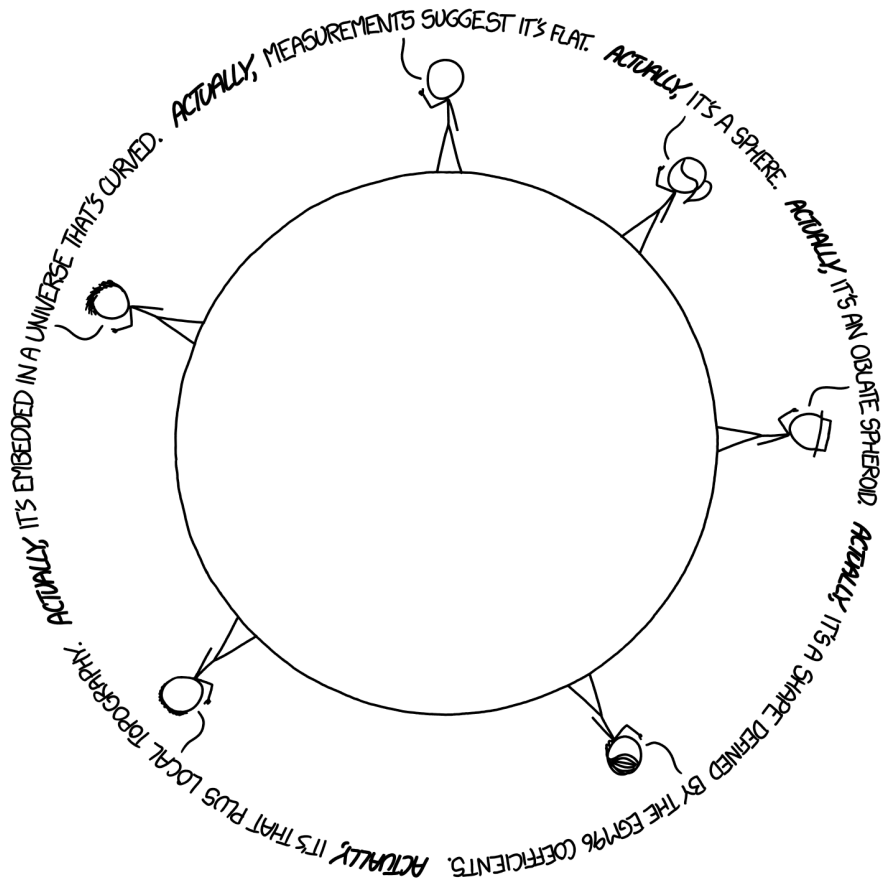


Corrado Caudek

***Appunti di Costruzione e
validazione di strumenti di
misura dell'efficacia
dell'intervento psicologico
in neuropsicologia –
B020881 (B213)***



Appunti di Costruzione e validazione di strumenti di
misura dell'efficacia dell'intervento psicologico in
neuropsicologia – AA 2021/2022





Indice

Elenco delle figure	xiii
Elenco delle tabelle	xv
Prefazione	xvii
I La teoria classica dei test	1
1 Fondamenti teorici	3
1.1 Valutazione psicometrica come ragionamento inferenziale	3
1.2 La Teoria Classica	5
1.3 Le due componenti del punteggio osservato	6
1.3.1 Il punteggio vero	6
1.3.2 Somministrazioni ripetute	7
1.3.3 Le assunzioni sul punteggio ottenuto	8
1.4 L'errore standard della misurazione σ_E	8
1.5 Assiomi della Teoria Classica	9
1.6 L'attendibilità del test	9
1.6.1 La varianza del punteggio osservato	10
1.6.2 La covarianza tra punteggio osservato e punteggio vero	10
1.6.3 Definizione e significato dell'attendibilità	11
1.7 Attendibilità e modello di regressione lineare	11
1.7.1 Simulazione	12
1.8 Misurazioni parallele e affidabilità	17
1.8.1 La correlazione tra misurazioni parallele	18
1.8.2 La correlazione tra due forme parallele del test	19
1.8.3 La correlazione tra punteggio osservato e punteggio vero	20
1.8.4 I fattori che influenzano l'attendibilità	20
1.9 Metodi alternativi per la stima del coefficiente di attendibilità	21

2	L'incertezza della misura	23
2.1	La stima dell'errore standard della misurazione	23
2.1.1	Interpretazione	24
2.1.2	Simulazione	26
2.2	Dimostrazione	29
2.3	Intervallo di confidenza per il punteggio vero e σ_E . . .	30
3	La stima del punteggio vero	31
3.1	Il paradosso di Kelley	31
3.2	L'errore standard della stima	36
3.3	Intervallo di confidenza per il punteggio vero	38
3.4	Cut-off	40
3.5	Procedure alternative	42
II	L'analisi fattoriale	43
4	Utilizzo e costruzione di test psicometrici	45
4.1	Caratteristiche dell'analisi fattoriale	46
4.1.1	L'analisi fattoriale esplorativa e confermativa . .	47
4.1.2	Assunzioni dell'analisi fattoriale	47
4.1.3	Sviluppo storico dell'analisi fattoriale	48
4.1.4	Fasi dell'analisi fattoriale	51
5	Il modello unifattoriale	53
5.1	Modello monofattoriale	53
5.2	Correlazione parziale	55
5.3	Principio base dell'analisi fattoriale	58
5.3.1	Vincoli sulle correlazioni	60
5.3.2	Teoria dei Due Fattori	60
5.3.3	Annullamento della tetrade	62
5.3.4	Metodo del centroide	63
5.3.5	Introduzione a <code>lavaan</code>	63
6	Il modello statistico dell'analisi fattoriale	75
6.1	Modello monofattoriale	75
6.2	Covarianza tra un indicatore e il fattore comune	76
6.3	Espressione fattoriale della varianza	76
6.4	Covarianza tra due variabili manifeste	80
6.5	Correlazioni osservate e correlazioni riprodotte dal modello	80
6.6	Bontà di adattamento del modello ai dati	81

6.7	L'errore standard della misurazione e il modello fattoriale	82
6.7.1	Un caso concreto	84
7	Il modello multifattoriale	91
7.1	Modello multifattoriale: fattori ortogonali	91
7.1.1	Assunzioni del modello multifattoriale	92
7.1.2	Interpretazione dei parametri del modello	93
7.1.3	Covarianza tra variabili e fattori	93
7.1.4	Espressione fattoriale della varianza	94
7.1.5	Espressione fattoriale della covarianza	95
7.2	Modello fattoriale: Fattori obliqui	101
7.2.1	Covarianza teorica tra variabili e fattori	101
7.2.2	Espressione fattoriale della varianza	102
7.2.3	Covarianza teorica tra due variabili	103
7.3	EFA con <code>lavaan</code>	110
8	Path Analysis	121
8.1	Path diagram	121
8.2	Path analysis e regressione multipla	122
8.3	Effetti diretti e indiretti	124
8.4	Le regole di Wright	124
8.4.1	Scomposizione delle correlazioni (covarianze)	125
8.4.2	Scomposizione della varianza	126
8.5	Come calcolare i coefficienti di percorso?	127
8.6	Path analysis e software	128
8.7	Oltre la regressione multipla	132
9	Attendibilità e modello fattoriale	133
9.1	Modello fattoriale e CTT	134
9.2	Classi di modelli	135
9.2.1	Indicatori congenerici	135
9.2.2	Indicatori tau-equivalenti	136
9.2.3	Indicatori paralleli	137
9.2.4	Indicatori strettamente paralleli	137
9.3	Metodo dei minimi quadrati non pesati	138
9.4	Varianza del punteggio totale di un test	139
9.5	Stima dell'attendibilità	140
9.5.1	Coefficiente omega	140
9.5.2	Coefficiente α di Cronbach	143
9.5.3	La formula "profetica" di Spearman-Brown	147

9.6	Attendibilità e scala di misura	151
9.7	Quale indice usare?	152
10	CFA: confronto tra modelli	153
10.1	Modello congenerico	154
10.2	Modello tau-equivalente	155
10.3	Modello parallelo	155
10.4	Il test del χ^2	157
III	Costruzione di strumenti psicometrici	159
11	Tipologie dei test psicometrici	161
11.0.1	Test orientati al criterio	161
11.0.2	Test basati sulla norma	162
11.1	Variabili latenti e sviluppo di uno strumento psicometrico	163
11.1.1	Chiarificazione del costrutto di interesse	163
11.1.2	Operazionalizzazione del costrutto di interesse	165
11.1.3	Rassegna della letteratura rilevante	166
11.1.4	Analisi concettuale del costrutto	167
11.1.5	Metodi di ricerca qualitativi	168
11.2	Lo sviluppo dello strumento	168
11.2.1	Formato	168
11.2.2	Composizione tipografica	170
11.2.3	Istruzioni ai soggetti	170
11.2.4	La costruzione degli item	172
11.3	Numero delle opzioni di risposta	174
11.3.1	Item a codifica inversa	175
11.4	Numero di item	176
11.5	Numero di soggetti	177
12	Valutazione della matrice di correlazione	179
12.1	Ispezione visiva della matrice di correlazione	179
12.2	Sfericità di Bartlett	180
12.3	Test di adeguatezza campionaria di Kaiser-Meyer-Olkin	180
12.3.1	Matrice anti-immagine	181
13	L'estrazione dei fattori	183
13.1	Metodo delle componenti principali	183
13.2	Metodo dei fattori principali	188
13.3	Metodo dei fattori principali iterato	190
13.3.1	Casi di Heywood	191

13.4 Metodo di massima verosimiglianza	192
14 Il numero dei fattori	195
14.1 Quota di varianza spiegata	195
14.2 Valore medio degli autovalori	196
14.3 Scree test	197
14.4 Parallel analysis	197
14.5 Test di bontà di adattamento	198
15 La rotazione fattoriale	207
15.1 Indeterminatezza della soluzione fattoriale	207
15.2 Parsimonia e semplicità	210
15.2.1 Il criterio della “struttura semplice”	210
15.3 Rotazione nello spazio geometrico	211
15.3.1 Rotazione ortogonale	211
15.3.2 Vincoli alla rotazione	212
15.3.3 Metodo grafico	213
15.3.4 Metodi di rotazione ortogonale	218
15.3.5 Metodi di rotazione obliqua	219
15.4 Matrice dei pesi fattoriali e matrice di struttura	220
16 Come valutare e rifinire la soluzione fattoriale	225
16.1 Valutazione della matrice pattern	225
16.1.1 Item con basse saturazioni su tutti i fattori	225
16.1.2 Item con saturazioni eevate su più di un fattore	226
16.2 Valutazione dell’attendibilità	226
16.2.1 La procedura split-half	226
16.2.2 L’indice α di Cronbach	227
16.2.3 L’attendibilità quale stabilità temporale	228
16.2.4 Forme parallele	228
16.3 Selezione di un sottoinsieme di item	229
16.4 Attendibilità e numero di item	232
16.4.1 Numero di item e affidabilità	235
16.5 Analisi degli item	235
16.5.1 Difficoltà degli item	236
16.5.2 Correzione per guessing	237
16.5.3 Discriminatività	238
16.5.4 Potere discriminante dell’item e analisi fattoriale	239
16.5.5 Punteggio sull’item e punteggio totale	239
16.5.6 Relazioni tra coppie di item	240

16.5.7	Ridondanza	240
16.5.8	Massimizzazione della varianza del punteggio totale	240
17	Invarianza di misura	243
17.1	Intercette degli item	243
17.2	Terminologia	245
17.3	Un esempio concreto	245
18	Indici di bontà dell'adattamento	249
18.1	Stima del modello	249
18.2	Massima verosimiglianza	249
18.3	Identificabilità del modello	251
18.4	Bontà dell'adattamento	252
18.4.1	Chi quadrato	252
18.5	Misure di adeguamento per il confronto	254
18.5.1	CFI	254
18.6	Misure di adeguamento parsimonioso	254
18.6.1	TLI	254
18.7	Misure di adeguamento assoluto	255
18.7.1	RMSEA	255
18.7.2	RMRS	255
18.8	Interpretazione	256
18.9	Una nota di cautela	256
19	Curve di crescita latente	259
19.1	Dati longitudinali	259
19.1.1	Un esempio concreto	260
19.1.2	Un secondo esempio	270
IV	Appendici	1
	Appendice	3
A	Simbologia di base	3
B	Lo studio del caso singolo	5
B.1	Motivazione	5
B.2	Il confronto tra caso singolo e un gruppo di controllo . .	6
B.2.1	Il metodo di Crawford e Howell (1998)	6
B.2.2	Il metodo di Barton et al. (2002)	8

<i>Contents</i>	xi
B.3 Lo studio delle dissociazioni	9
B.4 La misura del cambiamento	12
B.4.1 Cambiamento nel caso singolo e confronto con un campione di controllo	15
C La valutazione dell'intervento psicologico	19
C.1 Efficacia dell'intervento ed efficienza clinica	20
D La dimensione dell'effetto del trattamento	23
D.1 Risposta continua	24
D.1.1 Indici d di Cohen e g di Hedges	24
D.2 Risposta binaria	25
D.2.1 Differenza tra due proporzioni	28
D.2.2 Rischio relativo	29
D.2.3 Odds ratio	29
D.2.4 Logit	30
D.2.5 Rapporto delle quote	32
D.2.6 Logaritmo del rapporto delle quote	33
D.3 Tipi fondamentali di indagine	35
D.3.1 Studi prospettici	35
D.3.2 Studi trasversali	35
D.3.3 Studi longitudinali retrospettivi	36



Elenco delle figure

1.1	Simulazione della relazione tra punteggio osservato e punteggio vero per 100 individui in base alle assunzioni della CTT.	14
8.1	Esempio di path diagram.	122
8.2	Path diagram per il modello di regressione multipla con due regressori.	123
8.3	Path diagram per una relazione a catena.	124
9.1	Modello fattoriale che mette in relazione i punteggi osservati e i punteggi veri.	133
15.1	Modello fattoriale con due fattori comuni ortogonali. . .	220
15.2	Modello fattoriale con due fattori comuni dopo una rotazione obliqua.	222
19.1	Modello di crescita latente.	264



Elenco delle tabelle

5.1	Matrice di correlazioni nella quale tutte le correlazioni parziali tra le variabili Y al netto dell'effetto di ξ sono nulle.	59
5.3	Factor Loadings	70
7.1	Factor Loadings	100
7.3	Factor Loadings	105
16.1	Linee guida per l'interpretazione dell'indice di discriminatività D	238



Prefazione

La presente dispensa contiene il materiale delle lezioni dell'insegnamento di *Costruzione e validazione di strumenti di misura dell'efficacia dell'intervento psicologico in neuropsicologia* B020881 (B213) rivolto agli studenti del secondo anno del Corso di Laurea Magistrale in Psicologia Clinica e della Salute e Neuropsicologia (curriculum: assessment e intervento psicologici in neuropsicologia - E21), A.A. 2021-2022. L'insegnamento si propone di fornire agli studenti un'introduzione all'assessment psicologico, ovvero un insieme di conoscenze/competenze che si pongono all'intersezione tra psicometria, statistica e informatica.

Nello specifico, l'insegnamento si focalizzerà sull'analisi fattoriale confermativa (*confirmatory factor analysis*, CFA) e sull'analisi fattoriale esplorativa, (*explorative factor analysis*, EFA), cioè sugli strumenti che vengono usati durante il processo di sviluppo dei test psicometrici, ovvero che vengono usati per esaminare la struttura latente di una scala psicologica (ad esempio un questionario). In questo contesto, la CFA viene utilizzata per verificare il numero di dimensioni sottostanti gli indicatori (fattori) e l'intensità delle relazioni item-fattore (saturazioni fattoriali). La CFA consente anche di capire di come dovrebbe essere svolto lo scoring di un test. Quando la struttura latente è multifattoriale (cioè, a due o più fattori), il numero di fattori è indicativo del numero di sottoscale e di come esse dovrebbero essere codificate. La CFA è un importante strumento analitico anche per altri aspetti della valutazione psicometrica. Può essere utilizzata per stimare l'affidabilità di scala dei test psicometrici in modo da evitare i problemi della teoria classica dei test (ad es. alpha di Cronbach). Dati i recenti progressi nell'analisi dei dati categoriali, ora la CFA offre un quadro analitico comparabile a quello offerto dalla teoria di risposta agli item (IRT). In effetti, secondo [Brown \(2015\)](#), la CFA offre una maggiore flessibilità analitica rispetto al modello IRT tradizionale.

Un costrutto è un concetto teorico che può essere operazionalizzato nei termini di un fattore. In psicologia clinica, psichiatria e neuropsicologia,

ad esempio, i disturbi mentali sono costrutti manifestati da vari insiemi di sintomi che sono riportati dal paziente o osservati da altri. La CFA è uno strumento analitico indispensabile per la validazione dei costrutti psicologici. I risultati della CFA possono fornire prove convincenti della validità convergente e discriminante dei costrutti teorici. La validità convergente è indicata dall'evidenza che diversi indicatori di costrutti teoricamente simili o sovrapposti sono fortemente correlati. La validità discriminante è indicata dai risultati che mostrano che gli indicatori di costrutti teoricamente distinti sono altamente incorrelati. Un punto di forza fondamentale degli approcci CFA per la costruzione e la validazione di uno strumento psicometrico è che le risultanti stime di validità convergente e discriminante sono corrette per l'errore di misurazione. Pertanto, la CFA fornisce un quadro analitico migliore rispetto ai metodi tradizionali che non tengono conto dell'errore di misurazione (ad esempio, gli approcci ordinari ai minimi quadrati come la correlazione/regressione multipla, i quali presuppongono che le variabili nell'analisi siano prive di errori di misurazione).

Spesso, parte della covariazione delle misure osservate è dovuta a fonti diverse dai fattori latenti di interesse. Questa covariazione aggiuntiva spesso riflette la varianza del metodo utilizzato per la misurazione. Gli effetti del metodo possono verificarsi anche all'interno di un'unica modalità di valutazione. Ad esempio, effetti del metodo sono solitamente presenti nei questionari che contengono una combinazione di elementi formulati positivamente e negativamente. Sfortunatamente, l'EFA non è in grado di stimare gli effetti del metodo. In effetti, l'uso di EFA quando esistono effetti del metodo può produrre risultati fuorvianti, ovvero suggerire la presenza di fattori aggiuntivi che corrispondono invece ad artefatti della misurazione. Nella CFA, invece, gli effetti del metodo possono essere specificati come parte della teoria dell'errore del modello di misurazione.

Un altro punto di forza della CFA è la sua capacità di affrontare il problema della generalizzabilità del modello di misurazione tra gruppi di individui o nel tempo. La valutazione dell'invarianza della misura è un aspetto importante dello sviluppo del test. Se un test è destinato a essere somministrato in una popolazione eterogenea, si dovrebbe stabilire che le sue proprietà di misurazione sono equivalenti in sottogruppi della popolazione (es. sesso, razza). Si dice che un test è distorto quando alcuni dei suoi elementi non misurano il costrutto sottostante in modo comparabile tra gruppi di rispondenti. Il test fornisce una stima distorta se,

ad esempio, per un dato livello di vera intelligenza, gli uomini tendono a ottenere un punteggio di QI più alto rispetto alle donne. Il problema della generalizzabilità della validità del costrutto tra i gruppi può essere affrontato nella CFA esaminando gruppi multipli mediante modelli MI-MIC (indicatori multipli, cause multiple). Inoltre, è possibile chiedersi se il modello di misurazione sia equivalente tra i gruppi. Le soluzioni CFA a gruppi multipli vengono anche utilizzate per esaminare l'invarianza della misurazione longitudinale. Questo è un aspetto molto importante dell'analisi delle variabili latenti dei progetti di misure ripetute. In assenza di tale valutazione, non è possibile determinare se il cambiamento temporale in un costrutto sia dovuto a un vero cambiamento dei rispondenti o a cambiamenti nel modo di rispondere alla scala nel tempo. L'analisi a gruppi multipli può essere applicata a qualsiasi tipo di modello CFA. Ad esempio, queste procedure possono essere incorporate nell'analisi dei dati multitratto-multimetodo per esaminare la generalizzabilità della validità del costrutto tra gruppi.

In questo insegnamento la discussione delle tecniche della CFA sarà preceduta da un'introduzione relativa alla EFA e la teoria classica dei test. La EFA, infatti, può essere concepita il metodo che viene utilizzato nei primi passi dello sviluppo di una scala psicometria, mentre la teoria classica dei test rappresenta la cornice teorica di partenza, di cui la CFA e i modelli di equazioni strutturali costituiscono uno sviluppo.

L'insegnamento pone una grande enfasi non solo sulla comprensione dei concetti teorici necessari per la costruzione e la validazione di uno strumento di misura in psicologia, ma anche sulla capacità di applicare tali concetti in situazioni concrete. Di conseguenza, la discussione dei concetti sarà sempre accompagnata da applicazioni pratiche. Tali applicazioni richiedono l'uso di un software. In questo insegnamento useremo R ([R Core Team, 2021](#)) quale linguaggio di programmazione probabilistica e, tra gli altri, il pacchetto `lavaan` che consente di svolgere le analisi statistiche della CFA e della EFA ([Beaujean, 2014](#)). La teoria classica dei test verrà descritta con riferimento al classico testo di [Lord and Novick \(1968\)](#). Questa dispensa, inoltre, segue da vicino la trattazione della CFA fornita nei testi di [McDonald \(2013\)](#) e di [Brown \(2015\)](#).

Trattando di argomenti avanzati, questo insegnamento presuppone la conoscenza di base dei concetti fondamentali della teoria delle probabilità; presuppone inoltre il possesso delle conoscenze di base necessarie per procedere all'utilizzo di R. Informazioni su tali argomenti sono forniti

xx

nella dispensa di Psicometria (A.A. 2021-2022).

Prefazione

Corrado Caudek
Marzo 2022

Parte I

La teoria classica dei test



1

Fondamenti teorici

1.1 Valutazione psicometrica come ragionamento inferenziale

In apparenza, i test psicometrici sono solo dei test. Somministriamo un test, otteniamo un punteggio ed è naturale pensare che sia tutto lì. Nonostante le apparenze, la valutazione psicologica e neuropsicologica non consiste soltanto nell'assegnare dei punteggi: si tratta di ragionare su ciò che osserviamo di quello che le persone dicono, fanno o producono, in maniera tale da giungere a delle concezioni più ampie di tali persone a proposito di aspetti che non abbiamo – e spesso non possiamo – osservare. Più specificamente, possiamo considerare la valutazione psicologica e neuropsicologica come un esempio di ragionamento che fa uso di modelli probabilistici per giungere a delle spiegazioni, previsioni o conclusioni.

I dati osservati diventano un'evidenza quando sono ritenuti rilevanti per l'inferenza desiderata attraverso l'instaurazione di relazioni tra i dati e l'obiettivo dell'inferenza. Spesso utilizziamo dati provenienti da più fonti. Queste possono essere di tipo simile (ad esempio, item di test aventi lo stesso formato) o di tipo molto diverso (ad esempio, il curriculum di un richiedente oltre al colloquio, la storia medica della famiglia di un paziente, ...). Le evidenze possono essere contraddittorie (ad esempio, uno studente riesce a svolgere un compito difficile ma fallisce in un altro facile) e quasi sempre non sono del tutto conclusive.

Queste caratteristiche hanno due implicazioni. In primo luogo, è difficile capire cosa le evidenze implicano. I processi inferenziali sono sempre complessi. In secondo luogo, a causa della natura non conclusiva delle evidenze disponibili, non siamo mai del tutto certi delle nostre inferenze. Per affrontare tale incertezza, la teoria psicometrica ci fornisce gli strumenti che ci possono aiutare nel processo inferenziale, dai dati disponibili alle decisioni che prendiamo.

Un secolo fa, la relazione tra prestazioni osservate, da un lato, e l'abilità inosservabile del rispondente, dall'altro, iniziò a essere formalizzata nei termini dell'*errore di misurazione*. Gulliksen (1961) ha descritto "il problema centrale della teoria dei test" come "la relazione tra l'abilità dell'individuo e il suo punteggio osservato sul test" (p. 101). Tale caratterizzazione è valida ancora oggi, con una definizione opportunamente ampia di "abilità" e di "punteggio sul test" che sia in grado di comprendere le diverse forme di assessment psicologico e neuropsicologico. Comprendere e essere in grado di rappresentare la relazione tra le prestazioni osservate e la capacità soggiacente è dunque fondamentale per le forme di ragionamento che vengono impiegate nella valutazione psicologica e neuropsicologica.

Come risultato dell'errore di misurazione, i ragionamenti che compiamo nella valutazione psicologica e neuropsicologica costituiscono un esempio di ragionamento in condizioni di incertezza. A causa della natura imperfetta della misurazione e dell'incompletezza dell'informazione disponibile, le nostre inferenze sono incerte e possono essere sempre invalidate o riviste. Ragionare da ciò che è parziale (ciò che vediamo uno paziente dire, fare o produrre) a ciò che è generale (la "vera" abilità del paziente) è necessariamente incerto, e le nostre inferenze o conclusioni sono sempre prone ad errori.

Quali strumenti devono essere impiegati per affrontare la nostra incertezza sulla relazione che intercorre tra prestazioni osservate e abilità soggiacenti? Secondo Lewis, molti dei progressi nella teoria psicometrica sono resi possibili "trattando lo studio della relazione tra le risposte agli item di un test e il tratto ipotizzato di un individuo come problema di inferenza statistica" (Lewis, 1986). Una connessione diretta tra errore di misura e approccio probabilistico è stata anche proposta da Samejima: "There may be an enormous number of factors eliciting a student's specific overt reactions to a stimulus, and, therefore, it is suitable, even necessary, to handle the situation in terms of the probabilistic relationship between the two" (Samejima, 1983).

Questo punto di vista è diventato quello dominante nella psicomетria moderna e sottolinea l'utilità di utilizzare il linguaggio e gli strumenti della teoria della probabilità per comunicare il carattere parziale dei dati di cui dispone lo psicologo e l'incertezza delle inferenze che ne derivano.

I reattivi psicologici possono essere costruiti e validati mediante vari approcci probabilistici: la Teoria Classica dei test (*classical test theory*, in

breve CTT) e la teoria di risposta all'item (*item response theory*, in breve IRT) sono quelli più noti. Recentemente, il problema della valutazione psicologica è stato anche formulato in un'ottica bayesiana. In questo insegnamento esamineremo gli approcci della CTT e dell'IRT, ma non quello bayesiano.

1.2 La Teoria Classica

La CTT nasce alla fine dell'Ottocento (Alfred Binet e altri, 1894) allo scopo di studiare l'attendibilità e la validità dei risultati dei questionari utilizzati per valutare le caratteristiche psico-sociali, non direttamente osservabili, delle persone esaminate. L'impiego su vasta scala e lo sviluppo della CTT ha inizio negli anni Trenta, anche se il modello formale su cui tale teoria si basa viene proposta da Spearman all'inizio del Novecento (Spearman, 1904). La tecnica dell'analisi fattoriale esplorativa (*Exploratory Factor Analysis*, EFA), verrà poi affinata da Thurstone (1947) alla fine della seconda guerra mondiale. Tra la fine degli anni '60 e gli inizi degli anni '70, Jöreskog (1969) sviluppa l'analisi fattoriale confermativa (*Confirmatory Factor Analysis*, CFA). Negli anni '70, l'analisi fattoriale viene integrata con la path analysis nel lavoro di Jöreskog (1978) che dà origine ai modelli di equazioni strutturali (*Structural Equation Modeling*, SEM).

Iniziamo qui ad esaminare queste tecniche psicometriche prendendo in esame, per prima, la teoria classica dei test. Seguiremo la trattazione proposta da Lord and Novick (1968).

L'equazione fondamentale alla quale si riconduce la teoria classica dei test è quella che ipotizza una relazione lineare e additiva tra il punteggio osservato di un test (X), la misura della variabile latente (T) e la componente casuale dell'errore (E). Un punto cruciale nella CTT è l'entità della varianza dell'errore. Minore è la varianza dell'errore, più accuratamente il punteggio reale viene riflesso dai nostri punteggi osservati. In un mondo perfetto, tutti i valori di errore sarebbero uguali a 0. Cioè, ogni partecipante otterrebbe il punteggio esatto. Questo però non è possibile. Pertanto, abbiamo una certa varianza negli errori. La corrispondente deviazione standard di tali errori ha il un nome: si chiama *errore standard di misurazione*, indicato da σ_E . Uno dei principali obiettivi della CTT è

quello di ottenere una stima di σ_E .

1.3 Le due componenti del punteggio osservato

CTT si occupa delle relazioni tra X , T ed E . La CTT si basa su un modello relativamente semplice in cui il punteggio osservato, il punteggio vero (cioè l'abilità inosservabile del rispondente) e l'errore aleatorio di misurazione sono legati da una relazione lineare. Indicati con $T_{\nu j}$ (*true score*) l'abilità latente da misurare dell'individuo ν nella prova j , con $X_{\nu j}$ la variabile osservata (*observed score*) per l'individuo ν nella prova j e con $E_{\nu j}$ l'errore aleatorio di misurazione, il modello si rappresenta con

$$X_{\nu j} = T_{\nu} + E_{\nu j}.$$

Dunque, in base alla (1.3) il punteggio osservato $X_{\nu j}$ differisce da quello vero $T_{\nu j}$ a causa di una componente di errore aleatorio $E_{\nu j}$. Uno degli obiettivi centrali della CTT è quello di quantificare l'entità di tale errore. Vedremo come questa quantificazione verrà fornita nei termini dell'attendibilità del test. L'attendibilità (o affidabilità) rappresenta l'accuratezza con cui un test può misurare il punteggio vero (Coaley, 2014):

- Se l'attendibilità è grande, σ_E è piccolo: X ha un piccolo errore di misurazione e sarà vicino a T .
- Se l'attendibilità è piccola, σ_E è grande: X presenta un grande errore di misurazione e si discosterà molto da T .

1.3.1 Il punteggio vero

La @ref(eq:observed-true_plus-error) ci dice che il punteggio osservato è dato dalla somma di due componenti: una componente sistematica (il punteggio vero) e una componente aleatoria (l'errore di misurazione). Ma che cos'è il punteggio vero? La CTT considera un reattivo psicologico come una selezione aleatoria di item da un universo/popolazione di item attinenti al costrutto da misurare (Nunnally, 1994; Kline, 2013). Se il reattivo psicologico viene concepito in questo modo, il punteggio vero diventa il punteggio che un rispondente otterrebbe se fosse misurato su tutto l'universo degli item proprio del costrutto in esame. L'errore di

misurazione riflette dunque il grado in cui gli item che costituiscono il test non riescono a rappresentare l'intero universo degli item attinenti al costrutto.

In maniera equivalente, il punteggio vero può essere concepito come il punteggio non “distorto” da componenti estranee al costrutto, ovvero da effetti di apprendimento, fatica, memoria, motivazione, eccetera. Essendo concepita come del tutto aleatoria (ovvero, priva di qualunque natura sistematica), la componente aleatoria non introduce dei bias nella tendenza centrale della misurazione.

Il punteggio vero è concepito come un punteggio inosservabile che corrisponde al valore atteso di infinite realizzazioni del punteggio ottenuto:

$$T = \mathbb{E}(X) \equiv \mu_X \equiv \mu_T.$$

In altri termini, secondo la definizione di Lord e Novick (1968), e facendo riferimento alla seconda definizione presentata sopra, il punteggio vero è concepito come la media dei punteggi che un soggetto otterrebbe se il test venisse somministrato ripetutamente nelle stesse condizioni, in assenza di effetti di apprendimento e/o fatica.

1.3.2 Somministrazioni ripetute

Nella formulazione del modello della CTT si possono distinguere due tipi di esperimenti aleatori: uno che considera l'unità di osservazione (l'individuo) come campionaria, l'altro che considera il punteggio, per un determinato individuo, come campionario. Un importante risultato è dato dall'unione dei due esperimenti, ovvero dalla dimostrazione che i risultati della CTT, la quale è stata sviluppata ipotizzando ipotetiche somministrazioni ripetute del test allo stesso individuo sotto le medesime condizioni, si generalizzano al caso di una singola somministrazione del test ad un campione di individui ([Allen and Yen, 2001](#)). In base a questo risultato, se consideriamo la somministrazione del test ad una popolazione di individui, allora diventa più facile dare un contenuto empirico alle quantità della CTT:

- σ_X^2 è la varianza del punteggio osservato nella popolazione,
- σ_T^2 è la varianza dei punteggi vero nella popolazione,
- σ_E^2 è la varianza della componente d'errore nella popolazione.

1.3.3 Le assunzioni sul punteggio ottenuto

La CTT *assume* che la media del punteggio osservato X sia uguale alla media del punteggio vero,

$$\mu_X \equiv \mu_T,$$

in altri termini, assume che il punteggio osservato fornisca una stima statisticamente corretta dell'abilità latente (punteggio vero). In pratica, il punteggio osservato non sarà mai uguale all'abilità latente, ma corrisponde solo ad uno dei possibili punteggi che il soggetto può ottenere, subordinatamente alla sua abilità latente. L'errore della misura è la differenza tra il punteggio osservato e il punteggio vero: $E \equiv X - T$.

In base all'assunzione secondo cui il valore atteso dei punteggi è uguale alla media del valore vero, segue che

$$\mathbb{E}(E) = \mathbb{E}(X - T) = \mathbb{E}(X) - \mathbb{E}(T) = \mu_X - \mu_T = 0,$$

ovvero, il valore atteso degli errori è uguale a zero.

1.4 L'errore standard della misurazione σ_E

La radice quadrata di σ_E^2 , ovvero la deviazione standard degli errori, è la quantità fondamentale della CTT ed è chiamata *errore standard della misurazione*. La stima dell'errore standard della misurazione costituisce uno degli obiettivi più importanti della CTT. Ricordiamo che la deviazione standard è simile (non identica) alla media del valore assoluto degli scarti dei valori di una distribuzione dalla media. Possiamo dunque utilizzare questa proprietà per descrivere il modo in cui la CTT interpreta σ_E .

L'*errore standard della misurazione* σ_E ci dice qual è, approssimativamente, la variazione attesa del punteggio osservato, se il test venisse somministrato un'altra volta al rispondente nelle stesse condizioni.

1.5 Assiomi della Teoria Classica

La CTT *assume* che gli errori siano delle variabili aleatorie incorrelate tra loro

$$\rho(E_i, E_k | T) = 0, \quad \text{con } i \neq k,$$

e incorrelate con il punteggio vero,

$$\rho(E, T) = 0,$$

le quali seguono una distribuzione gaussiana con media zero e deviazione standard pari a σ_E :

$$E \sim \mathcal{N}(0, \sigma_E).$$

La quantità σ_E è detta errore standard della misurazione.

Sulla base di tali assunzioni la CTT deriva la formula dell'attendibilità di un test. Si noti che le assunzioni della CTT hanno una corrispondenza puntuale con le assunzioni su cui si basa il modello di regressione lineare.

1.6 L'attendibilità del test

Il concetto di attendibilità è strettamente legato alla riproducibilità della misurazione: si riferisce al grado di stabilità, di coerenza interna e di precisione di una procedura di misurazione. Affinché una misurazione psicologica sia utile, deve produrre lo stesso risultato se viene applicata ripetutamente un determinato rispondente. Altri termini che vengono usati sono: affidabilità, costanza e credibilità.

Vedremo nel seguito come il coefficiente di attendibilità fornisce una stima della quota della varianza del punteggio osservato che può essere attribuita all'abilità latente ("punteggio vero", cioè privo di errore di misurazione). In generale, un coefficiente di attendibilità maggiore di 0.80 viene ritenuto soddisfacente perché indica che l'80% o più della varianza

dei punteggi ottenuti è causata da ciò che il test intende misurare, anziché dall'errore di misurazione.

Per definire l'attendibilità, la CTT si serve di due quantità:

- la varianza del punteggio osservato,
- la correlazione tra punteggio osservato e punteggio vero.

Vediamo come queste quantità possano essere ottenute sulla base delle assunzioni del modello statistico che sta alla base della CTT.

1.6.1 La varianza del punteggio osservato

La varianza del punteggio osservato X è uguale alla somma della varianza del punteggio vero e della varianza dell'errore di misurazione.

Dimostrazione. La varianza del punteggio osservato è uguale a

$$\sigma_X^2 = \mathbb{V}(T + E) = \sigma_T^2 + \sigma_E^2 + 2\sigma_{TE}.$$

Dato che $\sigma_{TE} = \rho_{TE}\sigma_T\sigma_E = 0$, in quanto $\rho_{TE} = 0$, ne segue che

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2.$$

□

1.6.2 La covarianza tra punteggio osservato e punteggio vero

La covarianza tra punteggio osservato X e punteggio vero T è uguale alla varianza del punteggio vero.

$$\begin{aligned} \text{Dimostrazione. } \sigma_{XT} &= \mathbb{E}(XT) - \mathbb{E}(X)\mathbb{E}(T) \\ &= \mathbb{E}[(T + E)T] - \mathbb{E}(T + E)\mathbb{E}(T) \\ &= \mathbb{E}(T^2) + \underbrace{\mathbb{E}(ET)}_{=0} - [\mathbb{E}(T)]^2 - \underbrace{\mathbb{E}(E)}_{=0}\mathbb{E}(T) \\ &= \mathbb{E}(T^2) - [\mathbb{E}(T)]^2 \\ &= \sigma_T^2. \end{aligned}$$

□

Da ciò segue che la correlazione tra punteggio osservato X e punteggio vero T è uguale al rapporto tra la deviazione standard del punteggio vero e la deviazione standard del punteggio osservato:

$$\rho_{XT} = \frac{\sigma_{XT}}{\sigma_X \sigma_T} = \frac{\sigma_T^2}{\sigma_X \sigma_T} = \frac{\sigma_T}{\sigma_X}. \quad (1.1)$$

1.6.3 Definizione e significato dell'attendibilità

La CTT definisce attendibilità di un test (o di un item) come il quadrato della correlazione tra punteggio osservato X e punteggio vero T , ovvero come il rapporto tra la varianza del punteggio vero e la varianza del punteggio osservato:

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}.$$

Questa è la quantità fondamentale della CTT e misura il grado di variazione del punteggio vero rispetto alla variazione del punteggio osservato[2].

Dato che $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$, in base alla (1.6.3) possiamo scrivere

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}. \quad (1.2)$$

Questo significa che il coefficiente di attendibilità assume valore 1 se la varianza degli errori σ_E^2 è nulla e assume valore 0 se la varianza degli errori è uguale alla varianza del punteggio osservato. Ciò significa che il coefficiente di attendibilità è un numero contenuto nell'intervallo compreso tra 0 e 1.

1.7 Attendibilità e modello di regressione lineare

Il modello di regressione lineare sta alla base della CTT. Infatti si può dire che tutte le proprietà della CTT che abbiamo discusso in precedenza non sono altro che le caratteristiche di un modello di regressione lineare nel quale

- la variabile dipendente è costituita dai punteggi osservati X , e
- la variabile indipendente corrisponde ai punteggi veri T .

Se rappresentiamo la CTT in questo modo, il coefficiente di attendibilità $\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}$ non diventa altro che la quota di varianza del punteggio osservato X che viene spiegata dal punteggio vero T in base ad un modello lineare con pendenza unitaria e intercetta nulla. Nei termini di una tale rappresentazione, il coefficiente di attendibilità misura la forza della relazione lineare tra X e T e corrisponde al coefficiente di determinazione del seguente modello di regressione:

$$X = 0 + 1 \cdot T + E.$$

1.7.1 Simulazione

Per dare un contenuto concreto alle affermazioni precedenti, consideriamo la seguente simulazione svolta in R. In tale simulazione il punteggio vero T e l'errore E verranno creati in modo tale da soddisfare i due vincoli della CTT: T e E saranno delle variabili gaussiane e tra loro incorrelate. Nella simulazione generiamo 100 coppie di valori X e T con i seguenti parametri: $T \sim \mathcal{N}(\mu_T = 12, \sigma_T^2 = 6)$, $E \sim \mathcal{N}(\mu_E = 0, \sigma_E^2 = 3)$. A tale fine usiamo le seguenti istruzioni:

```
set.seed(123)
library("MASS")
n <- 100
Sigma <- matrix(c(6, 0, 0, 3), byrow = TRUE, ncol = 2)
Sigma
#>      [,1] [,2]
#> [1,]    6    0
#> [2,]    0    3
```

```
mu <- c(12, 0)
mu
#> [1] 12  0
```



```
Y <- mvrnorm(n, mu, Sigma, empirical = TRUE)
T <- Y[, 1]
E <- Y[, 2]
```

Le istruzioni precedenti creano un insieme di valori tali per cui le medie e la matrice di varianze-covarianze assumono esattamente i valori indicati. Possiamo dunque immaginare tale insieme di dati come la nostra “popolazione”.

Secondo la CTT, il punteggio osservato è $X = T + E$. Simuliamo dunque il punteggio osservato X nel modo seguente:

```
X <- T + E
```

Le prime 6 osservazioni così ottenute sono:

```
head(cbind(T, E, X))
#>           T           E           X
#> [1,] 11.148 -1.5708  9.577
#> [2,] 13.138 -0.3335 12.804
#> [3,] 10.391  2.5457 12.937
#> [4,] 11.452 -0.1955 11.257
#> [5,]  9.978 -0.4920  9.486
#> [6,] 10.730  2.9609 13.691
```

Un diagramma di dispersione è fornito nella figura seguente:

```
tibble(X, T) %>%
  ggplot(aes(X, T)) +
  geom_point()
```

Secondo la CTT, il valore atteso di T è uguale al valore atteso di X . Verifichiamo questa assunzione della CTT nei nostri dati:

```
mean(T)
#> [1] 12
mean(X)
#> [1] 12
```

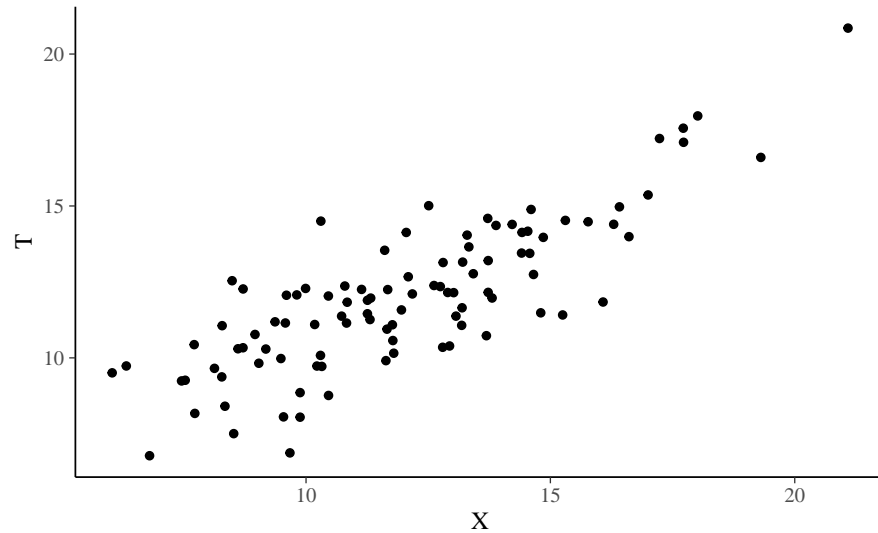


Figura 1.1: Simulazione della relazione tra punteggio osservato e punteggio vero per 100 individui in base alle assunzioni della CTT.

L'errore deve avere media zero, varianza σ_E^2 e deve essere incorrelato con T :

```
mean(E)
#> [1] 4.061e-18
var(E)
#> [1] 3
cor(T, E)
#> [1] -1.947e-16
```

Ricordiamo che la radice quadrata della varianza degli errori è chiamata errore standard della misurazione, σ_E . La quantità $\sqrt{\sigma_E^2}$ fornisce una misura della dispersione del punteggio osservato attorno al valore vero, nella condizione ipotetica di effettuare ripetute somministrazioni del test.

```
sqrt(3)
#> [1] 1.732
```

Dato che T e E sono incorrelati, ne segue che la varianza del punteggio osservato X è uguale alla somma della varianza del punteggio vero T e

della varianza degli errori E :

```
var(X)
#> [1] 9
var(T) + var(E)
#> [1] 9
```

La varianza del punteggio vero T è uguale alla covarianza tra il punteggio vero T e il punteggio osservato X :

```
cov(T, X)
#> [1] 6
var(T)
#> [1] 6
```

La correlazione tra il punteggio osservato e il punteggio vero è uguale al rapporto tra la deviazione standard del punteggio vero e la deviazione standard del punteggio osservato:

```
cor(X, T)
#> [1] 0.8165
sd(T) / sd(X)
#> [1] 0.8165
```

Focalizziamoci ora sull'attendibilità. Per la CTT, l'attendibilità è uguale al quadrato del coefficiente di correlazione tra il punteggio vero T e il punteggio osservato X :

```
cor(X, T)^2
#> [1] 0.6667
```

La motivazione di questa simulazione è quella di mettere in relazione il coefficiente di attendibilità, calcolato con le formule della CTT, con il modello di regressione lineare. Analizziamo dunque i dati della simulazione mediante il seguente modello di regressione lineare:

$$X = a + bT + E$$

```

fm <- lm(X ~ T)
summary(fm)
#>
#> Call:
#> lm(formula = X ~ T)
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -4.197 -1.101  0.052  1.155  4.239
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept) 8.53e-15   8.75e-01      0        1
#> T           1.00e+00   7.14e-02     14   <2e-16 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 1.74 on 98 degrees of freedom
#> Multiple R-squared:  0.667, Adjusted R-squared:  0.663
#> F-statistic: 196 on 1 and 98 DF, p-value: <2e-16

```

Si noti che la retta di regressione ha intercetta 0 e pendenza 1. Questo è coerente con l'assunzione $\mathbb{E}(X) = \mathbb{E}(T)$. Ma il risultato più importante di questa simulazione è il seguente: il coefficiente di determinazione ($R^2 = 0.67$) del modello di regressione $X = 0 + 1 \times T + E$ è identico al coefficiente di attendibilità che abbiamo calcolato con la formula $\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}$:

```

var(T) / var(X)
#> [1] 0.6667

```

Ciò ci consente di attribuire al coefficiente di attendibilità la seguente interpretazione: l'attendibilità di un test non è altro che la quota di varianza del punteggio osservato X che viene spiegata dalla regressione di X sul punteggio vero T in un modello di regressione dove $\alpha = 0$ e $\beta = 1$.

Che cosa si può concludere dai risultati di questa simulazione? Possiamo dire che, in base alla CTT,

- c'è una relazione lineare tra il punteggio osservato X e il punteggio vero T ; tale relazione lineare ha pendenza unitaria e intercetta zero.
- La CTT fa proprie le assunzioni del modello di regressione lineare: incorrelazione tra variabile esplicativa T ed errore E , e indipendenza e gaussianità degli errori.
- Come conseguenza di tali assunzioni, il coefficiente di attendibilità non è altro che la quota di varianza del punteggio osservato X che viene spiegata dal punteggio vero tramite una regressione lineare, ovvero non è altro che il coefficiente di determinazione del modello di regressione $X = \alpha + \beta T + E$, dove $\alpha = 0$ e $\beta = 1$.

Vedremo in seguito come sia possibile formulare la CTT nei termini del modello statistico dell'analisi fattoriale. Nel linguaggio dell'analisi fattoriale, la varianza dell'errore σ_E^2 viene chiamata *specificità* (*uniqueness*).

1.8 Misurazioni parallele e affidabilità

L'equazione $\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}$ definisce il coefficiente di attendibilità ma non fornisce gli strumenti per calcolarlo, dato che la varianza del punteggio vero σ_T^2 è una quantità incognita. Il metodo utilizzato dalla CTT per ottenere una stima (empirica) dell'attendibilità è quello delle forme parallele del test. Se è possibile elaborare versioni alternative dello stesso test che risultino equivalenti tra loro in termini di contenuto, modalità di risposta e caratteristiche statistiche, allora diventa anche possibile stimare il coefficiente di attendibilità.

Secondo la CTT, due test $X = T + E$ e $X' = T' + E'$ si dicono misurazioni parallele della stessa abilità latente se il punteggio vero T è uguale al punteggio vero T' e se la varianza degli errori $\mathbb{V}(E)$ è uguale alla varianza degli errori $\mathbb{V}(E')$.

Se il punteggio vero è uguale al valore atteso del punteggio osservato, $T = \mathbb{E}(X)$, allora devono essere uguali anche le medie dei punteggi osservati delle due forme parallele del test, $\mathbb{E}(X) = \mathbb{E}(X')$.

Dimostrazione. Consideriamo l'eguaglianza dei valori attesi dei punteggi osservati in due forme parallele del test: $\mathbb{E}(X) = \mathbb{E}(X')$. Risulta immediato che

$$\mathbb{E}(X) = \mathbb{E}(T + E) = \mathbb{E}(T) + \mathbb{E}(E) = T,$$

dato che $\mathbb{E}(E) = 0$ e T non è una variabile aleatoria. Inoltre, $\mathbb{E}(X') = T$, dato che $T = T'$. Ne segue che $\mathbb{E}(X) = \mathbb{E}(X')$. \square

In maniera corrispondente, anche le varianze dei punteggi osservati di due misurazioni parallele devono essere uguali, $\mathbb{V}(X) = \mathbb{V}(X')$.

Dimostrazione. Per la misurazione parallela X abbiamo

$$\mathbb{V}(X) = \mathbb{V}(T + E) = \mathbb{V}(T) + \mathbb{V}(E);$$

per la misurazione parallela X' abbiamo

$$\mathbb{V}(X') = \mathbb{V}(T' + E') = \mathbb{V}(T') + \mathbb{V}(E').$$

Dato che $\mathbb{V}(E) = \mathbb{V}(E')$ e che $T = T'$, ne segue che $\mathbb{V}(X) = \mathbb{V}(X')$. \square

Per costruzione, inoltre, gli errori E e E' devono essere incorrelati con T e tra loro.

1.8.1 La correlazione tra misurazioni parallele

Un'ulteriore assunzione della CTT è la seguente. La CTT assume che, data una serie di misurazioni parallele X_1, X_2, X_3, \dots e un arbitrario test Z , si ha

$$\rho(X_1, X_2) = \rho(X_1, X_3) = \rho(X_2, X_3) = \dots$$

e

$$\rho(X_1, Z) = \rho(X_2, Z) = \rho(X_3, Z) = \dots$$

ovvero, tutte le misurazioni parallele sono correlate tra loro nella stessa misura e ciascuna misurazione parallela correla nella stessa misura con qualunque altro test.

L'assunzione precedente può essere espressa, in maniera equivalente, come segue. Si consideri la matrice di correlazioni calcolata su tutto il dominio degli item (ovvero, la matrice delle correlazioni tra ciascuna

coppia di item nel dominio del costrutto). La correlazione media di questa matrice quantifica la capacità media di ciascun item di rappresentare il costrutto. La CTT assume che la correlazione di ciascun item con ciascuno degli altri sia costante (ovvero, uguale per qualunque coppia di item). Detto in altri termini: secondo la CTT ciascun item rappresenta il costrutto nella stessa misura. Questa è un'assunzione molto forte che si riflette, come vedremo, nella formula del coefficiente α di Cronbach utilizzata per misurare l'attendibilità come consistenza interna. È un'assunzione molto forte che raramente viene soddisfatta in pratica.

Secondo la CTT, dunque, forme parallele del test devono avere lo stesso valore atteso e la stessa varianza. Inoltre, ciascuna forma parallela deve correlare nella stessa misura con qualunque altro test. In che modo si differenziano allora le forme parallele del test? L'unica differenza tra le forme parallele del test riguarda il punteggio osservato: a causa dell'errore di misurazione $X \neq X'$.

Il concetto di forme parallele del test è estremamente importante per la CTT perché attraverso tale nozione diventa possibile giungere ad una stima empirica dell'attendibilità. Prima di presentare questo ultimo passaggio algebrico è però necessario calcolare la correlazione tra due misurazioni parallele.

1.8.2 La correlazione tra due forme parallele del test

Secondo la CTT, la correlazione tra due misurazioni parallele è uguale al rapporto tra la varianza del punteggio vero e la varianza del punteggio osservato. Ricordiamo che la varianza del punteggio osservato è uguale nelle due forme parallele del test: $\mathbb{V}(X) = \mathbb{V}(X')$.

Dimostrazione. Assumendo, senza perdita di generalità, che $\mathbb{E}(X) = \mathbb{E}(X') = \mathbb{E}(T) = 0$, possiamo scrivere

$$\begin{aligned} \rho_{XX'} &= \frac{\sigma(X, X')}{\sigma(X)\sigma(X')} \\ &= \frac{\mathbb{E}(XX')}{\sigma(X)\sigma(X')} \\ &= \frac{\mathbb{E}[(T + E)(T + E')]}{\sigma(X)\sigma(X')} \\ &= \frac{\mathbb{E}(T^2) + \mathbb{E}(TE') + \mathbb{E}(TE) + \mathbb{E}(EE')}{\sigma(X)\sigma(X')}. \end{aligned}$$

Ma $\mathbb{E}(TE) = \mathbb{E}(TE') = \mathbb{E}(EE') = 0$; inoltre, $\sigma(X) = \sigma(X') = \sigma_X$. Dunque,

$$\rho_{XX'} = \frac{\mathbb{E}(T^2)}{\sigma_X \sigma_X} = \frac{\sigma_T^2}{\sigma_X^2}. \quad (1.3)$$

□

Si noti come la (1.3) e l'equazione che definisce il coefficiente di attendibilità, ovvero $\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}$, riportano tutte e due la stessa quantità a destra dell'uguale. Otteniamo così un importante risultato. Il coefficiente di attendibilità, ovvero il quadrato del coefficiente di correlazione tra il punteggio osservato e il punteggio vero, è uguale alla correlazione tra il valore osservato di due misurazioni parallele:

$$\rho_{XT}^2 = \rho_{XX'}. \quad (1.4)$$

Tale risultato è importante perché consente di esprimere la quantità inosservabile ρ_{XT}^2 nei termini della quantità $\rho_{XX'}$ che può essere calcolata sulla base del punteggio osservato. Quindi, la stima di ρ_{XT}^2 si riduce alla stima di $\rho_{XX'}$. Per questa ragione, la (1.4) è forse la formula più importante della CTT.

1.8.3 La correlazione tra punteggio osservato e punteggio vero

Consideriamo ora la correlazione tra punteggio osservato e punteggio vero. La (1.4) si può scrivere come

$$\rho_{XT} = \sqrt{\rho_{XX'}}.$$

Tale risultato si può interpretare dicendo che la correlazione tra punteggio osservato e punteggio vero è uguale alla radice quadrata del coefficiente di attendibilità.

1.8.4 I fattori che influenzano l'attendibilità

Considerando le precedenti tre equazioni

$$\rho_{XT}^2 = \rho_{XX'}, \quad \rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}, \quad \rho_{XT}^2 = 1 - \frac{\sigma_E^2}{\sigma_X^2},$$

possiamo dire che ci sono tre modi equivalenti per concludere che l'attendibilità di un test è alta. L'attendibilità di un test è alta

1. quando la correlazione tra misurazioni parallele è alta,
2. quando la varianza del punteggio vero è grande relativamente alla varianza del punteggio osservato,
3. quando la varianza dell'errore di misura è piccola relativamente alla varianza del punteggio osservato.

Tali considerazioni hanno importanti implicazioni per le scelte che devono guidare la costruzione di un test. Si consideri, in particolare, l'equazione $\rho_{XT}^2 = \rho_{XX'}$. Se interpretiamo $\rho_{XX'}$ come la correlazione tra due item, allora tale equazione ci fornisce un criterio per la scelta degli item da includere in un test: dobbiamo includere nel test gli item che correlano maggiormente tra loro. In questo modo, infatti, l'attendibilità del test aumenterà perché gli item inclusi nel test risultano maggiormente correlati con il punteggio vero.

1.9 Metodi alternativi per la stima del coefficiente di attendibilità

Come si stima in pratica l'affidabilità? Un modo grossolano (e molto impreciso) consiste nel somministrare allo stesso gruppo di individui lo stesso test in due differenti momenti e di calcolare il coefficiente di correlazione dei punteggi totali (*test-retest reliability*). McDonald (2013) afferma che tale procedura può essere giustificata in due modi diversi. La prima giustificazione è basata sull'assunzione che il valore vero non varia tra le due somministrazioni del test. Se le cose stanno in questo modo, gli errori saranno indipendenti e la correlazione tra il punteggio osservato nelle due somministrazioni ci fornirà una stima di $\rho_{XX'}$. Il problema è che non disponiamo di nessuno strumento per distinguere questa situazione ideale dal caso in cui viene violata l'assunzione dell'invarianza del punteggio vero. Una seconda giustificazione del metodo test-retest ci porta a definire il punteggio vero di retest come la componente del punteggio osservato che non varia tra le due somministrazioni. Il tal senso, il coefficiente di attendibilità viene concepito come un coefficiente di stabilità temporale. In generale, maggiore è l'intervallo temporale tra le

due somministrazioni, minore sarà il valore del coefficiente di stabilità temporale. Uno dei problemi del metodo test-retest è che due somministrazioni successive di un test ci forniscono soltanto un sottoinsieme delle possibili informazioni che verrebbero raccolte da uno studio longitudinale che copre un periodo temporale maggiore. Se tale studio longitudinale venisse eseguito, potremmo trovare la funzione che descrive la variazione del punteggio osservato in funzione del tempo. In generale, tale funzione non può essere descritta da un singolo parametro. Resta aperta la domanda di quale sia relazione tra questa funzione e il coefficiente di attendibilità.

Se sono disponibili due forme parallele dello stesso test, l'affidabilità può essere calcolata mediante il coefficiente di correlazione dei punteggi totali dei due test (*parallel-forms reliability*), valendo l'uguaglianza $\rho_{XX'} = \rho_{XT}^2$. Anche questo metodo, come il metodo del test-retest, non è esente da errori.

Il metodo di stima più diffuso è quello conosciuto come Cronbach's alpha (*internal consistency reliability*) originariamente ricavato da [Kuder and Richardson \(1937\)](#) per item dicotomici e poi generalizzato da [Cronbach \(1951\)](#) per item a risposte ordinali di qualunque tipo. L'idea su cui si basa consiste nel fatto che ogni singolo item del test, se confrontato con tutti gli altri, può essere usato per stimarne l'affidabilità. L'analisi degli item viene utilizzata per ottenere una stima della consistenza interna del test e valuta la misura in cui gli item del test sono espressione dello stesso costrutto.

2

L'incertezza della misura

Lord e Novick (1968) fanno notare come l'errore $E = X - T$ sia la variabile aleatoria di primario interesse per la CTT, in quanto lo scopo è stimare il punteggio vero di ciascun rispondente e confrontare le stime ottenute nel caso di rispondenti diversi. La grandezza dell'errore che si commette utilizzando il punteggio osservato quale misura del punteggio vero può essere quantificata mediante la deviazione standard di E , ovvero mediante ciò che viene chiamato *l'errore standard della misurazione*, σ_E (*Standard Error of Measurement, SEM*). Ma come è possibile stimare σ_E ?

2.1 La stima dell'errore standard della misurazione

L'errore standard della misurazione quantifica il grado di incertezza presente nei punteggi di un test. Può essere dimostrato che una stima dell'errore standard della misurazione (σ_E) è data da:

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}},$$

dove σ_X è la deviazione standard dei punteggi ottenuti in un campione di rispondenti e $\rho_{XX'}$ è il coefficiente di attendibilità. Per stimare σ_E è dunque necessario sottrarre da uno l'attendibilità del test, prendere la radice quadrata della differenza e moltiplicare la radice quadrata per la deviazione standard dei punteggi del test.

Si noti che l'errore standard della misurazione σ_E è direttamente associato all'attendibilità del test: l'errore standard della misurazione diminuisce al crescere dell'attendibilità del test. Se l'attendibilità del test è uguale a 0 σ_E diventa uguale alla deviazione standard del punteggio osservato del test. Se l'attendibilità del test è uguale a 1 σ_E diventa

uguale a zero: se il test è perfettamente affidabile non ci sono errori e σ_E è uguale a zero.

2.1.1 Interpretazione

McDonald afferma che il termine E segue una *propensity distribution*, ovvero rappresenta le fluttuazioni casuali nel tempo di un rispondente, che corrispondono a fluttuazioni di umore, motivazione, ecc. L'errore standard della misura fornisce una stima della deviazione standard di tali punteggi, ovvero una stima della deviazione standard dei punteggi che un singolo individuo otterrebbe nel caso di ipotetiche infinite somministrazioni di un test (o di forme parallele di un test) sotto le stesse identiche condizioni, se il punteggio vero rimane costante.

La CTT assume i punteggi ottenuti da un individuo, nel caso di ipotetiche infinite somministrazioni di un test nelle stesse identiche condizioni, abbiano una distribuzione normale centrata sul valore vero. L'errore standard della misurazione è la stima della deviazione standard di una tale distribuzione di punteggi ipotetici. Maggiore è l'errore standard della misurazione, maggiore è l'errore che si compie usando il test per valutare l'abilità latente del rispondente.

Il coefficiente di attendibilità, la varianza dell'errore e l'errore standard della misurazione sono tutti indicatori diretti o indiretti della precisione del test. Tuttavia, questi indici forniscono informazioni diverse sul grado di precisione del test:

- l'errore standard della misurazione ci consente di fare inferenze sulla precisione del punteggio osservato di un singolo rispondente, ma non è possibile assegnare tale interpretazione al coefficiente di attendibilità;
- l'errore standard della misurazione è espresso nella stessa unità di misura del punteggio osservato, mentre la varianza di E è espressa nei termini del quadrato del punteggio osservato;
- l'attendibilità corrisponde ad un rapporto tra varianze e dunque è un numero puro (privo di unità di misura).

Supponiamo che un test di intelligenza produca un punteggio medio pari a 100 con una deviazione standard di 15. Supponiamo inoltre che il test abbia una attendibilità pari a 0.73. Si calcoli l'errore standard della misurazione.

Applicando la formula dell'errore standard della misurazione, otteniamo

$$\begin{aligned}
\sigma_E &= \sigma_X \sqrt{1 - \rho_{XX'}} \\
&= 15 \sqrt{1 - 0.73} \\
&= 7.79.
\end{aligned}$$

Il valore di 7.79 significa che, se immaginiamo di somministrare molte volte il test ad un rispondente, sotto le stesse identiche condizioni, ci aspettiamo che i valori ottenuti differiscano tra loro, in media, di circa 8 punti tra le successive somministrazioni del test. Inoltre, se immaginiamo di somministrare molte volte il test ad un rispondente, sotto le stesse identiche condizioni, ci aspettiamo che il 95% dei punteggi così ottenuti sia compreso nell'intervallo

punteggio vero del rispondente $\pm 1.96 \cdot \text{errore standard della misurazione}$.

Questa è una proprietà della distribuzione gaussiana.

Per il caso presente, questo intervallo è uguale a $2 \cdot 1.96 \cdot 7.79 = 30.54$ punti. In altre parole, ci possiamo aspettare che, nel caso di somministrazioni ripetute del test sotto le stesse identiche condizioni, i punteggi del QI di un singolo rispondente varino tra loro all'interno di un intervallo di 30 punti. Ciò significa che, se il test avesse un'attendibilità pari a 0.73, e se la deviazione standard dei punteggi del test nella popolazione fosse pari a 15, la somministrazione di un tale test ad un singolo individuo sarebbe di scarsa utilità, a causa dell'enorme errore di misurazione. Per fare un confronto con i dati di questo esempio, la Full Scale IQ (FSIQ) della WAIS-IV ([Wechsler, 2008](#)) ha un'attendibilità split-half pari a 0.98, con errore standard di misurazione pari a 2.16.

Continuando con l'esempio precedente, per gli ipotetici dati riportati sopra, poniamoci ora la seguente domanda: qual è la probabilità che un rispondente ottenga un punteggio minore o uguale a 116 nel test, se il suo punteggio vero è uguale a 120?

Il problema si risolve rendendosi conto che i punteggi del rispondente si distribuiscono normalmente attorno al punteggio vero di 120, con una deviazione standard uguale a 7.79. Dobbiamo dunque trovare l'area sottesa alla normale $\mathcal{N}(120, 7.79)$ nell'intervallo $[-\infty, 116]$. Utilizzando , la soluzione si trova nel modo seguente:

```
pnorm(116, 120, 7.79)
#> [1] 0.3038
```

Se la variabile aleatorie corrispondente al punteggio osservato segue una distribuzione $\mathcal{N}(120, 7.79)$, la probabilità che il rispondente ottenga un punteggio minore o uguale a 116 è dunque uguale a 0.30.

Poniamoci ora la seguente domanda: quale intervallo di valori centrato sul punteggio vero contiene, con una probabilità di 0.95, i punteggi che il rispondente otterrebbe in ipotetiche somministrazioni ripetute del test sotto le stesse identiche condizioni?

Dobbiamo trovare i quantili della distribuzione $\mathcal{N}(120, 7.79)$ a cui sono associate le probabilità di 0.025 e 0.975. La soluzione è dunque data da:

```
qnorm(c(.025, .975), 120, 7.79)
#> [1] 104.7 135.3
```

L'intervallo cercato è dunque $[104.7, 135.3]$.

2.1.2 Simulazione

Ritorniamo ora alla simulazione precedente nella quale abbiamo messo in relazione il modello della CTT con il modello di regressione lineare. In base a tale simulazione, poniamoci lo scopo di chiarire il significato dell'errore standard della misurazione.

Impostiamo la simulazione come abbiamo fatto in precedenza. Chiamiamo X il valore osservato in un test. Per la CTT, il punteggio osservato X è costituito da due componenti, la componente vera T e la componente d'errore E . Si suppone che gli errori siano gaussiani e incorrelati con la componente vera. Immaginiamo di somministrare 200 volte il test ad un individuo sotto le stesse identiche condizioni.

```
library("MASS")
library("arm")
set.seed(123)
n <- 200
Sigma <- matrix(
```

```
c(
  11, 0,
  0, 4
),
byrow = TRUE, ncol = 2
)
mu <- c(100, 0)
Y <- mvrnorm(n, mu, Sigma, empirical = TRUE)
T <- Y[, 1]
E <- Y[, 2]
```

Verifichiamo l'incorrelazione tra T ed E :

```
cor(T, E)
#> [1] -1.069e-16
```

I valori ottenuti sono la somma del valore vero e della componente d'errore:

```
X <- T + E
```

Per questi dati, il coefficiente di attendibilità è uguale a:

```
rx <- cor(X, T)^2
rx
#> [1] 0.7333
```

Possiamo ora calcolare l'errore standard della misurazione utilizzando la (2.1):

```
sd(X) * sqrt(1 - rx)
#> [1] 2
```

Si noti che tale valore non è altro che la deviazione standard degli errori della misurazione:

```
sd(E)
#> [1] 2
```

Ovvero, nei termini del modello di regressione $X = 0 + 1 \cdot T + E$, l'errore standard della misurazione corrisponde all'errore standard della regressione:

```
fm <- lm(formula = X ~ T)
summary(fm)
#>
#> Call:
#> lm(formula = X ~ T)
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -4.432 -1.404 -0.147  1.219  7.212
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept) -1.93e-13   4.29e+00    0.0      1
#> T           1.00e+00   4.29e-02   23.3 <2e-16 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 2.01 on 198 degrees of freedom
#> Multiple R-squared:  0.733, Adjusted R-squared:  0.732
#> F-statistic: 545 on 1 and 198 DF, p-value: <2e-16
```

Si noti che, nell'output di R fornito sopra, l'errore standard della regressione, ovvero `residual sd`, corrisponde a 2.01 anziché a 2.0. Ciò si verifica in quanto R ha calcolato una stima della deviazione standard dei residui nella popolazione utilizzando, al denominatore, $n - 2$. Nel nostro caso è invece necessario dividere per n in quanto i dati della simulazione sono quelli della popolazione, non di un campione.

2.2 Dimostrazione

Poniamoci ora il problema di derivare la formula dell'errore standard della misurazione. Per derivare la formula $\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}}$ sono necessari due passi: prima dobbiamo trovare la varianza del punteggio vero; poi dobbiamo esprimere il punteggio osservato come la somma della varianza del punteggio vero e la varianza dell'errore.

Dimostrazione. In base alla definizione del coefficiente di attendibilità $\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2}$ possiamo scrivere $\sigma_T^2 = \rho_{XX'} \sigma_X^2$, dove X e X' sono due forme parallele di un test. Ricordiamo che misurazioni parallele hanno le seguenti proprietà: $\mathbb{E}(X) = \mathbb{E}(X')$ e $\mathbb{V}(X) = \mathbb{V}(X')$. Dato che $\sigma_X = \sigma_{X'}$, l'equazione precedente diventa $\sigma_T^2 = \rho_{XX'} \sigma_X \sigma_{X'}$. Utilizzando la definizione della covarianza tra X e X' , ovvero, $\sigma_{XX'} = \rho_{XX'} \sigma_X \sigma_{X'}$, possiamo concludere che la varianza del punteggio vero è uguale alla covarianza tra due misurazioni parallele:

$$\sigma_T^2 = \sigma_{XX'}.$$

Essendo l'attendibilità del test il rapporto tra la varianza del punteggio vero e la varianza del punteggio osservato, ed essendo che la varianza del punteggio vero uguale alla covarianza tra due misurazioni parallele, possiamo concludere che l'attendibilità aumenta all'aumentare della covarianza media tra gli item del test. Si noti come questo importante risultato della CTT dipenda dall'ipotesi di omogeneità delle varianze degli item del test.

Calcoliamo ora la varianza di E . La varianza del punteggio osservato è uguale a $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$. Sulla base della definizione di attendibilità $\sigma_T^2 = \rho_{XX'} \sigma_X^2$, la varianza del punteggio osservato si può scrivere come $\sigma_X^2 = \rho_{XX'} \sigma_X^2 + \sigma_E^2$, da cui

$$\begin{aligned} \sigma_E^2 &= \sigma_X^2 - \sigma_X^2 \rho_{XX'} \\ &= \sigma_X^2 (1 - \rho_{XX'}). \end{aligned}$$

La varianza degli errori della misurazione $\sigma_E^2 = \sigma_X^2 (1 - \rho_{XX'})$ è dunque uguale al prodotto di due fattori: il primo fattore è la varianza del punteggio osservato; il secondo fattore è uguale a uno meno la correlazione

tra due forme parallele del test. Possiamo così calcolare una quantità incognita, σ_E^2 , nei termini di due quantità osservabili, σ_X^2 e $\rho_{XX'}$. \square

2.3 Intervallo di confidenza per il punteggio vero e σ_E

Uno degli usi che vengono fatti dell'errore standard della misurazione è quello di costruire, con essi, gli intervalli di confidenza per il punteggio vero. Tale uso, però, non è corretto (Charter, 1996). Gli intervalli di confidenza costruiti usando l'errore standard della misurazione vengono talvolta incorrettamente interpretati in modo tale da suggerire che l'intervallo di confidenza al $(1 - \alpha)\%$ identifica una gamma di valori, *centrata sul valore osservato*, entro il quale cadono i punteggi veri del test nel $(1 - \alpha)\%$ di ipotetiche somministrazioni ripetute del test. Ma le cose non stanno così. In realtà, come abbiamo detto sopra, l'errore standard della misurazione è la deviazione standard, calcolata rispetto al valore vero, di ipotetiche misurazioni ripetute dello stesso test. Si può ribadire questo concetto nel modo seguente: “In spite of Dudek (1979)’s reminder that the SEM should not be used to construct confidence intervals, many test manuals, computer-scoring programs, and texts in psychology and education continue to do so. Because authors of many textbooks and manuals make these errors, it is understandable that those who learned from and look to these sources for guidance also make these errors. In summary, the SEM should not be used to construct confidence intervals for test scores” (p. 1141). Sembra piuttosto chiaro.

3

La stima del punteggio vero

Uno degli scopi principali della valutazione psicologica è quello di stimare il punteggio vero del rispondente. Il punteggio osservato X differisce dal punteggio vero T a causa della presenza dell'errore della misurazione: $X = T + E$. Poniamoci ora il problema di utilizzare i concetti della Teoria Classica per stimare il punteggio vero di un rispondente utilizzando il suo punteggio osservato e l'attendibilità del test. Questa stima è utile soprattutto quando è necessario costruire un intervallo di confidenza per il punteggio vero.

Per costruire l'intervallo di confidenza del punteggio vero dobbiamo utilizzare due quantità:

- una stima del punteggio vero,
- l'errore standard della stima (ovvero, una stima della deviazione standard della distribuzione delle stime del punteggio vero che si otterrebbe se il test venisse somministrato infinite volte sotto le stesse condizioni).

Iniziamo con il problema della stima del punteggio vero.

3.1 Il paradosso di Kelley

Nella sua monografia del 1954, "Clinical versus statistical prediction: A theoretical analysis and a review of the evidence", Paul Meehl suscitò un grande scalpore con una convincente dimostrazione del fatto che metodi meccanici di combinazione dei dati, come ad esempio la regressione multipla, sono in grado di fornire delle predizioni migliori di quanto sia in grado di fare la diagnosi clinica eseguita da esperti. L'enorme quantità di letteratura che è stata prodotta in seguito a tale contributo ha fornito forti e univoche evidenze a sostegno di questa osservazione.

È interessante notare che Robyn Dawes (2005) ha pubblicato un articolo su *Journal of Clinical Psychology* (61, 1245–1255) dal titolo seguente: “The ethical implications of Paul Meehl’s Work on comparing clinical versus actuarial prediction methods”. L’argomento principale sostenuto da Dawes è che, date le evidenze molto convincenti che sono disponibili, non è etico usare il giudizio clinico in preferenza all’uso di modelli statistici di previsione. Citiamo dall’abstract:

Whenever statistical prediction rules [...] are available for making a relevant prediction, they should be used in preference to intuition. [...] Providing service that assumes that clinicians “can do better” simply based on self-confidence or plausibility in the absence of evidence that they can actually do so is simply unethical.

Sulla base di quanto detto sopra, e in riferimento ai nostri scopi presenti, si pone dunque il problema di capire come sia possibile utilizzare il modello di regressione per ottenere una stima del punteggio vero di un rispondente. A questo proposito si deve notare che è necessario tenere in considerazione il fatto che le nostre variabili indipendenti sono corrotte dall’errore di misurazione, mentre il modello di regressione tradizionale presuppone che le variabili indipendenti siano misurate senza errori. Le considerazioni seguenti sono state proposte da Kelly negli anni ’20.

Come dimostrato in seguito, la formula di Kelley si basa sull’equivalenza algebrica secondo la quale l’attendibilità è uguale al quadrato del coefficiente di correlazione tra i punteggi osservati e i punteggi veri. In base alla formula di Kelley, il punteggio vero di un rispondente può essere stimato nel modo seguente mediante il modello di regressione:

$$\hat{T} = \mu_x + \rho(X - \mu_x),$$

laddove X è il punteggio osservato, μ_x è la media dei punteggi ottenuti da tutti i rispondenti di un campione e ρ è l’attendibilità del test.

Quando l’attendibilità è perfetta ($\rho = 1$), il punteggio vero è uguale al punteggio osservato. Quando l’attendibilità è zero (tutta la varianza è dovuta all’errore della misurazione), allora la stima migliore del punteggio vero è data dalla media del campione. Quando $0 < \rho < 1$, la stima del punteggio vero corrisponde ad un valore che si discosta dal punteggio osservato nella direzione della media del campione. La stima del punteggio vero, dunque, esibisce la proprietà della regressione verso la media del punteggio osservato, in funzione dell’attendibilità del test[¹].

La (3.1) può essere interpretata dicendo che, per stimare il punteggio vero di un rispondente, partiamo dalla media della distribuzione della popolazione dei rispondenti e ci spostiamo nella direzione del punteggio osservato di un rispondente. Tuttavia, non raggiungiamo il valore del punteggio osservato: la quantità di cui ci spostiamo è proporzionale all'attendibilità. In altre parole, a seconda della dimensione di ρ , la stima del punteggio vero di un individuo è dovuta, in parte, a dove si trova l'individuo in relazione al gruppo di appartenenza: la stima del punteggio vero dell'individuo si sposterà verso l'alto se l'individuo è collocato sotto la media del gruppo di appartenenza e si sposterà verso il basso se l'individuo è collocato al di sopra della media del gruppo di appartenenza. Questa equazione è stata chiamata il *paradosso di Kelley*.

È importante sottolineare che l'interpretazione precedente rivela che la formula di Kelley contraddice la nozione intuitiva secondo cui il punteggio osservato può essere utilizzato quale stima del punteggio vero (cioè, $\hat{T} = X$). Tale ragionamento ingenuo sarebbe corretto se l'attendibilità del test fosse perfetta ($\rho = 1$). All'altra estremità dello spettro, quando $\rho = 0$, la formula di Kelley ci suggerisce μ_x quale stima del punteggio vero, il che è equivale a dire che il punteggio osservato deve essere ignorato – infatti se la varianza di X è solamente dovuta all'errore di misurazione, allora il test è del tutto inutile quale strumento inferenziale per differenziare le abilità dei rispondenti. Fortunatamente, in pratica è molto improbabile che $\rho = 0$. Se ρ cade tra gli estremi di 0 e 1, allora il punteggio vero stimato sarà compreso tra il punteggio osservato e μ_x . Per capire cosa esso catturi, possiamo citare Kelley(1947), che osservò:

This is an interesting equation in that it expresses the estimate of true ability as the weighted sum of two separate estimates, – one based upon the individual's observed score, X_1 (X nella notazione corrente) and the other based upon the mean of the group to which he belongs, M_1 (μ_x nella notazione corrente). If the test is highly reliable, much weight is given to the test score and little to the group mean, and vice versa.

Dimostrazione. Come si arriva all'equazione di Kelley? Abbiamo visto in precedenza come l'equazione che mette in relazione il punteggio osservato con il punteggio vero non è altro che un modello di regressione con intercetta nulla e pendenza unitaria: $X = 0 + 1 \cdot T + E$. In questo caso, però, il problema è diverso, in quanto noi vogliamo *predire* il punteggio vero sulla base del punteggio osservato per mezzo di un modello di regressione (Nunnally, 1978). Avendo quale scopo quello di “predire”

il punteggio vero T sulla base del punteggio osservato X , il modello di regressione diventa

$$T = \alpha + \beta X + \varepsilon.$$

Se esprimiamo le variabili come deviazioni dalla media, $x = X - \bar{X}$ e $\tau = T - \mathbb{E}(T)$, allora l'intercetta diventa uguale a zero e il modello diventa $\tau = \beta x + \varepsilon$, ovvero $\hat{\tau} = \beta x$. Il problema è quello di calcolare il coefficiente β .

Nel modello $\hat{\tau} = \beta x$, la pendenza della retta di regressione è uguale a $\beta = \frac{\sigma_{\tau x}}{\sigma_x^2}$. Possiamo dunque scrivere il modello di regressione nel modo seguente:

$$\hat{\tau} = \frac{\sigma_{\tau x}}{\sigma_x^2} x. \quad (3.1)$$

La correlazione tra x (o X) e τ (o T) è uguale a $\rho_{\tau x} = \frac{\sigma_{\tau x}}{\sigma_x \sigma_\tau}$. Dunque $\sigma_{\tau x} = \rho_{\tau x} \sigma_x \sigma_\tau$ e l'equazione precedente diventa

$$\begin{aligned} \hat{\tau} &= \frac{\rho_{\tau x} \sigma_x \sigma_\tau}{\sigma_x^2} x \\ &= \rho_{\tau x} \frac{\sigma_\tau}{\sigma_x} x. \end{aligned}$$

In base alla definizione di attendibilità, la varianza del punteggio vero è $\sigma_\tau^2 = \sigma_x^2 \rho_{xx'}$. Dunque, la deviazione standard del punteggio vero diventa $\sigma_\tau = \sigma_x \sqrt{\rho_{xx'}}$. Sostituendo questo risultato nell'equazione precedente otteniamo

$$\begin{aligned} \hat{\tau} &= \rho_{\tau x} \frac{\sigma_x \sqrt{\rho_{xx'}}}{\sigma_x} x \\ &= \rho_{\tau x} \sqrt{\rho_{xx'}} x. \end{aligned}$$

In precedenza abbiamo visto che $\rho_{\tau x}^2 = \rho_{xx'}$, dunque

$$\begin{aligned} \hat{\tau} &= \rho_{\tau x} \sqrt{\rho_{xx'}} x \\ &= \sqrt{\rho_{xx'}} \sqrt{\rho_{xx'}} x \\ &= \rho_{xx'} x. \end{aligned}$$

In conclusione, una stima del punteggio vero si ottiene moltiplicando il punteggio osservato, espresso come deviazione dalla media, per il coefficiente di attendibilità.

Riscriviamo ora la formula appena ottenuta nei termini del punteggio grezzo X (non in termini di deviazioni dalla media. Per fare ciò, sommiamo \bar{X} così da ottenere

$$\hat{T} = \rho_{XX'}(X - \bar{X}) + \bar{X},$$

laddove \hat{T} è la stima del punteggio vero grezzo. Sviluppando otteniamo

$$\begin{aligned}\hat{T} &= \rho_{XX'}(X - \bar{X}) + \bar{X} \\ &= X\rho_{XX'} - \bar{X}\rho_{XX'} + \bar{X} \\ &= \bar{X}(1 - \rho_{XX'}) + X\rho_{XX'} \\ &= \bar{X} - \bar{X}\rho_{XX'} + X\rho_{XX'} \\ &= \bar{X} + \rho_{XX'}(X - \bar{X}).\end{aligned}$$

Per i dati campionari, la formula diventa:

$$\hat{T} = \bar{X} + r_{XX'}(X - \bar{X}),$$

dove X è il punteggio (grezzo) osservato, \bar{X} è la media dei punteggi osservati di un campione di rispondenti e $r_{XX'}$ è il coefficiente di attendibilità. \square

Esercizio 3.1. Posto un coefficiente di attendibilità pari a 0.80 e una media del test pari a $\bar{X} = 100$, si trovi una stima del punteggio vero per un rispondente con un punteggio osservato uguale a $X = 115$.

La stima del punteggio vero \hat{T} è uguale a

$$\begin{aligned}\hat{T} &= \bar{X} + r_{XX'}(X - \bar{X}) \\ &= 100 + 0.80 \cdot (115 - 100) = 112.\end{aligned}$$

3.2 L'errore standard della stima

Oltre a ottenere una stima del punteggio vero da un punteggio osservato, il modello di regressione di Kelley ci fornisce anche l'errore standard della stima. È chiaro che la stima del punteggio vero è difficile da interpretare se non è accompagnata da una qualche indicazione sulla precisione della stima. Tale informazione viene appunto fornita dall'*errore standard della stima*.

Se il test potesse essere somministrato ad un rispondente più volte sotto le identiche condizioni, sarebbe possibile ottenere in ciascuna somministrazione una stima del valore vero \hat{T} . A causa dell'errore della misurazione, il punteggio osservato non può che variare in ciascuna ipotetica somministrazioni del test e, di conseguenza, in ciascuna ipotetica somministrazione varierà anche la stima di \hat{T} . La deviazione standard di tali (ipotetiche) stime di \hat{T} è chiamata *errore standard della stima*. L'errore standard della stima, $\sigma_{\hat{T}}$, si calcola con la formula seguente:

$$\sigma_{\hat{T}} = \sigma_X \sqrt{\rho_{XX'}(1 - \rho_{XX'})}. \quad (3.2)$$

Dimostrazione. Per ricavare la si definisce ε l'errore che si commette quando si stima il punteggio vero \hat{T} con il punteggio osservato T (si veda Lord e Novick, 1968):

$$\varepsilon = T - \hat{T}.$$

Si presti attenzione alla notazione: $E = X - T$ indica l'errore della misurazione, ovvero la differenza tra il punteggio osservato e il punteggio vero. Invece $\varepsilon = T - \hat{T}$ indica la differenza tra il punteggio vero e la stima del punteggio vero. Avendo che $\hat{T} = \bar{X} + \rho_{XX'}(X - \bar{X})$, la varianza di $\varepsilon = T - \hat{T}$ si può scrivere come

$$\begin{aligned} \mathbb{V}(\varepsilon) &= \mathbb{V}(T - \hat{T}) \\ &= \mathbb{V}(T - \bar{X} - \rho_{XX'}X + \rho_{XX'}\bar{X}). \end{aligned}$$

Dato che la varianza di una variabile aleatoria non cambia sommando a tale variabile una costante, dobbiamo semplicemente calcolare

$$\mathbb{V}(\varepsilon) = \mathbb{V}(T - \rho_{XX'}X).$$

Dobbiamo trovare la varianza della somma di due variabili aleatorie, una delle quali moltiplicata per una costante. Dunque:

$$\mathbb{V}(\varepsilon) = \mathbb{V}(T) + \rho_{XX'}^2 \mathbb{V}(X) - 2\rho_{XX'} \text{Cov}(X, T),$$

ovvero, semplificando la notazione,

$$\sigma_\varepsilon^2 = \sigma_T^2 + \rho_{XX'}^2 \sigma_X^2 - 2\rho_{XX'} \sigma_{XT}.$$

La quantità $\rho_{XX'}$ è il coefficiente di attendibilità. Quindi

$$\sigma_\varepsilon^2 = \sigma_T^2 + \left(\frac{\sigma_T^2}{\sigma_X^2} \right)^2 \sigma_X^2 - 2 \frac{\sigma_T^2}{\sigma_X^2} \sigma_{XT}.$$

Semplificando otteniamo

$$\begin{aligned} \sigma_\varepsilon^2 &= \sigma_T^2 + \frac{\sigma_T^4}{\sigma_X^4} \sigma_X^2 - 2 \frac{\sigma_T^2}{\sigma_X^2} \sigma_{XT} \\ &= \sigma_T^2 + \sigma_T^2 \frac{\sigma_T^2}{\sigma_X^2} - \sigma_T^2 2 \frac{\sigma_{XT}}{\sigma_X^2} \\ &= \sigma_T^2 \left(1 + \frac{\sigma_T^2}{\sigma_X^2} - 2 \frac{\sigma_{XT}}{\sigma_X^2} \right). \end{aligned}$$

Dato che $\sigma_{XT} = \sigma_T^2$, l'equazione precedente diventa uguale a

$$\begin{aligned} \sigma_\varepsilon^2 &= \sigma_T^2 \left(1 + \frac{\sigma_T^2}{\sigma_X^2} - 2 \frac{\sigma_T^2}{\sigma_X^2} \right) \\ &= \sigma_T^2 \left(1 - \frac{\sigma_T^2}{\sigma_X^2} \right). \end{aligned}$$

L'errore standard della stima è dunque uguale a

$$\begin{aligned}
\sigma_\varepsilon &= \sigma_T \sqrt{1 - \frac{\sigma_T^2}{\sigma_X^2}} \\
&= \sigma_T \sqrt{\frac{\sigma_X^2 - \sigma_T^2}{\sigma_X^2}} \\
&= \frac{\sigma_T}{\sigma_X} \sqrt{\sigma_X^2 - \sigma_T^2}.
\end{aligned}$$

Dato che $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$, abbiamo

$$\begin{aligned}
\sigma_\varepsilon &= \frac{\sigma_T}{\sigma_X} \sqrt{\sigma_E^2} \\
&= \frac{\sigma_T}{\sigma_X} \sigma_E \\
&= \sqrt{\rho_{XX'}} \sigma_E.
\end{aligned}$$

Ricordando che l'errore standard della misurazione è $\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}}$, possiamo scrivere

$$\begin{aligned}
\sigma_\varepsilon &= \sqrt{\rho_{XX'}} \sigma_E \\
&= \sqrt{\rho_{XX'}} \sigma_X \sqrt{1 - \rho_{XX'}} \\
&= \sigma_X \sqrt{\rho_{XX'}(1 - \rho_{XX'})}.
\end{aligned}$$

□

Per dati campionari, l'errore standard della stima si calcola nel modo seguente:

$$s_{\hat{T}} = s_X \sqrt{r_{XX'}(1 - r_{XX'})},$$

dove s_X è deviazione standard del campione e $r_{XX'}$ è il coefficiente di attendibilità.

3.3 Intervallo di confidenza per il punteggio vero

L'errore standard della stima $\sigma_{\hat{T}}$ viene usato per calcolare l'intervallo di confidenza per il punteggio vero^[2]:

$$\hat{T} \pm z\sigma_{\hat{T}},$$

laddove \hat{T} è la stima del punteggio vero e z è il quantile della normale standardizzata al livello di probabilità desiderato. Se il campione è piccolo (minore di 30) è opportuno usare t anziché z .

Si osservi che l'intervallo $\hat{T} \pm z\sigma_{\hat{T}}$ è centrato sulla *stima puntuale del valore vero* e ha una ampiezza che dipende sia dal livello di copertura desiderato (da cui dipende il quantile $z_{\frac{\alpha}{2}}$), sia dal grado di precisione dello stimatore misurato dall'errore standard della stima, $\sigma_{\hat{T}} = \sigma_X \sqrt{\rho_{XX'}(1 - \rho_{XX'})}$. L'errore standard della stima diventa tanto più grande quanto minore è l'attendibilità $\rho_{XX'}$ del test.

L'intervallo di confidenza ricorda allo psicologo quanto sia imprecisa la misura che utilizza: tanto più grande è l'intervallo di confidenza, tanto maggiore è l'incertezza dell'interpretazione. L'intervallo di confidenza è lo strumento che consente allo psicologo di giungere ad una conclusione sapendo qual è la probabilità che tale conclusione sia sbagliata. Se la decisione è basata su un intervallo di confidenza al 95%, la probabilità di sbagliare è 0.05. Se lo psicologo vuole che la probabilità d'errore sia più piccola, può costruire un intervallo di confidenza utilizzando un valore α minore. La diminuzione di α , però, produce un aumento dell'ampiezza dell'intervallo di confidenza. Valori accettabili per α sono 0.1 e 0.05.

Esercizio 3.2. Charter (1996) discute l'effetto della variazione dell'attendibilità del test sull'ampiezza dell'intervallo di confidenza per il punteggio vero. Nell'esempio considera i punteggi del QI ($\mu = 100$, $\sigma = 15$) immaginando di variare il coefficiente di attendibilità del test tramite il quale il QI viene misurato. I valori esaminati sono 0.55, 0.65, 0.75, 0.85 e 0.95. Consideriamo, ad esempio, il caso di un punteggio osservato pari a $QI = 120$ e poniamo che $\rho_{xx'} = 0.65$. In tali circostanze, la stima del punteggio vero è pari a

$$\begin{aligned}\hat{T} &= \bar{X} + r_{XX'}(X - \bar{X}) \\ &= 100 + 0.65(120 - 100) \\ &= 113.\end{aligned}$$

L'errore standard della stima è uguale a

$$\begin{aligned}
 \sigma_{\hat{T}} &= \sigma_X \sqrt{r_{XX'}(1 - r_{XX'})} \\
 &= 15 \sqrt{0.65(1 - 0.65)} \\
 &= 7.15.
 \end{aligned}$$

L'intervallo di confidenza al 95% per la stima del punteggio vero diventa pertanto uguale a

$$113 \pm 1.96 \cdot 7.15 = [98.98, 127.02].$$

Lo stesso risultato si ottiene con la funzione `CI.tscore()` del pacchetto `psychometric` che richiede i seguenti argomenti: il punteggio osservato, la media del campione, la deviazione standard del campione e l'attendibilità del test:

```
library("psychometric")
CI.tscore(120, 100, 15, 0.65)
#>   SE.Est  LCL T.Score UCL
#> 1  7.155 98.98    113 127
```

3.4 Cut-off

Uno degli usi possibili degli intervalli di confidenza per il punteggio vero è quello di confrontare i limiti dell'intervallo di confidenza con un cut-off. Sono possibili tre alternative: il limite inferiore dell'intervallo di confidenza è maggiore del cut-off, il limite superiore dell'intervallo è minore del cut-off, oppure il valore del cut-off è contenuto all'interno dell'intervallo. Nel primo caso, lo psicologo afferma, con un grado di certezza $1 - \alpha$, che il valore vero del rispondente è superiore al cut-off. Nel secondo caso, lo psicologo afferma, con un grado di certezza $1 - \alpha$, che il valore vero del rispondente è inferiore al cut-off. Nel terzo caso lo psicologo non può concludere né che il valore vero sia inferiore né che sia superiore al cut-off.

Esercizio 3.3. Si considerino i punteggi del QI, per cui $\bar{X} = 100$ e $s_X = 15$. Sia l'attendibilità del test $\rho_{XX'} = 0.95$. Supponiamo che il

rispondente abbia un $QI = 130$. Poniamo che il cut-off per ammettere il rispondente ad un corso avanzato sia 120. Ci sono tre alternative: il valore vero del rispondente è sicuramente maggiore di 120; il valore vero del rispondente è sicuramente inferiore di 120; le evidenze disponibili ci lasciano in dubbio se il punteggio vero sia maggiore o minore di 120. Svolgiamo i calcoli per trovare l'intervallo di confidenza al livello di certezza del 95%:

```
xm <- 100
sx <- 15
rho <- .95
x <- 130
t.hat <- xm + rho * (x - xm)
t.hat
#> [1] 128.5
se.t <- sx * sqrt(rho * (1 - rho))
se.t
#> [1] 3.269
t.hat + c(1, -1) * qnorm(.025, 0, 1) * se.t
#> [1] 122.1 134.9
```

Dato che il limite inferiore dell'intervallo di confidenza è maggiore del cut-off, lo psicologo conclude che il punteggio vero del rispondente è maggiore di 120. Quindi, raccomanda che il rispondente sia ammesso al corso avanzato.

Continuiamo con l'esempio precedente, ma supponiamo che l'attendibilità del test abbia un valore simile a quello che solitamente si ottiene empiricamente, ovvero 0.80.

```
xm <- 100
sx <- 15
rho <- .8
x <- 130
t.hat <- xm + rho * (x - xm)
t.hat
#> [1] 124
se.t <- sx * sqrt(rho * (1 - rho))
se.t
```

```
#> [1] 6
t.hat + c(1, -1) * qnorm(.025, 0, 1) * se.t
#> [1] 112.2 135.8
```

In questo secondo esempio, l'intervallo di confidenza al 95% è [112.24, 135.76] e contiene il valore del cut-off. Dunque, la decisione dello psicologo è che non vi sono evidenze sufficienti che il vero valore del rispondente sia superiore al cut-off. Si noti come la diminuzione dell'attendibilità del test porta all'aumento delle dimensioni dell'intervallo di confidenza.

3.5 Procedure alternative

Non vi è un unico modo per costruire gli intervalli di confidenza per il punteggio vero. Charter e Feldt (1991) descrivono altri quattro approcci possibili, oltre a quello discusso qui, per costruire gli intervalli di confidenza per il punteggio vero. L'approccio che abbiamo descritto è accettato da tutti gli autori; le procedure alternative descritte da Charter e Feldt (1991), non sono invece accettate come valide da tutti gli autori.

La più comune delle procedure alternative descritte da Charter e Feldt (1991), che rappresenta l'approccio tradizionale a questo problema, centra l'intervallo di confidenza sul punteggio osservato di un rispondente e utilizza l'errore standard della misurazione per calcolare i limiti dell'intervallo di confidenza:

$$X_j \pm z_{\frac{\alpha}{2}} \sigma_E,$$

dove $\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}}$. Tale procedura è però stata criticata da diversi autori (es., Dudek, 1979).

Parte II

L'analisi fattoriale



4

Utilizzo e costruzione di test psicometrici

La maggior parte degli psicologi che utilizzano i test non costruiscono dei test ad hoc ma utilizzano i test già validati e interpretano i punteggi attenuti sulla base delle norme fornite nel manuale del test. Nella selezione del test che è più appropriato per il problema che lo psicologo deve affrontare, si devono considerare domande quali: Qual è il valore minimo di attendibilità che è richiesto? Ovvero, a che livello di precisione dobbiamo essere in grado di differenziare tra i rispondenti? Che tipo di validità è importante? Quali sono le caratteristiche della popolazione che il gruppo normativo deve rappresentare? Qual è il livello di istruzione necessario per completare il test? Qual è il tempo disponibile per la somministrazione del test e per lo scoring dei risultati? Quali sono i costi necessari per la somministrazione del test e per lo scoring dei risultati? Avendo dato una risposta a tali domande, la selezione del test solitamente si riduce ad una scelta tra poche alternative. Per giungere ad una decisione tra tali alternative è importante consultare la letteratura specialistica che discute le proprietà psicometriche dei test e la loro validità.

Ciascuno psicologo ha l'obbligo di dimostrare che il test che usa per un certo scopo costituisca lo strumento migliore, tra quelli disponibili, per giungere ad una decisione razionale e obiettiva relativamente al problema che si trova ad affrontare (si veda il Codice Deontologico). Se nessuno dei test disponibili si dimostra appropriato per misurare un determinato tratto psicologico, si procede alla costruzione di un nuovo reattivo. La costruzione di un test richiede sia conoscenze specialistiche di tipo psicometrico, sia una conoscenza specialistica del fenomeno considerato.

Le fasi di costruzione dei test comprendono la definizione delle aree di contenuto che andranno misurate dal test; la generazione degli item per ciascuna area in un numero di circa tre volte superiore a quello che ci si aspetta farà parte della versione finale del test; la somministrazione degli item ad un campione sufficientemente numeroso (centinaia di ri-

spondenti selezionati in modo tale che il campione sia rappresentativo della popolazione di interesse); l'analisi degli item che ci consente di selezionare gli item migliori; infine, la somministrazione della versione revisionata del test ad un nuovo campione per stabilire se la versione finale del test sia soddisfacente dal punto di vista psicometrico. In caso affermativo, il campione esaminato fornisce le norme del test e questa fase va sotto il nome di *standardizzazione del test*.

4.1 Caratteristiche dell'analisi fattoriale

La teoria psicomетria è l'unico strumento che abbiamo a disposizione per creare strumenti di misurazione validi e attendibili per l'assessment psicologico e neuropsicologico. La psicomетria comprende due diversi approcci metodologici per la costruzione dei reattivi psicologici: la teoria classica dei test e la teoria di risposta all'item (*item response theory*). Esamineremo qui gli aspetti di base della teoria classica dei test e ci focalizzeremo, in particolare, sull'analisi fattoriale.

Se vogliamo costruire un test psicomетrico per la valutazione di un particolare deficit psicologico o neuropsicologico, prima di iniziare lo studio, dobbiamo fornire una risposta ad una serie di domande. Per esempio, come dobbiamo selezionare gli item? Gli item scelti coprono tutto il dominio del fenomeno considerato? Quanti item dobbiamo usare? A quanti soggetti dobbiamo somministrare lo strumento? Quali analisi statistiche dobbiamo svolgere sui dati raccolti?

L'analisi fattoriale è uno strumento statistico che può essere utilizzato per trovare una risposta a queste domande. Chiariamo subito che l'analisi fattoriale è una tecnica statistica complessa che richiede l'uso di un software. Per gli scopi di questo insegnamento useremo il pacchetto *lavaan* del linguaggio statistico .

L'analisi fattoriale può essere pensata come una tecnica statistica per la ricerca di variabili latenti a partire da alcune variabili osservate. La distinzione tra variabili latenti e variabili osservate si basa sulla osservabilità, ossia sulla possibilità di rilevazione empirica. Le prime sono variabili non direttamente osservabili in quanto rappresentano concetti molto generali o complessi, mentre le seconde sono facilmente rilevabili.

In ogni caso, entrambe possono essere operazionalizzate, per cui anche nel caso delle variabili latenti c'è una sostanziale differenza con i concetti.

Ma che cos'è una variabile latente o fattore? Un fattore può essere descritto come una combinazione lineare di variabili manifeste tra loro associate le quali rappresentano una specifica dimensione di un costrutto psicologico, la quale si distingue da altre dimensioni dello stesso costrutto o dalle dimensioni di costrutti diversi (Tabachnick & Fidell, 2001). Per esempio, la Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV) dà una valutazione complessiva delle capacità cognitive di adolescenti e adulti e distingue tra quattro dimensioni dell'intelligenza: comprensione verbale, ragionamento visuo-percettivo, memoria di lavoro e velocità di elaborazione.

L'analisi fattoriale viene usata per lo sviluppo e la validazione di un test psicometrico e per stabilire la validità di costrutto di uno strumento per una specifica popolazione. Una volta che la struttura interna di un costrutto è stata chiarita, l'analisi fattoriale può essere usata per identificare le variabili esterne (per esempio, il genere o il livello di istruzione) che sono associate alle varie dimensioni del costrutto di interesse (Nunnally & Bernstein, 1994).

4.1.1 L'analisi fattoriale esplorativa e confermativa

Nei suoi primi modelli matematici, l'analisi fattoriale era una tecnica di analisi esplorativa appunto perché permetteva di “esplorare” le relazioni nascoste fra un gran numero di variabili. In tempi più recenti, la tecnica delle equazioni strutturali ha permesso di sviluppare una tecnica di analisi fattoriale di tipo confermativo, che cioè permette di verificare se effettivamente i fattori ipotizzati servono a spiegare le variabili misurate. In tempi ancora più recenti è stato notato che i vincoli posti dai modelli SEM sono, alle volte, troppo restrittivi per cui le tecniche dell'analisi fattoriale esplorativa vengono integrate con i modelli di equazioni strutturali in maniera tale da produrre un nuovo approccio alla definizione delle variabili latenti, ovvero quelli che si chiamano modelli ESEM (Exploratory Structural Equation Modeling).

4.1.2 Assunzioni dell'analisi fattoriale

L'assunzione di base dell'analisi fattoriale esplorativa è che, dato un insieme di variabili osservate, esista un insieme più piccolo di fattori latenti

i quali siano in grado di spiegare i legami, le interrelazioni e le dipendenze tra le variabili statistiche osservate. Dato che l'analisi fattoriale si propone di spiegare una matrice di correlazioni, molte delle assunzioni che stanno alla base del calcolo delle correlazioni si applicano anche all'analisi fattoriale: campioni di grande numerosità, il fatto che le variabili considerate abbiano una distribuzione di probabilità continua, la linearità nella relazione tra le variabili. In generale, queste assunzioni si riducono al requisito della normalità multivariata, ovvero il requisito per il quale tutte le variabili considerate e tutte le combinazioni lineari tra tali variabili sono distribuite normalmente.

Molto spesso, però, l'assunzione di normalità multivariata è violata a causa del fatto che, anziché essere delle variabili continue, gli item sono spesso variabili (discrete) che derivano da risposte a questionari fornite utilizzando scale di tipo Likert. Vedremo come questo problema può essere superato utilizzando le correlazioni polioriche.

4.1.3 Sviluppo storico dell'analisi fattoriale

L'analisi fattoriale è un metodo di analisi multivariata che, dal punto di vista storico, risulta fortemente interconnessa con i modelli psicometrici. Alcuni concetti basilari per i successivi sviluppi dell'analisi fattoriale si possono trovare in Galton (1888, 1889) con il concetto di *fonte comune* (*common source*) e in Pearson (1901) con l'introduzione delle *componenti principali* come metodo di riduzione dei dati. Si deve però a Spearman (1904) l'elaborazione del primo *modello di analisi fattoriale* secondo il quale le risposte fornite ad un insieme di test di abilità sono riconducibili ad un unico fattore generale di intelligenza.

L'analisi fattoriale è stata sviluppata da diversi gruppi di psicologi inglesi e americani che, tra l'inizio del '900 e il 1930, svilupparono una serie di tecniche statistiche per affrontare problemi quali quello di determinare le dimensioni dell'intelligenza (Garnett, 1919; Pearson, 1901; Spearman, 1904, 1923, 1927, 1929; Thurstone, 1935, 1937a, 1937b; Wilson, 1928).

In Inghilterra, Spearman e i suoi colleghi (Burt, 1939, 1941; Garnett, 1919; Ledermann, 1937, 1938; Spearman, 1904, 1922, 1923, 1927, 1928, 1929, 1930a, 1930b; Thomson, 1934, 1936, 1938) svilupparono il concetto della teoria dei due fattori dell'intelligenza umana secondo cui esiste un fattore generale, g , che esprime l'intelligenza generale, è ereditario e compare in maggiore o minor misura in tutti i test, e esistono inoltre tanti

fattori specifici, s , che rappresentano l'acquisizione specifica attraverso l'apprendimento e l'esperienza, e sono presenti in modo differenziato in ogni singolo test. Dato che i fattori specifici sono incorrelati tra loro, il fattore g spiega la maggior parte delle correlazioni tra i punteggi delle abilità mentali (Nunnally & Bernstein, 1994). Anche se successive ricerche non confermarono totalmente l'ipotesi di Spearman dell'esistenza di un fattore g , e particolarmente vennero alla luce dei fattori di gruppo, che potevano spiegare le correlazioni esistenti tra test simili (Garnett, 1919; Spearman, 1927), la maggior parte dello sviluppo iniziale dell'analisi fattoriale fu dedicata alla ricerca di un singolo fattore per l'intelligenza.

Spearman analizzò molti insiemi di dati nel tentativo di trovare conferma al modello uni-fattoriale: alcuni di questi, tuttavia, non risultarono compatibili con il modello dei due fattori. La generalizzazione del modello fattoriale al caso in cui sono presenti più fattori comuni si deve a Thurstone (1938, 1947). Nel modello di Thurstone non viene fatto più riferimento ad un fattore generale che influenza tutte le variabili osservate: si ipotizza invece la presenza di una molteplicità di fattori comuni i quali determinano le variabili osservate. I contributi di Thurstone riguardano alcuni degli aspetti fondamentali dell'analisi fattoriale esplorativa, quali il concetto di struttura semplice e i fattori obliqui.

Thurstone e collaboratori applicarono la teoria dei fattori multipli a vari problemi psicometrici (Thurstone, 1931, 1940, 1947, 1948, 1954) diventando ben presto il riferimento più importante per lo sviluppo delle procedure dell'analisi fattoriale. La centralità della scuola statunitense continuò fino alla fine degli anni '60, quando un numero di importanti psicometristi europei contribuirono a spostare il focus dell'assessment psicologico dall'approccio esplorativo a quello confermativo (Jöreskog, 1967, 1969, 1970; Jöreskog & Goldberger, 1972; Sörbom, 1974). I contributi statistici di Jöreskog e Sörbom hanno contribuito a sviluppare quelli che oggi chiamiamo i modelli di equazioni strutturali (Structural Equation Models, SEM), i quali costituiscono una delle tecniche più utilizzate per l'analisi dei dati nelle discipline psicologiche e sociali.

Un'ulteriore linea di ricerca venne sviluppata nel campo della biometria, ad opera di Sewall Wright (1934). In un tale approccio vengono utilizzati modelli di equazioni simultanee, che includono però solo variabili osservabili, per analizzare particolari schemi di rappresentazione dei nessi di influenza tra le variabili noti come *analisi dei percorsi* (*path analysis*).

A partire dalla metà degli anni cinquanta, con la diffusione dei centri

calcolo, l'analisi fattoriale è diventata uno dei metodi di analisi dei dati più largamente utilizzati in psicologia e, in particolare, nelle ricerche volte alla costruzione dei reattivi psicologici. Le tradizioni di ricerca dei modelli di equazioni simultanee con variabili osservate, da una parte, e dei modelli di equazioni strutturali con variabili osservate e latenti, dall'altra, sono rimaste sostanzialmente indipendenti fino agli anni '60 dello scorso secolo, quando metodologi delle scienze sociali, come Blalock (1961, 1963), Boudon (1965) e Duncan (1966) hanno cominciato a sottolineare i vantaggi derivanti dalla possibilità di combinare la semplicità di rappresentare nessi di influenza tra le variabili tramite i diagrammi tipici della *path analysis*, con il rigore derivante dalla specificazione delle equazioni simultanee, in modo tale da includere sia variabili osservate sia variabili latenti. La disponibilità di programmi per computer che permettono di tradurre in un linguaggio non matematico le complesse operazioni matematiche connesse alla risoluzione simultanea di sistemi di equazioni lineari con variabili latenti ha poi facilitato enormemente la diffusione delle tecniche dell'analisi fattoriale confermativa. A partire dalla fine degli anni sessanta, soprattutto grazie a Karl Jöreskog, il modello di analisi fattoriale si è ulteriormente raffinato e sviluppato verso un approccio confermativo orientato prevalentemente all'esame di ipotesi teoriche.

Sebbene l'analisi fattoriale confermativa (CFA) possa essere vista come una motivazione per la creazione di modelli di misurazione parsimoniosi, tali modelli spesso includono un certo livello di errore di misurazione sistematico che deriva dal fatto che gli item sono raramente indicatori puri dei corrispondenti costrutti, e quindi ci si può aspettare un certo grado di associazione rilevante per il costrutto anche tra item e costrutti non target, ma concettualmente associati (Morin et al., 2016). Quando tali associazioni restano inesprese nel modello, a causa dei vincoli restrittivi dell'approccio CFA (cioè, del fatto che gli item possono essere indicatori di un unico fattore), allora i risultati dell'analisi possono risultare sistematicamente distorti. In particolare, si può osservare una sovrastima delle associazioni tra i fattori. Inoltre, i vincoli eccessivamente restrittivi dei modelli CFA possono anche minare la bontà di adattamento dei modelli e la validità discriminante dei fattori (Marsh et al., 2010, 2014). I contemporanei modelli ESEM si pongono l'obiettivo di affrontare questo tipo di problemi.

4.1.4 Fasi dell'analisi fattoriale

Indipendentemente dal software statistico che viene utilizzato e dal livello di esperienza del ricercatore, lo sviluppo di un test psicometrico inizia con l'analisi fattoriale esplorativa. In essa si possono distinguere otto stadi successivi:

1. la specificazione del problema,
2. la scelta degli item,
3. la verifica dell'adeguatezza della matrice di correlazione,
4. l'estrazione iniziale dei fattori,
5. la rotazione dei fattori,
6. l'ulteriore selezione degli item,
7. l'interpretazione dei risultati,
8. la costruzione di un report della ricerca,
9. la validazione dei risultati.

Queste diverse fasi della costruzione e validazione di un reattivo psicologico verranno discusse nel seguito di queste dispense.



5

Il modello unifattoriale

In questo capitolo verranno presentate le basi teoriche dell'analisi fattoriale, ovvero di quel modello statistico che offre la possibilità di ricostruire le correlazioni osservate tra le variabili manifeste considerando le saturazioni delle variabili in uno o più fattori generali. Nell'analisi fattoriale p variabili manifeste (item) vengono concepite come condizionalmente indipendenti date m variabili latenti chiamate *fattori*. L'analisi fattoriale si pone lo scopo di interpretare i fattori come dei costrutti teorici inosservabili. Infatti, il desiderio di spiegare mediante il concetto di intelligenza le correlazioni osservate tra le prestazioni di un gruppo di individui in una serie di compiti è stato la forza trainante nello sviluppo originale dell'analisi fattoriale. L'analisi fattoriale consente di identificare i costrutti di cui gli item sono espressione e di stabilire in che misura ciascun item rappresenta il costrutto. Il modello unifattoriale ipotizza $m = 1$; il modello multifattoriale ipotizza $m > 1$. Lo scopo di questo capitolo è quello di introdurre il modello fattoriale che assume l'esistenza di un unico fattore comune latente.

5.1 Modello monofattoriale

Con p variabili manifeste y_i , il caso più semplice è quello di un solo fattore comune:

$$y_i = \mu_i + \lambda_i \xi + 1 \cdot \varepsilon_i \quad i = 1, \dots, p, \quad (5.1)$$

dove ξ rappresenta il fattore comune a tutte le y_i , ε_i sono i fattori specifici o unici di ogni variabile osservata e λ_i sono le saturazioni (o pesi) fattoriali le quali stabiliscono il peso del fattore latente su ciascuna variabile osservata.

Si noti che il modello di analisi fattoriale è solo apparentemente simile al modello di regressione. Infatti, sia il fattore comune ξ sia i fattori specifici ε_i sono inosservabili: tutto ciò che giace a destra dell'uguaglianza è dunque incognito. L'analisi di regressione e l'analisi fattoriale si differenziano non solo per tale aspetto, ma anche per il fatto di avere obiettivi diversi. L'analisi di regressione ha l'obiettivo di individuare le variabili esplicative, direttamente osservabili, che sono in grado di spiegare la maggior parte della *varianza* della variabile dipendente. Il problema dell'analisi unifattoriale, invece, è quello di identificare la variabile esplicativa inosservabile che è in grado di spiegare la maggior parte della *covarianza* tra le variabili osservate.

Nel caso di cinque variabili osservate e un solo fattore comune, ad esempio, il modello fattoriale (5.1) può essere rappresentato graficamente nel modo seguente.

Si suole assumere per comodità che $\mu = 0$, il che corrisponde a considerare le variabili y_i come ottenute dagli scarti dalle medie μ_i , per $i = 1, \dots, p$:

$$y_i - \mu_i = \lambda_i \xi + 1 \cdot \varepsilon_i. \quad (5.2)$$

Si assume che il fattore comune abbia media zero, $\mathbb{E}(\xi) = 0$, e varianza unitaria, $\mathbb{V}(\xi) = 1$, i fattori specifici abbiano media zero, $\mathbb{E}(\varepsilon_j) = 0$, varianza $\mathbb{V}(\varepsilon_i) = \psi_i$ e siano incorrelati tra loro, $\mathbb{E}(\varepsilon_i \varepsilon_k) = 0$, e con il fattore comune, $\mathbb{E}(\varepsilon_i \xi) = 0$. In questo modello, poiché i fattori specifici sono tra loro incorrelati, l'interdipendenza tra le variabili è completamente spiegata dal fattore comune.

Dalle ipotesi precedenti è possibile ricavare:

- la covarianza tra y_i e il fattore comune,
- la varianza della i -esima variabile osservabile y_i ,
- la covarianza tra due variabili y_i e y_k .

Questo sarà l'obiettivo della discussione presente in questo capitolo.

5.2 Correlazione parziale

Prima di discutere il modello statistico dell'analisi fattoriale, chiariamo il concetto di correlazione parziale. La nascita dell'analisi fattoriale viene di solito attribuita a Charles Spearman. Nel 1904, Spearman pubblicò un articolo dal titolo “General Intelligence, objectively determined and measured” dove propose la Teoria dei Due Fattori. Nel suo articolo del 1904, Spearman dimostrò come, mediante il metodo dell'annullamento della tetrade (*tetrad differences*), sia possibile identificare un fattore inosservabile a partire da una matrice di correlazioni. L'annullamento della tetrade rappresenta un'applicazione della teoria della correlazione parziale. Il problema è quello di stabilire se, controllando un insieme di variabili inosservabili ξ_j , dette fattori, le correlazioni tra le variabili osservabili Y_i , al netto degli effetti lineari delle ξ_j , diventino statisticamente nulle.

Consideriamo un esempio nel quale sono presenti solo tre variabili: Y_1 , Y_2 e F . In generale, la correlazione $r_{1,2}$ tra due variabili Y_1 e Y_2 può risultare dalla loro associazione con una terza variabile F . Per calcolare la correlazione parziale tra Y_1 e Y_2 al netto dell'effetto lineare di F è necessario trovare le componenti di Y_1 e di Y_2 che sono linearmente indipendenti da F .

Vediamo come si può ottenere questo risultato. La componente di Y_1 linearmente indipendente da F è data dai residui E_1 del modello

$$Y_1 = b_{01} + b_{11}F + E_1.$$

La componente di Y_2 linearmente indipendente da F è data dai residui E_2 del modello

$$Y_2 = b_{02} + b_{12}F + E_2.$$

La correlazione parziale $r_{1,2|F}$ è la correlazione di Pearson tra E_1 e E_2 , ovvero la correlazione tra le componenti di Y_1 e Y_2 linearmente indipendenti da F .

La correlazione parziale tra Y_1 e Y_2 al netto dell'effetto di F può essere calcolata direttamente dalle correlazioni semplici tra le tre variabili Y_1 , Y_2 e F mediante la seguente formula:

$$r_{1,2|F} = \frac{r_{12} - r_{1F}r_{2F}}{\sqrt{(1 - r_{1F}^2)(1 - r_{2F}^2)}} \quad (5.3)$$

Facciamo un esempio numerico. Sia f una variabile su cui misuriamo n valori

```
set.seed(123)
n <- 1000
f <- rnorm(n, 24, 12)
```

Siano y_1 e y_2 funzioni lineari di f , a cui viene aggiunta una componente d'errore gaussiano:

```
y1 <- 10 + 7 * f + rnorm(n, 0, 50)
y2 <- 3 + 2 * f + rnorm(n, 0, 50)
```

La correlazione tra y_1 e y_2 ($r_{12} = 0.355$) deriva dal fatto che \hat{y}_1 e \hat{y}_2 sono entrambe funzioni lineari di f :

```
Y <- cbind(y1, y2, f)
R <- cor(Y)
round(R, 3)
#>      y1      y2      f
#> y1 1.000 0.380 0.867
#> y2 0.380 1.000 0.423
#> f  0.867 0.423 1.000
```

Eseguiamo le regressioni di y_1 su f e di y_2 su F :

```
fm1 <- lm(y1 ~ f)
fm2 <- lm(y2 ~ f)
```

Nella regressione, ciascuna osservazione y_{i1} viene scomposta in due componenti linearmente indipendenti, i valori adattati \hat{y}_i e i residui, e_i : $y_i = \hat{y}_i + e_i$. Nel caso di y_1 abbiamo

```
round(head(cbind(y1, y1.hat = fm1$fit, e = fm1$res, fm1$fit + fm1$res)), 3)
#>      y1 y1.hat      e
#> 1  81.13 130.5 -49.375  81.13
#> 2 106.67 159.7 -53.037 106.67
#> 3 308.03 317.8  -9.813 308.03
#> 4 177.31 186.3  -8.971 177.31
#> 5  61.39 191.5 -130.089  61.39
#> 6 374.09 331.7  42.426 374.09
```

Lo stesso può dirsi di y_2 . La correlazione parziale $r_{12|f}$ tra y_1 e y_2 dato f è uguale alla correlazione di Pearson tra i residui e_1 e e_2 calcolati mediante i due modelli di regressione descritti sopra:

```
cor(fm1$res, fm2$res)
#> [1] 0.02829
```

La correlazione parziale tra y_1 e y_2 al netto di f è .02829.

Per i dati esaminati sopra, dunque, la correlazione parziale tra le variabili y_1 e y_2 diventa uguale a zero se la variabile f viene controllata (ovvero, se escludiamo da y_1 e da y_2 l'effetto lineare di f). Il fatto che la correlazione parziale sia zero significa che la correlazione che abbiamo osservato tra y_1 e y_2 ($r = 0.355$) non dipendeva dall'effetto che una variabile y esercitava sull'altra, ma bensì dal fatto che c'era una terza variabile, f , che influenzava sia y_1 sia y_2 . In altre parole, le variabili y_1 e y_2 sono condizionalmente indipendenti dato f . Ciò significa, come abbiamo visto sopra, che la componente di y_1 linearmente indipendente da f è incorrelata con la componente di y_2 linearmente indipendente da f .

La correlazione che abbiamo calcolato tra i residui di due modelli di regressione non è altro che la correlazione che viene calcolata applicando la (5.3). Infatti, inserendo nella (5.3) i valori delle correlazioni esaminate otteniamo

```
(R[1, 2] - R[1, 3] * R[2, 3]) /
  sqrt((1 - R[1, 3]^2) * (1 - R[2, 3]^2)) %>%
  round(3)
#> [1] 0.02828
```

In conclusione, possiamo dunque attribuire alla (5.3) la seguente interpretazione: la correlazione parziale tra le variabili y_1 e y_2 dato f non è altro che la correlazione tra le componenti di y_1 e y_2 da cui l'effetto lineare di f è stato rimosso.

5.3 Principio base dell'analisi fattoriale

Attualmente, l'inferenza statistica nell'analisi fattoriale spesso si svolge mediante il calcolo di stime della massima verosimiglianza ottenute mediante procedure iterative come l'algoritmo EM (Rubin & Thayer, 1982). All'inizio dell'analisi fattoriale, tuttavia, la procedura di estrazione dei fattori faceva leva sulle relazioni invarianti che il modello fattoriale impone agli elementi della matrice di covarianza delle variabili osservate. Il più conosciuto tra tali invarianti è la *tetrad* che si presenta nei modelli ad un fattore.

La *tetrad* è una combinazione di quattro correlazioni. Se l'associazione osservata tra le variabili dipende effettivamente dal fatto che le variabili in questione sono state causalmente generate da un fattore comune inosservabile, allora è possibile generare una combinazione delle correlazioni tra le variabili che porta all'annullamento della *tetrad*. In altre parole, l'analisi fattoriale si chiede se esiste un insieme esiguo di $m < p$ variabili inosservabili che rendono significativamente nulle tutte le correlazioni parziali tra le p variabili osservate al netto dei fattori comuni. Se il metodo della correlazione parziale consente di identificare m variabili latenti, allora lo psicologo conclude che tali fattori corrispondono agli m costrutti che intende misurare.

Per chiarire il metodo dell'annullamento della *tetrad* consideriamo la matrice di correlazioni riportata nella Tabella @ref(tab:corr_parziale). Nella tabella, la correlazione parziale tra ciascuna coppia di variabili y_i , y_j (con $i \neq j$) dato ξ è sempre uguale a zero. Ad esempio, la correlazione parziale tra y_3 e y_5 dato ξ è:

$$\begin{aligned}
r_{35|\xi} &= \frac{r_{35} - r_{3\xi}r_{5\xi}}{\sqrt{(1 - r_{3\xi}^2)(1 - r_{5\xi}^2)}} \\
&= \frac{0.35 - 0.7 \times 0.5}{\sqrt{(1 - 0.7^2)(1 - 0.5^2)}} = 0.
\end{aligned}$$

Lo stesso risultato si trova per qualunque altra coppia di variabili y_i e y_j , ovvero $r_{ij|\xi} = 0$.

Tabella 5.1: Matrice di correlazioni nella quale tutte le correlazioni parziali tra le variabili Y al netto dell'effetto di ξ sono nulle.

	ξ	y_1	y_2	y_3	y_4	y_5
ξ	1.00					
y_1	0.90	1.00				
y_2	0.80	0.72	1.00			
y_3	0.70	0.63	0.56	1.00		
y_4	0.60	0.54	0.48	0.42	1.00	
y_5	0.50	0.45	0.40	0.35	0.30	1.00

Possiamo dunque dire che, per la matrice di correlazioni della Tabella precedente, esiste un'unica variabile ξ la quale, quando viene controllata, spiega tutte le

$$p(p-1)/2 = 5(5-1)/2 = 10$$

correlazioni tra le variabili y . Questo risultato non è sorprendente, in quanto la matrice di correlazioni della Tabella esaminata è stata costruita in modo tale da possedere tale proprietà.

Ma supponiamo di essere in una situazione diversa, ovvero di avere osservato soltanto le variabili y_i e di non conoscere ξ . In tali circostanze ci possiamo porre la seguente domanda: “esiste una variabile inosservabile ξ la quale, se venisse controllata, renderebbe uguali a zero tutte le correlazioni parziali tra le variabili y ?” Se una tale variabile inosservabile esiste, ed è in grado di spiegare tutte le correlazioni tra le variabili osservate y , allora essa viene chiamata *fattore*.

Definizione 5.1. Un fattore è una variabile inosservabile in grado di rendere significativamente nulle tutte le correlazioni parziali tra le variabili manifeste.

5.3.1 Vincoli sulle correlazioni

Come si può stabilire se esiste una variabile inosservabile in grado di rendere nulle tutte le correlazioni parziali tra le variabili osservate? Riscriviamo la (5.3) per specificare la correlazione parziale tra le variabili y_i e y_j dato ξ :

$$r_{ij|\xi} = \frac{r_{ij} - r_{i\xi}r_{j\xi}}{\sqrt{(1 - r_{i\xi}^2)(1 - r_{j\xi}^2)}}$$

Affinché $r_{ij|\xi}$ sia uguale a zero è necessario che

$$r_{ij} - r_{i\xi}r_{j\xi} = 0$$

ovvero

$$r_{ij} = r_{i\xi}r_{j\xi}.$$

In altri termini, se esiste un fattore non osservato ξ in grado di rendere uguali a zero tutte le correlazioni parziali $r_{ih|\xi}$, allora la correlazione tra ciascuna coppia di variabili y deve essere uguale al prodotto delle correlazioni tra ciascuna y e il fattore latente ξ . Questo è il principio base dell'analisi fattoriale.

5.3.2 Teoria dei Due Fattori

Per fare un esempio concreto relativo al metodo dell'annullamento della tetrade, esaminiamo la matrice di correlazioni originariamente analizzata da Spearman. Spearman (1904) raccolse alcune misure di capacità intellettuale su un piccolo numero di studenti di una scuola superiore. Nello specifico, esaminò i voti di tali studenti nelle seguenti materie: studio dei classici (c), letteratura inglese (e) e abilità matematiche (m). Considerò anche la prestazione in un compito di discriminazione dell'altezza di suoni (*pitch discrimination*) (p), ovvero un'abilità diversa da quelle richieste nei test scolastici.

Secondo la Teoria dei Due Fattori, le prestazioni relative ad un determinato compito intellettuale possiedono una componente comune (detta fattore “g”) con le prestazioni in un qualunque altro compito intellettuale e una componente specifica a quel determinato compito. Il modello dell'intelligenza di Spearman prevede dunque due fattori, uno generale e uno specifico (detto fattore “s”). Il fattore “g” costituisce la componente invariante dell'abilità intellettuale, mentre il fattore “s” è una componente che varia da condizione a condizione.

Come è possibile stabilire se esiste una variabile latente in grado di spiegare le correlazioni tra le variabili osservate da Spearman? Lo strumento proposto da Spearman per rispondere a questa domanda è *l'annullamento della tetrade*. L'annullamento della tetrade utilizza i vincoli sulle correlazioni che derivano dalla definizione di correlazione parziale. In precedenza abbiamo visto che la correlazione parziale tra le variabili y indicizzate da i e j , al netto dell'effetto di ξ , è nulla se

$$r_{ij} = r_{i\xi}r_{j\xi}.$$

Nel caso dei dati di Spearman, dunque, le correlazioni parziali sono nulle se la correlazione tra “studi classici” e “letteratura inglese” è uguale al prodotto della correlazione tra “studi classici” e il fattore ξ e della correlazione tra “letteratura inglese” e il fattore ξ . Inoltre, la correlazione tra “studi classici” e “abilità matematica” deve essere uguale al prodotto della correlazione tra “studi classici” e il fattore ξ e della correlazione tra “abilità matematica” e il fattore ξ ; e così via.

Le correlazioni tra le variabili manifeste e il fattore latente sono dette *saturazioni fattoriali* e vengono denotate con la lettera λ . Se il modello di Spearman è corretto, avremo che

$$r_{ec} = \lambda_e \times \lambda_c,$$

dove r_{ec} è la correlazione tra “letteratura inglese” (e) e “studi classici” (c), λ_e è la correlazione tra “letteratura inglese” e ξ , e λ_c è la correlazione tra “studi classici” e ξ .

Allo stesso modo, la correlazione tra “studi classici” e “matematica” (m) dovrà essere uguale a

$$\lambda_c \times \lambda_m,$$

eccetera.

5.3.3 Annullamento della tetrade

Date le correlazioni tra tre coppie di variabili manifeste, il metodo dell'annullamento della tetrade rende possibile stimare i valori delle saturazioni fattoriali λ . Ad esempio, per le variabili c , m ed e , possiamo scrivere le seguenti tre equazioni in tre incognite:

$$\begin{aligned} r_{cm} &= \lambda_c \times \lambda_m, \\ r_{em} &= \lambda_e \times \lambda_m, \\ r_{ce} &= \lambda_c \times \lambda_e. \end{aligned}$$

Risolvendo il precedente sistema di equazioni lineari, il coefficiente di saturazione λ_m della variabile y_m nel fattore comune ξ , ad esempio, può essere calcolato a partire dalle correlazioni tra le variabili manifeste c , m , ed e nel modo seguente:

$$\lambda_m = \sqrt{\frac{r_{cm}r_{em}}{r_{ce}}}. \quad (5.4)$$

Lo stesso vale per le altre due saturazioni λ_c e λ_e .

Nel suo articolo del 1904, Spearman osservò le seguenti correlazioni tra le variabili Y_c , Y_e , Y_m e Y_p :

	Y_C	Y_E	Y_M	Y_P
Y_C	1.00	0.78	0.70	0.66
Y_E		1.00	0.64	0.54
Y_M			1.00	0.45
Y_P				1.00

Utilizzando la (5.4), mediante le correlazioni r_{cm} , r_{em} , e r_{ce} fornite dalla tabella precedente, la saturazione λ_m diventa uguale a:

$$\hat{\lambda}_m = \sqrt{\frac{r_{cm}r_{em}}{r_{ce}}} = \sqrt{\frac{0.70 \times 0.64}{0.78}} = 0.76.$$

È importante notare che il metodo dell'annullamento della tetrade produce risultati falsificabili. Infatti, ci sono modi diversi per calcolare la

stessa saturazione fattoriale. Se il modello fattoriale è corretto si deve ottenere lo stesso risultato in tutti i casi. Nel caso presente, la saturazione fattoriale λ_m può essere calcolata in altri due modi:

$$\begin{aligned}\hat{\lambda}_m &= \sqrt{\frac{r_{cm}r_{mp}}{r_{cp}}} = \sqrt{\frac{0.78 \times 0.45}{0.66}} = 0.69, \\ \hat{\lambda}_m &= \sqrt{\frac{r_{em}r_{mp}}{r_{ep}}} = \sqrt{\frac{0.64 \times 0.45}{0.54}} = 0.73.\end{aligned}$$

I tre valori che sono stati ottenuti sono molto simili. Qual è allora la stima migliore di λ_m ?

5.3.4 Metodo del centroide

La soluzione più semplice è quella di fare la media di questi tre valori ($\bar{\lambda}_m = 0.73$). Un metodo migliore (meno vulnerabile ai valori anomali) è dato dal rapporto tra la somma dei numeratori e dei denominatori:

$$\hat{\lambda}_m = \sqrt{\frac{0.70 \times 0.64 + 0.78 \times 0.45 + 0.64 \times 0.45}{0.78 + 0.66 + 0.54}} = 0.73$$

In questo caso, i due metodi danno lo stesso risultato. Le altre tre saturazioni fattoriali trovate mediante il metodo del centroide sono:

$$\hat{\lambda}_c = 0.97, \quad \hat{\lambda}_e = 0.84, \quad \hat{\lambda}_p = 0.65.$$

In conclusione,

$$\hat{\Lambda}' = (\hat{\lambda}_c, \hat{\lambda}_e, \hat{\lambda}_m, \hat{\lambda}_p) = (0.97, 0.84, 0.73, 0.65).$$

5.3.5 Introduzione a lavaan

Analizziamo ora nuovamente gli stessi dati usando un metodo di stima moderno (massima verosimiglianza), mediante le funzioni del pacchetto `lavaan`. La matrice completa dei dati di Spearman è messa a disposizione da [Kan et al. \(2019\)](#). Iniziamo a caricare i pacchetti necessari:

```
library("lavaan")
library("semPlot")
library("knitr")
library("kableExtra")
library("tidyr")
library("corrplot")
```

Specifichiamo il nome delle variabili manifeste

```
varnames <- c(
  "Classics", "French", "English", "Math", "Pitch", "Music"
)
```

e il loro numero

```
ny <- length(varnames)
```

Leggiamo la matrice di correlazione:

```
spearman_cor_mat <- matrix(
  c(
    1.00, .83, .78, .70, .66, .63,
    .83, 1.00, .67, .67, .65, .57,
    .78, .67, 1.00, .64, .54, .51,
    .70, .67, .64, 1.00, .45, .51,
    .66, .65, .54, .45, 1.00, .40,
    .63, .57, .51, .51, .40, 1.00
  ),
  ny, ny,
  byrow = TRUE,
  dimnames = list(varnames, varnames)
)
```

Specifichiamo l'ampiezza campionaria:

```
n <- 33
```

Definiamo il modello unifattoriale con la sintassi di `lavaan`, laddove l'operatore `=~` si può leggere dicendo che la variabile latente a sinistra dell'operatore è definita dalle variabili manifeste elencate a destra dell'operatore e separate dal segno `+`.

```
spearman_mod <- "
  g =~ Classics + French + English + Math + Pitch + Music
"
```

Adattiamo il modello ai dati:

```
fit1 <- lavaan::cfa(
  spearman_mod,
  sample.cov = spearman_cor_mat,
  sample.nobs = n,
  std.lv = TRUE
)
```

Il parametro `std.lv = TRUE` specifica che imponiamo una varianza pari a 1 a tutte le variabili latenti comuni (nel caso presente, solo una). Ciò consente di potere stimare le saturazioni fattoriali.

Possiamo esaminare la soluzione ottenuta con la seguente istruzione:

```
summary(
  fit1,
  fit.measures = TRUE,
  standardized = TRUE
)
#> lavaan 0.6-10 ended normally after 23 iterations
#>
#> Estimator ML
#> Optimization method NLMINB
#> Number of model parameters 12
#>
#> Number of observations 33
#>
#> Model Test User Model:
#>
```

```

#> Test statistic 2.913
#> Degrees of freedom 9
#> P-value (Chi-square) 0.968
#>
#> Model Test Baseline Model:
#>
#> Test statistic 133.625
#> Degrees of freedom 15
#> P-value 0.000
#>
#> User Model versus Baseline Model:
#>
#> Comparative Fit Index (CFI) 1.000
#> Tucker-Lewis Index (TLI) 1.086
#>
#> Loglikelihood and Information Criteria:
#>
#> Loglikelihood user model (H0) -212.547
#> Loglikelihood unrestricted model (H1) -211.091
#>
#> Akaike (AIC) 449.094
#> Bayesian (BIC) 467.052
#> Sample-size adjusted Bayesian (BIC) 429.622
#>
#> Root Mean Square Error of Approximation:
#>
#> RMSEA 0.000
#> 90 Percent confidence interval - lower 0.000
#> 90 Percent confidence interval - upper 0.000
#> P-value RMSEA <= 0.05 0.976
#>
#> Standardized Root Mean Square Residual:
#>
#> SRMR 0.025
#>
#> Parameter Estimates:
#>
#> Standard errors Standard

```

```

#> Information Expected
#> Information saturated (h1) model Structured
#>
#> Latent Variables:
#> Estimate Std.Err z-value P(>|z|)
#> g =~
#> Classics 0.942 0.129 7.314 0.000
#> French 0.857 0.137 6.239 0.000
#> English 0.795 0.143 5.545 0.000
#> Math 0.732 0.149 4.923 0.000
#> Pitch 0.678 0.153 4.438 0.000
#> Music 0.643 0.155 4.142 0.000
#> Std.lv Std.all
#>
#> 0.942 0.956
#> 0.857 0.871
#> 0.795 0.807
#> 0.732 0.743
#> 0.678 0.689
#> 0.643 0.653
#>
#> Variances:
#> Estimate Std.Err z-value P(>|z|)
#> .Classics 0.083 0.051 1.629 0.103
#> .French 0.234 0.072 3.244 0.001
#> .English 0.338 0.094 3.610 0.000
#> .Math 0.434 0.115 3.773 0.000
#> .Pitch 0.510 0.132 3.855 0.000
#> .Music 0.556 0.143 3.893 0.000
#> g 1.000
#> Std.lv Std.all
#> 0.083 0.086
#> 0.234 0.242
#> 0.338 0.349
#> 0.434 0.447
#> 0.510 0.526
#> 0.556 0.573
#> 1.000 1.000

```

È possibile semplificare l'output dalla funzione `summary()` in maniera tale da stampare solo la tabella completa delle stime dei parametri e degli errori standard, ecc.

```
kable(coef(fit1), booktabs = TRUE, format = "markdown")
```

	x
g=~Classics	0.94153
g=~French	0.85749
g=~English	0.79480
g=~Math	0.73208
g=~Pitch	0.67826
g=~Music	0.64325
Classics~~Classics	0.08322
French~~French	0.23441
English~~English	0.33799
Math~~Math	0.43375
Pitch~~Pitch	0.50965
Music~~Music	0.55593

Aniché stampare direttamente la tabella dei parametri, è meglio riformattarla con `kable` quando utilizziamo RMarkdown. Senza usare `kable`, l'output diventa:

```
parameterEstimates(fit1, standardized = TRUE)
#>      lhs op      rhs  est  se    z pvalue
#> 1      g =~ Classics 0.942 0.129 7.314 0.000
#> 2      g =~  French 0.857 0.137 6.239 0.000
#> 3      g =~ English 0.795 0.143 5.545 0.000
#> 4      g =~   Math 0.732 0.149 4.923 0.000
#> 5      g =~   Pitch 0.678 0.153 4.438 0.000
#> 6      g =~   Music 0.643 0.155 4.142 0.000
#> 7 Classics ~~ Classics 0.083 0.051 1.629 0.103
#> 8  French ~~  French 0.234 0.072 3.244 0.001
#> 9 English ~~ English 0.338 0.094 3.610 0.000
#> 10   Math ~~   Math 0.434 0.115 3.773 0.000
#> 11  Pitch ~~   Pitch 0.510 0.132 3.855 0.000
```



```

#> 12    Music ~~    Music 0.556 0.143 3.893 0.000
#> 13      g ~~      g 1.000 0.000    NA    NA
#>      ci.lower ci.upper std.lv std.all std.nox
#> 1      0.689    1.194 0.942 0.956 0.956
#> 2      0.588    1.127 0.857 0.871 0.871
#> 3      0.514    1.076 0.795 0.807 0.807
#> 4      0.441    1.024 0.732 0.743 0.743
#> 5      0.379    0.978 0.678 0.689 0.689
#> 6      0.339    0.948 0.643 0.653 0.653
#> 7     -0.017    0.183 0.083 0.086 0.086
#> 8      0.093    0.376 0.234 0.242 0.242
#> 9      0.154    0.522 0.338 0.349 0.349
#> 10     0.208    0.659 0.434 0.447 0.447
#> 11     0.251    0.769 0.510 0.526 0.526
#> 12     0.276    0.836 0.556 0.573 0.573
#> 13     1.000    1.000 1.000 1.000 1.000

```

Se invece usiamo `kable`, con gli opportuni parametri, otteniamo:

```

parameterEstimates(fit1, standardized = TRUE) %>%
  dplyr::filter(op == "~") %>%
  dplyr::select(
    "Latent Factor" = lhs,
    Indicator = rhs,
    B = est,
    SE = se,
    Z = z,
    "p-value" = pvalue,
    Beta = std.all
  ) %>%
  knitr::kable(
    digits = 3, booktabs = TRUE, format = "markdown",
    caption = "Factor Loadings"
  )

```

Tabella 5.3: *Factor Loadings*

Latent Factor	Indicator	B	SE	Z	p-value	Beta
g	Classics	0.942	0.129	7.314	0	0.956
g	French	0.857	0.137	6.239	0	0.871
g	English	0.795	0.143	5.545	0	0.807
g	Math	0.732	0.149	4.923	0	0.743
g	Pitch	0.678	0.153	4.438	0	0.689
g	Music	0.643	0.155	4.142	0	0.653

Esaminiamo la matrice delle correlazioni residue:

```
cor_table <- residuals(fit1, type = "cor")$cov
knitr::kable(
  cor_table,
  digits = 3,
  format = "markdown",
  booktabs = TRUE
)
```

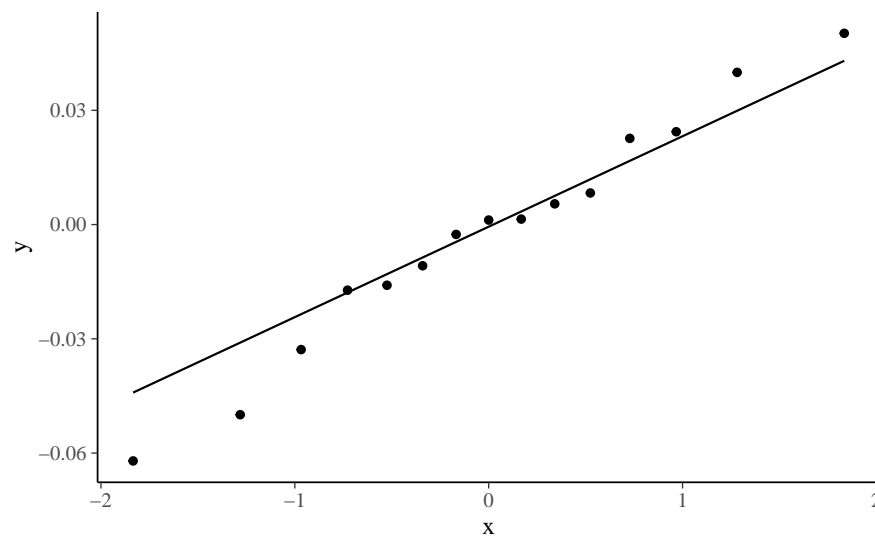
	Classics	French	English	Math	Pitch	Music
Classics	0.000	-0.003	0.008	-0.011	0.001	0.005
French	-0.003	0.000	-0.033	0.023	0.050	0.001
English	0.008	-0.033	0.000	0.040	-0.016	-0.017
Math	-0.011	0.023	0.040	0.000	-0.062	0.024
Pitch	0.001	0.050	-0.016	-0.062	0.000	-0.050
Music	0.005	0.001	-0.017	0.024	-0.050	0.000

Creiamo un qq-plot dei residui:

```
res1 <- residuals(fit1, type = "cor")$cov
res1[upper.tri(res1, diag = TRUE)] <- NA
v1 <- as.vector(res1)
v2 <- v1[!is.na(v1)]

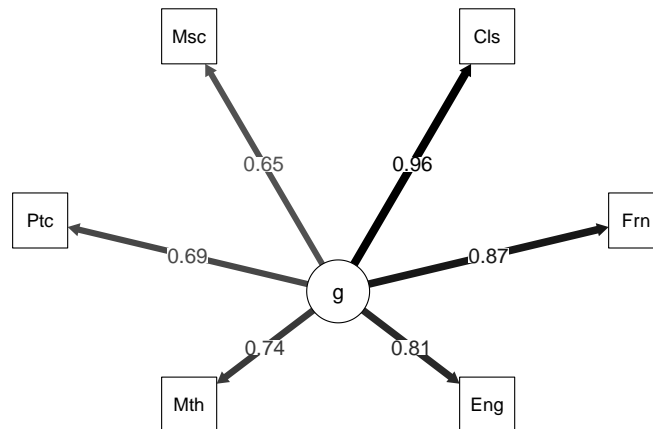
tibble(v2) %>%
```

```
ggplot(aes(sample = v2)) +
  stat_qq() +
  stat_qq_line()
```



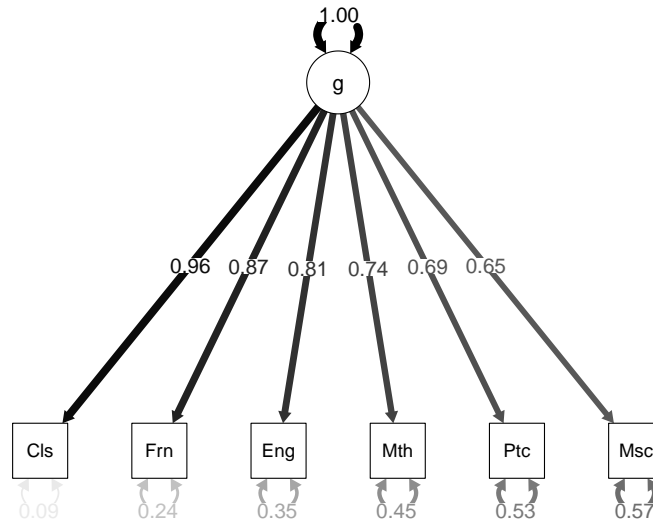
Il pacchetto **semPlot** consente di disegnare diagrammi di percorso per vari modelli SEM. La funzione **semPaths** prende in input un oggetto creato da **lavaan** e disegna il diagramma, con diverse opzioni disponibili. Il diagramma qui prodotto controlla le dimensioni dei caratteri/etichette, la visualizzazione dei residui e il colore dei percorsi/coefficienti. Sono disponibili queste e molte altre opzioni di controllo.

```
semPaths(
  fit1,
  residuals = FALSE,
  sizeMan = 7,
  "std",
  posCol = c("black"),
  edge.label.cex = 1.2,
  layout = "circle2"
)
```



In una versione alternativa del diagramma di percorso aggiungiamo anche le specificità:

```
semPaths(  
  fit1,  
  "std",  
  posCol = c("black"),  
  edge.label.cex = 1.2,  
  sizeMan = 7  
)
```



Il calcolo delle saturazioni fattoriali con il metodo del centroide aveva prodotto il risultato: $\hat{\Lambda} = (0.97, 0.84, 0.73, 0.65)$. Si noti la somiglianza con i valori ottenuti mediante il metodo di massima verosimiglianza.

Conclusioni

Nel presente capitolo abbiamo introdotto il metodo dell'annullamento della tetrade che consente di stimare le saturazioni di un modello monofattoriale. Abbiamo anche visto che il metodo dell'annullamento della tetrade non è altro che un'applicazione della correlazione parziale.

Possiamo dire che un tema cruciale nella costruzione dei test psicologici è quello di stabilire il numero di fattori/tratti che sono soggiacenti all'insieme degli indicatori che vengono considerati. La teoria classica dei test richiede che il test sia monofattoriale, ovvero che gli indicatori considerati siano l'espressione di un unico tratto latente. La violazione della monodimensionalità rende problematica l'applicazione dei principi della teoria classica dei test ai punteggi di un test che non possiede tale proprietà. L'esame della dimensionalità di un gruppo di indicatori rappresenta dunque una fase cruciale nel processo di costruzione di un test e, solitamente, questo esame è affrontato mediante l'analisi fattoriale. In questo capitolo abbiamo presentato le proprietà di base del modello

unifattoriale.

6

Il modello statistico dell'analisi fattoriale

6.1 Modello monofattoriale

Il punto di partenza dell'*analisi fattoriale esplorativa* è rappresentato da una matrice di dimensioni $p \times p$ (dove p è il numero di variabili osservate) che contiene i coefficienti di correlazione (o di covarianza) tra le variabili. Il punto di arrivo è rappresentato da una matrice di dimensioni $p \times k$ (dove k è il numero di fattori comuni che contiene i coefficienti (le *saturazioni*) che esprimono la relazione tra i fattori e le variabili osservate. Considereremo ora il modello matematico dell'analisi fattoriale esplorativa, con un solo fattore comune, che rappresenta il caso più semplice.

Con p variabili manifeste Y_i , il modello ad un fattore comune può essere espresso algebricamente nel modo seguente:

$$Y_i = \mu_i + \lambda_i \xi + \delta_i \quad i = 1, \dots, p$$

dove ξ rappresenta il fattore latente, chiamato anche *fattore comune*, poiché è comune a tutte le Y_i , i δ_i sono invece specifici di ogni variabile osservata e per tale ragione vengono chiamati *fattori specifici* o *unici*, e infine i λ_i sono detti *saturazioni* (o *pesi*) fattoriali poiché consentono di valutare il peso del fattore latente su ciascuna variabile osservata. Si suole assumere per comodità che $\mu = 0$, il che corrisponde a considerare le variabili Y_i come ottenute dagli scarti dalle medie μ_i per $i = 1, \dots, p$:

$$Y_i - \mu_i = \lambda_i \xi + \delta_i.$$

Si assume che il fattore comune abbia media zero, $\mathbb{E}(\xi) = 0$, e varianza unitaria, $\mathbb{V}(\xi) = 1$, che i fattori specifici abbiano media zero, $\mathbb{E}(\delta_j) = 0$, e varianza $\mathbb{V}(\delta_j) = \psi_j$, che i fattori specifici siano incorrelati tra loro,

$\mathbb{E}(\delta_i \delta_k) = 0$, e che i fattori specifici siano incorrelati con il fattore comune, $\mathbb{E}(\delta_i \xi) = 0$.

In questo modello, poiché i fattori specifici sono tra loro incorrelati, l'interdipendenza tra le variabili manifeste è completamente spiegata dal fattore comune. Dalle ipotesi precedenti è possibile ricavare la covarianza tra Y_i e il fattore comune, la varianza della i -esima variabile manifesta Y_i e la covarianza tra due variabili manifeste Y_i e Y_k .

6.2 Covarianza tra un indicatore e il fattore comune

Dal modello monofattoriale è possibile determinare l'espressione della covarianza teorica tra una variabile manifesta Y_i e il fattore comune ξ :

$$\text{Cov}(Y_i, \xi) = \mathbb{E}(Y_i \xi) - \mathbb{E}(Y_i) \mathbb{E}(\xi).$$

Dato che $\mathbb{E}(\xi) = 0$, possiamo scrivere

$$\begin{aligned} \text{Cov}(Y_i, \xi) &= \mathbb{E}(Y_i \xi) = \mathbb{E}[(\lambda_i \xi + \delta_i) \xi] \\ &= \mathbb{E}(\lambda_i \xi^2 + \delta_i \xi) \\ &= \lambda_i \underbrace{\mathbb{E}(\xi^2)}_{\mathbb{V}(\xi)=1} + \underbrace{\mathbb{E}(\delta_i \xi)}_{\text{Cov}(\delta_i, \xi)=0} \\ &= \lambda_i. \end{aligned}$$

Nel modello a un solo fattore, dunque, la saturazione λ_j rappresenta la covarianza la variabile manifesta Y_i e il fattore comune ξ e indica l'importanza del fattore nel determinare il punteggio osservato. Se le variabili Y_i sono standardizzate, la saturazione fattoriale λ_i corrisponde alla correlazione tra Y_i e ξ .

6.3 Espressione fattoriale della varianza

Nell'ipotesi che le variabili Y_i abbiano media nulla, la varianza di Y_i

$$\mathbb{V}(Y_i) = \mathbb{E}(Y_i^2) - [\mathbb{E}(Y_i)]^2 = \mathbb{E}(Y_i^2)$$

\end{equation}

è data da

$$\begin{aligned} \mathbb{V}(Y_i) &= \mathbb{E}[(\lambda_i \xi + \delta_i)^2] \\ &= \lambda_i^2 \underbrace{\mathbb{E}(\xi^2)}_{\mathbb{V}(\xi)=1} + \underbrace{\mathbb{E}(\delta_i^2)}_{\mathbb{V}(\delta_i)=\psi_i} + 2\lambda_i \underbrace{\mathbb{E}(\xi\delta_i)}_{\text{Cov}(\xi,\delta_i)=0} \\ &= \lambda_i^2 + \psi_i. \end{aligned}$$

La quantità λ_i^2 è denominata *comunalità* della i -esima variabile manifesta e corrisponde alla quota della varianza della Y_i spiegata dal fattore comune. Di conseguenza ψ_i è la parte residua della varianza di Y_i non spiegata dal fattore comune ed è denominata *unicità* di Y_i . Nel caso di variabili standardizzate, l'unicità diventa uguale a $\psi_i = 1 - \lambda_i^2$.

In definitiva, la varianza totale di una variabile osservata può essere divisa in una parte che ciascuna variabile condivide con le altre variabili e che è spiegata dal fattore comune (questa parte è chiamata *comunalità* ed è uguale al quadrato della saturazione della variabile osservata nel fattore comune, ovvero $h_i^2 = \lambda_i^2$) e in una parte che è spiegata dal fattore specifico e di errore (questa parte è chiamata *unicità* ed è uguale a $u_i = \psi_i$).

Esercizio 6.1. Riprendiamo l'analisi della matrice di correlazioni di Spearman. Nell'output prodotto dalla funzione `factanal()` viene riportata la quantità denominata **SS loadings**.

Tale quantità indica la porzione della varianza totale delle 4 variabili manifeste che viene spiegata dal fattore comune. Ciascuna variabile standardizzata contribuisce con un'unità di varianza; nel caso presente, dunque la varianza totale sarà uguale a 4. Si ricordi che per *varianza totale* si intende la somma delle varianze delle variabili manifeste, ovvero la *traccia* della matrice di varianze e covarianze. La quota della varianza totale spiegata dal modello è data dalla somma delle comunalità delle quattro variabili, ovvero dalla somma delle saturazioni fattoriali innalzate al quadrato.

```
Spearman <- matrix(c(
  1.0, .78, .70, .66,
```

```

      .78, 1.0, .64, .54,
      .70, .64, 1.0, .45,
      .66, .54, .45, 1.0
    ),
    byrow = TRUE, ncol = 4
  )
rownames(Spearman) <- c("C", "E", "M", "P")
colnames(Spearman) <- c("C", "E", "M", "P")
Spearman
#>      C      E      M      P
#> C 1.00 0.78 0.70 0.66
#> E 0.78 1.00 0.64 0.54
#> M 0.70 0.64 1.00 0.45
#> P 0.66 0.54 0.45 1.00

```

Eseguiamo l'analisi fattoriale:

```

fm <- factanal(covmat = Spearman, factors = 1)
fm
#>
#> Call:
#> factanal(factors = 1, covmat = Spearman)
#>
#> Uniquenesses:
#>      C      E      M      P
#> 0.086 0.329 0.460 0.539
#>
#> Loadings:
#> Factor1
#> C 0.956
#> E 0.819
#> M 0.735
#> P 0.679
#>
#>
#>          Factor1
#> SS loadings    2.587
#> Proportion Var 0.647
#>

```

```
#> The degrees of freedom for the model is 2 and the fit was 0.023
```

Le saturazioni fattoriali sono:

```
L <- c(fm$load[1], fm$load[2], fm$load[3], fm$load[4])
L
#> [1] 0.9563 0.8194 0.7350 0.6790
```

Facendo il prodotto interno otteniamo:

```
t(L) %*% L
#>      [,1]
#> [1,] 2.587
```

In termini proporzionali, la quota della varianza spiegata è uguale a $2.587/4 = 0.647$. Questa quantità è indicata nell'output con la denominazione **Proportion Var**.

Quale quota di varianza delle variabile manifeste viene spiegata dal modello ad un fattore comune? Ciascuna variabile manifesta ha un'unità di varianza. L'unicità (*uniqueness*) indica la proporzione della varianza della variabile considerata che non viene spiegata dalla soluzione fattoriale. La comunaltà è uguale a:

```
round(1 - fm$uniqueness, 3)
#>      C      E      M      P
#> 0.914 0.671 0.540 0.461
```

La componente di varianza non spiegata, o unicità, è invece uguale a:

```
round(fm$uniqueness, 3)
#>      C      E      M      P
#> 0.086 0.329 0.460 0.539
```

6.4 Covarianza tra due variabili manifeste

Nell'ipotesi che le variabili Y_i abbiano media nulla, la covarianza tra Y_i e Y_k

$$\text{Cov}(Y_i, Y_k) = \mathbb{E}(Y_i Y_k) - \mathbb{E}(Y_i)\mathbb{E}(Y_k) = \mathbb{E}(Y_i Y_k)$$

è uguale al prodotto delle corrispondenti saturazioni fattoriali.

$$\begin{aligned} \text{Cov}(Y_i, Y_k) &= \mathbb{E}(Y_i Y_k) = \mathbb{E}[(\lambda_i \xi + \delta_i)(\lambda_k \xi + \delta_k)] \\ &= \mathbb{E}(\lambda_i \lambda_k \xi^2 + \lambda_i \xi \delta_k + \lambda_k \delta_i \xi + \delta_i \delta_k) \\ &= \lambda_i \lambda_k \underbrace{\mathbb{E}(\xi^2)}_{\mathbb{V}(\xi)=1} + \lambda_i \underbrace{\mathbb{E}(\xi \delta_k)}_{\text{Cov}(\xi, \delta_k)=0} + \\ &\quad + \lambda_k \underbrace{\mathbb{E}(\delta_i \xi)}_{\text{Cov}(\delta_i, \xi)=0} + \underbrace{\mathbb{E}(\delta_i \delta_k)}_{\text{Cov}(\delta_i, \delta_k)=0} \\ &= \lambda_i \lambda_k \end{aligned}$$

6.5 Correlazioni osservate e correlazioni riprodotte dal modello

In generale possiamo affermare che il modello monofattoriale è adeguato se si verifica che $\text{Cov}(Y_i, Y_k | \xi) = 0$ ($i, k = 1, \dots, p$; $i \neq k$), ossia se il fattore comune spiega tutta la covarianza tra le variabili osservate. La matrice di correlazioni riprodotte dal modello è chiamata Σ e può essere espressa come:

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

In altri termini, il modello monofattoriale è adeguato se è nulla la differenza tra la matrice di correlazioni osservate e la matrice di correlazioni riprodotte dal modello. Per i dati di Spearman, le correlazioni riprodotte dal modello ad un fattore sono

```
round(L %*% t(L) + diag(fm$uniq), 3)
#>      [,1] [,2] [,3] [,4]
#> [1,] 1.000 0.784 0.703 0.649
#> [2,] 0.784 1.000 0.602 0.556
#> [3,] 0.703 0.602 1.000 0.499
#> [4,] 0.649 0.556 0.499 1.000
```

La matrice delle differenze tra le correlazioni campionarie e quelle riprodotte è

```
round(Spearman - (L %*% t(L) + diag(fm$uniq)), 3)
#>      C      E      M      P
#> C  0.000 -0.004 -0.003  0.011
#> E -0.004  0.000  0.038 -0.016
#> M -0.003  0.038  0.000 -0.049
#> P  0.011 -0.016 -0.049  0.000
```

Lo scarto maggiore tra le correlazioni campionarie e quelle riprodotte è uguale a 0.049. Si può dunque concludere che il modello monofattoriale spiega in maniera ragionevole i dati di Spearman.

6.6 Bontà di adattamento del modello ai dati

La verifica della bontà di adattamento del modello ai dati si determina mediante un test statistico che valuta la differenza tra la matrice di correlazioni (o di covarianze) osservata e la matrice di correlazioni (o covarianze) predetta dal modello fattoriale. L'ipotesi nulla che viene valutata è che la matrice delle correlazioni residue sia dovuta semplicemente agli errori di campionamento, ovvero che la matrice di correlazioni predetta dal modello $\Sigma(\theta)$ sia uguale alla matrice di correlazioni Σ nella popolazione.

La statistica test v è una funzione della differenza tra la matrice riprodotta $S(\theta)$ e quella osservata S

$$v = f[S(\theta) - S]$$

e si distribuisce come una χ^2 con ν gradi di libertà

$$\nu = p(p+1)/2 - q,$$

dove p è il numero di variabili manifeste e q è il numero di parametri stimati dal modello fattoriale (ovvero, λ e ψ). La statistica v assume valore 0 se i parametri del modello riproducono esattamente la matrice di correlazioni tra le variabili nella popolazione. Tanto maggiore è la statistica v tanto maggiore è la discrepanza tra le correlazioni osservate e quelle predette dal modello fattoriale. Un risultato statisticamente significativo (es., $p < .05$) rivela una discrepanza tra il modello e i dati. Il test del modello fattoriale mediante la statistica χ^2 segue dunque una logica diversa da quella utilizzata nei normali test di ipotesi statistiche: *un risultato statisticamente significativo indica una mancanza di adattamento del modello ai dati.*

L'applicazione del test χ^2 per valutare la bontà di adattamento del modello ai dati richiede che ciascuna variabile manifesta sia distribuita normalmente (più precisamente, richiede che le variabili manifeste siano un campione casuale che deriva da una normale multivariata).

Il limite principale della statistica v è che essa dipende dalle dimensioni del campione: al crescere delle dimensioni campionarie è più facile ottenere un risultato statisticamente significativo (ovvero, concludere che vi è un cattivo adattamento del modello ai dati). Per questa ragione, la bontà di adattamento del modello ai dati viene valutata da molteplici indici, non soltanto dalla statistica v .

6.7 L'errore standard della misurazione e il modello fattoriale

Per concludere, prendiamo nuovamente in esame la nozione dell'errore standard della misurazione, uno dei concetti centrali della CTT, e vediamo come tale concetto possa essere "ripensato" nel contesto del modello statistico dell'analisi fattoriale. Iniziamo con una dimostrazione.

Dimostrazione. Secondo la CTT, il punteggio X ottenuto dalla somministrazione del test è uguale a $X = T + E$, dove E è una variabile

aleatorie indipendente da T . Se consideriamo il rispondente i -esimo, il modello diventa $X_i = T_i + E_i$, dove T_i è il valore vero ed E_i è una variabile aleatoria con media 0.

Riscriviamo ora questa equazione nei termini di un modello monofattoriale con p variabili manifeste (item). Per ciascun item avremo:

$$\begin{aligned} Y_{1i} &= \lambda_1 \xi_i + \delta_{1i} \\ Y_{2i} &= \lambda_2 \xi_i + \delta_{2i} \\ &\dots \\ Y_{pi} &= \lambda_p \xi_i + \delta_{pi} \end{aligned}$$

Il punteggio totale X_i per il rispondente i -esimo è dato dalla somma dei punteggi osservati in ciascun item, ovvero

$$\begin{aligned} X_i &= \sum_{j=1}^p Y_{ji} = \sum_{j=1}^p \lambda_j \xi_i + \sum_{j=1}^p \delta_{ji} \\ &= \left(\sum_{j=1}^p \lambda_j \right) \xi_i + \sum_{j=1}^p \delta_{ji} \\ &= T_i + E_i \end{aligned}$$

Secondo la CTT, la varianza del punteggio osservato X_i si scompone in due componenti: $\sigma_{X_i}^2 = \sigma_{T_i}^2 + \sigma_{E_i}^2$. Nei termini del modello fattoriale, la varianza della componente vera del punteggio totale del test, $\sigma_{T_i}^2$, è data dal quadrato della somma delle saturazioni fattoriali:

$$\begin{aligned} \sigma_{T_i}^2 &= \mathbb{V} \left[\left(\sum_{j=1}^p \lambda_j \right) \xi_i \right] \\ &= \left(\sum_{j=1}^p \lambda_j \right)^2 \mathbb{V}(\xi_i) \\ &= \left(\sum_{j=1}^p \lambda_j \right)^2 \end{aligned}$$

Nei termini del modello fattoriale, se consideriamo il punteggio totale del test, la varianza della componente dell'errore della misurazione, $\sigma_{E_i}^2$, è data dalla somma delle unicità:

$$\begin{aligned}\sigma_{E_i}^2 &= \mathbb{V} \left(\sum_{j=1}^p \delta_{ji} \right) \\ &= \sum_{j=1}^p \mathbb{V}(\delta_{ji}) \\ &= \sum_{j=1}^p \Psi_j\end{aligned}$$

Nei termini del modello fattoriale, dunque, una stima dell'errore standard della misurazione del punteggio totale del test è data dalla radice quadrata della quantità precedente, ovvero:

$$\sigma_E = \sqrt{\sum_{j=1}^p \Psi_j} \quad (6.1)$$

□

6.7.1 Un caso concreto

Applichiamo ora il risultato precedente ad un caso concreto. Consideriamo i dati utilizzati nella validazione italiana del *Cognitive Style Questionnaire - Short Form* (CSQ-SF, Meins et al. 2012). Il CSQ-SF viene utilizzato per misurare la vulnerabilità all'ansia e alla depressione. È costituito da cinque sottoscale: *Internality*, *Globality*, *Stability*, *Negative consequences* e *Self-worth*.

Leggiamo i dati in R:

```
csq <- rio::import(here::here("data", "csq540.csv"))
```

Il numero di partecipanti è

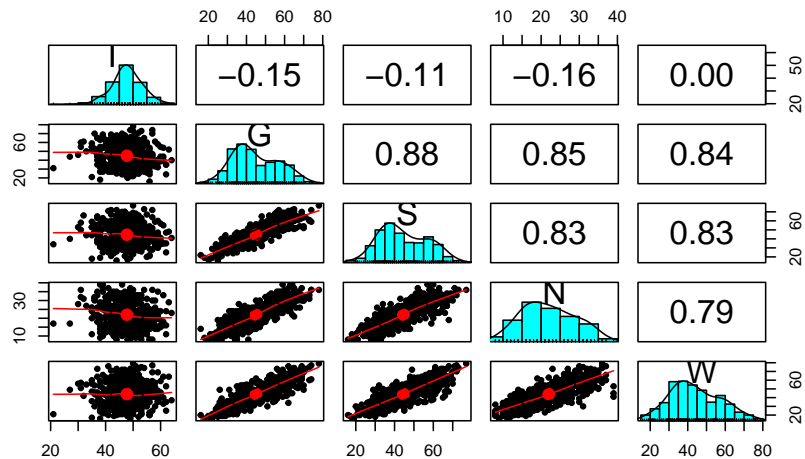
```
n <- nrow(csq)
n
#> [1] 540
```


Le statistiche descrittive si ottengono con la seguente istruzione:

```
psych::describe(csq, type = 2)
#>   vars  n mean   sd median trimmed  mad min max
#> I    1 540 47.76  5.78    48  47.87  4.45  21  64
#> G    2 540 45.00 11.94    42  44.55 11.86  16  78
#> S    3 540 44.60 12.18    42  44.24 13.34  16  77
#> N    4 540 22.01  6.92    21  21.86  7.41   8  39
#> W    5 540 44.05 13.10    43  43.66 13.34  16  79
#>   range skew kurtosis  se
#> I    43 -0.31    1.07 0.25
#> G    62  0.34   -0.70 0.51
#> S    61  0.27   -0.77 0.52
#> N    31  0.21   -0.74 0.30
#> W    63  0.31   -0.53 0.56
```

Esaminiamo la matrice di correlazione:

```
psych::pairs.panels(csq)
```



Specifichiamo il modello unifattoriale nella sintassi di `lavaan`:

```
mod_csq <- "
  F =~ NA*I + G + S + N + W
  F =~ 1*F
"
```

Adattiamo il modello ai dati:

```
fit <- lavaan::cfa(
  mod_csq,
  data = csq
)
```

Esaminiamo i risultati:

```
summary(
  fit,
  standardized = TRUE,
  fit.measures = TRUE
)
#> lavaan 0.6-10 ended normally after 26 iterations
#>
#> Estimator ML
#> Optimization method NLMINB
#> Number of model parameters 10
#>
#> Number of observations 540
#>
#> Model Test User Model:
#>
#> Test statistic 46.716
#> Degrees of freedom 5
#> P-value (Chi-square) 0.000
#>
#> Model Test Baseline Model:
#>
#> Test statistic 2361.816
#> Degrees of freedom 10
#> P-value 0.000
```

```

#>
#> User Model versus Baseline Model:
#>
#>    Comparative Fit Index (CFI)                0.982
#>    Tucker-Lewis Index (TLI)                  0.965
#>
#> Loglikelihood and Information Criteria:
#>
#>    Loglikelihood user model (H0)            -8741.781
#>    Loglikelihood unrestricted model (H1)     -8718.423
#>
#>    Akaike (AIC)                             17503.562
#>    Bayesian (BIC)                           17546.478
#>    Sample-size adjusted Bayesian (BIC)      17514.734
#>
#> Root Mean Square Error of Approximation:
#>
#>    RMSEA                                     0.124
#>    90 Percent confidence interval - lower    0.093
#>    90 Percent confidence interval - upper    0.158
#>    P-value RMSEA <= 0.05                   0.000
#>
#> Standardized Root Mean Square Residual:
#>
#>    SRMR                                     0.033
#>
#> Parameter Estimates:
#>
#>    Standard errors                          Standard
#>    Information                             Expected
#>    Information saturated (h1) model         Structured
#>
#> Latent Variables:
#>
#>           Estimate  Std.Err  z-value  P(>|z|)
#>    F =~
#>      I             0.725    0.253    2.867    0.004
#>      G            -11.322    0.384   -29.481    0.000
#>      S            -11.342    0.398   -28.513    0.000

```

```

#>      N           -6.163    0.233 -26.398    0.000
#>      W           -11.598    0.444 -26.137    0.000
#>      Std.lv  Std.all
#>
#>      0.725    0.126
#>     -11.322   -0.949
#>     -11.342   -0.932
#>      -6.163   -0.891
#>     -11.598   -0.886
#>
#> Variances:
#>               Estimate Std.Err  z-value  P(>|z|)
#>      F               1.000
#>      .I             32.840    2.000   16.420    0.000
#>      .G             14.038    1.473    9.532    0.000
#>      .S             19.508    1.718   11.353    0.000
#>      .N              9.847    0.725   13.573    0.000
#>      .W             36.892    2.685   13.737    0.000
#>      Std.lv  Std.all
#>      1.000    1.000
#>     32.840    0.984
#>     14.038    0.099
#>     19.508    0.132
#>      9.847    0.206
#>     36.892    0.215

```

Recuperiamo le specificità:

```

psi <- parameterEstimates(fit)$est[7:11]
psi
#> [1] 32.840 14.038 19.508 9.847 36.892

```

Stimiamo l'errore standard della misurazione con la (6.1):

```

sqrt(sum(psi))
#> [1] 10.64

```

Applichiamo ora la formula della TCT:

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX'}}.$$

Per trovare σ calcoliamo prima il punteggio totale:

```
tot_score <- rowSums(csq)
```

La deviazione standard di `tot_score` ci fornisce una stima di σ :

```
sigma <- sd(tot_score)
sigma
#> [1] 41.26
```

Per applicare la formula della TCT abbiamo bisogno dell'attendibilità. La stimiamo usando la funzione `reliability` del pacchetto `semTools` dall'oggetto creato da `lavaan::cfa()`:

```
rel <- semTools::reliability(fit)
rel
#>           F
#> alpha  0.8507
#> omega  0.9330
#> omega2 0.9330
#> omega3 0.9273
#> avevar 0.7917
```

Utilizzando Ω otteniamo:

```
sigma * sqrt(1 - rel[2])
#> [1] 10.68
```



7

Il modello multifattoriale

7.1 Modello multifattoriale: fattori ortogonali

La teoria dei due fattori ha orientato per diversi anni le ricerche sull'intelligenza, finché Thurstone (1945) non propose una sua modifica, conosciuta come teoria multifattoriale. Secondo Thurstone la covarianza tra le variabili manifeste non può essere spiegata da un unico fattore generale. Invece è necessario ipotizzare l'azione causale di diversi fattori, definiti comuni, i quali si riferiscono solo ad alcune delle variabili considerate.

Il modello plurifattoriale assume che ciascuna variabile manifesta sia espressa come funzione lineare di un certo numero m di fattori comuni, $\xi_1, \xi_2, \dots, \xi_m$, responsabili della correlazione con le altre variabili, e di un solo fattore specifico (termine d'errore), responsabile della variabilità della variabile stessa. Per p variabili manifeste, Y_1, Y_2, \dots, Y_p , il modello fattoriale diventa quello indicato dal sistema di equazioni lineari descritto di seguito. Idealmente, m dovrebbe essere molto più piccolo di p così da consentire una descrizione parsimoniosa delle variabili manifeste in funzione di pochi fattori soggiacenti.

Le variabili manifeste Y sono indicizzate da $i = 1, \dots, p$. Le variabili latenti ξ (fattori) sono indicizzate da $j = 1, \dots, m$. I fattori specifici δ sono indicizzati da $i = 1, \dots, p$. Le saturazioni fattoriali si distinguono dunque tramite due indici, i e j : il primo indice si riferisce alle variabili manifeste, il secondo si riferisce ai fattori latenti.

Indichiamo con μ_i , con $i = 1, \dots, p$ le medie delle p variabili manifeste Y_1, Y_2, \dots, Y_p . Se non vi è alcun effetto delle variabili comuni latenti, allora la variabile Y_{ijk} , dove k è l'indice usato per i soggetti, sarà uguale a:

$$\begin{cases} Y_{1k} &= \mu_1 + \delta_{1k} \\ &\vdots \\ Y_{ik} &= \mu_i + \delta_{ik} \\ &\vdots \\ Y_{pk} &= \mu_p + \delta_{pk} \end{cases}$$

Se invece le variabili manifeste rappresentano la somma dell'effetto causale di m fattori comuni e di p fattori specifici, allora possiamo scrivere:

$$\begin{cases} Y_1 - \mu_1 &= \lambda_{11}\xi_1 + \dots + \lambda_{1k}\xi_k \dots + \lambda_{1m}\xi_m + \delta_1 \\ &\vdots \\ Y_i - \mu_i &= \lambda_{i1}\xi_1 + \dots + \lambda_{ik}\xi_k \dots + \lambda_{im}\xi_m + \delta_i \\ &\vdots \\ Y_p - \mu_p &= \lambda_{p1}\xi_1 + \dots + \lambda_{pk}\xi_k \dots + \lambda_{pm}\xi_m + \delta_p \end{cases}$$

Nel precedente sistema di equazioni lineari,

- ξ_j , con $j = 1, \dots, m$, rappresenta la j -esima variabile inosservabile a fattore comune (ossia il j -esimo fattore comune a tutte le variabili Y_i);
- λ_{ij} rappresenta il parametro, detto *saturazione* o *peso* fattoriale, che riflette l'importanza del j -esimo fattore comune nella composizione della i -esima variabile osservabile;
- δ_i rappresenta il fattore specifico (o unico) di ogni variabile manifesta Y_i .

In conclusione, secondo il modello multifattoriale, le variabili manifeste Y_i , con $i = 1, \dots, p$, sono il risultato di una *combinazione lineare* di $m < p$ fattori inosservabili ad esse comuni ξ_j , con $j = 1, \dots, m$, e di p fattori specifici δ_i , con $i = 1, \dots, p$, anch'essi inosservabili e di natura residua.

7.1.1 Assunzioni del modello multifattoriale

Le variabili inosservabili a fattore comune ξ_j , con $j = 1, \dots, m$, in quanto latenti, non possiedono unità di misura. Pertanto, per semplicità si assume che abbiano media zero, $\mathbb{E}(\xi_j) = 0$, abbiano varianza unitaria, $\mathbb{V}(\xi_j) = \mathbb{E}(\xi_j^2) - [\mathbb{E}(\xi_j)]^2 = 1$, e siano incorrelate tra loro, $\text{Cov}(\xi_j, \xi_h) = 0$, con $j, h = 1, \dots, m$; $j \neq h$. Si assume inoltre che le variabili a fattore specifico δ_i siano tra loro incorrelate, $\text{Cov}(\delta_i, \delta_k) = 0$,

con $i, k = 1, \dots, p$, $i \neq k$, abbiano media zero, $\mathbb{E}(\delta_i) = 0$, e varianza uguale a $\mathbb{V}(\delta_i) = \psi_{ii}$. La varianza ψ_{ii} è detta *varianza specifica* o *unicità* della i -esima variabile manifesta Y_i . Si assume infine che i fattori specifici siano linearmente incorrelati con i fattori comuni, ovvero $\text{Cov}(\xi_j, \delta_i) = 0$ per ogni $j = 1, \dots, m$ e per ogni $i = 1, \dots, p$.

7.1.2 Interpretazione dei parametri del modello

Quale esempio, consideriamo il caso di $p = 5$ variabili osservabili e $m = 2$ fattori ortogonali. Se le variabili manifeste sono ‘centrate’ (ovvero, se a ciascuna di esse sottraiamo la rispettiva media), allora il modello multifattoriale diventa +

$$\begin{aligned} Y_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1, \\ Y_2 &= \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \delta_2, \\ Y_3 &= \lambda_{31}\xi_1 + \lambda_{32}\xi_2 + \delta_3, \\ Y_4 &= \lambda_{41}\xi_1 + \lambda_{42}\xi_2 + \delta_4, \\ Y_5 &= \lambda_{51}\xi_1 + \lambda_{52}\xi_2 + \delta_5. \end{aligned}$$

7.1.3 Covarianza tra variabili e fattori

Nell’ipotesi che le variabili Y_i abbiano media nulla, la covarianza tra Y_i e ξ_j è uguale alla saturazione fattoriale λ_{ij} :

$$\begin{aligned} \text{Cov}(Y_i, \xi_j) &= \mathbb{E}(Y_i \xi_j) \\ &= \mathbb{E}[(\lambda_{i1}\xi_1 + \dots + \lambda_{im}\xi_m + \delta_i)\xi_j] \\ &= \lambda_{i1} \underbrace{\mathbb{E}(\xi_1 \xi_j)}_{=0} + \dots + \lambda_{ij} \underbrace{\mathbb{E}(\xi_j^2)}_{=1} + \dots \\ &\quad + \lambda_{im} \underbrace{\mathbb{E}(\xi_m \xi_j)}_{=0} + \underbrace{\mathbb{E}(\delta_i \xi_j)}_{=0} \\ &= \lambda_{ij}. \end{aligned}$$

Anche nel modello multifattoriale, dunque, le saturazioni fattoriali rappresentano le covarianze tra le variabili e i fattori:

$$\text{Cov}(Y_i, \xi_j) = \lambda_{ij} \quad i = 1, \dots, p; \quad j = 1, \dots, m.$$

Naturalmente, se le variabili sono standardizzate, le saturazioni fattoriali diventano correlazioni:

$$r_{ij} = \lambda_{ij}.$$

7.1.4 Espressione fattoriale della varianza

Come nel modello monofattoriale, la varianza delle variabili manifeste si decompone in una componente dovuta ai fattori comuni, chiamata *comunalità*, e in una componente specifica alle Y_i , chiamata *unicità*. Nell'ipotesi che le variabili Y_i abbiano media nulla, la varianza di Y_i è uguale a

$$\mathbb{V}(Y_i) = \mathbb{E}[(\lambda_{i1}\xi_1 + \dots + \lambda_{im}\xi_m + \delta_i)^2]. \quad (7.1)$$

Come si sviluppa il polinomio precedente? Il quadrato di un polinomio è uguale alla somma dei quadrati di tutti i termini più il doppio prodotto di ogni termine per ciascuno di quelli che lo seguono. Il valore atteso del quadrato del primo termine è uguale a $\lambda_{i1}^2 \mathbb{E}(\xi_1^2)$ ma, essendo la varianza di ξ_1 uguale a 1, otteniamo semplicemente λ_{i1}^2 . Lo stesso vale per i quadrati di tutti i termini seguenti tranne l'ultimo. Infatti, $\mathbb{E}(\delta_i^2) = \psi_{ii}$. Per quel che riguarda i doppi prodotti, sono tutti nulli. In primo luogo perché, nel caso di fattori ortogonali, la covarianza tra i fattori comuni è nulla, $\mathbb{E}(\xi_j \xi_h) = 0$, con $j \neq h$. In secondo luogo perché i fattori comuni sono incorrelati con i fattori specifici, quindi $\mathbb{E}(\delta_i \xi_j) = 0$.

In conclusione,

$$\begin{aligned} \mathbb{V}(Y_i) &= \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2 + \psi_{ii} \\ &= \sum_{j=1}^m \lambda_{ij}^2 + \psi_{ii} \\ &= h_i^2 + \psi_{ii} \\ &= \text{comunalità} + \text{unicità}, \end{aligned}$$

la varianza della variabile manifesta Y_i è suddivisa in due parti: il primo addendo è definito comunalità poiché rappresenta la parte di variabilità della Y_i spiegata dai fattori comuni; il secondo addendo è invece definito varianza specifica (o unicità) poiché esprime la parte di variabilità della Y_i non spiegata dai fattori comuni.

7.1.5 Espressione fattoriale della covarianza

Per semplificare, consideriamo il caso particolare esaminato prima, ovvero quello con $p = 5$ variabili osservabili e $m = 2$ fattori ortogonali. Nell'ipotesi che le variabili Y_i abbiano media nulla, la covarianza tra Y_1 e Y_2 , ad esempio, è uguale a:

$$\begin{aligned}\text{Cov}(Y_1, Y_2) &= \mathbb{E}(Y_1 Y_2) \\ &= \mathbb{E}[(\lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1)(\lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \delta_2)] \\ &= \lambda_{11}\lambda_{21}\mathbb{E}(\xi_1^2) + \lambda_{11}\lambda_{22}\mathbb{E}(\xi_1\xi_2) + \lambda_{11}\mathbb{E}(\xi_1\delta_2) + \\ &\quad \lambda_{12}\lambda_{21}\mathbb{E}(\xi_1\xi_2) + \lambda_{12}\lambda_{22}\mathbb{E}(\xi_2^2) + \lambda_{12}\mathbb{E}(\xi_2\delta_2) + \\ &\quad \lambda_{21}\mathbb{E}(\xi_1\delta_1) + \lambda_{22}\mathbb{E}(\xi_2\delta_1) + \mathbb{E}(\delta_1\delta_2) \\ &= \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22}.\end{aligned}$$

In conclusione, la covarianza tra le variabili manifeste Y_l e Y_m riprodotta dal modello è data dalla somma dei prodotti delle saturazioni $\lambda_l\lambda_m$ nei due fattori.

Esercizio 7.1. Consideriamo i dati riportati da [Brown \(2015\)](#), ovvero otto misure di personalità raccolte su un campione di 250 pazienti che hanno concluso un programma di psicoterapia. Le scale sono le seguenti:

- anxiety (N1),
- hostility (N2),
- depression (N3),
- self-consciousness (N4),
- warmth (E1),
- gregariousness (E2),
- assertiveness (E3),
- positive emotions (E4).

```
varnames <- c("N1", "N2", "N3", "N4", "E1", "E2", "E3", "E4")
sds <- "5.7 5.6 6.4 5.7 6.0 6.2 5.7 5.6"

cors <- "
1.000
0.767 1.000
0.731 0.709 1.000
0.778 0.738 0.762 1.000"
```

```
-0.351 -0.302 -0.356 -0.318 1.000
-0.316 -0.280 -0.300 -0.267 0.675 1.000
-0.296 -0.289 -0.297 -0.296 0.634 0.651 1.000
-0.282 -0.254 -0.292 -0.245 0.534 0.593 0.566 1.000"
```

```
psychot_cor_mat <- getCov(cors, names = varnames)
n <- 250
```

Eseguiamo l'analisi fattoriale esplorativa con il metodo della massima verosimiglianza ipotizzando due fattori comuni incorrelati:

```
n_facs <- 2
fit_efa <- factanal(
  covmat = psychot_cor_mat,
  factors = n_facs,
  rotation = "varimax",
  n.obs = n
)
```

Esaminiamo le saturazioni fattoriali:

```
lambda <- fit_efa$loadings
lambda
#>
#> Loadings:
#>   Factor1 Factor2
#> N1  0.854 -0.228
#> N2  0.826 -0.194
#> N3  0.811 -0.233
#> N4  0.865 -0.186
#> E1 -0.202  0.773
#> E2 -0.139  0.829
#> E3 -0.158  0.771
#> E4 -0.147  0.684
#>
#>               Factor1 Factor2
#> SS loadings      2.923   2.526
```

```
#> Proportion Var    0.365    0.316
#> Cumulative Var    0.365    0.681
```

La soluzione fattoriale conferma la presenza di due fattori: il primo fattore satura sulle scale di neoturisticismo, il secondo sulle scale di estroversione.

La correlazione riprodotta r_{12} è uguale a $\lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22}$

```
lambda[1, 1] * lambda[2, 1] + lambda[1, 2] * lambda[2, 2]
#> [1] 0.7493
```

e corrisponde da vicino alla correlazione osservata 0.767.

L'intera matrice di correlazioni riprodotte è $\Lambda\Lambda^T + \psi$:

```
Rr <- lambda %*% t(lambda) + diag(fit_efa$uniq)
Rr %>%
  round(3)
#>      N1      N2      N3      N4      E1      E2      E3
#> N1  1.000  0.749  0.745  0.781 -0.348 -0.307 -0.311
#> N2  0.749  1.000  0.715  0.751 -0.317 -0.276 -0.281
#> N3  0.745  0.715  1.000  0.745 -0.344 -0.306 -0.308
#> N4  0.781  0.751  0.745  1.000 -0.318 -0.274 -0.280
#> E1 -0.348 -0.317 -0.344 -0.318  1.000  0.669  0.628
#> E2 -0.307 -0.276 -0.306 -0.274  0.669  1.000  0.661
#> E3 -0.311 -0.281 -0.308 -0.280  0.628  0.661  1.000
#> E4 -0.281 -0.254 -0.279 -0.254  0.558  0.587  0.550
#>      E4
#> N1 -0.281
#> N2 -0.254
#> N3 -0.279
#> N4 -0.254
#> E1  0.558
#> E2  0.587
#> E3  0.550
#> E4  1.000
```

La differenza tra la matrice di correlazioni riprodotte e la matrice di correlazioni osservate è uguale a:

```
(psychot_cor_mat - Rr) %>%
  round(3)
#>      N1      N2      N3      N4      E1      E2      E3
#> N1  0.000  0.018 -0.014 -0.003 -0.003 -0.009  0.015
#> N2  0.018  0.000 -0.006 -0.013  0.015 -0.004 -0.008
#> N3 -0.014 -0.006  0.000  0.017 -0.012  0.006  0.011
#> N4 -0.003 -0.013  0.017  0.000  0.000  0.007 -0.016
#> E1 -0.003  0.015 -0.012  0.000  0.000  0.006  0.006
#> E2 -0.009 -0.004  0.006  0.007  0.006  0.000 -0.010
#> E3  0.015 -0.008  0.011 -0.016  0.006 -0.010  0.000
#> E4 -0.001  0.000 -0.013  0.009 -0.024  0.006  0.016
#>      E4
#> N1 -0.001
#> N2  0.000
#> N3 -0.013
#> N4  0.009
#> E1 -0.024
#> E2  0.006
#> E3  0.016
#> E4  0.000
```

Esercizio 7.2. Consideriamo nuovamente i dati precedenti ma, in questo caso, eseguiamo un'analisi fattoriale confermativa. Usando `lavaan` il modello diventa:

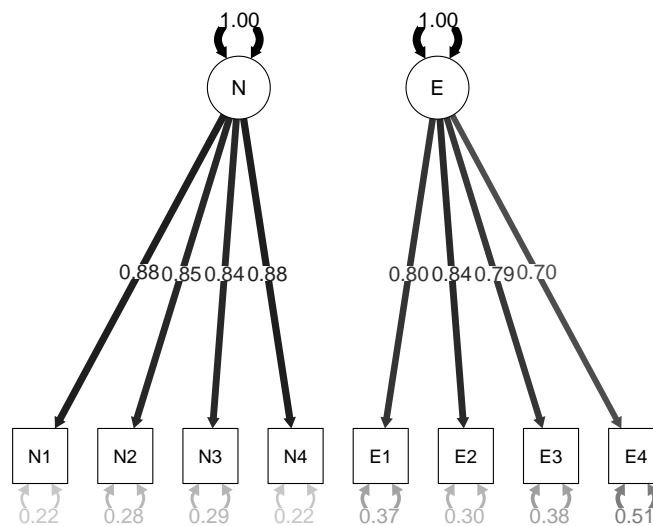
```
cfa_mod <- "
  N =~ N1 + N2 + N3 + N4
  E =~ E1 + E2 + E3 + E4
"
```

```
fit_cfa <- lavaan::cfa(
  cfa_mod,
  sample.cov = psychot_cor_mat,
  sample.nobs = n,
  orthogonal = TRUE,
```

```
std.lv = TRUE
)
```

Il path diagram si ottiene nel modo seguente:

```
semPaths(
  fit_cfa,
  "std",
  posCol = c("black"),
  edge.label.cex = 1.2,
  sizeMan = 7
)
```



Esaminiamo le saturazioni fattoriali:

```
parameterEstimates(fit_cfa, standardized = TRUE) %>%
  dplyr::filter(op == "~") %>%
  dplyr::select(
    "Latent Factor" = lhs,
    Indicator = rhs,
    B = est,
```

```

SE = se,
Z = z,
"p-value" = pvalue,
Beta = std.all
) %>%
knitr::kable(
  digits = 3, booktabs = TRUE, format = "markdown",
  caption = "Factor Loadings"
)

```

Tabella 7.1: *Factor Loadings*

Latent Factor	Indicator	B	SE	Z	p-value	Beta
N	N1	0.882	0.051	17.42	0	0.884
N	N2	0.847	0.052	16.34	0	0.849
N	N3	0.840	0.052	16.13	0	0.842
N	N4	0.882	0.051	17.43	0	0.884
E	E1	0.795	0.056	14.28	0	0.796
E	E2	0.838	0.054	15.37	0	0.839
E	E3	0.788	0.056	14.10	0	0.789
E	E4	0.697	0.058	11.94	0	0.699

Il risultato sembra sensato: le saturazioni su ciascun fattore sono molto alte. Tuttavia, la matrice delle correlazioni residue

```

cor_table <- residuals(fit_cfa, type = "cor")$cov
knitr::kable(
  cor_table,
  digits = 3,
  format = "markdown",
  booktabs = TRUE
)

```

	N1	N2	N3	N4	E1	E2	E3	E4
N1	0.000	0.017	-0.013	-0.003	-0.351	-0.316	-0.296	-0.282
N2	0.017	0.000	-0.006	-0.012	-0.302	-0.280	-0.289	-0.254

	N1	N2	N3	N4	E1	E2	E3	E4
N3	-0.013	-0.006	0.000	0.018	-0.356	-0.300	-0.297	-0.292
N4	-0.003	-0.012	0.018	0.000	-0.318	-0.267	-0.296	-0.245
E1	-0.351	-0.302	-0.356	-0.318	0.000	0.007	0.006	-0.022
E2	-0.316	-0.280	-0.300	-0.267	0.007	0.000	-0.011	0.007
E3	-0.296	-0.289	-0.297	-0.296	0.006	-0.011	0.000	0.015
E4	-0.282	-0.254	-0.292	-0.245	-0.022	0.007	0.015	0.000

rivela che il modello ipotizzato dall'analisi fattoriale confermativa non è adeguato.

7.2 Modello fattoriale: Fattori obliqui

Anche nel caso di fattori comuni correlati è possibile esprimere nei termini dei parametri del modello la covarianza teorica tra una variabile manifesta Y_i e uno dei fattori comuni, la covarianza teorica tra due variabili manifeste, e la comunaltà di ciascuna variabile manifesta. Dato però che i fattori comuni risultano correlati, l'espressione fattoriale di tali quantità è più complessa che nel caso di fattori comuni ortogonali.

7.2.1 Covarianza teorica tra variabili e fattori

In base al modello multifattoriale con m fattori comuni la variabile Y_i è

$$Y_i = \lambda_{i1}\xi_1 + \dots + \lambda_{im}\xi_m + \delta_i.$$

Poniamoci il problema di trovare la covarianza teorica tra la variabile manifesta Y_i e il fattore comune ξ_j . Come in precedenza, il problema si riduce a quello di trovare $\mathbb{E}(Y_i\xi_j)$. Ne segue che

$$\begin{aligned}
\text{Cov}(Y_i, \xi_j) &= \mathbb{E}(Y_i \xi_j) \\
&= \mathbb{E}[(\lambda_{i1}\xi_1 + \dots + \lambda_{ij}\xi_j + \dots + \lambda_{im}\xi_m + \delta_i)\xi_j] \\
&= \lambda_{i1} \underbrace{\mathbb{E}(\xi_1 \xi_j)}_{\neq 0} + \dots + \lambda_{ij} \underbrace{\mathbb{E}(\xi_j^2)}_{=1} + \dots \\
&\quad + \lambda_{im} \underbrace{\mathbb{E}(\xi_m \xi_j)}_{\neq 0} + \underbrace{\mathbb{E}(\delta_i \xi_j)}_{=0} \\
&= \lambda_{ij} + \lambda_{i1} \text{Cov}(\xi_1, \xi_j) + \dots + \lambda_{im} \text{Cov}(\xi_m, \xi_j).
\end{aligned}$$

Ad esempio, nel caso di tre fattori comuni ξ_1, ξ_2, ξ_3 , la covarianza tra Y_1 e ξ_1 diventa

$$\lambda_{11} + \lambda_{12} \text{Cov}(\xi_1, \xi_2) + \lambda_{13} \text{Cov}(\xi_1, \xi_3).$$

7.2.2 Espressione fattoriale della varianza

Poniamoci ora il problema di trovare la varianza teorica della variabile manifesta Y_i . In base al modello fattoriale, la variabile Y_i è specificata come nella (7.2.1). La varianza di Y_i è $\mathbb{V}(Y_i) = \mathbb{E}(Y_i^2) - [\mathbb{E}(Y_i)]^2$. Però, avendo espresso Y_i nei termini della differenza dalla sua media, l'espressione della varianza si riduce a $\mathbb{V}(Y_i) = \mathbb{E}(Y_i^2)$. Dobbiamo dunque sviluppare l'espressione

$$\mathbb{E}(Y_i^2) = \mathbb{E}[(\lambda_{i1}\xi_1 + \dots + \lambda_{im}\xi_m + \delta_i)^2].$$

In conclusione, la varianza teorica di Y_i è uguale a

$$\begin{aligned}
\mathbb{V}(Y_i) &= \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2 + \\
&\quad 2\lambda_{i1}\lambda_{i2}\text{Cov}(\xi_1, \xi_2) + \dots + 2\lambda_{i,m-1}\lambda_{im}\text{Cov}(\xi_{m-1}, \xi_m) + \\
&\quad \psi_{ii}.
\end{aligned}$$

Ad esempio, nel caso di tre fattori comuni, ξ_1, ξ_2, ξ_3 , la varianza di Y_1 è

$$\begin{aligned}
\mathbb{V}(Y_1) &= \lambda_{11}^2 + \lambda_{12}^2 + \lambda_{13}^2 + \\
&\quad 2\lambda_{11}\lambda_{12}\text{Cov}(\xi_1, \xi_2) + \\
&\quad 2\lambda_{11}\lambda_{13}\text{Cov}(\xi_1, \xi_3) + \\
&\quad 2\lambda_{12}\lambda_{13}\text{Cov}(\xi_2, \xi_3) + \\
&\quad \psi_{11}.
\end{aligned}$$

7.2.3 Covarianza teorica tra due variabili

Consideriamo ora il caso più semplice di due soli fattori comuni correlati e calcoliamo la covarianza tra Y_1 e Y_2 :

$$\begin{aligned}\mathbb{E}(Y_1 Y_2) &= \mathbb{E}[(\lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1)(\lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \delta_2)] \\ &= \mathbb{E}(\lambda_{11}\lambda_{21}\xi_1^2 + \lambda_{11}\lambda_{22}\xi_1\xi_2 + \lambda_{11}\xi_1\delta_2 + \\ &\quad + \lambda_{12}\lambda_{21}\xi_1\xi_2 + \lambda_{12}\lambda_{22}\xi_2^2 + \lambda_{12}\xi_2\delta_2 + \\ &\quad + \lambda_{21}\xi_1\delta_1 + \lambda_{22}\xi_2\delta_1 + \delta_1\delta_2).\end{aligned}$$

Distribuendo l'operatore di valore atteso, dato che $\mathbb{E}(\xi^2) = 1$ e $\mathbb{E}(\xi\delta) = 0$, otteniamo

$$\text{Cov}(Y_1, Y_2) = \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} + \lambda_{12}\lambda_{21}\text{Cov}(\xi_1, \xi_2) + \lambda_{11}\lambda_{22}\text{Cov}(\xi_1, \xi_2).$$

In termini matriciali si scrive

$$\Sigma = \Lambda\Phi\Lambda^T + \Psi,$$

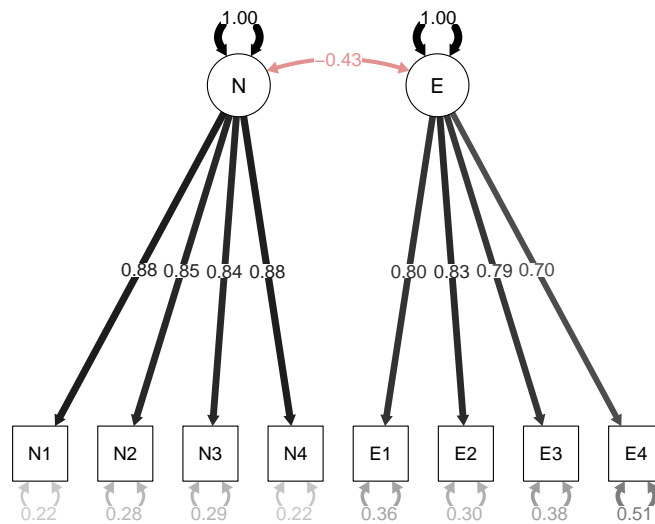
dove Φ è la matrice di ordine $m \times m$ di varianze e covarianze tra i fattori comuni e Ψ è una matrice diagonale di ordine p con le unicità delle variabili.

Esercizio 7.3. Consideriamo nuovamente i dati esaminati negli esercizi precedenti, ma questa volta il modello consente una correlazione tra i due fattori comuni:

```
fit2_cfa <- lavaan::cfa(
  cfa_mod,
  sample.cov = psychot_cor_mat,
  sample.nobs = n,
  orthogonal = FALSE,
  std.lv = TRUE
)
```

Visualizziamo il modello nel modo seguente:

```
semPaths(
  fit2_cfa,
  "std",
  posCol = c("black"),
  edge.label.cex = 1.1,
  sizeMan = 7
)
```



Esaminiamo le saturazioni fattoriali:

```
parameterEstimates(fit2_cfa, standardized = TRUE) %>%
  dplyr::filter(op == "=~") %>%
  dplyr::select(
    "Latent Factor" = lhs,
    Indicator = rhs,
    B = est,
    SE = se,
    Z = z,
    "p-value" = pvalue,
    Beta = std.all
  ) %>%
```

```
knitr::kable(
  digits = 3, booktabs = TRUE, format = "markdown",
  caption = "Factor Loadings"
)
```

Tabella 7.3: *Factor Loadings*

Latent Factor	Indicator	B	SE	Z	p-value	Beta
N	N1	0.883	0.051	17.47	0	0.885
N	N2	0.847	0.052	16.34	0	0.849
N	N3	0.842	0.052	16.19	0	0.844
N	N4	0.880	0.051	17.38	0	0.882
E	E1	0.800	0.055	14.46	0	0.802
E	E2	0.832	0.054	15.29	0	0.834
E	E3	0.788	0.056	14.15	0	0.789
E	E4	0.698	0.058	11.97	0	0.699

Le saturazioni sono simili a quelle che abbiamo trovato in precedenza, In questo caso, però, la matrice delle correlazioni residue è adeguata:

```
cor_table <- residuals(fit2_cfa, type = "cor")$cov
knitr::kable(
  cor_table,
  digits = 3,
  format = "markdown",
  booktabs = TRUE
)
```

	N1	N2	N3	N4	E1	E2	E3	E4
N1	0.000	0.016	-0.015	-0.002	-0.042	0.005	0.008	-0.013
N2	0.016	0.000	-0.007	-0.010	-0.006	0.028	0.002	0.004
N3	-0.015	-0.007	0.000	0.018	-0.062	0.006	-0.007	-0.035
N4	-0.002	-0.010	0.018	0.000	-0.010	0.053	0.007	0.023
E1	-0.042	-0.006	-0.062	-0.010	0.000	0.006	0.001	-0.027
E2	0.005	0.028	0.006	0.053	0.006	0.000	-0.007	0.010
E3	0.008	0.002	-0.007	0.007	0.001	-0.007	0.000	0.014

	N1	N2	N3	N4	E1	E2	E3	E4
E4	-0.013	0.004	-0.035	0.023	-0.027	0.010	0.014	0.000

Esercizio 7.4. Esaminiamo più da vicino la matrice di correlazioni riprodotta dal modello, nel caso di fattori obliqui. Le saturazioni fattoriali sono:

```
lambda <- inspect(fit2_cfa, what = "std")$lambda
lambda
#>      N      E
#> N1 0.885 0.000
#> N2 0.849 0.000
#> N3 0.844 0.000
#> N4 0.882 0.000
#> E1 0.000 0.802
#> E2 0.000 0.834
#> E3 0.000 0.789
#> E4 0.000 0.699
```

La matrice di intercorrelazioni fattoriali è

```
Phi <- inspect(fit2_cfa, what = "std")$psi
Phi
#>      N      E
#> N 1.000
#> E -0.435 1.000
```

Le varianze residue sono:

```
Psi <- inspect(fit2_cfa, what = "std")$theta
Psi
#>      N1      N2      N3      N4      E1      E2      E3      E4
#> N1 0.217
#> N2 0.000 0.280
#> N3 0.000 0.000 0.288
#> N4 0.000 0.000 0.000 0.222
#> E1 0.000 0.000 0.000 0.000 0.357
```

```
#> E2 0.000 0.000 0.000 0.000 0.000 0.305
#> E3 0.000 0.000 0.000 0.000 0.000 0.000 0.377
#> E4 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.511
```

Mediante i parametri del modello la matrice di correlazione si riproduce nel modo seguente:

$$\Sigma = \Lambda\Phi\Lambda^T + \Psi.$$

Le corrispondenti istruzioni R sono:

```
R_hat <- lambda %**% Phi %**% t(lambda) + Psi
R_hat %>%
  round(3)
#>   N1    N2    N3    N4    E1    E2    E3
#> N1  1.000
#> N2  0.751  1.000
#> N3  0.746  0.716  1.000
#> N4  0.780  0.748  0.744  1.000
#> E1 -0.309 -0.296 -0.294 -0.308  1.000
#> E2 -0.321 -0.308 -0.306 -0.320  0.669  1.000
#> E3 -0.304 -0.291 -0.290 -0.303  0.633  0.658  1.000
#> E4 -0.269 -0.258 -0.257 -0.268  0.561  0.583  0.552
#>   E4
#> N1
#> N2
#> N3
#> N4
#> E1
#> E2
#> E3
#> E4  1.000
```

Le correlazioni residue sono:

```
(psychot_cor_mat - R_hat) %>%
  round(3)
#>   N1    N2    N3    N4    E1    E2    E3
```

```
#> N1  0.000
#> N2  0.016  0.000
#> N3 -0.015 -0.007  0.000
#> N4 -0.002 -0.010  0.018  0.000
#> E1 -0.042 -0.006 -0.062 -0.010  0.000
#> E2  0.005  0.028  0.006  0.053  0.006  0.000
#> E3  0.008  0.002 -0.007  0.007  0.001 -0.007  0.000
#> E4 -0.013  0.004 -0.035  0.023 -0.027  0.010  0.014
#>    E4
#> N1
#> N2
#> N3
#> N4
#> E1
#> E2
#> E3
#> E4  0.000
```

Per fare un esempio, calcoliamo la correlazione predetta dal modello tra le variabili Y_1 e Y_2 :

```
lambda[1, 1] * lambda[2, 1] + lambda[1, 2] * lambda[2, 2] +
  lambda[1, 1] * lambda[2, 2] * Phi[1, 2] +
  lambda[1, 2] * lambda[2, 1] * Phi[1, 2]
#> [1] 0.7508
```

Questo valore si avvicina al valore contenuto dell'elemento (1, 2) della matrice di correlazioni osservate:

```
psychot_cor_mat[1, 2]
#> [1] 0.767
```

Usando le funzionalità di `lavaan` la matrice di correlazione predetta si ottiene con:

```
fitted(fit2_cfa)$cov
#>      N1      N2      N3      N4      E1      E2      E3
```



```
#> N1  0.996
#> N2  0.748  0.996
#> N3  0.743  0.713  0.996
#> N4  0.777  0.745  0.741  0.996
#> E1 -0.307 -0.295 -0.293 -0.306  0.996
#> E2 -0.320 -0.306 -0.305 -0.319  0.666  0.996
#> E3 -0.303 -0.290 -0.289 -0.302  0.630  0.656  0.996
#> E4 -0.268 -0.257 -0.255 -0.267  0.558  0.580  0.550
#>   E4
#> N1
#> N2
#> N3
#> N4
#> E1
#> E2
#> E3
#> E4  0.996
```

La matrice dei residui è

```
resid(fit2_cfa)$cov
#>   N1    N2    N3    N4    E1    E2    E3
#> N1  0.000
#> N2  0.016  0.000
#> N3 -0.015 -0.007  0.000
#> N4 -0.002 -0.010  0.018  0.000
#> E1 -0.042 -0.006 -0.062 -0.010  0.000
#> E2  0.005  0.028  0.006  0.053  0.006  0.000
#> E3  0.008  0.002 -0.007  0.007  0.001 -0.007  0.000
#> E4 -0.013  0.004 -0.035  0.023 -0.026  0.010  0.014
#>   E4
#> N1
#> N2
#> N3
#> N4
#> E1
#> E2
#> E3
```

```
#> E4 0.000
```

La matrice dei residui standardizzati è

```
resid(fit2_cfa, type = "standardized")$cov
#>   N1    N2    N3    N4    E1    E2    E3
#> N1 0.000
#> N2 1.674 0.000
#> N3 -1.769 -0.569 0.000
#> N4 -0.350 -1.152 1.746 0.000
#> E1 -1.214 -0.161 -1.646 -0.294 0.000
#> E2 0.154 0.794 0.168 1.626 0.637 0.000
#> E3 0.219 0.062 -0.191 0.193 0.075 -0.693 0.000
#> E4 -0.314 0.092 -0.824 0.552 -1.481 0.624 0.690
#>   E4
#> N1
#> N2
#> N3
#> N4
#> E1
#> E2
#> E3
#> E4 0.000
```

7.3 EFA con lavaan

Una funzionalità sperimentale di `lavaan` (ancora non ufficiale) è quella che consente di svolgere l'analisi fattoriale esplorativa con la funzione `efa()`. Consideriamo nuovamente i dati di [Brown \(2015\)](#), ovvero otto misure di personalità raccolte su un campione di 250 pazienti che hanno concluso un programma di psicoterapia.

```
# 1-factor model
f1 <- '
efa("efa")*f1 =~ N1 + N2 + N3 + N4 + E1 + E2 + E3 + E4
```

```
'
# 2-factor model
f2 <- '
efa("efa")*f1 +
efa("efa")*f2 =~ N1 + N2 + N3 + N4 + E1 + E2 + E3 + E4
'
```

```
efa_f1 <-
  cfa(
    model = f1,
    sample.cov = psychot_cor_mat,
    sample.nobs = 250,
    rotation = "oblimin"
  )
```

```
summary(
  efa_f1,
  fit.measures = TRUE,
  standardized = TRUE,
  rsquare = TRUE
)
#> lavaan 0.6-10 ended normally after 2 iterations
#>
#> Estimator ML
#> Optimization method NLMINB
#> Number of model parameters 16
#>
#> Rotation method OBLIMIN OBLIQUE
#> Oblimin gamma 0
#> Rotation algorithm (rstarts) GPA (100)
#> Standardized metric TRUE
#> Row weights None
#>
#> Number of observations 250
#>
#> Model Test User Model:
#>
```

```

#> Test statistic 375.327
#> Degrees of freedom 20
#> P-value (Chi-square) 0.000
#>
#> Model Test Baseline Model:
#>
#> Test statistic 1253.791
#> Degrees of freedom 28
#> P-value 0.000
#>
#> User Model versus Baseline Model:
#>
#> Comparative Fit Index (CFI) 0.710
#> Tucker-Lewis Index (TLI) 0.594
#>
#> Loglikelihood and Information Criteria:
#>
#> Loglikelihood user model (H0) -2394.637
#> Loglikelihood unrestricted model (H1) -2206.974
#>
#> Akaike (AIC) 4821.275
#> Bayesian (BIC) 4877.618
#> Sample-size adjusted Bayesian (BIC) 4826.897
#>
#> Root Mean Square Error of Approximation:
#>
#> RMSEA 0.267
#> 90 Percent confidence interval - lower 0.243
#> 90 Percent confidence interval - upper 0.291
#> P-value RMSEA <= 0.05 0.000
#>
#> Standardized Root Mean Square Residual:
#>
#> SRMR 0.187
#>
#> Parameter Estimates:
#>
#> Standard errors Standard

```

```

#> Information
#> Information saturated (h1) model Expected
#>                                     Structured
#> Latent Variables:
#>                                     Estimate Std.Err z-value P(>|z|)
#> f1 =~ efa
#> N1          0.879    0.051  17.333  0.000
#> N2          0.841    0.052  16.154  0.000
#> N3          0.841    0.052  16.175  0.000
#> N4          0.870    0.051  17.065  0.000
#> E1         -0.438    0.062  -7.041  0.000
#> E2         -0.398    0.063  -6.327  0.000
#> E3         -0.398    0.063  -6.342  0.000
#> E4         -0.364    0.063  -5.746  0.000
#> Std.lv Std.all
#>
#> 0.879 0.880
#> 0.841 0.842
#> 0.841 0.843
#> 0.870 0.872
#> -0.438 -0.439
#> -0.398 -0.398
#> -0.398 -0.399
#> -0.364 -0.364
#>
#> Variances:
#>                                     Estimate Std.Err z-value P(>|z|)
#> .N1          0.224    0.028   7.915  0.000
#> .N2          0.289    0.033   8.880  0.000
#> .N3          0.288    0.032   8.866  0.000
#> .N4          0.239    0.029   8.174  0.000
#> .E1          0.804    0.073  10.963  0.000
#> .E2          0.838    0.076  11.008  0.000
#> .E3          0.837    0.076  11.007  0.000
#> .E4          0.864    0.078  11.041  0.000
#> f1          1.000
#> Std.lv Std.all
#> 0.224 0.225

```

```
#>      0.289      0.290
#>      0.288      0.289
#>      0.239      0.240
#>      0.804      0.807
#>      0.838      0.841
#>      0.837      0.841
#>      0.864      0.867
#>      1.000      1.000
#>
#> R-Square:
#>                Estimate
#>      N1                0.775
#>      N2                0.710
#>      N3                0.711
#>      N4                0.760
#>      E1                0.193
#>      E2                0.159
#>      E3                0.159
#>      E4                0.133
```

```
efa_f2 <-
  cfa(
    model = f2,
    sample.cov = psychot_cor_mat,
    sample.nobs = 250,
    rotation = "oblimin"
  )
```

```
summary(
  efa_f2,
  fit.measures = TRUE,
  standardized = TRUE,
  rsquare = TRUE
)
#> lavaan 0.6-10 ended normally after 1 iterations
#>
#>      Estimator      ML
```

```

#> Optimization method NLMINB
#> Number of model parameters 23
#>
#> Rotation method OBLIMIN OBLIQUE
#> Oblimin gamma 0
#> Rotation algorithm (rstarts) GPA (100)
#> Standardized metric TRUE
#> Row weights None
#>
#> Number of observations 250
#>
#> Model Test User Model:
#>
#> Test statistic 9.811
#> Degrees of freedom 13
#> P-value (Chi-square) 0.709
#>
#> Model Test Baseline Model:
#>
#> Test statistic 1253.791
#> Degrees of freedom 28
#> P-value 0.000
#>
#> User Model versus Baseline Model:
#>
#> Comparative Fit Index (CFI) 1.000
#> Tucker-Lewis Index (TLI) 1.006
#>
#> Loglikelihood and Information Criteria:
#>
#> Loglikelihood user model (H0) -2211.879
#> Loglikelihood unrestricted model (H1) -2206.974
#>
#> Akaike (AIC) 4469.758
#> Bayesian (BIC) 4550.752
#> Sample-size adjusted Bayesian (BIC) 4477.840
#>
#> Root Mean Square Error of Approximation:

```

```

#>
#> RMSEA 0.000
#> 90 Percent confidence interval - lower 0.000
#> 90 Percent confidence interval - upper 0.048
#> P-value RMSEA <= 0.05 0.957
#>
#> Standardized Root Mean Square Residual:
#>
#> SRMR 0.010
#>
#> Parameter Estimates:
#>
#> Standard errors Standard
#> Information Expected
#> Information saturated (h1) model Structured
#>
#> Latent Variables:
#> Estimate Std.Err z-value P(>|z|)
#> f1 =~ efa
#> N1 0.874 0.053 16.592 0.000
#> N2 0.851 0.055 15.551 0.000
#> N3 0.826 0.054 15.179 0.000
#> N4 0.896 0.053 16.802 0.000
#> E1 -0.046 0.040 -1.138 0.255
#> E2 0.035 0.034 1.030 0.303
#> E3 0.000 0.040 0.010 0.992
#> E4 -0.006 0.049 -0.131 0.896
#> f2 =~ efa
#> N1 -0.017 0.032 -0.539 0.590
#> N2 0.011 0.035 0.322 0.748
#> N3 -0.035 0.036 -0.949 0.343
#> N4 0.031 0.031 0.994 0.320
#> E1 0.776 0.059 13.125 0.000
#> E2 0.854 0.058 14.677 0.000
#> E3 0.785 0.060 13.106 0.000
#> E4 0.695 0.063 10.955 0.000
#> Std.lv Std.all
#>

```



```

#>      0.874      0.876
#>      0.851      0.853
#>      0.826      0.828
#>      0.896      0.898
#>     -0.046     -0.046
#>      0.035      0.035
#>      0.000      0.000
#>     -0.006     -0.006
#>
#>     -0.017     -0.017
#>      0.011      0.011
#>     -0.035     -0.035
#>      0.031      0.031
#>      0.776      0.778
#>      0.854      0.855
#>      0.785      0.787
#>      0.695      0.697
#>
#> Covariances:
#>                                Estimate Std.Err z-value P(>|z|)
#> f1 ~~
#> f2                                -0.432    0.059  -7.345    0.000
#> Std.lv Std.all
#>
#>     -0.432     -0.432
#>
#> Variances:
#>                                Estimate Std.Err z-value P(>|z|)
#> .N1                                0.218    0.028   7.790    0.000
#> .N2                                0.279    0.032   8.693    0.000
#> .N3                                0.287    0.032   8.907    0.000
#> .N4                                0.216    0.029   7.578    0.000
#> .E1                                0.361    0.044   8.226    0.000
#> .E2                                0.292    0.043   6.787    0.000
#> .E3                                0.379    0.046   8.315    0.000
#> .E4                                0.509    0.053   9.554    0.000
#> f1                                1.000
#> f2                                1.000

```

```

#>   Std.lv  Std.all
#>   0.218   0.219
#>   0.279   0.280
#>   0.287   0.289
#>   0.216   0.217
#>   0.361   0.362
#>   0.292   0.293
#>   0.379   0.381
#>   0.509   0.511
#>   1.000   1.000
#>   1.000   1.000
#>
#> R-Square:
#>               Estimate
#>   N1             0.781
#>   N2             0.720
#>   N3             0.711
#>   N4             0.783
#>   E1             0.638
#>   E2             0.707
#>   E3             0.619
#>   E4             0.489

```

```

# define the fit measures
fit_measures_robust <- c(
  "chisq", "df", "pvalue",
  "cfi", "rmsea", "srmr"
)

# collect them for each model
rbind(
  fitmeasures(efa_f1, fit_measures_robust),
  fitmeasures(efa_f2, fit_measures_robust)
) %>%
# wrangle
data.frame() %>%
mutate(
  chisq = round(chisq, digits = 0),

```

```
df = as.integer(df),
pvalue = ifelse(pvalue == 0, "< .001", pvalue)
) %>%
mutate_at(vars(cfi:srmr), ~ round(., digits = 3))
#>   chisq df          pvalue cfi rmsea srmr
#> 1   375 20          < .001 0.71 0.267 0.187
#> 2    10 13 0.709310449320062 1.00 0.000 0.010
```

I risultati mostrano come la funzione `efa()` sia effettivamente in grado di distinguere in maniera chiara tra i due fattori. Il confronto tra modelli mostra la superiorità del modello a due fattori.



8

Path Analysis

La *path analysis* è un metodo per decomporre la correlazione (o la covarianza) in componenti differenti al fine di studiare i processi causali sottostanti. La *path analysis* comprende due parti principali: la rappresentazione grafica delle interrelazioni esistenti tra le variabili e la scomposizione delle correlazioni (o covarianze) nei termini dei parametri del modello.

8.1 Path diagram

Il path diagram fornisce una rappresentazione grafica delle relazioni esistenti tra le variabili oggetto di interesse. In tale diagramma, le variabili non osservate o latenti sono racchiuse in un cerchio o ellisse; le variabili osservate sono racchiuse in un quadrato o rettangolo. Due classi di variabili vengono rappresentate in un path diagram: quelle che non ricevono effetti causali da altre variabili e quelle che li ricevono. Una variabile *esogena* (cioè esterna) svolge sempre e soltanto funzione di variabile indipendente, ovvero di variabile che causa un effetto. Una variabile *endogena* (cioè interna) può essere effetto di alcune variabili e contemporaneamente causa per altre, oppure può svolgere solo il ruolo di variabile dipendente. Le fonti causali delle variabili endogene sono interne al path diagram; le fonti causali delle variabili esogene sono esterne al path diagram. La distinzione tra variabili esogene e endogene ha delle ovvie assonanze con la distinzione tra variabili indipendenti e dipendenti propria dei modelli lineari.

Le frecce che connettono le variabili nel diagramma denotano nessi causali o mere associazioni. Una freccia orientata rappresenta un nesso causale tra le variabili implicate: la variabile che riceve la freccia dipende dalla variabile da cui parte la freccia. Una freccia curva a due direzioni indi-

ca, invece, un'associazione non causale tra due variabili. Il fatto che due variabili non siano collegate nel diagramma equivale ad assumere che tali variabili siano incorrelate. Un esempio è fornito nella Figura 8.1 la quale rende esplicite le relazioni tra tre variabili latenti e nove variabili manifeste.

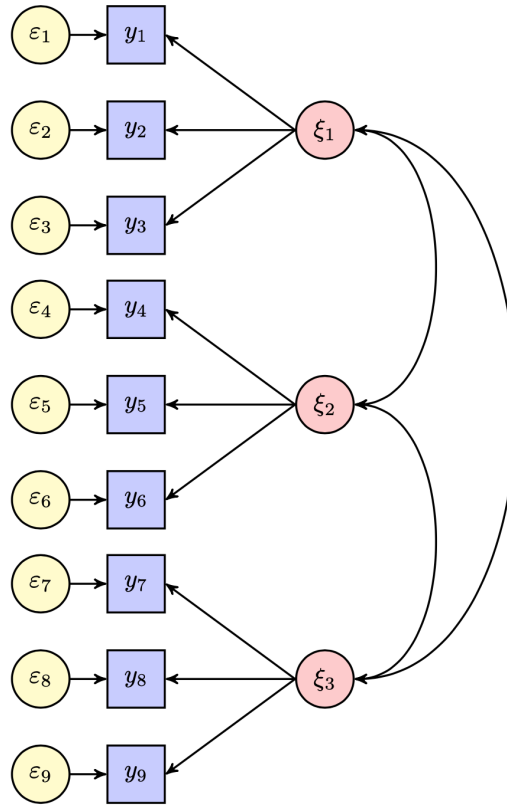


Figura 8.1: Esempio di path diagram.

8.2 Path analysis e regressione multipla

Vi è una stretta relazione tra path analysis e regressione multipla, tanto che la regressione può essere considerata un caso particolare di path analysis. Per semplicità, si supponga che le variabili siano state standardizzate, anche se la stessa analisi può essere condotta per variabili grezze.

Il path diagram mostra la relazione tra tutte le variabili, comprendendo anche i fattori di disturbo, e fornisce dunque la rappresentazione grafica di un sistema di equazioni simultanee. Nel caso di due regressori, il modello di regressione multipla può essere rappresentato tramite il path diagram riportato nella Figura 8.2.

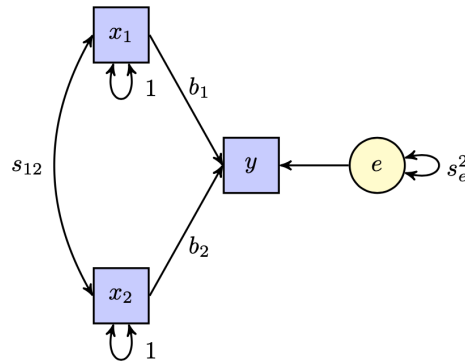


Figura 8.2: Path diagram per il modello di regressione multipla con due regressori.

I coefficienti di percorso associati alle frecce orientate esprimono la portata del nesso causale e corrispondono ai pesi beta (ovvero ai coefficienti parziali di regressione standardizzati). Le frecce non orientate esprimono la portata della pura associazione tra variabili e dunque corrispondono alle correlazioni.

Nel caso di due variabili esogene x_1 e x_2 , il modello di regressione diventa

$$y = b_1 x_1 + b_2 x_2 + 1 \cdot e$$

dove y è la variabile endogena ed e è il fattore di disturbo. Tale modello di regressione può essere rappresentato graficamente come indicato nella figura precedente.

Nella figura, le frecce dritte indicano un'influenza causale dalla variabile da cui parte la freccia a quella a cui la freccia arriva. A tali frecce dritte sono associati i coefficienti di percorso b_1 e b_2 (ovvero i pesi beta). Il coefficiente 1 rappresenta l'effetto del fattore di disturbo e sulla variabile endogena y , implicito nelle equazioni e reso esplicito nella figura.

Si noti che si hanno tante equazioni quante sono le variabili endogene. Nel caso presente, c'è un'unica equazione in quanto vi è una sola variabile

endogena (ovvero la y , le cui cause sono interne al path diagram). All'interno di ciascuna equazione, inoltre, ci saranno tanti termini quante sono le frecce dritte che puntano verso la variabile endogena. Nell'esempio, ci sono tre termini, uno per ciascun freccia dritta.

8.3 Effetti diretti e indiretti

La path analysis fornisce un metodo per distinguere tra i diversi tipi di effetti che influenzano le variabili: l'effetto diretto, l'effetto indiretto e l'effetto totale. Gli effetti diretti sono quelli non mediati da altre variabili. Gli effetti indiretti operano attraverso l'intervento di almeno una variabile. L'effetto totale è la somma di tutti gli effetti diretti e indiretti.

Nella Figura 8.3 la variabile y_1 ha un effetto diretto sulla y_2 . La variabile y_1 ha un effetto indiretto sulla y_3 in quanto non c'è una freccia causale che colleghi direttamente la variabile y_1 alla y_3 . La variabile y_1 è una variabile esogena e le variabili y_2 e y_3 sono variabili endogene.

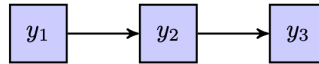


Figura 8.3: Path diagram per una relazione a catena.

Nella Figura 8.3, la variabile x_1 ha un effetto diretto sulla y , ma anche un effetto indiretto sulla y derivante dalla correlazione tra x_1 e x_2 . In un path diagram, l'effetto diretto è rappresentato da una freccia dritta (es., b_1). L'effetto indiretto tra due variabili è rappresentato da un percorso composto che include una o più frecce dritte e non più di una linea curva – per es., $s_{12}b_2$.

8.4 Le regole di Wright

Lo scopo della path analysis è quello di decomporre la correlazione (o la covarianza) nei termini della somma di tutti i percorsi (diretti e indiretti) che legano le due variabili tramite i coefficienti detti *path coefficients*.

Usando il path diagram, Sewall Wright (1921, 1934) enunciò le regole che, attraverso le cosiddette *tracing rules*, legano le correlazioni (o covarianze) delle variabili ai parametri del modello. Le tracing rules possono essere espresse nei termini seguenti:

- è possibile procedere prima all'indietro lungo una freccia e poi in avanti, seguendo la direzione di una freccia, ma non si può andare prima avanti e poi tornare indietro;
- un percorso composto non deve transitare due volte per la stessa variabile (non devono esserci loop);
- un percorso non può comprendere più di una linea curva.

Si chiama “percorso” il tracciato che unisce due variabili; è costituito da sequenze di frecce direzionali e di curve non direzionali. A ciascun percorso legittimo (ovvero, che soddisfa le regole di Wright) viene assegnato un valore numerico pari al prodotto dei coefficienti incontrati sul percorso medesimo. I coefficienti di percorso possono essere o coefficienti parziali di regressione standardizzati, se il legame ha una direzione, oppure coefficienti di correlazione, se il legame è bidirezionale.

8.4.1 Scomposizione delle correlazioni (covarianze)

Il principio di base è stato espresso da Sewall Wright (1934) nel modo seguente: “Any correlation between variables in a network of sequential relations can be analyzed into contributions from all the paths (direct or through common factors) by which the two variables are connected, such that the value of each contribution is the product of the coefficients pertaining to the elementary paths. If residual correlations are present (represented by bidirectional arrows) one (but never more than one) of the coefficients thus multiplied together to give the contribution of the connecting path, may be a correlation coefficient. The others are all path coefficients.”

Possiamo così enunciare la regola di scomposizione della correlazione.

Definizione 8.1. La correlazione fra due variabili può essere decomposta in tanti addendi quanto sono i percorsi che le collegano; ogni addendo è dato dal prodotto dei coefficienti incontrati sul percorso.

Si consideri il diagramma rappresentato nella Figura 8.2. La variabile endogena è la y . Le variabili esogene, correlate tra loro, sono x_1 e x_2 .

Il diagramma di percorso corrisponde alla seguente equazione:

$$y = 0.50x_1 + 0.40x_2 + e$$

dove le variabili x_1 e x_2 sono incorrelate con e .

La correlazione tra y e x_1 è uguale alla somma dell'effetto diretto che x_1 esercita sulla y e dell'effetto indiretto che x_1 esercita sulla y tramite la correlazione con x_2 . In base alle regole di Wright, x_1 e y risultano collegate da due percorsi legittimi: il percorso costituito dalla freccia dritta $x_1 \rightarrow y$; il percorso composto dalla freccia dritta $x_2 \rightarrow y$ e dalla curva non direzionale $x_1 \leftrightarrow x_2$. Il valore numerico del primo percorso è 0.50. Il valore numerico del secondo percorso è 0.50×0.40 . La correlazione tra le variabili x_1 e y è dunque uguale alla somma dei valori numerici dei due percorsi legittimi che legano x_1 alla y :

$$\begin{aligned} r_{x_1,y} &= \beta_{y,x_1} + r_{x_1,x_2}\beta_{y,x_2} \\ &= 0.50 + 0.50 \times 0.40 = 0.70. \end{aligned}$$

La correlazione tra x_2 e y è invece uguale a:

$$\begin{aligned} r_{yx_2} &= \beta_{yx_2} + r_{x_1x_2}\beta_{yx_1} \\ &= 0.40 + 0.50 \times 0.50 = 0.65. \end{aligned}$$

8.4.2 Scomposizione della varianza

La varianza di una variabile endogena si decompone in una quota di varianza spiegata dalle variabili agenti causalmente su di essa e in una quota di varianza non spiegata.

Definizione 8.2. La varianza spiegata è data dalla somma di tanti addendi quanti sono i percorsi che consentono di collegare la variabile a se stessa rispettando le tracing rules di Wright.

Facendo riferimento alla Figura 8.2, si possono individuare quattro percorsi legittimi che collegano y a se stessa:

1. $0.50 \times 1.00 \times 0.50$,
2. $0.40 \times 1.00 \times 0.40$,
3. $0.50 \times 0.50 \times 0.40$,

$$4. \quad 0.40 \times 0.50 \times 0.50.$$

La varianza della variabile endogena y che viene spiegata dalle variabili esogene x_1 e x_2 è dunque uguale a
 $0.25 + 0.16 + 0.10 + 0.10 = 0.61$.

Inoltre, dato che le variabili rappresentate nel diagramma sono standardizzate, la varianza complessiva della y è uguale a 1.00. La varianza della y non spiegata dalle variabili x_1 e x_2 è quindi uguale a
 $1 - 0.61 = 0.39$.

8.5 Come calcolare i coefficienti di percorso?

Data una matrice di correlazione, i coefficienti di percorso possono essere calcolati risolvendo un sistema di equazioni simultanee. Si supponga che, per le tre variabili della figura precedente, vi sia la seguente matrice di correlazione:

	y	x_1	x_2
y	1.00		
x_1	0.70	1.00	
x_2	0.65	0.50	1.00

Esprimendo le tre correlazioni nei termini dei coefficienti del path diagram otteniamo:

$$\begin{cases} r_{x_1 x_2} &= 0.50 \\ r_{y x_2} &= \beta_{y x_2} + 0.50 \beta_{y x_1} = 0.65 \\ r_{x_1 y} &= \beta_{y x_1} + 0.50 \beta_{y x_2} = 0.70 \end{cases} \quad (8.1)$$

Risolvendo il sistema di equazioni simultanee, si ottengono i valori dei coefficienti di percorso:

$$\begin{aligned} \beta_{y x_1} &= 0.50 \\ \beta_{y x_2} &= 0.40 \end{aligned}$$

8.6 Path analysis e software

Vengono simulate 100 osservazioni su tre variabili. Imponendo un effetto causale diretto delle variabili x_1 e x_2 sulla y e una correlazione > 0 tra le variabili x_1 e x_2 , otteniamo i seguenti dati:

```
set.seed(3)
n <- 100
x1 <- rnorm(n, 100, 9)
x2 <- x1 + rnorm(n, 0, 10)
cor(x1, x2)
#> [1] 0.5346
y <- 10 + 3 * x1 + 1.5 * x2 + rnorm(n, 0, 15)
dd <- data.frame(y, x1, x2)
print(cor(dd), 3)
#>      y      x1      x2
#> y  1.000 0.831 0.786
#> x1 0.831 1.000 0.535
#> x2 0.786 0.535 1.000
lower <- "
  1
  .831 1
  .786 .535 1
"

library("lavaan")
dat.cov <- getCov(lower, names = c("y", "x1", "x2"))
dat.cov
#>      y      x1      x2
#> y  1.000 0.831 0.786
#> x1 0.831 1.000 0.535
#> x2 0.786 0.535 1.000
```

Data una matrice di correlazioni e data la specificazione delle relazioni tra le variabili, la funzione `sem()` contenuta nel pacchetto `lavaan` consente di stimare i coefficienti di percorso. Le relazioni tra le variabili sono specificate con la sintassi descritta nel manuale.

```

# the model
mr.model <- "y ~ x1 + x2"
fit <- sem(mr.model, sample.cov = dat.cov, sample.nobs = n)
summary(fit, standardized = TRUE)
#> lavaan 0.6-10 ended normally after 1 iterations
#>
#> Estimator ML
#> Optimization method NLMINB
#> Number of model parameters 3
#>
#> Number of observations 100
#>
#> Model Test User Model:
#>
#> Test statistic 0.000
#> Degrees of freedom 0
#>
#> Parameter Estimates:
#>
#> Standard errors Standard
#> Information Expected
#> Information saturated (h1) model Structured
#>
#> Regressions:
#>
#> Estimate Std.Err z-value P(>|z|)
#> y ~
#> x1 0.575 0.045 12.710 0.000
#> x2 0.478 0.045 10.571 0.000
#> Std.lv Std.all
#>
#> 0.575 0.575
#> 0.478 0.478
#>
#> Variances:
#>
#> Estimate Std.Err z-value P(>|z|)
#> .y 0.145 0.020 7.071 0.000
#> Std.lv Std.all
#> 0.145 0.146

```

```
summary(lm(scale(y) ~ scale(x1) + scale(x2)))
#>
#> Call:
#> lm(formula = scale(y) ~ scale(x1) + scale(x2))
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -0.8720 -0.2462  0.0163  0.2571  0.9038
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept) -7.47e-16   3.87e-02    0.0      1
#> scale(x1)    5.75e-01   4.60e-02   12.5   <2e-16 ***
#> scale(x2)    4.79e-01   4.60e-02   10.4   <2e-16 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.387 on 97 degrees of freedom
#> Multiple R-squared:  0.853, Adjusted R-squared:  0.85
#> F-statistic: 283 on 2 and 97 DF, p-value: <2e-16
1 - .145
#> [1] 0.855
```

Consideriamo nuovamente la seguente matrice di correlazioni:

```
lower <- "
  1
.70 1
.65 .50 1
"
```

Converto tali dati in una matrice simmetrica.

```
dat.cov <- getCov(lower, names = c("y", "x1", "x2"))
dat.cov
#>      y  x1  x2
#> y  1.00 0.7 0.65
```

```
#> x1 0.70 1.0 0.50
#> x2 0.65 0.5 1.00
```

Specifico il modello:

```
mr_model <- "y ~ x1 + x2"
```

Adatto il modello ai dati:

```
fit <- sem(mr_model, sample.cov = dat.cov, sample.nobs = n)
```

Esamino i risultati:

```
summary(fit, standardized = TRUE)
#> lavaan 0.6-10 ended normally after 1 iterations
#>
#> Estimator ML
#> Optimization method NLMINB
#> Number of model parameters 3
#>
#> Number of observations 100
#>
#> Model Test User Model:
#>
#> Test statistic 0.000
#> Degrees of freedom 0
#>
#> Parameter Estimates:
#>
#> Standard errors Standard
#> Information Expected
#> Information saturated (h1) model Structured
#>
#> Regressions:
#> Estimate Std.Err z-value P(>|z|)
#> y ~
#> x1 0.500 0.072 6.934 0.000
```

```

#>      x2              0.400    0.072    5.547    0.000
#>      Std.lv Std.all
#>
#>      0.500    0.500
#>      0.400    0.400
#>
#> Variances:
#>              Estimate Std.Err z-value P(>|z|)
#>      .y              0.386    0.055    7.071    0.000
#>      Std.lv Std.all
#>      0.386    0.390

```

Con la funzione `sem()` del pacchetto `lavaan` abbiamo dunque replicato i risultati ottenuti in precedenza. Il valore 0.386 rappresenta la quota di varianza della y non spiegata dalle variabili esogene.

8.7 Oltre la regressione multipla

In generale, lo psicologo ha a che fare con diagrammi di percorso nei quali sono presenti variabili non osservabili (latenti) e quindi l'approccio della regressione multipla non può essere applicato. È necessario invece descrivere il diagramma di percorso mediante un insieme di equazioni strutturali, definendo un numero di equazioni almeno altrettanto grande quanto il numero delle incognite. Tale soluzione viene solitamente fornita da un software. Consideriamo di seguito alcuni esempi in cui vengono applicate le regole di Wright per diagrammi di percorso che non possono essere descritti nei termini di un modello di regressione multipla. Un esempio di path diagram che non si riduce al modello di regressione multipla è quello fornito nella Figura 8.1.

La path analysis è anche usata in quel campo della psicologia interessato alla misurazione dei costrutti psicologici quali i tratti della personalità, le capacità cognitive e i disturbi psicopatologici. Questa è la ragione per cui la discutiamo qui.

9

Attendibilità e modello fattoriale

McDonald (2013) mostra come la teoria classica dei test possa essere messa in relazione con il modello dell'analisi fattoriale. La figura 9.1 descrive nei termini del modello fattoriale la relazione che intercorre tra i punteggi Y ottenuti dalla somministrazione di un test con cinque item e i punteggi veri.

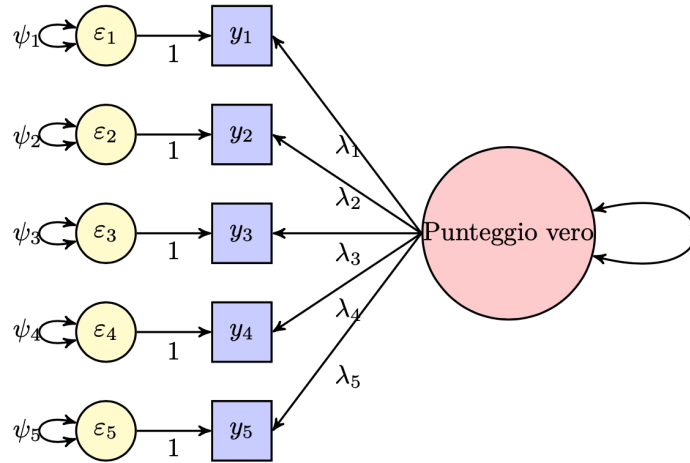


Figura 9.1: Modello fattoriale che mette in relazione i punteggi osservati e i punteggi veri.

Il metodo delle forme parallele proposto dalla teoria classica dei test fornisce una risposta solo in parte soddisfacente al problema della stima del coefficiente di attendibilità. Ricordiamo che il metodo delle forme parallele consiste nel somministrare due questionari X e X' , espressione dello stesso costrutto, nella stessa occasione allo stesso campione di soggetti. In tali circostanze $\rho_{XT}^2 = \rho_{XX'}$. Affinché la relazione definita dall'equazione precedente sia vera, le due forme del test devono essere parallele, nel senso descritto della teoria classica dei test. In pratica, però, è impossibile somministrare lo stesso test due volte agli stessi rispondenti *nelle*

medesime condizioni. È dunque necessario basare la stima del coefficiente di attendibilità sui dati acquisiti mediante un'unica somministrazione del test.

Vi sono vari metodi per la stima dell'attendibilità nel caso di un'unica somministrazione di un test. Considereremo qui tre metodi che possono essere applicati mediante l'utilizzo dell'analisi fattoriale: l' α di Cronbach, l' ω di McDonald e il metodo di Spearman-Brown. Il coefficiente α è la misura più utilizzata per la stima dell'attendibilità quale coerenza interna, o omogeneità. Vedremo come tale indice costituisca il limite inferiore dell'attendibilità di un test, se alcune assunzioni sono soddisfatte, mentre risulta uno stimatore distorto dell'attendibilità se le assunzioni che descriveremo risultano violate.

Per discutere i diversi metodi di stima dell'attendibilità quale coerenza interna è prima necessario distinguere tra tre diverse forme che il modello mono-fattoriale può assumere. Queste tre forme sono quelle del modello con indicatori congenerici, τ -equivalenti e paralleli.

9.1 Modello fattoriale e CTT

Sia X_1, X_2, \dots, X_p , con $p > 2$, un insieme di item osservati. I punteggi ottenuti su tali item sono costituiti da una componente di punteggio vero e da una componente d'errore:

$$\begin{aligned} X_1 &= T_1 + E_1, \\ X_2 &= T_2 + E_2, \\ &\dots \\ X_p &= T_p + E_p. \end{aligned}$$

Seguendo McDonald (1999), tale scomposizione in una componente vera e in una componente d'errore può essere espressa nei termini dei parametri del modello fattoriale. L'espressione $X_i = T_i + E_i$ può infatti essere riscritta come

$$X_i = \lambda_i \xi + \delta_i, \quad i = 1, \dots, p,$$

dove X_i denota il punteggio osservato per l'item i -esimo, λ_i è il peso fattoriale i -esimo, ξ è il fattore comune e δ_i è la componente erratica del punteggio osservato i -esimo. Valgono le assunzioni del modello monofattoriale. Ovvero, si assume che ξ e δ_i siano incorrelati per ciascun item i -esimo e che δ_i e δ_k siano incorrelati per ciascuna coppia $i \neq k$.

9.2 Classi di modelli

Si possono distinguere tre importanti casi del modello mono-fattoriale:

- il modello con indicatori congenerici,
- il modello con indicatori τ -equivalenti,
- il modello con indicatori paralleli.

Il modello con indicatori congenerici rappresenta il caso più generale, mentre gli indicatori τ -equivalenti e paralleli sono casi particolari, ovvero impongono restrizioni al modello con indicatori congenerici.

9.2.1 Indicatori congenerici

Indicatori *congenerici* misurano lo stesso costrutto, ma non necessariamente nella stessa misura. Nel caso di indicatori congenerici, nel modello mono-fattoriale non viene imposto alcun vincolo né sulle saturazioni fattoriali né sulle specificità:

$$\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_p,$$

$$\psi_{11} \neq \psi_{22} \neq \dots \neq \psi_{pp}.$$

Il modello mono-fattoriale con indicatori congenerici è dunque

$$X_i = \lambda_i \xi + \delta_i. \quad (9.1)$$

Dalle assunzioni precedenti possiamo derivare la matrice Σ riprodotta in base al modello congenerico la quale risulta essere uguale a

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p}, \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p}, \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}.$$

Si noti come tutte le varianze e tutte le covarianze siano tra loro diverse.

9.2.2 Indicatori tau-equivalenti

Nel caso di indicatori τ -equivalenti, si ha che

$$\lambda_1 = \lambda_2 = \dots = \lambda_p = \lambda,$$

$$\psi_{11} \neq \psi_{22} \neq \dots \neq \psi_{pp}.$$

Il modello monofattoriale con indicatori τ -equivalenti diventa dunque

$$X_i = \lambda\xi + \delta_i, \quad (9.2)$$

ovvero

$$X_i = \tau + \delta_i, \quad (9.3)$$

dove $\tau = \lambda\xi$ è l'attributo comune scalato nell'unità di misura dell'indicatore. Secondo il modello (9.2), tutte le $p(p-1)$ covarianze tra gli item del test devono essere uguali, ovvero

$$\sigma_{ik} = \lambda^2 = \sigma_T^2, \quad (9.4)$$

per $i \neq k$. Gli elementi sulla diagonale principale della matrice di varianze e covarianze saranno invece

$$\sigma_{ii} = \lambda^2 + \psi_{ii} = \sigma_T^2 + \psi_{ii}. \quad (9.5)$$

La matrice Σ riprodotta in base al modello τ -equivalente è dunque uguale a

$$\Sigma = \begin{bmatrix} \sigma_T^2 + \psi_{11} & \sigma_T^2 & \dots & \sigma_T^2 \\ \sigma_T^2 & \sigma_T^2 + \psi_{22} & \dots & \sigma_T^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_T^2 & \sigma_T^2 & \dots & \sigma_T^2 + \psi_{pp} \end{bmatrix}. \quad (9.6)$$

Tutte le covarianze sono uguali, mentre le varianze sono tra loro diverse.

9.2.3 Indicatori paralleli

Nel caso di indicatori paralleli si ha che

$$\begin{aligned} \lambda_1 &= \lambda_2 = \dots = \lambda_p = \lambda, \\ \psi_{11} &= \psi_{22} = \dots = \psi_{pp} = \psi. \end{aligned}$$

Il modello costituito da indicatori paralleli impone dunque un'ulteriore restrizione che riguarda le varianze degli item, ovvero:

$$\sigma_{ii} = \lambda^2 + \psi = \sigma_T^2 + \sigma^2.$$

La struttura di varianze e covarianze imposta dal modello per indicatori paralleli è dunque tale da richiedere l'uguaglianza tra tutte le covarianze tra gli item e l'uguaglianza tra tutte le varianze degli item. La matrice Σ riprodotta in base al modello con indicatori paralleli è dunque uguale a

$$\Sigma = \begin{bmatrix} \sigma_T^2 + \sigma^2 & \sigma_T^2 & \dots & \sigma_T^2 \\ \sigma_T^2 & \sigma_T^2 + \sigma^2 & \dots & \sigma_T^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_T^2 & \sigma_T^2 & \dots & \sigma_T^2 + \sigma^2 \end{bmatrix}.$$

9.2.4 Indicatori strettamente paralleli

L'aggiunta di un ulteriore vincolo a quelli definiti dal modello costituito da indicatori paralleli, ovvero quello dell'eguaglianza delle medie, definisce gli indicatori detti *strettamente paralleli* (McDonald, 1999).

9.3 Metodo dei minimi quadrati non pesati

Nel modello uni-fattoriale, la varianza di ciascun indicatore viene scomposta nella somma di due componenti: la componente σ_T^2 dovuta all'effetto del fattore latente comune e la componente ψ dovuta all'effetto del fattore specifico. McDonald (2013) illustra come sia possibile stimare tali componenti dai dati osservati. Tali stime vengono poi utilizzate per calcolare la coerenza interna del test tramite le formule degli indici α di Cronbach e ω di McDonald.

In precedenza abbiamo visto come la varianza del punteggio vero sia uguale alla covarianza tra due forme parallele dello stesso test: $\sigma_T^2 = \sigma_{XX'}$. Se gli indicatori sono τ -equivalenti, la matrice la matrice Σ riprodotta dal modello è uguale a

$$\Sigma = \begin{bmatrix} \sigma_T^2 + \psi_{11} & \sigma_T^2 & \dots & \sigma_T^2 \\ \sigma_T^2 & \sigma_T^2 + \psi_{22} & \dots & \sigma_T^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_T^2 & \sigma_T^2 & \dots & \sigma_T^2 + \psi_{pp} \end{bmatrix},$$

ovvero, tutte le covarianze sono tra loro uguali. Nel caso di indicatori τ -equivalenti, dunque, una stima $\hat{\sigma}_T^2$ di σ_T^2 è data dalla media delle covarianze della matrice \mathbf{S} :

$$\hat{\sigma}_T^2 = \frac{1}{p(p-1)} \sum_{i \neq k} \sum s_{ik}. \quad (9.7)$$

Tale metodo di stima di σ_T^2 viene chiamato “metodo dei minimi quadrati non pesati” (McDonald, 2013).

Inoltre, nel caso di indicatori τ -equivalenti, la stima di ψ_{ii} nella (9.5) è data da

$$\hat{\psi}_{ii} = s_{ii} - \hat{\sigma}_T^2,$$

per ciascun item.

Nel caso di *indicatori paralleli*, la stima di σ_T^2 è ancora data dalla (9.7), ovvero dalla media delle covarianze della matrice Σ . La stima del valore costante ψ è invece data da

$$\hat{\psi} = \frac{1}{p} \sum_i (s_{ii} - \hat{\sigma}_T^2) \quad (9.8)$$

9.4 Varianza del punteggio totale di un test

Il punteggio totale Y di un test omogeneo è uguale alla somma dei punteggi X_i sui p item di cui è composto il test: $Y = \sum_{i=1}^p X_i$. Poniamoci ora il problema di descrivere la varianza del punteggio totale del test nei termini dei parametri del modello uni-fattoriale. Nel caso di un modello congenerico ad un fattore comune, la varianza del punteggio totale Y del test può essere scomposta in due componenti: il quadrato della somma delle saturazioni fattoriali, corrispondente alla varianza attribuibile al punteggio vero (ovvero la quota di varianza derivante dall'attributo di cui gli item sono indicatori) e la somma delle varianze specifiche dei p indicatori, corrispondente alla varianza degli errori della misura del punteggio totale del test, ovvero

$$\mathbb{V}(Y) = \left(\sum_i \lambda_i \right)^2 + \sum_i \psi_{ii} \quad (9.9)$$

Dimostrazione. Per un modello congenerico, la varianza del punteggio totale Y è uguale a:

$$\begin{aligned} \mathbb{V}(Y) &= \mathbb{V} \left[\sum_i (\lambda_i \xi + \delta_i) \right] \\ &= \mathbb{V} [(\lambda_1 \xi + \delta_1) + (\lambda_2 \xi + \delta_2) + \dots + (\lambda_p \xi + \delta_p)] \\ &= \mathbb{V} \left[\left(\sum_i \lambda_i \right) \xi + \sum_i \delta_i \right] \\ &= \left(\sum_i \lambda_i \right)^2 \underbrace{\mathbb{V}(\xi)}_{=1} + \sum_i \mathbb{V}(\delta_i) \\ &= \left(\sum_i \lambda_i \right)^2 + \sum_i \psi_{ii}. \end{aligned}$$

□

9.5 Stima dell'attendibilità

9.5.1 Coefficiente omega

Avendo scomposto la varianza del punteggio totale di un test come indicato nella (9.9)

$$\mathbb{V}(Y) = \left(\sum_i \lambda_i \right)^2 + \sum_i \psi_{ii}.$$

McDonald (1999) definisce il coefficiente di attendibilità ω come il rapporto tra la varianza “vera” (attribuibile all’attributo comune) e la varianza totale. Nei termini dei parametri del modello uni-fattoriale, il coefficiente ω diventa:

$$\begin{aligned} \omega &= \frac{(\sum_{i=1}^p \lambda_i)^2}{\sigma_Y^2} \\ &= \frac{(\sum_{i=1}^p \lambda_i)^2}{(\sum_{i=1}^p \lambda_i)^2 + \sum_{i=1}^p \psi_{ii}} \end{aligned}$$

Il coefficiente ω consente dunque di stimare il coefficiente di attendibilità nei termini dei parametri del modello fattoriale congenerico, utilizzando i dati ottenuti in un’unica somministrazione del test.

9.5.1.1 Un esempio concreto

Per illustrare la procedura per il calcolo del coefficiente ω , McDonald (1999) utilizza i dati derivanti dalla somministrazione del test *Satisfaction With Life Scale* (SWLS) a 215 rispondenti. Tale test è costituito da 14 item ma, per semplificare la discussione, McDonald ne utilizza solo 5.

```
SWLS <- matrix(
  c(
    2.565, 1.424, 1.481, 1.328, 1.529,
    1.424, 2.493, 1.267, 1.051, 1.308,
    1.481, 1.267, 2.462, 1.093, 1.360,
    1.328, 1.051, 1.093, 2.769, 1.128,
```



```

    1.529, 1.308, 1.360, 1.128, 3.355
  ),
  ncol = 5, byrow = TRUE
)
SWLS
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 2.565 1.424 1.481 1.328 1.529
#> [2,] 1.424 2.493 1.267 1.051 1.308
#> [3,] 1.481 1.267 2.462 1.093 1.360
#> [4,] 1.328 1.051 1.093 2.769 1.128
#> [5,] 1.529 1.308 1.360 1.128 3.355

```

Eseguiamo l'analisi fattoriale con il metodo della massima verosimiglianza:

```
fa <- factanal(covmat = SWLS, factors = 1, n.obs = 215)
```

Le saturazioni fattoriali sono:

```

fa$load
#>
#> Loadings:
#>      Factor1
#> [1,] 0.817
#> [2,] 0.694
#> [3,] 0.726
#> [4,] 0.591
#> [5,] 0.643
#>
#>      Factor1
#> SS loadings    2.438
#> Proportion Var 0.488

```

Le specificità sono uguali a

```

fa$uniq
#> [1] 0.3330 0.5182 0.4732 0.6512 0.5867

```

Il coefficiente ω

$$\omega = \frac{(\sum_{i=1}^p \lambda_i)^2}{(\sum_{i=1}^p \lambda_i)^2 + \sum_{i=1}^p \psi_{ii}}$$

può essere calcolato nel modo seguente:

```
(sum(fa$load))^2 / (sum((fa$load))^2 + sum(fa$uniq))
#> [1] 0.8245
```

Nel caso presente, il coefficiente di attendibilità $\omega = 0.82$ ci dice che l'82% della varianza del punteggio totale Y del test viene spiegato dal fattore comune latente.

9.5.1.2 Coefficiente ω e assunzioni della teoria classica dei test

Il calcolo di ω è basato sull'assunzione (tipica della teoria classica dei test) che $\psi_{ik} = 0$ per $i \neq k$. Tale assunzione però potrebbe non essere soddisfatta nel caso di dati empirici. In tal caso, come indicato da Bollen (1980), la (9.10) diventa

$$\omega = \frac{(\sum_{i=1}^p \lambda_i)^2}{(\sum_{i=1}^p \lambda_i)^2 + \sum_{i=1}^p \psi_{ii} + \sum_{i,k,i \neq k}^p \psi_{ik}}. \quad (9.10)$$

L'appropriatezza dell'assunzione dell'incorrelazione dei fattori specifici può essere verificata mediante un'analisi fattoriale confermativa. Se vi sono molte coppie di fattori specifici correlati, allora può essere necessario introdurre nel modello dei fattori aggiuntivi che rendano conto di queste covarianze. In questo caso, la scala non sarà più unidimensionale: la presenza di più fattori indica la presenza di più sottoscale. Il problema presentato sopra, tuttavia, non sempre può essere risolto individuando delle sottoscale perché, anche in tal caso, possono rimanere delle covarianze tra i fattori specifici che non sono spiegate dai fattori che individuano le sottoscale. In questi casi, per calcolare ω sarà necessario utilizzare la (9.10).

McDonald (1999) attribuisce al coefficiente ω le seguenti interpretazioni: ω è uguale al quadrato della correlazione tra la Y e il fattore comune ξ o, in maniera equivalente, tra la Y e il punteggio vero (in base alla

definizione di attendibilità: $\rho_{XT}^2 = \sigma_\tau^2 / \sigma_X^2$); ω è uguale alla correlazione tra due test Y e Y' aventi la stessa somma (o media) delle saturazioni nel modello ad un fattore e la stessa somma (o media) delle varianze specifiche nel modello ad un fattore; ω è uguale al quadrato della correlazione tra il punteggio totale di p item e il punteggio medio di un insieme infinito di item di un dominio omogeneo di cui i p item costituiscono un sottoinsieme.

9.5.2 Coefficiente α di Cronbach

Il coefficiente ω consente di stimare il coefficiente di attendibilità nel caso di un modello monofattoriale congenerico. Invece, il coefficiente α fornisce una stima del coefficiente di attendibilità nel caso di un modello con indicatori τ -equivalenti.

Se p item soddisfano il modello di τ -equivalenza, la varianza di ciascun item può essere scomposta in una componente attribuibile al valore vero e in una componente d'errore, come indicato nella (9.5), ovvero, $\sigma_{ii} = \lambda^2 + \psi_{ii} = \sigma_T^2 + \sigma_i^2$. In base al principio di τ -equivalenza, le varianze e covarianze riprodotte dal modello uni-fattoriale hanno le caratteristiche descritte nella matrice (9.6). Dato che tutti gli item hanno la stessa saturazione fattoriale λ , la formula per il calcolo del coefficiente ω si riduce a

$$\omega = \frac{(\sum_i \lambda_i)^2}{(\sum_i \lambda_i)^2 + \sum_i \psi_{ii}} = \frac{p^2 \lambda^2}{\sigma_Y^2} = \frac{p^2 \sigma_T^2}{\sigma_Y^2}$$

dove Y è il punteggio totale del test.

Usando il metodo dei minimi quadrati non pesati, una stima di ω può essere ottenuta nel modo seguente:

$$\hat{\omega} = \frac{p^2 \hat{\sigma}_T^2}{s_Y^2} \quad (9.11)$$

dove una stima di σ_T^2 viene fornita dalla (9.7), ovvero

$$\hat{\sigma}_T^2 = \frac{1}{p(p-1)} \sum_{i \neq k} s_{ik} \quad (9.12)$$

Inserendo la (9.12) nella (9.11), otteniamo

$$\hat{\omega} = \frac{p}{p-1} \frac{\sum \sum_{i \neq k} s_{ik}}{s_Y^2} \quad (9.13)$$

In conclusione, nel caso di indicatori τ -equivalenti, una stima del coefficiente ω è data da

$$\hat{\omega} = \frac{p}{p-1} \left(1 - \frac{\sum_i s_{ii}}{s_Y^2} \right) \quad (9.14)$$

La stima dell'attendibilità fornita dalla (9.14) trova il suo corrispettivo per i valori della popolazione nell'equazione seguente:

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\sum_{i=1}^p \sigma_{ii}}{\sigma_Y^2} \right) = \frac{p}{p-1} \frac{\sum_{i \neq k}^p \text{Cov}(X_i, X_k)}{\mathbb{V}(Y)} \quad (9.15)$$

La (9.15) definisce quello che è conosciuto come il coefficiente α .

Il coefficiente α fu scoperto da Guttman nel 1945 e incorrettamente attribuito a Cronbach. Viene spesso chiamato coefficiente α di Guttman-Cronbach, o G-C α .

Se gli indicatori soddisfano i requisiti del modello di τ -equivalenza, i coefficienti α e ω sono uguali. Se il modello di τ -equivalenza è appropriato, il coefficiente α fornisce un limite inferiore del coefficiente ω (ovvero, fornisce una sottostima di ω): $\omega \geq \alpha$. A causa del fatto che fornisce una stima conservativa del coefficiente di attendibilità, α viene preferito ad ω da alcuni ricercatori. Si noti però che α possiede tale carattere conservativo solo nel caso in cui le assunzioni del modello τ -equivalente siano soddisfatte.

9.5.2.1 Un esempio concreto

consideriamo nuovamente la matrice di varianze e covarianze SWLS. Il coefficiente α si calcola usando la (9.14) e, per i dati presenti, risulta essere uguale a

```
p <- 5
alpha <- (p / (p - 1)) * (1 - tr(SWLS) / sum(SWLS))
alpha
#> [1] 0.8191
```

Lo stesso risultato si ottiene utilizzando la funzione `alpha()` contenuta nel pacchetto `psych`:

```
alpha(SWLS)
#>
#> Reliability analysis
#> Call: alpha(x = SWLS)
#>
#>   raw_alpha std.alpha G6(smc) average_r S/N median_r
#>      0.82      0.82    0.79      0.48 4.6      0.49
#>
#> Reliability if an item is dropped:
#>   raw_alpha std.alpha G6(smc) average_r S/N var.r
#> V1      0.75      0.76    0.70      0.44 3.1 0.0027
#> V2      0.78      0.79    0.74      0.48 3.7 0.0060
#> V3      0.78      0.78    0.73      0.47 3.5 0.0055
#> V4      0.81      0.81    0.77      0.52 4.3 0.0027
#> V5      0.80      0.80    0.75      0.50 3.9 0.0057
#>   med.r
#> V1 0.44
#> V2 0.49
#> V3 0.48
#> V4 0.52
#> V5 0.50
#>
#> Item statistics
#>      r r.cor r.drop
#> V1 0.83 0.79 0.71
#> V2 0.77 0.68 0.62
#> V3 0.78 0.71 0.64
#> V4 0.70 0.58 0.53
#> V5 0.74 0.63 0.57
```

9.5.2.2 Violazione dell'assunto di tau-equivalenza

Il coefficiente α , la misura di attendibilità maggiormente usata in psicometria, è basato sull'assunzione che il modello di misurazione sia τ -equivalente. Come indicato sopra, se tale assunzione è soddisfatta, α fornisce un limite inferiore dell'attendibilità del test. Nei casi in cui tale assunzione venga violata, però, α può perdere tale carattere conservativo

e può fornire una *sovrastima* dell'attendibilità del test (Sijtsma, 2009).

NKano e Azuma (2003) riportano i risultati di una simulazione che mette in evidenza le conseguenze che risultano dalla violazione dell'assunzione di incorrelazione tra le componenti specifiche del modello monofattoriale. Questi autori trovano che, quando il principio dell'incorrelazione dei fattori specifici è violato, allora le stime dell'attendibilità ottenute mediante il coefficiente α sono affette da un errore sistematico. Tale errore sistematico aumenta all'aumentare del numero di coppie di fattori specifici che risultano tra loro correlati. In queste circostanze, dunque, il coefficiente α non fornisce più una stima conservativa dell'attendibilità.

In conclusione, il coefficiente ω fornisce una stima adeguata dell'attendibilità nel caso di un modello di misurazione congenerico. L'utilizzo del coefficiente α per la stima dell'attendibilità richiede un modello di misurazione τ -equivalente. L'esistenza di fattori specifici correlati invalida sia il coefficiente α , sia il coefficiente ω calcolato in base alla (9.10). In tali circostanze l'attendibilità deve essere stimata utilizzando una diversa equazione (Kano & Azuma, 2003; Komaroff, 1997).

Questa discussione mette in evidenza un aspetto importante: il coefficiente α fornisce una stima conservativa dell'attendibilità di un test solo se le variabili osservate sono associate alle variabili latenti come indicato dal modello di misurazione τ -equivalente. Se le assunzioni del modello τ -equivalente sono violate (per esempio, l'assunzione dell'incorrelazione degli errori), allora α porta ad una sovrastima della stima dell'attendibilità del test.

Sijtsma (2009), tra gli altri, sconsiglia l'uso di α per la stima dell'attendibilità del test in quanto, nelle applicazioni reali, *le assunzioni di τ -equivalenza e dell'incorrelazione degli errori risultano spesso violate*. La violazione dell'assunzione di τ -equivalenza porta ad una stima conservativa dell'attendibilità, mentre la violazione dell'assunzione dell'incorrelazione degli errori porta ad una stima liberale dell'attendibilità. In entrambi i casi, l'errore sistematico può essere sostanziale.

Un secondo problema è che α viene spesso preso quale misura della “struttura interna” di un test e quindi come evidenza che gli item del test “misurino la stessa cosa.” Tale interpretazione di α è sbagliata, in quanto α non fornisce alcuna informazione a questo proposito. Non è semplice fornire ad α una chiara interpretazione, anche nel caso in cui siano soddisfatte le assunzioni del modello di misurazione su cui si basa.

9.5.3 La formula “profetica” di Spearman-Brown

L'attendibilità può essere stimata mediante il coefficiente ω se il modello di misurazione è congenerico e tramite il coefficiente α , se il modello di misurazione è τ -equivalente. Chiediamoci ora come possa essere misurata l'attendibilità nel caso di un modello di misurazione costituito da indicatori paralleli.

Si considerino p item paralleli, tali per cui $\lambda_1 = \lambda_2 = \dots = \lambda_p = \lambda$ e $\psi_{11} = \psi_{22} = \dots = \psi_{pp} = \psi$. In tal caso, la quota di varianza del punteggio totale del test che viene spiegata dalla variabile latente è uguale a

$$\left(\sum_i \lambda_i \right)^2 = (p\lambda)^2 = p^2\lambda^2.$$

L'attendibilità di un singolo item è data da

$$\rho_1 = \frac{\lambda^2}{\lambda^2 + \psi} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}.$$

Per p item paralleli abbiamo

$$\begin{aligned} \rho_p &= \frac{p^2\lambda^2}{p^2\lambda^2 + p\psi} \\ &= \frac{p^2\lambda^2}{p(p\lambda^2 + \psi)} \\ &= \frac{p\lambda^2}{p\lambda^2 + \psi} \\ &= \frac{p\lambda^2}{(p-1)\lambda^2 + (\lambda^2 + \psi)} \end{aligned}$$

ovvero,

$$\begin{aligned} \rho_p &= \frac{p \frac{\lambda^2}{\lambda^2 + \psi}}{(p-1) \frac{\lambda^2}{\lambda^2 + \psi} + \frac{\lambda^2 + \psi}{\lambda^2 + \psi}} \\ &= \frac{p\rho_1}{(p-1)\rho_1 + 1}, \end{aligned}$$

ricordando che l'attendibilità di ciascun singolo item è $\rho_1 = \frac{\lambda^2}{\lambda^2 + \psi}$.

La (9.16) esprime l'attendibilità di un test costituito da p item paralleli nei termini dell'attendibilità di un solo item ed è tradizionalmente conosciuta come la formula di Spearman-Brown (*Spearman-Brown prophecy formula*). Nel caso di item paralleli si ha che

$$\omega = \alpha = \rho_p.$$

9.5.3.1 Un esempio concreto

Poniamoci ora il problema di calcolare l'attendibilità del test SWLS ipotizzando che gli item siano paralleli e utilizzando la formula di Spearman-Brown. La matrice di correlazione è:

```
R <- cov2cor(SWLS)
round(R, 3)
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 1.000 0.563 0.589 0.498 0.521
#> [2,] 0.563 1.000 0.511 0.400 0.452
#> [3,] 0.589 0.511 1.000 0.419 0.473
#> [4,] 0.498 0.400 0.419 1.000 0.370
#> [5,] 0.521 0.452 0.473 0.370 1.000
```

Seguendo McDonald (1999), supponiamo di calcolare l'attendibilità di un singolo item (ρ_1) come la correlazione media tra gli item:

```
rr <- NULL
k <- 1
for (i in 1:p) {
  for (j in 1:p) {
    if (j != i) {
      rr[k] <- R[i, j]
    }
    k <- k + 1
  }
}
ro_1 <- mean(rr, na.rm = TRUE)
ro_1
#> [1] 0.4798
```


Applicando la formula di Spearman-Brown, la stima dell'attendibilità del test diventa pari a

```
(p * ro_1) / ((p - 1) * ro_1 + 1)
#> [1] 0.8218
```

9.5.3.2 Attenuazione

All'aumentare dell'errore di misurazione, la correlazione tra due variabili tende a diminuire. L'errore di misurazione, dunque, “maschera” l'associazione esistente tra le variabili. Tale fenomeno va sotto il nome di *attenuazione*.

Lord e Novick (1967) notano che, volendo determinare la relazione esistente tra due costrutti, uno psicologo può costruire opportune scale per misurarli. Se la relazione tra queste scale è lineare, allora il grado di associazione tra le scale può essere misurato dal coefficiente di correlazione. Le scale, però, contengono una componente di errore e, quindi, la correlazione empirica tra le due scale assume un valore minore della “reale” correlazione tra i costrutti. In tali circostanze, possono essere usate opportune formule per stimare il valore della correlazione disattenuata tra i tratti latenti.

Si può dimostrare che la correlazione tra i punteggi veri di due costrutti, T_x e T_y , può essere espressa nei termini della correlazione ρ_{XY} tra i punteggi osservati X e Y , e nei termini dei coefficienti di attendibilità $\rho_{XX'}$, $\rho_{YY'}$ dei due test:

$$\rho(T_X, T_Y) = \frac{\rho_{XY}}{\sqrt{\rho_{XX'}\rho_{YY'}}} \quad (9.16)$$

Inoltre, può essere dimostrato che la correlazione tra i punteggi di un test e i punteggi veri di un secondo test può essere espressa nei termini delle correlazioni tra i punteggi osservati dei due test e del coefficiente di attendibilità del secondo test:

$$\rho(X, T_Y) = \frac{\rho_{XY}}{\sqrt{\rho_{YY'}}}. \quad (9.17)$$

9.5.3.3 Correlazioni disattenuate

Le (9.16) e (9.17) consentono di calcolare le cosiddette *correlazioni disattenuate*. L'idea è che le correlazioni tra i punteggi veri di due test

sono sottostimate dalle correlazioni tra i punteggi osservati dei test, a causa dell'errore di misura. Se le attendibilità dei test sono conosciute, le (9.16) e (9.17) possono essere usate per stimare le correlazioni tra i corrispondenti punteggi veri. La teoria dell'attenuazione costituisce un'ulteriore applicazione del coefficiente di attendibilità nell'ambito della teoria classica dei test.

Le correlazioni disattenuate sono state usate già a partire dal 1904 da Spearman. Nell'esempio di Spearman, X era una misura di discriminazione dell'altezza di un suono (*pitch discrimination*) e Y era una misura di intelligenza fornita da un insegnante. La correlazione tra queste due misure era $\hat{\rho}_{XY} = 0.38$. Le attendibilità delle due misure erano pari a, rispettivamente, $\hat{\rho}_{XX'} = 0.25$ e $\hat{\rho}_{YY'} = 0.55$. In base alla (9.17)

$$\rho(X, T_Y) = \frac{\rho_{XY}}{\sqrt{\rho_{YY'}}}$$

la correlazione predetta tra i valori veri di pitch discrimination e i valori empirici dell'intelligenza è

$$\hat{\rho}(X, T_Y) = \frac{0.38}{\sqrt{0.25}} = 0.76.$$

In base alla (9.16)

$$\rho(T_X, T_Y) = \frac{\rho_{XY}}{\sqrt{\rho_{XX'}\rho_{YY'}}}$$

la correlazione tra i valori veri di pitch discrimination e i valori veri dell'intelligenza è

$$\hat{\rho}(T_X, T_Y) = \frac{0.38}{\sqrt{0.25 \times 0.55}} = 1.025.$$

Si noti come i limiti di questa procedura emergano già dall'esempio fornito da Spearman: le correlazioni disattenuate possono facilmente produrre una sovrastima.

Questa formula originò una controversia tra Charles Spearman e Karl Pearson. In un suo articolo del 1904 (lo stesso anno dei famosi articoli di Spearman), Pearson riportò diverse correlazioni nell'intorno di 0.5 che riguardavano la misurazione empirica di caratteristiche quali la vivacità

e l'introspezione. Spearman criticò l'articolo di Pearson affermando che le osservazioni probabilmente contenevano un sostanziale errore di misurazione, il che determinava il fatto che fossero così basse. Le corrispondenti correlazioni disattenuate erano, secondo Spearman, probabilmente molto più alte. Tale critica venne del tutto ignorata da Pearson sulla base del fatto che la formula di Spearman poteva condurre a correlazioni maggiori di uno. Inoltre, Pearson non accettava i riferimenti a quantità inosservabili. Spearman, d'altra parte, eseguì diversi studi su variabili psicologiche alle quali applicò la sua formula per le correlazioni disattenuate. In molti casi, trovò che le correlazioni disattenuate erano vicine ad uno. Questo suggeriva che tali variabili psicologiche erano indicatori dello stesso fenomeno. Queste considerazioni spinsero Spearman a procedere in questa direzione, giungendo ad inventare l'analisi fattoriale così com'è riportata nell'articolo del 1904 *“General intelligence”, objectively determined and measured*.

McDonald (1999) afferma che le correlazioni disattenuate devono essere usate con cautela. Un metodo migliore per calcolare le correlazioni tra le variabili latenti (ovvero, le correlazioni non “inquinata” dagli errori di misura) è quello di costruire un modello di equazioni strutturali nel quale diverse ipotesi possono essere direttamente verificate, compresa quella della correlazione tra le variabili latenti.

9.6 Attendibilità e scala di misura

McDonald (2013) fa notare che i coefficienti ω e α , ma non il coefficiente di Spearman-Brown, dipendono dalla scala di misura degli item. Stimare ω utilizzando una matrice di correlazione anziché una matrice di varianze e di covarianze è equivalente a stimare il coefficiente di attendibilità di una somma di item standardizzati. Il risultato ottenuto mediante ω e α non si generalizza però al caso in cui si voglia valutare l'attendibilità del punteggio totale di un test calcolato sui valori grezzi degli item.

Il modello ad un fattore comune non dipende dall'unità di misura degli indicatori e può essere adattato sia ad una matrice di correlazione sia ad una matrice di varianze e di covarianze. Il calcolo dei coefficienti ω e α , invece, deve essere fatto sulla soluzione trovata utilizzando una matrice di varianze e di covarianze.

9.7 Quale indice usare?

L'indice di attendibilità più diffuso in letteratura è il coefficiente α di Cronbach. Affinché α fornisca una stima dell'attendibilità del test, però, gli item devono essere τ -equivalenti. Il modello di τ -equivalenza richiede l'unidimensionalità del tratto latente. In pratica, tale assunzione viene spesso violata, dato che la maggior parte dei test, oltre ad un fattore generale, misurano anche altri fattori. Anche nel caso di un test unidimensionale, le comunaltà degli item non sono mai uguali tra loro, violando così l'assunzione di τ -equivalenza. In tali circostanze, se risulta soddisfatta l'assunzione di incorrelazione degli errori, il coefficiente α sottostima l'attendibilità del test. Se invece l'assunzione di incorrelazione degli errori non risulta soddisfatta, allora il coefficiente α sovrastima l'attendibilità del test. Per tali ragioni, l'utilità del coefficiente α di Cronbach è molto limitata e, in generale, è preferibile usare il coefficiente ω (McDonald, 1999). Altre alternative sono gli indici *glb* (*Greatest Lower Bound*; si veda, ad esempio, Ten Berge e Sočan, 2004) e β (Revelle, 1979).

10

CFA: confronto tra modelli

In un modello CFA, i parametri possono essere stimati senza vincoli, possono essere fissi o possono essere stimati sulla base di alcuni vincoli. Un parametro libero è sconosciuto e il ricercatore consente all'algoritmo di stima di trovare il suo valore ottimale che, insieme agli altri parametri del modello, riduce al minimo le differenze tra le matrici di varianze-covarianze osservate e quelle predette dal modello. Un parametro fisso è pre-specificato dal ricercatore ad un valore specifico, più comunemente 1.0 (ad esempio, per definire la metrica di una variabile latente) o 0 (ad esempio, l'assenza di saturazioni fattoriali o di covarianze di errore). Come per un parametro libero, anche un parametro vincolato è sconosciuto; tuttavia, un tale parametro non può assumere un valore qualsiasi, ma deve rispettare le restrizioni sui suoi valori che il ricercatore ha imposto. I vincoli più comuni sono i vincoli di uguaglianza, in cui i parametri non standardizzati devono assumere valori uguali (ad esempio, in diversi gruppi).

Consideriamo un esempio discusso da [Brown \(2015\)](#). Viene qui considerato un set di dati in cui le prime tre misure osservate (X1, X2, X3) sono indicatori di un costrutto latente corrispondente alla Memoria uditiva e le seconde tre misure (X4, X5, X6) sono indicatori di un altro costrutto latente, Memoria visiva. Le tre misure usate quali indicatori del costrutto di memoria uditiva sono: X1 = memoria logica, X2 = associazione verbale a coppie, X3 = liste di parole; le tre misure usate come indicatori del costrutto di memoria visiva sono: X4 = immagini di facce, X5 = foto di famiglia, X6 = generiche riproduzioni visive. I dati sono i seguenti:

```
sds <- "2.610  2.660  2.590  1.940  2.030  2.050"

cors <- "
1.000
```

```

0.661  1.000
0.630  0.643  1.000
0.270  0.300  0.268  1.000
0.297  0.265  0.225  0.805  1.000
0.290  0.287  0.248  0.796  0.779  1.000"

covs <- getCov(cors, sds = sds, names = paste("x", 1:6, sep = ""))

```

Adattiamo i cinque modelli discussi da [Brown \(2015\)](#).

10.1 Modello congenerico

```

model.congeneric <- "
  auditorymemory =~ x1 + x2 + x3
  visualmemory   =~ x4 + x5 + x6
"

```

```

fit.congeneric <- cfa(
  model.congeneric,
  sample.cov = covs,
  sample.nobs = 200,
  std.lv = TRUE
)

```

L'output (qui non fornito) si ottiene con:

```

summary(
  fit.congeneric,
  fit.measures = TRUE,
  standardized = TRUE,
  rsquare = TRUE
)

```

10.2 Modello tau-equivalente

Solo memoria auditiva:

```
model.tau.a <- "  
  auditorymemory =~ x1 + v1*x1 + v1*x2 + v1*x3  
  visualmemory   =~ x4 + x5 + x6  
"
```

```
fit.tau.a <- cfa(  
  model.tau.a,  
  sample.cov = covs,  
  sample.nobs = 200,  
  std.lv = TRUE  
)
```

Memoria auditiva e visiva:

```
model.tau.av <- "  
  auditorymemory =~ x1 + v1*x1 + v1*x2 + v1*x3  
  visualmemory   =~ x4 + v2*x4 + v2*x5 + v2*x6  
"
```

```
fit.tau.av <- cfa(  
  model.tau.av,  
  sample.cov = covs,  
  sample.nobs = 200,  
  std.lv = TRUE  
)
```

10.3 Modello parallelo

Solo memoria auditiva:

```
model.parallel.a <- "  
  auditorymemory =~ x1 + v1*x1 + v1*x2 + v1*x3  
  visualmemory   =~ x4 + v2*x4 + v2*x5 + v2*x6  
  x1 ~~ v3 * x1  
  x2 ~~ v3 * x2  
  x3 ~~ v3 * x3  
"
```

```
fit.parallel.a <- cfa(  
  model.parallel.a,  
  sample.cov = covs,  
  sample.nobs = 200,  
  std.lv = TRUE  
)
```

Memoria auditiva e visiva:

```
model.parallel.av <- "  
  auditorymemory =~ x1 + v1*x1 + v1*x2 + v1*x3  
  visualmemory   =~ x4 + v2*x4 + v2*x5 + v2*x6  
  x1 ~~ v3 * x1  
  x2 ~~ v3 * x2  
  x3 ~~ v3 * x3  
  
  x4 ~~ v4 * x4  
  x5 ~~ v4 * x5  
  x6 ~~ v4 * x6  
"
```

```
fit.parallel.av <- cfa(  
  model.parallel.av,  
  sample.cov = covs,  
  sample.nobs = 200,  
  std.lv = TRUE  
)
```


10.4 Il test del χ^2

Il confronto tra modelli nidificati procede attraverso il test χ^2 . Tale test si basa su una proprietà delle variabili casuali distribuite come χ^2 : la differenza tra due v.c. X_1 e X_2 che seguono la distribuzione χ^2 , rispettivamente con ν_1 e ν_2 , con $\nu_1 > \nu_2$, è una variabile casuale che segue la distribuzione χ^2 con gradi di libertà pari a $\nu_1 - \nu_2$.

Un modello nidificato è un modello che impone dei vincoli sui parametri del modello di partenza. L'imposizione di vincoli sui parametri ha la conseguenza che vi sarà un numero minore di parametri da stimare. Il confronto tra i modelli si esegue valutando in maniera relativa la bontà di adattamento di ciascun modello per mezzo della statistica chi-quadrato. La statistica così calcolata avrà un numero di gradi di libertà uguale alla differenza tra i gradi di libertà dei due modelli.

Nel caso dell'esempio in discussione, abbiamo

```
anova(
  fit.congeneric,
  fit.tau.a,
  fit.tau.av,
  fit.parallel.a,
  fit.parallel.av,
  test = "chisq"
)
#> Chi-Squared Difference Test
#>
#>           Df  AIC  BIC Chisq Chisq diff Df diff
#> fit.congeneric    8 4702 4745  4.88
#> fit.tau.a         10 4699 4735  5.66      0.78      2
#> fit.tau.av        12 4695 4725  5.88      0.22      2
#> fit.parallel.a   14 4691 4714  5.98      0.10      2
#> fit.parallel.av  16 4690 4707  9.28      3.30      2
#>           Pr(>Chisq)
#> fit.congeneric
#> fit.tau.a           0.68
#> fit.tau.av           0.90
```

```
#> fit.parallel.a      0.95  
#> fit.parallel.av     0.19
```

I test precedenti indicano come non vi sia una perdita di adattamento passando dal modello congenerico al modello più restrittivo (ovvero, il modello parallelo per entrambi i fattori). Per questi dati, dunque, può essere adottato il modello più semplice, cioè il modello parallelo.

Parte III

Costruzione di strumenti psicometrici



11

Tipologie dei test psicometrici

Prima di iniziare lo sviluppo di un test psicometrico, il ricercatore deve decidere quale tipologia di strumento sia più utile per affrontare il problema che ha di fronte. Possiamo infatti distinguere tra test orientati al criterio e test riferiti alla norma.

11.0.1 Test orientati al criterio

I test orientati al criterio hanno quale scopo il confronto fra gruppi precostituiti di individui. Gli item del test vengono selezionati in base alla loro capacità empirica di discriminare fra gruppi criterio, ad esempio, malati/sani, bocciati/promossi, schizofrenici/depressi. I test orientati al criterio sono costruiti utilizzando metodi empirici e non teorici. Al vantaggio di una chiara utilità pratica si accompagna il grande svantaggio di identificare fattori aventi una scarsa validità di costrutto, i quali risultano inutili per la comprensione dei processi psicologici.

Il processo di sviluppo della scala è semplice, in quanto si devono selezionare gli item che mostrano punteggi differenti in gruppi-criterio noti. Se i gruppi criterio possono essere individuati con chiarezza è sempre teoricamente possibile sviluppare test in grado di discriminarli.

Tuttavia, non sempre i gruppi possono essere definiti in modo attendibile; oppure, la definizione dei gruppi criterio potrebbe avere senso solo all'interno di una teoria, ma non sia generalizzabile ad altre tradizioni teoriche. In questo caso, il test rischia di essere eccessivamente specifico, dimostrandosi utile solo nelle condizioni per cui è stato sviluppato, ma con scarsa capacità di potere essere utilizzato in condizioni diverse.

Lo svantaggio principale dei test orientati al criterio è che il significato psicologico dei punteggi è ignoto. Non avendo una teoria sulle variabili psicologiche che distinguono due gruppi, un buon test discriminante non ci aiuta a capire perché tali gruppi siano diversi. Non è possibile sapere quanti costrutti siano coinvolti nella determinazione di un pun-

teggio. Inoltre, due punteggi uguali non implicano la presenza dei medesime meccanismi psicologici. Date queste ambiguità, utilizzando questi test non è possibile neppure aumentare le nostre conoscenze in maniera incrementale.

Il problema maggiore per lo sviluppo di questi strumenti è la definizione del criterio: qual è la variabile numerica che discrimina nella maniera maggiore tra i gruppi in esame (malati/sani)?

La batteria di item iniziale deve essere sufficientemente grande e non è necessario che gli item abbiano validità di contenuto o validità di facciata. In generale, la batteria di item deve essere più grande di quelle usate per il metodo fattoriale o di analisi degli item: mancando criteri teorici per la scelta degli item, la scelta iniziale degli item è molto arbitraria ed è dunque necessario partire da un numero molto elevato di item. Ciò è meno vero quando gli item hanno una certa validità di facciata o di contenuto. In seguito, semplicemente si selezionano gli item che discriminano efficacemente fra i gruppi, o gli item fortemente associati con il punteggio criterio. È necessario poi replicare su un diverso campione la capacità discriminativa degli item selezionati.

11.0.2 Test basati sulla norma

Le misure riferite ad una norma indicano la posizione del rispondente in riferimento alla distribuzione di punteggi ottenuti nello stesso test da un campione di grandi dimensioni e rappresentativo della popolazione di riferimento. La maggioranza dei test di personalità, attitudinali e cognitivi sono test basati sulla norma.

La metrica utilizzata per tale confronto può avere caratteristiche diverse. I punteggi standardizzati (con media 0 e varianza unitaria) calcolati rispetto al gruppo di riferimento sono spesso convertiti in una scala diversa, per esempio aventi media pari a 500 e deviazione standard di 100 (punteggi SAT), o aventi media pari a 100 e deviazione standard di 15 (es. punteggi WAIS-VI). È facile operare tale trasformazione. Il punteggio Y di un rispondente può essere trasformato nel modo seguente in un punteggio standard X_i , avente una media target pari a μ_s e una deviazione standard σ_s

$$X_i = \mu_s + z_i \sigma_s$$

dove z è il punteggio standardizzato $z = \frac{Y - \bar{Y}}{s_Y}$.

11.1 Variabili latenti e sviluppo di uno strumento psicometrico

Quando uno psicologo sviluppa una scala di misura è meno interessato agli item della scala che ai costrutti che si intendono misurare: “Scale items are usually a means to the end of construct assessment ...they are necessary because many constructs cannot be assessed directly” (p. 2). Dato che non sono osservabili direttamente, i costrutti sono detti *variabili latenti*. I costrutti sono interpretati come le cause (non visibili) che fanno in modo che gli item assumano un determinato valore per un determinato rispondente in un certo momento del tempo. Mentre alcune variabili, quali l'altezza, il peso, il battito cardiaco, la temperatura, possono essere misurate direttamente, i costrutti psicologici, quali l'ansia, la personalità, la qualità della vita, possono solo essere misurati indirettamente esaminando gli effetti che hanno sugli indicatori osservabili del costrutto. Gli item che vengono misurati tramite uno strumento sono gli indicatori osservati o empirici degli attributi del costrutto. Il dolore, ad esempio, è un costrutto psicologico non direttamente osservabile. Tuttavia, al dolore sono associati molteplici indicatori che sono direttamente osservabili, quali il pallore, la sudorazione profusa, ecc.

Allo scopo di misurare le variabili latenti del costrutto di interesse, lo psicologo deve identificare gli indicatori empirici del costrutto che possono essere direttamente osservabili. L'identificazione di tali indicatori empirici avviene attraverso (1) la chiarificazione del costrutto di interesse, (2) l'operazionalizzazione del costrutto, (3) la rassegna della letteratura rilevante, (4) l'analisi concettuale del costrutto.

11.1.1 Chiarificazione del costrutto di interesse

Vi sono diverse domande a cui lo psicologo deve rispondere prima di iniziare la selezione degli item, altrimenti rischia di produrre uno strumento con una scarsa validità di costrutto.

1. Qual è lo scopo dello strumento? Che cosa lo strumento dovrà misurare?

2. Quali altri costrutti sono associati al costrutto di interesse? In che misura essi si distinguono dal costrutto di interesse? A tali domande non è semplice dare una risposta se il costrutto di interesse è complesso e astratto.
3. Lo strumento da costruire intende misurare le caratteristiche generali del costrutto di interesse, o intende focalizzarsi su alcuni specifici aspetti del costrutto?

Esempio 11.1. Watson et al. (2007) si sono posti il problema di costruire uno strumento atto misurare la depressione superando i limiti degli strumenti già esistenti, quali il Beck Depression Inventory—II (BDI—II; Beck, Steer, & Brown, 1996) e il Center for Epidemiological Studies Depression Scale (CES—D; Radloff, 1977). La scala costruita dagli autori prende il nome di Inventory of Depression and Anxiety Symptoms (IDAS).

Per rispondere alla prima domanda, Watson et al. (2007) fanno notare che gli strumenti esistenti comprendono contenuti non specifici, ovvero non direttamente associati alla depressione. Infatti, sia il BDI-2 sia il CES—D contengono item che fanno riferimento a vari tipi di ansia. Di conseguenza, la validità discriminante di questi strumenti risulta compromessa. Inoltre, gli strumenti esistenti non contengono item che coprono tutto il dominio del costrutto della depressione maggiore, così come specificato dal Diagnostic and Statistical Manual of Mental Disorders (4th ed.). Infine, un'altra limitazione degli strumenti esistenti è il fatto che essi sono stati creati per produrre un singolo item della severità dei sintomi e quindi ignorano l'eterogeneità e la multidimensionalità del fenomeno depressivo. Questo si riflette sul fatto che gli strumenti esistenti manifestano una struttura fattoriale poco chiara, nel senso che autori diverse hanno trovato soluzioni fattoriali diverse. Lo strumento che Watson et al. (2007) intendono sviluppare vuole superare queste difficoltà costruendo una scala che direttamente rifletta, in ciascuna delle sue sottoscale, gli aspetti distintivi della depressione, a differenza di quanto accade per gli strumenti BDI—II e CES—D.

Per rispondere alla seconda domanda, Watson et al. (2007) fanno notare come la depressione sia inserita in una rete nomologica di costrutti che include, in primo luogo, l'ansia. Diversamente dagli strumenti già esistenti, BDI—II e CES—D, Watson et al. (2007) si propongono esplicitamente di creare scale che riflettano gli aspetti specifici della depressione, distinti dall'ansia. Per fare questo, Watson et al. (2007) iniziano con il considera-

re un ampio insieme di item che rappresentano sintomi associati all'ansia. In questo modo viene perseguito l'obiettivo, all'interno dello strumento, di esaminare la relazione tra i sintomi d'ansia e quelli della depressione in modo da creare scale distinte per tali dimensioni così da aumentare a validità discriminante dello strumento.

Per rispondere alla terza domanda, Watson et al. (2007) affermano di volere sviluppare uno strumento che, nel suo punteggio generale, rifletta le caratteristiche generali della depressione mentre, quando vengono considerate le varie sottoscale che lo costituiscono, consente di misurare con precisione ciascuna delle dimensioni del costrutto esaminato.

11.1.2 Operazionalizzazione del costrutto di interesse

La definizione concettuale fornisce il significato teorico generale del costrutto. L'operazionalizzazione è invece una definizione del costrutto che ne consenta la misurazione (Vogt, 1993). Gli indicatori osservabili o empirici sono il prodotto finale di tale processo di operazionalizzazione (Keck, 1998) e diventano gli item dello strumento. Se il costrutto di interesse è stato sviluppato all'interno di un approccio teorico ben articolato, allora diventa più semplice stabilire quali siano le dimensioni che caratterizzano il costrutto, in che modo esse si possano manifestare, e in che modo possano essere misurate. Tuttavia, molti costrutti psicologici vengono spesso caratterizzati in maniera diversa da approcci teorici differenti.

Esempio 11.2. Per chiarire il costrutto di depressione, Watson et al. (2007) fanno riferimento al DSM-IV il quale elenca nove criteri sintomatici per un episodio depressivo maggiore: (1) umore depresso per la maggior parte del giorno, quasi ogni giorno, come riportato dal soggetto o come osservato dagli altri, (2) marcata diminuzione di interesse o piacere per tutte, o quasi tutte, le attività per la maggior parte del giorno, quasi ogni giorno (come riportato dal soggetto o come osservato dagli altri), (3) significativa perdita di peso, senza essere a dieta, o aumento di peso, oppure diminuzione o aumento dell'appetito quasi ogni giorno, (4) insonnia o ipersonnia quasi ogni giorno, (5) agitazione o rallentamento psicomotorio quasi ogni giorno (osservabile dagli altri, non semplicemente sentimenti soggettivi di essere irrequieto o rallentato), (6) faticabilità o mancanza di energia quasi ogni giorno, (7) sentimenti di autosvalutazione o di colpa eccessivi o inappropriati (che possono essere deliranti), quasi ogni giorno, (8) ridotta capacità di pensare o di concentrarsi, o indecisione, quasi ogni giorno (come impressione soggettiva o osservata

dagli altri), (9) pensieri ricorrenti di morte, ricorrente ideazione suicidaria senza un piano specifico, o un tentativo di suicidio, o l'ideazione di un piano specifico per commettere suicidio.

Per massimizzare l'utilità dell'IDAS, Watson et al. (2007) includono item molteplici per ciascuno dei nove criteri sintomatici per un episodio depressivo maggiore. Allo scopo di assicurare che un numero sufficiente di indicatori venga incluso nello strumento per ciascuna di queste dimensioni potenziali, nell'insieme di item preso in considerazione inizialmente, Watson et al. (2007) organizzano gli item potenziali in gruppi chiamati *homogeneous item composites* (HIC). Essi fanno comunque notare come la costruzione di questi HIC non forza l'emergenza di un corrispondente fattore, ma soltanto consente di campionare tutto il dominio potenziale del costrutto.

11.1.3 Rassegna della letteratura rilevante

È importante per lo psicologo conoscere la maggior parte possibile della letteratura rilevante prima di iniziare il processo di costruzione di un nuovo strumento. Una sistematica rassegna della letteratura consente allo psicologo di valutare e organizzare i risultati empirici provenienti da fonti diverse che sono utili per individuare i potenziali indicatori empirici del costrutto. La rassegna della letteratura consente di sintetizzare le scoperte in un campo di ricerca, mette in evidenza gli aspetti metodologici associati al costrutto di interesse, chiarisce quali siano gli approcci teorici all'interno dei quali il costrutto è stato discusso e consente di mettere in evidenza, quando è opportuno, la "dimensione dell'effetto" attraverso le meta-analisi.

Esempio 11.3. Nel caso dell'articolo di Watson et al. (2007), gran parte dell'introduzione è dedicata alla rassegna della letteratura che viene discussa allo scopo di mettere in evidenza i limiti degli strumenti esistenti, considerare quali sono le caratteristiche degli item utilizzati, mettere in relazione gli indicatori utilizzati dagli strumenti esistenti con gli approcci teorici disponibili in relazione alla depressione e all'ansia, discutere le soluzioni fattoriali che sono state ottenute dai dati raccolti tramite gli strumenti esistenti, considerare quali aree di contenuto del costrutto non sono state adeguatamente indagate dagli strumenti esistenti.

11.1.4 Analisi concettuale del costrutto

L'analisi concettuale del costrutto è un altro metodo che può essere usato per determinare gli indicatori empirici del costrutto di interesse.

È necessario stabilire quali siano gli attributi del costrutto di interesse, includendo la specificazione degli antecedenti e delle conseguenze che derivano da esso. Si devono identificare tutti gli usi che, nella letteratura specialistica, sono stati fatti del costrutto in esame. Infine, è necessario elencare tutti gli indicatori empirici che siano mai stati utilizzati per il costrutto esaminato.

Esempio 11.4. Allo scopo di campionare efficacemente l'intero dominio del costrutto, Watson et al. (2007) hanno definito 20 HIC: Depressed Mood, Loss of Interest or Pleasure, Appetite Disturbance, Sleep Disturbance, Psychomotor Problems, Fatigue/Anergia, Worthlessness/Guilt, Cognitive Problems, Suicidal Ideation, Hopelessness, Melancholic Depression, Angry/Irritable Mood, High Energy/High Positive Affect, Anxious Mood, Worry, Panic, Agoraphobia, Social Anxiety, Traumatic Intrusions, Obsessive-Compulsive Symptoms.

Tredici HIC (per un totale di 117 item) raggruppavano gli indicatori rilevanti per la depressione. Tra questi, nove HICs (per un totale di 79 items) facevano riferimento ai sintomi di base della depressione maggiore così come descritta nel DSM-IV (depressed mood, loss of interest or pleasure, appetite disturbance, sleep disturbance, psychomotor problems, fatigue/anergia, worthlessness and guilt, cognitive problems, suicidal ideation). I quattro rimanenti HIC facevano riferimento alla presenza di sintomi della Hopelessness (Abramson, Metalsky, & Alloy, 1989), ai sintomi specifici della depressione malinconica (Joiner et al., 2005), allo stato d'animo di rabbia/irritabilità (il quale rappresenta una forma alternativa di depressione tra gli adolescenti; DSM-IV, American Psychiatric Association, 1994, p. 327), e infine ad indicatori di energia e affetto positivo (i quali sono stati specificamente associati alla depressione; Mineka et al., 1998).

Gli altri sette HIC (per un totale di 63 item) sono stati introdotti per valutare sintomi associati all'ansia. Essi sono stati raggruppati nei termini dello stato d'animo ansioso, della worry, del panico, dell'agorafobia, dell'ansia sociale e delle intrusioni traumatiche associate al PTSD.

11.1.5 Metodi di ricerca qualitativi

Metodi di ricerca qualitativi posso anche essere usati allo scopo di identificare i potenziali indicatori empirici del costrutto. In particolare, possono essere usati i metodi della ricerca fenomenologica, dell'indagine naturalistica, i focus group e lo studio del caso singolo.

L'indagine fenomenologica is pone l'obiettivo di descrivere il costrutto dal punto di vista di chi fa di esperienza di esso (Carpenter, 1999). Utili a questo proposito sono ovviamente le descrizioni che i soggetti forniscono della propria esperienza.

Nell'indagine naturalistica, lo psicologo osserva le conseguenze del costrutto così come si manifestano nel mondo naturale. Uno strumento possibile di raccolta dati è l'intervista con il paziente.

Il focus group, originariamente sviluppato in ambito economico per ottenere opinioni su un determinato prodotto (Morse & Field, 1995), ha le caratteristiche di 'a semi-structured group session, moderated by a group leader, held in an informal setting with the purpose of collecting information on a designated topic' (Carry, 1994, p. 226).

Un'altra possibile fonte di informazioni è costituita dagli studi sul caso singolo.

11.2 Lo sviluppo dello strumento

Una volta selezionati gli indicatori empirici del costrutto, deve essere scelta una modalità di presentazione che consenta la raccolta efficiente dei dati. Ciascuno strumento può essere descritto in base a sei caratteristiche: (1) formato, (2) composizione tipografica, (3) istruzioni ai soggetti, (4) la costruzione degli item, (5) formato di risposta, e (6) numero di item.

11.2.1 Formato

I formati di scala più usati sono lo scaling Thurstoniano, lo scaling di Guttman, le scale a differenziale semantico, le scale di valutazione grafica, semantic differential scales, graphic rating scales, le scale visive di tipo analogico (visual analog scales) e le scale Likert. Ci concentriamo

qui sulle scale Likert per la loro importanza nei test psicometrici basati sull'analisi fattoriale.

11.2.1.1 Scala Likert

Sviluppata nel 1932 da Rensis Likert per misurare gli atteggiamenti, una scala Likert è una scala ordinale usata dai rispondenti per valutare il grado di accordo o disaccordo con l'affermazione che viene loro proposta. Di solito le alternative di risposta sono cinque o sette, da “molto d'accordo” a “fortemente contrario.”

Essendo una scala ordinale, le risposte possono essere ordinate, ma le distanze tra i livelli della scala non sono quantificabili. Quindi le distanze tra i livelli “sempre,” “spesso” e “talvolta” non sono necessariamente uguali. In altri termini, non possiamo assumere che le differenze tra i livelli di risposta siano equidistanti anche se le differenze tra i valori numerici assegnati ai livelli della scala lo sono.

C'è una lunga controversia sulla possibilità di trattare i valori numerici di una scala ordinale come se essi provenissero da una scala ad intervalli. In altri termini, ci si è chiesti se sia appropriato usare statistiche descrittive quali la media e la deviazione standard per i dati a questo livello di scala, e ci si è chiesti se sia appropriato usare i test parametrici per dati a livello di scala Likert. È risaputo che i test non parametrici, i quali non fanno assunzioni sulla forma della distribuzione della popolazione da cui abbiamo campionato i dati, hanno una potenza statistica nettamente inferiore ai test parametrici. Inoltre, concetti quali quelli di media e varianza non hanno senso se i livelli di una scala Likert non vengono considerati a livello di scala ad intervalli. Per queste ragioni alcuni autori ritengono problematico non potere trattare i dati provenienti da scale di tipo Likert come se fossero a livello di scala ad intervalli.

È stato risposto a tali difficoltà che sufficienti evidenze mostrano come risulti giustificato trattare i dati a livello di scala Likert come se fossero a livello di scala ad intervalli quando la numerosità campionaria è sufficientemente grande e quando i dati si distribuiscono in maniera approssimativamente normale. Altri autori (es. Jöreskog & Sörbom, 1996) ritengono invece che le scale tipo Likert vadano considerate in ogni caso come ordinali, e debbano essere analizzate di conseguenza. Nel caso dell'analisi fattoriale e dei modelli di equazioni strutturali questo significa semplicemente che l'analisi si deve basare sul calcolo delle correlazioni polioriche.

In conclusione, la procedura che sta alla base delle scale Likert consiste nella somma dei punti attribuiti ad ogni singola domanda. I vantaggi della scala Likert consistono nella sua semplicità e applicabilità, mentre i suoi svantaggi sono il fatto che i suoi elementi vengono trattati come scale cardinali pur essendo ordinali e il fatto che il punteggio finale non rappresenta una variabile cardinale.

11.2.2 Composizione tipografica

Criteri da considerare nella formattazione tipografica del test di un test psicometrico sono la facilità di lettura, la chiarezza e l'organizzazione. La formattazione dovrebbe tenere in considerazione l'età dei rispondenti e la potenziale difficoltà di lettura.

11.2.3 Istruzioni ai soggetti

Le istruzioni devono essere chiare e concise. Oltre ad illustrare la consegna, esse forniscono una cornice di riferimento che deve essere comune a tutti i rispondenti. Le istruzioni seguono un formato simile al seguente:

Lo studio ha come obiettivo generale ... In particolare, con la ricerca che qui presentiamo, si intendono ottenere dati relativi a ... Nel caso tu decida di partecipare allo studio, questa ricerca prevede l'attuazione dei seguenti trattamenti ... La ricerca durerà ... e vi parteciperanno ... individui. Dalla partecipazione a questa ricerca sono prevedibili i seguenti benefici ... La partecipazione allo studio non comporta alcun rischio. Sei del tutto libero/a di non partecipare allo studio. La tua adesione a questo programma di ricerca è completamente volontaria e potrà essere ritirata in qualsiasi momento.

Ai sensi del Decreto Legislativo 30 giugno 2003 n. 196 in materia di protezione dei dati personali, tratteranno i tuoi dati esclusivamente in funzione della realizzazione dello studio. Lo psicologo che ti seguirà nello studio ti identificherà con un codice: i dati che ti riguardano raccolti nel corso dello studio, ad eccezione del nominativo, saranno registrati, elaborati e conservati unitamente a tale codice, alla data di nascita, al genere. Soltanto il supervisore del progetto di ricerca potrà collegare questo codice al tuo nominativo. I dati, trattati mediante strumenti anche elettronici, saranno diffusi solo in forma rigorosamente anonima, ad esempio attraverso pubblicazioni scientifiche, statistiche e convegni scientifici. La tua partecipazione allo studio implica che il gruppo di ricerca

che organizza lo studio e il Comitato etico potranno conoscere i dati che ti riguardano solo attraverso modalità tali da garantire la riservatezza della tua identità.

Potrai esercitare i diritti di cui all'art. 7 del Codice in materia di protezione dei Dati Personali (es. accedere ai tuoi dati personali, integrarli, aggiornarli, rettificarli, opporsi al loro trattamento per motivi legittimi, ecc.) rivolgendoti direttamente al responsabile della ricerca. Potrai interrompere in ogni momento e senza fornire alcuna giustificazione la tua partecipazione allo studio: in tal caso, i dati raccolti verranno distrutti.

Se lo richiederai, alla fine dello studio potranno esserti comunicati i risultati ottenuti in generale e, in particolare, quelli che ti riguardano. Per ulteriori informazioni e comunicazioni durante la ricerca puoi rivolgerti a ...

Potrai segnalare qualsiasi fatto che riterrai opportuno evidenziare, relativamente alla ricerca che ti riguarda, al Comitato Etico dell'Università degli Studi di Firenze. La segnalazione dovrà essere inoltrata all'attenzione di ...

È inoltre necessario che i partecipanti completino una dichiarazione di assenso (consenso informato). Ad esempio:

Io sottoscritto ... dichiaro di aver ricevuto dal Dottor ... esaurienti spiegazioni in merito alla richiesta di partecipazione alla ricerca in oggetto, secondo quanto riportato nella scheda informativa qui allegata, copia della quale mi è stata prima d'ora consegnata (indicare data e ora della consegna).

Dichiaro altresì di aver potuto discutere tali spiegazioni, porre tutte le domande che ho ritenuto necessarie e di aver ricevuto risposte soddisfacenti, come pure di aver avuto la possibilità di informarmi in merito ai particolari dello studio con persona di mia fiducia.

Accetto dunque liberamente di partecipare alla ricerca, avendo compreso completamente il significato della richiesta e i rischi e benefici che possono derivare da questa partecipazione. Acconsento al trattamento dei miei dati personali per gli scopi della ricerca nei limiti e con le modalità indicate nell'informativa fornitami con il presente documento.

Sono stato informato/a, inoltre, del mio diritto ad avere libero accesso alla documentazione relativa alla ricerca ed alla valutazione espressa dal Comitato Etico dell'Università degli Studi di Firenze.

11.2.4 La costruzione degli item

La scelta di item tecnicamente adeguati sul piano strutturale e linguistico non è un problema statistico. La formulazione verbale degli item è molto importante in quanto essa contribuisce all'errore di misura. Per ridurre gli errori di misura, gli item devono essere formulati nella forma più chiara e meno ambigua possibile. È ovviamente necessario impiegare contenuti coerenti con la definizione del costrutto, ma non ci sono regole semplici per generare item che fanno emergere il costrutto che si cerca di misurare. Vanno certamente evitati contenuti che inducano atteggiamenti difensivi e/o ostili nei rispondenti. La formulazione verbale deve inoltre essere appropriata al livello di scolarità dei rispondenti. Pett, Lackey e Sullivan (2003) forniscono le raccomandazioni seguenti.

- Evitare affermazioni che si riferiscono al passato a meno che il costrutto faccia direttamente riferimento al passato.
- Evitare affermazioni fattuali. Evitare affermazioni su cui quasi tutti (o quasi nessuno) sono d'accordo.
- Evitare l'uso di pronomi personali con un significato ambiguo.
- Selezionare item che potenzialmente possano coprire l'intera gamma delle possibili risposte concernenti il costrutto di interesse.
- Se viene fatto riferimento ad argomenti sensibili, la formulazione verbale deve essere la più neutra possibile.
- Utilizzare un linguaggio chiaro, semplice, diretto. Utilizzare frasi corte, altrimenti non ne è chiaro il senso.
- Evitare affermazioni ambigue o interpretabili in più modi.
- Evitare formulazioni sintattiche complesse.
- Evitare l'uso di parole a bassa frequenza o l'uso di una terminologia che potrebbe non essere capita dai rispondenti.
- Disporre gli item aventi un contenuto sensibile verso la fine dello strumento.
- Fare riferimento a comportamenti specifici e non generali.
- Evitare la duplicazione delle domande.

11.2.4.1 Desiderabilità sociale

Quando si sviluppa lo strumento è necessario tenere in considerazione il fatto che i rispondenti tendono a fornire risposte socialmente desiderabili piuttosto che risposte veritiere (Rosenthal & Rosnow, 1991; Waltz et al., 1991). La Desiderabilità sociale non soltanto introduce dei bias nello strumento ma può anche comprometterne la validità.

La Desiderabilità Sociale si riferisce al bisogno provato da alcuni individui di approvazione sociale e accettazione, e alla credenza di poterle ottenere attraverso comportamenti appropriati e culturalmente accettati (Marlowe & Crowne, 1961). La Desiderabilità Sociale consiste nella tendenza a fornire risposte molto positive quando vengono poste domande su di sé, con l'obiettivo di risultare positivamente agli occhi dell'altro. Marlowe e Crowne (1960) hanno proposto la scala di valutazione MCSCS (Marlowe-Crowne Social Desirability Scale), largamente utilizzata per indagare questo costrutto. Un'altra scala di valutazione molto utilizzata è la BIDR (Balanced Inventory of Desirable Responding, 1991) proposta da Paulhus: tale scala contiene 40 item, volti a rilevare la gestione delle impressioni e l'autoinganno.

11.2.4.2 Item marker

Quando si anticipa la presenza di più costrutti latenti, è utile utilizzare nell'insieme degli item alcuni item marker, ovvero item che correlano molto con un solo fattore e pochissimo con altri. Questo facilita l'interpretazione dei fattori. I marker consentono infatti di attribuire ai fattori un nome (etichetta) coerente con l'area semantica cui i marker fanno riferimento.

11.2.4.3 Campionamento del dominio

Il campionamento del dominio può essere inteso sia come campionamento del contenuto o come campionamento del comportamento. L'adeguatezza del campionamento del contenuto riguarda il fatto che l'insieme degli item sia o meno in grado di rappresentare il dominio di contenuto di interesse. Questa caratteristica è un indice dell'adeguatezza del test nel misurare ciò che intende misurare e dovrebbe garantire che le risposte agli item possano rappresentare una stima della quantità di costrutto posseduta dal rispondente.

Il campionamento del comportamento riguarda invece il grado in cui le risposte a un test costituiscono un campione adeguato dei comportamenti che il test intende misurare. In questo caso ci si chiede se il test riflette i comportamenti che intende valutare e possiede dunque un valore descrittivo del comportamento del rispondente.

Un item mal formulato determina una distorsione delle risposte e non può essere considerato rappresentativo di nessun dominio di contenuto né di nessun universo di comportamenti.

Per la generazione iniziale degli item è molto importante considerare il parere della popolazione target e degli esperti. Interviste accuratamente strutturate e a risposta aperta con esperti o potenziali soggetti permettono non solo di verificare che gli item siano rappresentativi o rilevanti per il costrutto, ma anche che siano formulati correttamente. Questo processo può anche suggerire sfaccettature ulteriori rispetto a quelle progettate inizialmente e la necessità di raffinare il costrutto. Nello sviluppo di un test è molto utile ascoltare il parere di persone inserite nel contesto applicativo del test anche per sapere qual è la terminologia specifica da utilizzare nella formulazione degli item, o se gli item sono chiari, o se la scala di risposta è di facile compilazione.

È anche utile rivolgersi a giudici esterni aventi una conoscenza approfondita del dominio di contenuto per ottenere una prospettiva esterna e autorevole che aiuti nell'individuazione degli item da eliminare e di quelli che richiedono un affinamento.

Gli item di un test dovrebbero essere distribuiti in modo che riflettano la relativa importanza delle varie sfaccettature del costrutto target (Nunnally & Bernstein, 1994). Se gli item per una certa sfaccettatura sono troppi o troppo pochi, i punteggi e le inferenze ottenute da questi punteggi saranno distorte.

11.3 Numero delle opzioni di risposta

Un item è costituito da due parti: l'item stem, cioè il testo che contiene la domanda o l'affermazione da valutare e le alternative di risposta.

In una scala di tipo Likert, le categorie di risposta si dicono a parziale autonomia semantica, ovvero sono tali per cui le modalità di risposta devono essere confrontate con le altre affinché il rispondente sia in grado di stabilire il loro valore. A ciascuna modalità di risposta viene attribuito un punteggio (4, 3, 2, 1 oppure 3, 2, 1, 0), e la somma (o media) dei punteggi alle risposte di ciascun individuo sull'intera batteria rappresenta la posizione dell'individuo sul concetto indagato.

Per esempio,

Fortemente d'accordo 7 6 5 4 3 2 1 Fortemente in disaccordo

oppure,

Molto Abbastanza Poco Per niente

Il numero ottimale delle opzioni di risposta è stato dibattuto a lungo. Per esempio, Schutz e Rucker (1975) hanno trovato che “the number of available response categories does not materially affect the cognitive structure derived from the results” (p. 323), il che suggerisce che il numero di opzioni di risposta ha poco effetto sui risultati ottenuti. Tale conclusione, tuttavia, è stata contraddetta da altri ricercatori. Per esempio, Garner (1960) ha suggerito risultati massimamente informativi si ottengono utilizzando più di 20 opzioni di risposta. D'altra parte, Green e Rao (1970), hanno trovato che i risultati migliori si ottengono con sei o sette alternative di risposta, con un guadagno molto piccolo all'aumentare delle categorie di risposta al di là di sette. In un articolo molto citato, Preston e Colman (2000) hanno esaminato le risposte fornite da un campione di rispondenti variando il numero di opzioni di risposta pari a 2, 3, ..., 11, e 101. Dopo avere calcolato l'attendibilità test-retest e la validità dello strumento, oltre al potere discriminante degli item, hanno concluso che le scale a 2, 3 e 4 passi hanno prestazioni piuttosto basse, avendo gli indici calcolati valori molto maggiori per le scale di risposta con un numero maggiore di opzioni di risposta. In particolare, i risultati dello studio suggeriscono che scale di valutazione con 7, 9 o 10 opzioni di risposta sono da preferire rispetto ad altri numeri di alternative di risposta.

Oltre alle scale Likert è possibile usare le risposte auto-ancoranti, ovvero quelle in cui gli item prevedono solo due aggettivi di risposta, estremi (per esempio, “per niente” e “molto”), legati da un segmento continuo in cui il rispondente deve scegliere la propria posizione. Un esempio è la Visual Analogue Scale usata nella misura dell'umore. Tali scale sono molto più rare delle scale Likert.

11.3.1 Item a codifica inversa

Alcuni item correlano fortemente in maniera negativa con gli altri item e con il punteggio totale del test. Tali item richiedono una codifica inversa. Ad esempio, due item del questionario S.T.A.I per la valutazione dell'ansia sono codificati nel modo seguente.

“Sono preoccupata.”

Per nulla Un po' Abbastanza Moltissimo

con valori 1, 2, 3 e 4, rispettivamente.

“Mi sento bene.”

Per nulla Un po' Abbastanza Moltissimo

con valori 4, 3, 2, e 1, rispettivamente.

In ambito psicometrico si è soliti ritenere che due proprietà contrarie giacciono sullo stesso continuum latente. Nella costruzione di un test psicologico viene dunque consigliato di utilizzare sia item con contenuto orientato nella direzione del costrutto (per cui punteggi alti nell'item sono il riflesso di un alto livello del costrutto) sia nella direzione opposta (per cui punteggi alti nell'item sono il riflesso di un basso livello del costrutto). Nel primo caso si parla di *straight item*, nel secondo di *reverse item*. Lo scopo centrale degli item reverse è quello di contrastare l'acquiescenza, ovvero di rallentare il soggetto nella compilazione del test, evitando di rispondere in maniera automatica, così da prestare maggiore attenzione al contenuto degli item.

11.4 Numero di item

Un test psicometrico, oltre ad essere valido, deve minimizzare l'errore di misura. L'attendibilità di uno strumento sia dall'attendibilità di ciascun item, sia dal numero di item che lo compongono. Tratteremo di questo argomento in un capitolo successivo.

Kline (1986) suggerisce di costruire un numero di item almeno doppio del numero di item che andranno a costituire il test finale. La lunghezza del test dipende dal suo scopo. Un Test di valutazione delle abilità per la scuola primaria non deve richiedere più di trenta minuti per essere completato, altrimenti la fatica e la noia finiscono per distorcere i risultati dello strumento. Lo stesso si può dire per un test di personalità per soggetti adulti. Idealmente, un test dovrebbe essere il più breve possibile, a patto di raggiungere un livello adeguato di validità. Come regola

euristica, Kline (1986) suggerisce la soglia di almeno 50 item nella forma finale del test.

11.5 Numero di soggetti

Vi è poco accordo su quale sia la grandezza del campione necessaria per lo svolgimento dell'analisi fattoriale. Nunnally (1978) ha suggerito che il campione deve essere costituito da almeno 10 soggetti per ciascun item. Comrey e Lee (1992) hanno fornito le seguenti indicazioni: 50—very poor, 100—poor, 200—fair, 300—good, 500—very good, 1,000 or more—excellent. Secondo altri autori “as a general rule of thumb, it is comforting to have at least 300 cases for factor analysis” (Tabachnick e Fidell, 2001).



12

Valutazione della matrice di correlazione

L'analisi fattoriale, essendo una tecnica di analisi multivariata, richiede la comprensione di almeno alcuni concetti di base dell'algebra matriciale. A livello minimale è necessario capire che cosa sono i vettori e le matrici, che cosa è il determinante di una matrice, e in che modo possano essere eseguite le operazioni algebriche su vettori e matrici (si veda l'Appendice).

Dopo avere padroneggiato gli aspetti di base dell'algebra matriciale, possiamo valutare il determinante della matrice da fattorializzare: se il determinante è nullo non si può fare l'analisi. Se invece il determinante non è nullo, ci possiamo chiedere se le correlazioni campionarie siano sufficientemente grandi da giustificare l'analisi fattoriale. Se le correlazioni tra gli item sono di modesta entità, allora la soluzione fornita dall'analisi fattoriale non consentirà una descrizione parsimoniosa delle informazioni contenute nella matrice delle correlazioni. Per rispondere a questa domanda si ricorre all'ispezione visiva della matrice di correlazione e si possono svolgere due test: il test della sfericità di Bartlett e il Kaiser-Meyer-Olkin test.

12.1 Ispezione visiva della matrice di correlazione

L'ispezione visiva della matrice di correlazione viene effettuata per vedere che vi siano blocchi di correlazioni alte fra loro e basse con le alte.

12.2 Sfericità di Bartlett

Il test della sfericità di Bartlett verifica l'ipotesi $H_0 : R = I$ (ovvero che gli item siano tra loro incorrelati) tramite la formula:

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2p + 5) \right] \ln |R|,$$

in cui n è il numero dei soggetti, p il numero delle variabili e $|R|$ il determinante della matrice di correlazione. La statistica del test della sfericità di Bartlett si distribuisce come χ^2 con $p(p - 1)/2$ gradi di libertà. Se il risultato del test è significativo (ovvero, se il valore della statistica test è sufficientemente grande), significa che R ha correlazioni sufficientemente elevate da non essere paragonabili a 0; se è non significativo le correlazioni sono basse e non si distinguono da 0. Il limite di questo test è che dipende dal numero delle variabili e dalla numerosità del campione, quindi tende ad essere significativo all'aumentare del campione e del numero delle variabili anche se ci sono correlazioni basse.

12.3 Test di adeguatezza campionaria di Kaiser-Meyer-Olkin

Henry Kaiser (1970) ha introdotto una misura di adeguatezza campionaria per le matrici sottoposte ad analisi fattoriale. In seguito, Kaiser and Rice (1974) hanno modificato questo indice che da allora è conosciuto come l'indice Kaiser-Meyer-Olkin (KMO). Il test di adeguatezza campionaria KMO è dato da

$$\text{KMO} = \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j r_{ij}^2 + \sum_i \sum_j p_{ij}^2},$$

dove r_{ij} sono le correlazioni osservate e p_{ij} sono le correlazioni parzializzate su tutte le altre. Se le correlazioni parzializzate sono piccole, KMO tende a 1. Secondo Kaiser (1970), se $\text{KMO} > 0.90$, l'adeguatezza campionaria è eccellente, fra .80 e .90 è buona, fra .70 e .80 è accettabile, fra .60 e .70 è mediocre, se è inferiore a .60 è meglio non fare l'analisi.

12.3.1 Matrice anti-immagine

La matrice delle correlazioni parzializzate riporta il valore di ogni correlazione dopo aver eliminato il contributo di tutte le altre variabili non implicate. Una correlazione parzializzata alta significa che due variabili sono molto correlate fra loro, ma non hanno legami con nessun'altra variabile. L'analisi fattoriale richiede invece che ci siano più di due variabili per fattore. La matrice anti-immagine contiene i complementi a 1 della correlazione parzializzata fra due variabili rispetto a tutte e altre. Valori alti indicano correlazioni parziali basse e viceversa.



13

L'estrazione dei fattori

Motivazione

Lo scopo dell'analisi fattoriale è quello di descrivere in maniera parsimoniosa le relazioni che intercorrono tra un grande numero di item. Ci si chiede se è possibile identificare un piccolo numero di variabili latenti che, quando vengono controllate, rendono uguali a zero le correlazioni parziali tra gli item. Abbiamo visto nel capitolo precedente come sia possibile determinare il numero dei fattori comuni. Chiediamoci ora come sia possibile stimare le saturazioni fattoriali che corrispondono alle correlazioni (o covarianze) tra gli item e i fattori.

In termini matriciali, il modello multifattoriale si scrive

$$\Sigma = \Lambda\Phi\Lambda^T + \Psi$$

dove Φ è la matrice di ordine $m \times m$ di varianze e covarianze tra i fattori comuni e Ψ è una matrice diagonale di ordine p con le unicità delle variabili. Poniamoci ora il problema di stimare Λ .

13.1 Metodo delle componenti principali

L'analisi fattoriale eseguita mediante il metodo delle componenti principali, nonostante il nome, non è un'analisi delle componenti principali. Il metodo delle componenti principali costituisce invece un'applicazione del teorema di scomposizione spettrale di una matrice. Il *teorema spettrale* afferma che “data la matrice simmetrica $\mathbf{S}_{p \times p}$, è sempre possibile trovare una matrice $\mathbf{C}_{p \times p}$ ortogonale tale che

$$\mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}^T$$

con \mathbf{D} diagonale.” Il teorema specifica inoltre che gli elementi presenti sulla diagonale di \mathbf{D} sono gli autovalori di \mathbf{S} , mentre le colonne di \mathbf{C} rappresentano i rispettivi autovettori normalizzati associati agli autovalori di \mathbf{S} .

Facciamo un esempio numerico utilizzando i dati discussi da Rencher(2002). Brown, Williams e Barlow (1984) hanno raccolto le valutazioni di una ragazza dodicenne relativamente a sette persone di sua conoscenza. Ciascuna persona veniva valutata su una scala a nove punti rispetto a cinque variabili: *kind*, *intelligent*, *happy*, *likeable* e *just*. La matrice di correlazione per tali variabili è riportata di seguito:

```
R <- matrix(c(
  1.000, .296, .881, .995, .545,
  .296, 1.000, -.022, .326, .837,
  .881, -.022, 1.000, .867, .130,
  .995, .326, .867, 1.000, .544,
  .545, .837, .130, .544, 1.000
),
ncol = 5, byrow = T, dimnames = list(
  c("K", "I", "H", "L", "J"),
  c("K", "I", "H", "L", "J")
)
)
```

Gli autovalori e gli autovettori si calcolano con la funzione `eigen()`:

```
e <- eigen(R)
print(e, 3)
#> eigen() decomposition
#> $values
#> [1] 3.263377 1.538382 0.167969 0.030030 0.000242
#>
#> $vectors
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] -0.537 -0.186 -0.1899 -0.125 0.791
#> [2,] -0.287 0.651 0.6849 -0.120 0.103
```

```
#> [3,] -0.434 -0.474  0.4069  0.614 -0.212
#> [4,] -0.537 -0.169 -0.0953 -0.629 -0.527
#> [5,] -0.390  0.538 -0.5658  0.444 -0.204
```

Come indicato in precedenza, la matrice \mathbf{R} può essere espressa come $\mathbf{R} = \mathbf{C}\mathbf{D}\mathbf{C}^T$:

```
e$vector %**% diag(e$values) %**% t(e$vector)
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 1.000 0.296 0.881 0.995 0.545
#> [2,] 0.296 1.000 -0.022 0.326 0.837
#> [3,] 0.881 -0.022 1.000 0.867 0.130
#> [4,] 0.995 0.326 0.867 1.000 0.544
#> [5,] 0.545 0.837 0.130 0.544 1.000
```

Esaminiamo ora gli autovalori. I primi due autovalori spiegano da soli il 96% della varianza campionaria:

```
(e$values[1] + e$values[2]) / 5
#> [1] 0.9604
```

Usando i primi due autovalori e i primi due autovettori sarà dunque possibile riprodurre in maniera soddisfacente la matrice \mathbf{R} operando nel contempo una riduzione di dimensionalità dei dati.

Per fattorizzare $\mathbf{R} = \mathbf{C}\mathbf{D}\mathbf{C}^T$ nella forma $\hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}^T$ iniziamo a scrivere

$$\mathbf{D} = \mathbf{D}^{1/2}\mathbf{D}^{1/2}$$

dove

$$\mathbf{D}^{1/2} = \begin{bmatrix} \sqrt{\theta_1} & 0 & \dots & 0 \\ 0 & \sqrt{\theta_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{\theta_p} \end{bmatrix}$$

Viene qui usata la notazione θ per denotare gli autovalori anziché il tradizionale λ per evitare la confusione con la notazione λ_{jl} usata per le saturazioni fattoriali. In questo modo, possiamo scrivere

$$\begin{aligned}
\mathbf{R} &= \mathbf{C}\mathbf{D}\mathbf{C}^\top \\
&= \mathbf{C}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{C}^\top \\
&= (\mathbf{C}\mathbf{D}^{1/2})(\mathbf{C}\mathbf{D}^{1/2})^\top
\end{aligned}$$

Non possiamo però limiarci a definire $\hat{\Lambda} = \mathbf{C}\mathbf{D}^{1/2}$ in quanto $\mathbf{C}\mathbf{D}^{1/2}$ è di ordine $p \times p$ e non otteniamo quindi una riduzione di dimensioni. Quello che cerchiamo è una matrice $\hat{\Lambda}$ di ordine $p \times m$ con $m < p$. Dunque, definiamo la matrice $\mathbf{D}_1 = \text{diag}(\theta_1, \theta_2, \dots, \theta_m)$ come la matrice contenente gli m autovalori più grandi di \mathbf{R} e $\mathbf{C}_1 = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$ come la matrice contenente i rispettivi autovettori. Mediante il metodo delle componenti principali, le saturazioni fattoriali $\hat{\Lambda}$ vengono quindi stimate nel modo seguente:

$$\begin{aligned}
\hat{\Lambda} &= \mathbf{C}_1\mathbf{D}_1^{1/2} \\
&= (\sqrt{\theta_1}\mathbf{c}_1, \sqrt{\theta_2}\mathbf{c}_2, \dots, \sqrt{\theta_m}\mathbf{c}_m)
\end{aligned}$$

Per l'esempio presente, con $m = 2$ e $p = 5$, avremo

$$\begin{bmatrix} \hat{\lambda}_{11} & \hat{\lambda}_{12} \\ \hat{\lambda}_{21} & \hat{\lambda}_{22} \\ \hat{\lambda}_{31} & \hat{\lambda}_{32} \\ \hat{\lambda}_{41} & \hat{\lambda}_{42} \\ \hat{\lambda}_{51} & \hat{\lambda}_{52} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ c_{51} & c_{52} \end{bmatrix} \begin{bmatrix} \sqrt{\theta_1} & 0 \\ 0 & \sqrt{\theta_2} \end{bmatrix}$$

Le saturazioni fattoriali stimate sono dunque uguali a

$$\begin{bmatrix} \sqrt{\theta_1}c_{11} & \sqrt{\theta_2}c_{12} \\ \sqrt{\theta_1}c_{21} & \sqrt{\theta_2}c_{22} \\ \sqrt{\theta_1}c_{31} & \sqrt{\theta_2}c_{32} \\ \sqrt{\theta_1}c_{41} & \sqrt{\theta_2}c_{42} \\ \sqrt{\theta_1}c_{51} & \sqrt{\theta_2}c_{52} \end{bmatrix}$$

Svolgendo i calcoli con R otteniamo:

```

L <- cbind(
  e$variables[, 1] * sqrt(e$values[1]),
  e$variables[, 2] * sqrt(e$values[2])
)
```

```
)

round(L, 3)
#>      [,1]  [,2]
#> [1,] -0.970 -0.231
#> [2,] -0.519  0.807
#> [3,] -0.785 -0.588
#> [4,] -0.971 -0.210
#> [5,] -0.704  0.667
```

La matrice di correlazione riprodotta (con le comunalità sulla diagonale principale) diventa

```
round(L %*% t(L), 3)
#>      [,1]  [,2]  [,3]  [,4]  [,5]
#> [1,] 0.993  0.317  0.896  0.990  0.529
#> [2,] 0.317  0.921 -0.067  0.335  0.904
#> [3,] 0.896 -0.067  0.961  0.885  0.160
#> [4,] 0.990  0.335  0.885  0.987  0.543
#> [5,] 0.529  0.904  0.160  0.543  0.940
```

Possiamo ora capire il motivo del nome “metodo delle componenti principali.” Le saturazioni fattoriali sono proporzionali agli autovettori di \mathbf{R} . Tuttavia, dopo la rotazione delle saturazioni fattoriali, l’interpretazione dei fattori è diversa da quella che viene assegnata ai risultati dell’analisi delle componenti principali.

È possibile condurre l’analisi fattoriale con il metodo delle componenti principali sia utilizzando la matrice \mathbf{S} di varianze-covarianze sia la matrice \mathbf{R} delle correlazioni. Tuttavia, le soluzioni ottenute usando \mathbf{S} o \mathbf{R} non sono legate da una relazione algebrica: il metodo delle componenti principali non è invariante rispetto ai cambiamenti di scala delle osservazioni. Un altro svantaggio del metodo delle componenti principali è che non fornisce un test di bontà di adattamento. Tale test può essere invece svolto quando la soluzione viene trovata con il metodo della massima verosimiglianza.

13.2 Metodo dei fattori principali

Il *metodo dei fattori principali* (*principal factor method*, anche detto *principal axis method*) è uno dei metodi maggiormente usati per la stima delle saturazioni fattoriali e delle comunalità. Il metodo delle componenti principali trascura la specificità Ψ e si limita a fattorializzare le covarianze di \mathbf{S} o le correlazioni di \mathbf{R} . Il metodo dei fattori principali utilizza una procedura simile al metodo delle componenti principali, utilizzando però una matrice ridotta di varianze-covarianze $\mathbf{S} - \hat{\Psi}$ in cui una stima delle comunalità viene sostituita alle varianze presenti sulla diagonale principale. Nel caso della matrice ridotta di correlazioni $\mathbf{R} - \hat{\Psi}$, per la comunalità i -esima $\sum_j \lambda_{ij}^2$ si sceglie il quadrato del coefficiente di correlazione multipla tra Y_i e tutte le altre $p - 1$ variabili. Tale valore si può trovare nel modo seguente:

$$\hat{h}_i^2 = R_i^2 = 1 - \frac{1}{r^{ii}}$$

dove r^{ii} è l'elemento diagonale i -esimo di \mathbf{R}^{-1} . Nel caso di una matrice ridotta di varianze-covarianze $\mathbf{S} - \hat{\Psi}$, le comunalità possono essere stimate calcolando

$$\hat{h}_i^2 = s_{ii} - \frac{1}{r^{ii}}$$

dove s_{ii} è l'elemento diagonale i -esimo di \mathbf{S} .

Affinché le stime comunalità possano essere calcolate come descritto sopra, la matrice \mathbf{R} deve essere non singolare. Nel caso in cui \mathbf{R} sia singolare, per la stima della comunalità i -esima, \hat{h}_i^2 , si utilizza il valore assoluto del più elevato coefficiente di correlazione lineare tra Y_i e le altre variabili.

Scelta la stima della comunalità, la matrice ridotta di varianze-covarianze si ottiene sostituendo alle varianze sulla diagonale principale le stime delle comunalità:

$$\mathbf{S} - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & \hat{h}_2^2 & \dots & s_{2p} \\ \dots & \dots & \dots & \dots \\ s_{p1} & s_{p2} & \dots & \hat{h}_p^2 \end{bmatrix}$$

In maniera equivalente, la matrice ridotta di correlazioni si ottiene nel modo seguente:

$$\mathbf{R} - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \dots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & \hat{h}_p^2 \end{bmatrix}$$

Verranno ora svolti i calcoli necessari per la stima dei coefficienti di saturazione con il metodo dei fattori principali utilizzando la matrice di correlazione dell'esempio precedente. Quale stima della comunaltà i -esima, verrà utilizzato il valore assoluto più elevato nella riga i -esima della matrice \mathbf{R} . Per i dati dell'esempio, le stime delle comunaltà sono dunque pari a 0.995, 0.837, 0.881, 0.995 e 0.837.

Inserendo tali valori nella diagonale principale, otteniamo la matrice ridotta delle correlazioni $\mathbf{R} - \hat{\Psi}$:

```
R1 <- R
h.hat <- c(.995, .837, .881, .995, .837)
R1[cbind(1:5, 1:5)] <- h.hat
R1
#>      K      I      H      L      J
#> K 0.995  0.296  0.881  0.995  0.545
#> I 0.296  0.837 -0.022  0.326  0.837
#> H 0.881 -0.022  0.881  0.867  0.130
#> L 0.995  0.326  0.867  0.995  0.544
#> J 0.545  0.837  0.130  0.544  0.837
```

Gli autovalori della matrice ridotta di correlazioni $\mathbf{R} - \hat{\Psi}$ sono:

```
ee <- eigen(R1)
round(ee$values, 3)
#> [1]  3.202  1.394  0.029  0.000 -0.080
```

La somma degli autovalori è uguale a

```
sum(ee$values)
#> [1] 4.545
```

I primi due autovalori di $\mathbf{R} - \hat{\Psi}$ sono:

```
round(ee$vectors[, 1:2], 3)
#>      [,1]  [,2]
#> [1,] -0.548 -0.177
#> [2,] -0.272  0.656
#> [3,] -0.431 -0.461
#> [4,] -0.549 -0.159
#> [5,] -0.373  0.549
```

Moltiplicando tali valori per la radice quadrata dei rispettivi autovalori si ottengono le stime delle saturazioni fattoriali:

```
round(ee$vectors[, 1:2] %*% sqrt(diag(ee$values[1:2])), 3)
#>      [,1]  [,2]
#> [1,] -0.981 -0.209
#> [2,] -0.487  0.774
#> [3,] -0.772 -0.544
#> [4,] -0.982 -0.187
#> [5,] -0.667  0.648
```

Tale risultato replica quello riportato da Rencher (2002).

13.3 Metodo dei fattori principali iterato

Solitamente, per migliorare la stima della comunaltà, la diagonale della matrice $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$ viene ottenuta per iterazione. Dopo avere trovato $\hat{\Lambda}$ a partire da $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$ come indicato in precedenza, utilizzando le stime delle saturazioni fattoriali così ottenute possiamo stimare le comunaltà nel modo seguente:

$$\hat{h}_i^2 = \sum_{j=1}^m \hat{\lambda}_{ij}^2.$$

I valori di \hat{h}_i^2 vengono quindi sostituiti nella diagonale della matrice ridotta $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$. A partire da questa nuova matrice, usando il metodo descritto in precedenza, possiamo così ottenere una nuova stima delle saturazioni fattoriali $\hat{\Lambda}$. Mediante questa nuova stima di $\hat{\Lambda}$, possiamo procedere ad una nuova stima delle comunaltà. Tale processo continua iterativamente sino alla convergenza. Gli autovalori e gli autovettori della versione finale di $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$ vengono infine usati per stimare i pesi fattoriali. Il metodo dei fattori principali iterato e il metodo delle componenti principali producono risultati molto simili quando m assume un piccolo valore (questo si verifica quando le correlazioni sono alte) e quando p (il numero delle variabili) è grande.

13.3.1 Casi di Heywood

Tra gli inconvenienti del metodo dei fattori principali iterato vi è il fatto che può talvolta portare a soluzioni inammissibili (quando viene fattorizzata la matrice \mathbf{R}) caratterizzate da valori di comunaltà maggiori di uno (*caso di Heywood*). Se $\hat{h}_i^2 > 1$ allora $\hat{\psi}_i < 0$ il che è chiaramente assurdo in quanto una varianza non può assumere un valore negativo. Solitamente, quando la stima di una comunaltà è maggiore di uno, il processo iterativo viene interrotto e il programma riporta che non può essere trovata una soluzione.

Nell'esempio presente viene utilizzata la funzione `factor.pa()` contenuta nel pacchetto `psych` per trovare la soluzione dei fattori principali mediante il metodo iterativo:

```
pa <- fa(R, nfactors = 2, rotate = "none", fm = "pa")
pa
#> Factor Analysis using method = pa
#> Call: fa(r = R, nfactors = 2, rotate = "none", fm = "pa")
#> Standardized loadings (pattern matrix) based upon correlation matrix
#>   PA1   PA2   h2    u2 com
#> K 0.98 -0.21 1.01 -0.008 1.1
#> I 0.48  0.74 0.77  0.230 1.7
#> H 0.78 -0.56 0.92  0.085 1.8
```

```

#> L 0.98 -0.19 0.99 0.010 1.1
#> J 0.69 0.69 0.95 0.049 2.0
#>
#>
#> PA1 PA2
#> SS loadings 3.22 1.41
#> Proportion Var 0.64 0.28
#> Cumulative Var 0.64 0.93
#> Proportion Explained 0.70 0.30
#> Cumulative Proportion 0.70 1.00
#>
#> Mean item complexity = 1.5
#> Test of the hypothesis that 2 factors are sufficient.
#>
#> The degrees of freedom for the null model are 10 and the objective function was 12
#> The degrees of freedom for the model are 1 and the objective function was 5.6
#>
#> The root mean square of the residuals (RMSR) is 0.01
#> The df corrected root mean square of the residuals is 0.04
#>
#> Fit based upon off diagonal values = 1

```

Si noti che, in questo caso, le unicità assumono valori negativi, il che suggerisce che la soluzione è impropria.

13.4 Metodo di massima verosimiglianza

L'applicazione del metodo di massima verosimiglianza è indicata quando si può assumere che le variabili manifeste seguono una distribuzione normale multivariata. Sotto tali condizioni, tale metodo produce le stime dei pesi fattoriali che più verosimilmente hanno prodotto le correlazioni osservate. Gli stimatori di massima verosimiglianza sono preferibili a quelli ottenuti con altri metodi, sempre che siano pienamente realizzate le premesse. La funzione F da minimizzare rappresenta una misura di "distanza" tra la matrice di covarianza osservata e quella predetta dal modello. Uguagliando a zero le derivate di F rispetto a Λ e Ψ si ottengono le equazioni per le stime di massima verosimiglianza di $\hat{\Lambda}$ e

$\hat{\Psi}$. Risolvendo tali equazioni rispetto alle incognite $\hat{\Lambda}$ e $\hat{\Psi}$ si ricavano le stime di massima verosimiglianza.

Non esistendo una soluzione analitica per queste equazioni, si ricorre a procedimenti numerici iterativi che talvolta presentano problemi di convergenza. La soluzione, pur presentando la possibilità di fornire delle stime di comunaltà superiori a 1 (caso di Heywood), è equivariante rispetto a cambiamenti di scala: le stime di massima verosimiglianza sono indipendenti dall'unità di misura delle variabili manifeste. Pertanto, si ottiene la stessa soluzione sia che si analizzi la matrice delle varianze e covarianze, sia che si analizzi la matrice delle correlazioni.

Consideriamo nuovamente i dati dell'esempio precedente. Le istruzioni sono le seguenti:

```
factanal(covmat = R, factors = 2, rotation = "none", n.obs = 225)
#>
#> Call:
#> factanal(factors = 2, covmat = R, n.obs = 225, rotation = "none")
#>
#> Uniquenesses:
#>      K      I      H      L      J
#> 0.005 0.268 0.055 0.008 0.005
#>
#> Loadings:
#>      Factor1 Factor2
#> K  0.955  -0.289
#> I  0.528   0.673
#> H  0.720  -0.653
#> L  0.954  -0.287
#> J  0.764   0.642
#>
#>
#>              Factor1 Factor2
#> SS loadings      3.203   1.457
#> Proportion Var    0.641   0.291
#> Cumulative Var    0.641   0.932
#>
#> Test of the hypothesis that 2 factors are sufficient.
#> The chi square statistic is 648.1 on 1 degree of freedom.
#> The p-value is 5.81e-143
```



14

Il numero dei fattori

Sono stati proposti quattro criteri per determinare il numero m di fattori da estrarre (Rencher, 2002). Tali criteri sono elencati di seguito.

- Scegliere m tale per cui la varianza spiegata dal modello fattoriale superi una soglia predeterminata, per esempio l'80% della varianza totale, $tr(\mathbf{S})$ o $tr(\mathbf{R})$.
- Scegliere m uguale al numero di autovalori aventi un valore maggiore del valore medio degli autovalori. Per \mathbf{R} il valore medio degli autovalori è 1; per \mathbf{S} è $\sum_{j=1}^p \theta_j / p$.
- Usare lo *scree test*.
- Valutare l'ipotesi che m sia il numero corretto di fattori, $H_0 : \Sigma = \Lambda\Lambda^T + \Psi$, dove Λ è di ordine $p \times m$.

14.1 Quota di varianza spiegata

Il primo criterio si applica soprattutto al metodo delle componenti principali. La proporzione della varianza capionaria spiegata dal fattore j -esimo estratto da \mathbf{S} è uguale a

$$\sum_{i=i}^p \hat{\lambda}_{ij}^2 / tr(\mathbf{S}).$$

Nel caso in cui i fattori vengano estratti da \mathbf{R} avremo

$$\sum_{i=i}^p \hat{\lambda}_{ij}^2 / p.$$

Nel caso di fattori incorrelati, ciascun fattore contribuisce con una quota complessiva di varianza spiegata pari alla somma dei quadrati delle sa-

turazioni fattoriali contenute nella matrice $\hat{\Lambda}$: $\sum_{i=1}^p \sum_{j=1}^m \hat{\lambda}_{ij}^2$. Nel caso del metodo delle componenti principali, tale somma è anche uguale alla somma dei primi m autovalori, o alla somma di tutte le p comunaltà:

$$\sum_{i=1}^p \sum_{j=1}^m \hat{\lambda}_{ij}^2 = \sum_{i=1}^p \hat{h}_i^2 = \sum_{j=1}^m \theta_j$$

Sulla base di queste considerazioni, il numero m di fattori viene scelto in modo da spiegare una quota sufficientemente grande di \mathbf{S} o \mathbf{p} .

Il numero dei fattori può essere determinato in questo modo anche nel caso in cui l'analisi fattoriale venga eseguita con il metodo dei fattori principali (ovvero, nel caso in cui vengano usate le stime delle comunaltà per generare la matrice ridotta $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$). In questo caso, però, è possibile che alcuni autovalori della matrice $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$ assumano valore negativo. In tali circostanze, è possibile che la proporzione cumulativa della varianza $\sum_{j=1}^m \theta_j / \sum_{j=1}^p \theta_j$ assuma un valore maggiore di 1.0 per $j < p$.

La proporzione cumulativa della varianza si riduce poi a 1.0 quando vengono considerati anche i successivi autovalori negativi. Di conseguenza, può succedere che, utilizzando la matrice $\mathbf{S} - \hat{\Psi}$ o $\mathbf{R} - \hat{\Psi}$, il criterio definito in base alla quota della varianza spiegata venga raggiunto per un valore m minore di quello che verrebbe trovato utilizzando la matrice \mathbf{S} o \mathbf{R} .

Nel caso del metodo dei fattori principali iterato, m viene specificato precedentemente a ciascuna iterazione e $\sum_i \hat{h}_i^2$ viene ottenuto dopo ciascuna iterazione calcolando $\text{tr}(\mathbf{S} - \hat{\Psi})$. Per scegliere m , come per il metodo delle componenti principali, possono essere usati gli autovalori di \mathbf{S} o \mathbf{R} .

14.2 Valore medio degli autovalori

Il calcolo del valore medio degli autovalori è una procedura euristica implementata in molti software. In una variante di tale metodo, m viene scelto in modo tale da uguagliare il numero degli autovalori positivi della matrice ridotta $\mathbf{R} - \hat{\Psi}$ (in tale matrice vi sono solitamente degli autovalori

negativi). Tale variante ha però lo svantaggio di produrre solitamente un numero di fattori troppo grande.

14.3 Scree test

Lo scree test è basato su una rappresentazione grafica degli autovalori di **S** o **R**. Si costruisce un grafico che rappresenta gli autovalori ordinati in maniera decrescente in funzione del numero dei fattori. I punti che rappresentano gli autovalori vengono collegati con una spezzata. Il valore m viene determinato in corrispondenza di quel fattore oltre il quale il dislivello tra fattori successivi diventa esiguo e la spezzata tende a diventare orizzontale.

14.4 Parallel analysis

La Parallel Analysis è un metodo alternativo allo scree test^[1]. Nella Parallel Analysis, il criterio usato per decidere il numero di fattori da estrarre viene determinato dal confronto con la media degli autovalori generati da un campione casuale di variabili standardizzate. Tale confronto ha lo scopo di controllare le variazioni dovute agli errori di campionamento. Anche se, nel caso di variabili incorrelate, tutti gli autovalori di una matrice di correlazione dovrebbero avere un valore pari a uno, come conseguenza della variabilità campionaria in qualunque campione finito vi sono comunque uno o più autovalori empirici maggiori di uno.

Tale fatto può essere illustrato mediante la seguente simulazione di Monte Carlo. Si consideri una matrice di correlazione calcolata su $p = 10$ variabili casuali mutuamente indipendenti, ciascuna costituita da $n = 20$ osservazioni.

```
n <- 20
nsim <- 1000
e1 <- rep(0, nsim)
for (i in 1:nsim) {
  Y <- cbind(
```

```

      rnorm(n), rnorm(n), rnorm(n), rnorm(n), rnorm(n),
      rnorm(n), rnorm(n), rnorm(n), rnorm(n), rnorm(n)
    )
    e <- eigen(cor(Y))
    e1[i] <- e$values[1]
  }
  max(e1)
#> [1] 3.333

```

Per i dati di questa simulazione, l'autovalore maggiore ha un valore pari a 3.53, anche se i dati sono del tutto casuali. La Parallel Analysis tiene conto di questo fatto e determina m confrontando gli autovalori empirici con le loro “controparti casuali.” Vanno a determinare m solo gli autovalori empirici che hanno un valore superiore ai corrispondenti autovalori generati da una matrice di dati dello stesso ordine composta da colonne mutualmente incorrelate. Nel caso dell'esempio presente, per esempio, l'autovalore maggiore dovrà avere un valore maggiore di 3.53 (anziché di 1.00 o del punto di flesso della spezzata dello scree test).

14.5 Test di bontà di adattamento

Se si assume la normalità distributiva dei dati è possibile valutare la bontà di adattamento attraverso il test del rapporto di verosimiglianze. L'ipotesi nulla postula che la matrice di covarianza delle Y abbia la forma specificata dal modello fattoriale, ossia

$$H_0 : \Sigma = \Lambda \Lambda^T + \Psi,$$

ovvero che m fattori comuni siano sufficienti per spiegare la struttura di interdipendenza della variabile casuale Y oggetto di osservazione campionaria. L'alternativa è che m fattori comuni non siano sufficienti a tale spiegazione

$$H_1 : \Sigma \neq \Lambda \Lambda^T + \Psi,$$

dove Λ è di ordine $p \times m$. Se l'ipotesi nulla non viene rifiutata vuol dire che il modello fornisce un buon adattamento ai dati.

Sotto l'ipotesi nulla, il test ha una distribuzione asintotica, per $n \rightarrow \infty$, di tipo chi quadrato con gradi di libertà pari a:

$$\nu = \frac{1}{2} [(p - m)^2 - (p - m)].$$

Tale risultato di natura asintotica, valido per n grande, può essere migliorato, per ottenere una approssimazione migliore, sostituendo ν con:

$$\nu^* = n - 2 - \frac{2p - 1}{6} - \frac{2}{3}m.$$

Il rifiuto di H_0 implica che m è troppo piccolo e un numero maggiore di fattori è necessario. Solitamente si inizia l'analisi considerando un numero di fattori molto piccolo: $m^* = 1$ e si prende come ipotesi per il test che il numero di fattori sia m^* . Se l'ipotesi nulla è accettata il procedimento si arresta, altrimenti si passa a considerare $m^* + 1$ fattori e si prosegue con lo stesso ragionamento. Il procedimento si arresta non appena si verifica una delle seguenti situazioni: è stata accettata l'ipotesi H_0 per un certo valore di m , oppure, $\nu = \frac{1}{2} [(p - m)^2 - (p - m)] = 0$, ossia la variabile χ^2 dovrebbe avere zero gradi di libertà, che non è possibile.

Per poter applicare il test che determina il buon grado di adattamento del modello fattoriale occorre che i gradi di libertà della statistica del chi-quadrato siano positivi. Questo sta a significare che il numero di fattori comuni non può superare il più grande numero intero che soddisfa la seguente equazione:

$$m < \frac{1}{2} (2p + 1 - \sqrt{8p + 1})$$

per un numero fissato p di variabili manifeste. In pratica, quando n è grande, il test basato sul rapporto di verosimiglianze rivela un numero di fattori maggiore degli altri metodi descritti in precedenza. Alcuni considerano dunque il valore m indicato dal test quale limite superiore del numero dei fattori che rivestono una qualche importanza pratica.

Per alcuni campioni di dati, la scelta di m non è ovvia. Questa indeterminazione costituisce un limite dell'analisi fattoriale. Solitamente, si procede con utilizzando un certo metodo per la scelta di m (diciamo

lo scree test) e valuta la proporzione di varianza spiegata di ciascun item e, dopo un'appropriata rotazione, l'interpretabilità della soluzione ottenuta. Se le comunaltà o l'interpretabilità dei fattori non sembrano adeguati, si procede con un numero maggiore di fattori. Tale procedura è certamente soggettiva e i limiti della soluzione che viene ottenuta sono evidenti.

Per altri campioni di dati, la scelta di m consente una maggiore certezza. Questo avviene quando tutti i metodi che abbiamo descritto prima forniscono la stessa risposta. In questi casi, possiamo essere più certi della soluzione dell'analisi fattoriale.

Esempio 14.1. Per confrontare i quattro metodi discussi per la scelta del numero m di fattori usiamo qui una matrice di correlazioni calcolata sulla WAIS. Le 11 sottoscale sono le seguenti:

- X1 = Information
- X2 = Comprehension
- X3 = Arithmetic
- X4 = Similarities
- X5 = Digit.span
- X6 = Vocabulary
- X7 = Digit.symbol
- X8 = Picture.completion
- X9 = Block.design
- X10 = Picture.arrangement
- X11 = Object.assembly

```
varnames <- c(
  "IN", "CO", "AR", "SI", "DS", "VO", "SY", "PC",
  "BD", "PA", "OA", "AG", "ED"
)
temp <- matrix(c(
  1, 0.67, 0.62, 0.66, 0.47, 0.81, 0.47, 0.60, 0.49, 0.51, 0.41,
  -0.07, 0.66, .67, 1, 0.54, 0.60, 0.39, 0.72, 0.40, 0.54, 0.45,
  0.49, 0.38, -0.08, 0.52, .62, .54, 1, 0.51, 0.51, 0.58, 0.41,
  0.46, 0.48, 0.43, 0.37, -0.08, 0.49, .66, .60, .51, 1, 0.41,
  0.68, 0.49, 0.56, 0.50, 0.50, 0.41, -0.19, 0.55, .47, .39, .51,
  .41, 1, 0.45, 0.45, 0.42, 0.39, 0.42, 0.31, -0.19, 0.43,
  .81, .72, .58, .68, .45, 1, 0.49, 0.57, 0.46, 0.52, 0.40, -0.02,
```

```

0.62, .47, .40, .41, .49, .45, .49, 1, 0.50, 0.50, 0.52, 0.46,
-0.46, 0.57, .60, .54, .46, .56, .42, .57, .50, 1, 0.61, 0.59,
0.51, -0.28, 0.48, .49, .45, .48, .50, .39, .46, .50, .61, 1,
0.54, 0.59, -0.32, 0.44, .51, .49, .43, .50, .42, .52, .52, .59,
.54, 1, 0.46, -0.37, 0.49, .41, .38, .37, .41, .31, .40, .46, .51,
.59, .46, 1, -0.28, 0.40, -.07, -.08, -.08, -.19, -.19, -.02,
-.46, -.28, -.32, -.37, -.28, 1, -0.29, .66, .52, .49, .55, .43,
.62, .57, .48, .44, .49, .40, -.29, 1
), nrow = 13, ncol = 13, byrow = TRUE)

colnames(temp) <- varnames
rownames(temp) <- varnames

wais_cor <- temp[1:11, 1:11]
wais_cor
#>      IN  CO  AR  SI  DS  VO  SY  PC  BD  PA
#> IN 1.00 0.67 0.62 0.66 0.47 0.81 0.47 0.60 0.49 0.51
#> CO 0.67 1.00 0.54 0.60 0.39 0.72 0.40 0.54 0.45 0.49
#> AR 0.62 0.54 1.00 0.51 0.51 0.58 0.41 0.46 0.48 0.43
#> SI 0.66 0.60 0.51 1.00 0.41 0.68 0.49 0.56 0.50 0.50
#> DS 0.47 0.39 0.51 0.41 1.00 0.45 0.45 0.42 0.39 0.42
#> VO 0.81 0.72 0.58 0.68 0.45 1.00 0.49 0.57 0.46 0.52
#> SY 0.47 0.40 0.41 0.49 0.45 0.49 1.00 0.50 0.50 0.52
#> PC 0.60 0.54 0.46 0.56 0.42 0.57 0.50 1.00 0.61 0.59
#> BD 0.49 0.45 0.48 0.50 0.39 0.46 0.50 0.61 1.00 0.54
#> PA 0.51 0.49 0.43 0.50 0.42 0.52 0.52 0.59 0.54 1.00
#> OA 0.41 0.38 0.37 0.41 0.31 0.40 0.46 0.51 0.59 0.46
#>      OA
#> IN 0.41
#> CO 0.38
#> AR 0.37
#> SI 0.41
#> DS 0.31
#> VO 0.40
#> SY 0.46
#> PC 0.51
#> BD 0.59
#> PA 0.46

```

```
#> OA 1.00
```

Il primo metodo per la determinazione di m richiede di estrarre tanti fattori quanti sono necessari per spiegare una quota predeterminata della varianza totale. Supponiamo di porre il criterio pari all'80% della varianza totale. La soluzione ottenuta in questo modo ci porterebbe a mantenere $m = 5$ fattori:

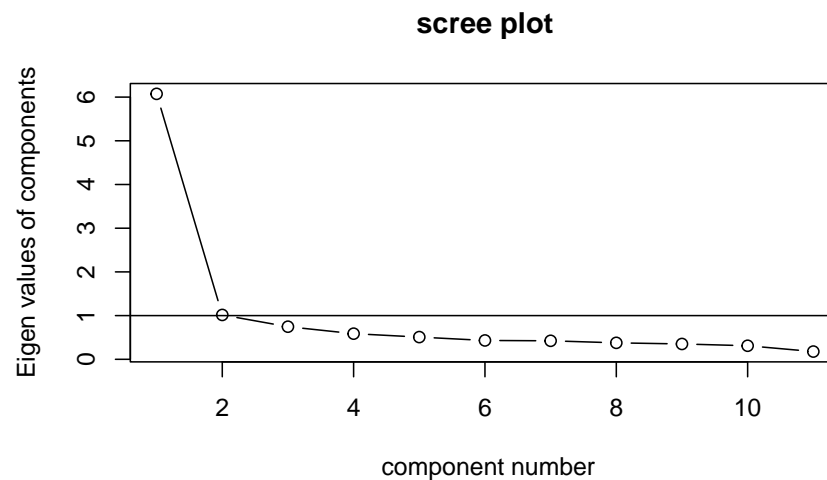
```
out <- eigen(wais_cor)
sum(out$val[1:4]) / sum(out$val)
#> [1] 0.7657
sum(out$val[1:5]) / sum(out$val)
#> [1] 0.8119
```

Il secondo metodo suggerisce di mantenere tutti gli autovalori superiori al valore medio degli autovalori (che, nel caso di \mathbf{R} è uguale a 1). Nel caso presente, $m = 2$:

```
round(out$values, 3)
#> [1] 6.074 1.015 0.746 0.587 0.508 0.431 0.423 0.377
#> [9] 0.351 0.310 0.177
```

Il terzo metodo, lo scree test, può essere eseguito usando la funzione `VSS.scree()` contenuta nel pacchetto `psych`.

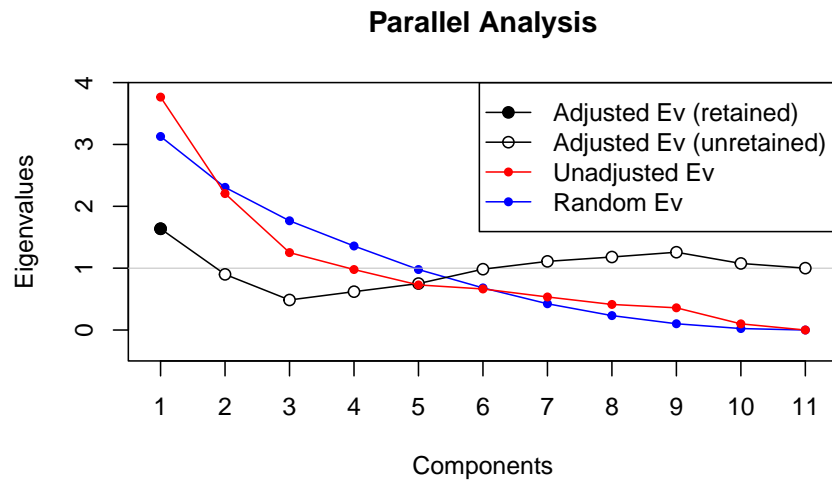
```
VSS.scree(wais_cor)
```



Lo scree test suggerisce una soluzione a $m = 1$ fattori, come indicato nella figura precedente. Il terzo metodo, nella versione della Parallel Analysis, può essere eseguito usando la funzione `paran()` contenuta nel pacchetto `paran`. La Parallel Analysis indica una soluzione a $m = 1$ fattore.

```
library("paran")
paran(wais_cor, graph = TRUE)
#>
#> Using eigendecomposition of correlation matrix.
#> Computing: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
#>
#>
#> Results of Horn's Parallel Analysis for component retention
#> 330 iterations, using the mean estimate
#>
#> -----
#> Component    Adjusted    Unadjusted    Estimated
#>              Eigenvalue  Eigenvalue    Bias
#> -----
#> 1            1.636878    3.765744     2.128865
#> -----
#>
```

```
#> Adjusted eigenvalues > 1 indicate dimensions to retain.
#> (1 components retained)
```



Il quarto metodo consiste nell'applicazione di un test inferenziale relativo al numero di fattori. Anche questo metodo indica una soluzione a sei fattori:

```
factanal(covmat = wais_cor, factors = 5, n.obs = 933)
#>
#> Call:
#> factanal(factors = 5, covmat = wais_cor, n.obs = 933)
#>
#> Uniquenesses:
#>   IN   CO   AR   SI   DS   VO   SY   PC   BD
#> 0.235 0.389 0.117 0.419 0.600 0.109 0.277 0.308 0.334
#>   PA   OA
#> 0.472 0.456
#>
#> Loadings:
#>   Factor1 Factor2 Factor3 Factor4 Factor5
#> IN  0.745   0.264   0.301   0.192   0.118
```



```

#> CO 0.667 0.278 0.244 0.129 0.111
#> AR 0.378 0.236 0.814 0.145
#> SI 0.591 0.332 0.207 0.252 0.121
#> DS 0.288 0.208 0.366 0.341 0.155
#> VO 0.865 0.216 0.207 0.229
#> SY 0.251 0.364 0.153 0.708
#> PC 0.425 0.548 0.156 0.216 0.375
#> BD 0.246 0.708 0.230 0.201 0.107
#> PA 0.355 0.457 0.163 0.325 0.245
#> OA 0.211 0.664 0.128 0.205
#>
#>
#>          Factor1 Factor2 Factor3 Factor4 Factor5
#> SS loadings    2.799   1.986   1.176   1.043   0.280
#> Proportion Var 0.254   0.181   0.107   0.095   0.025
#> Cumulative Var 0.254   0.435   0.542   0.637   0.662
#>
#> Test of the hypothesis that 5 factors are sufficient.
#> The chi square statistic is 12.46 on 10 degrees of freedom.
#> The p-value is 0.256

```

Le differenze tra i risultati ottenuti con i quattro metodi descritti sopra suggeriscono la presenza di una componente di arbitrarietà nella scelta della soluzione da adottare.



15

La rotazione fattoriale

Nel capitolo 13 abbiamo visto come sia possibile ottenere la soluzione fattoriale non ruotata per il numero di fattori comuni che meglio riassume l'informazione contenuta nella matrice di correlazioni (o covarianze). La soluzione non ruotata non garantisce l'identificazione di aggregati omogenei e interpretabili di variabili osservate. Si tende dunque a ricorrere alla rotazione degli assi fattoriali nella ricerca di una soluzione più facilmente interpretabile di quella ottenuta in prima istanza.

15.1 Indeterminatezza della soluzione fattoriale

Il problema della rotazione si pone perché la matrice delle saturazioni non presenta un'unica soluzione e, attraverso la sua trasformazione matematica, si possono ottenere infinite matrici dello stesso ordine. Tale fatto va sotto il nome di *indeterminatezza della soluzione fattoriale*.

La matrice delle saturazioni fattoriali Λ non risulta univocamente definita in quanto non esiste una soluzione unica alla determinazione delle saturazioni fattoriali. Una matrice di correlazioni R consente di determinare soluzioni fattoriali diverse, ovvero matrici aventi lo stesso numero di fattori comuni ma una diversa configurazione di saturazioni fattoriali, oppure matrici di saturazioni fattoriali corrispondenti ad un diverso numero di fattori comuni.

Siano Λ_1 e Λ_2 due matrici aventi lo stesso numero di righe e colonne, ma contenenti saturazioni fattoriali diverse. Λ_1 è definita dai valori seguenti

```
l1 <- matrix(c(
  0.766, -0.232,
  0.670, -0.203,
```

```

    0.574, -0.174,
    0.454,  0.533,
    0.389,  0.457,
    0.324,  0.381
  ),
  byrow = TRUE, ncol = 2
)
```

mentre per Λ_2 abbiamo

```

l2 <- matrix(c(
  0.783,  0.163,
  0.685,  0.143,
  0.587,  0.123,
  0.143,  0.685,
  0.123,  0.587,
  0.102,  0.489
),
  byrow = TRUE, ncol = 2
)
```

Esaminiamo la matrice delle correlazioni riprodotte dalle due matrici di pesi fattoriali (con le comunaltà sulla diagonale di R):

```

l1 %*% t(l1)
#>      [,1] [,2] [,3] [,4] [,5] [,6]
#> [1,] 0.6406 0.5603 0.4801 0.2241 0.1920 0.1598
#> [2,] 0.5603 0.4901 0.4199 0.1960 0.1679 0.1397
#> [3,] 0.4801 0.4199 0.3598 0.1679 0.1438 0.1197
#> [4,] 0.2241 0.1960 0.1679 0.4902 0.4202 0.3502
#> [5,] 0.1920 0.1679 0.1438 0.4202 0.3602 0.3002
#> [6,] 0.1598 0.1397 0.1197 0.3502 0.3002 0.2501
l2 %*% t(l2)
#>      [,1] [,2] [,3] [,4] [,5] [,6]
#> [1,] 0.6397 0.5597 0.4797 0.2236 0.1920 0.1596
#> [2,] 0.5597 0.4897 0.4197 0.1959 0.1682 0.1398
#> [3,] 0.4797 0.4197 0.3597 0.1682 0.1444 0.1200
#> [4,] 0.2236 0.1959 0.1682 0.4897 0.4197 0.3496
```

```
#> [5,] 0.1920 0.1682 0.1444 0.4197 0.3597 0.2996
#> [6,] 0.1596 0.1398 0.1200 0.3496 0.2996 0.2495
```

Come si vede, viene ottenuto lo stesso risultato utilizzando matrici Λ con lo stesso numero m di colonne ma saturazioni fattoriali diverse.

Si consideri ora il caso di matrici Λ corrispondenti a soluzioni fattoriali con un diverso numero di fattori comuni. Siano Λ_1 e Λ_2 due matrici aventi lo stesso numero di righe ma un numero diverso di colonne:

```
l1 <- matrix(c(
  0.9,
  0.7,
  0.5,
  0.3
),
byrow = TRUE, ncol = 1
)

l2 <- matrix(c(
  0.78, 0.45,
  0.61, 0.35,
  0.43, 0.25,
  0.25, 0.15
),
byrow = TRUE, ncol = 2
)
```

Si noti che la stessa matrice di correlazioni riprodotte (con le comunali-
tà sulla diagonale principale) viene generata dalle saturazioni fattoriali
corrispondenti ad un numero diverso di fattori comuni:

```
l1 %*% t(l1)
#>      [,1] [,2] [,3] [,4]
#> [1,] 0.81 0.63 0.45 0.27
#> [2,] 0.63 0.49 0.35 0.21
#> [3,] 0.45 0.35 0.25 0.15
#> [4,] 0.27 0.21 0.15 0.09
```

```

l2 %*% t(l2)
#>      [,1]  [,2]  [,3]  [,4]
#> [1,] 0.8109 0.6333 0.4479 0.2625
#> [2,] 0.6333 0.4946 0.3498 0.2050
#> [3,] 0.4479 0.3498 0.2474 0.1450
#> [4,] 0.2625 0.2050 0.1450 0.0850

```

15.2 Parsimonia e semplicità

Come si raggiunge allora una qualche certezza sui risultati dell'analisi fattoriale? Il problema dell'*indeterminazione fattoriale* si affronta scegliendo la soluzione che soddisfa i seguenti due criteri: *criterio della parsimonia*: se sia un modello ad un fattore comune sia un modello a due fattori comuni possono spiegare la covarianza tra le variabili si deve accettare quello ad un fattore; *criterio della semplicità*: a parità di numero di fattori, sono da preferire le strutture più semplici della matrice Λ (Thurstone, 1947).

Il criterio della parsimonia è facilmente applicabile: se due soluzioni fattoriali aventi un numero diverso di fattori riproducono allo stesso modo la matrice \mathbf{S} o \mathbf{R} , si sceglie la soluzione con il numero minore di fattori. D'altra parte, se vi sono diverse soluzioni fattoriali con lo stesso numero m di fattori, il criterio della semplicità ci guida nella scelta della trasformazione più appropriata della matrice $\hat{\Lambda}$. La trasformazione della matrice $\hat{\Lambda}$ va sotto il nome di *rotazione*. A seconda che i fattori ruotati risultino o meno incorrelati, si distingue tra metodi di rotazione ortogonale o obliqua dei fattori.

15.2.1 Il criterio della “struttura semplice”

Tramite la rotazione degli assi fattoriali miriamo alla “struttura semplice” della matrice delle saturazioni fattoriali: poche ma forti saturazioni diverse da zero e assenza di variabili saturate da più di un fattore. Il criterio della “struttura semplice” è stato originariamente proposto da Thurstone (1947) secondo il quale tale criterio viene raggiunto quando:

- nella matrice fattoriale ruotata, ogni variabile deve avere almeno un

peso nullo;

- ogni fattore deve avere almeno m saturazioni nulle (m : numero dei fattori comuni);
- per ciascuna coppia di fattori vi devono essere saturazioni basse su un fattore e saturazioni alte sull'altro;
- nel caso di molti fattori, per ciascuna coppia di fattori una grande proporzione di saturazioni dovrebbe essere nulla;
- per ciascuna coppia di fattori, vi dovrebbero essere solo poche saturazioni di entità non trascurabile su entrambi i fattori.

Nella pratica, il requisito della struttura semplice viene perseguito, non tanto seguendo le indicazioni di Thursone, quanto bensì cercando di massimizzare il numero di saturazioni nulle o quasi nulle nella matrice $\hat{\Lambda}$. Uno dei grandi vantaggi che derivano dall'ottenimento della struttura semplice è la facilitazione nell'interpretazione dei fattori (Cattell, 1978).

L'esame delle saturazioni fattoriali contenute nella matrice $\hat{\Lambda}^*$ ruotata consente infatti di fornire un'interpretazione ai fattori. Per poter interpretare un fattore, dobbiamo chiederci quali sono le variabili che risultano maggiormente associate con tale fattore e quanto forti siano tali legami. Se i coefficienti di impatto di un fattore sono positivi e piuttosto elevati su un sottoinsieme di variabili osservate, da ciò deduciamo che il fattore rappresenta ciò che hanno in comune le variabili che saturano sul fattore. Ovviamente, l'interpretazione si complica nel caso di variabili che saturano su più fattori.



15.3 Rotazione nello spazio geometrico

15.3.1 Rotazione ortogonale

Come è stato notato nella sezione precedente, la matrice Λ non è *identificabile* poiché non esiste una soluzione unica alla determinazione delle saturazioni fattoriali: qualunque matrice $\hat{\Lambda}^* = \hat{\Lambda}\mathbf{T}$, dove \mathbf{T} è una matrice ortonormale di ordine m , è in grado di riprodurre la matrice di varianze-covarianze allo stesso modo di $\hat{\Lambda}$. La matrice $\hat{\Lambda}$ è pertanto determinata a meno della moltiplicazione per una matrice ortonormale.

Definizione 15.1. Geometricamente, i pesi fattoriali costituiscono le coordinate di un punto (ci sono tanti punti quante sono le p variabili

manifeste) in uno spazio avente un numero di dimensioni pari al numero m dei fattori.

Dal punto di vista geometrico, il problema dell'indeterminazione fattoriale si può descrivere facendo riferimento alla rotazione rigida dei punti che rappresentano le saturazioni fattoriali attorno l'origine del sistema di coordinate. Tale rotazione rigida lascia invariate le distanze tra i punti (ed è equivalente ad una rotazione (contraria) del sistema di assi cartesiani) e dà luogo ad un nuovo insieme di valori per i pesi fattoriali. Ciascuno di questi insiemi di pesi fattoriali così ottenuti produce la medesima matrice di correlazioni riprodotte dal modello fattoriale. L'indeterminazione fattoriale nasce dal fatto che sono possibili infinite rotazioni diverse degli assi.

15.3.2 Vincoli alla rotazione

Il problema della non identificabilità di $\hat{\Lambda}$ viene generalmente risolto imponendo dei vincoli alla rotazione. Il criterio che ci guida nella scelta di una delle possibili trasformazioni della matrice dei pesi fattoriali è quello della *semplicità* della matrice $\hat{\Lambda}$ (Thurstone, 1947), ovvero la vicinanza dei suoi elementi ai valori 0 e 1. Quanto più ciò si verifica tanto più risulta semplice l'interpretazione dei fattori comuni nei termini delle variabili. L'identificazione dei fattori risulta infatti semplificata se ciascuno di essi è fortemente correlato con un numero limitato di variabili ed è poco correlato con le altre.

Le rotazioni ortogonali lasciano immutate le comunalità nel caso di fattori incorrelati. Questo accade perché qualunque rotazione rigida rispetto all'origine preserva le distanze tra i punti identificati dai pesi fattoriali e, nel caso di fattori incorrelati, la comunalità non è nient'altro che la distanza dall'origine (al quadrato):

$$\hat{h}_i^2 = \sum_{j=1}^m \hat{\lambda}_{ij}^2$$

Rotazioni non ortogonali, però, mutano la quota di varianza spiegata da ciascun fattore, essendo questa data da

$$\frac{\sum_{j=1}^p \hat{\lambda}_{ij}^2}{\text{tr}(\mathbf{S})}$$

oppure da

$$\frac{\sum_{i=1}^p \hat{\lambda}_{ij}^2}{\text{tr}(\mathbf{R})}$$

laddove $\text{tr}(\mathbf{R}) = p$, con $i = 1, \dots, p$ (numero di item) e $j = 1, \dots, m$ (numero di fattori).

Diversi algoritmi sono stati proposti per la rotazione ortogonale dei fattori. Inizieremo ad esaminare una possibile soluzione al problema dell'indeterminazione fattoriale mediante il metodo grafico. Esamineremo poi i metodi Quartimax e Varimax.

15.3.3 Metodo grafico

Come si può ruotare il sistema degli assi? Se ci sono solo $m = 2$ fattori, per ottenere la loro rappresentazione geometrica utilizziamo un sistema di coordinate bidimensionale. L'ispezione visiva del diagramma delle saturazioni fattoriali ci può guidare alla scelta della rotazione più appropriata. Le righe di $\hat{\Lambda}$ corrispondono a coppie di pesi fattoriali ($\hat{\lambda}_{i1}, \hat{\lambda}_{i2}$, con $i = 1, \dots, p$) che possono essere interpretate come le coordinate di p punti (tanti quanti le variabili manifeste). Gli assi del diagramma vengono ruotati di un angolo ϕ in modo tale da portarli il più vicino possibile ai punti presenti nel grafico. Le nuove coordinate ($\hat{\lambda}_{i1}^*, \hat{\lambda}_{i2}^*$) vengono calcolate come $\hat{\Lambda}^* = \hat{\Lambda}\mathbf{T}$, dove

$$\mathbf{T} = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}$$

è una matrice ortogonale 2×2 .

Si considerino i dati di Brown, Williams e Barlow (1984) discussi da Rencher (2002). Ad una bambina di dodici anni è stato chiesto di valutare sette dei suoi conoscenti su cinque variabili: *kind*, *intelligent*, *happy*, *likeable* e *just*. Per queste cinque variabili, la matrice di correlazioni è

```
R <- matrix(c(
  1.00, .296, .881, .995, .545,
  .296, 1.000, -.022, .326, .837,
  .881, -.022, 1.000, .867, .130,
```

```

      .995, .326, .867, 1.000, .544,
      .545, .837, .130, .544, 1.00
    ),
    ncol = 5, byrow = TRUE, dimnames = list(
      c("K", "I", "H", "L", "J"), c("K", "I", "H", "L", "J")
    )
  )
)

```

Dalla matrice **R** estraiamo due fattori con il metodo delle componenti principali:

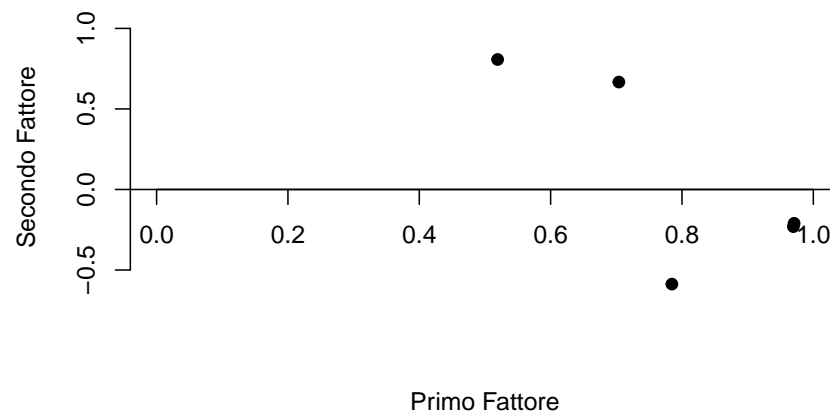
```

library("psych")
f.pc <- principal(R, 2, rotate = FALSE)
f.pc
#> Principal Components Analysis
#> Call: principal(r = R, nfactors = 2, rotate = FALSE)
#> Standardized loadings (pattern matrix) based upon correlation matrix
#>   PC1   PC2   h2    u2 com
#> K 0.97 -0.23 0.99 0.0067 1.1
#> I 0.52  0.81 0.92 0.0792 1.7
#> H 0.78 -0.59 0.96 0.0391 1.9
#> L 0.97 -0.21 0.99 0.0135 1.1
#> J 0.70  0.67 0.94 0.0597 2.0
#>
#>
#>          PC1  PC2
#> SS loadings      3.26 1.54
#> Proportion Var    0.65 0.31
#> Cumulative Var    0.65 0.96
#> Proportion Explained 0.68 0.32
#> Cumulative Proportion 0.68 1.00
#>
#> Mean item complexity = 1.6
#> Test of the hypothesis that 2 components are sufficient.
#>
#> The root mean square of the residuals (RMSR) is 0.03
#>
#> Fit based upon off diagonal values = 1

```

Nella seguente figura, i punti rappresentano le cinque coppie di pesi fattoriali non ruotati:

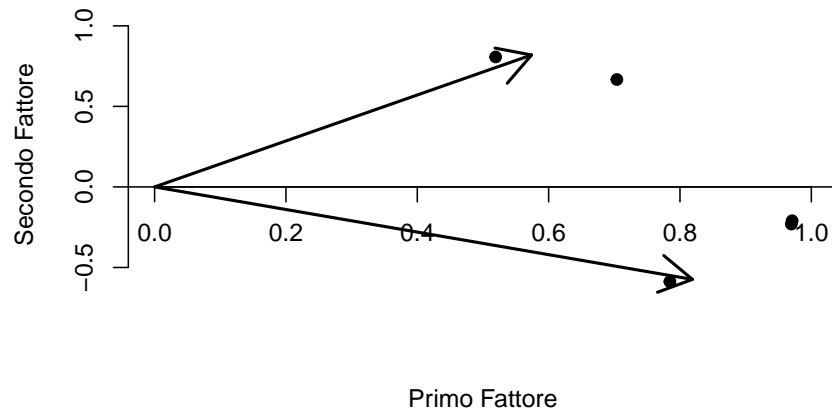
```
plot(
  f.pc$load[, 1], f.pc$load[, 2],
  bty = "n", xaxt = "n",
  xlab = "Primo Fattore", ylab = "Secondo Fattore",
  ylim = c(-.6, 1), xlim = c(0, 1), pch = 19
)
axis(1, pos = c(0, 0))
abline(0, 0)
```



Rencher (2002) nota che, per questi dati, una rotazione ortogonale di -35° ci porterebbe ad avvicinare gli assi ai punti nel diagramma. Per verificare questo, disegniamo sul diagramma i nuovi assi dopo una rotazione di -35° . Le istruzioni R sono le seguenti:

```
plot(
  f.pc$load[, 1], f.pc$load[, 2],
  bty = "n", xaxt = "n",
  xlab = "Primo Fattore", ylab = "Secondo Fattore",
  ylim = c(-.6, 1), xlim = c(0, 1), pch = 19
)
```

```
)  
axis(1, pos = c(0, 0))  
abline(0, 0)  
  
ar <- matrix(c(  
  0, 0,  
  0, 1,  
  0, 0,  
  1, 0  
) , ncol = 2, byrow = TRUE)  
  
angle <- 35  
rad <- angle * pi / 180  
T <- matrix(c(  
  cos(rad), -sin(rad),  
  sin(rad),  cos(rad)  
) , ncol = 2, byrow = TRUE)  
  
round(ar %*% T, 3)  
#>      [,1]  [,2]  
#> [1,] 0.000 0.000  
#> [2,] 0.574 0.819  
#> [3,] 0.000 0.000  
#> [4,] 0.819 -0.574  
  
arrows(0, 0, 0.574, 0.819, lwd = 2)  
arrows(0, 0, 0.819, -0.574, lwd = 2)
```



Nella figura precedente, le due frecce rappresentano gli assi ruotati. È chiaro come tale rotazione di -35° ha effettivamente l'effetto di avvicinare gli assi ai punti del diagramma. Se usiamo dunque il valore $\phi = -35^\circ$ nella matrice di rotazione, possiamo calcolare le saturazioni fattoriali della soluzione ruotata $\hat{\Lambda}^* = \hat{\Lambda}\mathbf{T}$. Le saturazioni fattoriali ruotate non sono altro che la proiezione ortogonale dei punti sugli assi ruotati:

```
angle <- -35
rad <- angle * pi / 180
T <- matrix(c(
  cos(rad), -sin(rad),
  sin(rad),  cos(rad)
), ncol = 2, byrow = TRUE)
round(f.pc$load %*% T, 3)
#>      [,1] [,2]
#> K  0.927 0.367
#> I -0.037 0.959
#> H  0.980 -0.031
#> L  0.916 0.385
#> J  0.194 0.950
```

La soluzione ottenuta in questo modo riproduce quella riportata da Rencher (2002).

15.3.4 Metodi di rotazione ortogonale

Un tipo di rotazione ortogonale molto utilizzato è la rotazione Varimax (Kaiser, 1958). La matrice $\hat{\Lambda}$ è semplificata in modo tale che le varianze dei quadrati degli elementi λ_{ij} appartenenti a colonne diverse di $\hat{\Lambda}$ siano massime. Se le saturazioni fattoriali in una colonna di $\hat{\Lambda}$ sono simili tra loro, la varianza sarà prossima a zero. Tale varianza è tanto più grande quanto più i quadrati degli elementi λ_{ij} assumono valori prossimi a 0 e 1. Amplificando le correlazioni più alte e riducendo quelle più basse, la rotazione Varimax agevola l'interpretazione di ciascun fattore.

Usando la funzione `factanal()` del modulo `base`, la rotazione Varimax può essere applicata alla soluzione ottenuta mediante il metodo di massima verosimiglianza. Usando le funzioni `principal()` e `factor.pa()` disponibili nel pacchetto `psych`, la rotazione Varimax può essere applicata alle soluzioni ottenute mediante il metodo delle componenti principali e il metodo del fattore principale. Ad esempio, otteniamo:

```
f_pc <- principal(R, 2, n.obs = 7, rotate = "varimax")
f_pc
#> Principal Components Analysis
#> Call: principal(r = R, nfactors = 2, rotate = "varimax", n.obs = 7)
#> Standardized loadings (pattern matrix) based upon correlation matrix
#>      RC1   RC2   h2    u2 com
#> K 0.95  0.30 0.99 0.0067 1.2
#> I 0.03  0.96 0.92 0.0792 1.0
#> H 0.97 -0.10 0.96 0.0391 1.0
#> L 0.94  0.32 0.99 0.0135 1.2
#> J 0.26  0.93 0.94 0.0597 1.2
#>
#>                                     RC1  RC2
#> SS loadings                        2.81 1.99
#> Proportion Var                      0.56 0.40
#> Cumulative Var                      0.56 0.96
#> Proportion Explained                0.58 0.42
#> Cumulative Proportion               0.58 1.00
#>
#> Mean item complexity = 1.1
#> Test of the hypothesis that 2 components are sufficient.
#>
```

```
#> The root mean square of the residuals (RMSR) is 0.03
#> with the empirical chi square 0.12 with prob < 0.73
#>
#> Fit based upon off diagonal values = 1
```

Il metodo Quartimax (Neuhaus e Wringley, 1954) opera una semplificazione della matrice $\hat{\Lambda}$ massimizzando le covarianze tra i quadrati degli elementi λ_{ij} appartenenti a righe diverse, subordinatamente alla condizione che la varianza delle righe rimanga inalterata.

15.3.5 Metodi di rotazione obliqua

Parlare di rotazione obliqua significa usare un termine improprio: per definizione, infatti, una rotazione implica una trasformazione ortogonale che preserva le distanze. Secondo Rencher (2002), un termine migliore sarebbe *trasformazione obliqua*. Il termine rotazione obliqua, comunque, fa parte dell'uso corrente.

Nella rotazione obliqua, gli assi della soluzione ruotata non devono rimanere ortogonali e quindi possono più facilmente avvicinarsi ai raggruppamenti di punti nello spazio delle saturazioni fattoriali (assumendo che dei raggruppamenti esistano). Vari metodi analitici sono stati proposti per ottenere una rotazione obliqua. Qui esamineremo brevemente solo uno di essi, il metodo Direct Oblimin.

Il criterio usato nel metodo Direct Oblimin (Jennrich e Sampson, 1966) è il seguente:

$$\sum_{ij} \left(\sum_v \lambda_i^2 \lambda_j^2 - w \frac{1}{p} \sum_v \lambda_i^2 \sum_v \lambda_j^2 \right)$$

dove \sum_{ij} si riferisce alla somma su tutte le coppie di fattori ij . In questo caso si procede ad una minimizzazione piuttosto che a una masssimizzazione.

15.4 Matrice dei pesi fattoriali e matrice di struttura

Nella rotazione ortogonale i fattori sono incorrelati. In tali circostanze, le correlazioni tra le variabili e i fattori sono uguali alle saturazioni fattoriali. In un *path diagram*, infatti, vi è un unico percorso legittimo (in base alle regole di Wright) che collega le variabili manifeste ai fattori.

Si consideri la situazione presentata nella figura 15.1, con due variabili latenti incorrelate (ξ_1 e ξ_2) e quattro variabili manifeste (y_1, y_2, y_3, y_4). Siano $\lambda_{11}, \lambda_{12}, \lambda_{13}$ e λ_{14} le saturazioni fattoriali delle variabili nel primo fattore; siano $\lambda_{21}, \lambda_{22}, \lambda_{23}$ e λ_{24} le saturazioni fattoriali delle variabili nel secondo fattore.

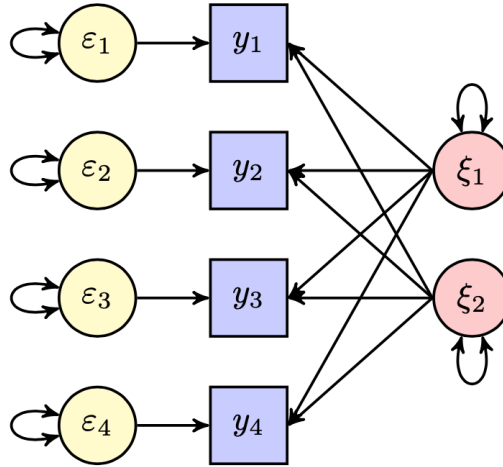


Figura 15.1: Modello fattoriale con due fattori comuni ortogonali.

In un diagramma di percorso, la correlazione tra due variabili contenute è uguale alla somma dei valori numerici di tutti i percorsi legittimi che collegano le variabili. Se i fattori comuni sono incorrelati (come nella figura 15.1, allora c'è un unico percorso che collega ciascuna variabile manifesta a ciascun fattore comune. Le correlazioni tra variabili manifeste e fattori comuni sono dunque uguali ai pesi fattoriali.

In queste circostanze, la soluzione fattoriale contenuta nella matrice delle saturazioni fattoriali rappresenta le correlazioni fra variabili e fattori.

Tale matrice viene detta matrice “di struttura.” Le saturazioni possono essere interpretate in maniera equivalente ai pesi beta del modello di regressione multipla, i quali stimano il contributo specifico di ciascun fattore comune nel determinare la varianza spiegata degli item (Tabachnick & Fidell, 2001).

Invece, nel caso della rotazione obliqua, la soluzione fattoriale ruotata produce un insieme di fattori comuni fra loro correlati. Di conseguenza, i pesi contenuti nella matrice delle saturazioni fattoriali non rappresentano le correlazioni fra variabili e fattori. Nel caso di una rotazione obliqua è perciò necessario specificare tre matrici diverse:

- la matrice dei pesi fattoriali $\hat{\Lambda}$, detta *matrice pattern* (*factor pattern matrix*, o “configurazione,” o “matrice dei modelli”),
- la *matrice di struttura* (*factor structure matrix*), che è la matrice delle correlazioni tra variabili manifeste e fattori,
- la *matrice di intercorrelazione fattoriale* $\hat{\Phi}$ (*factor intercorrelation matrix*), che è la matrice che esprime le correlazioni tra i fattori.

La matrice pattern rappresenta i coefficienti parziali di regressione della variabile sul fattore, al netto degli altri fattori. Nel caso della rotazione obliqua, è la matrice che viene usata per determinare in che grado viene raggiunta la “struttura semplice”.

Esaminiamo ora più in dettaglio la soluzione fattoriale prodotta da una rotazione obliqua. Gli assi che rappresentano i fattori non sono ortogonali (ovvero, i fattori sono correlati) e, in un diagramma di percorso, le variabili manifeste sono collegate ai fattori attraverso due percorsi distinti. Tali percorsi rappresentano l’effetto “diretto” e “indiretto” dei fattori sulle variabili. Nel caso di una rotazione obliqua, come abbiamo detto sopra, le saturazioni fattoriali non coincidono con le correlazioni tra variabili e fattori.

Si consideri la figura 15.2. Nel caso di una rotazione obliqua, la correlazione tra i due fattori comuni viene rappresentata mediante la freccia non direzionata ϕ_{12} che collega ξ_1 e ξ_2 . In tali circostanze, ci sono due percorsi legittimi (in base alle regole di Wright) che consentono di collegare ciascuna variabile manifesta ad un fattore comune. Nel caso della variabile y_1 e il fattore ξ_1 , ad esempio, i percorsi sono: la freccia causale λ_{11} che rappresenta l’effetto diretto di ξ_1 su y_1 e il percorso composto che rappresenta l’effetto indiretto di ξ_1 su y_1 . Il valore numerico di tale percorso composto è uguale al prodotto $\lambda_{21}\phi_{12}$. Nei termini dell’analisi

dei percorsi, dunque, la correlazione tra ξ_1 e y_1 è uguale alla somma dei valori numerici dei percorsi legittimi che collegano y_1 a ξ_1 , ovvero $\lambda_{11} + \lambda_{21}\phi_{12}$.

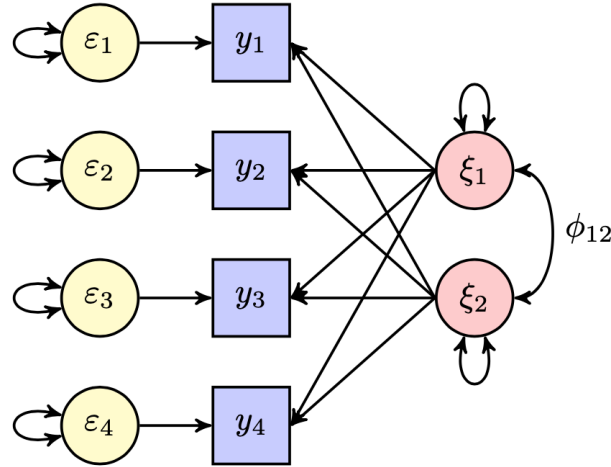


Figura 15.2: Modello fattoriale con due fattori comuni dopo una rotazione obliqua.

Per illustrare la rotazione obliqua, utilizziamo i dati presentati da Rencher (2002). Si consideri la seguente matrice di correlazione:

```
R <- matrix(
  c(
    1.00, 0.735, 0.711, 0.704,
    0.735, 1.00, 0.693, 0.709,
    0.711, 0.693, 1.00, 0.839,
    0.704, 0.709, 0.839, 1.00
  ),
  ncol = 4,
  byrow = TRUE
)
R
#>      [,1] [,2] [,3] [,4]
#> [1,] 1.000 0.735 0.711 0.704
#> [2,] 0.735 1.000 0.693 0.709
```

```
#> [3,] 0.711 0.693 1.000 0.839
#> [4,] 0.704 0.709 0.839 1.000
```

Iniziamo a calcolare una soluzione a due fattori usando il metodo delle componenti principali e una rotazione Varimax. I pesi fattoriali sono:

```
f1_pc <- principal(R, 2, rotate = "varimax")
f1_pc
#> Principal Components Analysis
#> Call: principal(r = R, nfactors = 2, rotate = "varimax")
#> Standardized loadings (pattern matrix) based upon correlation matrix
#>   RC1  RC2   h2   u2 com
#> 1 0.50 0.78 0.86 0.140 1.7
#> 2 0.47 0.81 0.88 0.124 1.6
#> 3 0.90 0.33 0.92 0.078 1.3
#> 4 0.89 0.35 0.92 0.083 1.3
#>
#>                                     RC1  RC2
#> SS loadings                        2.08 1.50
#> Proportion Var                      0.52 0.37
#> Cumulative Var                      0.52 0.89
#> Proportion Explained                0.58 0.42
#> Cumulative Proportion              0.58 1.00
#>
#> Mean item complexity = 1.5
#> Test of the hypothesis that 2 components are sufficient.
#>
#> The root mean square of the residuals (RMSR) is 0.06
#>
#> Fit based upon off diagonal values = 0.99
```

Si noti che i due fattori non sono molto distinti. La soluzione prodotta da una rotazione obliqua si ottiene nel modo seguente:

```
pr_oblimin <- principal(R, 2, rotate = "oblimin")
```

La matrice $\hat{\Lambda}$ dei pesi fattoriali si ricava nel modo seguente:

```
cbind(pr_oblimin$load[, 1], pr_oblimin$load[, 2])
#>      [,1]      [,2]
#> [1,]  0.03207  0.90186
#> [2,] -0.02543  0.95557
#> [3,]  0.96859 -0.01097
#> [4,]  0.94727  0.01328
```

La matrice $\hat{\Phi}$ di intercorrelazione fattoriale è

```
pr_oblimin$Phi
#>      TC1    TC2
#> TC1  1.000  0.787
#> TC2  0.787  1.000
```

La matrice di struttura si ottiene premoltiplicando la matrice Λ dei pesi fattoriali alla matrice Φ di intercorrelazione fattoriale:

$$\text{matrice di struttura} = \Lambda\Phi.$$

Per esempio, la correlazione tra la prima variabile e il primo fattore è

```
pr_oblimin$load[1, 1] + pr_oblimin$load[1, 2] * pr_oblimin$Phi[2, 1]
#>      TC1
#> 0.7418
```

L'intera matrice di struttura si trova nel modo seguente:

```
round(pr_oblimin$load %*% pr_oblimin$Phi, 3)
#>      TC1    TC2
#> [1,] 0.742 0.927
#> [2,] 0.727 0.936
#> [3,] 0.960 0.751
#> [4,] 0.958 0.759
```

16

Come valutare e rifinire la soluzione fattoriale

16.1 Valutazione della matrice pattern

La maggior parte di strumenti usati nell'assessment psicologico e neuropsicologico non valuta una singola dimensione psicologica, ma piuttosto misura molteplici aspetti di un costrutto. Di conseguenza, l'analisi fattoriale produce solitamente una soluzione con più di un fattore. Idealmente, dopo la rotazione, ciascun item saturerà fortemente in un singolo fattore e debolmente negli altri. In realtà, anche dopo la rotazione degli assi fattoriali, talvolta si presentano item che saturano debolmente in tutti i fattori, oppure item che saturano fortemente in più di un fattore.

Uno dei primi passi da compiere per valutare e rifinire la soluzione fattoriale è quello di valutare la matrice di struttura in base al criterio della "struttura semplice" e poi esaminare gli effetti delle azioni intraprese (es., eliminare alcuni item) nella matrice pattern. Ricordiamo che la matrice di struttura contiene le correlazioni tra item e fattori, mentre la matrice pattern contiene le saturazioni fattoriali.

16.1.1 Item con basse saturazioni su tutti i fattori

Prima di procedere all'analisi fattoriale è auspicabile esaminare la matrice di correlazioni tra gli item ed eliminare quegli item che sono insufficientemente correlati con tutti gli altri item nella matrice. Tuttavia, anche dopo questo screening iniziale, è possibile trovare item con basse saturazioni su tutti i fattori. Dal punto di vista pratico, per basse saturazioni si intendono quelle il cui valore assoluto è minore di 0.30 (Hair et al., 1995). Hair e collaboratori suggeriscono due soluzioni nel caso di item con basse saturazioni su tutti i fattori: (1) eliminare gli item con basse saturazioni, (2) valutare le comunaltà degli item problematici e il contributo specifico che forniscono allo strumento. Se un item ha una bassa comunaltà, o se il contributo di un item nei confronti del significa-

to generale dello strumento è di poca importanza, allora l'item dovrebbe essere eliminato. Si procede dunque alla creazione di una nuova soluzione fattoriale e i risultati vengono riesaminati.

Se vi sono degli item con basse saturazioni su tutti i fattori che però contribuiscono in maniera importante a determinare il significato della scala nel suo complesso, allora questi item dovrebbero essere mantenuti. Alle volte, per tali item è possibile creare delle sottoscale separate dalle altre.

16.1.2 Item con saturazioni elevate su più di un fattore

È comune trovare item che saturano su fattori multipli (pesi fattoriali $> .30$), specialmente nel caso di soluzioni fattoriali ottenuti dopo una rotazione obliqua. Kline (2000) suggerisce di eliminare tali item in quanto rendono difficile da interpretare il significato della scala che così si ottiene. Hair e collaboratori (1995) ritengono invece che tali item debbano essere mantenuti dato possono chiarire il significato dei fattori che la scala identifica.

16.2 Valutazione dell'attendibilità

All'interno del problema della costruzione di uno strumento vengono esaminati tre aspetti dell'attendibilità: la consistenza interna, la stabilità e l'equivalenza.

16.2.1 La procedura split-half

La consistenza interna misura il grado di coerenza tra gli item che costituiscono lo strumento o le sottoscale dello strumento. Se tutti gli item che costituiscono uno strumento o una sua sottoscala misurano la stessa cosa, allora saranno fortemente associati tra loro. La consistenza interna misurata con il metodo dello split-half si determina calcolando la correlazione di Pearson tra i punteggi ottenuti utilizzando ciascuna delle due metà degli item dello strumento. Usando un software, è meglio trovare la media delle correlazioni inter-item ricavabili a partire da tutte le possibili divisioni a metà dell'insieme di item che costituiscono lo strumento. La correlazione trovata in questo modo viene poi corretta utilizzando la

formula “profetica” di Spearman-Brown per tenere in considerazione il fatto che l'attendibilità è stata calcolata utilizzando soltanto metà degli item dello strumento.

Si noti che la formula di Spearman-Brown è basata sull'assunzione che le due metà dello strumento siano parallele, ovvero che abbiano identici punteggi veri e uguali varianze d'errore (questa assunzione comporta la conseguenza che le due metà degli item producono punteggi aventi uguali medie e varianze). Se queste assunzioni molto stringenti non vengono soddisfatte, allora ciò conduce ad una sovrastima dell'attendibilità della scala.

16.2.1.0.1 L'analisi della varianza

Se tutti gli item di uno strumento o di una sottoscala sono espressione dello stesso costrutto, allora ci dobbiamo aspettare che anche le medie dei punteggi sugli item siano uguali. Come è stato detto sopra, questa è infatti una delle assunzioni delle forme strettamente parallele di un test. È dunque possibile verificare questa assunzione mediante un'ANOVA che, appunto, sottopone a test l'ipotesi nulla dell'uguaglianza delle medie. Nel caso degli item di un test, dato che ciascun soggetto completa tutti gli item che costituiscono lo strumento, è appropriato usare un'ANOVA per misure ripetute che, nella sua declinazione più moderna, corrisponde ad un modello multilivello (*mixed-effect model*).

16.2.2 L'indice α di Cronbach

L'indice α di Cronbach è la misura più utilizzata per valutare l'attendibilità quale consistenza interna di uno strumento. L' α di Cronbach è stato interpretato come la proporzione di varianza della scala che può essere attribuita al fattore comune (DeVellis, 1991). Può anche essere interpretato come la correlazione stimata tra i punteggi della scala e un'altro strumento della stessa lunghezza tratto dall'universo degli item possibili che costituiscono il dominio del costrutto (Kline, 1986). La radice quadrata del coefficiente α di Cronbach rappresenta la correlazione stimata tra i punteggi ottenuti tramite lo strumento e i punteggi veri (Nunnally & Bernstein, 1994).

In precedenza abbiamo descritto una serie di limiti del coefficiente α di Cronbach. In generale, molti ricercatori suggeriscono di usare al suo posto l'indice ω di McDonald.

16.2.3 L'attendibilità quale stabilità temporale

La stabilità temporale viene valutata attraverso la procedura di test-retest. La correlazione tra le misure ottenute in due momenti negli stessi rispondenti ci fornisce l'attendibilità di test-retest.

Kline (2000) ha messo in evidenza come l'attendibilità di test-retest sia influenzata da molteplici fattori, tra cui le caratteristiche del campione, la maturità dei rispondenti, i cambiamenti nello stato emozionale, le differenze nelle condizioni di somministrazione del test, la possibilità di ricordare le risposte date in precedenza, la difficoltà degli item, la grandezza del campione e le caratteristiche del costrutto (ad esempio, stato vs. tratto).

Particolare attenzione deve essere rivolta all'intervallo temporale usato nella procedura di test-retest. Se il periodo di tempo che intercorre tra le due somministrazioni è troppo corto, i risultati possono risultare distorti a causa del fatto che i soggetti si ricordano le risposte date in precedenza. Questo può condurre ad una sovrastima dell'attendibilità test-retest (Pedhazur & Schmelkin, 1991). Un intervallo temporale troppo lungo tra le due somministrazioni ha invece come limite il fatto che, in questo caso, vi è un'alta possibilità che intervengano dei cambiamenti nei rispondenti rispetto al costrutto in esame. Alla luce di queste considerazioni è stato suggerito di utilizzare un intervallo temporale abbastanza breve, ovvero di una o due settimane (Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991). Se è necessario valutare la stabilità temporale nel corso di un lungo arco temporale, Nunnally e Bernstein (1994) suggeriscono di utilizzare un intervallo di sei mesi o maggiore.

16.2.4 Forme parallele

Per cercare di evitare i problemi associati all'attendibilità quale stabilità temporale, alcuni autori si sono posti il problema di esaminare la correlazione tra forme parallele (o equivalenti) dello strumento. La correlazione tra forme parallele di uno strumento va sotto il nome di coefficiente di equivalenza e fornisce una misura alternativa dell'attendibilità dello strumento (Burns & Grove, 2001; Pedhazur & Schmelkin, 1991; Polit & Hungler, 1999).

Nunnally e Bernstein (1994) suggeriscono di confrontare i risultati ottenuti con la somministrazione delle forme parallele lo stesso giorno con quelli ottenuti nel caso di un intervallo temporale di due settimane. Kli-

ne (2000) ritiene che l'attendibilità tra due forme parallele debba essere di almeno 0.9 perché, per valori inferiori, sarebbe difficile sostenere che le forme sono veramente parallele.

È tuttavia molto oneroso predisporre due forme parallele di uno strumento. Per questa ragione, il coefficiente di equivalenza viene raramente usato.

16.3 Selezione di un sottoinsieme di item

Tipicamente, la costruzione di un test viene realizzata somministrando un grande numero di item per poi selezionare gli item “migliori” che andranno a fare parte del test vero e proprio. Si supponga di somministrare inizialmente m item, quando si desidera che il test finale sia costituito da $p < m$ item. Un modo di affrontare questo problema potrebbe essere quello di calcolare l'attendibilità del test (coefficiente ω) per tutti i possibili sottoinsiemi di p item, così da individuare il sottoinsieme migliore. Questo modo di procedere, però, è problematico perché richiede la valutazione di un elevatissimo numero di possibilità. Per esempio, da un insieme iniziale neanche troppo numeroso di 100 item, il numero di sottoinsiemi di 20 item è uguale a

$$\binom{100}{20} = 5.36 \times 10^{20}$$

È dunque necessario trovare metodi alternativi che evitino una tale esplosione combinatoria.

A questo fine è utile calcolare la *quantità di informazione* di ciascun item:

$$\frac{\lambda_i^2}{\psi_{ii}}.$$

McDonald (2013) mostra che l'omissione di uno o più item produce sempre una riduzione dell'attendibilità del test (ovvero, una riduzione nel valore di ω). Tuttavia, tale riduzione è tanto più piccola quanto più piccola è la quantità di informazione degli item omessi. Il processo di

selezione degli item può dunque essere guidato da un semplice principio: si selezionano gli item aventi la quantità di informazione maggiore.

Esempio 16.1. Si consideri nuovamente la matrice di varianze e di covarianze SWLS:

```
varnames <- c("Y1", "Y2", "Y3", "Y4", "Y5")
SWLS <- matrix(c(
  2.565, 1.424, 1.481, 1.328, 1.529,
  1.424, 2.493, 1.267, 1.051, 1.308,
  1.481, 1.267, 2.462, 1.093, 1.360,
  1.328, 1.051, 1.093, 2.769, 1.128,
  1.529, 1.308, 1.360, 1.128, 3.355
),
ncol = 5, byrow = TRUE,
dimnames = list(varnames, varnames)
)
SWLS
#>      Y1      Y2      Y3      Y4      Y5
#> Y1 2.565 1.424 1.481 1.328 1.529
#> Y2 1.424 2.493 1.267 1.051 1.308
#> Y3 1.481 1.267 2.462 1.093 1.360
#> Y4 1.328 1.051 1.093 2.769 1.128
#> Y5 1.529 1.308 1.360 1.128 3.355
```

Utilizzando la funzione `sem()` contenuta nel pacchetto `sem` il modello ad un fattore viene definito nel modo seguente:

```
mod_1 <- "
  F =~ Y1 + Y2 + Y3 + Y4 + Y5
"
```

Otteniamo così una stima dei pesi fattoriali e delle varianze specifiche:

```
fit <- lavaan::cfa(
  mod_1,
  sample.cov = SWLS,
  sample.nobs = 215,
```

```
std.lv = TRUE
)
```

Esaminiamo la *quantità di informazione* fornita da ciascun item.

```
lambda <- inspect(fit, what = "std")$lambda
lambda
#>      F
#> Y1 0.817
#> Y2 0.694
#> Y3 0.726
#> Y4 0.591
#> Y5 0.643
```

```
theta <- diag(inspect(fit, what = "std")$theta)
theta
#>      Y1      Y2      Y3      Y4      Y5
#> 0.3330 0.5182 0.4732 0.6512 0.5867
```

Possiamo calcolare tale quantità facendo il rapporto tra ciascun peso fattoriale innalzato al quadrato e la sua varianza specifica:

```
for (i in 1:5) {
  print(lambda[i]^2 / theta[i])
}
#>      Y1
#> 2.003
#>      Y2
#> 0.9299
#>      Y3
#> 1.113
#>      Y4
#> 0.5356
#>      Y5
#> 0.7046
```

Tale risultato indica che il quarto item è il meno informativo e che il quinto item è il secondo meno informativo. Se un solo item deve essere

eliminato, dunque, questo sarà il quarto item. Se devono essere eliminati due item, dunque, andranno eliminati il quarto e il quinto item.

In conclusione, possiamo dire che l'omissione di un item provoca sempre una riduzione del valore di ω del test. Tale riduzione è però tanto minore quanto più piccola è la quantità di informazione dell'item omesso. Questo principio può essere usato per selezionare gli item del test.

16.4 Attendibilità e numero di item

Di quanto cambia l'attendibilità di uno strumento se viene variato il numero di item? Una risposta a questa domanda può essere fornita dalla formula profetica di Spearman-Brown. Supponiamo che nella formula di Spearman-Brown,

$$\rho_p = \frac{p\rho_1}{(p-1)\rho_1 + 1}, \quad (16.1)$$

ρ_1 rappresenti l'attendibilità di un test costituito da un certo numero di item. Ponendo $p = 2$, la (16.1) ci fornisce una stima dell'attendibilità di un test costituito da un numero doppio di item rispetto al test originario. Valori di p minori di 1, invece, vengono usati per predire le variazioni dell'attendibilità conseguenti ad una diminuzione nel numero degli item del test.

Ricordiamo che le predizioni della formula di Spearman-Brown sono accurate solo se la forma allungata o accorciata del test è parallela al test considerato. Per esempio, se ad un test con un coefficiente di attendibilità molto alto vengono aggiunti item aventi una bassa attendibilità, allora l'attendibilità del test allungato sarà minore di quella predetta dalla formula di Spearman-Brown.

Si consideri la scala SWLS. Chiediamoci come varia l'attendibilità della scala se il numero di item aumenta da 5 a 20. L'attendibilità della scala SWLS costituita da 5 item è 0.824. Applicando la formula di Spearman-Brown otteniamo:

```
(4 * 0.824) / ((4 - 1) * 0.824 + 1)
#> [1] 0.9493
```

In maniera alternativa, supponiamo che i 15 item aggiuntivi abbiano le stesse saturazioni fattoriali medie ($\bar{\lambda}$) e le stesse varianze specifiche medie ($\bar{\psi}$) rispetto agli item originali. Mediante gli item di cui disponiamo, definiamo l'attendibilità di un "item medio" come

$$\rho_1 = \frac{\bar{\lambda}^2}{\bar{\lambda}^2 + \bar{\psi}}.$$

Ovvero

```
rho_1 <- mean(lambda)^2 / (mean(lambda)^2 + mean(theta))
rho_1
#> [1] 0.4845
```

L'attendibilità predetta di un test costituito da 20 item sarà dunque uguale a

```
(20 * rho_1) / ((20 - 1) * rho_1 + 1)
#> [1] 0.9495
```

il che replica il risultato ottenuto precedentemente.

Un altro modo ancora per ottenere lo stesso risultato è quello di adattare ai dati il modello monofattoriale con item paralleli:

```
mod_2 <- "
  F =~ a*Y1 + a*Y2 + a*Y3 + a*Y4 + a*Y5
  Y1 =~ b*Y1
  Y2 =~ b*Y2
  Y3 =~ b*Y3
  Y4 =~ b*Y4
  Y5 =~ b*Y5
"
```

```
fit2 <- lavaan::cfa(
  mod_2,
  sample.cov = SWLS,
  sample.nobs = 215,
  std.lv = TRUE
)
```

da cui

```
lambda <- inspect(fit2, what = "std")$lambda
lambda
#>      F
#> Y1 0.689
#> Y2 0.689
#> Y3 0.689
#> Y4 0.689
#> Y5 0.689
```

```
theta <- diag(inspect(fit2, what = "std")$theta)
theta
#>      Y1      Y2      Y3      Y4      Y5
#> 0.5247 0.5247 0.5247 0.5247 0.5247
```

```
rho_1 <- lambda[1]^2 / (lambda[1]^2 + theta[2])
rho_1
#>      Y2
#> 0.4753
```

```
(20 * rho_1) / ((20 - 1) * rho_1 + 1)
#>      Y2
#> 0.9477
```

Anche se la formula di Spearman-Brown ha un ruolo centrale nella teoria classica dei test, si tenga conto che non rappresenta l'unico strumento che può essere utilizzato a questo proposito. La quantità detta *informazione dell'item* (*item information*), formulata dai modelli IRT, consente

di predire con maggiore precisione i cambiamenti nella qualità della misura a seguito dell'aggiunta o della cancellazione di un sottoinsieme di item.

16.4.1 Numero di item e affidabilità

La formula di Spearman-Brown può anche essere riarrangiata in maniera tale da consentirci di predire il numero degli item necessari per raggiungere un determinato livello di affidabilità:

$$p = \frac{\rho_p(1 - \rho_1)}{\rho_1(1 - \rho_p)}, \quad (16.2)$$

dove ρ_1 è l'attendibilità stimata di un "item medio," ρ_p è il livello desiderato di attendibilità del test totale e p è il numero di item del test allungato.

L'attendibilità della scala SWLS costituita da 5 item è $\omega = 0.824$. Quanti item devono essere aggiunti se si vuole raggiungere un livello di attendibilità pari a 0.95?

Ponendo $\rho_p = 0.95$ e $\rho_1 = 0.479$, in base alla (16.2) si ottiene che

```
(.95 * (1 - rho_1)) / (rho_1 * (1 - .95))
#>      Y2
#> 20.98
```

il test dovrà essere costituito da 21 item.

16.5 Analisi degli item

L'analisi degli item esamina le risposte fornite ai singoli item del questionario allo scopo di valutare la qualità degli item e del questionario nel suo complesso. Sotto al rubrica di analisi degli item possiamo raggruppare le procedure che possono essere utilizzate per descrivere la difficoltà degli item, le relazioni tra coppie di item, il punteggio totale del test, le relazioni tra gli item e il punteggio totale del test. Tali procedure vengono utilizzate per la selezione degli item al fine di costruire un questionario omogeneo, attendibile e dotato di validità predittiva.

L'esame delle caratteristiche psicometriche degli item che compongono il test non può però essere automatizzata ed eseguita solo sulla base dei risultati delle analisi statiche. L'analisi degli item, invece, va integrata con considerazioni di ordine teorico basate sulla centralità dell'item rispetto alla definizione del costrutto, agli scopi della misurazione e al modo in cui l'item è stato formulato e costruito. Se alcuni aspetti di un costrutto non vengono rappresentati da item che soddisfano i criteri descritti sopra, o se c'è un numero insufficiente di item per produrre uno strumento attendibile, allora alcuni item devono essere riscritti. Nello scrivere gli item, risultano utili le intuizioni che si sono guadagnate nello scrivere gli item che non hanno funzionato.

16.5.1 Difficoltà degli item

La più ovvia caratteristica psicometrica di un item in un test di prestazione massima è la proporzione di soggetti che risponde correttamente all'item. La proporzione p_j di soggetti che rispondono correttamente all'item j -esimo, o si dichiarano in accordo con l'affermazione espressa dall'item, fornisce una stima di π_j , ovvero il *livello di difficoltà* dell'item. In realtà, p_j dovrebbe essere chiamato "facilità dell'item" in quanto assume il suo valore maggiore (ovvero 1) quando tutti i rispondenti rispondono correttamente all'item considerato e il suo valore minimo (ovvero 0) quando le risposte sono tutte sbagliate. I valori p_j giocano un ruolo importante nelle procedure di selezione degli item.

La difficoltà degli item deve essere interpretata in riferimento alla probabilità di indovinare la risposta corretta. Si suppone, infatti, che i rispondenti tirino ad indovinare quando non conoscono la risposta alla domanda di un questionario. Nel caso di item dicotomici, per esempio, ci possiamo aspettare un valore p_j pari a 0.50 sulla base del caso soltanto; nel caso di item a risposta multipla con quattro opzioni di scelta, invece, p_j assume un valore pari a 0.25 quando i rispondenti tirano ad indovinare.

Se il test è composto, per la maggior parte, da item "facili," allora il test non sarà in grado di discriminare tra rispondenti con livelli di abilità diversi, in quanto quasi tutti i rispondenti saranno in grado di fornire una risposta corretta alla maggioranza degli item. Lo stesso si può dire per un test composto da item "difficili." Se consideriamo un test composto unicamente da item di difficoltà media, è facile rendersi conto che non potrà differenziare i rispondenti di abilità media da quelli con abilità

superiori alla media (dato che non ci sono item “difficili”), e neppure da quelli con abilità inferiori alla media (dato che non ci sono item “facili”).

In generale è buona pratica costruire test composti da item che coprano tutti i livelli di difficoltà. La scelta che viene usualmente fatta è quella di una dispersione moderata e simmetrica del livello di difficoltà attorno ad un valore leggermente superiore al valore che sta a metà tra il livello del caso (1.0 diviso per il numero di alternative) e il punteggio pieno (1.0). Per item che presentano quattro alternative di risposta, ad esempio, il livello del caso è pari a $1.00/4 = 0.25$. Il livello ottimale di difficoltà media è dunque uguale a

$$0.25 + (1.00 - 0.25)/2 = 0.62.$$

Per item dicotomici, il livello del caso è $1.00/2 = 0.50$ e il livello ottimale di difficoltà media è

$$0.50 + (1.00 - .50)/2 = 0.75.$$

In generale, item con livelli di difficoltà superiore a 0.90 o inferiore a 0.20 dovrebbero essere utilizzati con cautela.

16.5.2 Correzione per guessing

Alle volte i valori p_j sono calcolati introducendo una correzione per le risposte fornite casualmente dai soggetti (*guessing*). Si consideri un test a scelta multipla composto da item aventi ciascuno C alternative di risposta ed una sola risposta corretta. Si supponga che un rispondente risponda correttamente a R item e risponda in maniera sbagliata a W item.

La correzione per guessing si ottiene applicando una formula basata sul seguente ragionamento. Se assumiamo che un rispondente si limita a tirare ad indovinare allora, ogni C risposte, ci aspettiamo 1 risposta giusta e $C - 1$ risposte sbagliate. Per calcolare il punteggio totale del test in modo da eliminare il numero di risposte corrette ottenute tirando ad indovinare è necessario sottrarre 1 punto per ogni $C - 1$ item a cui è stata fornita una risposta corretta. Questo ragionamento conduce alla seguente formula:

$$FS = R - \frac{W}{C - 1}, \quad (16.3)$$

con $R = \#$ risposte corrette, $W = \#$ risposte sbagliate, $C = \#$ alternative di risposta. Per esempio, se $C = 5$, allora è necessario sottrarre un punto ogni 4, il che è proprio quello che fa la (16.3). La (16.3) per la correzione per guessing produce un punteggio totale corretto per il guessing identico a quello che si otterrebbe pesando ciascuna risposta con 1 punto per le risposte corrette e $-\frac{1}{C-1}$ punti per le risposte sbagliate – senza considerare le risposte non date.

La correzione per guessing rappresenta il tentativo di scomporre il numero totale di risposte corrette in due componenti: le risposte corrette dovute alle conoscenze del soggetto, le risposte che risultano corrette come effetto del caso. La stessa formula può anche essere utilizzata per calcolare la difficoltà degli item corretta per guessing.

16.5.3 Discriminatività

La discriminatività è una misura di quanto ogni item è in grado di distinguere i soggetti con elevati livelli nel costrutto da quelli con un livello basso. L'indice di discriminatività D per i test di prestazione massima si trova nel modo seguente. Dopo avere calcolato il punteggio totale al test, si dividono i soggetti in due gruppi: soggetti con basso punteggio e soggetti con alto punteggio. Una volta definiti i due gruppi, l'indice di discriminatività D sarà dato da:

$$D = P(\text{alto}) - P(\text{basso}),$$

dove $P(\text{alto})$ è la proporzione di soggetti che ha risposto correttamente all'item nel gruppo con punteggi alti e $P(\text{basso})$ è la proporzione di soggetti che ha risposto correttamente all'item nel gruppo con punteggi bassi. Il valore di D può variare da -1 a +1. Nella tabella seguente sono fornite le linee guida per l'interpretazione di questo indice (Ebel, 1965).

Tabella 16.1: Linee guida per l'interpretazione dell'indice di discriminatività D .

Valore di D	Commento
$D \geq 0.40$	Ottima, nessuna revisione
$0.30 \leq D < 0.40$	Buona, revisioni minime
$0.20 \leq D < 0.30$	Sufficiente, revisioni parziali
$D < 0.20$	Insufficiente, riformulazione o eliminazione

La discriminatività degli item di tipo Likert viene valutata con la medesima procedura degli item dei test di prestazione massima, anche se cambiano le procedure statistiche da utilizzare. Si può dividere la distribuzione dei punteggi totali (o punteggi medi) in quartili e confrontare il punteggio medio o mediano del quartile superiore con quello del quartile inferiore, oppure, se il test è orientato al criterio e lo scopo è selezionare gli item che discriminano meglio due gruppi precostituiti di soggetto, eseguire i medesimi confronti tra il gruppo target (ad esempio, pazienti) e quello “di controllo” (per esempio, popolazione generale). È consigliabile valutare la dimensione dell’effetto, ad esempio attraverso l’indice d di Cohen. La dimensione dell’effetto dovrebbe essere almeno moderata ($d > |0.50|$).

16.5.4 Potere discriminante dell’item e analisi fattoriale

Secondo McDondald (1999), la nozione di potere discriminante dell’item può essere trattata in maniera più precisa nell’ambito del modello monofattoriale. Se l’insieme di item a disposizione non è eccessivamente grande (200 o meno), infatti, è possibile procedere alla selezione degli item migliori tramite l’analisi fattoriale – ovvero, scegliendo gli item con le saturazioni maggiori.

16.5.5 Punteggio sull’item e punteggio totale

Il grado di associazione tra il punteggio sull’item e il punteggio totale viene considerato dalla teoria classica dei test quale indice del potere discriminante dell’item. Se il test fornisce una misura attendibile di un unico attributo, e se un item è fortemente associato al punteggio del test, allora l’item medesimo sarà in grado di distinguere tra rispondenti che ottengono un punteggio basso nel test e rispondenti che ottengono un punteggio alto nel test.

Nel caso di una forte associazione positiva tra il punteggio sull’item e il punteggio totale, la probabilità di risposta corretta sull’item è alta per rispondenti che ottengono un punteggio totale alto, e bassa per i rispondenti che ottengono un punteggio totale basso. Nel caso di una debole associazione tra il punteggio sull’item e il punteggio totale, invece, la probabilità di risposta corretta all’item non è predittiva del punteggio totale. Gli item con un basso potere discriminante dovrebbero dunque essere rimossi dal reattivo.

È necessario distinguere i casi in cui gli item sono dicotomici dal caso di item continui. Nel caso di item dicotomici e di un test unidimensionale, il potere discriminante viene calcolato mediante la correlazione biseriale o punto-biseriale.

16.5.6 Relazioni tra coppie di item

Le relazioni tra coppie di item sono importanti sia per la costruzione sia per l'analisi dei test. La teoria classica dei test definisce l'attendibilità di un test (o di un item) come il rapporto tra la varianza del punteggio vero e la varianza del punteggio osservato. Il coefficiente di attendibilità può però essere calcolato anche trovando la correlazione tra due forme parallele di un test (o tra due item). Inoltre, è possibile interpretare la correlazione tra due forme parallele di un test (o tra due item) come il quadrato del coefficiente di correlazione tra i punteggi osservati e i punteggi veri di un test (o di un item).

Molti indici sono disponibili per misurare il grado di associazione tra gli item. Per item quantitativi, possiamo usare la correlazione di Pearson o la covarianza. Per item qualitativi politomici ordinali, usiamo la correlazione policorica. Per item ordinali dicotomici, usiamo la correlazione tetracorica. Per item dicotomici usiamo, ad esempio, l'indice ϕ .

16.5.7 Ridondanza

Nel processo di raffinamento del test occorre tenere conto degli item ridondanti, ossia di quelli che sono troppo associati tra loro. La ridondanza può essere valutata con indici statistici quali la correlazione: se due o più item hanno tra loro una correlazione maggiore di $|0.70|$ viene mantenuto nell'item pool solo un item, dato che gli altri apportano la stessa informazione.

16.5.8 Massimizzazione della varianza del punteggio totale

Uno dei criteri che possono essere utilizzati per selezionare gli item che andranno a costituire la versione finale di un test è quello di massimizzare la varianza del punteggio totale. Più in particolare, si vuole massimizzare il rapporto tra la varianza del punteggio totale e la somma delle varianze dei punteggi dei p item. Dato che il coefficiente α di Cronbach ha la seguente forma:

$$\alpha = \frac{p}{p-1} \left[1 - \frac{\sum \sigma_{X_i}^2}{\sigma_T^2} \right],$$

la scelta di massimizzare il rapporto definito in precedenza avrà anche la conseguenza di massimizzare α .

McDondald (1999) fa notare che una procedura di selezione degli item basata sul principio della massimizzazione di α ha però dei limiti. In primo luogo, tale procedura è appropriata solo quando l'insieme di item è troppo grande per selezionare gli item in base all'esame delle saturazioni fattoriali ottenute applicando il modello monofattoriale. Inoltre, anche se il modello monofattoriale non può essere applicato per problemi pratici, a causa del numero di item troppo grande, tale modello deve essere comunque adeguato anche se si procede selezionando gli item sulla base di α . Ovviamente, però, tale assunzione non può essere verificata. In secondo luogo, McDondald (1999) nota che la procedura di selezione basata sulla massimizzazione di α non è adeguata neanche nel caso in cui il costrutto latente abbia una struttura gerarchica, con un fattore generale ed una serie di fattori subordinati. In quel caso, infatti, la selezione degli item basata sulla massimizzazione di α può portare all'eliminazione degli item che definiscono tutti i fattori subordinati tranne uno. Nel caso in cui il costrutto abbia una struttura gerarchica, dunque, la selezione degli item basata sulla massimizzazione di α deve essere accompagnata da considerazione relative al contenuto del costrutto.



17

Invarianza di misura

I precedenti esempi di CFA presentati in questa dispensa sono stati stimati all'interno di un singolo gruppo, hanno utilizzato come input un'unica matrice covarianza e hanno portato alla stima dei parametri del modello sui quali non è stata imposta alcuna restrizione. In questo capitolo, le analisi precedenti verranno estese considerando il problema dell'invarianza di misura. Quello che ci chiediamo è se sia sensato considerare la medesima struttura fattoriale in gruppi diversi.

17.1 Intercette degli item

In generale, i modelli di equazioni strutturali vengono utilizzati per modellare unicamente la matrice di covarianza delle variabili osservate in un set di dati. Ricordiamo che, quando abbiamo introdotto il modello dell'analisi fattoriale,

$$y_i = \mu + \lambda_j \xi_k + \delta_i,$$

per semplicità abbiamo ignorato la media μ degli indicatori esprimendo i dati osservati nei termini degli scarti dalla media, $y_i - \mu$, in quanto ciò lascia immutate le covarianze. Tuttavia, in alcune applicazioni (quali l'invarianza di misura, ad esempio), è utile considerare anche le medie delle variabili osservate. Per ottenere ciò è possibile fare esplicito riferimento alle intercette delle equazioni precedenti nella sintassi `lavaan` del modello. È possibile fare riferimento all'intercetta di una variabile manifesta nel modo seguente:

```
variable ~ 1
```

laddove la parte sinistra dell'espressione precedente contiene il nome della variabile osservata a cui si fa riferimento. Per esempio, nel caso di un modello a due fattori comuni, è possibile aggiungere le intercette delle variabili manifeste nel modo seguente:

```
mod <- "  
  # two-factor model  
  f1 =~ x1 + x2 + x3  
  f2 =~ x4 + x5 + x6  
  # intercepts  
  x1 ~ 1  
  x2 ~ 1  
  x3 ~ 1  
  x4 ~ 1  
  x5 ~ 1  
  x6 ~ 1  
"
```

Tuttavia, è più conveniente omettere le intercette nella specificazione del modello e aggiungere l'argomento `meanstructure = TRUE` nella funzione `cfa()`:

```
fit <- cfa(  
  mod,  
  data = d,  
  meanstructure = TRUE  
)
```

Si noti che la statistica chi-quadrato e il numero di gradi di libertà sono gli stessi in un modello con o senza `meanstructure`. Il motivo è che, nel caso di un modello con `meanstructure`, vengono introdotti nuovi dati (il valore della media per ciascuna delle p variabili osservate) ma vengono anche aggiunti p parametri da stimare (un'intercetta per ciascuna delle p variabili osservate). Il risultato finale è che la bontà dell'adattamento resta immutata. In pratica, l'unico motivo per aggiungere le intercette nella sintassi del modello è quello di stabilire dei vincoli nella stima di tali parametri.

17.2 Terminologia

La discussione dell'invarianza di misura nel contesto della CFA fa uso della seguente terminologia.

- L'invarianza configurale (*configural invariance*) indica una struttura fattoriale equivalente di due o più gruppi, ovvero la presenza dello stesso numero di fattori e di un pattern di saturazioni fattoriali simili tra i gruppi.
 - L'uguaglianza delle saturazioni fattoriali definisce l'"invarianza metrica" (*metric invariance*) o "invarianza fattoriale debole" (*weak factorial invariance*).
 - L'uguaglianza delle intercette degli indicatori definisce l'"invarianza scalare" o "invarianza fattoriale forte".
 - L'uguaglianza dei residui degli indicatori definisce l'"invarianza fattoriale rigorosa" (*strict factorial invariance*) .
-

17.3 Un esempio concreto

Consideriamo qui un esempio discusso da [Brown \(2015\)](#). Il modello CFA riguarda un modello di misurazione per la depressione maggiore così come definita nel DSM-IV. Ci sono 9 indicatori:

- MDD1, depressed mood;
- MDD2, loss of interest in usual activities;
- MDD3, weight/appetite change;
- MDD4, sleep disturbance;
- MDD5, psychomotor agitation/retardation;
- MDD6, fatigue/loss of energy;
- MDD7, feelings of worthlessness/guilt;
- MDD8, concentration difficulties;
- MDD9, thoughts of death/suicidality.

Leggiamo i dati:

```
d <- readRDS(
  here::here("data", "mdd_sex.RDS")
)
```

Il problema riguarda l'invarianza fattoriale in funzione del genere. Consideriamo il seguente modello:

```
model_mdd <- "
  MDD =~ mdd1 + mdd2 + mdd3 + mdd4 + mdd5 + mdd6 + mdd7 + mdd8 +
        mdd9
  mdd1 ~~ mdd2
"
```

Si noti la presenza di una correlazione residua tra gli indicatori `mdd1` e `mdd2`.

Esaminiamo dunque le varie forme di invarianza fattoriale:

```
# configural invariance
fit_ef <- cfa(
  model_mdd,
  data = d,
  group = "sex",
  meanstructure = TRUE
)
# equal factor loadings
fit_efl <- update(
  fit_ef,
  group.equal = c("loadings")
)
# equal indicator intercepts
fit_eii <- update(
  fit_efl,
  group.equal = c("loadings", "intercepts")
)
# equal indicator error variances
fit_eir <- update(
  fit_eii,
```

```

    group.equal = c("loadings", "intercepts", "residuals")
  )
  # equal factor variances
  fit_fv <- update(
    fit_eir,
    group.equal = c(
      "loadings", "intercepts", "residuals",
      "lv.variances"
    )
  )
  # equal latent means
  fit_fm <- update(
    fit_fv,
    group.equal = c(
      "loadings", "intercepts", "residuals",
      "lv.variances", "means"
    )
  )
)

```

Confrontiamo i modelli:

```

lavTestLRT(fit_ef, fit_efl, fit_eii, fit_eir, fit_fv, fit_fm)
#> Chi-Squared Difference Test
#>
#>      Df   AIC   BIC Chisq Chisq diff Df diff
#> fit_ef  52 27526 27785  98.9
#> fit_efl 60 27514 27735 102.8      3.93      8
#> fit_eii 68 27510 27695 115.3     12.47      8
#> fit_eir 77 27502 27645 125.0      9.71      9
#> fit_fv  78 27501 27639 125.8      0.79      1
#> fit_fm  79 27501 27635 127.7      1.92      1
#>      Pr(>Chisq)
#> fit_ef
#> fit_efl      0.86
#> fit_eii      0.13
#> fit_eir      0.37
#> fit_fv      0.37
#> fit_fm      0.17

```

Il confronto tra i precedenti modelli nidificati che introducono vincoli sempre più tringenti sui parametri indica che non vi è una “significativa” perdita di bontà dell’adattamento passando dal modello congenerico al modello che assume l’uguaglianza delle saturazioni fattoriali, delle intercette, delle varianze residue, delle varianze delle variabili latenti e delle medie nei due gruppi. Questi dati, dunque, forniscono forti evidenze di invarianza fattoriale tra maschi e femmine in relazione al costrutto di depressione maggiore.

18

Indici di bontà dell'adattamento

I passi principali nella CFA e nei modelli SEM comprendono la specificazione del modello, la stima dei parametri, la valutazione del modello e dei parametri e la modificazione del modello. Questa sequenza può essere ripetuta molte volte fino a quando non si trovi un modello considerato accettabile. La valutazione del modello viene eseguita calcolando gli indici di bontà dell'adattamento. In questo capitolo considereremo i principali indici di bontà dell'adattamento utilizzati nella letteratura.

18.1 Stima del modello

L'obiettivo della CFA è ottenere stime per ogni parametro del modello di misura (vale a dire, saturazioni fattoriali, varianze e covarianze fattoriali, varianze residue ed eventualmente covarianze degli errori) che producano una matrice di covarianza prevista (denotata da Σ) che assomiglia il più possibile alla matrice di covarianza delle variabili campionarie (denotata da S). Questo processo fa riferimento ad una funzione di adattamento che consente di ridurre al minimo la differenza tra Σ e S . Il metodo di stima più utilizzato nella CFA (e, in generale, nei modelli SEM) è la massima verosimiglianza (ML).

18.2 Massima verosimiglianza

L'equazione fondamentale dell'analisi fattoriale è

$$y = \Lambda x + z,$$

dove y è un vettore di p componenti (i punteggi del test), x è un vettore

di $k < p$ componenti (i punteggi dei fattori comuni), z è un vettore di p componenti (la componente unica dei punteggi del test) e Λ è una $p \cdot k$ matrice (di saturazioni fattoriali).

Dalle assunzioni del modello fattoriale deriva che

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

Si assume che il vettore casuale y abbia una distribuzione normale multivariata con matrice di covarianza Σ e che da tale distribuzione sia stato estratto un campione casuale di n osservazioni y_1, y_2, \dots, y_n . Il logaritmo della funzione di verosimiglianza per il campione è dato da

$$\log L = \frac{1}{2}n[\log |\Sigma| + \text{tr}(S\Sigma^{-1})].$$

L'equazione precedente viene vista come funzione di Λ e Ψ . Anziché massimizzare $\log L$, è equivalente e più conveniente minimizzare

$$F_k(\Lambda, \Psi) = \log |\Sigma| + \text{tr}[S\Sigma^{-1}] - \log |S| - p,$$

dove $|S|$ è il determinante della matrice di covarianza tra le variabili osservate, $|\Sigma|$ è il determinante della matrice di covarianza prevista e p è il numero di indicatori.

L'obiettivo della stima di massima verosimiglianza della CFA è trovare le stime dei parametri che rendono più probabili i dati osservati (o, al contrario, massimizzano la probabilità dei parametri dati i dati). Le stime dei parametri in un modello CFA si ottengono con una procedura iterativa. Cioè, l'algoritmo inizia con una serie iniziale di stime dei parametri (denominate valori iniziali o stime iniziali, che possono essere generate automaticamente dal software o specificate dall'utente) e raffina ripetutamente queste stime nel tentativo di minimizzare la differenza tra Σ e S . Il programma effettua controlli interni per valutare i suoi progressi nell'ottenere stime dei parametri che al meglio riproducono S . Si raggiunge la convergenza quando l'algoritmo produce una serie di stime dei parametri che non possono essere ulteriormente migliorate per ridurre la differenza tra Σ e S .

18.3 Identificabilità del modello

Un modello CFA deve essere formulato in modo tale da garantire la risolvibilità matematica dello stesso, ovvero deve essere tale da consentire una stima univoca dei parametri del modello. Detto in altre parole, la specificazione del modello ne deve garantire l'identificabilità.

Il problema dell'identificazione richiede, innanzitutto, di chiarire il concetto di gradi di libertà (*degrees of freedom*). Nel presente contesto, per gradi di libertà (df) intendiamo

$$\text{df} = \#(\text{unità di informazione}) - \#(\text{parametri da stimare}).$$

I dati che vengono analizzati da un modello CFA sono contenuti in una matrice di covarianza. Per una matrice di covarianza di ordine p , il numero di unità di informazione è

$$\frac{p(p+1)}{2}.$$

Affinché il modello sia identificabile, devono essere soddisfatte le seguenti condizioni.

1. Indipendentemente dalla complessità del modello (ad es. modelli ad un fattore rispetto a più fattori), l'unità di misura delle variabili latenti deve essere specificata (di solito fissandola a un valore di 1);
2. Indipendentemente dalla complessità del modello, il numero di unità di informazione (es. la matrice di covarianza degli indicatori) deve essere uguale o superiore al numero di parametri da stimare (es. saturazioni fattoriali, specificità, covarianze degli errori dell'indicatore, covarianze tra i fattori);
3. Nel caso di modelli ad un fattore è richiesto un minimo di tre indicatori. Quando vengono utilizzati tre indicatori, la soluzione a un fattore si dice "appena identificata" (*just-identified*); in tali condizioni non è possibile valutare la bontà dell'adattamento.
4. Nel caso di modelli a due o più fattori e due indicatori per costrutto latente, la soluzione è sovraidentificata, a condizio-

ne che ogni variabile latente sia correlata con almeno un'altra variabile latente e gli errori tra gli indicatori siano tra loro incorrelati. Tuttavia, poiché tali soluzioni sono suscettibili di scarsa identificazione empirica, viene raccomandato un minimo di tre indicatori per variabile latente.

In conclusione, una semplice e necessaria condizione per l'identificazione di un modello CFA è che vi siano più unità di informazione che parametri da stimare. Dunque, abbiamo che:

- se $df < 0$, il modello *non è identificato* e, in questo caso, non è possibile stimare i parametri;
- se $df = 0$, il modello è *appena identificato* o “satturo”; in questo caso, la matrice di covarianza riprodotta coincide con la matrice di covarianza delle variabili osservate e, di conseguenza, non esiste un residuo attraverso cui valutare la bontà dell'adattamento del modello;
- se $df > 0$, il modello è *sovra-identificato* ed esistono le condizioni per valutare la bontà dell'adattamento.

Le considerazioni precedenti ci fanno capire perché non si può fare un'analisi fattoriale con solo due indicatori e un fattore; in tali circostanze, infatti, ci sono $(2 \cdot 3)/2 = 3$ gradi di libertà, ma 4 parametri da stimare (due saturazioni fattoriali e due specificità). Il caso di tre item e un fattore definisce un modello “appena identificato”, ovvero, il caso in cui ci sono zero gradi di libertà. In tali circostanze è possibile stimare i parametri (ricordiamo il metodo dell'annullamento della tetrade), ma non è possibile un test di bontà dell'adattamento. Questo vuol dire che, in pratica, è necessario un numero minimo di quattro indicatori per ciascun fattore.

18.4 Bontà dell'adattamento

18.4.1 Chi quadrato

L'indice classico di bontà dell'adattamento dei modelli CFA è il χ^2 . Sotto determinate condizioni, la funzione di discrepanza $F_k(\Sigma, S)$ moltiplicata per n o $n - 1$ (a seconda dei software)

$$nF_k(\Lambda, \Psi) \quad \text{oppure} \quad (n - 1)F_k(\Lambda, \Psi)$$

con n uguale alla numerosità campionaria, si distribuisce come una χ^2 con gradi di libertà (*degrees of freedom*)

$$df = \frac{p(p+1)}{2} - t, \quad (18.1)$$

dove p è il numero di item (variabili osservate) e t è il numero di parametri da stimare.

Sebbene l'indice χ^2 sia stato il primo indice di adattamento ad essere sviluppato, esso è raramente usato nella ricerca applicata quale unico indice di adattamento del modello. Infatti,

- in molti casi (es. piccolo N , dati non normali) la distribuzione sottostante non è χ^2 (il che compromette i test di significatività statistica del modello basati su χ^2);
- χ^2 dipende fortemente dalla dimensione del campione; soluzioni fattoriali in corrispondenza di valori N grandi vengono regolarmente rifiutate sulla base di χ^2 anche quando differenze tra Σ e S sono trascurabili;
- χ^2 si basa sull'ipotesi molto stringente $\Sigma = S$. Come discusso di seguito, molti indici di adattamento alternativi si basano su standard meno stringenti come l'adattamento "ragionevole" e l'adattamento relativo a un modello di indipendenza.

Nonostante questi limiti, la statistica χ^2 viene comunque utilizzata per altri scopi, come il confronto di modelli nidificati, il calcolo di altri indici di adattamento (ad es. l'indice di Tucker–Lewis) e il calcolo del rapporto tra χ^2 e gradi di libertà.

Sebbene la statistica χ^2 sia riportata di routine nell'output dei software che svolgono la CFA, nella valutazione dell'adattamento del modello si fa solitamente affidamento su altri indici di adattamento. Tali indici possono essere suddivisi in tre categorie:

- *misure di adeguamento assoluto* – indicano l'abilità del modello di riprodurre i dati osservati;
- *misure di adeguamento per il confronto o comparative* – permettono di confrontare fra loro 2 o più modelli e di scegliere il migliore (statisticamente);
- *misure di adeguamento parsimonioso* – indici "aggiustati" in base ai gradi di libertà.

18.5 Misure di adeguamento per il confronto

18.5.1 CFI

Gli indici di adattamento comparativo [indicati anche come indici di adattamento incrementale; ad es. [Hu and Bentler \(1998\)](#)] valutano l'adattamento di una soluzione specificata dall'utente in relazione a un modello di base nidificato più ristretto. Tipicamente, questo modello di base è un modello “nullo” o “di indipendenza” in cui le covarianze tra tutti gli indicatori di input sono fissate a zero, ma nessun vincolo viene posto sulle varianze degli indicatori. Uno di questi indici, l'*indice di adattamento comparativo* (*comparative fit index*, CFI; Bentler, 1990), è calcolato come segue.

Sia $\delta = \chi^2 - \text{df}$, dove df sono i gradi di libertà di un particolare modello. Tanto più δ è prossimo allo zero tanto maggiore è la bontà dell'adattamento. La formula di CFI è

$$\text{CFI} = \frac{\delta_B - \delta_T}{\delta_B}, \quad (18.2)$$

dove il pedice T denota il modello target (cioè il modello in valutazione) e il pedice B denota il modello baseline (cioè il modello “nullo”).

18.6 Misure di adeguamento parsimonioso

18.6.1 TLI

Un indice che rientra in questa categoria è l'*indice Tucker-Lewis* (*Tucker-Lewis index*, TLI, anche chiamato indice di adattamento non normato). Il TLI si pone il problema di compensare per l'effetto della complessità del modello, ovvero include una funzione di penalizzazione per l'addizione di parametri che non migliorano notevolmente l'adattamento del modello. Il TLI è calcolato con la seguente formula:

$$\text{TLI} = \frac{(\chi_B^2/\text{df}_B) - (\chi_T^2/\text{df}_T)}{(\chi_B^2/\text{df}_B) - 1}, \quad (18.3)$$

dove χ_T^2 è il valore χ^2 del modello target, df_T sono i gradi di libertà del modello target, χ_B^2 è il valore χ^2 del modello baseline e df_B sono i gradi di libertà del modello base.

18.7 Misure di adeguamento assoluto

18.7.1 RMSEA

L'errore quadratico medio di approssimazione è una misura assoluta dell'adattamento perché non confronta la discrepanza del modello target rispetto a un modello di base, come CFI o TLI. Invece, RMSEA definisce δ come parametro di non centralità che misura il grado di errata specificazione. Ricordiamo dal CFI che $\delta = \chi^2 - df$, dove df sono i gradi di libertà del modello; tanto maggiore è δ tanto più grande è la mancanza di adattamento del modello ai dati. L'indice RMSEA si ottiene nel modo seguente:

$$\text{RMSEA} = \sqrt{\frac{\delta}{df(n-1)}}, \quad (18.4)$$

dove n corrisponde alla numerosità campionaria.

L'indice RMSEA fornisce una stima dell'errore di approssimazione che si commette quando la matrice delle correlazioni (o covarianze) osservate viene riprodotta tramite la matrice ricavata dalle saturazioni fattoriali. Questo indice rappresenta una stima della bontà di adattamento del modello nella popolazione, ponderata per i gradi di libertà e quindi è una misura che tiene in considerazione la parsimonia del modello.

18.7.2 RMRS

L'indice RMRS viene definito come la radice quadrata della media dei residui al quadrato. L'indice RMRS rappresenta la media della correlazione residua, cioè non spiegata dal modello, ed è ricavabile con la seguente formula:

$$\text{RMRS} = \sqrt{\frac{2 \sum_i \sum_j (r_{ij} - \hat{r}_{ij})^2}{q(q+1)}}, \quad (18.5)$$

dove q è il numero di variabili manifeste, e r_{ij} e \hat{r}_{ij} sono rispettivamente la correlazione osservata e la correlazione riprodotta tra le variabili i e j .

18.8 Interpretazione

Un valore RMSEA $< .05$ indica un “close fit” e quello $< .08$ suggerisce un ragionevole adattamento modello-dati. Bentler e Bonett (1980) raccomandano TLI $> .90$ per un adattamento accettabile.

L'interpretazione degli indici di bontà di adattamento trovati nella CFA o nella modellazione di equazioni strutturali può essere ottenuta usando le funzioni del pacchetto **effectsize**. Ad esempio (dal manuale):

```
structure <- " ind60 =~ x1 + x2 + x3
               dem60 =~ y1 + y2 + y3
               dem60 ~~ ind60 "
fit <- lavaan::sem(structure, data = lavaan::PoliticalDemocracy)
effectsize::interpret(fit)
#>      Name      Value Interpretation
#> 1      GFI 0.96664      satisfactory
#> 2     AGFI 0.91243      satisfactory
#> 3      NFI 0.97486      satisfactory
#> 4     NNFI 1.00010      satisfactory
#> 5      CFI 1.00000      satisfactory
#> 6     RMSEA 0.00000      satisfactory
#> 7     SRMR 0.02726      satisfactory
#> 8      RFI 0.95287      satisfactory
#> 9     PNFI 0.51993      satisfactory
#> 10     IFI 1.00005      satisfactory
```

18.9 Una nota di cautela

Nella letteratura SEM sono state sollevate forti argomentazioni contro l'applicazione di RMSEA, CFI e TLI e i loro valori di cutoff convenzionali

(ad esempio, [Barrett, 2007](#)). Tuttavia, prima che i ricercatori propongano e accettino alternative migliori, l'applicazione di questi indici di bontà dell'adattamento continuerà nella maggior parte degli studi SEM. [Xia and Yang \(2019\)](#) fanno notare come, in base alla consuetudine corrente, valori RMSEA più grandi e valori CFI e TLI più piccoli indicano un adattamento peggiore. Ciò spinge i ricercatori a modificare i loro modelli per cercare di ottenere indici migliori. Tuttavia, la pratica attuale si è evoluta in una fase in cui gli indici di adattamento servono come criteri (e gli unici criteri, in molte situazioni) per determinare se accettare o rifiutare un modello ipotizzato: se i valori degli indici di adattamento raggiungono la soglia “di pubblicabilità” (ad es. $RMSEA < .06$), allora non si ritiene più necessario migliorare il modello. In realtà, un'affermazione come “poiché i valori RMSEA, CFI e TLI suggeriscono un buon adattamento, questo modello è stato scelto come modello finale” non è sufficiente. Il raggiungimento di una serie di soglie desiderate di RMSEA, CFI e TLI è solo uno dei possibili indicatori che devono essere considerati nel processo di selezione di modelli. I ricercatori dovrebbero anche spiegare se esistono altre opzioni per migliorare il modello, perché tali opzioni sono o non sono adottate e quali sono le conseguenze scientifiche della scelta del modello in questione come quello finale.



19

Curve di crescita latente

19.1 Dati longitudinali

Un importante classe di modelli a variabili latenti è quella dei modelli delle curve di crescita latente (*latent growth models*, LGM). I modelli delle curve di crescita latente vengono spesso utilizzati per analizzare dati longitudinali. In questo tipo di dati, una misura di esito viene ottenuta in diversi momenti del tempo e si vuole studiare il cambiamento nel tempo. In molti casi, la traiettoria nel tempo può essere modellata come una semplice funzione lineare o quadratica. Le differenze individuali vengono catturate dagli *effetti random* che sono convenientemente rappresentati da variabili latenti (continue), spesso chiamate *fattori di crescita* (*growth factors*).

Possiamo pensare ai modelli LGM come ad un'estensione del modello CFA dotato di `meanstructure`. L'inclusione della `meanstructure` significa che non possiamo usare in input la matrice di covarianza campionaria, ma dobbiamo invece utilizzare i dati grezzi (ovvero, le singole osservazioni per ciascun partecipante). Un'altro requisito degli LGM è che i dati devono essere forniti del formato `wide`, il che significa che ogni colonna rappresenta la variabile di esito in un diverso momento nel tempo. Si presume che ogni osservazione o riga sia indipendente dalle altre; le colonne mostrano invece una dipendenza temporale tra loro.

I modelli LGM sono un caso speciale di CFA e corrispondono un modello CFA a due fattori in cui le saturazioni fattoriali sono fissate a valori predefiniti. Nello specifico, le saturazioni sul fattore che specifica le intercette delle funzioni individuali di crescita sono tutte fissate a 1; le saturazioni sul fattore che specifica le pendenze delle funzioni individuali di crescita sono fissate in modo tale da corrispondere ai diversi momenti del tempo, laddove la saturazione in corrispondenza del momento t_0 è uguale a 0 e la progressione procede con incrementi di 1 (ovvero, $0, 1, \dots, t - 1$).

Si giunge così alla seguente specificazione del modello fattoriale:

$$y_{it} = \tau_i + (1)\xi_1 + (t)\xi_2 + \delta_t, \quad (19.1)$$

con $t = 0, 1, 2, 3$ nel caso di quattro misurazioni temporali. Si noti che, per disegno, $\tau_i = 0$.

19.1.1 Un esempio concreto

L'esempio che discuteremo utilizza i dati di un tutorial `lavaan`, ovvero un campione di dati artificiali chiamato `Demo.growth` in cui un punteggio viene misurato su quattro punti temporali. I dati sono i seguenti:

```
data(Demo.growth)
glimpse(Demo.growth)
#> Rows: 400
#> Columns: 10
#> $ t1 <dbl> 1.7256, -1.9842, 0.3195, 0.7769, 0.4489, ~
#> $ t2 <dbl> 2.1424, -4.4006, -1.2691, 3.5314, -0.7728, ~
#> $ t3 <dbl> 2.7732, -6.0166, 1.5600, 3.1382, -1.5035, ~
#> $ t4 <dbl> 2.51596, -7.02962, 2.86853, 5.36374, 0.078~
#> $ x1 <dbl> -1.1641, -1.7454, 0.9202, 2.3595, -1.0887, ~
#> $ x2 <dbl> 0.1742, -1.5769, -0.1418, 0.7080, -1.0100, ~
#> $ c1 <dbl> -0.02768, -2.03197, 0.05237, 0.01911, 0.65~
#> $ c2 <dbl> 0.55492, 0.12533, -1.25774, 0.64738, 0.730~
#> $ c3 <dbl> 0.25448, -1.56423, -1.80339, -0.43238, -0.~
#> $ c4 <dbl> -1.00640, 1.22927, -0.32726, -1.03240, -0.~
```

La variabile dipendente è rappresentata dalle quattro colonne chiamate `t1`, `t2`, `t3` e `t4` che corrispondono alla serie temporale con quattro misurazioni per ciascun soggetto.

Trasformiamo i dati in formato `long`:

```
demo_growth_long <- Demo.growth %>%
  dplyr::select(t1, t2, t3, t4) %>%
  pivot_longer(
    cols = starts_with("t"),
    names_to = "t",
```



```

    values_to = "y"
  ) %>%
  as.data.frame()
demo_growth_long$time <- rep(0:3, 400)
demo_growth_long$id <- rep(1:400, each = 4)

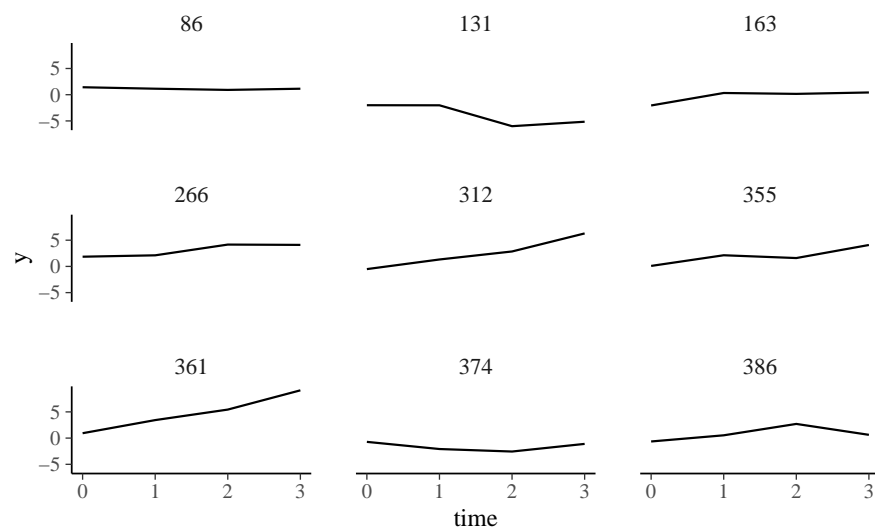
```

Esaminiamo un campione casuale di 9 soggetti. È presente una notevole variazione da soggetto a soggetto:

```

id_sel <- sample(1:400, 9)
d <- demo_growth_long[demo_growth_long$id %in% id_sel, ]
d %>%
  ggplot(aes(x = time, y = y)) +
  geom_line() +
  facet_wrap(~id)

```



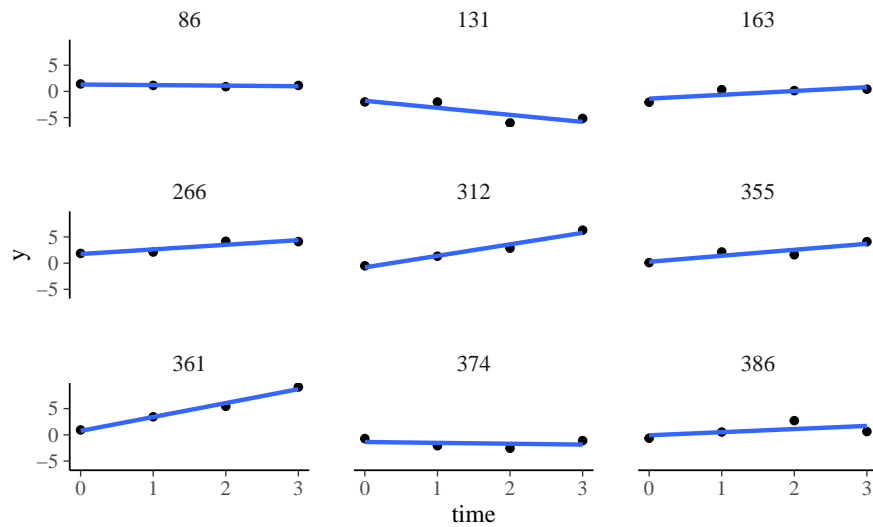
Notiamo che un semplice modello lineare è appropriato per rendere conto della variazione temporale della variabile risposta:

```

d %>%
  ggplot(aes(x = time, y = y)) +

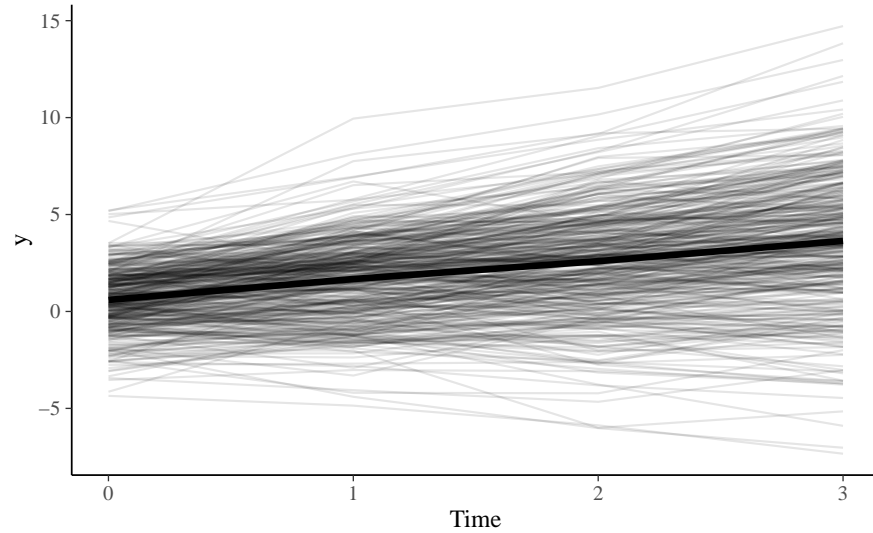
```

```
geom_point() +
stat_smooth(method = "lm", se = FALSE) +
facet_wrap(~id)
```



Complessivamente, i dati suggeriscono un andamento crescente della variabile risposta in funzione del tempo:

```
demo_growth_long %>%
  ggplot(aes(time, y, group = id)) +
  geom_line(alpha = 0.1) + # add individual line with transparency
  stat_summary( # add average line
    aes(group = 1),
    fun = mean,
    geom = "line",
    size = 1.5,
    color = "black"
  ) +
  labs(x = "Time", y = "y")
```



Per adattare un modello lineare di crescita a questi dati, specifichiamo il seguente modello a variabili latenti

$$y_j = \alpha_0 + \alpha_1 \lambda_j + \zeta_{00} + \zeta_{11} \lambda_j + \epsilon_j,$$

dove

- y_j è la variabile di interesse che cambia nel tempo, j .
- α_0 rappresenta l'intercetta della retta di regressione al tempo $t = 0$ (il punto di partenza della linea nera sopra).
- $\alpha_1 \lambda_j$ è il tasso medio di variazione nel tempo (la pendenza della linea nera nel grafico sopra). Qui λ_j è solo l'indice dei punti temporali considerati (0, 1, 2, 3).
- ζ_{00} è la varianza tra i soggetti nel punto $t = 0$.
- $\zeta_{11} \lambda_j$ è la varianza del tasso di variazione tra i soggetti.
- ϵ_j è la varianza di ciascun soggetto attorno alla sua retta di regressione.

Tali relazioni statistiche vengono rappresentate dal modello di equazioni strutturali della figura 19.1.

In altre parole, un lineare di crescita latente corrisponde ad un modello fattoriale con due variabili latenti: un fattore che corrisponde al “punteggio vero” delle intercette individuali e un fattore che corrisponde al “punteggio vero” delle pendenze individuali. Nella sintassi `lavaan` questo diventa:

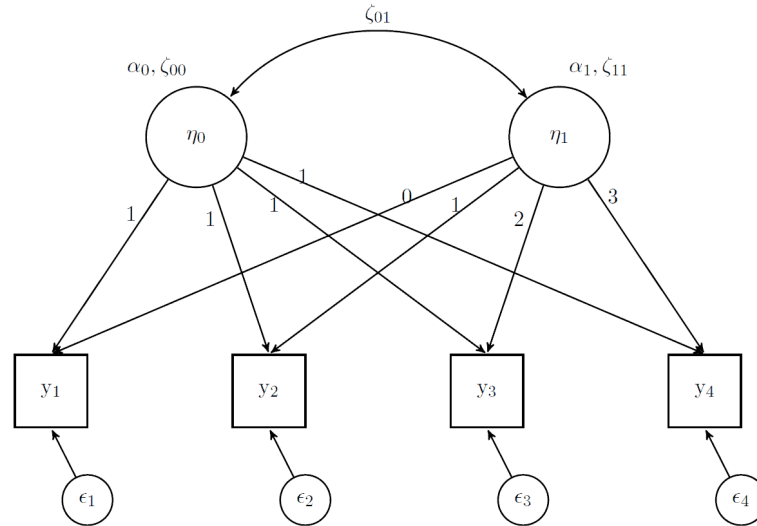


Figura 19.1: Modello di crescita latente.

```
model <- "
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4
"
```

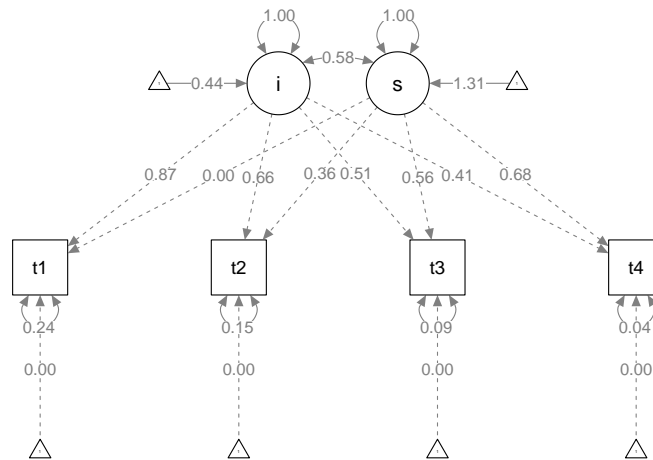
Si noti che, per il fattore η_0 (che rappresenta le intercette), i valori delle saturazioni fattoriali sono fissate a 1 – questo è il motivo per cui α_0 e ζ_{00} compaiono da soli nell’equazione precedente: in maniera esplicita sono $1 \cdot \alpha_0$ e $1 \cdot \zeta_{00}$. Le saturazioni per il fattore η_1 (che specifica le pendenze delle funzioni lineari) sono fissate ai valori dei diversi momenti nel tempo: qui i valori λ_j da 0 a 3. Abbiamo anche la correlazione tra η_0 e η_1 , rappresentata dalla doppia freccia ζ_{01} . Se questo parametro è positivo, questo significa che i partecipanti tendono a diventare sempre più diversi tra loro, al passare del tempo; un’interpretazione opposta si ha se il valore del parametro è negativo.

Adattiamo il modello ai dati:

```
fit <- growth(model, data = Demo.growth)
```

Esaminiamo il path diagram:

```
semPaths(
  fit, "std",
  posCol = c("black"),
  edge.label.cex = 0.9,
  sizeMan = 7,
  what = "path"
)
```



Ci sono 6 tipi di parametri di interesse:

```
kable(coef(fit), booktabs = TRUE, format = "markdown")
```

	x
t1~~t1	0.5953
t2~~t2	0.6760
t3~~t3	0.6349
t4~~t4	0.5076
i~~i	1.9320
s~~s	0.5869
i~~s	0.6179
i~1	0.6147

	x
s~1	1.0063

- l'intercetta $i = 0.615$ è il valore atteso della variabile risposta al momento t_0 ;
- la pendenza $s = 1.006$ è il tasso di cambiamento medio della variabile risposta nel tempo. Ad ogni successivo momento temporale, il valore medio della variabile risposta aumenta in media di 1.006 punti;
- varianza $i = 1.932$ misura la variazione tra i soggetti al momento t_0 (ci dice quanto sono diverse le intercette delle rette di regressione tra i soggetti);
- varianza $s = 0.587$ misura la variazione del tasso di crescita tra i soggetti (ci dice quanto sono diverse le pendenze delle rette di regressione tra i soggetti);
- varianze **t1**, ..., **t4**: i valori da 0.595 a 0.508 ci dicono la variazione tra i soggetti all'interno di ciascun momento del tempo;
- la covarianza tra **i** e **s** = 0.618 ci dice che i valori della variabile risposta diventano via via più diversi nel tempo tra i rispondenti (un valore negativo avrebbe l'interpretazione opposta).

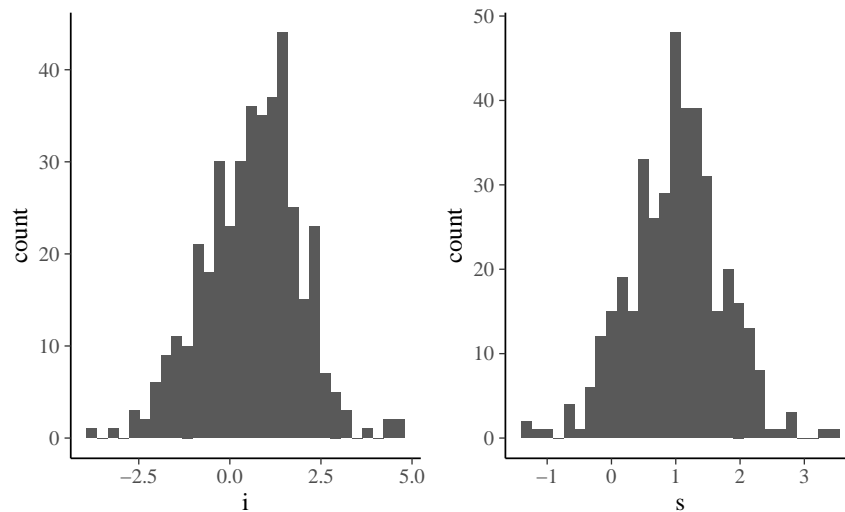
Le stime dell'intercetta e della pendenza della funzione di crescita per ciascun partecipante si ottengono nel modo seguente:

```
rand_eff <- as.data.frame(lavPredict(fit))
head(rand_eff)
#>      i      s
#> 1  1.2276  0.60278
#> 2 -2.6796 -1.38498
#> 3 -0.2955  0.88828
#> 4  1.1576  1.34051
#> 5 -0.4356  0.03193
#> 6 -1.3122  0.50259
```

Le distribuzioni delle stime individuali dell'intercetta e della pendenza della curva di crescita si ottengono nel modo seguente:

```
gi <- rand_eff %>%
  ggplot(aes(x = i)) +
```

```
geom_histogram()
gh <- rand_eff %>%
  ggplot(aes(x = s)) +
  geom_histogram()
gi + gh
```



Un buon modo per capire cosa fa il modello è quello di visualizzare i punteggi previsti. Useremo qui la funzione `predict()` per salvare un nuovo oggetto con i punteggi previsti a livello individuale per l'intercetta e la pendenza;

```
pred_lgm <- predict(fit)
head(pred_lgm)
#>           i           s
#> [1,]  1.2276  0.60278
#> [2,] -2.6796 -1.38498
#> [3,] -0.2955  0.88828
#> [4,]  1.1576  1.34051
#> [5,] -0.4356  0.03193
#> [6,] -1.3122  0.50259
```

Questi sono i valori previsti dal modello per ciascun partecipante. Se

calcoliamo la media di queste variabili otteniamo gli stessi risultati che sono stati riportati sopra:

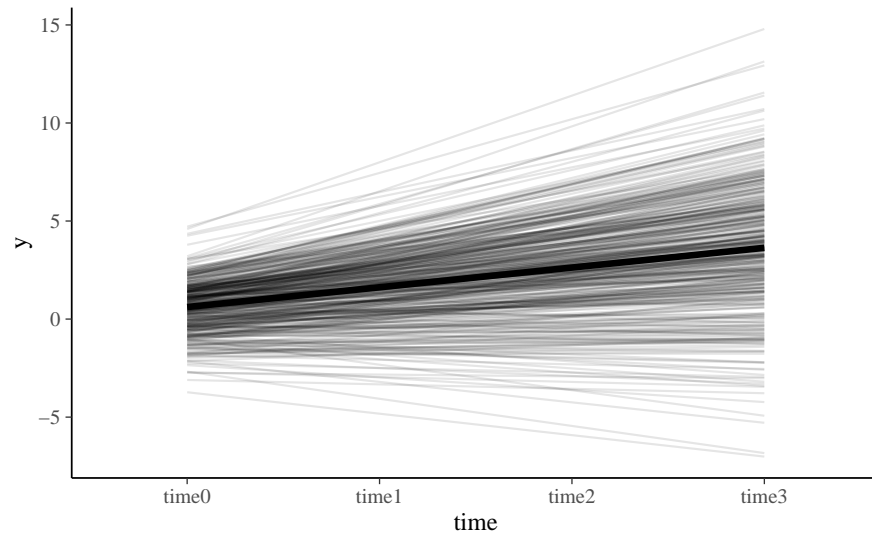
```
# average of the intercepts (first column)
mean(pred_lgm[, 1])
#> [1] 0.6147
# average of the slope (second column)
mean(pred_lgm[, 2])
#> [1] 1.006
```

Il cambiamento nel tempo previsto dal modello per il soggetto j -esimo è

$$y_j = \eta_0 + \lambda_j \eta_1.$$

Il cambiamento previsto per tutti i soggetti può essere visualizzato nel modo seguente:

```
# create long data for each individual
pred_lgm_long <- map(
  0:3, # loop over time
  function(x) pred_lgm[, 1] + x * pred_lgm[, 2]
) %>%
  reduce(cbind) %>% # bring together the wave predictions
  as.data.frame() %>% # make data frame
  setNames(str_c("time", 0:3)) %>% # give names to variables
  mutate(id = row_number()) %>% # make unique id
  gather(-id, key = time, value = pred) # make long format
# make graph
pred_lgm_long %>%
  ggplot(aes(time, pred, group = id)) + # what variables to plot?
  geom_line(alpha = 0.1) + # add a transparent line for each person
  stat_summary( # add average line
    aes(group = 1),
    fun = mean,
    geom = "line",
    size = 1.5,
    color = "black"
  ) +
  labs(y = "y", x = "time")
```

La linea nera più spessa rappresenta l'intercetta media e la pendenza media della curva di crescita del modello LGM. Ogni individuo ha una sua specifica intercetta e uno specifico tasso di cambiamento e questa diversità è catturata nelle componenti di varianza del modello

È anche possibile utilizzare le funzionalità di **lavaan** per verificare specifiche ipotesi di interesse relative ai dati longitudinali. Ad esempio, una possibile domanda riguarda l'uguaglianza degli errori nel tempo. Tale domanda può essere affrontata introducendo dei vincoli nella specificazione del modello LGM e, successivamente, confrontando la bontà dell'adattamento del modello vincolato e del modello generale.

Il modello vincolato è

```
model_eqerr <- "
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4
  t1 =~ a*t1
  t2 =~ a*t2
  t3 =~ a*t3
  t4 =~ a*t4
"
```

Adattiamo il modello vincolato:

```
fit_eqerr <- growth(model_eqerr, data = Demo.growth)
```

Confrontiamo il modello vincolato con il modello libero:

```
anova(fit, fit_eqerr)
#> Chi-Squared Difference Test
#>
#>          Df  AIC  BIC Chisq Chisq diff Df diff
#> fit          5 5528 5564  8.07
#> fit_eqerr    8 5524 5548  9.68      1.61      3
#>          Pr(>Chisq)
#> fit
#> fit_eqerr      0.66
```

Il test del rapporto di verosimiglianza (*likelihood ratio*) eseguito dalla funzione `anova()` produce un *p*-valore di 0.6573. Ciò significa che, nei dati esaminati, non si rileva una decremento della bontà dell'adattamento degna di nota nel passare dal modello libero al modello vincolato. Dunque, l'ipotesi dell'equaglianza della varianza degli errori nel tempo sembra ragionevole.

19.1.2 Un secondo esempio

Un modello leggermente più complesso aggiunge due regressori (*x1* e *x2*) che influenzano i fattori di crescita latenti. Inoltre, è stata aggiunta al modello una covariata *c* variabile nel tempo che influenza la misura del risultato nei quattro punti temporali.

```
model2 <- "
# intercept and slope with fixed coefficients
i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4
# regressions
i ~ x1 + x2
s ~ x1 + x2
# time-varying covariates
t1 ~ c1
t2 ~ c2
```

```

t3 ~ c3
t4 ~ c4
"

```

```
fit2 <- growth(model2, data = Demo.growth)
```

```
kable(coef(fit2), booktabs = TRUE, format = "markdown")
```

	x
i~x1	0.60839
i~x2	0.60411
s~x1	0.26224
s~x2	0.52173
t1~c1	0.14336
t2~c2	0.28900
t3~c3	0.32754
t4~c4	0.33049
t1~~t1	0.57982
t2~~t2	0.59559
t3~~t3	0.48141
t4~~t4	0.53521
i~~i	1.07946
s~~s	0.22376
i~~s	0.07475
i~1	0.58024
s~1	0.95758

I risultati mostrano che le due covariate x influenzano sia l'intercetta sia la pendenza della curva di crescita. Inoltre, vi sono evidenze di un effetto della covariata c .



Parte IV

Appendici



A

Simbologia di base

Per una scrittura più sintetica possono essere utilizzati alcuni simboli matematici.

- $\log(x)$: il logaritmo naturale di x .
- L'operatore logico booleano \wedge significa “e” (congiunzione forte) mentre il connettivo di disgiunzione \vee significa “o” (oppure) (congiunzione debole).
- Il quantificatore esistenziale \exists vuol dire “esiste almeno un” e indica l'esistenza di almeno una istanza del concetto/oggetto indicato. Il quantificatore esistenziale di unicità $\exists!$ (“esiste soltanto un”) indica l'esistenza di esattamente una istanza del concetto/oggetto indicato. Il quantificatore esistenziale \nexists nega l'esistenza del concetto/oggetto indicato.
- Il quantificatore universale \forall vuol dire “per ogni.”
- \mathcal{A}, \mathcal{S} : insiemi.
- $x \in A$: x è un elemento dell'insieme A .
- L'implicazione logica “ \Rightarrow ” significa “implica” (se ...allora). $P \Rightarrow Q$ vuol dire che P è condizione sufficiente per la verità di Q e che Q è condizione necessaria per la verità di P .
- L'equivalenza matematica “ \Leftrightarrow ” significa “se e solo se” e indica una condizione necessaria e sufficiente, o corrispondenza biunivoca.
- Il simbolo $|$ si legge “tale che.”
- Il simbolo \triangleq (o $:=$) si legge “uguale per definizione.”
- Il simbolo Δ indica la differenza fra due valori della variabile scritta a destra del simbolo.
- Il simbolo \propto si legge “proporzionale a.”
- Il simbolo \approx si legge “circa.”
- Il simbolo \in della teoria degli insiemi vuol dire “appartiene” e indica l'appartenenza di un elemento ad un insieme. Il simbolo \notin vuol dire “non appartiene.”
- Il simbolo \subseteq si legge “è un sottoinsieme di” (può coincidere con l'insieme stesso). Il simbolo \subset si legge “è un sottoinsieme proprio di.”

- Il simbolo $\#$ indica la cardinalità di un insieme.
- Il simbolo \cap indica l'intersezione di due insiemi. Il simbolo \cup indica l'unione di due insiemi.
- Il simbolo \emptyset indica l'insieme vuoto o evento impossibile.
- In matematica, argmax identifica l'insieme dei punti per i quali una data funzione raggiunge il suo massimo. In altre parole, $\operatorname{argmax}_x f(x)$ è l'insieme dei valori di x per i quali $f(x)$ raggiunge il valore più alto.
- a, c, α, γ : scalari.
- x, y : vettori.
- X, Y : matrici.
- $X \sim p$: la variabile casuale X si distribuisce come p .
- $p(\cdot)$: distribuzione di massa o di densità di probabilità.
- $p(y \mid x)$: la probabilità o densità di y dato x , ovvero $p(y = Y \mid x = X)$.
- $f(x)$: una funzione arbitraria di x .
- $f(X; \theta, \gamma)$: f è una funzione di X con parametri θ, γ . Questa notazione indica che X sono i dati che vengono passati ad un modello di parametri θ, γ .
- $\mathcal{N}(\mu, \sigma^2)$: distribuzione gaussiana di media μ e varianza σ^2 .
- $\text{Beta}(\alpha, \beta)$: distribuzione Beta di parametri α e β .
- $\mathcal{U}(a, b)$: distribuzione uniforme con limite inferiore a e limite superiore b .
- $\text{Cauchy}(\alpha, \beta)$: distribuzione di Cauchy di parametri α (posizione: media) e β (scala: radice quadrata della varianza).
- $\mathcal{B}(p)$: distribuzione di Bernoulli di parametro p (probabilità di successo).
- $\text{Bin}(n, p)$: distribuzione binomiale di parametri n (numero di prove) e p (probabilità di successo).
- $\mathbb{KL}(p \parallel q)$: la divergenza di Kullback-Leibler da p a q .

B

Lo studio del caso singolo

B.1 Motivazione

In anni recenti c'è stata una sostanziale ripresa di interesse nei confronti degli studi sul caso singolo in neuropsicologia. Tra le ragioni di ciò vi è la convinzione che la prestazione media di un gruppo di pazienti possa essere considerata come un artefatto statistico privo di significato il quale oscura differenze teoricamente importanti tra i pazienti. La versione più estrema di questo argomento è stata fatta da Caramazza e colleghi (ad esempio, Caramazza, 1986; Caramazza e McCloskey, 1988), i quali hanno sostenuto che i neuropsicologi dovrebbero studiare solo casi singoli. Vallar (2000) ha fornito un riassunto conciso di questa posizione: “[Hence] studies in groups of patients which aim at elucidating the neurological and functional architecture of mental processes are useless and harmful, since they provide misleading results. The only appropriate method is to study individual patients. (p. 334)

Questa è una visione minoritaria (anche se influente). Tuttavia, molti altri neuropsicologi hanno sottolineato l'importanza dello studio dei casi singoli (Capitani e Laiacina, 2000; Coltheart, 2001; Ellis & Young, 1996; Shallice, 1988). Ad esempio, la posizione di Vallar (2000) è meno estrema: “Single-case studies have a number of advantages, in comparison with group studies. The probability to produce significant theoretical advances is perhaps higher” (p. 332).

Gli studi sul caso singolo pongono però al neuropsicologo difficili problemi dal punto di vista statistico. Esaminiamo qui alcuni temi di tale discussione: il confronto tra caso singolo e un gruppo di controllo e lo studio delle dissociazioni.

B.2 Il confronto tra caso singolo e un gruppo di controllo

I neuropsicologi hanno spesso bisogno di confrontare un singolo caso con un piccolo gruppo di controllo. Tuttavia, il test t di Student per due campioni indipendenti non può essere applicato perché un gruppo è costituito da un'unica osservazione. Crawford e Garthwaite (2012) dimostrano che il test t di Student messo a punto da Crawford e Garthwaite (2007) fornisce un approccio migliore (in termini del controllo del tasso di errore di Tipo I) rispetto ad altre alternative comunemente usate.

Crawford e Garthwaite (2012) hanno esaminato il problema dello studio del caso singolo in neuropsicologia nel quale le inferenze riguardanti le prestazioni cognitive di un singolo caso vengono confrontate con un campione di controlli sani appaiati. Fino a poco tempo fa il metodo standard per l'inferenza sulla differenza tra un caso e un gruppo di controlli era quello di convertire il punteggio del caso in un punteggio z utilizzando la media campionaria e la deviazione standard del controllo, per poi valutare tale punteggio z con riferimento alla distribuzione normale.

Il problema di questo approccio è che non tiene in considerazione l'incertezza relativa alla media e alla deviazione standard del gruppo di controllo, trattando la media e la deviazione standard del gruppo di controllo come se fossero la media e la deviazione standard della popolazione del gruppo di controllo. Come conseguenza di ciò, l'errore di Tipo I viene inflazionato (in questo contesto si verifica un errore di tipo I quando si conclude che il punteggio del caso non è un'osservazione che appartiene alla popolazione dei punteggi dei controlli) e viene sovrastimata l'anormalità del punteggio del caso singolo. A tale problema sono state proposte diverse soluzioni.

B.2.1 Il metodo di Crawford e Howell (1998)

Questo metodo differisce dall'uso del punteggio z descritto sopra perché tratta la media e la deviazione standard del gruppo di controllo per quello che sono, ovvero statistiche campionarie, e poi utilizza il test t di Student per esaminare la differenza tra caso e controlli. La formula per il test t di Student da applicare nel caso presente è

$$t_{n-1} = \frac{x^* - \bar{x}}{s \sqrt{\frac{n+1}{n}}},$$

dove x^* è il punteggio del paziente, \bar{x} e s sono la media e la deviazione standard del gruppo di controllo, e n è l'ampiezza campionaria del gruppo di controllo.

Se il valore t ottenuto da questo test è minore del quantile t_{n-1} di ordine 0.05, allora si può concludere che il punteggio del caso è sufficientemente basso da consentirci di rifiutare ipotesi nulla che il punteggio del caso singolo appartiene alla popolazione di punteggi dei controlli. Se rifiutiamo l'ipotesi nulla possiamo così concludere che il caso esibisce un deficit relativo all'abilità in questione.

Il valore- p del test unidirezionale prodotto dalla (B.2.1) (a differenza del valore- p del test z discusso sopra) fornisce anche una stima puntuale statisticamente corretta dell'anormalità del punteggio del caso. Quindi se il valore- p è 0.023, allora possiamo affermare che solo il 2.3% della popolazione di controllo otterrà punteggi più bassi (il che significa che, in questo esempio, il punteggio del caso è anormalmente basso).

Esercizio B.1. Una versione bayesiana di questo test è implementata nella funzione `crawford.test()` del pacchetto `psycho`. Supponiamo che il caso mostri un QI pari a 61 mentre il QI dei controlli è $\{86, 100, 112, 95, 121, 102\}$. Ci chiediamo se il QI del caso sia anormalmente basso.

```
library("psycho")
patient <- 61 # The IQ of a patient
controls <- c(86, 100, 112, 95, 121, 102) # The IQs of a control group
crawford.test(patient, controls)
#> The Bayesian test for single case assessment (Crawford, Garthwaite, 2007) suggests that t
```

Utilizzando la (B.2.1) otteniamo un risultato quasi identico:

```
crawford_howell_test <- function(patient, controls) {
  n <- length(controls)
  s <- sd(controls)
  t <- (patient - mean(controls)) / (s * sqrt((n + 1) / n))
}
```

```

    list(t, 1 - pt(t, n - 1))
  }
  crawford_howell_test(patient, controls)
#> [[1]]
#> [1] -3.114
#>
#> [[2]]
#> [1] 0.9868

```

In entrambi i casi la risposta è che il 98.7% dei controlli ha un QI superiore a quello del caso. In tali circostanze possiamo dunque concludere che siamo in presenza di un deficit.

B.2.2 Il metodo di Barton et al. (2002)

Un metodo equivalente alla (B.2.1) è stato proposto da Barton et al. (2002). Questi autori confrontano il valore del caso con l'intervallo di confidenza costruito utilizzando i valori dei controlli:

$$\bar{x} \pm t_{n-1;0.975} \left(s \sqrt{\frac{n+1}{n}} \right),$$

laddove le quantità nella (B.2.2) hanno lo stesso significato che in precedenza. Se il valore del caso singolo cade al di fuori dell'intervallo definito dalla (B.2.2), allora possiamo concludere che il caso manifesta un deficit.

B.2.2.1 L'intervallo di confidenza “tradizionale”

Alcuni autori utilizzano una procedura simile a quella descritta nella (B.2.2), ma utilizzando la formula

$$\bar{x} \pm t_{n-1;0.975} \left(\sqrt{\frac{s^2}{n}} \right),$$

Secondo Crawford e Garthwaite (2012), questo approccio è sbagliato perché, all'aumentare di n , si finirebbe per decidere che il caso manifesta un deficit anche quando ci sono piccolissime differenze rispetto alla media dei controlli. Questo approccio, dunque, è da rifiutare.

B.2.2.2 Il test t di Student – versione 1

Alcuni autori utilizzano la formula del test t di Student nel modo seguente:

$$t_{n-1} = \frac{\bar{x} - x^*}{\frac{s}{\sqrt{n}}}.$$

Per esempio, Reinhold and Markowitsch (2007) descrivono tale procedura nel modo seguente: “Separately for each patient, all comparisons were computed by t-tests for one group with the score of each patient as the tested value (p. > 60)”.

Questo è un uso del tutto bizzarro del test t di Student e non trova giustificazione alcuna.

B.2.2.3 Il test t di Student – versione 2

Altri autori hanno usato la formula del test t di Student scritta nel modo seguente:

$$t_{n-1} = \frac{x^* - \bar{x}}{\frac{s}{\sqrt{n}}},$$

laddove, nuovamente, utilizziamo la media del campione dei controlli come se fosse la media della popolazione. Questo metodo è stato usato in molti studi sul caso singolo, ma conduce ad una grande inflazione dell'errore di I tipo e ad un'esagerazione dell'anormalità del valore del caso singolo.

B.3 Lo studio delle dissociazioni

Parallelamente all'aumento di interesse nei confronti dello studio del caso singolo, le dissociazioni hanno assunto un'importanza sempre maggiore nello sviluppo teorico in neuropsicologia. Ad esempio, Dunn e Kirsner (2003) hanno notato che “dissociations play an increasingly crucial role in the methodology of cognitive neuropsychology ...they have provided critical support for several influential, almost paradigmatic, models in the field” (p. 2).

Ciò è in parte dovuto ai limiti delle strategie alternative. Ad esempio, Vallar (2000) ha notato che le dissociazioni costituiscono “the most effective paradigm for investigating the modularity of the mental processes and their neural correlates”. Lo studio delle dissociazioni in rapporto al caso singolo pone però difficili problemi. Come notato da Vallar (2000), un ricercatore che studia un caso singolo può ritenere che “given the complex architecture of the cognitive system and the variability of the site and extent of naturally occurring lesions, it is very unlikely that two patients have similar functional deficits” (p. 334).

Di conseguenza, Coltheart (2001) ha posto la domanda retorica: “If every patient is unique, how can you replicate your results?” e ha concluso che in alcuni casi potrebbe essere che “the result is literally unreplicable, no matter how genuine”.

Storicamente, gran parte degli studi in neuropsicologia si sono posti il problema di dimostrare l'esistenza di associazioni tra diversi compiti cognitivi. Cioè, è stata posta enfasi sulla rilevazione di gruppi di sintomi cognitivi che si verificano insieme in modo affidabile al fine di identificare sindromi neurologiche o neuropsicologiche. L'insoddisfazione per la scarsa robustezza di tali evidenze è servita quale uno dei fattori trainanti per l'aumento di interesse nei confronti degli studi sul caso singolo. Infatti, indipendentemente dal numero di pazienti che dimostrano la co-occorrenza di deficit in due compiti, l'ipotesi che vi possa essere un processo cognitivo soggiacente comune viene facilmente falsificata dall'osservazione di un singolo paziente che presenta una dissociazione tra i due compiti. Per ritornare al famoso esempio di Karl Popper, l'affermazione “Tutti i cigni sono bianchi” può essere falsificata dalla scoperta di un solo cigno nero.

Tuttavia, sebbene i limiti delle prove associative siano ampiamente riconosciuti, è anche chiaro che ci sono difficoltà pratiche nella valutazione delle prove basate sull'osservazione delle dissociazioni. Ad esempio, supponiamo che ci siano evidenze a sostegno del fatto che un insieme di compiti cognitivi risulta associato con una funzione unitaria sottostante. Se tali evidenze devono essere falsificate da un'unica prova di segno contrario, allora dobbiamo essere sicuri della forza dell'evidenza del contro-esempio. Per continuare con l'esempio precedente, dobbiamo essere sicuri che il cigno sia effettivamente nero e non piuttosto un cigno bianco che, per qualche ragione, si è sporcato di nero. I criteri che ci portano a concludere che vi è una dissociazione nel caso singolo, dunque,

devono essere più stringenti di quelli che si usano nello studio di gruppi di pazienti. Ma quali devono essere tali criteri? Secondo Shallice (1988) evidenze di una ‘forte’ dissociazione sono fornite quando “neither task is performed at normal level, but Task I is performed very much better than Task II”.

Coltheart (2001) fornisce una definizione simile: “One can still speak of dissociations between two tasks even if performance is impaired on both tasks. If a patient is impaired at both Task A and Task B, but is significantly more impaired on the second task than on the first, that can be treated as a dissociation” (p. 12).

Nello studio dell’architettura funzionale alla base della cognizione umana, viene dato grande peso alla doppia dissociazione. Per stabilire l’esistenza di una doppia dissociazione occorrono due pazienti che mostrano pattern opposti di funzioni cognitive preservate e compromesse. Come afferma Coltheart (2001) “With double dissociations we need two patients: patient A who is impaired on Task X but normal on Task Y, and patient B who is normal on Task X but is impaired on Task Y” (p. 12).

Una singola dissociazione non viene considerata come una prova definitiva della divisione funzionale del sistema cognitivo perché i due compiti coinvolti possono attingere a un singolo processo sottostante ma semplicemente differire nella misura in cui vengono determinati da tale processo; cioè, le singole dissociazioni sono soggette agli artefatti derivanti dalla difficoltà del compito. L’esistenza di una doppia dissociazione, invece, è considerata dalla maggior parte dei ricercatori (ma non da tutti) come prova che esclude la difficoltà del compito come spiegazione alternativa dei risultati.

Molti studi su singoli casi usano più misure dei costrutti in esame (cioè usano misure di prestazione in compiti diversi ma correlati per esaminare i costrutti X e Y). In altre parole, il paziente viene confrontato con i controlli su una serie di compiti. Ciò è in linea con il fatto che i ricercatori sono interessati all’esistenza di una dissociazione tra funzioni, non nella dissociazione tra specifiche coppie di misure indirette e imperfette di queste funzioni. Quindi, i ricercatori cercano prove convergenti del deficit o della dissociazione. Tuttavia, l’integrazione di queste molteplici fonti di informazione è un compito estemamente complesso. Su tale aspetto vi è attualmente poca coerenza tra gli studi e i tentativi esistenti tendono ad essere qualitativi piuttosto che quantitativi. Lo sviluppo e la valutazione di un sistema quantitativo, in base al quale le probabilità

di una dissociazione potrebbero essere combinate o aggiornate a seconda delle diverse fasi di uno studio, darebbero un contributo significativo alla disciplina. Per la natura di questo problema, l'approccio basato su metodi statistici bayesiani piuttosto che frequentisti sembra essere la scelta più ovvia.

Esercizio B.2. Concludiamo questa discussione esaminando un caso concreto nel quale è necessario valutare se vi è una dissociazione tra due compiti, nel confronto tra un caso singolo e un piccolo gruppo di confronto. Anche in questo caso viene utilizzato il test t di Student nella versione di Crawford e Howell (1998). Tale test è implementato nella funzione `crawford_dissociation.test()` del pacchetto `psycho`.

```
case_X <- 132
case_Y <- 7
controls_X <- c(100, 125, 89, 105, 109, 99)
controls_Y <- c(7, 8, 9, 6, 7, 10)

result <- crawford_dissociation.test(
  case_X, case_Y, controls_X, controls_Y
)
#> The Crawford-Howell (1998) t-test suggests no dissociation between test X and test Y (t(5)
result
#>      t df      p
#> 1 1.625  5 0.1651
```

B.4 La misura del cambiamento

Un argomento ricorrente nella storia della psicometria è la misurazione del cambiamento. Ad esempio, ci si pone il problema dell'attendibilità della differenza tra due punteggi dei test, ovvero dell'attendibilità della differenza pre-test / post-test. Anche i metodi per studiare il cambiamento di un singolo partecipante sono stati molto discussi. Si possono distinguere studi che valutano le prestazioni di un singolo partecipante in due diversi momenti, oppure studi che esaminano una serie storica con più di due misurazioni.

È importante distinguere tra cambiamenti in un singolo soggetto e cambiamenti nella popolazione. In uno studio sul caso singolo, il neuropsicologo si concentra sul cambiamento di un singolo partecipante, mentre quando si esamina una popolazione di partecipanti il focus è su caratteristiche quali la media e la varianza di una popolazione. Ciò ci conduce ad una distinzione tra diverse interpretazioni che possono essere assegnate al concetto di precisione della misurazione. La teoria classica dei test si basa sul concetto di attendibilità, mentre nei modelli IRT viene usato il concetto di informazione. L'attendibilità è un concetto di precisione della misurazione che dipende dalle proprietà della popolazione. È infatti definita come la correlazione al quadrato dei punteggi osservati nel test e dei punteggi veri in una popolazione di partecipanti (Lord & Novick, 1968). L'informazione, d'altra parte, è un concetto di precisione indipendente dalla popolazione. In concetto di informazione si applica allo stimatore di massima verosimiglianza del valore del tratto latente in un singolo partecipante. La quantità di informazione sul valore del tratto latente di un singolo partecipante è alta se la varianza del valore del tratto latente stimato nel partecipante è piccola, mentre quantità di informazione è bassa se la varianza del valore del tratto latente è grande. Tale concetti si applica anche alle misure di cambiamento.

Quello che è importante notare è che una bassa attendibilità nel test non implica necessariamente una bassa precisione a livello di misurazione del caso singolo. Questa affermazione è valida nel senso che una bassa attendibilità dipendente dalla popolazione non implica necessariamente una bassa quantità di informazione a livello del singolo partecipante. Se l'attendibilità del test è bassa ma la quantità di informazione a livello del singolo partecipante è alta, allora le affermazioni sul cambiamento a livello della popolazione sono imprecise ma le affermazioni sul cambiamento a livello del partecipante possono essere precise. Questa distinzione tra attendibilità e informazione è importante per la misurazione del cambiamento. Per fare affermazioni sul cambiamento al livello del caso singolo, l'aspetto importante della precisione della misurazione è quello che riguarda la quantità di informazione, non l'attendibilità del test.

Un aspetto finale riguarda i cambiamenti a livello del punteggio totale o a livello dei singoli item. Il punteggio totale del test è uguale alla somma ponderata dei punteggi ai singoli item. La teoria classica dei test considera il punteggio totale osservato del test come una variabile continua, mentre il modello IRT considera il punteggio totale del test come una variabile discreta. Il modello IRT si applica agli item con

punteggio dicotomico (0 o 1), in cui il punteggio totale del test è dato dalla somma non ponderata dei punteggi dei singoli item.

Consideriamo ora il caso nel quale lo stesso test viene somministrato ad un singolo partecipante in molteplici occasioni. Poniamoci il problema di misurare il cambiamento dal punto di vista della teoria classica dei test. La teoria classica dei test considera il caso di un punteggio continuo e trae origine da un *Gedankenexperiment* (Lord & Novick, 1968, sezione 2.2) nel quale si suppone che il test venga somministrato ripetutamente allo stesso rispondente e che i punteggi così osservati siano tra loro indipendenti. Il punteggio vero del rispondente è definito come il valore atteso di questa *propensity distribution* (Lord & Novick, 1968, sezione 2.3). La varianza di tale *propensity distribution* è specifica per il singolo rispondente e indica la precisione della misurazione per l'individuo considerato; l'inverso di questa varianza è la quantità di informazione (Mellenbergh, 1996). Il modello classico si può estendere al caso di un cambiamento nella risposta del singolo soggetto.

Il test viene somministrato ad un rispondente in diverse occasioni e il punteggio vero del rispondente nella i -esima occasione è uguale al valore atteso dei punteggi nella i -esima *propensity distribution*. Si assume che le varianze delle *propensity distribution* del rispondente siano omogenee (vale a dire, si assume una eguale varianza in ciascuna delle occasioni). Inoltre, si assume che i punteggi dei test del rispondente in occasioni consecutive siano distribuiti in modo indipendente le une dalle altre. Si pone dunque il problema di come misurare la varianza dei punteggi in somministrazioni ripetute del test.

Sappiamo che la teoria classica dei test fornisce una risposta a questa domanda nei termini della formula seguente,

$$\sigma_E^2 = \sigma_X^2(1 - \rho_{XX'}),$$

dove σ_X^2 e $\rho_{XX'}$ sono, rispettivamente, la deviazione standard dei punteggi del test nella popolazione e l'attendibilità del test. Una stima dell'errore standard della misurazione σ_E^2 si ottiene utilizzando i valori campionari anziché quelli della popolazione riportanti nell'equazione precedente. Uno svantaggio di tale approssimazione è che la stima di σ_E^2 risulta essere uguale per ciascun partecipante, mentre quello che in realtà vorremmo conoscere è una stima della variabilità delle misure di uno specifico soggetto.

La teoria classica dei test procede poi valutando il cambiamento (per esempio, post/pre-test) del k -esimo partecipante nella t -esima occasione

nei termini della seguente equazione:

$$z_{kt} = \frac{\hat{\gamma}_{kt}}{\hat{\sigma}_E} \sqrt{2},$$

laddove

$$\hat{\gamma}_{kt} = x_{kt} - x_{k,t-1},$$

è la differenza dei punteggi osservati per il partecipante k -esimo in due diverse occasioni.

Esercizio B.3. Consideriamo un esempio basato sui dati di Meyers (1978). Meyers selezionò, da un campione più grande di 500 bambine di 13 anni d'età, un campione di 33 bambine che mostravano segni di ansia sociale. Le bambine selezionate furono assegnate in modo casuale a tre diverse condizioni (lista di attesa e due trattamenti diversi). Al pretest e circa 2 mesi dopo, al post-test, fu somministrato alle bambine un test di ansia sociale. Esaminiamo qui un test statistico relativo al cambiamento dei punteggi di una singola bambina. Per la bambina considerata, il punteggio al pre-test era pari a 8 e il punteggio post-test era uguale a 2, laddove un punteggio più basso indica un livello di ansia sociale minore. Il manuale del test riporta una stima dell'errore standard di misurazione pari a 2.6, nel caso di un campione casuale di 1.039 ragazze. Una stima del cambiamento dal pretest al post test è $\hat{\gamma}_{kt} = 2 - 8 = -6$. Usando la (B.4) otteniamo così un punteggio pari a

$$z_{k2} = -6/2.6\sqrt{2} = -1.63,$$

il quale non ci consente di rigettare l'ipotesi nulla di assenza di cambiamento al livello $\alpha = 0.05$ con un test monodirezionale.

B.4.1 Cambiamento nel caso singolo e confronto con un campione di controllo

Mellenbergh e van den Brink (1998) hanno sviluppato le considerazioni precedenti allo scopo di mettere a punto un test che consenta di valutare il cambiamento nel caso singolo mediante un confronto con un campione di controllo. Tale test è implementato nella funzione `mellenbergh.test()` contenuto nel pacchetto `psycho`. Vediamo qui sotto un esempio di applicazione di tale test.

Esercizio B.4. Si consideriano i seguenti dati

```
set.seed(12345)
t0 <- 80 # The IQ of a patient at baseline
t1 <- 100 # The IQ of a patient after the therapy
controls <- round(rnorm(20, 100, 15))
controls
#> [1] 109 111 98 93 109 73 109 96 96 86 98 127
#> [13] 106 108 89 112 87 95 117 104
```

Il test si svolge come indicato di seguito:

```
mellenbergh.test(t0, t1, controls = controls)
#> The Mellenbergh & van den Brink (1998) test suggests that the change is not significant (
```

Considerazioni conclusive

Crawford e Garthwaite (2012) esaminano il problema del confronto tra caso singolo e un gruppo di controllo e concludono che solo la soluzione discussa nel §B.2.1 e quella equivalente discussa nel §B.2.2 sono appropriate per effettuare il confronto. Le altre soluzioni presentate nel §B.2 e molto utilizzate in letteratura, sono inappropriate. Crawford e Garthwaite (2012) concludono che la maggior parte degli studi a caso singolo in neuro-psicologia rivelano una conoscenza sofisticata della teoria cognitiva da parte dei ricercatori. Inoltre, il disegno della ricerca è spesso basato su un'attenta analisi logica delle domande da affrontare e grande attenzione viene prestata allo sviluppo dei compiti sperimentali e dei materiali stimolo. Tuttavia, un'attenzione molto minore viene dedicata ai metodi statistici che vengono impiegati per rispondere alla domanda della ricerca. Secondo Crawford e Garthwaite (2012), è chiaro invece, anche sulla base di una serie di simulazioni Monte Carlo, che i metodi diversi da quello discusso nel §B.2.1, la versione bayesiana equivalente presentata da Crawford e Garthwaite (2007), dovrebbero essere abbandonati perché portano ad un'esagerazione del disturbo. La discussione precedente è stata poi estesa all'analisi statistica dei dati raccolti mediante l'uso di compiti diversi per studiare multicomponenzialità dei processi cognitivi

nel caso singolo e lo studio del cambiamento pre-test / post-test nel caso singolo.



C

La valutazione dell'intervento psicologico

La valutazione neuropsicologica dei pazienti che presentano disturbi cognitivi fornisce informazioni cruciali su cui basare la diagnosi e valutare se ci sono stati cambiamenti nelle condizioni di un individuo. Essa costituisce uno degli strumenti che possono essere usati per la valutazione degli effetti di un trattamento.

Ma come si misurano gli effetti di un trattamento psicologico, in generale, o neuropsicologico, in particolare? Consideriamo il caso di un intervento psicologico. La valutazione di un intervento psicologico è molto diversa rispetto alla valutazione sull'effetto dei farmaci, ad esempio, perché c'è un numero quasi infinito di fattori che possono essere esaminati per decidere se un determinato tipo di intervento psicologico ha portato ad un esito positivo, e semplicemente non c'è modo di controllarli tutti. I ricercatori non concordano neppure su cosa debba essere valutato quando si valutano gli effetti di un trattamento. Miglioramento nel benessere soggettivo? Diminuzione dei sintomi? Cambiamento di personalità? Miglioramento nei rapporti interpersonali? Miglioramento nelle capacità lavorative? Crescita personale, realizzazione e miglioramento della persona? Tutti i precedenti?

Molto spesso gli studi sugli effetti del trattamento considerano punteggi medi di misure di esito basate sui sintomi. Questo modo di quantificare l'esito del trattamento non tiene in considerazione il fatto che, in qualunque trattamento, si osserva un cambiamento in alcuni pazienti ma non in altri. Ciò è ulteriormente complicato dai problemi della comorbidità. È possibile infatti che un trattamento possa produrre un esito favorevole rispetto ad alcune dimensioni di un deficit ma, nel contempo, un esito negativo per altre. Anche se la maggior parte degli studi sugli effetti di un trattamento psicologico esamina la riduzione dei sintomi, il fine ultimo del trattamento dovrebbe essere quello di un miglioramento del funzionamento dell'individuo in contesti sociali e del livello di benessere percepito. Ma non è chiaro come ciò possa essere misurato.

C.1 Efficacia dell'intervento ed efficienza clinica

Indipendentemente dal problema di cosa deve essere misurato, è anche necessario chiedersi quali siano le finalità della valutazione. La letteratura psicologica distingue tra efficacia ed efficienza clinica del trattamento, laddove per efficacia (*efficacy*) si intende la capacità del trattamento di produrre un cambiamento, mentre l'efficienza clinica (*effectiveness*) può essere definita “la capacità di un intervento di produrre gli effetti benefici desiderati nella pratica clinica corrente” e fa anche riferimento al rapporto costi/benefici dell'intervento. Come descritto da Hunsley (2007), “Treatment efficacy studies involve methodological efforts to maximize the internal validity of a study. This commonly includes the use of design features, such as random assignment to treatment and control conditions, training of therapists to a specified level of competence in providing the treatment, and ensuring that all participants have the condition that the treatment was designed to address. Treatment effectiveness studies, on the other hand, strive to maximize external validity while maintaining an adequate level of internal validity (without which, of course, no viable conclusions could be drawn about the impact of the treatment). Most commonly, efforts to enhance external validity involve locating the treatment study within clinical service sites that provide ongoing health services, thus using clinicians who are routinely providing psychological services and patients who have been referred to the clinical settings” (p. 117).

I dati degli studi sull'*efficacy* e sull'*effectiveness* sono fondamentali per una piena comprensione dell'impatto potenziale di un trattamento. Una volta che un trattamento abbia dimostrato la sua *efficacy* attraverso la riproducibilità dei dati di un esperimento, il passo successivo è determinare l'impatto effettivo del trattamento nella pratica clinica. Deve essere infatti dimostrato che i trattamenti esaminati nelle condizioni controllate degli studi di ricerca (studi sull'*efficacy*) possano anche avere un adeguato impatto clinico quando vengono utilizzati nell'effettiva pratica clinica (studi sull'*effectiveness*).

Nel concetto di *effectiveness* si possono individuare quattro componenti: efficacia del trattamento, tollerabilità e sicurezza del trattamento, funzionamento del paziente, accettabilità dell'intervento da parte del paziente. Tutte le componenti interagiscono tra di loro e forniscono informazioni

sull'esito di un determinato intervento. Per quanto riguarda l'efficacia del trattamento, le variabili più importanti sono la scomparsa dei sintomi, il tasso di ricadute, la comorbidità. Gli elementi più importanti che incidono sulla tollerabilità e sicurezza del trattamento sono gli effetti collaterali, la sicurezza e la facilità di assunzione della terapia. Il funzionamento del paziente, il terzo dominio, include le normali attività quotidiane e la qualità della vita, elementi che possono essere quantificati e valutati in studi clinici randomizzati o in studi osservazionali. L'accettabilità di un intervento da parte del paziente infine, è il dominio della compliance, ed è strettamente legato alle altre tre componenti.

Sebbene ciascun dominio possa essere quantificabile singolarmente, è anche possibile misurare l'*effectiveness* in maniera globale. Uno strumento che può essere usato a tale scopo è la scala *Clinical Global Impression* (CGI), che rappresenta il modo più semplice per valutare e quantificare, in generale, l'*effectiveness* di un intervento. Inoltre, la scala della *Valutazione Globale del Funzionamento* (GAF) fornisce un punteggio singolo, composito della funzione psicologica, sociale ed occupazionale su un ipotetico continuum che va dal benessere mentale al funzionamento più deficitario causato dalla malattia.



D

La dimensione dell'effetto del trattamento

I risultati degli studi di ricerca sull'*efficacy* e sull'*effectiveness* degli interventi psicologici usano una metrica comune chiamata *dimensione dell'effetto*. La dimensione dell'effetto può essere calcolata per quasi tutti i tipi di disegni della ricerca e nel caso di quasi tutte le analisi statistiche. Le analisi statistiche basate sui confronti tra gruppi danno luogo a due tipi di indici che misurano la dimensione dell'effetto. Il primo tipo di dimensione dell'effetto riguarda le differenze tra le medie dei gruppi e si calcola nei termini della differenza tra le medie dei gruppi (per esempio gruppi con il trattamento e senza trattamento) divisa per la stima della deviazione standard raggruppata. Tali indici vanno sotto il nome di d , g o η^2 .

Il secondo tipo di dimensione dell'effetto comporta un confronto di gruppi in termini degli odds o della probabilità di un risultato. Un *odds ratio* (OR) viene calcolato per determinare l'associazione tra la condizione di gruppo (ad es., trattamento e non trattamento) e una variabile ad esito binario (ad esempio, la presenza o la non insorgenza di un evento, come la ricaduta). Un *rischio relativo* (RR) viene calcolato per confrontare la probabilità che un evento si verifichi in base ai due livelli del fattore di rischio considerato (ad es., trattamento/non trattamento, esposizione/non esposizione ad un fattore di rischio).

Si noti inoltre un punto importante: non è possibile confrontare direttamente la dimensione dell'effetto ottenuta in studi sull'*efficacy* e sull'*effectiveness* del trattamento perché, quasi sempre, tali studi vengono svolti mediante disegni sperimentali molto diversi. Gli studi sull'*efficacy* sono generalmente studi randomizzati controllati nei quali i risultati del trattamento vengono confrontati con i risultati di una condizione di controllo (ad esempio, nessun trattamento o una forma alternativa di trattamento). In questo caso, la dimensione dell'effetto si basa sul cambiamento che può essere attribuito all'effetto causale del trattamento. Pochi studi sull'*effectiveness* del trattamento sono invece studi

randomizzati controllati: tali studi consistono tipicamente in un'indagine entro i gruppi – cioè, in un confronto tra la condizione pre-trattamento e la condizione post-trattamento, senza alcuna condizione di controllo. Di conseguenza, la dimensione dell'effetto ottenuta in questo tipo di indagini è basata su un cambiamento dovuto a cause molteplici: oltre agli effetti del trattamento, ci sono agli effetti dovuti alla maturazione, alla regressione verso la media, alla remissione spontanea dei sintomi dovuta al passaggio di tempo e la reattività delle misure. Pertanto, è probabile che gli studi sull'*efficacy* portino ad una stima della dimensione dell'effetto maggiore rispetto ai valori che tipicamente vengono ottenuti negli studi sull'*effectiveness* del trattamento.

D.1 Risposta continua

D.1.1 Indici d di Cohen e g di Hedges

Per valutare la dimensione dell'effetto, Cohen (1962, 1988) ha introdotto una misura simile a un punteggio z in cui una di due medie campionarie viene sottratta dall'altra e il risultato è diviso per la deviazione standard della popolazione:

$$d = \frac{M_A - M_B}{\sigma}, \quad (\text{D.1})$$

laddove M_A e M_B sono le due medie campionarie e σ è la media della popolazione.

Hedges (1982) ha proposto una piccola modifica alla @ref(eq:d_cohen) nella quale la deviazione standard raggruppata sostituisce il parametro ignoto σ , ottenendo in questo modo la statistica g :

$$g = \frac{M_A - M_B}{s}. \quad (\text{D.2})$$

Al fine di evitare una sistematica sovrastima della dimensione dell'effetto in piccoli campioni, Borenstein, Hedges, Higgins e Rothstein (2009) hanno proposto la seguente correzione:

$$d_{unb} = d \left(1 - \frac{3}{4 \cdot \text{df} - 1} \right). \quad (\text{D.3})$$

La correzione è molto piccola quando l'ampiezza campionaria è grande (solo il 3% per 25 gradi di libertà) ma è sostanziale per piccoli campioni.

D.2 Risposta binaria

Supponiamo che la variabile risposta Y abbia due modalità, convenzionalmente chiamate successo (1) e insuccesso (0). Per esempio, un individuo con certi fattori di rischio può ammalarsi oppure no. Supponiamo che le osservazioni siano indipendenti e che $P(Y = 1) = \pi$ e $P(Y = 0) = 1 - \pi$. Spesso, ogni unità di osservazione è associata a un vettore (X_1, \dots, X_p) di variabili esplicative. L'obiettivo è studiare la relazione tra la probabilità $P(Y = 1)$ e le variabili esplicative. Il caso più semplice si ha quando esiste un solo carattere esplicativo X , con I modalità sconnesse (un fattore). In tal caso i dati binari sono raggruppati per modalità della variabile esplicativa e le risposte Y_1, \dots, Y_I sono il *numero di successi* nelle n_i prove indipendenti dell' i -esimo gruppo ($i = 1, \dots, I$). Se i dati sono raggruppati si possono presentare come una tavola di contingenza $I \times 2$:

	Successo	Insuccesso	Totale
1	Y_{11}	Y_{12}	Y_{1+}
2	Y_{21}	Y_{22}	Y_{2+}
\vdots	\vdots	\vdots	\vdots
I	Y_{I1}	Y_{I2}	Y_{I+}

Nelle applicazioni che considereremo i totali di riga non sono variabili aleatorie, ma costanti fisse per disegno. Pertanto

X	Successo	Insuccesso	Totale
x_1	Y_1	$n_1 - Y_1$	n_1
x_2	Y_2	$n_2 - Y_2$	n_2
\vdots	\vdots	\vdots	\vdots
x_I	Y_I	$n_I - Y_I$	n_I

Le situazioni in cui i totali di riga sono v.a. non saranno qui considerate.

Il caso fondamentale è quello della tavola 2×2 in cui vi è una sola variabile esplicativa binaria. Supponiamo che X sia il trattamento (1: presente; 0: controllo) e Y sia la risposta (1: successo; 0: insuccesso). La situazione è riassunta nella tavola seguente:

Trattamento	Risposta	
	Successo	Insuccesso
Presente	π_1	$1 - \pi_1$
Controllo	π_2	$1 - \pi_2$

dove π_1 e π_2 sono le probabilità condizionate $P(Y = \text{successo} \mid X = \text{trattamento presente})$ e $P(Y = \text{successo} \mid X = \text{controllo})$. Le proporzioni di successi nel campione forniscono delle stime di π_1 e π_2 :

$$\begin{aligned}\hat{\pi}_1 &= Y_1/n_1, \\ \hat{\pi}_2 &= Y_2/n_2,\end{aligned}$$

dove Y_i è il numero ottenuto di successi ed n_i è la numerosità delle prove per $X = x_i$. Per studiare l'*efficacy* del trattamento si calcolano i seguenti indici:

- la *differenza delle probabilità* $D = \pi_1 - \pi_2$;
- il *rapporto delle probabilità* $RR = \pi_1/\pi_2$, il cosiddetto *rischio relativo*;
- il *rapporto delle quote* (odds-ratio) detto anche *rapporto crociato* $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$;
- il *logaritmo del rapporto delle quote* $\log_e \theta$.

Esercizio D.1. Il Physicians' Health Study è stata un'indagine prospettiva svolta per verificare se l'uso regolare di Aspirina riduce la mortalità per malattie cardiovascolari. I partecipanti allo studio (dei medici volontari) venivano assegnati in modo casuale al trattamento (uso regolare di Aspirina, $n_1 = 11037$) o al placebo ($n_2 = 11034$). I soggetti non erano a conoscenza del tipo di trattamento cui erano stati assegnati.

```
aspirina <- matrix(
  c(104, 10933, 189, 10845),
  nrow = 2,
  byrow = TRUE
```

```

)
dimnames(aspirina) <- list(
  c("Aspirina", "Placebo"),
  c("Si", "No")
)
names(dimnames(aspirina)) <-
  c("Gruppo", "Infarto miocardico")

aspirina
#>           Infarto miocardico
#> Gruppo      Si      No
#> Aspirina 104 10933
#> Placebo  189 10845

```

L'evento chiamato convenzionalmente successo è la presenza di un infarto.

Calcolare la grandezza totale del campione e le proporzioni $\hat{\pi}_{ij}$ è facile:

```

tot <- sum(aspirina)
tot
#> [1] 22071
aspirina / tot
#>           Infarto miocardico
#> Gruppo      Si      No
#> Aspirina 0.004712 0.4954
#> Placebo  0.008563 0.4914

```

Le stime delle probabilità condizionate di successo dato X (proporzioni di riga) si ottengono nel modo seguente:

```

rowtot <- apply(aspirina, 1, sum)
rowtot
#> Aspirina Placebo
#>    11037    11034

```

```

rowprop <- sweep(aspirina, 1, rowtot, "/")
rowprop

```

```
#>          Infarto miocardico
#> Gruppo          Si      No
#> Aspirina 0.009423 0.9906
#> Placebo  0.017129 0.9829
```

quindi $\hat{\pi}_1 = 0.0094$ e $\hat{\pi}_2 = 0.0171$.

D.2.1 Differenza tra due proporzioni

Il modo più semplice per misurare l'effetto del trattamento sulla variabile risposta è la differenza delle probabilità

$$D = \pi_1 - \pi_2.$$

La differenza di probabilità assume valori compresi tra -1 e 1 ed è nulla se la risposta non dipende da X , cioè se il trattamento non ha effetto.

Esercizio D.2. Per l'esempio dell'aspirina, $\hat{\pi}_1 = 104/11037 = 0.0094$ è la proporzione di attacchi di cuore (successi) tra gli individui trattati con aspirina e $\hat{\pi}_2 = 189/11034 = 0.0171$ è la proporzione di attacchi di cuore tra gli individui trattati con il placebo. La differenza tra le proporzioni è $\hat{\pi}_1 - \hat{\pi}_2 = 0.0094 - 0.0171 = -0.0077$. L'errore standard stimato della differenza tra le proporzioni è

$$\sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}} = 0.0015.$$

Un intervallo di confidenza al 95% per la differenza tra le proporzioni è $-0.0077 \pm 1.96(0.0015)$ cioè $(-0.011, -0.005)$. Poiché l'intervallo contiene solo valori negativi si conclude che $\pi_1 < \pi_2$, per cui il trattamento con aspirina appare ridurre la probabilità di infarto.

Il risultato precedente si ottiene con nel modo seguente:

```
p1 <- 104 / 11037
p2 <- 189 / 11034
n1 <- 11037
n2 <- 11034
se <- sqrt((p1 * (1 - p1) / n1) + (p2 * (1 - p2) / n2))
```



```
se
#> [1] 0.00154
p1 - p2 + c(-1, 1) * qnorm(0.975) * se
#> [1] -0.010724 -0.004688
```

D.2.2 Rischio relativo

La differenza tra due proporzioni può rivestire un'importanza maggiore quando entrambe le proporzioni sono vicine a 0 o a 1 che quando esse sono vicine a 0.5. Per esempio, in uno studio che confronta la mortalità associata a due trattamenti, la differenza $0.010 - 0.001 = 0.009$ può essere più importante della differenza tra $0.410 - 0.401 = 0.009$. Infatti, la prima differenza coinvolge due proporzioni che stanno in rapporto di 10 a 1, mentre la seconda riguarda due proporzioni quasi uguali. Quindi la prima differenza appare di maggiore rilevanza. Nelle situazioni precedenti è utile calcolare il rischio relativo, cioè il rapporto

$$RR = \frac{\pi_1}{\pi_2}.$$

Il RR è sempre un numero non negativo. Se è uguale a 1 la risposta non dipende da X .

Esercizio D.3. Per le proporzioni precedenti, il rischio relativo è $0.010/0.001 = 10.0$ e $0.410/0.401 = 1.02$. Nell'esempio dell'aspirina il RR campionario è

$$0.0171/0.0094 = 1.82.$$

Quindi la proporzione campionaria di casi di infarto è più grande dell'82% per il gruppo dei trattati col placebo.

D.2.3 Odds ratio

Invece del RR si può misurare la dipendenza delle probabilità dal trattamento con il rapporto tra le *quote di scommessa* (gli odds). Gli odds di successo ω sono per definizione il rapporto tra la probabilità di successo (π) e la probabilità di insuccesso ($1 - \pi$):

$$\omega = \frac{\pi}{1 - \pi}.$$

La quota di scommessa ω è un indice non negativo che misura quanti successi ci si attendono per ogni insuccesso. Nel campo delle scommesse quando la quota dell'evento $Y = 1$ contro $Y = 0$ è ω , scommettendo su $Y = 0$ si vince ω volte la posta. Per esempio, se $\pi = 0.75$ la quota è $\omega = 0.75/0.25 = 3$; scommettendo su $Y = 0$ si riceve, in caso di vincita, 3 volte la posta. Se la quota è minore di 1, si può calcolare il reciproco ($1/\omega$) e interpretare il risultato come riferito all'evento complementare. Per esempio, se ω è 0.3333, uno si aspetta 0.3333 successi per ogni insuccesso ossia $1/0.3333 = 3$ insuccessi per ogni successo. La relazione inversa tra probabilità di successo e odds è $\pi = \frac{\omega}{1+\omega}$. Per esempio, se $\omega = 3$, la probabilità di successo sarà $\pi = \frac{3}{1+3}$ uguale a 0.75.

Esercizio D.4. Nell'esempio dell'aspirina l'odds di infarto per il gruppo aspirina è

$$\omega_1 = 104/10933 = 0.0095$$

e per il gruppo placebo è

$$\omega_2 = 189/10845 = 0.0174.$$

Dato l'odds $\omega_1 = 0.0095$, la probabilità di infarto per il gruppo aspirina è

$$\pi_1 = \frac{\omega_1}{1 + \omega_1} = \frac{0.0095}{1 + 0.0095} = 0.0094$$

ovvero

$$\pi_1 = \frac{104}{104 + 10933} = 0.0094.$$

D.2.4 Logit

Il logaritmo della quota (ϕ) o *logit* definito da

$$\phi = \log_e \omega = \log_e \frac{\pi}{1 - \pi} = \ln(\pi) - \ln(1 - \pi)$$

trasforma la probabilità $\pi \in (0, 1)$ in un numero $\phi \in \mathfrak{R}$. Il logit è simmetrico attorno allo 0 ed è privo di limite superiore e inferiore.

Per esempio, se $\pi = 0.75$ la quota è $\omega = 0.75/0.25 = 3$. Il logaritmo della quota (logit) è $\log_e 3 = 1.0986$. Vediamo di seguito alcuni valori rappresentativi che illustrano la relazione tra π e ϕ :

Probabilità	Odds	logit
π	$\omega = \frac{\pi}{1-\pi}$	$\phi = \ln \frac{\pi}{1-\pi}$
0.01	1/99=0.0101	-4.60
0.05	5/95=0.0526	-2.94
0.10	1/9=0.1111	-2.20
0.30	3/7=0.4286	-0.85
0.50	5/5=1	0.00
0.70	7/3 = 2.333	0.85
0.90	9/1 = 9	2.20
0.95	95/5 = 19	2.94
0.99	99/1 = 99	4.60

La trasformazione inversa del logit è

$$\pi = \frac{e^\phi}{1 + e^\phi},$$

dove $e \simeq 2.718$.

A un logit $\phi = 1.0986$, per esempio, corrisponde una probabilità

$$pi = \frac{e^\phi}{1 + e^\phi} = \frac{e^{1.0986}}{1 + e^{1.0986}} = 0.75$$

uguale a 0.75.

Si noti che il logit della probabilità complementare $1 - \pi = 0.25$ è -1.0986, cioè l'opposto. Se $\pi = 0.25$ la quota è $\omega = 0.25/0.75 = 0.3333$. Il logit di $\pi = 0.25$ è

$$\log_e(0.25/0.75) = \log_e 0.3333 = -1.0986.$$

La distribuzione campionaria della proporzione $\hat{\pi}$ è esattamente una Binomiale. Asintoticamente, è normale con media π e varianza asintotica stimata $\hat{\pi}(1-\hat{\pi})/n$. Lo stimatore del logit ha una distribuzione asintotica normale con media $\pi/(1-\pi)$ e varianza asintotica stimata $1/(n\hat{\pi}(1-\hat{\pi}))$.

Esercizio D.5. Per il campione di 11037 volontari sottoposti al trattamento aspirina la probabilità stimata di infarto miocardico è

$$\hat{\pi} = 104/11037 = 0.0094$$

Il logit empirico è

$$\ln\left(\frac{0.0094}{1 - 0.0094}\right) = -4.655$$

con errore standard asintotico

$$\sqrt{\frac{1}{11037 \times 0.0094 \times (1 - 0.0094)}} = 0.0985.$$

D.2.5 Rapporto delle quote

Una volta chiarito il concetto di odds, consideriamo ora il *rapporto delle quote* (*odds-ratio*) detto anche *rapporto crociato* (*cross-product ratio*). In una tabella 2×2 gli odds di successo nella riga i -esima sono $\omega_i = \pi_i/(1 - \pi_i)$. Il rapporto degli odds ω_1 e ω_2 nelle due righe

$$\theta = \frac{\omega_1}{\omega_2} = \frac{\pi_1/(1 - \pi_1)}{\pi_2/(1 - \pi_2)}$$

è chiamato rapporto delle quote.

Un rapporto delle quote può assumere solo valori non negativi. Se è uguale a 1 gli odds sono uguali e quindi sono uguali anche le probabilità, cioè la risposta è indipendente dal trattamento. I rapporti degli odds si valutano in rapporto a 1:

- se $1 < \theta < \infty$ gli odds sono più grandi nel gruppo 1 che nel gruppo 2, e quindi anche $\pi_1 > \pi_2$;
- se $0 < \theta < 1$ gli odds sono più piccoli nel gruppo 1 che nel gruppo 2 e $\pi_1 < \pi_2$.

Il rapporto degli odds non cambia se si permutano la variabile risposta e la variabile esplicativa; quindi θ tratta le variabili in modo simmetrico.

Esercizio D.6. Nell'esempio dell'aspirina il rapporto degli odds stimato è

$$\hat{\theta} = \frac{189/10845}{104/10933} = 1.832,$$

cioè gli odds a favore dell'infarto sono più grandi dell'83% per il gruppo placebo. Un rapporto degli odds di 1.832 non significa che π_2 è 1.832 volte π_1 , ma che gli odds $\pi_2/(1 - \pi_2)$ sono 1.832 volte gli odds $\pi_1/(1 - \pi_1)$. Tuttavia,

$$\theta = RR \frac{1 - \pi_1}{1 - \pi_2}.$$

Perciò quando la proporzione di successi è prossima a zero in entrambi i gruppi θ e il RR hanno valori simili. Si osservi infatti che il RR nell'esempio dell'aspirina è 1.83.

D.2.6 Logaritmo del rapporto delle quote

La distribuzione del rapporto degli odds è molto asimmetrica ed è conveniente usare la distribuzione del suo logaritmo $\log_e \theta$. Tale distribuzione è meno asimmetrica e più vicina alla normalità. Il logaritmo di θ è 0 in caso di indipendenza e l'interpretazione è simmetrica rispetto allo zero. Cioè se si invertono le righe o le colonne della tavola $\log_e \theta$ cambia il segno. Due valori di $\log_e \theta$ diversi solo per il segno rappresentano due livelli di associazione uguali. Raddoppiando $\log_e \theta$ corrisponde ad elevare al quadrato il rapporto delle quote.

L'errore standard asintotico del logaritmo del rapporto degli odds ha una formula semplice

$$\sqrt{\frac{1}{Y_{11}} + \frac{1}{Y_{12}} + \frac{1}{Y_{21}} + \frac{1}{Y_{22}}},$$

dove Y_{ij} sono le frequenze nelle celle della tavola di contingenza.

Un intervallo di confidenza al 95% per $\ln \theta$ è dato dalla stima $\ln \hat{\theta} \pm 1.96$ l'errore standard stimato. Per ottenere l'intervallo di confidenza al 95% per l'odds ratio esponenziamo i limiti L dell'intervallo di confidenza: e^L .

Esercizio D.7. Nell'esempio dell'aspirina il rapporto degli odds è

$$\hat{\theta} = \frac{189/10845}{104/10933} = 1.832$$

e il logaritmo del rapporto degli odds è

$$\log_e 1.832 = 0.605.$$

Il logaritmo del rapporto delle quote ha la seguente proprietà: se invertiamo l'ordine delle categorie di una delle variabili, il logaritmo del rapporto delle quote cambia semplicemente di segno:

$$\hat{\theta} = \frac{104/10933}{189/10845} = 0.5458$$

e quindi

$$\log_e 0.5458 = -0.605.$$

L'errore standard stimato del logaritmo dell'odds ratio è

$$\sqrt{\frac{1}{189} + \frac{1}{10845} + \frac{1}{104} + \frac{1}{10933}} = 0.1228.$$

Un intervallo di confidenza al 95% per $\log_e \theta$ è

```
se <- 0.1228
0.6054377 + c(-1, 1) * 1.96 * se
#> [1] 0.3647 0.8461
```

e il corrispondente intervallo di confidenza per θ è

```
exp(0.8462073)
#> [1] 2.331
exp(0.3646681)
#> [1] 1.44
```

Poiché non contiene il valore 1, i veri valori degli odds per il gruppo placebo e per il gruppo aspirina sono significativamente diversi: gli odds per l'infarto sono almeno il 44% in più rispetto al gruppo aspirina.

D.3 Tipi fondamentali di indagine

Si distinguono due tipi fondamentali di indagine: gli esperimenti e gli studi osservazionali. Negli esperimenti si studia l'effetto di uno o più trattamenti sulle risposte delle unità sperimentali. Negli esperimenti randomizzati ogni unità sperimentale viene assegnata casualmente dal ricercatore a una delle possibili modalità di trattamento. L'importanza della randomizzazione consiste nel produrre dei sottogruppi a seconda dei livelli del trattamento, in cui tutte le altre variabili, anche quelle non misurate, hanno approssimativamente la stessa distribuzione e quindi sono comparabili.

Negli studi osservazionali invece il ricercatore non può assegnare le unità ai trattamenti. Le indagini osservazionali si possono distinguere in studi longitudinali prospettici, trasversali (cross-sectional) e caso-controllo.

D.3.1 Studi prospettici

In un studio prospettico viene seguito un numero fisso di unità per ciascuna modalità della variabile esplicativa ($X = x_1, \dots, X = x_I$) e dopo un periodo prefissato si rilevano le proporzioni di successi negli I gruppi.

Esercizio D.8. Il Physicians' Health Study è stata un'indagine prospettica svolta per verificare se l'uso regolare di Aspirina riduce la mortalità per malattie cardiovascolari. I partecipanti allo studio (dei medici volontari) venivano assegnati in modo casuale al trattamento (uso regolare di Aspirina, $n_1 = 11037$) o al placebo ($n_2 = 11034$). Per cinque anni, i medici che parteciparono allo studio assunsero giornalmente una pastiglia di aspirina o un placebo. Alla fine dello studio l'incidenza di infarti miocardici venne misurata nei due gruppi.

D.3.2 Studi trasversali

Se si estrae un campione ad un tempo prefissato e si classificano le unità nella tabella $I \times 2$ a seconda delle modalità dei due caratteri si ha uno studio trasversale.

Esercizio D.9. Sulla base del General Social Survey (1984), 901 individui sono stati classificati in base alla soddisfazione lavorativa (soddi-

sfatto verso insoddisfatto) e a due categorie di reddito ($< \$15.000, \geq \15.000).

	Soddisfatto	Insoddisfatto
$< \$15.000$	391	104
$\geq \$15.000$	340	66

Ci si chiede: la soddisfazione lavorativa dipende dal reddito?

D.3.3 Studi longitudinali retrospettivi

Se si estraggono due campioni per $Y = 0$ e $Y = 1$ e si controlla nel passato se i soggetti appartengono al gruppo x_1, x_2, \dots, x_I allora si ha un studio retrospettivo. Un esempio tipico di indagine basata su un disegno retrospettivo è lo studio caso-controllo. Si considerino i dati seguenti:

	Cancro ai polmoni	
Mai fumato?	Casi	Controlli
Sì	688	650
No	21	59
Totale	709	709

La prima colonna si riferisce a 709 pazienti ricoverati in 20 ospedali londinesi per cancro polmonare. Ogni caso è stato appaiato a un controllo, cioè a un paziente ricoverato nello stesso ospedale per disturbi diversi dal cancro polmonare. Casi e controlli sono stati classificati poi a seconda che siano o siano stati fumatori oppure no.

Sarebbe naturale considerare il cancro polmonare come variabile risposta e il comportamento relativo al fumo quale variabile esplicativa, e confrontare le proporzioni di cancro polmonare tra fumatori e non fumatori, $P(\text{cancro} \mid \text{fumo})$. In questo studio, tuttavia, questo non ha senso dato che la distribuzione marginale della variabile risposta è fissa per disegno.

- Nello studio Aspirina erano i totali di riga a essere fissi – avevamo cioè un campione indipendente per ciascuna modalità della variabile esplicativa.

- Nel caso presente, invece, sono i totali di colonna ad essere fissi – ovvero, abbiamo un campione indipendente per ciascuna modalità della variabile risposta.

Invece è sensato usare le proporzioni nell'altro verso, cioè $688/709 = 0.970$ e $650/709 = 0.917$ come stime delle probabilità $P(X = \text{fumatore} \mid Y = \text{caso})$ e $P(X = \text{fumatore} \mid Y = \text{controllo})$. Se conoscessimo la proporzione di individui nella popolazione con cancro polmonare, potremmo stimare $P(Y = \text{cancro polmonare} \mid X = \text{fumatore})$ e $P(Y = \text{non cancro polmonare} \mid X = \text{fumatore})$. Ponendo $C = \text{cancro}$ e $F = \text{fumatore}$, con Bayes avremo

$$P(C \mid F) = \frac{P(C)P(F \mid C)}{P(C)P(F \mid C) + P(C^c)P(F \mid C^c)}.$$

Non possiamo però procedere in questo modo, dato che la probabilità $P(C)$ è ignota. Con questo disegno, inoltre, dalle probabilità $P(X = \text{fumatore} \mid Y = \text{cancro polmonare})$ e $P(X = \text{fumatore} \mid Y = \text{non cancro polmonare})$ non è neppure possibile calcolare il rischio relativo per la risposta Y .

Possiamo però stimare il rapporto degli odds

$$\frac{(688/709)/(21/709)}{(650/709)/(59/709)} = \frac{688 \times 59}{650 \times 21} = 2.97$$

e utilizzarlo per l'interpretazione di interesse (anche se lo studio è retrospettivo): gli odds per il cancro polmonare sono circa 3 volte più grandi tra i fumatori.

Se le probabilità condizionate di $Y = \text{cancro polmonare}$ dato $X = \text{fumatore}$ e $X = \text{non fumatore}$ sono vicine a zero, il RR ha un valore simile a quello di θ . Nell'esempio possiamo attenderci che tali probabilità siano piccole e quindi possiamo considerare il rapporto degli odds come un indicatore grezzo del RR . Possiamo perciò concludere dicendo che la frequenza relativa del cancro polmonare è circa 3 volte più grande per gli individui che hanno fumato rispetto a quelli che non hanno mai fumato.

Un intervallo di confidenza si calcola come indicato in precedenza.

```
se <- sqrt(1 / 688 + 1 / 650 + 1 / 21 + 1 / 59)
se
#> [1] 0.2599
```

```
theta <- (688 * 59) / (650 * 21)
theta
#> [1] 2.974
log(theta)
#> [1] 1.09
log(theta) + c(-1, 1) * 1.96 * se
#> [1] 0.5804 1.5993
exp(0.5803817)
#> [1] 1.787
exp(1.5992813)
#> [1] 4.949
```

Poiché non contiene il valore 1, i veri valori degli odds per il gruppo fumatori e per il gruppo non fumatori sono significativamente diversi: gli odds per il cancro ai polmoni sono almeno il 78% in più rispetto al gruppo non fumatori.

Considerazioni conclusive

Nel caso di una variabile risposta binaria e una variabile esplicativa con due sole modalità i dati si possono rappresentare in una tavola di contingenza 2×2 . Negli studi sperimentali l'efficacia del trattamento si può stabilire calcolando la differenza tra le proporzioni, il rapporto tra le proporzioni, il rapporto tra le quote e il logaritmo del rapporto tra le quote. I dati raccolti mediante un esperimento e mediante un'indagine osservazionale prospettico o trasversale hanno la stessa struttura. Le differenze nella risposta possono dunque essere analizzate mediante gli stessi indici descritti in precedenza. Negli studi osservazionali, però, le conclusioni sono molto meno stringenti. Nell'esperimento i gruppi sono in tutto confrontabili tranne per la modalità del trattamento e quindi eventuali differenze nella risposta non possono essere dovute ad altro che al trattamento. Negli studi osservazionali, invece, la mancanza di controllo sull'assegnazione dei trattamenti fa sì che i gruppi di unità non siano mai totalmente comparabili. Negli studi retrospettivi, come ad esempio i disegni caso-controllo, non ha senso calcolare la differenza tra le proporzioni condizionate alla variabile esplicativa e il rischio relativo. Possiamo

però calcolare il rapporto tra le quote e, per la proprietà simmetrica di questo indice (θ non cambia se vengono invertite la variabile sulle righe e quella sulle colonne), procedere poi all'interpretazione nella direzione di interesse. Dato che il logaritmo dell'odds ratio ha una distribuzione asintotica normale, un intervallo di confidenza per $\ln \theta$ può essere facilmente calcolato. Esponenziando i limiti dell'intervallo di confidenza per $\ln \theta$ si trovano i limiti dell'intervallo di confidenza per θ .



Bibliografia

- Allen, M. J. and Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5):815–824.
- Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Charter, R. A. (1996). Revisiting the standard errors of measurement, estimate, and prediction and their application to test scores. *Perceptual and Motor Skills*, 82(3):1139–1144.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334.
- Dudek, F. J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin*, 86(2):335—337.
- Gulliksen, H. (1961). Measurement of learning and mental abilities. Technical report, Princeton University.
- Hu, L.-t. and Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4):424—453.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2):183–202.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4):443–477.

- Kan, K.-J., van der Maas, H. L., and Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence*, 73:52–62.
- Kline, P. (2013). *Handbook of psychological testing*. Routledge.
- Kuder, G. F. and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160.
- Lewis, C. (1986). Test theory and psychometrika: The past twenty-five years. *Psychometrika*, 51(1):11–22.
- Lord, F. M. and Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Nunnally, J. C. (1994). *Psychometric theory*. McGraw-Hill.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Samejima, F. (1983). Constant information model on the dichotomous response level. In *New horizons in testing*, pages 287–308. Elsevier.
- Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15:201–293.
- Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of the vectors of mind.
- Wechsler, D. (2008). Wechsler adult intelligence scale—fourth edition (wais-iv). *San Antonio, TX: NCS Pearson*, 22(498):816–827.
- Xia, Y. and Yang, Y. (2019). Rmsea, cfi, and tli in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1):409–428.