0.039	0.453
T	MARSEILLE	0.288	0.750	0.285	0.747
U	MÜNCHEN	-0.073	0.011	-0.070	0.017
V	OSLO	0.556	-1.166	0.561	-1.168
W	PARIS	0.589	0.184	0.595	0.174
X	PRAHA	-0.185	-0.273	-0.180	-0.275
Y	ROMA	-0.305	0.814	-0.315	0.823
Z	SOFIJA	-1.238	0.190	-1.236	0.188
AA	STOCKHOLM	0.451	-1.339	0.452	-1.337
BB	STUTTGART	0.095	0.016	0.099	0.014
CC	VENEZIA	-0.281	0.389	-0.287	0.387
DD	WARSZAWA	-0.455	-0.669	-0.465	-0.673
EE	WIEN	-0.383	-0.130	-0.391	-0.136
FF	ZÜRICH	0.157	0.234	0.152	0.231

E' possibile a questo punto fare alcune osservazioni sui risultati fin qui ottenuti relativamente all'esempio presentato:

- tutte le analisi hanno registrato il valore massimo di *bontà di adattamento* dopo un numero di iterazioni piuttosto basso;

- tutti i diagrammi di Shepard hanno confermato una decisa corrispondenza tra dissomiglianze e distanze stimate;
- tutte le coordinate e, conseguentemente, le configurazioni ottenute risultano tra loro molto simili o comunque confrontabili.

In questo esempio, contrariamente alla situazione in cui si trova un ricercatore sociale, la possibilità di disporre di informazioni sulla configurazione reale consente di osservare come la configurazione ottenuta presenta una buona approssimazione delle reali posizioni reciproche delle città. Nel caso in cui si volesse adattare la mappa ottenuta all'orientamento utilizzato convenzionalmente nelle rappresentazioni geografiche, la configurazione necessiterà di una rotazione degli assi.
Una più attenta e approfondita osservazione rivela però come, in realtà, la localizzazione delle diverse città non è del tutto corretta se confrontata con quella reale. Ciò è sicuramente dovuto al fatto che le distanze stradali tra le città tengono conto di altre dimensioni, quali per esempio i dislivelli del territorio, di cui una ipotesi bidimensionale non ha potuto tenere conto[9].

5.1.2 Numero delle dimensioni

Nel definire il modello di misurazione non sempre si è sicuri della dimensionalità identificata per il costrutto; in alcuni casi tale dimensionalità può non essere nota; in questi casi l'analisi multidimensionale serve proprio a individuare il numero di dimensioni significative per l'attributo considerato. A tale proposito è importante distinguere tra:

- *dimensionalità vera o corretta*, ovvero quella realmente presente nel costrutto e, si ipotizza, nei dati; essa è ignota e ipotetica ed è determinabile statisticamente in termini probabilistici;
- *dimensionalità appropriata*, ovvero quella che si ottiene dopo l'analisi e che si ritiene adatta all'interpretazione dei dati.

La decisione riguardante la dimensionalità appropriata è molto importante e rappresenta una questione più sostanziale che statistica; inoltre la decisione, anche se può richiedere l'utilizzo di un criterio statistico, introduce elementi soggettivi difficilmente verificabili. Per questo motivo la mancanza di ipotesi sul numero di dimensioni che individuano il costrutto rappresenta un fattore di debolezza dell'applicazione che si intende fare.
Uno dei criteri utilizzabili per stabilire la dimensionalità appropriata è quella che tiene conto del numero di oggetti sottoposti ad analisi:

$$R \leq \left(\frac{K}{4}\right)$$

dove
R numero di dimensioni
K numero di oggetti.

E' comunque possibile identificare alcuni elementi che possono essere presi in considerazione nel prendere una decisione sul numero di dimensioni da utilizzare: bontà di adattamento, interpretabilità della configurazione, facilità di utilizzo, equilibrio tra parsimonia delle dimensioni e stabilità (Kruskal, 1978).

- <u>Bontà di adattamento</u>. L'incremento del numero di dimensioni consente di aumentare l'adattamento tra dati e modello; ciò contrasta con lo spirito dei metodi per l'analisi dimensionale e di strutture latenti che è quello di rappresentare gli oggetti in un numero minimo di dimensioni (in genere due). In generale, se per riprodurre la matrice delle distanze si definisce un numero di dimensioni:
  - *basso* (2 o 3), la rappresentazione degli oggetti nello spazio è più facile e più interpretabile ma può registrare un basso adattamento;
  - *alto*, la rappresentazione spaziale risulterà essere di difficile interpretazione anche se registrerà un alto adattamento; nell'adattamento potrebbe rientrare anche quello al rumore.

---

[9] Per inciso, si ricordi che dal punto di vista matematico per spiegare perfettamente le prossimità/distanze tra le 32 città sarebbe necessario una configurazione con 31 dimensioni!

In fase esplorativa può essere utile verificare la bontà di adattamento sugli stessi dati per diverse ipotesi di dimensionalità; successivamente i valori registrati dalle diverse *objective function* in corrispondenza dei diversi livelli di dimensionalità; la dimensione appropriata è individuata in corrispondenza del punto in cui la linea che lega i valori si appiattisce; tale grafico, interpretabile come lo *scree plot* (visto in precedenza) consente di decidere quante dimensioni occorrono per rappresentare gli oggetti e spiegare le distanze nel modo migliore. Per evitare decisioni in favore di soluzioni degenerate occorre verificare che in tale grafico si verifichino le seguenti condizioni:

- all'aumentare della dimensionalità deve aumentare l'adattamento,
- i punti devono formare un modello convesso: la linea che collega due punti qualunque deve essere sempre al di sopra dei punti intermedi.

▪ Chiarezza e interpretabilità. Quando la configurazione finale non risulta essere chiara è possibile effettuare diverse analisi con un numero diverso di dimensioni ed esaminare le configurazioni finali; è possibile che, applicando tale approccio, emergano più soluzioni interpretabili. Nel caso in cui i punti nella configurazione non consentono di individuare alcuna struttura, i dati possono essere considerati casuali.

▪ Facilità di utilizzo. Anche se le configurazioni con un numero di dimensioni maggiore presentano una maggiore significatività statistica, utilizzare configurazioni bi-dimensionali consente una migliore e più facile gestione dei risultati. Le configurazioni con molte dimensioni sono probabilmente utili solamente quando si utilizzano contemporaneamente altre tecniche supplementari finalizzate a identificare strutture comprensibili e interessanti.

▪ Stabilità. Se la dimensionalità risulta essere stabile attraverso diverse applicazioni, allora tale soluzione può essere considerata valida. Ciò vale anche nel caso in cui la configurazione presenta un valore di *stress* non molto basso: la consistenza della soluzione messa alla prova dalle diverse configurazioni giustifica la sua adozione. Una possibile strategia per verificare la stabilità della configurazione ottenuta può essere quella di suddividere casualmente i casi in due gruppi e di confrontare i risultati ottenuti dalla applicazione dell'analisi a ciascuno di essi. Si ricordi, comunque, che l'osservazione della stabilità non implica necessariamente la generalizzabilità della soluzione e del modello osservato.

5.1.3 Alcune varianti analitiche

Di seguito sono presentate alcune varianti analitiche al MDS visto.

➢ Analisi delle coordinate principali. Quando la relazione tra prossimità e distanze è di tipo lineare è possibile adottare il metodo delle coordinate principali; questo tipo di approccio è molto vicino a quello delle *componenti principali* (Maggino, 2005), del quale può essere considerata una variante; infatti si può affermare che il metodo delle coordinate principali non è altro che quello delle componenti principali applicato non ad una matrice di correlazione o di covarianza ma ad una matrice i cui elementi sono rappresentati da misure di prossimità (Maggino, 2005). L'identificazione delle coordinate principali è equivalente all'identificazione delle componenti principali se la misura di somiglianza è proporzionale alla distanza euclidea. L'applicazione del metodo consente di determinare il numero di dimensioni sottostanti alla matrice di somiglianza (per esempio verificare le dimensioni che i soggetti utilizzano per giudicare la somiglianza tra un particolare insieme di oggetti).

➢ Analisi *procruste* [10]. Questo approccio consente di confrontare configurazioni ottenute

[10] E' forse l'unica tecnica statistica il cui nome deriva dalla mitologia greca. Procuste, nome che in greco significa "il tenditore", allungava la gente come un fabbro allunga il ferro battendolo con il martello; secondo altre fonti Procuste, con lo stesso significato di *procoptas*, era soltanto l'epiteto, di Damaste, "il costrittore". Altri ancora raccontavano che il fabbro apportatore di morte viveva sul monte Coridallo sul quale conduceva la Via Sacra che andava da Eleusi a Atene

applicando metodi diversi sugli stessi dati di prossimità in uno spazio euclideo (il confronto può avvenire anche tra coppie di punti); tale analisi utilizza particolari tecniche che dilatano, trasferiscono, riflettono e ruotano una delle due configurazioni di punti in modo da farla convergere e coincidere il più possibile con l'altra;

- Approcci probabilistici. Si tratta di modelli che tentano di trasformare l'approccio multidimensionale da esplorativo/descrittivo ad induttivo/confermativo. Uno tra tali modelli utilizza particolari stime di massima verosimiglianza per:
  - verificare la significatività del numero di dimensioni,
  - definire regioni di confidenza per gli oggetti e per gli individui.

  Per poter applicare tali modelli in modo corretto è necessario soddisfare alcuni assunti piuttosto restrittivi riguardanti le prossimità (devono essere misurate su scala a rapporti e devono essere tra loro statisticamente indipendenti).

5.1.4 Multidimensional Scaling e sue estensioni

Il *MDS* ha trovato estensioni in modelli di analisi più complessi (Kruskal, 1978). Eccone alcune.

Multi-Dimensional Scaling delle differenze individuali

Tale metodo è finalizzato all'analisi contemporanea di più matrici di preferenza/prossimità; ciascuna delle matrici contiene i valori di preferenza/prossimità indicati da un soggetto rispetto ad un gruppo di elementi; in pratica la matrice sottoposta ad analisi è unica ma tridimensionale. Tale approccio assume una differenza significativa nei soggetti nell'esprimere giudizi di preferenza/prossimità. In particolare si assume che la differenza e la variabilità osservata tra le preferenze individuali sia spiegabile dalla presenza di dimensioni latenti; in altre parole si assume che esista uno spazio di riferimento comune a tutti gli individui (o gruppi di individui) e che le differenze tra gli individui siano espresse da un diverso peso assegnato alle dimensioni dello spazio comune[11]. Questo approccio può essere applicato anche per studiare la presenza di strutture latenti in un gruppo in diversi momenti (indagini longitudinali) o in condizioni diverse (ricerche sperimentali).
L'output della procedura è rappresentato sia dalla matrice delle coordinate degli oggetti che dalla matrice dei casi.
La versione più nota di questo approccio è chiamata *INDSCAL*, dal nome del programma predisposto nel 1970 da Carrol e Chang, definito come *modello euclideo ponderato* (*weighted euclidean model, WEM*). Dal punto di vista matematico tale modello non introduce alcun elemento innovativo rispetto al modello metrico; è possibile infatti dimostrare come il modello classico rappresenti un caso particolare del modello euclideo ponderato. In ciascuna delle S matrici sono registrate le prossimità tra gli oggetti i e j secondo l'individuo s $(\delta_{ij,s})$. La relazione ipotizzata tra prossimità e distanze è quella già osservata per il modello metrico:

$$d_{ij,s} = \delta_{ij,s} = \sqrt{\sum\left(x_{ir,s} - x_{jr,s}\right)^2}$$

---

nell'Antica Grecia. Lì lavorava Procuste con i suoi strumenti tra i quali non c'era un'incudine normale ma un letto scavato nella roccia, o fatto da un fabbro, sul quale egli stendeva i viandanti per "lavorarli" con il suo martello; poiché il letto era sempre troppo grande egli doveva allungare colui che vi giaceva. Altre fonti successive riportano che egli avesse avuto due letti, uno grande e uno piccolo: nel grande obbligava i piccoli allungandoli, in quello corto gli altri segando loro le estremità sporgenti. (da K.Kerényi, *Gli dei e gli eroi della Grecia*, 1958, trad.it. 1963 Il Saggiatore, Milano, 1976, Garzanti, Milano).

[11] Questo metodo è particolarmente adatto all'analisi del differenziale semantico, tecnica di rilevazione molto usata nei questionari attraverso la quale ciascun individuo attribuisce un significato e un'importanza a coppie di aggettivi in funzione del modo in cui si sono formate le proprie convinzioni politiche, sociali, culturali del proprio gruppo di appartenenza.

Questo approccio utilizza una diversa misura di adattamento (percentuale di varianza spiegata, anziché valore dello *stress*); contrariamente al metodo di base, in questo caso l'aggiunta di una dimensione non sempre consente di aumentare in modo significativo la percentuale totale di varianza spiegata[12]. Sia l'impostazione che l'interpretazione di questo approccio è comunque piuttosto complessa.

Scaling MDS ripetuto (_RMDS_)

Questo metodo si applica in situazioni simili alla precedente con la differenza che la terza dimensione riguarda osservazioni ripetute sulle prossimità.

Scaling MDS asimmetrico (_AMDS_)

Tale metodo si applica quando la matrice di prossimità è asimmetrica, ovvero quando la prossimità tra *a* e *b* è diversa dalla prossimità tra *b* e *a*. L'asimmetria può dipendere dalla natura del fenomeno analizzato (per esempio il primo oggetto influisce sul secondo); l'*Asymmetric MultiDimensional Analysis* (*AMDS*) consente di inserire nel modello geometrico tale asimmetria. Anche in questo caso si assume che le distanze tra gli stimoli siano funzione delle dissomiglianze ma l'obiettivo è quello di determinare per ciascuno stimolo un peso diverso per ciascuna delle dimensioni. Uno degli approcci analitici che possono essere applicati nell'ambito di questo metodo è quello detto *ASYMSCAL*: oltre alla matrice che contiene le coordinate degli oggetti, l'output prevede anche una seconda matrice contenente i pesi per ciascuno degli stimoli.
Nell'applicare tale metodo è importante tenere presente che l'asimmetria può essere anche prodotta da errori di misurazione (fluttuazioni nelle misure); in questo caso l'asimmetria non può essere presa in considerazione nel modello geometrico; è possibile rimediare a tale asimmetria calcolando, per esempio, la media tra i valori al di sopra e quelli al di sotto della diagonale principale; la nuova matrice risulterà essere così simmetrica e potrà essere sottoposta all'approccio classico.

Scaling MDS pesato (_WMDS_)

Questo approccio consente di inserire nel modello geometrico le differenze esistenti tra i casi quando sono considerate significative. Secondo tale modello per ciascun caso le distanze tra i punti sono funzione delle dissomiglianze e ciascuna distanza è ottenuta applicando specifici pesi per ciascuna dimensione. La strategia analitica più comunemente applicata è quella detta *Weighted MultiDimensional Scaling* (*WMDS*): l'output è rappresentato sia dalla matrice delle coordinate degli oggetti che dalla matrice contenente i pesi attribuiti agli oggetti per ciascun caso. La natura del modello geometrico è determinata dai contenuti delle matrici dei pesi. Se tutte le matrici dei pesi

- sono identiche, i risultati sono identici a quelli prodotti dal metodo *RMDS*;
- presentano valori diversi da zero solo nella diagonale principale, allora il risultato corrisponde al modello *INDSCAL*;
- sono simmetriche, allora i risultati corrispondono all'approccio *IDIOSCAL* o *GEMSCAL*.

[12] Sulla base di questo approccio generale sono stati sviluppati altri modelli *WEM* estesi, tra i quali si ricordano:
- *IDIOSCAL* (*Individual DIfference in Orientation SCAling*), che consente di considerare le dimensioni dello spazio come non ortogonali tra loro (e quindi correlate);
- *PINDIS* (*Procrustean INdividual DIfference Scaling*);
- *ALSCAL* (*Alternative Least squares SCALing*), che utilizza metodi alternativi alla tecnica dei minimi quadrati e consente di analizzare dati che sono misurati su qualunque tipo di scala, con osservazioni incomplete o dati *missing*, simmetrici o asimmetrici, con misure ripetute e non, continui e discreti;
- *SMACOF* (*Scaling by MAjorizing a COmplicated Function*), proposto da de Leeuw come metodo alternativo a quello dei minimi quadrati per la minimizzazione di un indice detto *S-Stress*;
- *CANDECOMP* (*CANonical DECOMPosition*), generalizzazione del modello *INDSCAL* a matrici più complesse;
- *PARAFAC-1* (*PARAllel profiles FACtor analysis*), modello che utilizza una matrice di covarianza.
- *PARAFAC-2* (*PARAllel profiles FACtor analysis*), caso particolare di *IDIOSCAL*.
- *CANDELINC* (*CANonical DEcomposition with LINear Constrain*), corrisponde al modello *CANDECOMP* con l'aggiunta di alcune restrizioni.
- *DEDICOM* (*DEcomposition into DIrectional COMponents*), analizza matrici di prossimità asimmetriche.

5.2 LA MAPPA DELLE PREFERENZE: L'UNFOLDING

Tra i diversi modelli sviluppati per la misura delle preferenze vi è l'***unfolding*** che ha l'obiettivo di rappresentare soggetti e stimoli in uno spazio comune – in genere unidimensionale – in modo tale che le distanze relative tra loro riflettano la prossimità psicologica tra stimoli e individui. L'approccio analitico, definito da Coombs (1950; McIver, 1979), consente di ricavare una scala unidimensionale di preferenza a partire dagli ordinamenti che i soggetti fanno degli oggetti.
Il procedimento prevede la somministrazione di una serie di stimoli che devono essere ordinati da ciascun soggetto secondo un criterio di preferenza. Ciascun ordinamento individuale è chiamato *scala I*.[13] L'assunto alla base del metodo indica che le dimensioni sottostanti l'ordinamento osservato possono essere determinate dopo aver identificato il *punto ideale* della scala sul quale il soggetto si colloca. Il punto ideale viene identificato in base alle disuguaglianze nell'ordinamento espresso dal soggetto. L'obiettivo è quindi quello di verificare se le diverse scale individuali *I* possono ritrovarsi in un'unica scala *J*.[14]
Nel caso di verifica positiva si può concludere che i soggetti impiegano un criterio comune di valutazione dei diversi stimoli. In caso contrario possono essersi verificate due diverse situazioni:
- nella valutazione degli stimoli i soggetti impiegano criteri multipli,
- i soggetti rispondono agli stimoli in modo casuale; non vi è alcun attributo sottostante.

Poniamo che due soggetti abbiano espresso le loro preferenze relativamente a cinque oggetti (stimoli) – *a, b, c, d, e* – e di poter rappresentare le loro preferenze su un'unica dimensione.
Nella seguente figura i singoli ordinamenti sono presentati in verticale con le linee I_1 e I_2, che riportano gli ordinamenti riferiti dai due soggetti, rispettivamente *cbade* e *decba*, mentre in orizzontale è posta la scala *J*.
Si assume che la "forza" della preferenza di ciascun soggetto su un'unica dimensione può essere rappresentata da una distribuzione normale. In tale modello, più lontano è un oggetto dalla media della distribuzione di preferenza di un soggetto, meno tale oggetto sarà preferito.

Se gli assiomi delle distanze (Maggino, 2005) risultano essere accettabili, allora la direzione non entra nel calcolo delle preferenze. E' possibile a questo punto procedere secondo due diverse prospettive di valutazione:
- Unfolding (distendere): secondo questa prospettiva è possibile utilizzare gli ordinamenti di preferenza individuali (scale *I*) per determinare la scala *J* (forza della preferenza). Nella figura si evidenzia come viene a determinarsi la scala *J*: si identificano le porzioni (dette rette di *unfolding*) delle scale I_1 e I_2 che inclinandosi definiscono la scala *J*. Questa mantiene l'essenziale integrità delle scale individuali *I* nel senso che un particolare stimolo risulterà più vicino ad un dato soggetto se questo lo ha preferito rispetto ad un altro. Osserviamo che per I_1
  - lo stimolo *c* è preferito allo stimolo *b*: sulla scala *J, c* è più vicino ad I_1 di *b*,
  - lo stimolo *d* è preferito allo stimolo *c*: sulla scala *J, d* è più vicino ad I_1 di *c*.

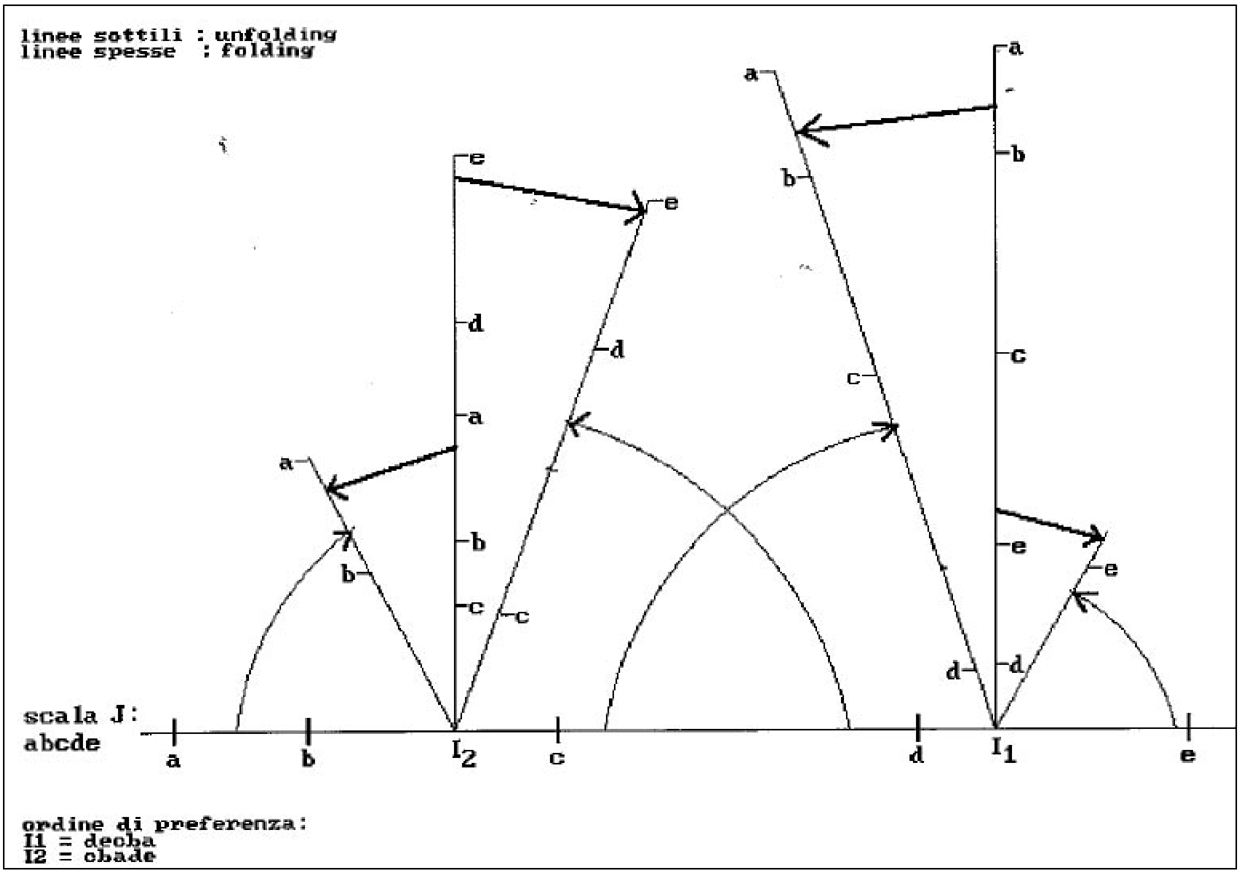
  Tale relazione preferenza-distanza può essere ritrovata anche per la scala individuale I_2.

[13] La matrice di input dell'*unfolding* è *two-mode two-ways*. L'elemento generico a_{ij} rappresenta quindi la preferenza espressa dal *j*-esimo individuo rispetto all'*i*-esimo oggetto. Caratteristica di questo modello è quella di combinare in un'unica rappresentazione spaziale i due *mode* della matrice: gli *N* oggetti (come succede per gli altri modelli) e i punti ideali degli *m* soggetti. Questa analisi è detta *joint space analysis* con riferimento alla individuazione dei punti ideali dei soggetti nella rappresentazione grafica.

[14] A tale proposito, il modello fa un'importante distinzione tra:
- scala *J* qualitativa, rappresentata dal semplice ordinamento degli stimoli (non è nota la distanza tra gli stimoli);
- scala *J* quantitativa, definibile quando dall'ordinamento degli stimoli è possibile inferire anche la loro distanza.

Quindi entrambi gli ordinamenti individuali possono essere *spiegati* sulla stessa dimensione.

- Folding (piegare): secondo questa prospettiva è possibile utilizzare la scala *J* per ricavare gli ordinamenti di preferenza individuali (scale *I*). Nella figura si evidenzia come sia possibile a partire dalla scala *J*, individuare porzioni (dette rette di *folding*) che possano essere *piegate* per ricomporre le scale (ordinamenti) individuali I_1 e I_2. Ciò viene fatto piegando la scala *J* rispetto al punto ideale che rappresenta ciascun individuo.



Nella figura le frecce aiutano ad identificare i due procedimenti: le frecce rette si riferiscono alla procedura di *unfolding* mentre le frecce curve si riferiscono alla procedura di *folding*.
In genere si utilizzano poche scale individuali (*I*) in quanto l'applicazione del modello risulta più complessa in presenza di un grande numero di scale *I* (McIver, 1979).

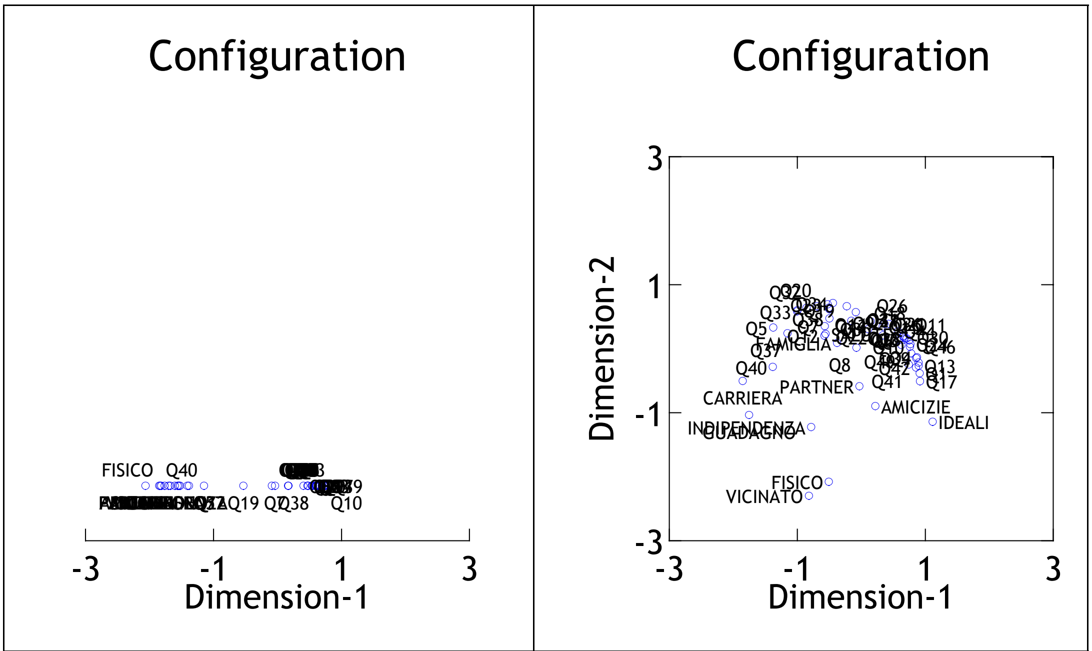
Il modello multidimensionale

Come si è visto, il modello di *unfolding* ha l'obiettivo di rappresentare – su un unico continuum metrico – stimoli e soggetti a partire dalle preferenze espresse dai soggetti stessi. Ciò assume che i soggetti nell'esprimere le preferenze per gli stimoli utilizzino un solo criterio.
Alcuni allievi di Coombs hanno esteso il modello unidimensionale al caso multidimensionale, applicabile quando si ipotizza che le preferenze siano espresse dai soggetti sulla base di più criteri.
La struttura teorica rimane la stessa anche se risulta, ovviamente, più complessa la struttura geometrica. L'operazione consiste nel posizionare i punti relativi agli stimoli e agli individui in uno spazio *R* dimensionale, usando le distanze, euclidee o non.
Poniamo che gli stimoli siano rappresentati da candidati alle elezioni e che i giudici siano elettori cui è stato chiesto di metterli in ordine di preferenza; se l'ideologia sarà l'unico criterio usato nel processo di valutazione e di giudizio, allora le preferenze dei giudici potranno essere rappresentate in uno spazio unidimensionale. Se invece gli elettori scelgono i candidati non solo sulla base di una ideologia politica ma anche in termini di attributi personali, di qualità professionali, ecc., sarà necessario individuare uno spazio multidimensionale.
Anche se il modello di *unfolding* multidimensionale è ben sviluppato, tutti gli algoritmi di verifica dell'adattamento del modello ai dati sono problematici. La difficoltà emerge in quanto per stimare

una grande quantità di informazioni (matrice $n*m$ delle coordinate dei punti-soggetto e matrice $k*m$ delle coordinate dei punti-stimolo) si utilizza una piccola quantità di informazioni (matrice $n*k$). Ne consegue che è possibile ottenere molte configurazioni dei punti che presentano un adattamento accettabile ai dati. Questo vuol dire che la stima di uno spazio congiunto multidimensionale deve essere considerata con una certa cautela in quanto è possibile ottenere soluzioni degenerate (presenza di minimo locale).

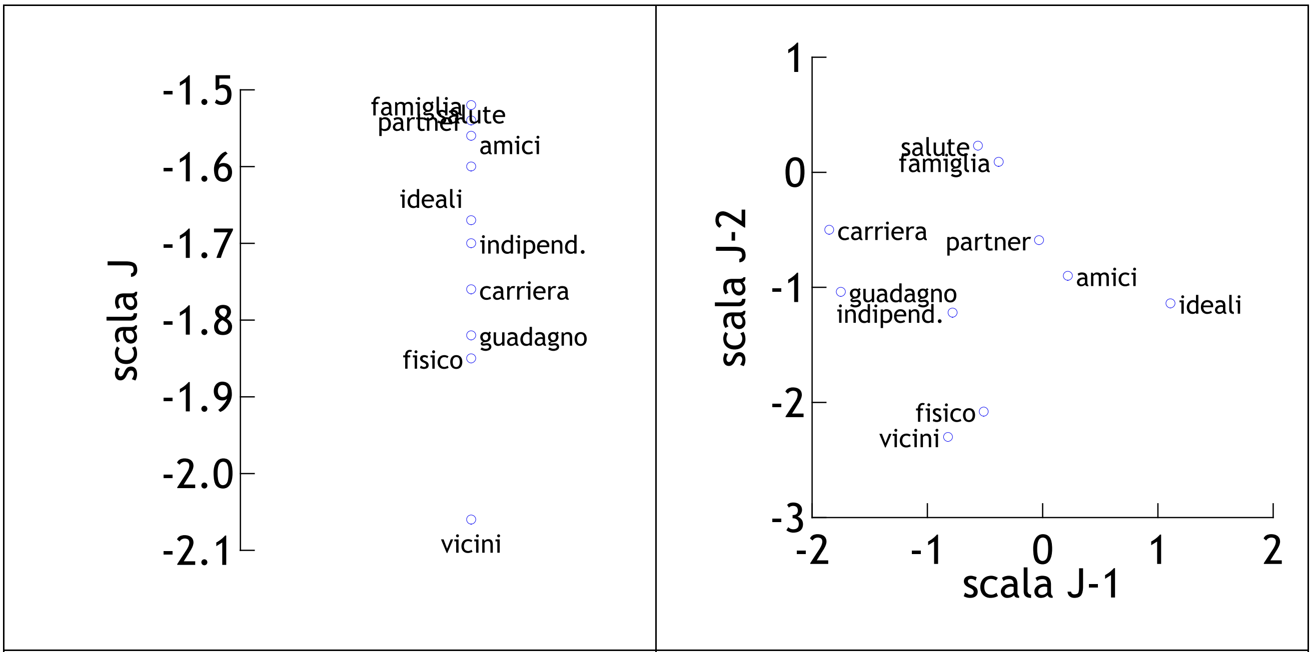
5.2.1 Un'applicazione

Riprendiamo i dati presentati nell'applicazione del modello Thurstone e relativi all'importanza che i soggetti attribuiscono ad alcuni ambiti della vita individuale. Abbiamo provato ad applicare il modello *unfolding* assumendo la presenza di uno spazio – nel quale collocare sia gli stimoli che i soggetti – prima unidimensionale e successivamente bidimensionale. In termini di adattamento del modello, il risultato è piuttosto soddisfacente sia nel caso unidimensionale (0.85872 la proporzione di varianza spiegata dalla configurazione finale) che nel caso bidimensionale (0.92826 la proporzione di varianza spiegata dalla configurazione finale).
Di seguito sono presentate le due diverse rappresentazioni ottenute collocando sia gli item che i soggetti prima sull'unico continuum (una scala *J*) e poi nello spazio bidimensionale (due scale *J*):



UNFOLDING UNIDIMENSIONALE	*UNFOLDING* BIDIMENSIONALE

Entrambe le rappresentazioni risultano naturalmente di difficile lettura per la presenza contemporanea di molti punti (dieci item e ca. cinquanta individui); per questo motivo può essere utile visualizzare entrambe le soluzioni solamente in relazione alla posizione degli item (a tal fine è stata applicata la stessa tecnica di rappresentazione utilizzata nel caso dei risultati ottenuti nell'applicazione del modello Thurstone):



UNFOLDING UNIDIMENSIONALE	*UNFOLDING* BIDIMENSIONALE

Mentre il risultato ottenuto con l'*unfolding* unidimensionale sembra essere perfettamente allineato a quello ottenuto dall'applicazione del modello Thurstone (v.), l'*unfolding* bidimensionale sembra aggiungere un nuovo elemento di interpretazione relativamente ai criteri di classificazione utilizzati dai soggetti intervistati (motivo per il quale tale modello spiega una maggiore porzione di varianza); infatti, mentre il continuum identificato dalla scala J-2 è molto simile alla scala J unidimensionale, quello identificato dalla scala J-1 sembra rappresentare una dimensione interpretabile in termini di *materiale* ←→ *spirituale*.

\- - - - -

A differenza di altri modelli di *scaling*, quali l'*additivo* o il *cumulativo*, l'*unfolding* non ha trovato grande applicazione nella ricerca sociale e risulta per questo sotto-utilizzato nonostante le sue caratteristiche potenzialmente potrebbero costituire un valido contributo alla misura e allo studio di dimensioni soggettive. Tra i motivi della scarsa applicazione possiamo trovare la difficoltà nella raccolta dei dati richiesti e di osservazione di scale *J* prevalenti, la mancanza di misure di bontà di adattamento e di affidabilità. Risulta inoltre difficile trovare *package* statistici che consentano l'analisi in presenza di molti stimoli.

6. UN MODELLO PER L'ANALISI DELLE PREFERENZE: LA CONJOINT ANALYSIS

L'analisi congiunta rappresenta un modello[1] utilizzato per analizzare come i soggetti sviluppano le proprie preferenze. Essa si basa sulla semplice premessa che gli individui valutano il valore di un oggetto (prodotto/servizio/idea), reale o ipotetico, combinando quantità separate di valori relativi a ciascun attributo.
In altre parole, il modello assume che le preferenze individuali siano frutto della combinazione di valutazioni che riguardano attributi diversi e distinti.
Il modello di misurazione su cui si basa l'analisi ha l'obiettivo di decomporre e valutare i singoli attributi che hanno condotto alla scelta di preferenza dell'individuo, a differenza di altri modelli di *scaling*, che consentono di ricostruire una misura sintetica per ciascun individuo relativamente al costrutto misurato.
Pur essendo nato e sviluppato in ambiti sperimentali in cui è coinvolto il comportamento umano, il modello congiunto ha successivamente trovato vaste applicazioni come quelle riguardanti ambiti in cui vengono studiati modelli individuali di decisione (Louviere, 1988; Malhotra, 1993).
Fin dalla metà degli anni 70, l'analisi congiunta ha attirato molta attenzione come metodo che consente di descrivere le decisioni e le preferenze che i consumatori manifestano tra diversi prodotti, servizi, ecc. Per questo l'analisi congiunta ha guadagnato molti favori nel mondo dell'industria. Durante gli anni 90, le applicazioni dell'analisi congiunta sono aumentate ulteriormente anche in altri campi di studio. L'utilizzazione nelle ricerche di *marketing* nello sviluppo di nuovi prodotti ha condotto alla sua adozione anche ad altri campi.
Ciò ha coinciso da una parte con lo sviluppo di metodi alternativi finalizzati alla misura delle scelte da parte dei consumatori e dall'altro alla diffusione di programmi computerizzati che consentivano di integrare l'intero processo, dalla generazione delle combinazioni dei valori delle variabili indipendenti che devono essere valutati per creare i simulatori di scelta al fine di prevedere le scelte dei consumatori attraverso un vasto numero di formulazioni alternative di prodotti o servizi.
L'analisi congiunta è particolarmente adatta per comprendere e studiare le reazioni e le valutazione dei consumatori rispetto a combinazioni predeterminate di attributi; tali combinazioni rappresentano potenziali prodotti o servizi. L'obiettivo di tali studi è quello di sviluppare prodotti che presentino attributi che idealmente si adattano alle preferenze dei consumatori.

Il modello assume che alla base di una scelta di preferenza vi sia il concetto di *utilità*. L'utilità rappresenta un giudizio soggettivo di preferenza unico per ciascun individuo. [2]
Nell'analisi congiunta si assume che l'utilità sia basata sul valore posto su ciascuno dei valori degli attributi ed espresso in una relazione che riflette il modo in cui l'utilità è formulata per ciascuna combinazione di attributi. E' possibile sommare i valori di utilità associati con ciascuna caratteristica di un oggetto per giungere alla determinazione dell'utilità totale. E' quindi possibile assumere che l'oggetto con i valori di utilità più alti sia anche il più preferito e presenta la migliore possibilità di essere scelto.
La *conjoint analysis* è forse l'unico metodo di analisi statistica per quale si richiede che il ricercatore costruisca un insieme (reale o ipotetico) di oggetti attraverso la combinazione di valori degli attributi. Tali combinazioni vengono quindi presentate ai soggetti che esprimono delle valutazioni totali. Questo vuol dire che i soggetti si trovano in una situazione molto realistica

[1] Com'è accaduto per il modello fattoriale, anche il metodo di verifica del modello congiunto si è successivamente diffuso come metodo di analisi generale multivariata; l'analisi congiunta rappresenta una famiglia di tecniche e di metodi, tutti teoricamente basati sui modelli di integrazione dell'informazione e sulla misurazione funzionale.

[2] Per semplificare la presentazione, da questo punto in poi utilizzeremo il termine "oggetto" per indicare un prodotto, un servizio, un'idea, ecc

(ciascun soggetto sceglie tra i diversi oggetti presentati).
Dato che il ricercatore ha costruito gli oggetti ipotetici in un particolare modo, è possibile determinare, a partire dagli ordinamenti dei soggetti, l'influenza di ciascun attributo e di ciascun valore di ciascun attributo sul giudizio di utilità espresso.
Perché l'applicazione dell'analisi congiunta possa avere esito positivo, il ricercatore deve essere in grado di descrivere l'oggetto in termini sia di attributi che di valori rilevanti per ciascun attributo. Ogni specifico attributo è detto *fattore*. I valori possibili per ciascun fattore sono detti *livelli*.

Caratteristiche del modello

Sono tre le caratteristiche statistiche che contraddistinguono il modello congiunto (Hair, 1998; Louviere, 1988; Malhotra, 1993):

- Natura decompositiva. Il modello congiunto è detto *decompositivo* in quanto la preferenza totale del soggetto relativamente all'oggetto, creato attraverso la specificazione dei valori (livelli) per ciascun attributo (fattore), viene decomposta per determinare l'importanza da attribuire a ciascun attributo, attraverso la verifica del modello. Tale approccio è detto lineare in quanto basato sulla combinazione lineare degli effetti dei fattori (variabili indipendenti) sulla scelta del soggetto (variabile dipendente); l'obiettivo è quello di sviluppare un modello predittivo.
- Verifica e stima del modello a livello individuale. L'originalità del modello congiunto sta nella possibilità di definire un modello di previsione delle preferenze per ciascun soggetto separatamente (stime disaggregate); è comunque possibile fare stime anche per gruppi omogenei, rispetto a determinate caratteristiche, di individui (stime aggregate). La decisione sul livello di stima da adottare dipende dalle informazioni che se ne vogliono trarre. La verifica del livello di accuratezza è fatta sulla base degli errori di previsione (aggregata o disaggregata) o dell'analisi dei residui.
- Flessibilità, dell'approccio alla verifica è dato da:
  (1) la capacità di utilizzare variabili dipendenti metriche e non metriche,
  (2) l'utilizzo delle variabili predittive di tipo categorico,
  (3) l'adozione di assunti, riguardanti le relazioni tra le variabili indipendenti e la variabile dipendente, generalmente accettabili. In pratica la verifica del modello congiunto consente di prevedere separatamente gli effetti per ciascun livello della variabile indipendente senza assumere che siano correlati tra loro. Inoltre il modello additivo può essere anche di tipo non-lineare (come nel caso in cui si preveda uno *scaling* curvilineo in cui dopo aver aggiunto un termine positivo se ne aggiunge uno negativo, e poi uno positivo, ecc.).

Applicazione

Dal punto di vista pratico il procedimento richiede che il ricercatore individui, per l'oggetto di interesse, i *fattori* che consentono di descrivere un oggetto; per ciascun fattore occorre quindi definire i *livelli*, ovvero i valori che il fattore può assumere. Successivamente si procede alla definizione ipotetica di più versioni dell'oggetto ottenute combinando i diversi valori (livelli) individuati per tutti gli attributi (fattori) definiti. Ciascuna combinazione è spesso definita *scenario*.
Gli scenari vengono presentati ai soggetti identificati che devono fornire le proprie valutazioni; è possibile richiedere due diverse forme di valutazione:
- *ranking*: si chiede a ciascun soggetto di ordinare gli scenari in ordine di preferenza,
- *rating*: si chiede a ciascun soggetto di indicare il proprio livello di preferenza su una scala di *rating*.

Se il ricercatore ha costruito gli scenari creando specifiche e appropriate combinazioni di fattori e livelli, l'analisi delle preferenze consente di individuare le ragioni che hanno guidato nelle preferenze e di comprendere la struttura di preferenza dei soggetti; in particolare, l'obiettivo è quello di determinare
- l'importanza e il peso di ciascun fattore nella decisione totale,
- quanto i diversi livelli di ciascun fattore influenzano la formazione della preferenza totale

(utilità).
In particolare, l'utilità totale, o *total worth*, espressa da un soggetto rispetto ad un oggetto, è composta da valori parziali, detti *part-worth*, relativi a ciascun livello per ciascun fattore. Il modello congiunto può essere quindi così formalizzato:

$$total \cdot worth = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(part - worth_{ij} \right)$$

dove
m numero di fattori individuati
n numero di livelli per ciascun fattore (valore quindi che varia per ciascun fattore).

Con le stime dei valori parziali è possibile stimare la preferenza di un soggetto per qualsiasi tipo di combinazione dei fattori. Inoltre la struttura di preferenza potrebbe rivelare qual è (o quali sono) il fattore (o i fattori) che determinano l'utilità totale e la scelta finale.
E' possibile anche combinare le scelte di più soggetti per rappresentare un ambiente detto "competitivo".

6.1 LA DEFINIZIONE E LA VERIFICA DEL MODELLO

6.1.1 Definizione degli obiettivi

Il principale obiettivo dell'analisi è quello di studiare le decisioni e le preferenze individuali
- determinando i contributi delle variabili indipendenti (fattori) e dei corrispondenti livelli,
- stabilendo un modello che risulti valido per poter predire i giudizi individuali relativi alla possibile accettazione di una combinazione di attributi, anche quelli inizialmente non valutati dall'individuo.

Ciò richiede che l'obiettivo di ricerca sia formulato intorno a due principali questioni:
1. *Definizione dell'utilità totale dell'oggetto*: per rappresentare in modo accurato il processo di preferenza individuale è necessario che vengano individuati e descritti tutti gli attributi dell'oggetto studiato e che ne determinano l'utilità e il valore; gli attributi da prendere in considerazione dovrebbero potenzialmente aggiungere o togliere qualcosa all'utilità totale dell'oggetto; in altre parole, è necessario considerare sia fattori positivi che fattori negativi; la mancanza di uno dei due tipi potrebbe dare risultati molto distorti.
2. *Specificazione dei fattori determinanti*: ciò richiede l'identificazione dei criteri-chiave coinvolti nel processo di scelta. Il rischio è quello di considerare importanti molti fattori che però si rivelano non in grado di differenziare le scelte in quanto non variano sostanzialmente tra gli oggetti[3].

6.1.2 Disegno

6.1.2.1 Selezione della metodologia di analisi

Dopo aver identificato e definito gli attributi che costituiscono l'utilità dell'oggetto, è necessario scegliere il tipo di modello congiunto più adatto; a tale proposito è possibile identificare tre diversi

[3] Si pensi a tale proposito all'acquisto di un'automobile: la sicurezza è sicuramente un fattore importante nella scelta ma non risulta essere determinante in quanto praticamente tutte le automobili rispondono agli standard di sicurezza definita.

approcci che si differenziano tra loro rispetto a 'numero massimo consentito di attributi', 'livello di analisi' e 'forma di modello permessa' (Hair, 1998; Louviere, 1988; Malhotra, 1993):

Caratteristiche	*Metodologia*		
	Tradizionale	*Adaptive*	*Basato sulla scelta*
Numero massimo di attributi	9	30	6
Livello di analisi	individuale	individuale	aggregato
Forma del modello	additivo	additivo	Additivo+effetti di interazione

- Approccio *tradizionale*: semplice modello additivo contenente fino a 9 fattori stimati a livello individuale.
- Approccio *adaptive*: modello sviluppato per poter gestire un grande numero di fattori (anche fino a 30) stimati a livello individuale.
- Approccio *basato sulla scelta*: definito per stimare i fattori a livello aggregato in forma additiva, tenendo conto anche delle interazioni.

6.1.2.2 Disegno dello stimolo: selezione e definizione dei fattori e dei livelli

Un momento particolarmente importante è quello del disegno degli stimoli ciascuno dei quali è definito da una diversa combinazione di livelli per tutti i fattori (Hair, 1998; Louviere, 1988; Malhotra, 1993).

Specificazione dei fattori[4]

- Misurazione. I fattori devono presentarsi in forma *comunicativa*, utilizzando lo strumento espressivo più appropriato (parole, figure, ecc.), e *operativa*, (i fattori devono essere distinguibili e non devono rappresentare concetti vaghi).
- Numero di fattori. Il numero di fattori da includere nel modello influenza direttamente il livello di validità e affidabilità. All'aumentare del numero di fattori e di livelli, aumenta il numero degli stimoli da definire e, conseguentemente, dei parametri da stimare. Il numero minimo di stimoli che deve essere valutato da un soggetto, è dato da

  (numero minimo di stimoli) = (numero totale di livelli per tutti i fattori) – (numero di fattori)+1
- Multicollinearità[5]. La multicollinearità tra fattori denota una perdita di indipendenza concettuale dei fattori; in questi casi si può creare attributi più complessi, che combinino gli aspetti degli attributi correlati; in ogni caso è più efficiente definire i disegni sperimentali e tecniche di stima per fattori ortogonali.

[4] Nelle applicazioni del modello congiunto in indagini di mercato occorre tenere presente un particolare fattore rappresentato dal prezzo, che va considerato come componente distinta dal valore. Quando si decide di inserire tale fattore, occorre essere consapevoli dell'impatto che il suo inserimento ha nella verifica del modello e della sua interpretazione.

[5] Questa rappresenta quanto di una variabile può essere previsto o spiegato dalle altre variabili impiegate nell'analisi. Per questo l'inserimento di variabili concettualmente irrilevanti può avere molti effetti pericolosi, anche se le variabili aggiuntive non immettono direttamente errori sistematici nei risultati del modello.
Nell'analisi di regressione multipla, la *multicollinearità* esiste quando due o più variabili indipendenti sono molto correlate; ciò rende difficile se non impossibile determinare i loro effetti separati sulla variabile dipendente.

Specificazione dei livelli

La definizione dei livelli rappresenta un momento importante del modello congiunto in quanto rappresentano le misure reali utilizzate per identificare gli stimoli. Nel definire i livelli occorre tenere presente i seguenti fattori:

- Misurazione. I livelli devono essere rappresentati in forma comunicativa, operativa, accettabile, rilevante dal punto di vista pratico.
- Numero di livelli. Il numero di livelli deve essere bilanciato tra i diversi fattori (è stato dimostrato che l'importanza relativa stimata aumenta all'aumentare del numero di livelli).
- Variabilità. La specificazione dei livelli dovrebbe condurre a visioni che siano realistiche (evitando quelle solo positive o favorevoli).

6.1.2.3 Definizione della forma del modello

Come si è detto, il principale obiettivo dell'analisi congiunta è quello di individuare e spiegare la struttura di preferenza individuale a partire delle valutazioni totali di un gruppo di stimoli, In questa prospettiva il ricercatore deve prendere due decisioni riguardanti:

- il disegno degli stimoli (regola di composizione), in termini di relazione tra fattori (modello additivo e di interazione),
- l'analisi delle valutazioni individuali, in termini di relazione tra fattori e tra *part-worth*.

Il disegno degli stimoli

In pratica occorre ipotizzare come i fattori sono tra loro correlati nel processo di decisione di ciascun soggetto. Il *modello additivo* è il più comune e descrive la situazione in cui il soggetto, nell'attribuire all'oggetto un valore totale per una determinata combinazione di attributi, somma i valori parziali di ciascun attributo. Quando si assume che determinate combinazioni di livelli pesano più/meno della loro semplice somma, si parla di effetti *additivi e interattivi*. L'aggiunta di effetti interattivi al modello può far diminuire la potenza predittiva in quanto la riduzione dell'efficienza statistica (presenza di più *parth-worth* da stimare) non è compensata dall'incremento della potenza predittiva dovuto all'interazione. In genere gli effetti interattivi spiegano meno varianza di quelli additivi.

La scelta tra le diverse forme determina il tipo e il numero di trattamenti o di stimoli che il soggetto deve valutare. La forma additiva richiede meno valutazioni da parte del soggetto e consente di eseguire stime in maniera più semplice. D'altra parte la forma che considera anche le interazioni potrebbe rappresentare, in molti casi, in maniera più accurata le reali valutazioni dei soggetti.

L'analisi delle valutazioni individuali

In pratica occorre ipotizzare come sono tra loro correlati i livelli dei fattori. Sono ammesse tre forme di relazione:

a) Relazione lineare, la forma più semplice e la più restrittiva; con essa si stima solamente un singolo *part-worth* che viene poi moltiplicato per il valore del livello per arrivare a *parth-worth* separati per ciascun livello;
b) Relazione quadratica: è descritta da una semplice relazione curvilinea;
c) Valori parziali separati: rappresenta la forma più generale; consente stime separate per ciascun livello con la conseguenza di avere il più alto numero di valori stimati.

La scelta del tipo di relazione dovrebbe essere fatta sulla base del modello concettuale e dell'ipotesi di partenza. Operare la scelta empiricamente, sulla base dei risultati che si ottengono, può sicuramente incrementare la capacità predittiva del modello ma ne riduce la validità.

6.1.3 Modalità di raccolta dei dati

Dopo aver specificato i fattori e i livelli ed aver definito la forma base del modello, si procede alla scelta del tipo di presentazione degli stimoli, il tipo di risposta e il metodo di raccolta dei dati. L'obiettivo è quello di presentare a ciascun individuo le combinazioni di attributi nella maniera più realistica ed efficiente possibile (Hair, 1998; Malhotra, 1993). La presentazione può utilizzare tecniche diverse (descrizioni verbali o grafiche).

6.1.3.1 Metodi di presentazione

- *Metodo "Trade-Off"*: si confrontano due attributi per volta ordinando tutte le combinazioni dei livelli.
  Vantaggi: semplice da somministrare.
  Svantaggi: l'utilizzazione di soli due fattori per volta rende la presentazione poco realistica e concreta; l'elevato numero di giudizi richiesto all'intervistato può favorire la tendenza dei soggetti al *response-set*; non consente l'utilizzo di tecniche grafiche; consente l'utilizzo delle sole risposte non-metriche.
- *Metodo "Full-Profile"*: gli stimoli, descritti separatamente; vengono ordinati o valutati dall'intervistato.
  Vantaggi. Utilizzando un livello per ciascun fattore, consente una descrizione realistica dello stimolo; consente di ridurre il numero dei confronti; il metodo consente anche di utilizzare più tipi di giudizi di preferenza (metrici e non-metrici).
  Svantaggi. L'utilizzo di un alto numero di fattori può indurre il soggetto a semplificare il processo di decisione, che si concentra solo su pochi fattori (sono infatti consigliati non più di sei fattori); un elemento che può influenzare le scelte e le risposte dei soggetti è rappresentato dall'ordine di presentazione dei fattori.
- *Metodo "pairwise combination"*. Confrontando due stimoli, il soggetto indica su una scala di *rating* il livello di preferenza di un profilo rispetto all'altro. Rispetto al precedente metodo, i profili non contengono tutti gli attributi ma vengono definiti selezionando solo alcuni attributi alla volta; rispetto al primo metodo i confronti non avvengono tra singoli attributi ma tra profili con molti attributi.

6.1.3.2 Creazione degli stimoli

Una volta selezionati fattori e livelli e scelto il metodo di presentazione, è necessario creare gli stimoli o i profili che i soggetti devono valutare (Hair, 1998; Malhotra, 1993).
Metodo "Trade-Off": vengono utilizzate tutte le possibili combinazioni degli attributi. Il numero delle matrici (M) dipende dal numero di fattori (m):

$$M = \frac{m(m-1)}{2}$$

tenendo presente che ogni matrice riguarda un numero di risposte uguali al prodotto dei livelli dei fattori.
Metodo "Full-Profile" e *metodo "pairwise combination"*: richiedono la valutazione di un singolo stimolo (primo metodo) o di una coppia di stimoli (secondo metodo). Con un numero basso di fattori e di livelli, è possibile chiedere ai soggetti di valutare tutti i possibili stimoli (*factorial design*). Con un numero elevato di fattori e livelli il disegno fattoriale non è praticabile; in questi casi, può essere opportuno suddividere tutti gli stimoli in sottogruppi (*fractional factorial design*).

6.1.3.3 Selezione di una misura di preferenza

La misurazione delle preferenze può avvenire con i seguenti tipi di misure (Hair, 1998; Malhotra, 1993):
- *Ranking*, che si ottiene chiedendo al soggetto di ordinare gli oggetti in ordine di preferenza; rappresenta la misura più affidabile anche perché semplice da applicare soprattutto in presenza di numero basso di stimoli; presenta dei problemi al momento della somministrazione.
- *Rating*, che si ottiene chiedendo al soggetto di valutare gli stimoli su una scala di preferenza (in genere da 0 a 10); tale misura è piuttosto semplice da applicare, anche in presenza di un alto numero di stimoli e consente l'utilizzazione di tecniche di analisi quantitativa; d'altra parte questo tipo di misura può rendere meno discriminanti i giudizi dati dai soggetti.

Il metodo *trade-off* consente solo l'utilizzo di dati ordinali (*ranking*) mentre il metodo *full-profile* consente di dati sia ordinali (ranking) che metrici (rating).

6.1.3.4 Rilevazione

La tecnica di rilevazione maggiormente utilizzata è quella dell'intervista individuale. Nuove tecniche consentono di effettuare la rilevazione anche per posta o per telefono. Le nuove tecniche informatiche inoltre consentono ulteriori sviluppi in questo senso.

6.1.4 Stima del modello e valutazione dell'adattamento complessivo

6.1.4.1 Selezione della tecnica di stima

La selezione della tecnica di stima del modello dipende dal tipo di misura di preferenza adottata.
Le valutazioni di tipo ordinale (*rankiing*) richiedono un particolare approccio dell'analisi della varianza specifico per dati ordinali; tale approccio, conosciuto con il nome di *MONANOVA* (*MONotonic ANalysis Of VAriance*).
Le valutazioni di tipo numerico (*rating*) consentono l'utilizzo di molte tecniche di stima per i *parth-worth* per ciascun livello, comprese quelle basate sulla regressione.

6.1.4.2 Valutazione della bontà di adattamento del modello

L'obiettivo di questa fase dell'analisi è quello di accertare la consistenza delle previsioni del modello rispetto all'insieme delle valutazioni o delle preferenze, espresse in termini individuali o considerate a livello aggregato. Per i dati di tipo *rank-order* vengono utilizzati gli indici di cograduazione mentre per i dati metrici (*rating*) si utilizza il coefficiente di correlazione.

6.1.5 Interpretazione dei risultati

6.1.5.1 Livello di analisi

In genere l'interpretazione dei risultati può avvenire a due livelli (Hair, 1998; Malhotra, 1993)

- *Disaggregato*: si verifica l’adattamento del modello per ciascun soggetto; l’interpretazione riguarda l’esame le stime dei *part-worth*, valutando la loro dimensione e la posizione reciproca di tale valori per ciascun livello di un fattore. Più alto è il valore del *part-worth* (negativo o positivo), maggiore è l’impatto che il corrispondente livello ha sull’utilità totale; la rappresentazione grafica dei valori *parth-worth* dei diversi livelli per ciascun fattore consente di identificare la presenza di modelli.
- *Aggregato*: l’analisi adatta un unico modello ad un insieme di risposte; tale approccio ha senso solo se l’aggregazione si riferisce ad gruppo omogeneo rispetto al tipo di indagine svolta.

6.1.5.2 Valutazione

La valutazione dei può essere fatta a due livelli (Hair, 1998; Malhotra, 1993).
- *Valutazione dell’impatto di ciascun livello attraverso le stime dei parth-worth.*. come abbiamo visto la valutazione delle stime dei *parth-worth* viene fatta per ciascun fattore, attraverso la valutazione della loro dimensione e della loro rilevanza pratica e teorica. Per permettere confronti a livello sia intra-individuale (tra fattori diversi) che inter-individuale (tra livelli dello stesso fattore), i valori dei *parth-worth* vengono convertiti su una scala comune (per esempio a 100 punti).
- *Valutazione dell’importanza relativa degli attributi*. Per effettuare tale valutazione si procede nel modo seguente:
  1. determinazione del *range* dei *part-worth* per ciascun fattore,
  2. somma di tutti i *range* per tutti i fattori,
  3. calcolo del rapporto tra *range* di ciascun fattore e somma dei *range*; ciò rappresenta l’importanza del fattore rispetto agli altri (tale valore può essere convertito in percentuale).

  Maggiore è il punteggio di importanza di un fattore maggiore è la sua utilità totale. Ciò consente anche di fare confronti tra individui. Se un range è molto piccolo intorno a valori estremi (molto bassi o molto alti) è consigliabile eliminare il fattore dall’analisi.

6.1.6 Validazione dei risultati

E’ possibile identificare due livelli di validazione:
- *validazione interna*: verifica dell’adeguatezza del modello adottato (additivo o interattivo); in genere tale verifica è fatta confrontando, nello studio preliminare, il modello adottato con quello alternativo;
- *validazione esterna*: riguarda la capacità del modello di prevedere le scelte reali e in termini specifici la questione della rappresentatività campionaria; anche se non esiste alcuna valutazione dell’errore campionario nei modelli individuali, è necessario sempre assicurare la rappresentatività del campione; ciò è particolarmente importante quando i risultati vengono utilizzati a scopi di segmentazione o simulazione di scelte.

6.2 LE APPLICAZIONI

Di solito i risultati dei modelli congiunti, stimati a livello sia individuale che aggregato, sono utilizzati nei casi che richiedono la rappresentazione di processi decisionali. Le applicazioni più comuni del modello congiunto con le sue rappresentazioni delle strutture di preferenza individuali sono: la segmentazione, l'analisi di profittabilità e le simulazioni (Hair, 1998; Malhotra, 1993).

- **Segmentazione**. Una delle applicazioni più comuni dei risultati ottenuti a livello individuale riguarda l'individuazione di segmenti di soggetti che presentano valori simili di *worth* parziale o totale. Le utilità dei valori parziali stimati possono essere utilizzate, isolatamente o in combinazione con altre variabili (per esempio demografiche), per ottenere raggruppamenti di soggetti che si presentano simili nelle loro preferenze.

- **Analisi di redditività** (*profitability*). Un completamento alla decisione è rappresentato dall'analisi della redditività marginale che può rappresentare un supporto alla decisione riguardante l'oggetto proposto. Se si conosce il costo di ciascuna caratteristica, è possibile combinare il costo di ciascun "prodotto" con la porzione di mercato attesa e il suo volume di vendita per prevedere la sua applicabilità. Tale processo dovrebbe puntare ad una combinazione di pochi attributi con la maggiore profittabilità al fine di aumentare il margine di profitto in corrispondenza del basso costo dei componenti particolari. In questa analisi possono essere utilizzati i risultati sia a livello individuale che a livello aggregato.

- **Simulazioni**. A questo punto dell'analisi si conosce solo il *parth-worth* relativi degli attributi e l'impatto dei livelli specifici ma non si è ancora raggiunto il vero obbiettivo dell'analisi, ovvero predire la quota di preferenze che uno stimolo reale o ipotetico è in grado probabilmente di catturare nei diversi scenari di mercato. Tale analisi potrebbe essere indicata con il termine inglese "*what if*" ("e se"). La simulazione delle scelte prevede tre momenti:
  1. stima e validazione dei modelli per ciascun soggetto (o gruppo),
  2. selezione degli insiemi di stimoli da verificare secondo i possibili scenari competitivi,
  3. simulazione delle scelte di tutti i soggetti (o i gruppi) per uno specifico insieme di stimoli e previsione della quota di preferenze di ciascuno stimolo attraverso l'aggregazione delle loro scelte.

  Dopo l'aver stimato il modello congiunto, il ricercatore può specificare un certo numero di insiemi di stimoli per simulare le scelte dei consumatori. In questi casi la valutazione può riguardare 1) l'impatto dell'introduzione di un prodotto all'interno di un mercato esistente; 2) il potenziale aumentato a partire da strategie multiprodotto e multi-etichetta, compresa la stima di cannibalismo; 3) l'impatto dell'eliminazione di un prodotto dal mercato. In ogni caso il ricercatore fornisce un insieme di stimoli che rappresentano il mercato e le conseguenti scelte simulate dei rispondenti.

Nello ambito dello studio delle preferenze individuali, questo modello presenta quindi delle potenzialità degne di attenzione. A tale proposito può essere interessante osservare una particolare applicazione presentata in Maggino F. (2005) *Importanza delle dimensioni di qualità della vita nelle preferenze dei cittadini: un'applicazione sperimentale dell'analisi congiunta.* Firenze University Press, Archivio E-Prints.

7. CONFRONTO TRA MODELLI DI SCALING.

La validazione di una scala di autovaluzione dell'autosufficienza fisica in una popolazione anziana

L'esempio qui presentato riguarda la validazione di una particolare scala di autovalutazione dell'autosufficienza fisica negli anziani. Tale validazione ha riguardato un campione piuttosto ampio di anziani residenti in Toscana e in Molise, per un totale di 3389 soggetti[1]. L'obiettivo era quello di mettere a punto uno strumento che consentisse di misurare non tanto l'efficienza fisica oggettivamente valutata quanto il livello di autosufficienza nell'ambiente di vita e nelle attività quotidiane così come vissuto e percepito dall'anziano.

7.1 LA SCALA

La scala, messa a punto integrando gli item proposti dall'*Organizzazione Mondiale della Sanità* nel 1979, può essere considerata un test di *performance* per le capacità fisiche richieste nello svolgimento delle attività di base della vita quotidiana (batteria di item funzionali). Queste capacità possono essere fatte risalire a un unico fattore di idoneità fisica (*fitness*) a condurre una vita "piena". Per l'anziano, presumibilmente ritirato dal lavoro, è possibile restringere il significato della scala al *possesso di capacità fisiche* per lo svolgimento delle attività quotidiane; le capacità fisiche si manifestano nello svolgere, in piena autonomia, le funzioni essenziali di:

- cura della persona, che si identifica con l'igiene personale (lavarsi, fare il bagno, pettinarsi, farsi la barba);
- cura del proprio abbigliamento, che nella società moderna consiste soprattutto nel vestirsi e spogliarsi, in quanto la vera e propria cura degli abiti rappresenta un elemento di delega anche in situazioni di piena autonomia funzionale;
- preparazione del cibo: gli elementi connessi con quest'area possono essere delegati, ma l'incapacità completa fisica di prepararsi in caso di necessità un pasto caldo può rappresentare un grave disagio esistenziale;
- cura della casa: anche le faccende domestiche - pesanti e non - sono elementi delegabili il cui svolgimento è legato a fattori culturali e di genere; ma, come per la preparazione del cibo, l'incapacità a svolgerle diventa "handicap" in caso di perdita del familiare o della persona incaricata di queste funzioni;
- approvvigionamento dei generi di prima necessità, la "spesa": rappresenta un elemento a sé, per il quale valgono le stesse considerazioni fatte ai punti precedenti.

Presupposto di base, per ciascuna di queste capacità, è la maggiore o minore sicurezza nello spostarsi in casa e fuori nell'ambiente esterno con i propri mezzi, o con quelli del trasporto pubblico, il cui uso richiede una dose aggiuntiva di capacità fisica. Sono tutte capacità indispensabili per le necessità della vita quotidiana, da prendere in considerazione in assenza e in presenza di ostacoli (come pavimenti sconnessi, soglie, gradini, scale più o meno lunghe, isolamento geografico). Questi spostamenti sono stati visti anche come indicatori della capacità di utilizzare mezzi pubblici, condizione che rende possibili normali scambi e incontri sociali con familiari ed amici.
La costruzione e la validazione di un tale tipo di scala ha l'obiettivo di:

- costruire un indicatore sintetico di autosufficienza,
- rilevare e valutare con un'accuratezza nota i soggetti che presentano un livello critico di disabilità; ciò viene fatto suddividendo il punteggio in livelli critici (valori-soglia) che, facendo riferimento al carico assistenziale (tipo e frequenza dell'aiuto richiesto), consentono l'individuazione e la definizione di diverse aree di non autosufficienza fisica.

Gli item sono stati scelti in modo da coprire tutti gli aspetti di base dell'attività quotidiana;

[1] Tale validazione ha coinvolto più ricerche (Tesi, 1990 e 1993).

successivamente gli item sono stati selezionati e ordinati secondo il modello scalare legato al criterio della diversa difficoltà; ciò consente di ottenere una valutazione in cui le condizioni per superare un item sono "teoricamente" richieste per superare l'item precedente. Ogni item deve avere la capacità di dividere il gruppo in due sezioni, e tutti gli individui che hanno superato item più difficili si devono trovare compresi nel gruppo che ha superato l'item più semplice. In altre parole alla base della costruzione della scala vi è quindi un modello scalare di tipo cumulativo che, come sappiamo, richiede una verifica basata sullo scalogramma di Guttman. Una delle più interessanti applicazioni del modello deterministico, infatti, è quella che consente la costruzione di una scala per la misurazione di capacità fisiche quando è possibile ipotizzare che le abilità per le funzioni di ogni giorno dipendono da capacità comuni, individuabili nei vari aspetti dell'attività quotidiana. In genere gli scalogrammi utilizzati nelle indagini sulle capacità fisiche vengono costruiti con significato

- positivo: possesso delle capacità fisiche, capacità di accudire a se stessi o
- negativo: disabilità, dipendenza nello svolgimento delle attività quotidiane.

In una scala con significato positivo si attribuisce il punteggio alto alla risposta che denota il superamento della prova.

In questa sede si è proceduto alla verifica dell'adattamento non solo del modello additivo-deterministico ma anche di altri. Ciò ha consentito di mettere in rilievo particolari aspetti dell'attributo misurato e, soprattutto, della scala utilizzata; in particolare; schematicamente, l'analisi ha riguardato:

Verifica della	Attraverso la verifica del livello di adattamento del modello
➢ omogeneità degli item	additivo
➢ cumulabilità (scalabilità)	deterministico
➢ dimensionalità	- fattoriale - *multidimensional scaling*
➢ scalabilità multidimensionale	POSAC

Gli item individuati sono i seguenti (in ordine di somministrazione):

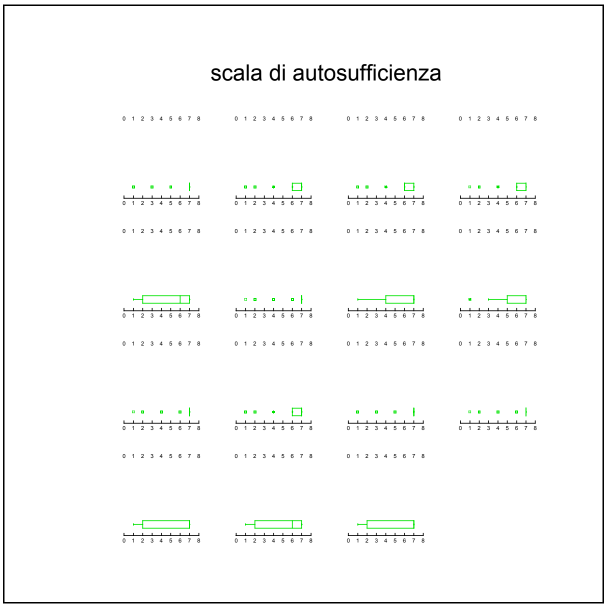
ITEM	Livelli di capacità misurati per ciascun item				
	1	2	3	4	5
	Senza difficoltà	Con difficoltà ma senza aiuto	Con aiuto per parte dell'azione	Con aiuto totale	Non lo fa per abitudine
1. Spostarsi per le stanze					
2. Uscire di casa					
3. Fare le scale					
4. Camminare per almeno 400 m					
5. Fare la spesa					
6. Lavarsi viso e braccia					
7. Fare il bagno o la doccia					
8. Vestirsi e spogliarsi					
9. Mangiare da solo					
10. Prepararsi un pasto caldo					
11. Usare il gabinetto					
12. Alzarsi e andare a letto					
13. Lavori domestici leggeri					
14. Lavori domestici pesanti					
15. Tagliarsi le unghie dei piedi					
	In alcuni item è stato espresso il codice 5 per identificare l'incapacità di eseguire una determinata attività per abitudine, tradizione familiare, culturale, ecc., e non per incapacità fisica (si pensi ai lavori domestici per gli uomini)				

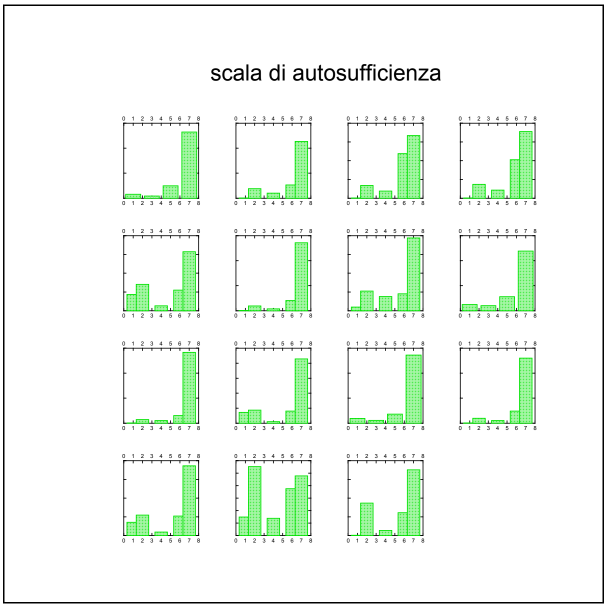
7.2 L'ANALISI

7.2.1 Analisi descrittiva dei singoli item

Di seguito possiamo osservare la distribuzione di frequenza relativa ai 15 item di autosufficienza.

ITEM	Livelli di capacità misurati per ciascun item					
	Senza difficoltà	Con difficoltà ma senza aiuto	Con aiuto per parte dell'azione	Con aiuto totale	Non lo fa per abitudine	Totale
1. Spostarsi per le stanze	2646	497	89	157		3389
2. Uscire di casa	2270	531	201	381	6	3389
3. Fare le scale	1669	1187	191	339	3	3389
4. Camminare per almeno 400 m	1777	1024	217	370	1	3389
5. Fare la spesa	1574	548	129	704	434	3389
6. Lavarsi viso e braccia	2720	408	73	187	1	3389
7. Fare il bagno o la doccia	1939	450	379	524	97	3389
8. Vestirsi e spogliarsi	2384	557	199	248		3389
9. Mangiare da solo	2831	308	107	142	1	3389
10. Prepararsi un pasto caldo	2140	403	53	435	358	3389
11. Usare il gabinetto	2727	363	112	187		3389
12. Alzarsi e andare a letto	2604	482	108	194	1	3389
13. Lavori domestici leggeri	1863	523	96	549	358	3389
14. Lavori domestici pesanti	955	752	277	1106	299	3389
15. Tagliarsi le unghie dei piedi	1757	613	139	869	11	3389





Osservando le distribuzioni di frequenza è possibile fare le prime valutazioni sulle performance rilevate secondo le affermazioni degli anziani. Osserviamo subito un risultato confortante per la salute degli anziani intervistati: la concentrazione delle risposte sui valori alti degli item indica una generale e positiva tendenza a svolgere le diverse attività senza difficoltà; tale concentrazione non appare però uniforme, ad indicare una possibile differenziazione degli item lungo il continuum di autosufficienza; inoltre, per alcuni item, si osservano distribuzioni con una forma irregolare di difficile interpretazione, in particolare per gli item 13, 14, 15.

7.2.2 Verifica dell'omogeneità

La definizione di autosufficienza qui adottata assume che tale caratteristica sia unidimensionale; conseguentemente la scala costruita assume che il punteggio totale sia monotonamente legato con la dimensione misurata. Per verificare ciò è necessario analizzare innanzitutto la consistenza interna del gruppo di item individuati. L'analisi della consistenza interna ha prodotto risultati piuttosto soddisfacenti; confrontando però i valori dello *split-half* per gli item suddivisi secondo due diversi metodi, *1^a^ metà-2^a^ metà* e *odd-even*, osserviamo subito come l'ordine degli item riflette in qualche modo una differenza tra gli item; infatti il primo metodo registra valori più bassi rivelando una diversa risposta dei soggetti ai due gruppi di domande.

INTERNAL CONSISTENCY DATA		
	HALF1-HALF2	ODD-EVEN
SPLIT-HALF CORRELATION	0.869	0.936
SPEARMAN-BROWN COEFFICIENT	0.930	0.967
GUTTMAN (RULON) COEFFICIENTE	0.926	0.960
COEFFICIENT *ALPHA*- ALL ITEM	0.943	
COEFFICIENT *ALPHA*	0.926 - 0.858	0.891 - 0.883

Item	Item-Total R	Item-reliability Index	Item-Total R Excl Item	Alpha Excl Item
1	0.820	0.611	0.799	0.939
2	0.859	0.880	0.834	0.936
3	0.857	0.814	0.834	0.937
4	0.851	0.838	0.826	0.937
5	0.706	1.082	0.635	0.943
6	0.805	0.623	0.781	0.939
7	0.785	0.979	0.741	0.938
8	0.862	0.775	0.842	0.937
9	0.766	0.552	0.740	0.940
10	0.660	0.958	0.586	0.944
11	0.827	0.652	0.806	0.938
12	0.834	0.671	0.813	0.938
13	0.713	1.049	0.647	0.942
14	0.663	0.926	0.592	0.943
15	0.782	0.993	0.737	0.939

7.2.3 Verifica della scalabilità

7.2.3.1 Verifica attraverso il modello deterministico

L'analisi della consistenza interna, come sappiamo, può condurre ad una valutazione finale ambigua perché uguali punteggi possono essere ottenuti con modelli di risposte (profili) diversi. Ciò vale soprattutto nel caso di item non dicotomici in cui sia importante non solo rilevare le differenze fra gli item ma anche pesare la "distanza", in termini di scalabilità, fra due item contigui. Quando è possibile assumere una gradualità degli item (per superare un item sono necessarie le stesse capacità necessarie per superare un item più facile più un livello di capacità ulteriore) è possibile validare la scala attraverso il confronto della distribuzione reale delle risposte dei soggetti del campione con un modello teorico di perfetta scalabilità tra item e di perfetta predicibilità di una prova sulla prova seguente Per verificare ciò si procede al confronto ripetuto per tutti gli item; ciò consente una valutazione globale della predicibilità della scala fatta sulla base degli errori osservati nel confronto tra il modello teorico e quanto è stato osservato.

Nel nostro caso il calcolo degli errori di scalabilità (per ogni soggetto e per ogni item) è reso complesso dal fatto che le risposte ai vari item non sono dicotomiche, ma con più livelli (4 o 5) di difficoltà variabili da item a item. Così, per attribuire lo stesso peso ad ogni item, qualunque sia il numero di livelli presenti nelle risposte, è stato assegnato ai due livelli estremi sempre lo stesso valore (1 e 7). Prevedendo un massimo di 7 livelli per ogni item, i punteggi intermedi sono stabiliti secondo un criterio di simmetria. Di seguito è presentato lo schema di ricodifica dei singoli item (i nuovi codici sono in blu).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	⇐ numero di identificazione degli item
Nuova codifica ⇓															Codici originari ⇓
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	1
5	6	6	6	6	6	6	5	6	6	5	6	6	6	6	2
3	4	4	4	4	4	4	3	4	4	3	4	4	4	4	3
1	2	2	2	2	2	2	1	2	2	1	2	2	2	2	4
	1	1	1	1	1	1		1	1		1	1	1	1	5
4	5	5	5	5	5	5	4	5	5	4	5	5	5	5	⇐ numero di livelli per ciascun item

Per ottenere un ordine gerarchico iniziale degli item è stato definito come *difficoltà crescente* il fatto che l'attività espressa da ogni item viene svolta da una percentuale decrescente di soggetti. La possibilità di trasporre questa progressiva difficoltà degli item così ricavata nella valutazione di singoli soggetti presuppone che in ciascun individuo le capacità fisiche richieste per il superamento di un item siano sufficienti per superare tutti gli item precedenti.

Dopo aver riunito in un unico codice le risposte relative alle modalità *senza difficoltà* e *con difficoltà ma senza aiuto* è stato ottenuto un riordinamento scalare degli item secondo la percentuale decrescente di soggetti che sono in grado di svolgere le attività relative all'item. I due riordinamenti (risposte originali e ricodificate riunendo le due modalità di risposta) coincidono per la maggior parte degli item. Di seguito vediamo una tabella nella quale a ciascun item è associato:

- l'ordine con cui è stato somministrato (colonna *a*),
- l’ordine scalare rispetto alle risposte non ricodificate (colonna *b*) e la frequenza di risposte tipo “1”;
- l’ordine scalare rispetto alle risposte ricodificate (colonna *c*) e la frequenza di risposte tipo “1” e “2”.

(a)	ITEM	*Senza difficoltà*		*Senza difficoltà e Con difficoltà ma senza aiuto*	
		Frequenza	(b)	Frequenza	(c)
1	1. Spostarsi per le stanze	2646	4	3143	1
2	2. Uscire di casa	2270	7	2801	8
3	3. Fare le scale	1669	13	2856	7
4	4. Camminare per almeno 400 m	1777	11	2801	9
5	5. Fare la spesa	1574	14	2122	14
6	6. Lavarsi viso e braccia	2720	3	3128	3
7	7. Fare il bagno o la doccia	1939	9	2389	11
8	8. Vestirsi e spogliarsi	2384	6	2942	6
9	9. Mangiare da solo	2831	1	3139	2
10	10. Prepararsi un pasto caldo	2140	8	2543	10
11	11. Usare il gabinetto	2727	2	3090	4
12	12. Alzarsi e andare a letto	2604	5	3086	5
13	13. Lavori domestici leggeri	1863	10	2386	12
14	14. Lavori domestici pesanti	955	15	1707	15
15	15. Tagliarsi le unghie dei piedi	1757	12	2370	13

E' stato adottato l'ordinamento ottenuto con la ricodifica (colonna *c*).

Tutti gli item hanno registrato coefficienti di riproducibilità superiori a 0.97 mentre l’intera scala ha prodotto un coefficiente di riproducibilità di 0.991. Vediamo di seguito i risultati nel dettaglio.

Item		confrontato con	ERRORI 1 punto	> 2 punti	RIPRODUCIBILITA' CR_{wi} (min richiesto 0.85)
1.	Spostarsi per le stanze		0	0	1 - 0 / 3389*6 = 1.000
9.	Mangiare da solo	1	662	346	1 - 346 / 3383*6 = 0.980
6.	Lavarsi viso e braccia	9	114	68	1 - 68 / 3387*6 = 0.997
11.	Usare il gabinetto	6	164	25	1 - 25 / 3124*6 = 0.999
12.	Alzarsi e andare a letto	11	607	107	1 - 107 / 3337*6 = 0.995
8.	Vestirsi e spogliarsi	12	125	9	1 - 9 / 3105*6 = 0.999
3.	Fare le scale	8	813	186	1 - 186 / 3355*6 = 0.991
2.	Uscire di casa	3	714	88	1 - 88 / 3384*6 = 0.996
4.	Camminare per almeno 400 m.	2	169	144	1 - 144 / 3355*6 = 0.993
10.	Prepararsi un pasto caldo	4	782	191	1 - 191 / 3387*6 = 0.991
7.	Fare il bagno o la doccia	10	555	404	1 - 404 / 3386*6 = 0.980
13.	Lavori domestici leggeri	7	537	366	1 - 366 / 3366*6 = 0.982
15.	Tagliarsi le unghie	13	578	388	1 - 388 / 3388*6 = 0.981
5.	Fare la spesa	15	451	285	1 - 285 / 3344*6 = 0.986
14.	Lavori domestici pesanti	5	428	268	1 - 268 / 3380*6 = 0.987
			totale= 2875		

Somma frequenze modali = 31856
Somma frequenze non-modali = 18978

$$CR_w \Rightarrow 1-\frac{n_S}{n * p_{mm}} = 1-\frac{2875}{3389*15*6} = 0.991$$

$$CS \Rightarrow 1-\frac{\sum n_{ie}}{me} = 1-\frac{2875}{18978} = 0.85$$

$$INP\% \Rightarrow [CR - \min(CR_i)]*100 = (0.991-0.98)*100 = 1.2$$

$$MMR \Rightarrow \frac{\sum nm_i}{N} = \frac{31856}{15*3389} = 0.63$$

dove

n_s	scarti negativi tra due item contigui
n	numero risposte (o numero di soggetti)
p_{mm}	punteggio_max - punteggio_min
nm_i	numero di risposte nella categoria modale (la più *scelta*) dell'item *i*
N	numero di risposte (*nitem* * *nsogg*)
n_{ie}	numero di errori dell'item *i*
me	errori marginali, somma di tutte le frequenze non-modali.

A questo punto è stato calcolato il coefficiente di predicibilità (*CP*) e il numero di previsioni realizzate per ciascun soggetto, ottenendo i seguenti risultati[2]:

[2] Ricordiamo che

$$CP_j = \sum \frac{pr_j}{pp_i}$$

dove
pr previsioni realizzate
pp previsioni possibili.

Coefficiente di predicibilità	Previsioni realizzate	Numero di soggetti
0.40	6	2
0.46	7	7
0.53	8	60
0.60	9	186
0.66	10	374
0.73	11	694
0.80	12	614
0.86	13	538
0.93	14	914

Gli item analizzati si sono rivelati nella loro applicazione pratica una scala con buone caratteristiche di scalabilità e di predicibilità.

7.2.3.2 Verifica attraverso il modello probabilistico

Come sappiamo il concetto di cumulabilità è presente anche nell'approccio probabilistico basato sull'*Item Response Theory*.
Proviamo a verificare se l'applicazione di tale modello, nella versione logistica con due parametri (difficoltà e discriminazione), produce risultati confrontabili con quelli ottenuti con il modello deterministico.
La preparazione della matrice dei dati per l'analisi ha richiesto l'applicazione dei seguenti passaggi:

a. disposizione dei dati in una matrice in cui le righe rappresentano i soggetti e le colonne gli item;
b. ordinamento della matrice per punteggi decrescenti e attribuzione del codice 1 alle risposte *corrette* e 0 alle risposte *scorrette* o (se significativo) alle *nulle*;
c. calcolo dei marginali di riga e colonna; eliminazione delle righe e delle colonne "complete" (ovvero quelle con risposte tutte corrette o con risposte tutte scorrette) e delle righe che presentano dei dati *missing*.

Per poter applicare il punto *b* si è proceduto alla ricodifica degli item; in particolare per dicotomizzare le risposte per ciascun item è stato seguito il seguente criterio[3]:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	⇐ item
Codifica originaria ⇓															Nuova codifica ⇓
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	1
5	6	6	6	6	6	6	5	6	6	5	6	6	6	6	0
3	4	4	4	4	4	4	3	4	4	3	4	4	4	4	
1	2	2	2	2	2	2	1	2	2	1	2	2	2	2	
	1	1	1	1	1	1		1	1		1	1	1	1	

Degli iniziali 3389 casi solo 2305 sono stati utilizzati per l'analisi per l'eliminazione di quelli che presentavano dati *missing*, il punteggio totale massimo o il punteggio totale minimo.

[3] E', naturalmente, possibile individuare altri criteri di dicotomizzazione ugualmente o più validi.

ITEM	(1) P	(2) b_i	(3) $E(b_i)$	(4) a_i	(5) $E(a_i)$
1	0.8547	-1.3625	0.0334	1.4325	0.0616
2	0.6915	-0.5286	0.0220	1.9376	0.0826
3	0.4308	0.3005	0.0200	1.8563	0.0778
4	0.4777	0.1644	0.0188	2.0746	0.0871
5	0.3896	0.4166	0.0207	1.7484	0.0740
6	0.8868	-1.5360	0.0339	1.6181	0.0768
7	0.5479	-0.0795	0.0246	1.3809	0.0555
8	0.7414	-0.7235	0.0233	1.9235	0.0841
9	0.9349	-2.1318	0.0499	1.1709	0.0539
10	0.6351	-0.7268	0.0542	0.5079	0.0252
11	0.8898	-1.4847	0.0297	2.1165	0.1144
12	0.8364	-1.2195	0.0303	1.5820	0.0690
13	0.5150	0.0037	0.0301	1.0118	0.0430
14	0.1210	1.4091	0.0303	1.4450	0.0634
15	0.4690	0.1729	0.0253	1.2897	0.0535
Media item utilizzati	0.6281	-0.4884	0.0298	1.5397	0.0681
Deviazione standard	0.2253	0.9135	0.0099	0.4214	0.0205
Numero item utilizzati	15	15	15	15	15

(1) *P*
(2) Difficoltà dell'item *i*
(3) Errore standard della difficoltà dell'item *i*
(4) Discriminazione dell'item *i*
(5) Errore standard della discriminazione dell'item *i*
(6)

Notiamo subito come nessun item ha superato il valore soglia di 0.25 nell'errore standard di difficoltà.

A scopo esemplificativo di seguito è presentata anche la lista delle stime riguardanti una parte dei 2305 soggetti utilizzati nell'analisi. Come si può notare vi sono soggetti che pur avendo lo stesso punteggio totale registrano diversi punteggi sulla logistica di capacità.

```
Listing of estimated item-response abilities and their standard errors.
All data below are based on 15 usable items.

                        Total        Mean         IRT        Std.
                Case    Score        Score      Ability      Error
    2    *******Unusable Case****** zero or perfect total score
                  4      8.0000     0.5333      0.4127     0.2945
    6    *******Unusable Case****** zero or perfect total score
                  8     11.0000     0.7333      0.1902     0.2786
    9    *******Unusable Case****** zero or perfect total score
                 10      8.0000     0.5333      0.4003     0.2931
   12    *******Unusable Case****** zero or perfect total score
                 13     11.0000     0.7333      0.2713     0.2822
                 14      6.0000     0.4000     -0.4386     0.2860
.....................................................................
               3376     10.0000     0.6667      0.7825     0.3668
               3377      7.0000     0.4667      0.1312     0.2775
 3378    *******Unusable Case****** zero or perfect total score
               3379     10.0000     0.6667     -0.8843     0.2912
               3380     14.0000     0.9333      0.1491     0.2777
               3383      4.0000     0.2667      1.2744     0.5177
 3386    *******Unusable Case****** zero or perfect total score
               3388     12.0000     0.8000      1.1461     0.4764
               3389     11.0000     0.7333      0.2390     0.2805

Mean                     9.4217     0.6181      0.0076     0.3554
Std Dev                  3.9217     0.2595      0.9940     0.1056
N cases                    1584       1584        1584       1584
```

In sintesi i 2305 soggetti hanno registrato i seguenti risultati:

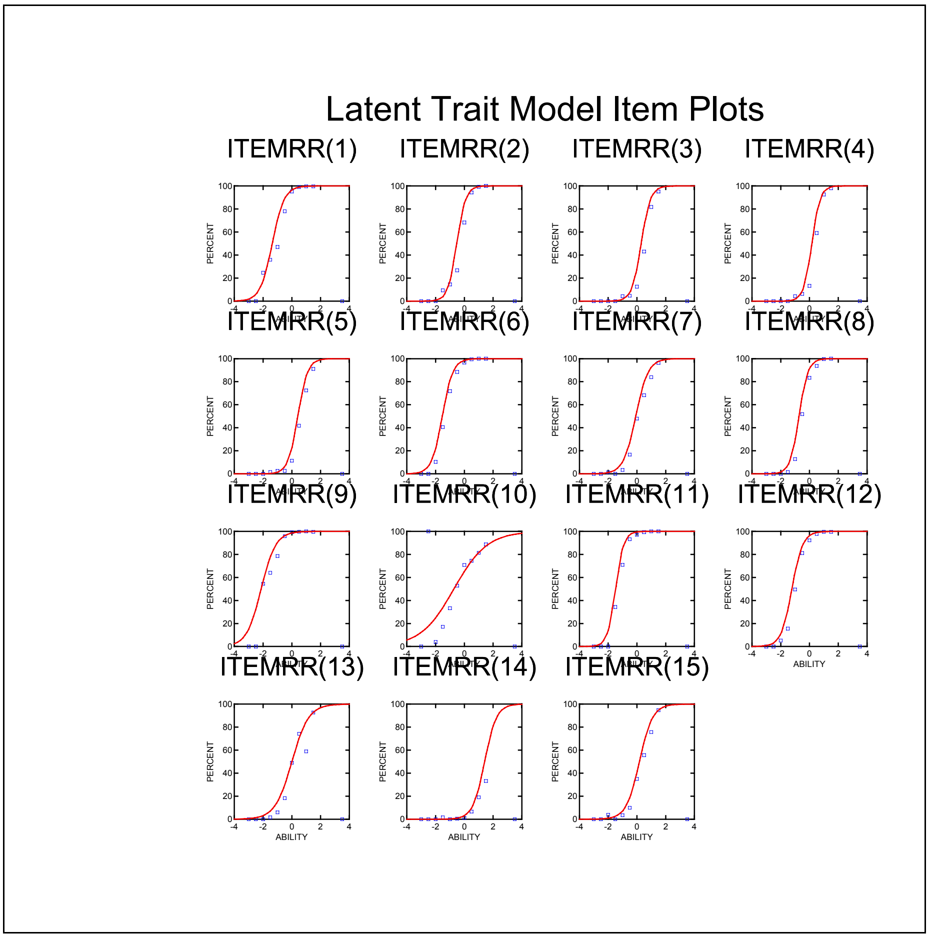
	Media	Deviazione standard
Punteggio medio	9.422	3.922
Media punteggi medi	0.628	0.261
Capacità media	0.000	1.000
Errore medio	0.357	0.106

Punteggio totale	Punteggio medio	Freq.	Freq. Cum.	%	% cum.
1	0.067	117	117	5.1	5.1
2	0.133	67	184	2.9	8.0
3	0.200	72	256	3.1	11.1
4	0.267	83	339	3.6	14.7
5	0.333	89	428	3.9	18.6
6	0.400	132	560	5.7	24.3
7	0.467	135	695	5.9	30.2
8	0.533	138	833	6.0	36.1
9	0.600	168	1001	7.3	43.4
10	0.667	151	1152	6.6	50.0
11	0.733	234	1386	10.2	60.1
12	0.800	270	1656	11.7	71.8
13	0.867	290	1946	12.6	84.4
14	0.933	359	2305	15.6	100.0

Ricordiamo che l'esame degli scarti standard (o dei loro quadrati) consente di identificare
- quali soggetti hanno dato una risposta poco plausibile,
- quali item non presentano una sufficiente affidabilità

che così possono essere eliminati dall'analisi.

Per fare ciò ci si dovrebbe riferire ai due coefficienti t_j e t_i che indicano il livello di incoerenza o *misfit*.
Come sappiamo la procedura di calcolo dei parametri ha carattere iterativo, in quanto ogni volta che le risposte relative ad soggetto o ad un item vengono considerate poco plausibili (alto *misfit*) il vettore di riga o di colonna corrispondente viene cancellato dalla matrice e si ripete l'intera procedura. Verifichiamo l'adattamento al modello logistico degli item utilizzando i grafici che mostrano la percentuale di punteggi *corretti* per ciascuno dei livelli di *capacità*. Ricordiamo che gli asterischi indicano per ciascun item la percentuale attesa di risposte corrette secondo il modello logistico e che i valori numerici delle proporzioni osservate (*P*) e attese (*E(P)*) sono presentati a destra.
Dall'osservazione di tali grafici si nota subito come l'item 10 presenti dei notevoli problemi di adattamento.



7.2.3.3 Confronto tra i risultati dei due approcci

Per confrontare lo scaling degli item ottenuto dalla applicazione dei due diversi modelli si è proceduto al calcolo delle associazioni tra i diversi parametri e coefficienti prodotti.
I risultati di tale confronto, relativamente agli item e ai soggetti, sono riassunti in due matrici presentate di seguito nella quale compaiono i valori delle associazioni calcolate per mezzo di tre coefficienti:

- *tau* di Kendall
- *rho* di Spearman
- *r* di Pearson.

Confronto dei risultati relativi agli item

	P	*bi*	*ebi*	*ai*	*eai*	*cr*	*irt*	*gutt*
r di Pearson								
P	1.00							
bi	-0.99	1.00						
ebi	0.40	-0.52	1.00					
ai	0.09	0.00	-0.74	1.00				
eai	0.24	-0.16	-0.55	0.95	1.00			
cr	0.45	-0.37	-0.12	0.52	0.59	1.00		
IRT	-0.81	0.84	-0.54	-0.04	-0.22	-0.43	1.00	
GUTT	-0.89	0.89	-0.29	-0.25	-0.36	-0.57	0.79	1.00
rho di Spearman								
P	1.00							
bi	-0.99	1.00						
ebi	0.57	-0.63	1.00					
ai	0.04	0.03	-0.63	1.00				
eai	0.18	-0.13	-0.50	0.98	1.00			
cr	0.44	-0.42	0.02	0.59	0.64	1.00		
IRT	-0.87	0.88	-0.64	-0.04	-0.19	-0.43	1.00	
GUTT	-0.88	0.88	-0.37	-0.20	-0.33	-0.58	0.79	1.00
tau di Kendall								
P	1.00							
bi	-0.96	1.00						
ebi	0.43	-0.47	1.00					
ai	0.03	0.01	-0.49	1.00				
eai	0.12	-0.09	-0.39	0.91	1.00			
cr	0.35	-0.31	0.02	0.41	0.44	1.00		
IRT	-0.68	0.71	-0.51	-0.01	-0.11	-0.33	1.00	
GUTT	-0.75	0.75	-0.29	-0.09	-0.18	-0.46	0.62	1.00
p: probabilità di rispondere correttamente all'item (P)								
bi: parametro di difficoltà dell'item *i* (b_i)								
ebi: errore standard del parametro di difficoltà $(E(b_i))$								
ai: parametro di discriminazione dell'item (a_i)								
eai: errore standard del parametro di discriminazione $(E(a_i))$								
cr: coefficiente di riproducibilità (CR)								
irt: ordinamento ottenuto con l'applicazione del modello logistico								
gutt: ordinamento ottenuto con l'applicazione del modello deterministico								

L'analisi di tali risultati ci consente innanzi tutto registrare un'alta concordanza tra gli ordinamenti degli item rispetto alla difficoltà ottenuti con i due approcci (alta correlazione tra le due serie ordinate *irt* e *gutt*). L'impressione avuta da questi dati è confermata anche confrontando direttamente le due graduatorie:

item ⇒		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ranghi ottenuti con l'applicazione dell'approccio	logistico	4	8	12	14	11	2	9	6	1	7	3	5	15	10	13
	deterministico	1	8	7	9	14	3	11	6	2	10	4	5	12	15	13

Proseguendo nel confronto i risultati ottenuti attraverso l'applicazione dei due modelli, è possibile notare come il coefficiente di riproducibilità (*cr*) presenta una correlazione piuttosto alta con il parametro di discriminazione (*a*) e una correlazione tendenzialmente negativa con il parametro di difficoltà (*b*).

Confronto dei risultati relativi ai soggetti[4]

Mettendo a confronto le valutazioni che i due approcci fanno dei soggetti misurati, osserviamo la relazione esistente tra coefficiente di predicibilità e livello di capacità.

	pauto	*cp*	*tp*	*mp*	*dj*	*edj*
r di Pearson						
pauto	1.000					
cp	0.58	1.00				
tp	0.84	0.74	1.00			
mp	0.84	0.74	1.00	1.00		
dj	0.81	0.75	0.98	0.99	1.00	
edj	0.17	0.44	0.34	0.34	0.43	1.00
rho di Spearman						
pauto	1.00					
cp	0.68	1.00				
tp	0.84	0.80	1.00			
mp	0.84	0.80	1.00	1.00		
dj	0.80	0.79	0.99	0.99	1.00	
edj	0.29	0.48	0.47	0.47	0.48	1.00
tau di Kendall						
pauto	1.00					
cp	0.54	1.00				
tp	0.68	0.68	1.00			
mp	0.68	0.68	1.00	1.00		
dj	0.62	0.64	0.94	0.94	1.00	
edj	0.17	0.36	0.29	0.29	0.32	1.00
pauto: punteggio di autosufficienza (approccio deterministico)						
cp: coefficiente di predicibilità						
tp: punteggio totale						
mp: punteggio medio						
dj: capacità del soggetto (d_j)						
edj: errore standard della capacità $(E(d_j))$						

Osserviamo subito l'alta correlazione che sia il punteggio di autosufficienza (*pauto*) che il coefficiente di predicibilità (*cp*) registrano con la capacità registrata da ciascun soggetto secondo il modello *IRT*, rispettivamente 0.812 e 0.747.

E' possibile notare a questo punto come il modello deterministico (secondo l'approccio di Guttman) e quello logistico (secondo la teoria dell'*item response*), pur prendendo origine da considerazioni e ipotesi abbastanza differenziate, giungano ad una valutazione sostanzialmente confrontabile della scala e ad una misurazione sostanzialmente coerente e concorde sia degli item che dei soggetti.

7.2.4 Verifica della dimensionalità

I risultati finora osservati confermano la natura ordinale della caratteristica misurata. Alcuni aspetti emersi ci autorizzano ad ipotizzare che l'autosufficienza fisica, così com'è stata definita e com'è percepita, non sia perfettamente unidimensionale. In particolare si potrebbe ipotizzare la presenza di due componenti (fattori). A tal fine verifichiamo se il criterio fattoriale e quello di *scaling*

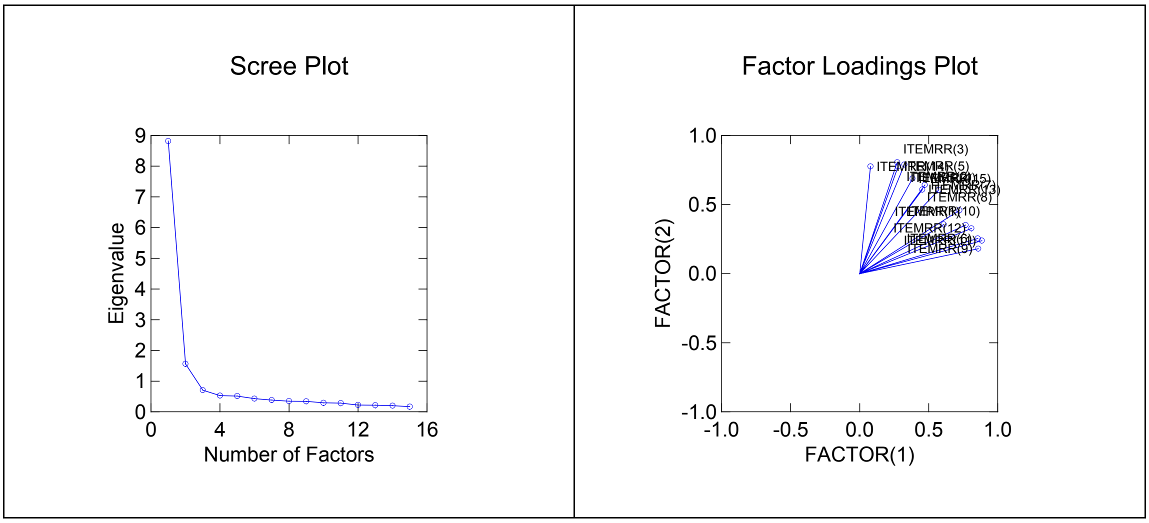
[4] Ricordiamo che per il modello logistico l'analisi viene effettuata solo sui soggetti che non presentano i punteggi estremi (massima e minima autosufficienza).

multidimensionale si adattano ai nostri dati.

Analisi fattoriale

L'applicazione dell'analisi fattoriale esplorativa (Maggino, 2005) sembra confermare la nostra ipotesi: i due fattori estratti spiegano quasi la stessa quantità di varianza per un totale di 70% di varianza totale.

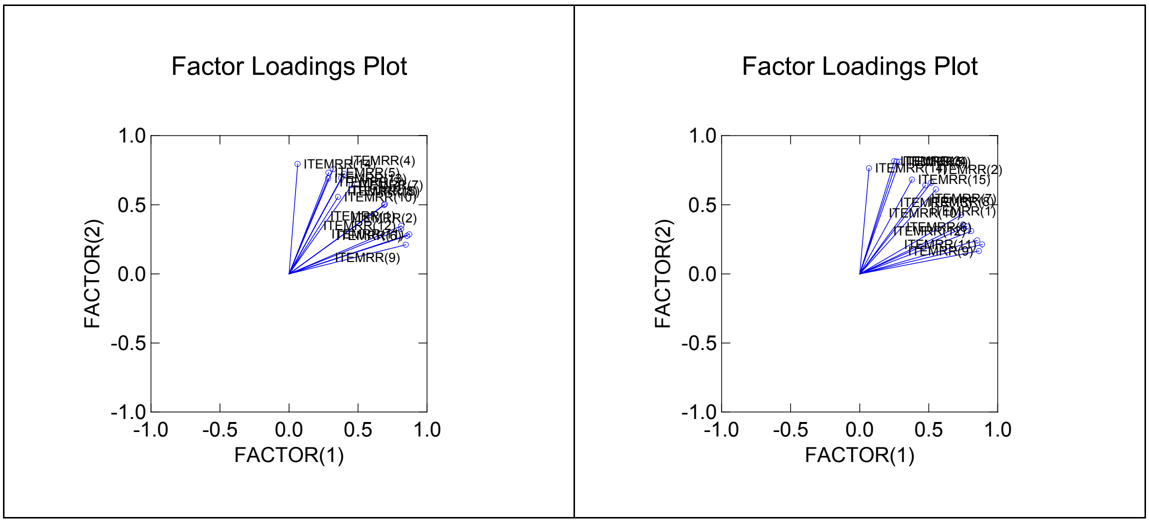
Rotated Loading Matrix (VARIMAX)		
ITEM 1	0.7675	0.3519
ITEM 2	0.5729	0.6064
ITEM 3	0.2719	0.8061
ITEM 4	0.3242	0.7895
ITEM 5	0.2785	0.7776
ITEM 6	0.8553	0.2545
ITEM 7	0.4717	0.6413
ITEM 8	0.7223	0.4580
ITEM 9	0.8592	0.1806
ITEM 10	0.6091	0.3556
ITEM 11	0.8849	0.2390
ITEM 12	0.8094	0.3275
ITEM 13	0.4534	0.6086
ITEM 14	0.0785	0.7757
ITEM 15	0.3830	0.6884
Varianza spiegata	5.5554	4.8247
Perc. Varianza spiegata	37.0362	32.1650



I due fattori estratti si riferiscono a due aspetti dell'autosufficienza tra loro ordinali: mentre il primo riguarda attività ed azioni che richiedono capacità minime (spostarsi per le stanze ed uscire di casa, lavarsi viso e braccia, vestirsi e spogliarsi, mangiare e prepararsi un pasto caldo da soli, usare il bagno, alzarsi e andare a letto), il secondo riguarda attività che richiedono capacità fisiche elevate (camminare e fare la spesa, fare il bagno o la doccia, tagliarsi le unghie dei piedi, fare lavori domestici).

In realtà i suggerimenti che ci pervenivano dai risultati precedenti facevano ipotizzare la presenza, o l'interferenza, di un fattore culturale accanto a quello riguardante strettamente le capacità fisiche. In particolare un indicatore di ciò poteva essere identificato con la variabile genere. Per questo motivo abbiamo provato a riapplicare il criterio fattoriale ai due gruppi separati.

Rotated Loading Matrix (VARIMAX)				
	maschi		femmine	
ITEM 1	0.812	0.324	0.749	0.357
ITEM 2	0.690	0.498	0.518	0.660
ITEM 3	0.321	0.758	0.251	0.815
ITEM 4	0.398	0.723	0.288	0.811
ITEM 5	0.286	0.732	0.267	0.812
ITEM 6	0.856	0.275	0.852	0.242
ITEM 7	0.455	0.644	0.476	0.641
ITEM 8	0.694	0.505	0.736	0.427
ITEM 9	0.846	0.211	0.864	0.165
ITEM 10	0.355	0.556	0.776	0.343
ITEM 11	0.871	0.286	0.887	0.213
ITEM 12	0.814	0.347	0.808	0.309
ITEM 13	0.284	0.691	0.552	0.611
ITEM 14	0.062	0.795	0.068	0.764
ITEM 15	0.378	0.694	0.379	0.682
Varianza spiegata	5.392	4.877	5.782	4.905
Perc. Varianza spiegata	35.944	32.515	38.545	32.702

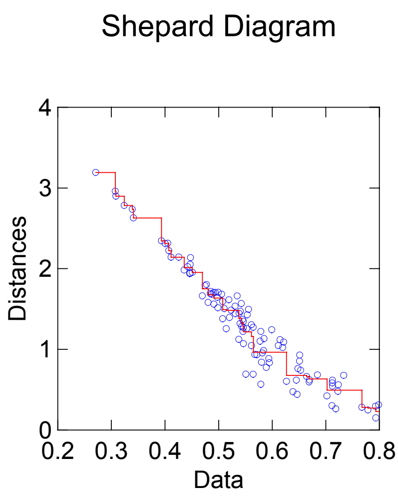


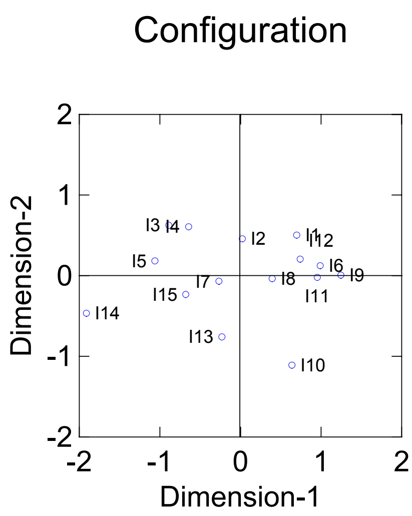
Come si può osservare la sostanziale uguaglianza tra le strutture fattoriali per maschi e femmine trova un'eccezione nell'item 8 ("vestirsi e spogliarsi") e nell'item 13 ("fare lavori domestici leggeri") che, il primo per i maschi e il secondo per le femmine, compaiono in entrambi i fattori; l'item che però rivela avere un comportamento molto legato al genere è il numero 10 ("prepararsi un pasto caldo") che per i maschi, al contrario di quanto succede per le femmine, definisce il fattore legato alle capacità elevate.

Il criterio di *scaling* multidimensionale

Anche se i dati sono del tipo stimolo-unico proviamo ad applicare il criterio di *scaling* multidimensionale; in tale applicazione la matrice di somiglianza sottoposta ad analisi è quella dei correlazione tra gli item; si assume una relazione monotona tra somiglianze e distanze (funzione di trasformazione monotona). I risultati rilevano un elevato adattamento tra matrice di somiglianza (osservata) e quella di distanze (calcolata).

Monotonic Multidimensional Scaling Minimizing Kruskal STRESS (form 1) in 2 dimensions Stress of final configuration is: 0.08803 Proportion of variance (RSQ) is: 0.96222					
	Dim 1	Dim 2		Dim 1	Dim 2
ITEM 1	.70	.50	ITEM 8	.39	-.04
ITEM 2	.02	.46	ITEM 9	1.25	.00
ITEM 3	-.89	.62	ITEM 10	.64	-1.11
ITEM 4	-.64	.61	ITEM 11	.95	-.02
ITEM 5	-1.06	.18	ITEM 12	.74	.20
ITEM 6	.99	.12	ITEM 13	-.23	-.76
ITEM 7	-.27	-.07	ITEM 14	-1.91	-.47
			ITEM 15	-.68	-.23

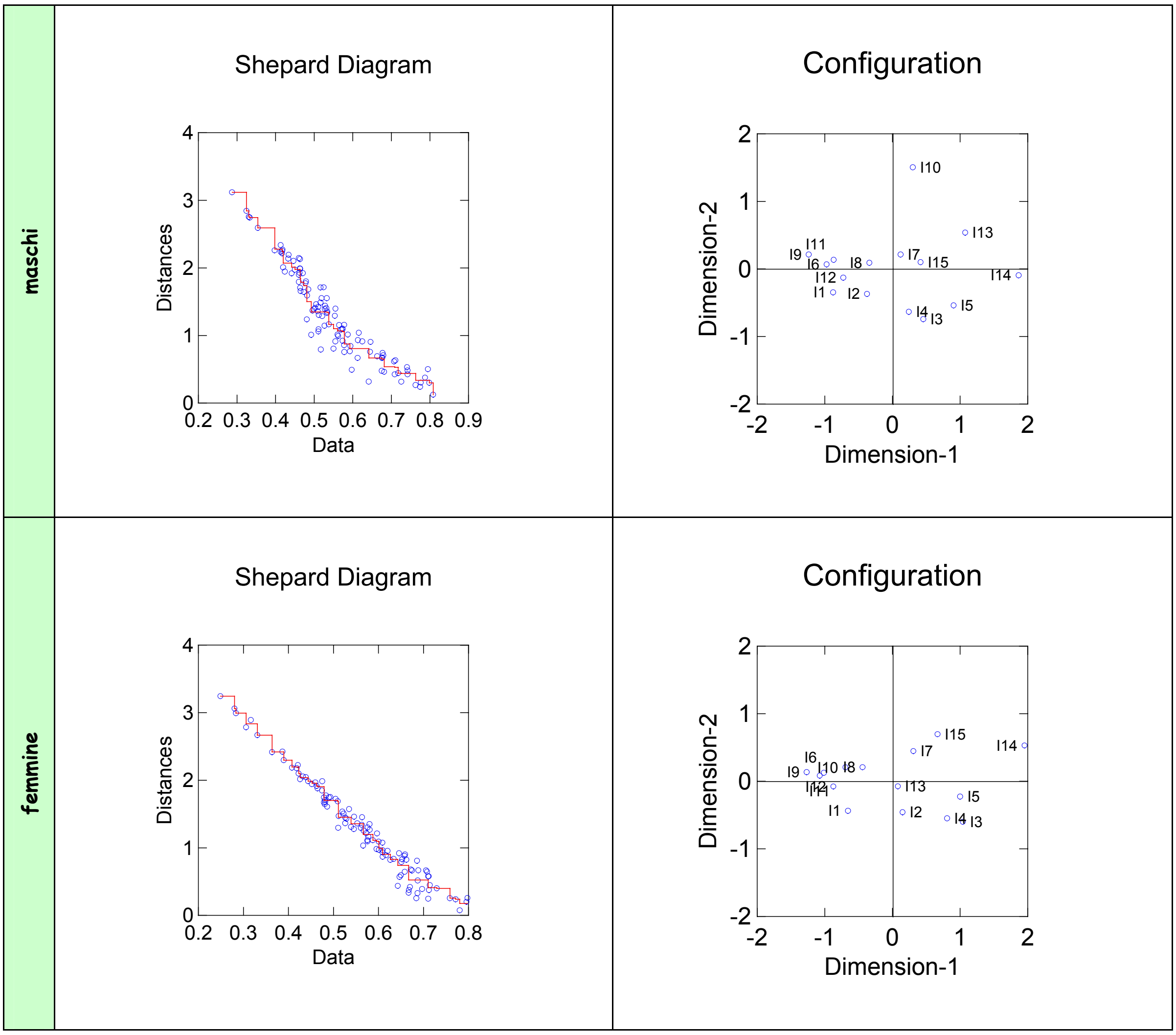




I risultati di questa analisi risultano essere piuttosto interessanti. La configurazione finale conferma la struttura fattoriale precedentemente vista: rispetto alla *dimensione-1* osserviamo in corrispondenza della parte positiva gli item che prima avevamo osservato nel fattore 1 e in corrispondenza della parte negativa gli item che precedentemente avevamo osservato nel fattore 2; rispetto al criterio fattoriale qui appare più evidente il carattere ordinale/cumulativo degli item.
Osservando i risultati relativamente alla *dimensione-2* è possibile rilevare come gli item sembrano disporsi in modo più legato alle abitudini che alle capacità fisiche (si notino nella parte negativa gli item "prepararsi un pasto caldo" e quelli relativi ai lavori domestici). Avendo individuato nella variabile "genere" quella che può discriminare rispetto alla dimensione culturale, proviamo a ripetere l'analisi separatamente per maschi e femmine.

Monotonic Multidimensional Scaling Minimizing Kruskal STRESS (form 1) in 2 dimensions		
	Maschi	Femmine
Stress of final configuration	0.09985	0.06547
Proportion of variance (RSQ)	0.95271	0.98193

Coordinate in due dimensioni					
		Maschi		Femmine	
		Dim 1	Dim 2	Dim 1	Dim 2
ITEM	1	-.88	-.35	-.66	-.44
	2	-.37	-.37	.15	-.46
	3	.46	-.74	1.04	-.60
	4	.24	-.63	81	-.55
	5	.91	-.54	1.00	-.23
	6	-.97	.07	-1.01	.13
	7	.12	.21	.31	.45
	8	-.34	.09	-.44	.21
	9	-1.24	.21	-1.27	.13
	10	.30	1.50	-.69	.21
	11	-.87	.13	-1.07	.08
	12	-.72	-.13	-.87	-.08
	13	1.08	.54	.08	-.08
	14	1.87	-.09	1.95	.53
	15	.42	.10	.67	.70



Attraverso questa analisi appare abbastanza chiaro come per il gruppo dei maschi il significato della dimensione 2 sia quasi del tutto attribuibile all'item 13 ("fare lavori domestici leggeri") e, soprattutto, l'item 10 ("prepararsi un pasto caldo") mentre per il gruppo delle femmine la seconda dimensione sembra quasi non avere alcun significato, infatti rispetto a questa dimensione gli item risultano essere concentrati in pochi valori.

7.2.5 Nuove ipotesi

A questo punto si può dire che nonostante la verifica dimensionale abbia confermato la natura ordinale della caratteristica misurata, la presenza di due particolari item può far insorgere due diverse ipotesi:

- si tratta di item affetti da *bias*; in questo caso si può assumere che l'errore è prodotto da una differenza significativa tra gruppi (maschi e femmine) rispetto alla difficoltà relativa dell'item; in altre parole risulta comparativamente più difficile rispondere "correttamente" per un gruppo anziché per un altro;
- si tratta di item che misurano un'altra dimensione.

Per verificare la prima ipotesi utilizziamo la tecnica *TID (Transformed Item Difficulties)* mentre la verifica del livello di adattamento del modello di scalogramma multidimensionale (POSAC) consentirà provare la seconda.

7.2.5.1 Verifica della presenza di item affetti da bias

L'applicazione della tecnica detta *Transformed Item Difficulties (TID)* consente di individuare la presenza di item affetti da *bias* attraverso la verifica della presenza o l'assenza dell'interazione gruppi*item. Nel nostro caso i gruppi individuati sono maschi e femmine.
Dato il significato dei nostri item, si considera indice di difficoltà, *p*, la proporzione di soggetti che hanno riferito di non essere capaci a svolgere una determinata funzione. Per individuare il valore di *p* si è deciso di procedere ad un diverso accorpamento, rispetto a quanto fatto precedentemente, delle categorie utilizzate, ovvero:

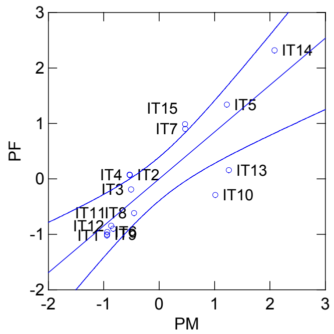
1-p	• Senza difficoltà • Con difficoltà ma senza aiuto
p	• Con aiuto per parte dell'azione • Con aiuto totale • Non lo fa per abitudine

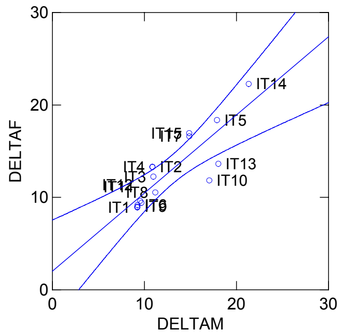
Di seguito valori di *p* e di *1-p* per i due gruppi:

ITEM		maschi		femmine	
		1-p	p	1-p	p
11	Usare il gabinetto	1505	.132	.1585	.167
1	Spostarsi per le stanza	1530	.107	.1613	.139
2	Uscire di casa	1439	.198	.1362	.390
3	Fare le scale	1434	.203	.1422	.330
4	Camminare per almeno 400 m.	1440	.197	.1361	.391
5	Fare la spesa	1052	.585	.1070	.682
6	Lavarsi viso e braccia	1529	.108	.1599	.153
7	Fare il bagno o la doccia	1218	.419	.1171	.581
8	Vestirsi e spogliarsi	1422	.215	.1520	.232
9	Mangiare da solo	1530	.107	.1609	.143
10	Prepararsi un pasto caldo	1098	.539	.1445	.307
12	Alzarsi e andare a letto	1513	.124	.1573	.179
13	Lavori domestici leggeri	1044	.593	.1342	.410
14	Lavori domestici pesanti	862	.775	.845	.907
15	Tagliarsi le unghie	1219	.418	.1151	.601

I valori di *p* vengono quindi standardizzati (*valori delta*). Di seguito vediamo due diagrammi di dispersione che mettono in relazione i valori *p* (primo diagramma) e i valori *delta* (secondo diagramma) per il gruppo di maschi e il gruppo delle femmine. Secondo questo approccio il livello di dispersione dei punti nel grafico così costruito è considerato una misura dell'interazione *gruppo*item*, una specie di coefficiente di correlazione inverso. La retta tracciata, che può essere considerata l'asse maggiore dell'ellisse descritta dai punti, serve come indice della relazione bivariata dei valori *p*/*delta* dei due gruppi. Esso diviene l'informazione base a partire dalla quale individuare la presenza di item *biased*. Per ciascuna retta sono determinati gli intervalli di confidenza dell'asse maggiore. Gli item che nel grafico appaiono al di fuori di tali intervalli possono essere giudicati "deviati".

Confronto tra il gruppo dei maschi e quello delle femmine per i valori:	
p	*delta*





Come si può osservare gli item, già segnalati come item deboli dall'analisi dello scalogramma, sono il 10 e il 13, risultati più "difficili" per i maschi, e il 15 e il 7 (per entrambi i gruppi) che per tutti richiedono capacità fisiche difficilmente rilevabili tra gli anziani.
Occorre però tener presente che in questo caso gli item risultati *biased* non sono di per sé non validi ma rivelano in realtà la probabile presenza di un'altra dimensione.

7.2.5.2 Verifica della scalabilità multidimensionale

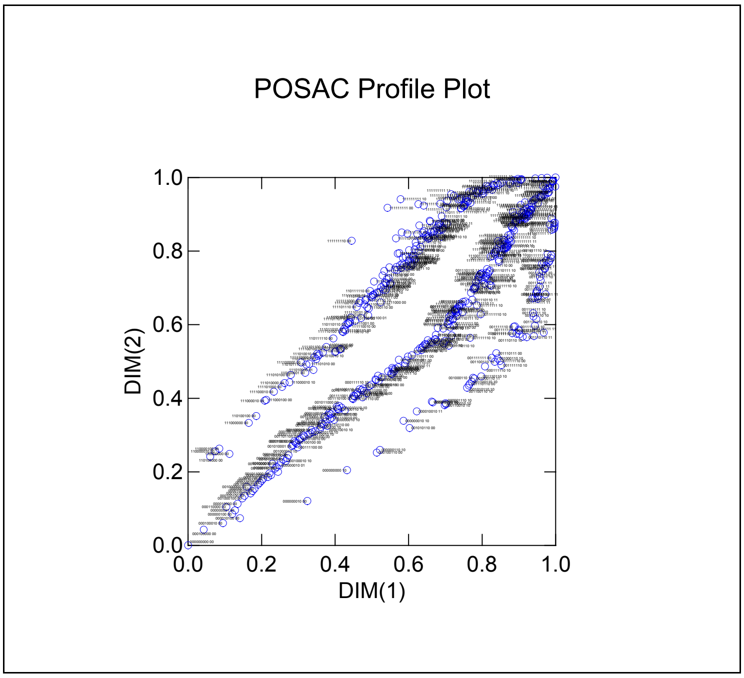
La misura del livello di adattamento del criterio di scalogramma multidimensionale consentirà di chiarire ulteriormente la presenza di due dimensioni. Per fare ciò, essendo la matrice dei dati molto grande, si utilizza la procedura *POSAC*; questa, come sappiamo, consente di determinare l'adattamento dei profili osservati in uno spazio bidimensionale.
Ricordiamo che l'obiettivo del *POSAC* è quello di verificare se è possibile assegnare due punteggi a ciascun profilo, in modo tale che sia possibile rappresentare la relazione tra due profili confrontando semplicemente i loro corrispondenti profili di coordinate. Si ottiene una rappresentazione perfetta quando i punteggi individuati descrivono perfettamente l'ordine e la confrontabilità dei profili originari.
Nel caso in cui non sia possibile ottenere per un certo scalogramma una perfetta rappresentazione in uno spazio bidimensionale, l'approccio *POSAC* definisce un criterio di bontà di adattamento che consente di stabilire qual è il tipo di collocazione bidimensionale che meglio descrive le relazioni d'ordine osservate tra i profili. A tal fine è stato definito un coefficiente, detto *coefficiente di corretta rappresentazione* (*CORREP*), basato sulla proporzione di profili rappresentati in modo corretto, tenendo conto della frequenza registrata da ciascun profilo. Il valore del coefficiente di corretta rappresentazione va da 0 a 1 (soluzione perfetta) ed è molto sensibile all'*approssimazione iniziale* utilizzata.

In altri casi il procedimento per definire l'approssimazione iniziale prevede che vengano eseguiti in successione i seguenti momenti:

a. calcolo della matrice dei coefficienti di debole monotonicità (*weak monotonicity coefficients*, *wm*);
b. identificazione dei due item (i_0 e j_0) che presentano la minore correlazione positiva (item estremi);
c. determinazione della posizione di ciascun profilo.

Reordered item weak monotonicity coefficients															
item	10	13	9	11	6	8	12	1	7	15	2	5	14	4	3
10	1.000														
13	0.874	1.000													
9	0.974	0.975	1.000												
11	0.964	0.977	0.987	1.000											
6	0.93	0.963	0.983	0.983	1.000										
8	0.889	0.921	0.986	0.990	0.985	1.000									
12	0.897	0.950	0.978	0.987	0.968	0.964	1.000								
1	0.890	0.963	0.955	0.965	0.968	0.956	0.961	1.000							
7	0.819	0.859	0.985	0.980	0.993	0.962	0.948	0.941	1.000						
15	0.804	0.847	0.980	0.985	0.976	0.968	0.956	0.939	0.918	1.000					
2	0.805	0.913	0.957	0.957	0.960	0.925	0.951	0.993	0.923	0.916	1.000				
5	0.787	0.858	0.986	0.975	0.978	0.944	0.958	0.976	0.899	0.872	0.974	1.000			
14	0.766	0.929	0.972	1.000	0.989	0.973	0.987	0.976	0.934	0.918	0.975	0.911	1.000		
4	0.724	0.849	0.958	0.964	0.967	0.933	0.959	0.991	0.880	0.869	0.976	0.931	0.944	1.000	
3	0.717	0.834	0.951	0.954	0.944	0.919	0.962	0.984	0.873	0.879	0.978	0.909	0.938	0.972	1.000

Final loss value	3350.8184
Proportion of profile pairs correctly represented	0.7701
Score-distance weighted coefficient	0.8520

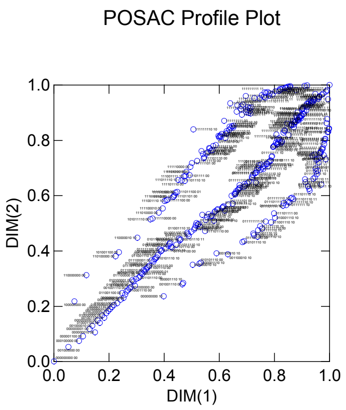


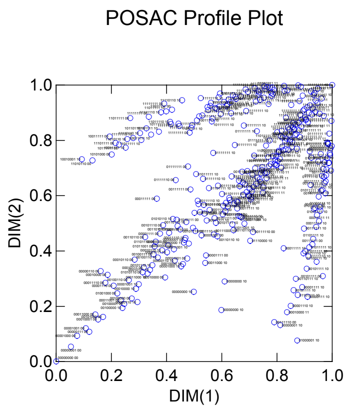
Maschi															
Reordered item weak monotonicity coefficients															
item	10	9	11	6	8	12	13	1	7	15	2	5	14	4	3
10	1.000														
9	0.969	1.000													
11	0.977	0.989	1.000												
6	0.938	0.986	0.987	1.000											
8	0.899	0.988	0.994	0.988	1.000										
12	0.916	0.978	0.987	0.973	0.974	1.000									
13	0.783	0.978	0.975	0.961	0.906	0.940	1.000								
1	0.913	0.965	0.975	0.975	0.966	0.979	0.964	1.000							
7	0.827	0.987	0.978	0.993	0.963	0.944	0.811	0.945	1.000						
15	0.807	0.987	0.985	0.970	0.971	0.964	0.822	0.953	0.917	1.000					
2	0.859	0.964	0.967	0.969	0.948	0.967	0.935	0.996	0.929	0.941	1.000				
5	0.765	0.994	0.979	0.973	0.947	0.965	0.811	0.971	0.882	0.852	0.968	1.000			
14	0.759	0.974	1.000	0.990	0.968	0.984	0.912	0.981	0.904	0.895	0.970	0.862	1.000		
4	0.741	0.951	0.966	0.970	0.935	0.966	0.841	0.995	0.873	0.873	0.983	0.908	0.938	1.000	
3	0.714	0.944	0.952	0.950	0.922	0.968	0.815	0.994	0.883	0.881	0.983	0.886	0.916	0.973	1.000

Femmine															
Reordered item weak monotonicity coefficients															
item	15	7	8	10	6	11	9	13	12	14	1	5	2	4	3
15	1.000														
7	0.916	1.000													
8	0.965	0.961	1.000												
10	0.944	0.934	0.946	1.000											
6	0.982	0.993	0.982	0.958	1.000										
11	0.985	0.981	0.985	0.976	0.980	1.000									
9	0.972	0.983	0.985	0.986	0.980	0.986	1.000								
13	0.904	0.917	0.939	0.969	0.968	0.979	0.976	1.000							
12	0.947	0.950	0.953	0.939	0.963	0.988	0.978	0.961	1.000						
14	0.938	0.963	0.979	0.982	0.987	1.000	0.968	0.996	0.990	1.000					
1	0.923	0.936	0.948	0.936	0.960	0.956	0.946	0.969	0.941	0.969	1.000				
5	0.885	0.910	0.939	0.957	0.982	0.971	0.976	0.933	0.950	0.956	0.978	1.000			
2	0.896	0.920	0.914	0.929	0.956	0.954	0.954	0.940	0.938	0.977	0.990	0.977	1.000		
4	0.860	0.885	0.937	0.913	0.964	0.966	0.969	0.918	0.951	0.945	0.985	0.948	0.968	1.000	
3	0.873	0.858	0.918	0.909	0.937	0.958	0.959	0.904	0.956	0.954	0.972	0.926	0.972	0.969	1.000

	maschi	femmine
Final loss value	953.5039	255.3252
Proportion of profile pairs correctly represented	0.7365	0.8296
Score-distance weighted coefficient	0.8340	0.9518

Maschi	Femmine





Confrontando i risultati prodotti, sia numerici che grafici, sul campione totale e suddiviso per genere, notiamo subito come il migliore adattamento è quello ottenuto nel campione femminile ovvero quello che, secondo le nostre ipotesi è anche meno affetto dalla dimensione "culturale".
Può essere curioso a questo punto confrontare l'ordinamento degli item rispetto ai valori dei coefficienti di *weak monotonicity*.

	Ordinamenti degli item secondo i valori dei coefficienti di *weak monotonicity*		
	Totale	Maschi	Femmine
1. Spostarsi per le stanze	8	8	3
2. Uscire di casa	11	11	13
3. Fare le scale	15	15	15
4. Camminare per almeno 400 m	14	14	14
5. Fare la spesa	12	12	12
6. Lavarsi viso e braccia	5	4	5
7. Fare il bagno o la doccia	9	9	2
8. Vestirsi e spogliarsi	6	5	3
9. Mangiare da solo	3	2	7
10. Prepararsi un pasto caldo	1	1	4
11. Usare il gabinetto	4	3	6
12. Alzarsi e andare a letto	7	6	9
13. Lavori domestici leggeri	2	7	8
14. Lavori domestici pesanti	13	13	10
15. Tagliarsi le unghie dei piedi	10	10	1

Notiamo subito come l'ordinamento ottenuto sul campione totale e quello ottenuto sul campione di maschi sono abbastanza omogenei tra loro e molto diversi da quello ottenuto sul campione di femmine. In particolare vediamo come ciò riguarda gli item gli item 10, 15, 7 e 9.

7.2.6 Individuazione dei valori-soglia

Da questa scala è stato possibile ottenere non solo una valutazione dell'autosufficienza dei singoli soggetti del campione in termini di punteggio globale ma anche altri tre punteggi relativi alle aree di non autosufficienza identificate sulla base della tipologia delle attività che le caratterizzano ma soprattutto della diversità dei bisogni assistenziali che ne derivano.
Infatti, poiché uno dei principali obbiettivi dello studio era l'identificazione dei bisogni assistenziali, è stato analizzato il significato delle singole attività rilevate dalla scala. Il riordinamento degli item

secondo il modello della "scalabilità" ha messo in evidenza che le attività svolte senza difficoltà dal maggior numero di soggetti e il cui deterioramento è da considerarsi indicativo di una grave disabilità, riguardano le attività relative alla "cura di sé" che devono essere espletate più volte al giorno e in tempi non prevedibili. Al contrario le attività più "difficili" comprendono le attività da svolgersi in modo saltuario. In posizione intermedia si ritrovano attività che vengono svolte quotidianamente a tempi prevedibili. Cosí, il nuovo ordinamento delle attività permette di individuare tre aree di attività diverse per:

- periodicità temporale
- tipo di intervento richiesto.

IDENTIFICAZIONE DELLE AREE DI NON AUTOSUFFICIENZA			
(b)	(a)		autosufficienza nelle attività
1	3	Lavarsi viso e braccia	per la cura di sé
2	1	Spostarsi per le stanze	
3	2	Mangiare da solo	
4	4	Usare il gabinetto	
5	5	Alzarsi e andare a letto	
6	6	Vestirsi e spogliarsi	
7	10	Prepararsi un pasto caldo	
11	12	Fare lavori domestici leggeri	quotidiane
12	11	Fare il bagno o la doccia	
13	14	Portare una borsa della spesa	
14	13	Tagliarsi le unghie dei piedi	
8	8	Uscire di casa	saltuarie
9	9	Camminare per almeno 400 m.	
10	7	Fare le scale	
15	15	Fare lavori domestici pesanti	
(a) ordine di presentazione			
(b) ordine ottenuto con il criterio deterministico applicato sui dati non dicotomizzati (non presentato in questa sede)			

I primi 7 item, indicativi di un livello minimale di autosufficienza comprendono attività che sono essenziali alla cura di sé che devono essere svolte più volte al giorno, a intervalli brevi e/o imprevedibili. Carenze nella possibilità di compiere tali attività sono proprie degli individui in impellenti condizioni di bisogno che necessitano di interventi complessi, pressoché continui a domicilio e la cui messa in opera richiede, almeno in parte, personale specializzato. Tali servizi non possono prescindere da una valutazione dell'assistenza fornita dai familiari o dalla rete sociale. Infatti questi soggetti per la loro incapacità di vivere da soli, sono presenti nel nostro studio, dunque al loro domicilio, solo perché possono contare su adeguati supporti familiari o informali.
Gli item 11,12,13,14 identificano attività che devono essere svolte 1-2 volte al giorno. I deficit in tali attività possono essere colmati da una struttura familiare solida senza alcun ausilio esterno. Per questo, la decisione di fornire un servizio di assistenza deve derivare da una valutazione dei "supporti informali". Anche nei casi in cui tale assistenza si renda necessaria (per esempio: soggetti che vivono da soli o con un coniuge anziano) sarà comunque realizzabile con un impegno limitato perché un operatore può assistere più di un soggetto nel corso della giornata. In alcuni casi, il deficit è colmabile con servizi generali (pasti a domicilio, servizi di lavanderia ecc.) che forniscono l'intervento richiesto con strutture e personale generico.
Gli item 8,9,10,15 si riferiscono ad attività saltuarie che devono essere svolte con intervalli anche variabili di 2-7 giorni. I soggetti con deficit solo in quest'area sono spesso in grado di vivere autonomamente purché dispongano di un aiuto periodico, anche esterno, familiare o istituzionale.
Per rendere interpretabili i dati della scala sono stati individuati 3 livelli di autosufficienza nell'area generale; la validazione di ciascuna delle tre sottoaree è stato effettuato un confronto con i risultati ottenuti attraverso una rilevazione parallela effettuata mediante un questionario somministrato alla persona (generalmente un familiare) che presta la maggiore assistenza. Ciò ha consentito di individuare le soglie che nella scala continua dell'autosufficienza corrispondono ai vari livelli di assistenza. Del questionario sul carico assistenziale sono stati considerati, per questa validazione,

gli item relativi a:

- bisogno di aiuto fisico per la cura di sé
- bisogno di aiuto per i compiti domestici
- bisogno di sorveglianza diurna e notturna

Ai fini di una classificazione dei soggetti sulla base del bisogno di assistenza i gruppi sono definiti come:

1. autosufficienti
2. parzialmente disabili: necessitano di aiuto saltuario
3. gravemente disabili: necessitano di assistenza continuativa.

Per la definizione di queste soglie i punteggi dell'autosufficienza sono stati messi in relazione con le variabili del questionario che consentivano di valutare il carico assistenziale. Il punto mediano per ciascuna tipologia di assistenza rappresenta, in questo tipo di distribuzioni, l'indice di tendenza centrale più stabile. La trasformazione *logit* condotta per una verifica di questa prima analisi ha confermato la stabilità del punto mediano. Con la trasformazione *logit*, infatti, si ricava un punto centrale, ossia un valore analogo, come significato, alla mediana, ma rispetto a questa ottenuto dall'interpolazione della funzione $0.5\log\left(\frac{p_i}{1-p_i}\right)$, in cui p_i è la proporzione di soggetti con un dato punteggio o meno. Il *logit 50* rappresenta così un indice centrale più stabile della mediana, specialmente in presenza di vari soggetti a pari merito.

VALORE CENTRALE (50° CENTILE) PER LA VARIABILE ASSISTENZA DIURNA E NOTTURNA								
		Frequenza dell'assistenza						
		Diurna e notturna			Notturna		Diurna	
Frequenza di intervento		+++	++	+	++	+	++	+
Punteggio di autosufficienza	Globale	27	40	55	57	62	59	75
	Cura di sé	29	52	68	70	69	71	80

\+ saltuaria
++ periodica
+++ continua

In questo modo la similitudine fra mediana e *logit 50* e la vicinanza fra i quartili e la mediana concorrono a fissare i punti di soglia dei vari gruppi di autosufficienza definiti in termini di livelli di assistenza.

I punti di divisione sono stati ricavati dalle distribuzioni di frequenza dei punteggi in tabelle di contingenza con le variabili del questionario dell'assistenza.

<table>
<tr><td colspan="6">SOGLIE FRA LE CATEGORIE DELL'AUTOSUFFICIENZA GLOBALE (CURA DI SÉ, ATTIVITA' QUOTIDIANE, ATTIVITA SALTUARIE) E DELLA SOLA CURA DI SÉ.</td></tr>
<tr><td colspan="2" rowspan="2"></td><td colspan="4">GRUPPO</td></tr>
<tr><td>Dipendenza completa</td><td colspan="2">Dipendenza parziale</td><td>Autosufficienza</td></tr>
<tr><td colspan="2">Frequenza di intervento</td><td>continuo</td><td>periodico</td><td>saltuario+</td><td>no</td></tr>
<tr><td rowspan="2">Punteggio di autosufficienza</td><td>Globale</td><td colspan="4">55 65 80</td></tr>
<tr><td>Cura di sé</td><td colspan="4">65 75 (*)</td></tr>
<tr><td colspan="6">(*) La soglia fra dipendenza limitata e autosufficienza per la cura di sé non viene riportata perché presenta un valore troppo alto e instabile: la cura di sé è richiesta con continuità, e non è risolvibile con interventi saltuari. Gli alti valori delle soglie dell'autosufficienza nella cura di sé sono logicamente dovuti al fatto che una perdita di autonomia anche lieve in quest'area comporta un disagio che deve essere necessariamente colmato.</td></tr>
</table>

Il nostro studio mette chiaramente in evidenza, per esempio, che i 6 item che riguardano le attività della cura di sé permettono da soli di identificare una fascia di popolazione in stato di più impellente bisogno di aiuto fisico per la vita quotidiana.

E' importante sottolineare come i livelli di autosufficienza sono stati definiti sulla base dello studio del rapporto tra autosufficienza e supporti informali ovvero mediante l'identificazione delle modalità *assistenziali* che gli individui, non potendo attendere l'attuazione dei servizi istituzionali, hanno già messo spontaneamente in atto.