

Testing psicologico

Statistica applicata alla misurazione in psicologia
con laboratorio in R

Ottobre 2019
(ver. 1.19.10)

bozza non corretta

Note e appunti per un corso introduttivo di teoria e tecniche dei test
con applicazioni ed esercitazioni in R

Antonio Calcagni
DPSS - Università degli studi di Padova

Suggerimenti, commenti, segnalazione di errori sono graditi
e possono essere inviati a: antonio.calcagni@unipd.it

bozza non corretta

Copyright © 2019 Antonio Calcagni. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation. A copy of the license (<https://www.gnu.org/licenses/fdl-1.3.html>) is available at the end of this book.

Indice

1	Introduzione alla misurazione	3
1.1	Lessico metrologico	3
1.1.1	Analisi degli errori	4
1.1.2	Proprietà del processo misurativo	6
1.2	Stimare il misurando, l'accuratezza e la precisione	8
1.2.1	Stima del misurando	8
1.2.2	Stima della precisione	10
1.2.3	Stima dell'accuratezza	11
1.3	Analisi dell'incertezza	13
1.3.1	Valutazione dell'incertezza tipo A e tipo B	14
1.4	Misurazione indiretta e propagazione dell'incertezza	15
1.5	Note di sintesi	16
1.6	Glossario minimale	16
1.7	Laboratorio in R	17
2	Richiami di statistica per la misurazione	19
2.1	Esperimenti aleatori	20
2.2	Assegnazioni di $\mathbb{P}()$	20
2.3	Variabili aleatorie	21
2.3.1	Esempi di costruzione di variabili aleatorie discrete	22
2.3.2	Funzioni di densità discrete e continue, funzione di ripartizione	24
2.3.3	Valori medi di variabili aleatorie	27
2.3.4	Variabili aleatorie standardizzate	30
2.3.5	Disuguaglianze notevoli (approfondimento)	30
2.3.6	Modelli probabilistici	31
2.3.7	Due teoremi limite fondamentali	32
2.4	Note di sintesi	34
2.5	Laboratorio in R	36
3	La misurazione in psicologia	37
3.1	Livelli di misurazione	37
3.2	Lessico psicometrico	38
3.3	Modelli per la misurazione in psicologia	39

4	La teoria classica dei test	41
4.1	Definizioni e proprietà iniziali	42
4.2	Misure parallele e non parallele	43
4.3	Determinare $\text{Var}[T]$ e $\text{Var}[E]$	45
4.4	Validità di una misura	46
4.5	Dalla validità di una misura all'attendibilità	47
4.6	Misure composite	48
4.7	Attendibilità di misure composite	49
4.7.1	Attendibilità e lunghezza del test	50
4.8	Stime dell'attendibilità	53
4.8.1	α di Cronbach	53
4.8.2	Altre tecniche per la valutazione dell'attendibilità	54
4.9	Fattori che influenzano l'attendibilità	55
4.10	Stima dei punteggi veri	55
4.11	Esempio di applicazione della TCT	56
4.12	Selezione delle misure osservabili	57
4.12.1	Misure osservabili dicotomiche	58
4.12.2	Misure osservabili politomiche	61
4.13	Punteggi totali e standardizzati	63
4.14	Note di sintesi	64
4.15	Laboratorio in R	65
5	Analisi fattoriale confermativa	67
5.1	Definizione del modello	68
5.1.1	Modello ad una sola variabile latente	68
5.1.2	Modello con più variabili latenti	70
5.1.3	Assunzioni del modello	73
5.1.4	Matrice di covarianza implicata dal modello	74
5.1.5	CFA in sintesi	77
5.2	Stima dei parametri del modello	77
5.3	Valutazione del modello CFA	79
5.3.1	Test del χ^2	79
5.3.2	RMSEA	80
5.3.3	Altri indici	80
5.3.4	Indici di modifica	80
5.3.5	Analisi dei residui e correlazione tra gli errori	81
5.4	Studio della dimensionalità della WISC-IV	82
5.5	Note di sintesi	89
5.6	Alcune Risorse utili	89
5.7	Laboratorio in R	90
6	Invarianza di misurazione	91
6.1	Schema generale di analisi dell'invarianza	92
6.1.1	Invarianza configurale	93
6.1.2	Invarianza debole	95
6.1.3	Invarianza forte	97

6.1.4	Invarianza esatta	99
6.2	Note di sintesi	101
6.3	Alcune Risorse utili	101
6.4	Laboratorio in R	101
A	Elementi essenziali di Algebra Lineare	103
A.1	Matrici, vettori, scalari	103
A.2	Alcune operazioni di algebra lineare	104
A.3	Note finali	107
B	Elementi essenziali del linguaggio R	109
B.1	Scrittura di espressioni	109
B.2	Vettori	110
B.2.1	Creazione di vettori	110
B.2.2	Selezione elementi di un vettore	111
B.2.3	Funzioni ed operazioni tra vettori	112
B.3	Matrici	112
B.3.1	Creazione di matrici	112
B.3.2	Selezione di elementi di una matrice	113
B.3.3	Funzioni ed operazioni tra matrici	113
B.4	DataFrames	114
B.4.1	Creazione di DataFrames	114
B.4.2	Selezione di elementi di un DataFrame	115
B.4.3	Funzioni con DataFrames	116
B.5	Liste	117
B.5.1	Creazione di Liste	117
B.5.2	Selezione di elementi di una lista	118
B.6	Tipi di vettori	118
B.6.1	Vettori numerici	118
B.6.2	Vettori logici	118
B.6.3	Vettori di caratteri	119
B.6.4	Fattori	119
C	Licenza GNU GPL	121
	Bibliografia	130

bozza non corretta

Prefazione

Questo documento contiene note e appunti per un corso introduttivo di teoria e tecniche dei test (testing psicologico). L'esposizione degli argomenti privilegia un approccio statistico con il tentativo, quando possibile, di connettere taluni argomenti (ad esempio, la teoria classica dei test) ai principali risultati ottenuti nel campo della metrologia. La trattazione di alcuni argomenti ha carattere meramente introduttivo (ad esempio, teoria della misurazione e degli errori) mentre altri argomenti fondamentali (ad esempio, la teoria della risposta all'item - IRT) non sono presenti in questa versione. La dispensa è stata realizzata utilizzando le lezioni del corso di "Testing Psicologico" tenuto a partire dall'a.a. 2018-2019 presso la Scuola di Psicologia dell'Università di Padova.

L'esposizione dei contenuti segue una struttura modulare progressiva e consta di quattro parti fondamentali: introduzione ai concetti fondamentali della misurazione (Capitolo 1), elementi di statistica inferenziale alla base dei principali risultati della misurazione (Capitolo 2), misurazione psicologica secondo la teoria classica dei test (Capitolo 3), modello statistico dell'analisi fattoriale confermativa per la misurazione psicologica (Capitolo 4). Ogni modulo è organizzato e pensato per essere autonomo sebbene la loro organizzazione proceda secondo uno schema progressivo. Ogni argomento è corredato da applicazioni ed esercitazioni mediante l'ambiente statistico R: una trattazione dei temi riguardanti la misurazione psicologica non può difatti procedere senza offrire strumenti concreti per affrontare le domande che il mondo della ricerca e quello lavorativo pongono agli studiosi di tali discipline. Sebbene non esaustivo rispetto all'ampio panorama offerto dai problemi che sorgono all'interno del testing psicologico e, più in generale, della psicometria e della statistica applicata, il laboratorio di R offerto all'interno di questo corso vuole raggiungere obiettivi minimali, fornendo agli studenti l'acquisizione di un linguaggio essenziale per la risoluzione di molti quesiti e problemi di carattere applicativo.

La presente dispensa è stata realizzata in sinergia con il lavoro svolto dagli studenti del corso dell'a.a. 2019-2020 che hanno provveduto a sistematizzare pazientemente appunti e lucidi delle lezioni. Il presente lavoro è stato realizzato in \LaTeX mentre le figure presenti nei capitoli unitamente agli esempi sono sviluppati in R. I codici utilizzati per gli esempi, le figure e i laboratori di ciascun capitolo sono disponibili alla pagina: <https://github.com/antcalcagni/psychological-testing>.

Padova, Ottobre 2019

bozza non corretta

1

Introduzione alla misurazione

«When you can measure what you are speaking about and express it in a number, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of science»

– Lord Kevin, 3 May 1883

1.1 Lessico metrologico

La **metrologia** è la scienza che si occupa della misurazione da una prospettiva scientifica, includendo la parte teorica (modelli matematici della misurazione), sperimentale (modi per la misurazione) e tecnica (strumenti per misurare). Si fonda sulla statistica e i suoi principi sono condivisi a livello internazionale. Esistono infatti diversi enti di riferimento per la misurazione, tra cui il *Bureau International des Poids et Mesures* (BIPM). Il BIPM è un ente intergovernativo dove le conoscenze sono condivise in un panorama internazionale, e comprende diverse unità di lavoro. Ogni Stato vi partecipa attraverso enti metrologici statali. Il BIPM si serve della *Guide to the Expression of Uncertainty in Measurement* (GUM) per fornire gli standard di misurazione. Il GUM è la guida per l'espressione e il modellamento dell'incertezza, ovvero il risultato del lavoro congiunto all'interno del BIPM, ed è interamente basata su un approccio statistico alla misurazione.

La **misurazione** può essere definita, utilizzando un approccio operativo, come un insieme di regole attuate per ottenere informazioni e produrre conoscenza su una quantità oggetto di interesse. Il concetto di quantità può essere definito come quella proprietà di un fenomeno a cui viene associato un numero (e dunque una *metrica*). L'attribuzione di numeri rende possibile alcune operazioni matematiche (sulla base dello spazio metrico scelto) le quali, a loro volta, devono poter rappresentare alcune delle azioni ammissibili sulle proprietà che sono oggetto dello studio. Quello che si viene a creare, dunque, è un

Hanno contribuito alla realizzazione del presente capitolo:

Niccolò Cao, Sara Dall'Oglio, Samuele Lovera, Lavinia Mattioli, Marta Pisani, Clara Pomoni, Carolina Romascu, Stefania Spada.

sistema di relazioni tra il sistema che rappresenta l'oggetto di interesse (e le proprietà di quest'ultimo) e il sistema metrico. Il legame che lega i due sistemi si attua mediante un *morfismo*.

Misurare una quantità richiede un riferimento metrologico, ovvero un'unità di misura convenzionale utilizzata come *riferimento* per le misurazioni (ad esempio, il metro per la lunghezza, il chilogrammo per il peso). In questo senso, dire che una matita è lunga 10 cm, significa che è lunga 10^{-2} (= 0.01) l'unità di riferimento che è il metro.

La quantità oggetto del processo misurativo è detta **misurando** e verrà indicata con la lettera η . Utilizzando il metodo sperimentale, l'obiettivo della misurazione è fornire una stima $\hat{\eta}$ precisa ed accurata del misurando. Per avere questa stima è indispensabile servirsi di un campione di misurazioni $\mathbf{x} = (x_1, \dots, x_n)$ estratto casualmente da una popolazione di riferimento Ω_η che rappresenta il misurando. Per ottenere una buona stima del misurando $\hat{\eta}$ occorre ripetere più volte la misurazione (ripetibilità della misurazione). Questo garantisce, in condizioni di ottimalità, la riduzione dell'errore di misurazione dovuto al processo di campionamento da Ω_η .

La differenza tra il risultato di una misurazione e il suo valore vero ($\hat{\eta} - \eta$) è detta **trueness** (più è grande la differenza, più la misurazione sarà inesatta e maggiore l'errore della stessa). Il *valore vero* η non è conoscibile e per determinare $\hat{\eta}$ occorre estrarre casualmente dalla popolazione Ω_η - che rappresenta il misurando - un campione di misurazioni \mathbf{x} . Il campione è finito ed ha cardinalità n . Ottenuto un campione di misurazioni, uno stimatore di η è la **media campionaria**:

$$\hat{\eta} = \text{mean}(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

sapendo che $\lim_{n \rightarrow \infty} \hat{\eta} \rightarrow \eta$.

1.1.1 Analisi degli errori

L'errore di una misurazione emerge dalla differenza tra il risultato della quantità misurata (la stima) e il suo valore vero:

$$\epsilon_\eta = (\hat{\eta} - \eta)$$

L'errore può essere causato da scarsa accuratezza, non ripetibilità delle misurazioni, scarsa risoluzione degli strumenti utilizzati per la misurazione, condizioni ambientali e malfunzionamento degli strumenti di misurazione. Nella teoria classica degli errori possiamo individuare tre tipologie generali di errore:

- errore casuale (*random error*)
- errore sistematico (*systematic error*)
- errore spurio (*mistake*)

Di seguito dettaglieremo alcune delle loro caratteristiche.

Errore casuale (*random error*)

Definire l'**errore casuale** δ_η implica la definizione di *ripetibilità* della misurazione. L'errore casuale, infatti, è intrinseco al misurando e discende dal processo di campionamento delle misure stesse. Un fatto empirico di rilievo è che fare con il fatto che aumentando la numerosità del campione n , ossia il numero di misurazioni effettuate, l'errore casuale δ_η tenderà a diminuire progressivamente all'aumentare di n . Perciò, idealmente, agendo sulla numerosità del campione è possibile far tendere questa quantità a zero, vale a dire:

$$\lim_{n \rightarrow \infty} \delta_\eta = 0$$

Ciò garantisce che $\lim_{n \rightarrow \infty} \hat{\eta} \rightarrow \eta$. Tuttavia tale fatto empirico non comporta l'eliminazione dell'incertezza casuale che ne risulta. Di fatti, poiché in relazione con le caratteristiche naturali dei misuranti (o della realtà empirica che li contiene), eliminare δ_η è non empiricamente realizzabile. Per tale motivo, l'errore casuale deve essere gestito attraverso modelli statistici opportuni che permettano di definire alcune proprietà per rendere una misurazione quanto più precisa possibile. Un esempio tipico della gestione statistica dell'errore è assumere che questo sia in media nullo ripetendo un numero elevato di volte la misurazione.

Per comprendere come l'errore casuale sia intrinseco al misurando, consideriamo l'esperimento del lancio di m sfere mediante un lanciatore a molla. La misurazione in questo caso consiste nel lanciare, una ad una, le m sfere nelle medesime condizioni sperimentali. La figura 1.1 mostra le distanze delle sfere dal loro punto di lancio iniziale unitamente alla distribuzione di tali distanze. È rilevante notare come la distanza dal punto di lancio sia differente per ciascuna delle m sfere sebbene le leggi che governano il moto delle sfere (leggi di Newton) rimangano invariate ad ogni lancio.

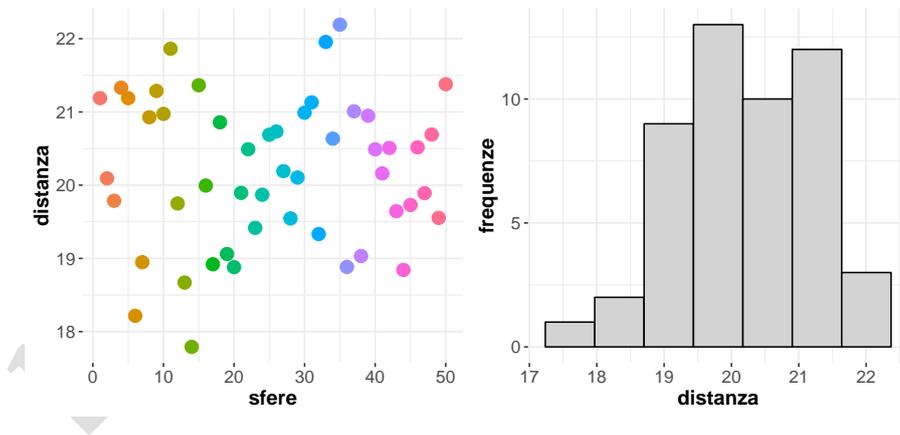


FIGURA 1.1: *Esemplificazione dell'errore casuale mediante lancio di m sfere nelle medesime condizioni sperimentali. Nota: i grafici rappresentano le distanze delle sfere dal punto di lancio.*

L'errore sistematico (*systematic error*)

L'**errore sistematico** κ_η è legato alla presenza di distorsioni nel processo misurativo che si aggiungono come quantità costanti ad ogni misurazione con il risultato di indebolire l'accuratezza della stessa. Diverse sono le cause: strumento difettoso, uso dello

strumento in condizioni diverse da quelle di taratura, errore dell'operatore durante la misurazione, ecc. L'errore sistematico ha un comportamento opposto a quello dell'errore casuale: all'aumentare delle misurazioni n , κ_η tenderà a un valore diverso da zero che distorce la convergenza della stima del misurando alla quantità ignota da misurare. La distorsione, pari a κ , è tale per cui si ha il seguente fatto:

$$\lim_{n \rightarrow \infty} \hat{\eta} = \eta + \kappa \quad \text{e} \quad \frac{1}{n} \sum_{i=1}^n x_i \neq \eta$$

In tal caso, per stimare adeguatamente η mediante un campione di misurazioni \mathbf{x} , occorrerà - *quando possibile* - sottrarre alla stima $\hat{\eta}$ la quantità κ affinché si abbia ancora che $\hat{\eta} \rightarrow \eta$ quando $n \rightarrow \infty$.

L'errore spurio (*mistake*)

L'**errore spurio** è una fonte di errore causata solitamente dall'azione dell'operatore che effettua la misurazione (osservatore) o dall'uso di uno strumento difettoso. Si comporta allo stesso modo dell'errore sistematico per cui distorce l'accuratezza della misurazione. Le misurazioni soggette a errore spurio non possono essere considerate ai fini dell'analisi dei dati misurativi. La figura 1.2 rappresenta un istogramma di misurazioni soggette ad errore spurio. Notiamo come la gran parte delle misurazioni si situa attorno ad un baricentro (segmento di linea rossa tratteggiata) mentre una componente residuale di misurazioni si situa all'esterno del gruppo principale di misurazioni. Una rappresentazione grafica del genere è indicatrice di misurazioni contenenti errore spurio.

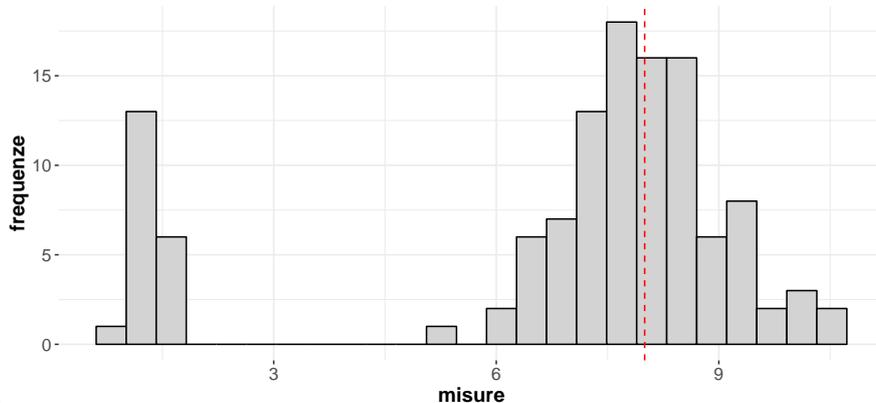


FIGURA 1.2: *Esemplificazione dell'errore spurio: grafico di n misurazioni di η . La presenza di misurazioni errate è visibile dallo scostamento di queste dalla distribuzione principale delle misurazioni. Nota: la linea rossa tratteggiata indica il vero valore del misurando.*

1.1.2 Proprietà del processo misurativo

Per valutare una misurazione si possono considerare due caratteristiche notevoli:

- l'**accuratezza**, sinonimo di correttezza, indica quanto le misurazioni (x_1, \dots, x_n) si avvicinano al valore vero η

- la **precisione**, ossia la dispersione che le misurazioni (x_1, \dots, x_n) hanno attorno al loro valore medio $\hat{\eta}$

Le figure 1.3-1.4 riportano una rappresentazione grafica di quattro campioni di misurazioni rilevate in tempi diversi. In particolare, la figura 1.3 mostra un caso di misurazioni affette da medesima precisione (come evidenziato dalla scatola del box-plot che rappresenta la differenza interquartile) e scarsa accuratezza (come evidenziato dallo scostamento in mediana dalla linea rossa tratteggiata, indicante il vero valore del misurando). Al contrario, la figura 1.4 mostra un caso di misurazioni affette da buona accuratezza e differenti precisioni. La figura 1.5 invece, riporta il caso di diverse misurazioni effettuate per il medesimo misurando η ed affette da diverso grado di precisione e accuratezza. In generale, per valutare la precisione e l'accuratezza è possibile utilizzare una rappresentazione grafica ad istogramma o a scatole e baffi (box-plot).

In generale, dato un campione di n misure $\mathbf{x} = (x_1, \dots, x_n)$:

- l'errore casuale δ_η influenza la precisione delle misurazioni e rallenta/velocizza la convergenza al valor vero $\hat{\eta} \rightarrow \eta$
- l'errore sistematico influenza l'accuratezza delle misurazioni e distorce la convergenza $\hat{\eta} \rightarrow \eta$ di una quantità costante κ tale per cui si ha che $\frac{1}{n} \sum_{i=1}^n x_i \neq \eta^*$

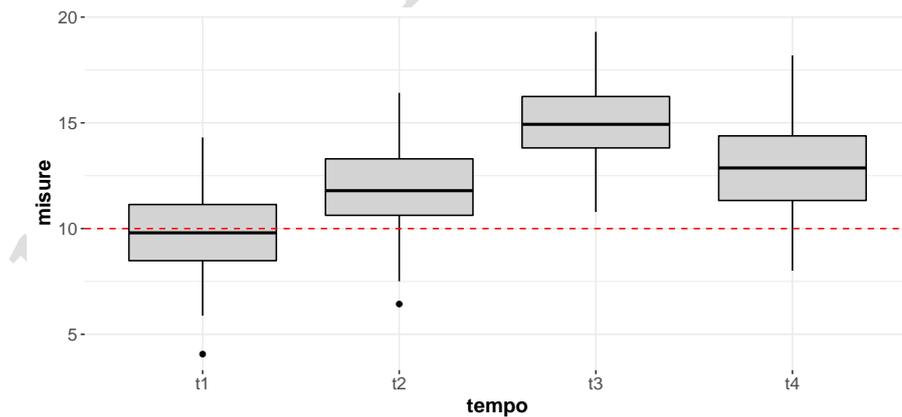


FIGURA 1.3: Andamento di n misure effettuate in tempi diversi: lo scostamento da η (accuratezza) indica una possibile presenza di errore sistematico κ_η mentre la precisione (differenza interquartile rappresentata dalle scatole dei grafici) resta invariata al variare del tempo. Nota: la linea rossa tratteggiata indica il vero valore di misura η .

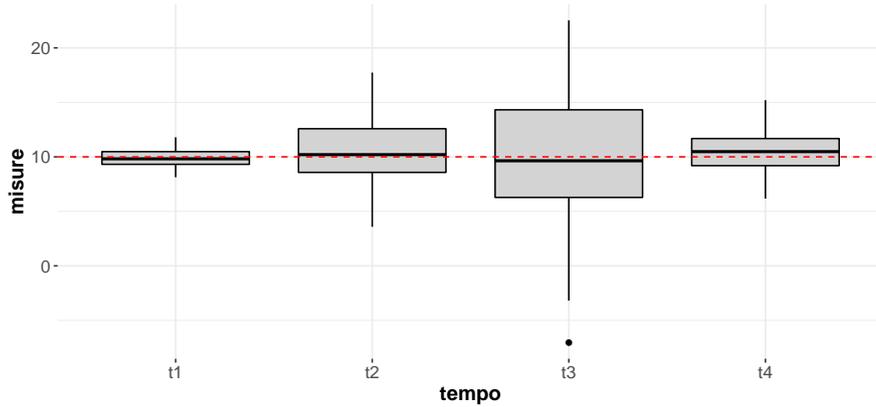


FIGURA 1.4: Andamento di n misure effettuate in tempi diversi: la variazione della precisione indica la presenza di errore casuale δ_η mentre l'accuratezza resta invariata al variare del tempo. Nota: la linea rossa tratteggiata indica il vero valore di misura η .

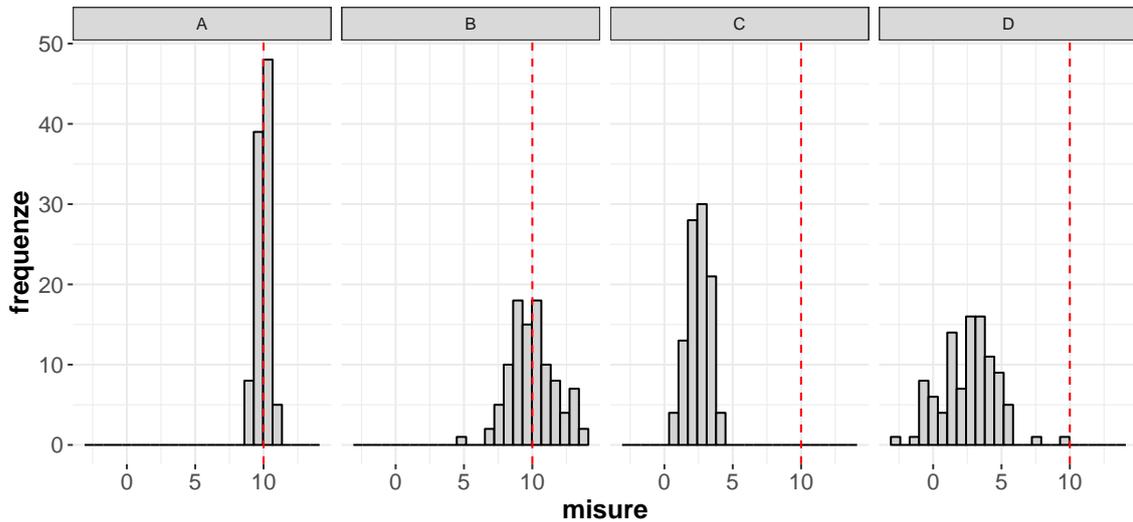


FIGURA 1.5: Differenze tra accuratezza e precisione. Note: la linea rossa tratteggiata indica il vero valore di misura η , (A) accurato e preciso, (B) accurato e impreciso, (C) inaccurato e preciso, (D) inaccurato e impreciso.

1.2 Stimare il misurando, l'accuratezza e la precisione

1.2.1 Stima del misurando

Dato un insieme di misurazioni $\mathbf{x} = (x_1, \dots, x_n)$ effettuate nelle medesime condizioni, una stima $\hat{\eta}_x$ della quantità di interesse η è ottenuta tramite la media aritmetica:

$$\hat{\eta}_x = \frac{1}{n} \sum_{i=1}^n x_i$$

Alcune proprietà notevoli della media aritmetica:

1. *Internalità*: la stima è compresa fra il valore minimo e quello massimo, ovvero $\min(x) \leq \hat{\eta}_x \leq \max(x)$
2. *Invarianza* rispetto alla somma
3. $(\sum_{i=1}^n x_i - \hat{\eta}) = 0$
4. $\min_{c \in \mathcal{C}} \sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \hat{\eta}_x)^2$ quando c è la media aritmetica
5. *associatività*: se $\mathbf{x} = \underbrace{\{(x_1, x_2), \dots, (x_{n-1}, x_n)\}}_{m \text{ sottogruppi}}$ allora $\hat{\eta}_x = \frac{1}{\sum_{j=1}^m n_j} \sum_{j=1}^m \hat{\eta}_j n_j$

Alcuni limiti della media aritmetica:

- è attratta dai valori anomali della distribuzione
- non è utile quando le distribuzioni delle misurazioni sono non unimodali (le distribuzioni unimodali hanno un solo picco)
- non può essere individuata mediante mezzi grafici e non può essere determinata solo attraverso una semplice osservazione dei dati
- non si può applicare ai dati qualitativi, è necessario che i dati siano di tipo quantitativo
- per calcolare la media aritmetica sono necessarie tutte le osservazioni (no osservazioni mancanti)

La figura 1.6 esemplifica due casi in cui la stima $\hat{\eta}$ effettuata mediante media aritmetica non è una buona stima del misurando η a causa della presenza di errore spurio (figura 1.6-A) o misurazioni bimodali (figura 1.6-B).

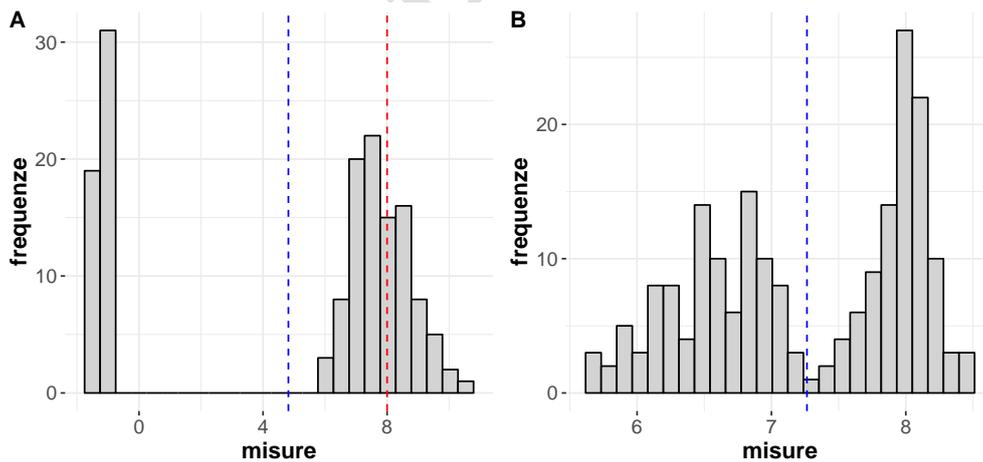


FIGURA 1.6: *Istogrammi di alcune misurazioni: (A): caso in cui $\hat{\eta}$ è attratta da misurazioni anomali. (B): caso in cui $\hat{\eta}$ è calcolata per misurazioni con distribuzioni bimodali. Nota: la linea rossa tratteggiata indica il vero valore di misura η mentre quella blu indica la media aritmetica $\hat{\eta}$.*

1.2.2 Stima della precisione

Campo di variazione

Un primo modo per calcolare la precisione è mediante il *campo di variazione*:

$$\Delta_x = \max(\mathbf{x}) - \min(\mathbf{x})$$

Proprietà notevoli:

1. $\Delta_x \geq 0 \rightarrow$ (il campo di variazione non può assumere valori negativi)
2. $\Delta_x = 0$ se $\mathbf{x} = (x, \dots, x)$, ossia se le misurazioni sono tutte uguali tra loro in magnitudine
3. se $\mathbf{y} = a + \mathbf{x}$, allora $\Delta_y = \Delta_x \rightarrow$ (ad esempio, qualora fosse presente un errore sistematico, la precisione della misura non cambierebbe)
4. se $\mathbf{y} = b\mathbf{x}$, allora $\Delta_y = b\Delta_x \rightarrow$ (ad esempio, se le misurazioni fossero moltiplicate per una costante b , anche il campo di variazione verrebbe moltiplicato per la medesima costante, ottenendo quindi una riscalatura della precisione).

Alcuni limiti del campo di variazione:

- dipende unicamente dai valori estremi delle misurazioni
- assume lo stesso valore per distribuzioni con medesimi estremi

Scarto quadratico medio

Un secondo modo per calcolare la precisione è quello di usare lo scarto quadratico medio (*standard deviation*):

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (x_i - \hat{\eta}_x)^2}$$

Proprietà notevoli:

1. $\sigma_x \geq 0 \rightarrow$ (lo scarto quadratico medio non può assumere valori negativi)
2. $\sigma_x = 0$ se $\mathbf{x} = (x, \dots, x) \rightarrow$ (quando tutte le misurazioni sono uguali, queste coincidono con la media e lo scarto è pari a zero)
3. se $\mathbf{y} = a + \mathbf{x}$, allora $\sigma_y = \sigma_x \rightarrow$ (ad esempio, qualora fosse presente un errore sistematico, la precisione della misura non cambierebbe)
4. se $\mathbf{y} = b\mathbf{x}$, allora $\sigma_y = b\sigma_x \rightarrow$ (ad esempio, se le misurazioni fossero moltiplicate per una costante b , anche lo s.q.m. verrebbe moltiplicato per la medesima costante, ottenendo quindi una riscalatura della precisione)

Coefficiente di variazione

Un terzo modo per calcolare la precisione è utilizzando il coefficiente di variazione (*errore relativo*):

$$CV(\mathbf{x}) = \frac{\sigma_x}{|\hat{\eta}_x|}$$

che è spesso rappresentato in scala percentuale $CV \cdot 100\%$.

Alcuni vantaggi derivanti dall'uso di tale indice sono i seguenti:

- è un indice adimensionale utile per confrontare misurazioni effettuate con diverse unità di misura (è un numero puro)
- è utile per confrontare misurazioni con diverse medie (indipendentemente dall'unità di misura)
- è utile per confrontare in circostanze diverse la variabilità relativa di un dato fenomeno.

Lo svantaggio principale si ha nel caso in cui:

$$CV(\mathbf{x}) \rightarrow \infty \quad \text{quando} \quad \hat{\eta}_x \rightarrow 0$$

vale a dire, nel caso in cui le misure del fenomeno studiato hanno media nulla o prossima allo zero il coefficiente di variazione tende ad infinito.

Rapporto segnale-rumore

Un quarto modo per calcolare la precisione è mediante l'utilizzo del rapporto segnale-rumore (*signal-to-error ratio*):

$$SNR(\mathbf{x}) = \frac{\hat{\eta}_x}{\sigma_x}$$

che valuta il rapporto tra quantità di segnale (quantificato da $\hat{\eta}_x$) e rumore (quantificato da σ_x). In generale, un valore di $SNR(\mathbf{x}) \geq 5$ indica un'ottima capacità di distinguere il segnale dal rumore. Il principale svantaggio di quest'indice è dovuto al fatto che $\mathbf{x} \in \mathbb{R}_+^n$, ossia esso ha senso solo per misurazioni sempre positive.

La figura 1.7 mostra un caso di applicazione degli indici CV e SNR per quantificare la dispersione.

1.2.3 Stima dell'accuratezza

La quantificazione dell'accuratezza di un processo misurativo è fatta sfruttando la ripetibilità della misurazione. In particolare, consideriamo quanto segue:

dove diversi campioni di misurazioni di ampiezza n sono rilevati J volte (ripetendo cioè il processo misurativo J volte). In questo caso, le medie delle diverse misurazioni $\hat{\eta} = (\hat{\eta}_{x1}, \dots, \hat{\eta}_{xj})$ possono essere utilizzate per stimare con maggiore accuratezza il valore vero η . Sotto determinate condizioni, infatti, la distribuzione di $\hat{\eta}$ è più accurata e precisa e per $J \rightarrow \infty$ si ha che:

$$\hat{\eta}^0 = \left(\frac{1}{J} \sum_{j=1}^J \hat{\eta}_{xj} \right) \rightarrow \eta$$

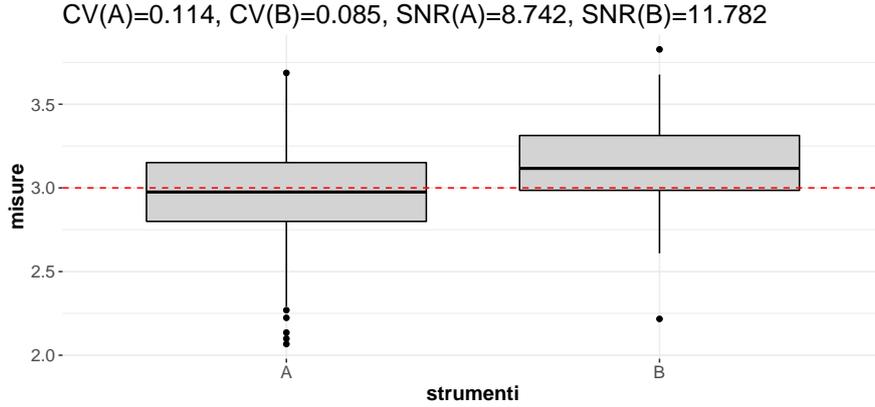


FIGURA 1.7: Grafico di n misurazioni di η effettuate con due strumenti A e B. Da notare come $\hat{\eta}_A$ è più vicina a η rispetto a $\hat{\eta}_B$. Tuttavia lo strumento A presenta maggiore variabilità rispetto allo strumento B, evidenziato dal fatto che $CV(A) > CV(B)$. Allo stesso modo, lo strumento B è migliore nell'estrarre il segnale dal rumore, infatti $SNR(A) < SNR(B)$. Lo strumento B può essere preferito allo strumento A, correggendolo opportunamente $\eta = \hat{\eta}_B - \kappa_B$ (κ_B è una distorsione sistematica). Nota: la linea rossa tratteggiata indica il vero valore di misura η

ripetizione	campione	stima di η
1	$(x_1, \dots, x_n)_1$	$\hat{\eta}_{x_1}$
\vdots	\vdots	\vdots
j	$(x_1, \dots, x_n)_j$	$\hat{\eta}_{x_j}$
\vdots	\vdots	\vdots
J	$(x_1, \dots, x_n)_J$	$\hat{\eta}_{x_J}$

mentre

$$\hat{\sigma}_{\hat{\eta}_0} = \sqrt{\frac{1}{J} \sum_{j=1}^J (\hat{\eta}_{x_j} - \hat{\eta}^0)^2} \rightarrow 0$$

Premesso ciò, in generale, dato un campione di n misurazioni, l'errore della stima (*standard error*) $\sigma_{\hat{\eta}}$ è calcolato come

$$\sigma_{\hat{\eta}} = \frac{\sigma_x}{\sqrt{n}}$$

dove σ_x è la deviazione standard delle n misurazioni. Il fattore $\frac{1}{\sqrt{n}}$ indica che l'errore diventa più piccolo all'aumentare di n . Per indicare l'accuratezza di una misurazione si usa la notazione $\hat{\eta} \pm \sigma_{\hat{\eta}}$ (ad esempio: 13.46 ± 0.04) che non indica altro che il range della stima di η :

$$\hat{\eta} - \sigma_{\hat{\eta}} \leq \eta \leq \hat{\eta} + \sigma_{\hat{\eta}}$$

Nota bene:

- σ_x indica la dispersione delle misurazioni (x_1, \dots, x_n) intorno a $\hat{\eta}_x$.
- $\sigma_{\hat{\eta}}$ indica un'informazione circa la concordanza tra $\hat{\eta}$ e η^* .

L'interpretazione dell'accuratezza dunque risulterà come segue: minore è la lunghezza dell'intervallo $\hat{\eta} \pm \sigma_{\hat{\eta}}$, maggiore sarà la *concordanza* (vicinanza) tra $\hat{\eta}$ e η . La figura

1.8 esemplifica il comportamento dell'accuratezza quando aumenta la numerosità delle misurazioni n .

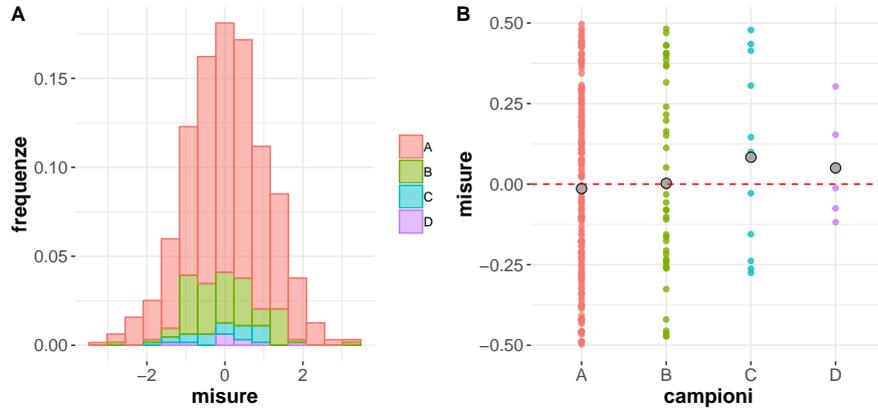


FIGURA 1.8: Accuratezza della misurazione e stima del misurando. (A) Istogrammi di quattro campioni di misure con $n_A = 500$, $n_B = 100$, $n_C = 25$, $n_D = 10$. Notiamo come all'aumentare di n la distribuzione delle misurazioni si centra intorno ad un valore medio. (B) Stessi campioni visti di profilo: il cerchio grigio indica la media calcolata sul campione. Per n piccolo, la media del campione si allontana da quella vera. Nota: la linea rossa tratteggiata indica il vero valore di misura η mentre in entrambi i grafici $\hat{\eta}_A = 0.05 \pm 0.04$, $\hat{\eta}_B = -0.01 \pm 0.09$, $\hat{\eta}_C = -0.03 \pm 0.16$, $\hat{\eta}_D = 0.14 \pm 0.27$

1.3 Analisi dell'incertezza

Ogni misurazione è portatrice di incertezza e ciò è dovuto a diversi fattori (del misurando, dello strumento, ambientali, ecc). Prima delle regole GUM, l'incertezza di misura era rappresentata mediante la **teoria degli errori classica** (errore casuale *vs* sistematico) che dava adito a diverse criticità, tra cui la scelta del modello matematico per la valutazione dell'errore sistematico e la scelta del metodo per combinare gli errori casuale e sistematico. La teoria classica presenta altresì un limite nella definizione stessa di misurando: il vero valore η è un ideale e pertanto impossibile da conoscere esattamente (non è nemmeno possibile calcolare il valore esatto dell'errore di misura). Da ciò ne consegue che le conoscenze fornite da un processo di misurazione, cioè il risultato della misurazione, sono sempre incomplete. Per tale motivo, alla fine del ventesimo secolo, la teoria dell'errore classica è stata pian piano affiancata dalla **teoria dell'incertezza**, che introduce il concetto di incertezza come *parametro*, associato al risultato di una misurazione, che caratterizza la dispersione dei valori che potrebbero ragionevolmente essere attribuiti al misurando [33]. Secondo le regole GUM, tale parametro è determinato statisticamente. Infatti, poiché si presume che il risultato di una misurazione sia stato corretto per tutti gli effetti sistematici riconosciuti dall'osservatore, il risultato di una misurazione dovrebbe essere influenzato solo da effetti casuali. L'introduzione del concetto di incertezza, in generale, supera anche la distinzione classica tra fonte di errore casuale e sistematica per introdurre invece quella di **incertezza totale**.

Più specificatamente, secondo il vocabolario internazionale di metrologia, l'incertezza è definita come:

«[...] non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. Measurement uncertainty includes components arising from systematic effects, such as components associated with corrections [...]. The parameter may be, for example, a standard deviation called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability. [...] Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by Type A evaluation of measurement uncertainty from the statistical distribution of the quantity values from series of measurements and can be characterized by standard deviations. The other components, which may be evaluated by Type B evaluation of measurement uncertainty, can also be characterized by standard deviations, evaluated from probability density functions based on experience or other information.» (definition 2.26) [4].

Secondo questa definizione, l'incertezza nel risultato di una misurazione è generalmente costituita da diversi componenti che possono essere raggruppate in due categorie principali:

- incertezza di **tipo A** che raggruppa l'incertezza di tipo statistico (quella attribuibile in definitiva alla fonte casuale dell'errore)
- incertezza di **tipo B** che comprende invece tutte quelle fonti di incertezza non statistica

1.3.1 Valutazione dell'incertezza tipo A e tipo B

Sia $\mathbf{x} = (x_1, \dots, x_n)$ un campione di n misurazioni utilizzato per portare informazioni sul misurando η . In generale:

$$\underbrace{\eta}_{\text{misurando}} = f(\underbrace{\mathbf{x}}_{\text{input}})$$

ossia il misurando, ciò che si vuole misurare, è espresso in funzione (nota) delle misurazioni empiriche \mathbf{x} (input). Come discusso in precedenza, la stima migliore di η è la media dell'input \mathbf{x} mentre la sua incertezza è espressa dalle quantità $\hat{\sigma}_{\hat{\eta}}$ e $\hat{\sigma}_x$. Poiché l'incertezza tipo A comprende l'incertezza statistica (dovuta cioè al processo di misurazione), si riconducono a questa categoria tutte le fonti casuali che influenzano la misura e di cui se ne può tenere conto tramite il calcolo dai loro parametri statistici (media e deviazione standard).

Al contrario, l'incertezza tipo B esprime tutte quelle fonti di incertezza non comprese nel tipo A, ovvero informazioni a priori sui *bias* dello strumento o il processo di misurazione, errore sistematico, quello che l'osservatore reputa essere un errore di misurazione (ad esempio, il tremore della mano dell'osservatore). Indichiamo con u_x tale incertezza.

L'incertezza totale della misurazione, secondo GUM, è dunque quantificata utilizzando la *regola della quadratura*

$$\hat{\sigma}_{tot} = \sqrt{\hat{\sigma}_x^2 + u_x^2}$$

dove $\hat{\sigma}_x$ è l'incertezza dovuta alla ripetibilità della misurazione. Quando $n = 1$, ossia nel caso di una misurazione singola:

$$\hat{\sigma}_{tot} = \sqrt{u_x^2}$$

ossia l'incertezza totale è espressa solo dall'incertezza dovuta allo strumento. Per un approfondimento sull'analisi dell'incertezza si veda [16] (in particolare §7.6-7.7), [33], [12].

1.4 Misurazione indiretta e propagazione dell'incertezza

Quando il misurando è rappresentato in funzione di una serie di quantità misurate $(\mathbf{x}_1, \dots, \mathbf{x}_J)$ dove il generico \mathbf{x}_j è un vettore di n misurazioni come di consueto:

$$\eta = f(\mathbf{x}_1, \dots, \mathbf{x}_J)$$

si parla di **misurazione indiretta**. In questo caso il valore stimato del misurando è ottenuto applicando la funzione f sulle J misurazioni mentre l'incertezza $\hat{\sigma}_{\hat{\eta}}$ è ottenuta usando le regole della **propagazione dell'incertezza** (per una rassegna si veda [17], §4.1.3 e §4.2.4).

In generale, la regola di propagazione quando η è misurato indirettamente è la seguente:

$$\hat{\sigma}_{\hat{\eta}} = \sum_{j=1}^J c_j^2 \hat{\sigma}_{x_j} + 2 \sum_{l=1}^{J-1} \sum_{h=l+1}^J c_l c_h \hat{\sigma}_{x_l} \hat{\sigma}_{x_h} \hat{\rho}_{x_l x_h}$$

dove:

- $\sum_{j=1}^J c_j^2 \hat{\sigma}_{x_j}$ indica il contributo della j -esima misurazione
- $\sum_{l=1}^{J-1} \sum_{h=l+1}^J c_l c_h \hat{\sigma}_{x_l} \hat{\sigma}_{x_h} \hat{\rho}_{x_l x_h}$ indica la correlazione tra coppie di misurazioni (l, h) con $l \neq h$
- $c_j := \frac{\partial f}{\partial x_j}$ è il coefficiente di sensibilità (sensitivity coefficient) ed è calcolabile usando ad esempio le espressioni indicate in [17] (Tabelle 4.1-4.2)
- quando x_l e x_h sono indipendenti il secondo addendo diventa zero e la formula per propagare l'incertezza si semplifica apportando solo contributi delle singole misurazioni:

$$\hat{\sigma}_{\hat{\eta}} = \sum_{j=1}^J c_j^2 \hat{\sigma}_{x_j}$$

Di seguito si riporta un'applicazione esemplificativa per il calcolo della propagazione dell'incertezza in un caso di semplice risoluzione.

Si vuole misurare η utilizzando due misure X_1 e X_2 . A tal fine si raccolgono due campioni \mathbf{x}_1 e \mathbf{x}_2 di n misurazioni indipendenti con $\hat{\sigma}_{x_1} = 2.1$ e $\hat{\sigma}_{x_2} = 0.4$. È noto che la funzione che esprime il misurando è la seguente: $\eta = 2.36 + \ln(X_1) + \ln(X_2)$.

L'incertezza delle misure di input è propagata sul misurando usando la seguente formulazione (si veda [12], Tabella 2, formula 14):

$$\hat{\sigma}_{\hat{\eta}}^2 = \left(\frac{\hat{\sigma}_{x_1}}{\bar{X}_1} \right)^2 + \left(\frac{\hat{\sigma}_{x_2}}{\bar{X}_2} \right)^2$$

Sapendo che $\bar{X}_1 = 4.1$ e $\bar{X}_2 = 9.3$ si ottiene quanto segue:

- $\hat{\sigma}_{\hat{\eta}} = \sqrt{\hat{\sigma}_{\hat{\eta}}^2} = 0.514$
- $\hat{\eta} = 2.36 + \log(4.1) + \log(9.3) = 6.001$

1.5 Note di sintesi

- (i) La teoria degli errori classica e l'approccio GUM offrono due approcci per la gestione dell'incertezza di misurazione
- (ii) La GUM offre una rappresentazione statistica dell'incertezza di misurazione
- (iii) Entrambi sono (inconsapevolmente) condivisi dalla gran parte della metodologia classica per la costruzione dei test comportamentali
- (iv) L'incertezza di tipo B può essere agevolmente modellata da approcci probabilistici di tipo Bayesiano
- (v) L'incertezza di tipo B può essere utilizzata per modellare diverse fonti di disturbo nel testing psicologico (es.: fake data)
- (vi) Diversi modelli statistici possono innestarsi all'interno della GUM per caratterizzare meglio il problema dell'incertezza [33]
- (vii) In GUM la teoria della probabilità è impiegata per modellare il processo di misurazione
- (viii) La misurazione è rappresentata come esperimento aleatorio
- (ix) Le leggi statistiche fondamentali (es.: Teorema del limite centrale) sono alla base della quantificazione del misurando

1.6 Glossario minimale

In generale η indica il misurando, la quantità da misurare. Il suo valore vero è generalmente ignoto e non definibile se non si conosce la teoria di riferimento, quindi la scala su cui η si esprime ed eventuali relazioni che ha con altre proprietà misurabili. La stima empirica del misurando è denotata con $\hat{\eta}$ ed è ottenibile come media di misure ottenute secondo il principio ideale del campionamento ripetuto. Il simbolo \mathbf{x} indica un campione di misurazioni, estratto casualmente dalla popolazione di riferimento del fenomeno in oggetto Ω_{η} . Questa rappresenta il misurando, poiché sarà proprio la popolazione le cui caratteristiche vogliamo conoscere e studiare. La quantità ϵ_{η} è l'errore di misura,

inevitabile in tutte le misurazioni per diversi motivi (strumento, condizioni empiriche, processo di campionamento). Si definisce come $\epsilon_\eta = (\hat{\eta} - \eta)$. La quantità di errore δ indica l'errore casuale mentre κ denota l'errore sistematico.

1.7 Laboratorio in R

I laboratori associati a questo capitolo sono i seguenti:

- o **lab1.R**: Richiami sull'utilizzo del linguaggio R
- o **lab2.R**: (A) Esplorazione grafica di misurazioni empiriche, (B) Stima di η e della precisione, (C) Misurazione indiretta e propagazione dell'errore

Il loro contenuto può essere visualizzato e scaricato all'indirizzo <https://github.com/antcalcagni/psychological-testing/tree/master/labs>.

bozza non corretta

bozza non corretta

2

Richiami di statistica per la misurazione

«Accept uncertainty. Be thoughtful, open, and modest. Remember “ATOM”»
– Wasserstein, Schirm & Lazar, 2019

La misurazione è alla base del processo di scoperta scientifica. Essa è soggetta inevitabilmente ad incertezza, come sottolineato sia dalla teoria degli errori classica che dall’approccio GUM. Le fonti di tale incertezza sono molteplici: ed hanno a che fare, ad esempio, con lo strumento di misura (ad esempio, una taratura errata), l’ambiente entro cui la misurazione avviene (ad esempio, la mutevolezza dell’ambiente sperimentale), il soggetto che effettua la misurazione e, non per ultime, le proprietà stesse intrinseche al misurando. Tali fonti, se hanno una preminenza su quelle che sono le misurazioni fisiche, hanno un effetto ancora più marcato su quelle che da più voci sono state definite *soft measurements* [3, 13, 21]. Dal momento che l’incertezza si rivela essere una quantità ineliminabile all’interno del processo misurativo, risulta necessario individuare modelli rigorosi capaci di controllarne le caratteristiche ed il comportamento. Senza avere la presunzione di effettuare un’elencazione completa ed esaustiva delle teorie e dei modelli per la gestione dell’incertezza, Klir [19] elenca alcuni di quelli che rientrano a pieno titolo nei modelli matematici per la rappresentazione dell’incertezza, tra cui la teoria della *probabilità* (Kolmogorov, De Finetti), la teoria della *possibilità* (Zadeh, Dubois) e la teoria delle *credenze* (Dempster, Shafer). Tra questi modelli, la teoria della probabilità occupa un posto di rilievo e, considerando gli sviluppi matematici occorsi negli ultimi due secoli nel campo, può essere ritenuta *a fortiori* il modello standard per la rappresentazione e gestione dell’incertezza.

Hanno contribuito alla realizzazione del presente capitolo:
Elisa Berlin, Margherita Biondi, Giulia Coccato, Sharon Luisa Delle Monache, Seila Mainardis, Angela Pesce, Giulia Vomiero

Nel seguito verranno riportati alcuni *richiami* elementari di teoria della probabilità mentre si rimanda a testi specializzati, ad esempio [6, 32], per una trattazione integrale.¹

2.1 Esperimenti aleatori

Un **esperimento aleatorio** (o prova) è un esperimento i cui **esiti** sono soggetti ad incerti. Mentre gli esiti possibili sono certi, ciò che è invece soggetto ad incertezza è l'*accadere* o meno di un certo esito. La collezione di tutti gli esiti possibili di una prova aleatoria (enumerabile) è detta **spazio campionario** Ω mentre sottoinsiemi A_1, \dots, A_m di Ω sono detti **eventi**. Ad esempio, nel caso del lancio di due monete lo spazio campionario (finito ed enumerabile) è $\Omega = \{cc, tt, tc, ct\}$ mentre l'evento secondo cui il "primo esito del lancio è testa" è definito da $A_1 = \{tt, tc\}$. Allo stesso modo, l'evento "almeno un esito è croce" è rappresentato da $A_2 = \{cc, tc, ct\}$. In generale, l'insieme di sottoinsiemi di Ω rappresenta l'insieme delle affermazioni possibili circa gli eventi che possono accadere con la prova aleatoria. Poiché gli eventi sono definiti in termini di insiemi, è possibile manipolarli attraverso le tipiche operazioni insiemistiche definite dall'algebra di Boole, ad esempio mediante *unione*, *intersezione*, *negazione* (o complementazione). Data una collezione di eventi A_1, \dots, A_m di Ω , la **probabilità** $\mathbb{P}(\cdot)$ è una funzione di insiemi tale che:

- (i) $0 \leq \mathbb{P}(A_j) \leq 1$
- (ii) $\mathbb{P}\left(\bigcup_{j=1}^m A_j\right) = \sum_{j=1}^m \mathbb{P}(A_j)$

La proprietà (i) asserisce che la probabilità di un evento (un sottoinsieme di Ω) è un numero reale compreso tra zero ed uno con la convenzione che quando la probabilità dell'evento tende a zero allora l'evento è poco probabile. La proprietà (ii) asserisce che quando $A_{j-1} \cap A_j = \emptyset$ per $(j = 2, \dots, m)$, ossia nel caso di eventi disgiunti a due a due (eventi incompatibili), la probabilità dell'unione di eventi è pari alla somma delle probabilità dei singoli eventi. L'unione di tutti i sottoinsiemi di Ω deve essere uguale ad Ω stesso (escludendo le intersezioni se presenti) e, conseguentemente, la somma dei singoli eventi dello spazio campionario deve essere pari ad uno.

2.2 Assegnazioni di $\mathbb{P}(\cdot)$

La misura $\mathbb{P}(\cdot)$ può essere assegnata in diversi modi:

- **Assegnazione classica:** la probabilità è assegnata come rapporto tra numero di casi a favore dell'evento target ed il numero di eventi possibili (rapporto tra cardinalità di insiemi), secondo la definizione che segue:

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}$$

¹Una trattazione ragionata della teoria della probabilità richiederebbe l'introduzione, almeno ad un livello base, di *teoria della misura* all'interno della quale definire le funzioni misurabili ed in particolare la misura di integrale secondo Lebesgue. Poiché una tale trattazione esula dallo scopo specifico ed introduttivo del presente capitolo, si rimanda a testi più dettagliati sull'argomento (ad esempio [31]).

dove $\#$ indica la cardinalità di un insieme.

È una definizione che possiamo definire *pre-empirica* o *analitica* nel senso che la sua definizione prescinde dall'esecuzione dell'esperimento (ad esempio, nel lancio di una moneta non truccata, l'insieme Ω e l'insieme dei sottoinsiemi di Ω è già noto). L'assegnazione classica, dunque, costruisce $\mathbb{P}()$ sulla base del sistema di eventi noto in generale, ossia prima dell'esperimento aleatorio, secondo un approccio di tipo analitico.

- **Assegnazione frequentista:** la probabilità è assegnata usando una successione di prove (aleatorie) indipendenti e si calcola dividendo la frequenza dell'evento target con il numero di prove complessivamente effettuate:

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \frac{f_A}{n}$$

dove f_A indica la frequenza relativa (osservata) di esiti a favore dell'evento target A .

Tale definizione richiede che la prova sia sottoposta al principio della ripetibilità. Un esempio notevole di un tale esperimento aleatorio è proprio quello della misurazione, che è ripetibile secondo l'accezione frequentista.

- **Assegnazione soggettivista:** la probabilità è assegnata secondo il grado di fiducia che si ha sul sistema degli eventi (in termini di scommessa, evidenza, ecc). In altri termini, tale assegnazione è connessa alla valutazione che un individuo effettua circa la verificabilità di un esito di per sé incerto [29].

Qualunque sia l'assegnazione della misura di probabilità, da cui ne discende evidentemente l'interpretazione, il sistema matematico di rappresentazione e calcolo alla sua base non muta. Sebbene impossibile da enucleare in questa breve trattazione, un discorso a parte merita l'interpretazione logicista della probabilità, secondo cui ad ogni misura di probabilità corrisponde un corpus di evidenza nel mondo empirico, ossia un fatto oggettivo da cui è possibile presumere la misura stessa della probabilità. Così, ad esempio, asserire che la probabilità di un'era glaciale tra n anni è pari a p significa rintracciare quell'insieme di evidenze oggettive, fattuali, tra cui il tasso attuale di riscaldamento globale, ad esempio, che rendano ragione dell'oggettività stessa dell'assegnazione p [27]. Problemi simili a questo, di natura epistemologica e filosofica, possono essere rintracciati in [14].

Per una trattazione della probabilità condizionale e dei modelli per prove indipendenti si vedano le fonti [6, 29].

2.3 Variabili aleatorie

Per casi empirici di rilievo spesso è difficoltoso costruire spazi di probabilità e/o utilizzarli per il calcolo effettivo. A tale riguardo, l'uso delle **variabili aleatorie** (v.a.) fornisce una rappresentazione matematica semplificata dell'esperimento aleatorio che, attraverso la definizione di un insieme di modelli probabilistici generali (ad esempio, modello Gaussiano o Normale, Esponenziale, Binomiale, ecc) utili per rappresentare e

studiare diversi tipi di fenomeni aleatori, permette agevolmente di fare alcune operazioni importanti come la somma di esperimenti aleatori.² L'introduzione delle variabili aleatorie, infatti, permette la definizione di una sorta di *algebra* delle variabili aleatorie. Una definizione non formale, qualitativa, della v.a. è la seguente: una regola, una mappa X , che fa corrispondere ad ogni elemento ω di Ω (o dei suoi sottoinsiemi) un numero reale x .³ Una v.a. è dunque una funzione di insiemi (dominio) a valori reali (codominio). L'insieme dei valori che X può assumere è detto **supporto** S ; se S è finito e numerabile (ad esempio, insieme dei numeri naturali \mathbb{N}) la v.a. X è detta **discreta**; al contrario, se S è infinito (come nel caso dell'insieme numeri reali \mathbb{R}) X è detta **continua**. La figura 2.1 mostra una tipica rappresentazione di variabile aleatoria.

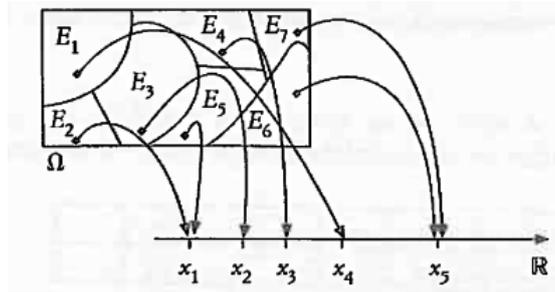


FIGURA 2.1: Variabile aleatoria: Corrispondenza tra eventi dello spazio campionario Ω e insieme dei numeri reali. Quest'ultimo, in figura rappresentato dalla retta orientata, è il supporto della variabile aleatoria. La figura è ripresa, senza modifiche, da [29].

L'introduzione delle variabili aleatorie permette di lavorare con spazi di probabilità più semplici formati sempre da elementi in \mathbb{R} . In tal modo, stesse variabili aleatorie possono essere usate come *modelli* per diversi spazi campionari, questi ultimi definiti in maniera anche diversa tra loro. Variabili aleatorie con supporto in \mathbb{R} (retta reale) sono dette **univariate**, con supporto in \mathbb{R}^2 sono dette **bivariate** mentre v.a. aventi supporto in \mathbb{R}^k sono dette **multiple** (o k -variate). Ad esempio, la v.a. (X, Y) è bivariata ed il suo supporto $S_{XY} = S_X \times S_Y$ è la *griglia* formata dal prodotto cartesiano dei supporti delle v.a. componenti dove, a ciascun elemento della griglia, è associato una probabilità congiunta $\mathbb{Q}(X = x, Y = y)$.

2.3.1 Esempi di costruzione di variabili aleatorie discrete

Di seguito alcuni esempi notevoli di costruzione di v.a. discrete.

1. Numero di teste in tre lanci di monete indipendenti

Sia $\Omega = \{ttt, ttc, tct, ctt, ccc, cct, ctc, tcc\}$, con $\#\Omega = 2^3$, lo spazio campionario dell'esperimento dotato di $\mathbb{P}()$ secondo l'assegnazione classica e $X :=$ "numero di

²In questo senso, studiare il processo misurativo richiede la possibilità di utilizzare un modello probabilistico noto per rappresentare il suo comportamento aleatorio. Ad esempio, il fenomeno dell'*errore casuale* è rappresentato mediante il modello Gaussiano (o Normale).

³L'aggettivo *aleatorio* fa riferimento al processo aleatorio che sottende la variabile aleatoria non alla variabile di per se. Inoltre, il termine *variabile* non indica propriamente una variabile (come siamo soliti pensare) ma una *funzione misurabile*.

teste”, con $S_X = \{0, 1, 2, 3\}$, la v.a. corrispondente. Questa è associata a Ω nel modo seguente:

$$X = 0: \{ccc\}$$

$$X = 1: \{ctc, tcc, cct\}$$

$$X = 2: \{ttc, tct, ctt\}$$

$$X = 3: \{ttt\}$$

Dove, ad esempio, il caso $X = 0$ è interpretato come segue: quando la v.a. assume valore zero l'evento intercettato nello spazio di probabilità originario è “croce-croce-croce”. La v.a. X definisce un nuovo spazio di probabilità (S_X, \mathbb{Q}) in corrispondenza allo spazio originario (Ω, \mathbb{P}) dove $\mathbb{Q}()$ è la misura di probabilità *indotta* dalla v.a. X sugli elementi x di S_x . In questo esempio, la misura indotta - definita ancora come rapporto tra cardinalità di insiemi secondo la definizione classica - è la seguente:

$$\mathbb{Q}(X=0) = 1/8$$

$$\mathbb{Q}(X=1) = 3/8$$

$$\mathbb{Q}(X=2) = 3/8$$

$$\mathbb{Q}(X=3) = 1/8$$

Dove, ad esempio, il caso $\mathbb{Q}(X=0) = 1/8$ è da intendersi nel seguente modo: la probabilità che la v.a. assuma valore zero è pari a $1/8$.

2. Lancio di un dado a sei facce

Sia $\Omega = \{1, \dots, 6\}$, con $\#\Omega = 6$, lo spazio campionario dell'esperimento dotato di $\mathbb{P}()$ secondo l'assegnazione classica e $X :=$ “faccia pari o dispari”, con $S_X = \{0, 1\}$, la v.a. corrispondente. Questa è associata a Ω nel modo seguente:

$$X = 0: \{2, 4, 6\}$$

$$X = 1: \{1, 3, 5\}$$

dove gli interi zero ed uno hanno qui un carattere arbitrario (avremmo potuto scegliere, infatti, qualsivoglia coppia di numeri a denotare gli eventi di X). La probabilità indotta $\mathbb{Q}()$ è in questo caso la seguente: $\mathbb{Q}(X = 0) = \mathbb{Q}(X = 1) = \frac{1}{2}$.

3. Lancio di una moneta con arresto

Sia $\Omega = \{t, c\}$ lo spazio campionario dell'esperimento dotato di $\mathbb{P}()$ secondo l'assegnazione classica, con $\mathbb{P}(t) = \theta$ e $\mathbb{P}(c) = (1 - \theta)$ dove $\theta \in [0, 1]$ è una probabilità (se la moneta non è truccata, ad esempio, $\theta = 0.5$). Definiamo la v.a. $X :=$ “numero di lanci necessari affinché si verifichi T” dove $S_X = \{1, 2, \dots, n\}$ è il supporto di X i cui valori indicano il numero di lanci prima dell'uscita del lato testa t (ad esempio, $X = 2$ indica che l'evento testa si è verificato dopo due lanci della moneta). L'associazione tra X e Ω è come segue:

$$X = 1: \{t\}$$

$$X = 2: \{ct\}$$

$$\begin{aligned}
 X = 3 &: \{cct\} \\
 &\vdots \\
 X = n &: \{ \overbrace{c \dots}^{n-1 \text{ volte}} t \}
 \end{aligned}$$

dove il caso $X = 1 : \{t\}$ indica che t è occorso alla prima estrazione mentre il caso $X = n : \{c, \dots, t\}$ indica che l'evento t si verifica alla n -esima estrazione. La probabilità $\mathbb{Q}()$, tenuto conto che ogni lancio della moneta è indipendente, è definita come segue per ogni valore di S_X :

$$\begin{aligned}
 \mathbb{Q}(X = 1) &= \theta \\
 \mathbb{Q}(X = 2) &= (1 - \theta)\theta \\
 \mathbb{Q}(X = 3) &= (1 - \theta)(1 - \theta)\theta \\
 &\vdots \\
 \mathbb{Q}(X = n) &= (1 - \theta)^{n-1}\theta
 \end{aligned}$$

Utilizzando quest'ultimo esempio, illustriamo come la probabilità di taluni eventi possa essere calcolata ricorrendo all'uso delle v.a. Chiediamoci: qual è la probabilità che esca testa dopo quattro lanci indipendenti, ossia $\mathbb{Q}(X > 4)$? Notiamo che l'evento $X > 4$ definito su S_X è in corrispondenza all'evento $\{ccct\}$ in Ω . Utilizzando il passaggio al complementare per facilitare il calcolo otteniamo quanto segue:

$$\begin{aligned}
 \mathbb{Q}(X > 4) &= 1 - \mathbb{Q}(X \leq 4) \\
 &= 1 - [\theta + \theta(1 - \theta) + \theta(1 - \theta)^2 + \theta(1 - \theta)^3] \\
 &= 1 - (4\theta + 6\theta^2 - 4\theta^3 + \theta^4) \\
 &\dots\text{sviluppando le potenze dei binomi} \\
 &= (1 - \theta)^4
 \end{aligned}$$

Per cui, nell'esempio di una moneta non truccata $\theta = 0.5$, la probabilità cercata è pari a $(1 - 0.5)^4 = 0.0625$.

2.3.2 Funzioni di densità discrete e continue, funzione di ripartizione

Una v.a. è completamente determinata dalla sua funzione di densità che può essere discreta (nel caso di v.a. discrete) o continua (nel caso di v.a. continue). La funzione di densità permette di calcolare le probabilità associate ai diversi comportamenti delle v.a. e, di conseguenza, di calcolare le diverse probabilità associate agli esiti aleatori di interesse. Più in generale, per ben identificare una v.a. X , occorre determinare:

- S_X , ovvero l'insieme dei valori (discreto o continuo) che X può assumere. A questo proposito, le v.a. possono essere definite:

1. univariate se $S_X \in \mathbb{R}$
2. bivariate se $S_X \in \mathbb{R}^2$
3. k-variate se $S_X \in \mathbb{R}^k$

- la modalità con cui \mathbb{Q} assegna le probabilità su S_X :

1. **funzione di densità discreta** $p(x)$ per v.a. discrete caratterizzata dalle seguenti proprietà principali:

$$p(X = x) \geq 0$$

$$\sum_{x \in S_X} p(X = x) = 1$$

2. **funzione di densità continua** $f(x)$ per v.a. continue caratterizzata dalle seguenti proprietà:

$$f(x_0 \leq X \leq x_0 + \delta) = \int_{x_0}^{x_0 + \delta} f(x) dx \quad \text{con } \delta > 0$$

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

$$f(x_0) = 0$$

3. **funzione di ripartizione** $F(x)$ per v.a. discrete e continue e permette di calcolare la distribuzione di probabilità di X , allo scopo di ottenere completa conoscenza sul comportamento della variabile stessa. Viene definita anche distribuzione cumulata e corrisponde nel caso discreto a:

$$F(x_0) = \mathbb{Q}(X \leq x_0) = \sum_{\substack{x \leq x_0 \\ x \in S_X}} p(x)$$

mentre nel caso continuo:

$$F(x_0) = \mathbb{Q}(X \leq x_0) = \int_{-\infty}^{x_0} f(x) dx$$

La funzione di ripartizione (CDF) non è decrescente: se $x_1 < x_2$, allora $F(x_1) < F(x_2)$. Infatti, sia con X discreta (utilizzando la densità discreta e sommando le probabilità), sia con X continua (utilizzando la densità continua e integrando la funzione di densità), $F(x)$ cresce o rimane costante al variare dei valori assunti da X . Inoltre, si ha che

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow \infty} F(x) = 1$$

Le figure 2.2-2.4 raffigurano rispettivamente tre esempi di funzioni di densità discrete, continue e di ripartizione.

Un esempio di utilizzo della densità continua per il calcolo della probabilità di eventi è il seguente. Una v.a. X continua ha distribuzione di probabilità $f(x) = 3x^2$ quando $x \in [0, 1]$ e $f(x) = 0$ quando $x \notin [0, 1]$. Valutiamo la probabilità $\mathbb{Q}(0.5 < X < 1)$ corrispondente all'evento $\{0.5 < X < 1\}$ mediante calcolo dell'integrale definito su $f(x)$:

$$\begin{aligned} \int_{0.5}^1 f(x) dx &= \int_{0.5}^1 3x^2 dx = \\ &= 3 \left[\frac{x^3}{3} \right]_{0.5}^1 = 3 \left(\frac{1^3}{3} - \frac{0.5^3}{3} \right) = 0.875 \end{aligned}$$

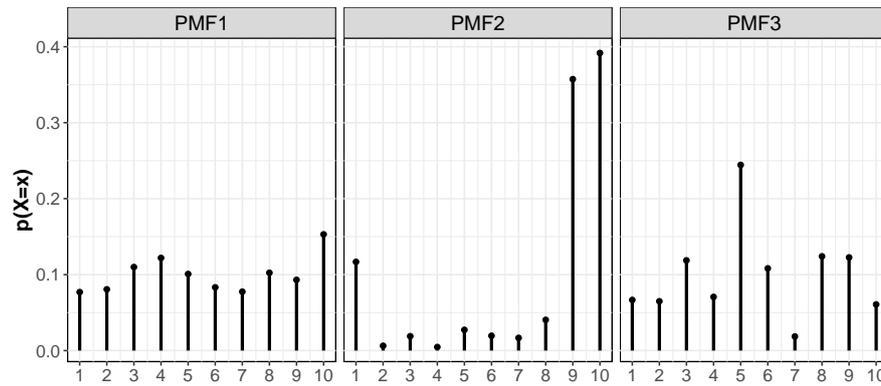


FIGURA 2.2: Densità discreta (PMF): Esempi di tre distribuzioni di probabilità per X discreta.

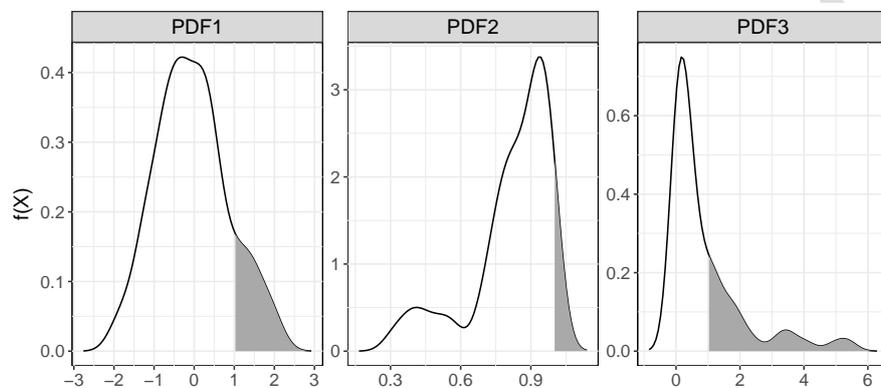


FIGURA 2.3: Densità continua (PDF): Esempi di tre distribuzioni di probabilità per X continua. L'area evidenziata in grigio corrisponde all'intervallo $[x_0, x_0 + \delta]$ con $\delta > 0$.

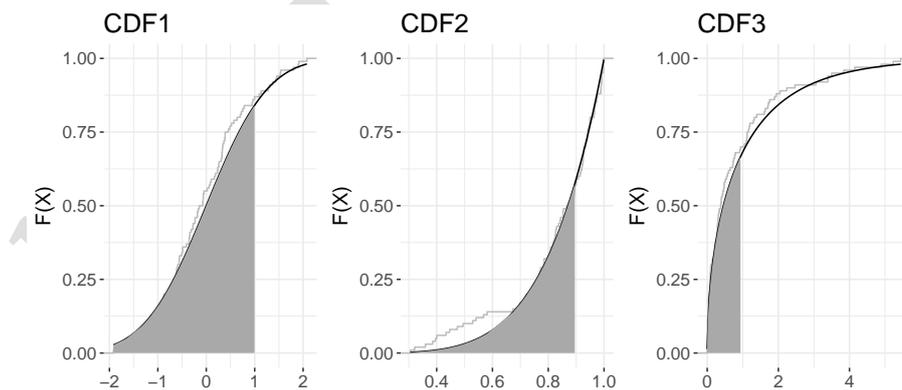


FIGURA 2.4: Funzione di ripartizione (CDF): Esempi di tre funzioni di ripartizione per X continua. L'area evidenziata in grigio corrisponde all'intervallo $[x_0, x_0 + \delta]$ con $\delta > 0$.

La probabilità ricercata è pari a $\mathbb{Q}(0.5 < X < 1) = 0.875$. In generale, il calcolo

di probabilità di eventi mediante densità discrete o continue è effettuato ricorrendo alla funzione di ripartizione $F(X \leq x_0)$, un tempo tabulata in tabelle, oggi ormai disponibile in diversi software tramite semplici funzioni. Alcuni esempi ed applicazioni in R sono disponibili nel file `lab3.R`.

2.3.3 Valori medi di variabili aleatorie

La v.a. X esprime in sintesi un modello matematico di un **esperimento aleatorio**. Il suo comportamento, se è nota $F(X)$, può essere rappresentato da *indici* che esprimono alcune sue caratteristiche, come la posizione e la dispersione definite attraverso la media (o valore atteso) e la varianza della stessa variabile aleatoria.

Più in dettaglio, il **valore atteso** di una v.a. X , denotato come $\mathbb{E}[X]$, indica l'aspettativa che si ha *in media* rispetto agli esiti dell'esperimento aleatorio associato ed è definito nel caso *discreto* da:

$$\mathbb{E}[X] = \sum_{x \in S_X} x p(x)$$

mentre nel caso *continuo*:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Il valore atteso non deve essere confuso con la *media aritmetica* usata in statistica descrittiva: mentre quest'ultima è una sintesi della distribuzione di frequenze di una variabile a livello campionario, il valore atteso è l'*esito medio atteso* a seguito di un esperimento aleatorio. La **varianza** di una v.a. X , denotata come $\mathbb{V}\text{ar}[X]$, indica la dispersione attesa degli esiti (molto dispersi, poco dispersi) dell'esperimento aleatorio associato a X ed è definita nel caso *discreto* da:

$$\mathbb{V}\text{ar}[X] = \sum_{x \in S_X} (X - \mathbb{E}[X])^2 p(x)$$

mentre nel caso *continuo*:

$$\mathbb{V}\text{ar}[X] = \int_{-\infty}^{+\infty} (X - \mathbb{E}[X])^2 f(x) dx$$

Quando siamo in presenza di v.a. almeno doppie è possibile definire la **covarianza** come indice di sintesi del comportamento congiunto di due esperimenti aleatori. Date due v.a. X_1 e X_2 con media e varianza finita $\mathbb{E}[X_1] = \mu_{X_1}$, $\mathbb{E}[X_2] = \mu_{X_2}$ e $\mathbb{V}\text{ar}[X_1] = \sigma_{X_1}^2$, $\mathbb{V}\text{ar}[X_2] = \sigma_{X_2}^2$ la covarianza tra X_1 e X_2 è definita come di seguito:

$$\text{Cov}[X_1, X_2] = \sigma_{X_1 X_2} = \mathbb{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]$$

mentre la **matrice di covarianza** di X è definita da:

$$\text{Cov}[X] = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 \end{bmatrix}$$

dove la diagonale contiene le varianze, i triangoli sotto e sopra la diagonale sono simmetrici e se $X_1 \perp X_2$ (variabili indipendenti) allora $\sigma_{X_1 X_2} = \sigma_{X_2 X_1} = 0$ (la matrice diventa

diagonale). La covarianza è un indicatore che può prendere qualsiasi valore reale. I valori negativi indicano che la covarianza tra le due variabili aleatorie è negativa in quanto al crescere di una decresce anche l'altra, e i valori positivi invece indicano che la covarianza tra le due variabili aleatorie è positiva perché al crescere di una cresce anche l'altra. Nel caso k -variato $X = (X_1, \dots, X_j, \dots, X_k)$, la matrice di covarianza si generalizza come segue:

$$\text{Cov}[X] = \Sigma_X = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1j} & \dots & \sigma_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{j1} & \dots & \sigma_j^2 & \dots & \sigma_{jk} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \dots & \sigma_{kj} & \dots & \sigma_{kk}^2 \end{bmatrix}$$

e conserva le stesse proprietà viste per il caso bivariato. La figura 4.1 mostra una rappresentazione grafica della covarianza per il caso bivariato.

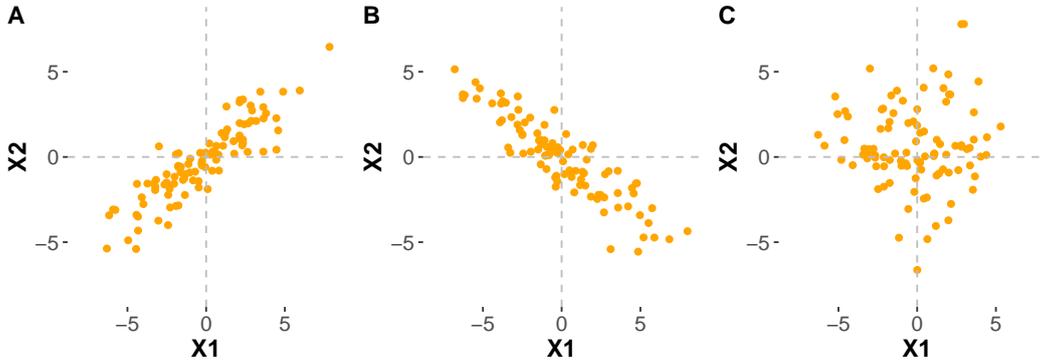


FIGURA 2.5: *Grafico a dispersione bivariato: Rappresentazione grafica dei valori osservati congiuntamente delle variabili X_1 e X_2 . (A): Grafico con associazione lineare positiva $\sigma_{X_1X_2} = 6.003$. (B): Grafico con associazione lineare negativa $\sigma_{X_1X_2} = -6.003$. (C): Grafico con nessuna associazione lineare $\sigma_{X_1X_2} \approx 0$. Nota: in tutti e tre i grafici $\mu_{X_1} = \mu_{X_2} = 0$ e $\sigma_{X_1}^2 = 9.140$ e $\sigma_{X_2}^2 = 5.165$.*

Occorre notare come l'operatore $\text{Cov}[X_1, X_2]$ è lineare: quando tende ad assumere valori prossimi a zero non indica assenza di associazione tra le variabili ma - al contrario - che non vi è *associazione lineare*. La figura 4.2, infatti, mostra pattern di associazione non lineare per i quali gli indici di covarianza sono pari a zero.

Analogamente alla covarianza, è possibile definire l'indice **correlazione** che si ottiene standardizzando la covarianza come segue:

$$\text{Cor}[X_1, X_2] = \rho_{X_1X_2} = \frac{\text{Cov}[X_1, X_2]}{\sqrt{\text{Var}[X_1]}\sqrt{\text{Var}[X_2]}}$$

Come per il caso precedente, la corrispondente **matrice di correlazione** è definita come di seguito:

$$\text{Cor}[X] = P_X = \begin{bmatrix} 1 & \rho_{X_1X_2} \\ \rho_{X_2X_1} & 1 \end{bmatrix}$$

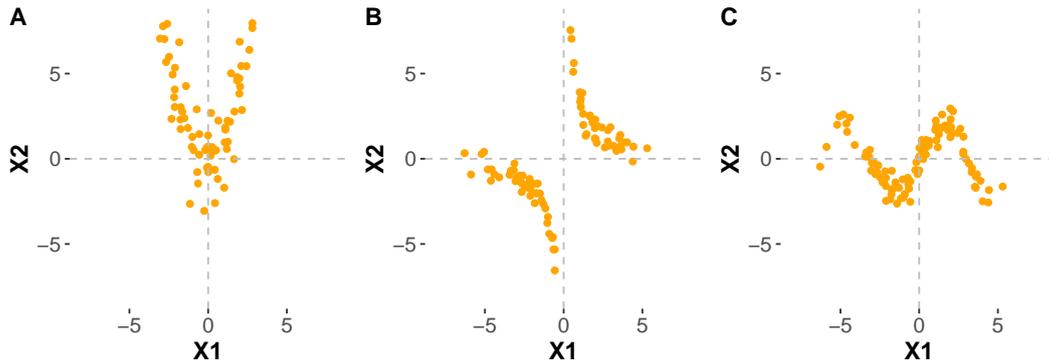


FIGURA 2.6: Grafico a dispersione bivariato: Rappresentazione grafica dei valori osservati congiuntamente delle variabili X_1 e X_2 . (A): Grafico con associazione non lineare (parabola) $\sigma_{X_1X_2} \approx 0$. (B): Grafico con associazione non lineare (iperbole) $\sigma_{X_1X_2} \approx 0$. (C): Grafico con associazione non lineare (sinusoide) $\sigma_{X_1X_2} \approx 0$.

dove la diagonale è unitaria, i triangoli sotto e sopra la diagonale sono simmetrici e se $X_1 \perp X_2$ (variabili indipendenti) allora $\rho_{X_1X_2} = \rho_{X_2X_1} = 0$. La generalizzazione della matrice di correlazione al caso k -variato è immediata. A differenza della covarianza, la correlazione prende valori nell'intervallo reale ristretto $[-1, 1]$ ma conserva la stessa interpretazione vista per la covarianza.

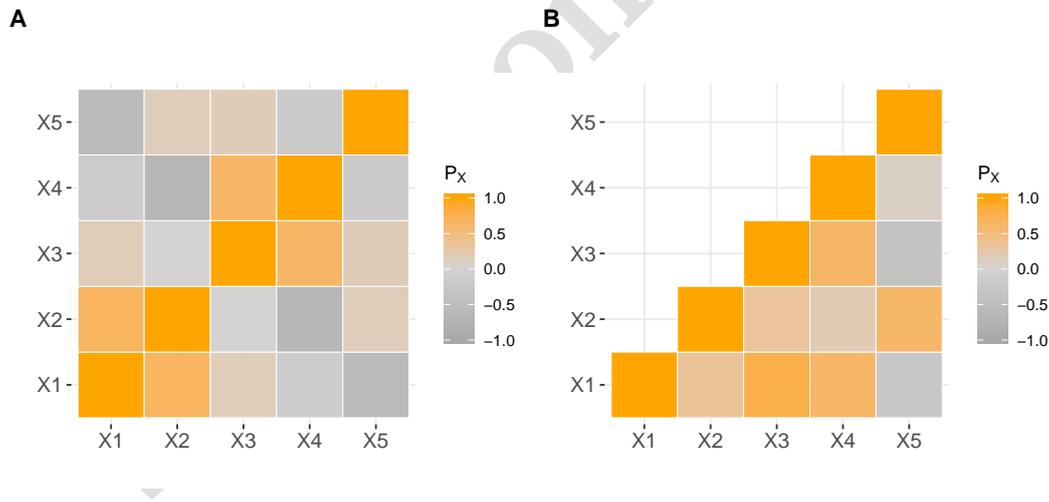


FIGURA 2.7: Grafico della matrice di correlazione: Rappresentazione grafica di una matrice di correlazione \mathbf{P}_X mediante gradiente dei colori: $\rho_{ij} \approx -1$ (grigio scuro), $\rho_{ij} \approx 0$ (grigio chiaro), $\rho_{ij} \approx 1$ (arancione). (A): matrice di correlazione piena (parte triangolare superiore e inferiore). (B): matrice di correlazione senza parte triangolare superiore. Si nota che la diagonale è di colore arancione poiché $\text{diag}[\mathbf{P}_X] = \mathbf{1}$.

2.3.4 Variabili aleatorie standardizzate

Quando una v.a. X ha $\text{Var}[X] = \sigma^2$ e valore atteso $\mathbb{E}[X] = \mu$ finiti, ossia quando esistono ed è possibile calcolarne il valore, la v.a.

$$Z = \frac{X - \mu}{\sqrt{\sigma^2}}$$

è detta **v.a. standardizzata** e si ottiene sottraendo a ciascun esito della v.a. originale X la media attesa μ ed eliminando l'influenza della dispersione di tali esiti σ^2 . La nuova v.a. Z si caratterizza per i seguenti risultati:

$$\mathbb{E}[Z] = \mu = 0 \quad \text{e} \quad \text{Var}[Z] = \sigma^2 = 1$$

Questa tipologia di variabili è sovente utilizzata per rappresentare l'**errore di misura** per il quale si richiede che $\mathbb{E}[X] = \mu = 0$. Ciò significa che dopo un elevato numero di misurazioni ci si attende che in media le misurazioni abbiano errore pari a zero ed una dispersione contenuta.

2.3.5 Disuguaglianze notevoli (approfondimento)

Spesso nel calcolo delle probabilità può essere utile ricorrere ad alcuni risultati importanti che semplificano di gran lunga il calcolo circa eventi di particolare interesse. Alcuni di questi sono noti come disuguaglianze di probabilità. Per meglio comprendere il loro ruolo, introduciamo il tutto con l'esempio seguente.

La soddisfazione settimanale X dei clienti misurata su una scala 0-100 è un v.a. non negativa con valore atteso $\mathbb{E}[X] = 50$. Chiediamoci: qual è la probabilità che la settimana successiva la soddisfazione sia superiore a 75, ossia $\mathbb{Q}(X > 75)$?

Per poter calcolare la probabilità richiesta occorre conoscere qualcosa in più circa X , ad esempio la varianza e/o la sua distribuzione di probabilità. In questo caso non abbiamo ulteriori informazioni oltre al valore atteso della v.a. Un modo per fornire una risposta al quesito è quello di utilizzare la **disuguaglianza di Markov**:

$$\mathbb{Q}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

con $a > 0$, la quale sfrutta la conoscenza del solo valore atteso per fornirci un limite superiore (*upper bound*) alla probabilità richiesta. In questo caso, applicando la definizione

$$\mathbb{Q}(X > 75) \leq \frac{50}{75} \leq 0.667$$

Dunque la probabilità che la soddisfazione sia superiore a 75 è inferiore o al massimo uguale a 0.667.

Supponiamo per un momento di conoscere anche la varianza di X , ossia $\text{Var}[X] = 76.23$. Per rispondere al quesito precedente tenendo anche conto dell'informazione fornita dalla varianza, utilizziamo la **disuguaglianza di Chebyshev**:

$$\mathbb{Q}(|X - \mathbb{E}[X]| > a) \leq \frac{\text{Var}[X]}{a^2}$$

che applicata al nostro caso:

$$\begin{aligned}\mathbb{Q}(X > 75) &= \mathbb{Q}\left([X > 25 + 50] \cup \underbrace{[X < 25 - 50]}_{\text{evento impossibile}}\right) \\ &= \mathbb{Q}(|X - 50| > 25) \leq \frac{76.23}{25^2} \leq 0.121\end{aligned}$$

La probabilità del limite superiore ottenuto con quest'ultima formulazione risulta *più bassa* di quella ottenuta precedentemente. Questo è coerente con il fatto che l'aggiunta di informazione ulteriore sulla v.a. X permette di determinare in maniera più fine il comportamento di quest'ultima.

Proviamo a determinare ora la probabilità che la settimana successiva la soddisfazione sia invece tra 40 e 60, ossia che $\mathbb{Q}(40 < X < 60)$. Utilizzando ancora la disuguaglianza di Chebyshev, nota la varianza di X , scriviamo:

$$\begin{aligned}\mathbb{Q}(40 < X < 60) &= \overbrace{\mathbb{Q}(40 - 50 < (X - 50) < 60 - 50)}^{\text{aggiungiamo/togliamo il baricentro dell'intervallo}} = \mathbb{Q}(-10 < (X - 50) < 10) \\ &= \mathbb{Q}(|X - 50| < 10) \geq 1 - \frac{76.23}{10^2} \geq 1 - 0.762 \geq 0.237\end{aligned}$$

Nell'ultimo passaggio abbiamo utilizzato la definizione alternativa della disuguaglianza:

$$\mathbb{Q}(|X - \mathbb{E}[X]| < a) \geq 1 - \frac{\text{Var}[X]}{a^2}$$

Quindi la probabilità che la soddisfazione nella settimana successiva sia tra 40 e 60 è almeno pari a 0.237. Notiamo come tale probabilità sia abbastanza bassa considerando che $\mathbb{E}[X] = 50$ e stiamo utilizzando un intervallo intorno al valore atteso abbastanza contenuto. La ragione risiede nella varianza di X elevata che fa sì che le realizzazioni casuali della v.a. siano disperse intorno al valore atteso. Infatti, se la varianza fosse stata invece $\text{Var}[X] = 4.21$ avremmo ottenuto $\mathbb{Q}(|X - 50| < 10) \geq 0.957$, un valore ragionevolmente più elevato poiché la dispersione dell'esperimento aleatorio è ora più contenuta. Quest'ultimo risultato ci indica che almeno nel 95% dei casi potremmo avere esiti di X nell'intervallo considerato.

2.3.6 Modelli probabilistici

Diversi esperimenti aleatori, anche se rappresentati con spazi campionari differenti, possono essere rappresentati da medesimi modelli probabilistici (es.: Normale, Binomiale, Gamma). Molti di questi modelli appartengono ad una stessa *famiglia parametrica* di modelli (es.: famiglia esponenziale) il cui comportamento è governato da un insieme di **parametri** a valore reale, di solito indicati genericamente con θ . Ad esempio, l'estrazione di una pallina da un'urna, la scelta di acquisto di un prodotto sono esperimenti aleatori (diversi tra loro) che possono essere rappresentati entrambi con un modello *Binomiale*. La rappresentazione dell'errore di misura legato ad un termostato oppure ad uno strumento psicometrico è fatta invece con un modello *Normale*. In generale,

diremo che una v.a. X (discreta o continua) si distribuisce secondo un certo modello F parametrizzato mediante θ :

$$X \sim F(x; \theta)$$

Alternativamente diremo che X segue F in legge oppure che X è distribuita secondo la legge F , ecc. Con F intenderemo la funzione di ripartizione (che è l'integrale della funzione di densità). Le tabelle 2.1-2.2 riportano alcuni modelli notevoli di v.a. mentre una lista più dettagliata di modelli può essere consultata all'indirizzo https://en.wikipedia.org/wiki/List_of_probability_distributions. La figura 2.8 riporta le densità di probabilità continue per alcuni modelli continui.

Modello	Notazione	S_X	θ	$\mathbb{E}[X]$	$\text{Var}[X]$
Uniforme	$U(x; \alpha, \beta)$	$[\alpha, \beta]$	$-\infty < \alpha < \beta < \infty$	$\frac{1}{2}(a + b)$	$\frac{1}{12}(b - a)^2$
Normale	$N(x; \mu, \sigma^2)$	\mathbb{R}	$\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{\{0\}}^+$	μ	σ^2
Esponenziale	$\text{Exp}(x; \lambda)$	\mathbb{R}^+	$\lambda \in \mathbb{R}^+$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Chi-quadrato	$\chi^2(x; k)$	\mathbb{R}^+	$k \in \mathbb{N}$	k	$2k$
Laplace	$\text{Lapl}(x; \mu, b)$	\mathbb{R}	$\mu \in \mathbb{R}, b \in \mathbb{R}^+$	μ	$2b^2$

TABELLA 2.1: Alcuni modelli probabilistici notevoli per v.a. continue

Modello	Notazione	S_X	θ	$\mathbb{E}[X]$	$\text{Var}[X]$
Binomiale	$\text{Bin}(x; n, \pi)$	$\{0, 1, \dots, n\}$	$n \in \mathbb{N}, \pi \in [0, 1]$	$n\pi$	$n\pi(1 - \pi)$
Poisson	$\text{Poi}(x; \lambda)$	\mathbb{N}	$\lambda \in \mathbb{R}^+$	λ	λ
Geometrica	$G(x; \pi)$	$\{1, \dots, n\}$	$\pi \in [0, 1]$	$\frac{1}{\pi}$	$\frac{1-\pi}{\pi^2}$

TABELLA 2.2: Alcuni modelli probabilistici notevoli per v.a. discrete

2.3.7 Due teoremi limite fondamentali

Nel capitolo 1 abbiamo più volte affermato che ripetere il processo di misurazione un numero molto elevato di volte, nelle medesime condizioni, concorre a diminuire l'errore associato alla misurazione e che la media di tali misurazioni è una buona stima del misurando (all'aumentare infatti del numero di misurazioni la media di queste tende ad avvicinarsi al vero valore da misurare). Alla base di tali risultati vi sono due teoremi fondamentali che caratterizzano il comportamento delle v.a. sotto certe condizioni limite, ad esempio nel caso in cui il numero di prove dell'associato esperimento aleatorio tenda all'infinito.

Legge (debole) dei grandi numeri

Sia X_1, \dots, X_n una *successione* (elenco ordinato) di v.a. indipendenti e identicamente distribuite con $\mathbb{E}[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$ (finiti ed uguali per ogni termine della

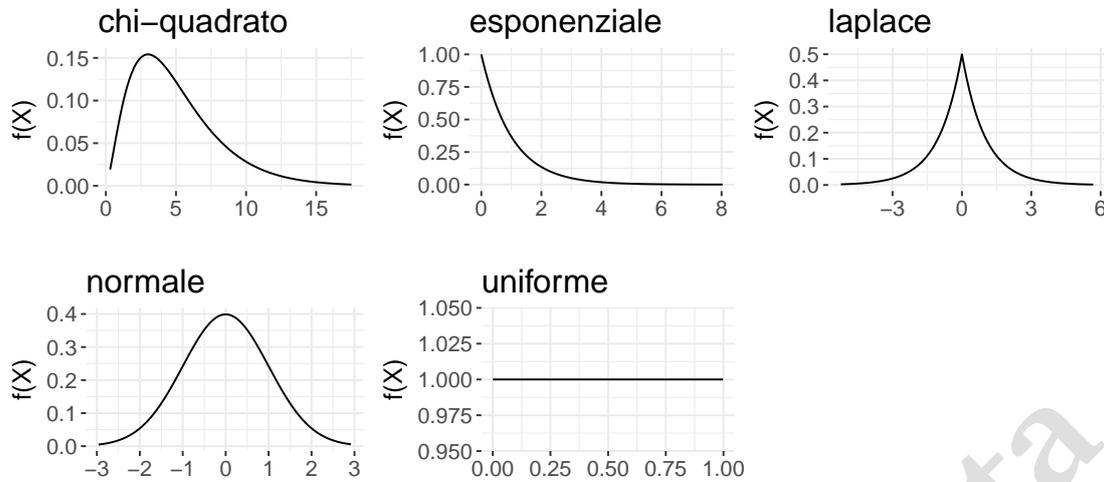


FIGURA 2.8: Densità di probabilità (PDF): alcuni esempi di distribuzioni di probabilità.

successione) e sia

$$\bar{X}_n = \frac{1}{n} \sum_i X_i$$

la v.a. media degli n termini della successione, allora per ogni $\epsilon \in \mathbb{R}^+$ (piccolo a piacere) si ha

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mu - \epsilon < \bar{X}_n < \mu + \epsilon) = 1$$

ossia che al crescere di n la probabilità che la media delle v.a. X_1, \dots, X_n si trovi in un *intorno piccolo* di μ è pari ad 1. In altre parole, ripetendo la misurazione un numero infinito di volte, la media delle misurazioni sarà vicina al *valore vero del misurando* con probabilità crescente e proporzionale a n .

La figura 2.9 rappresenta graficamente il risultato principale della legge debole dei grandi numeri: come possiamo notare, l'intorno $I_{\epsilon,k}$ cresce in lunghezza al crescere di n :

$$\mathbb{P}(\bar{X}_{n=5} \in I_{\epsilon,k}) = 0.162$$

$$\mathbb{P}(\bar{X}_{n=10} \in I_{\epsilon,k}) = 0.228$$

$$\mathbb{P}(\bar{X}_{n=100} \in I_{\epsilon,k}) = 0.614$$

$$\mathbb{P}(\bar{X}_{n=1000} \in I_{\epsilon,k}) = 0.998$$

Teorema del limite centrale

Sia X_1, \dots, X_n una *successione* di v.a. indipendenti e identicamente distribuite con $\mathbb{E}[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$ (finiti ed uguali per ogni termine della successione), sia \bar{X}_n la v.a. media e sia

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

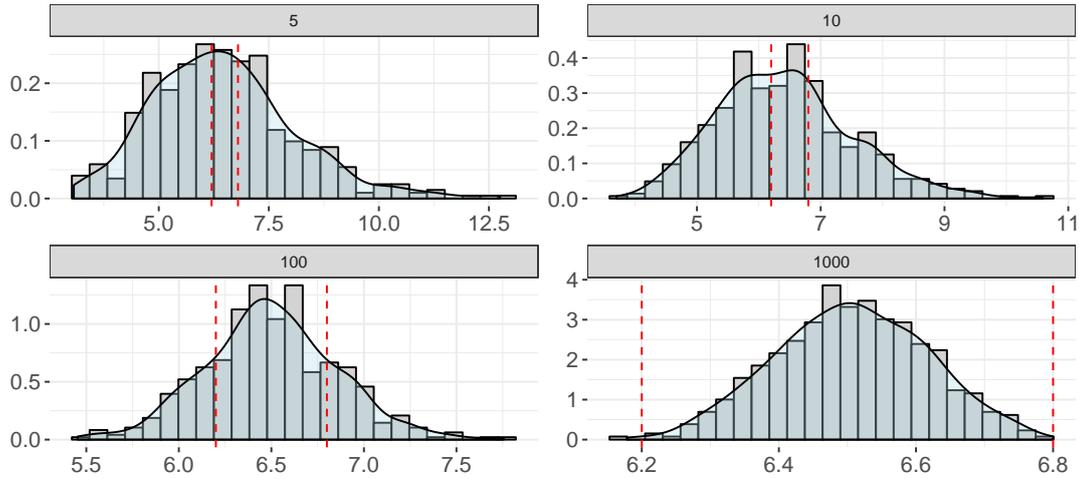


FIGURA 2.9: Legge debole dei grandi numeri: Distribuzione di frequenza e stima della funzione di densità per $n = 5, 10, 100, 1000$. In tutte le prove, $X_i \sim \chi^2(k = 7.5)$ mentre le linee tratteggiate indicano l'intorno $I_{\epsilon, k} = [k \pm \epsilon]$ con $\epsilon = 0.3$.

la v.a. standardizzata con $\mathbb{E}[Z_n] = 0$ e $\text{Var}[Z_n] = 1$. Per ogni $x \in \mathbb{R}$ si ha:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) = \mathbb{P}(Z \leq x)$$

con $Z \sim N(0, 1)$. L'enunciato asserisce che la distribuzione limite Z_n ha sempre la stessa distribuzione (funzione di ripartizione) indipendentemente dalla distribuzione di X_1, \dots, X_n e questa distribuzione è quella Normale con media nulla e varianza unitaria. Vi è una relazione con la legge debole presentata in precedenza. infatti, mentre quest'ultima ha evidenziato come i valori della v.a. \bar{X}_n possono essere vicini o lontani a μ , il teorema del limite centrale è di aiuto nel quantificare tale vicinanza che è nell'ordine di $\mu \pm \frac{\sigma}{\sqrt{n}}$. Le figure 2.10-2.11 mostrano una rappresentazione grafica del risultato del teorema del limite centrale rispetto alla funzione di densità e a quella di ripartizione.

Occorre notare come il teorema del limite centrale presenta moltissime applicazioni: dal calcolo delle probabilità e dall'approssimazione di distribuzioni di probabilità, all'inferenza statistica e alla verifica di ipotesi. Inoltre, costituisce anche la base della **legge degli errori accidentali**, secondo la quale, al crescere del numero n di misurazioni, l'errore di misura si distribuisce come una normale centrata sullo zero.

2.4 Note di sintesi

- (i) La misurazione è un processo aleatorio
- (ii) Il calcolo delle probabilità, la teoria delle variabili aleatorie, l'inferenza statistica[†] sono elementi di base per lo studio della misurazione
- (iii) Il misurando (misurazione diretta) o gli input che costituiscono il misurando (misurazione indiretta) sono variabili aleatorie dotate di distribuzione di probabilità

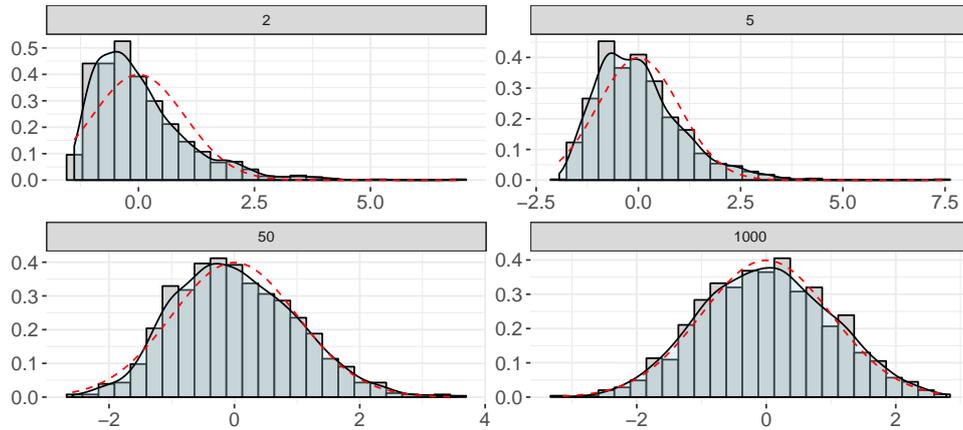


FIGURA 2.10: Teorema del limite centrale: Distribuzione di frequenza e stima della funzione di densità per $n = 2, 5, 50, 1000$. In tutte le prove, $X_i \sim \text{Exp}(\lambda = 1.75)$. Come possiamo notare al crescere di n , la densità Z_n tende in distribuzione alla normale standardizzata $N(0, 1)$ rappresentata dalla curva rossa tratteggiata.

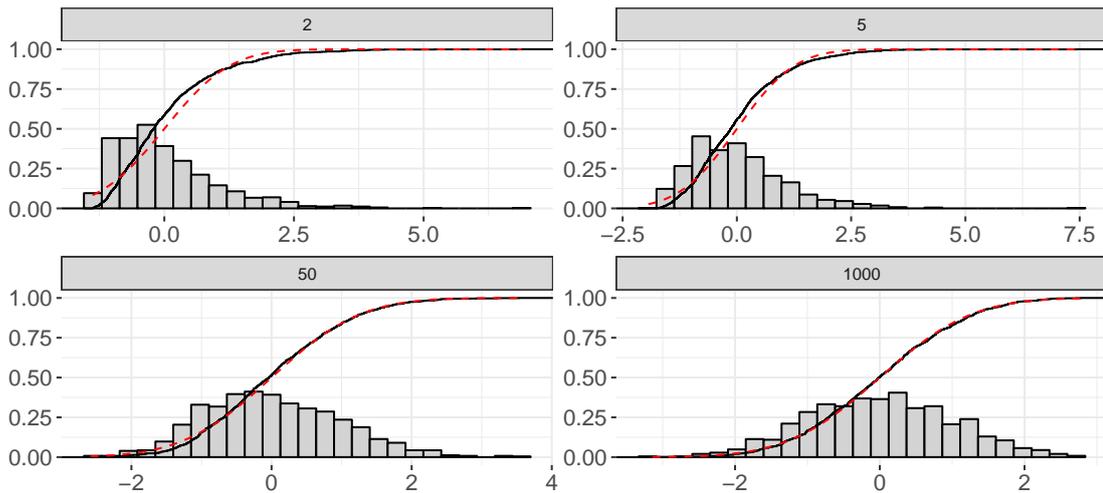


FIGURA 2.11: Teorema del limite centrale: Distribuzione di frequenza e stima della funzione di ripartizione per $n = 2, 5, 50, 1000$. In tutte le prove, $X_i \sim \text{Exp}(\lambda = 1.75)$. Come possiamo notare al crescere di n , la funzione di ripartizione Z_n tende in distribuzione alla funzione di ripartizione normale standardizzata rappresentata dalla curva rossa tratteggiata.

- (iv) Misurare significa estrarre da una popolazione X un campione di misure $\mathbf{x} = (x_1, \dots, x_n)$ (istanze) attraverso le quali studiare X (inferenza statistica)
- (v) Misurare significa studiare la legge (probabilistica) F che governa il misurando $\eta \sim F(x; \boldsymbol{\theta})$
- (vi) Nella valutazione dell'incertezza (teoria degli errori, incertezza tipo A) è centrale il ruolo giocato dal teorema del limite centrale e dalla legge dei grandi numeri,

nonché dalle condizioni di ripetibilità degli esperimenti aleatori

2.5 Laboratorio in R

I laboratori associati a questo capitolo sono i seguenti:

- o **lab3.R**: (A) Calcolo delle probabilità con approccio classico, (B) Calcolo delle probabilità con approccio frequentista, (C) Calcolo delle probabilità con alcuni modelli notevoli
- o **lab4.R**: (A) Introduzione ai metodi Monte Carlo, (B) Calcolo di statistiche mediante Monte Carlo, (C) Propagazione dell'incertezza via Monte Carlo

Il loro contenuto può essere visualizzato e scaricato all'indirizzo <https://github.com/antcalcagni/psychological-testing/tree/master/labs>.

bozza non corretta

3

La misurazione in psicologia

Dopo aver affrontato la misurazione dal punto di vista della teoria degli errori casuali e della teoria GUM alla gestione dell'incertezza (capitolo primo) e aver richiamato alcuni di quelli che sono i fondamenti statistici alla base della misurazione (capitolo secondo), verrà ora dato spazio ad una delle principali teorie presenti nella misurazione in psicologia, ossia la teoria classica dei test. Misurare in psicologia (e nelle scienze sociali e comportamentali in generale) ha un significato non sempre chiaro ed è per questo che il dibattito negli anni ha assunto sempre più una prospettiva pragmatica, incentrata questa sul *come* si misura, a scapito della riflessione ontologica, incentrata quest'ultima sul *cosa* si misura. Quest'ultimo aspetto è noto come il *problema della misurabilità* in psicologia [24]. Per una trattazione più approfondita su tali questioni - che non verranno trattate in questo capitolo - si veda, ad esempio, [25, 26]. Una trattazione dei principali argomenti propri della ricerca psicometrica è presente in [30]. Una lettura italiana utile nel fornire un'agevole introduzione ai principali modelli teorici per la misurazione in psicologia è certamente [2]. Il lettore interessato alla fondazione statistica di molti modelli usati per la misurazione in psicologia potrà infine consultare [20]. Anche se non affrontato in questo corso, il problema della misurabilità e della sua fattiva *realizzabilità* riveste tuttavia un ruolo centrale nel progresso della psicologia all'interno delle discipline scientifiche. Affrontare il problema della misurazione in psicologia, anche al costo di ridimensionare la pretesa scientifica di alcuni risultati raggiunti, è un nodo cruciale per l'intera costruzione della psicologia come scienza, nonché del suo dialogo (comprensibile) con le altre discipline. La misurazione, difatti, investe tutto il problema del metodo scientifico, della sua applicabilità e dei suoi limiti.

3.1 Livelli di misurazione

Rispetto a quanto avviene nelle discipline fisiche, in quelle psicosociali la misurazione è un processo che spesso risulta meno preciso e più inficiato da errore. Questo perché nelle prime discipline la misurazione riguarda sostanzialmente variabili di tipo fisico diretta-

Hanno contribuito alla realizzazione del presente capitolo:

Yoyce Pellizzer, Anna Frena, Giulia Zaupa, Maria Camilot, Anita Zanolla, Michelle Lorenzin e Alice Bettamin, Maria Lucia Cavallo, Rachele Mencuccini, Nicolò Turco, Chiara Grimoldi, Ambra Perugini

mente misurabili, mentre nelle seconde l'oggetto di misurazione è spesso rappresentato da caratteristiche che non sono direttamente osservabili ma possono essere solo inferite dai comportamenti osservabili. Si deve a Stevens [34] la definizione di differenti regole di corrispondenza a seconda delle caratteristiche della variabile oggetto di studio; queste hanno consentito di definire quattro scale, ciascuna delle quali associata ad un particolare livello di misurazione. Queste possono essere ordinate gerarchicamente secondo un livello crescente di complessità, dove ogni nuova scala di misurazione incorpora al suo interno le proprietà della precedente e ne aggiunge di altre. Il primo livello di misurazione è rappresentato dalla **scale nominale**, nella quale l'unica relazione che è possibile stabilire tra gli elementi di un sistema empirico è la categorizzazione. Le categorie di una variabile misurata su scala nominale possono essere confrontate solo in termini di uguaglianza-differenza. Spesso tale livello viene definito di tipo qualitativo. Il secondo livello di misurazione è rappresentato dalla **scala ordinale**, la cui proprietà aggiuntiva rispetto alla precedente è che le variabili, oltre a essere categorizzate, sono anche ordinate in modo crescente o decrescente rispetto al grado della caratteristica considerata. Nella scala ordinale la regola di corrispondenza utilizzata per associare i numeri alle proprietà di oggetti/eventi è che i numeri vengono assegnati in modo da riflettere l'ordine di graduatoria tra oggetti/eventi rispetto alla caratteristica in esame. Il terzo livello è quello delle **scale a intervalli equivalenti** nella quale, oltre alle proprietà definite per le due scale precedenti, è definibile un'unità di misura, ovvero si può quantificare la presenza o assenza di una caratteristica e il grado con cui essa è presente. La regola per l'assegnazione di valori numerici a oggetti/eventi è che la differenza tra i valori numerici assegnati deve corrispondere alla differenza tra i livelli della caratteristica posseduta da soggetti o eventi. Quarto livello è la **scala a rapporti equivalenti** nella quale l'unica proprietà aggiuntiva rispetto alle precedenti è la presenza di un elemento d'intensità nulla, che costituisce un punto di origine naturale per la scala (ovvero uno zero assoluto). Alcuni importanti concetti che caratterizzano le scale di misura sono quelli di **trasformazioni ammissibili** e di **significanza e invarianza delle statistiche**. Con il concetto di trasformazione ammissibile si fa riferimento a quelle funzioni che consentono il passaggio da una scala di misura ad un'altra definite sullo stesso sistema empirico. Per significanza di una statistica si intende invece l'ammissibilità di questa statistica rispetto alla scale di misurazione della variabile. A differenza del termine *misurazione*, quello di *rilevazione* può essere utilizzato soprattutto quando si ha come unità di analisi non una proprietà direttamente osservabile ma una variabile latente. A questo riguardo, con il termine *scaling* si può intendere il processo di rilevazione di una proprietà latente. Lo scaling può essere unidimensionale e in tale caso l'obiettivo è quello di rilevare un solo attributo latente. Può essere anche multidimensionale e in tale caso l'obiettivo è di rilevare due o più costrutti latenti. Il processo di scaling consente di differenziare e di ordinare lungo un continuum oggetti/eventi in relazione ad una caratteristica che è stata rilevata.

3.2 Lessico psicometrico

Solitamente il ricercatore non è interessato al punteggio del soggetto nel singolo item (o quesito), quanto al punteggio complessivo ottenuto su tutti gli item utilizzati per misurare uno specifico costrutto. Utilizzare più di un item presenta degli indubbi vantaggi sia

per quanto riguarda l'attendibilità sia per quanto riguarda la validità del test. E' bene quindi distinguere tra due livelli fondamentali negli strumenti che misurano caratteristiche psicologiche: gli item, i test e le scale. Un test rappresenta un procedimento tramite il quale è possibile ottenere una misurazione di una caratteristica psicologica. In un test psicologico un campione del comportamento di un soggetto viene osservato e descritto attraverso una procedura standardizzata che da origine ad una certa quantificazione. Gli item sono gli elementi minimali di un test psicologico. E' possibile distinguere tra diverse tipologie di item a seconda della caratteristica psicologica che viene misurata. Se si tratta di caratteristiche come la personalità o gli atteggiamenti, tali item non prevedono una risposta corretta, ma danno al rispondente la possibilità di riportare il proprio punto di vista. Oltre ad item di contenuto verbale è possibile avere anche item che presentano stimoli non verbali. Se si tratta invece di caratteristiche cognitive come le attitudini e l'intelligenza, solitamente l'item è costituito da un quesito che prevede più alternative di risposta delle quali solo una è quella corretta. Gli item vengono combinati tra loro per ottenere un punteggio complessivo che riflette la presenza di quella determinata caratteristica oggetto di studio nel soggetto cui è stato somministrato il test.

In sintesi, dunque, un **test** è uno strumento di misura usato per quantificare oggetti di interesse psicologico non osservabili (costrutti latenti, ovvero i *misurandi*). Generalmente è caratterizzato da un insieme di **item** (o domande/indicatori) che possono essere aggregati insieme per formare una **scala**. Tale scala è utilizzata per quantificare un **costrutto latente** ed è un *proxy* di quest'ultimo. Un test è solitamente formato da più scale, spesso correlate tra loro. Il numero di costrutti latenti, individuato dalle scale, è detto **dimensione di un test**. La tabella 3.1 riporta alcuni termini di raccordo tra il lessico metrologico utilizzato nel primo capitolo e quello psicometrico di solito utilizzato nella ricerca psicologica.

metrologia	psicometria
misurando	costrutto/tratto latente
misura	scala (se aggregata)
misura	misura osservabile (non aggregata)
misura	item/indicatore
precisione di una misura	attendibilità di una scala
misurazione indiretta	misure composite
strumento di misura	test

TABELLA 3.1: Raccordo tra terminologia psicometrica e metrologica

3.3 Modelli per la misurazione in psicologia

Diversi sono i modelli utilizzati all'interno dell'approccio psicometrico alla misurazione in psicologia. Tra questi, ad esempio, la *teoria classica dei test*, la *teoria della risposta all'item*, la *teoria della generalizzabilità*, la *teoria dello scaling geometrico*, ecc. In questo capitolo verrà affrontato solo uno di questi modelli, ossia la teoria classica dei test. Tale

scelta è dovuta a ragioni di ordine storico (è stato il primo modello ad essere formalizzato) e pragmatico (è il modello attualmente più utilizzato nella costruzione dei test). Gli altri modelli, soprattutto quello legato alla teoria alla risposta all'item (IRT), richiederebbero per certi versi una comprensione degli aspetti di base facilmente rintracciabili nella teoria classica dei test. L'esposizione della teoria classica dei test seguirà la formalizzazione presente in [2, 20].

bozza non corretta

4

La teoria classica dei test

La teoria classica dei test (TCT), conosciuta anche come teoria dell'errore causale, modello classico o come modello del punteggio vero, è il modello standard utilizzato per la costruzione e la valutazione dei test. La TCT è ampiamente basata sulla teoria degli errori accidentali presentata nel modulo A ed ha diverse connessioni con l'approccio GUM alla gestione dell'incertezza. La prima formulazione della TCT si deve a Spearman [35] che postula la decomposizione del punteggio osservato di un test X in due diverse componenti, ossia il punteggio vero T e l'errore casuale E , legati mediante una relazione lineare e additiva. Ciò che sta al di sotto di tale decomposizione è l'idea che in ogni misurazione sia presente una componente d'errore, quindi il punteggio osservato non coincide con quello vero, anzi, se si ripetesse la rilevazione infinite volte, si potrebbe ottenere ogni volta un punteggio diverso. Quindi il punteggio osservato X è solo uno dei possibili punteggi che il soggetto può ottenere. Il modello sottolinea come l'errore sia intrinseco nel processo stesso di misurazione e di come quindi, tutte le volte che si effettua una misurazione esista una dimensione di incertezza sul risultato finale dell'operazione che non può essere eliminata del tutto. La riduzione dell'incertezza che si commette nel processo di misurazione è fondamentale, in quanto solo quando l'errore commesso è contenuto, le misure ottenute consentono di prevenire a conclusioni significative e sensate. L'idea che ogni misurazione sia inficiata da una componente di errore, ma che questo errore sia casuale e non sistematico, è fondamentale, poiché le assunzioni della TCT si sviluppano a partire da questo concetto. Il modello teorico alla base della TCT ipotizza che un test venga somministrato per un numero elevato n di volte allo stesso soggetto. A partire da questo modello teorico, il punteggio osservato può essere considerato una variabile che assume dei valori secondo una distribuzione di probabilità contenente il punteggio vero. Un determinato punteggio riportato da un soggetto al test può allora essere considerato come un campione proveniente da una popolazione n di punteggi possibili, popolazione che contiene il punteggio vero e che assume una forma normale. Il punteggio vero rappresenta il valore atteso della distribuzione di probabilità associata ai punteggi osservati.

Hanno contribuito alla realizzazione del presente capitolo:

Yoyce Pellizzer, Anna Frena, Giulia Zaupa, Maria Camilot, Anita Zanolla, Michelle Lorenzin e Alice Bettamin, Maria Lucia Cavallo, Rachele Mencuccini, Nicolò Turco, Chiara Grimoldi, Ambra Perugini

4.1 Definizioni e proprietà iniziali

Consideriamo un insieme di misure (items) indicizzate tramite $j = 1, \dots, p$ e un insieme di individui (persone) $i = 1, \dots, n$ (estratti casualmente da una popolazione \mathcal{P}) su cui le misurazioni sono effettuate usando un determinato test. Alla j -esima misurazione è associata la realizzazione della variabile aleatoria (X_j, T_j, E_j) i cui esiti sono ottenuti campionando su \mathcal{P} . I termini della tripla sono così definiti: X_j è la v.a. che governa la realizzazione della j -esima misura *osservabile*, T_j è la v.a. che governa la realizzazione della j -esima misura *vera* mentre E_j è la v.a. che governa la realizzazione della j -esima componente di *errore*. Della tripla (X_j, T_j, E_j) conosciamo solo X_j mentre le altre componenti sono in relazione a questa attraverso la definizione fondamentale della TCT:

$$X_j = T_j + E_j$$

Le seguenti proprietà valgono per la tripla considerata:

- (i) $\mathbb{E}[E_j] = 0$
- (ii) $\text{Cor}[E_j, T_j] = 0$
- (iii) $\text{Cor}[E_j, E_{j'}] = 0$

dove (i) indica che l'errore ha carattere casuale e non sistematico, (ii) che l'errore di misura e il misurando latente sono incorrelati, (iii) che due misurazioni differenti j e j' hanno errori incorrelati. La proprietà (ii) è necessaria affinché si possa quantificare T_j attraverso X_j , vale a dire che si possa isolare dalla varianza osservata di X_j la componente di errore E_j da quella latente e vera T_j sottesa all'osservato. La proprietà (iii) invece è richiesta per la sola semplificazione del modello rispetto ai calcoli successivi che verranno richiesti. Dato che X_j è una v.a. e considerati gli assunti (i)-(iii) appena enucleati, è possibile caratterizzare il comportamento di X_j utilizzando i suoi primi due momenti, ossia valore atteso e varianza (vedi Capitolo 2). Per cui:

$$(iv) \mathbb{E}[X_j] = \mathbb{E}[T_j + E_j] = \mathbb{E}[T_j] + \mathbb{E}[E_j] \xrightarrow{0}$$

$$(v) \text{Var}[X_j] = \text{Var}[T_j + E_j] = \text{Var}[T_j] + \text{Var}[E_j] + 2\text{Cov}[T_j, E_j] \xrightarrow{0}$$

dove la proprietà (iv) è derivata grazie alla linearità del valore atteso¹ mentre la (v) è ottenuta applicando la definizione di varianza di una combinazione lineare². In generale, da quanto appreso nel Capitolo 2 (paragrafo 2.3.3), covarianza e correlazione sono in relazione tra loro secondo la relazione nota³, ragione per cui nel proseguo scriveremo spesso l'una in luogo dell'altra e viceversa. In termini qualitativi, la proprietà (iv) ci consente di derivare il valore vero del misurando (non osservabile) T_j attraverso l'osservabile X_j mentre la (v) ci offre un modello lineare e additivo per studiare la varianza degli item osservati X_j in termini di una componente vera ed una di errore. La proprietà (v) sarà utile per definire uno degli obiettivi della TCT, vale a dire quello di minimizzare

¹ $\mathbb{E}[A + B] = \mathbb{E}[A] + \mathbb{E}[B]$

² $\text{Var}[A + B] = \text{Var}[A] + \text{Var}[B] + 2\text{Cov}[A, B]$

³ $\text{Cov}[A, B] = \text{Cor}[A, B] \sqrt{\text{Var}[A]} \sqrt{\text{Var}[B]}$

la componente di varianza d'errore $\mathbb{V}\text{ar}[E]$ e di massimizzare la varianza del misurando $\mathbb{V}\text{ar}[T]$. Difatti, un test con una tale proprietà sarebbe in grado di isolare bene la componente d'errore da quella vera che si vuole quantificare.

Allo stadio attuale di presentazione della TCT non possediamo ancora una formulazione per definire la varianza vera $\mathbb{V}\text{ar}[T_j]$ e quella d'errore $\mathbb{V}\text{ar}[E_j]$ sebbene sappiamo che tali componenti siano in relazione tra loro e che siano presenti nella variabilità osservata $\mathbb{V}\text{ar}[X_j]$. Per poter offrire una definizione di tali varianze occorre ricorrere a quelle che sono definite *misure parallele*.

4.2 Misure parallele e non parallele

In termini qualitativi, due misure (o test) sono parallele quando sono *equivalenti* nel senso che una può essere utilizzata al posto dell'altra e viceversa. Ciò è dovuto al fatto che tali misure quantificano lo stesso misurando o costrutto latente. Affinché due misure siano dette parallele è necessario che siano soddisfatte talune proprietà essenziali, violate le quali si indebolisce progressivamente l'equivalenza tra le misure fino a raggiungere un livello in cui il confronto tra le due può avvenire solo in alcuni contesti particolari. In generale, sono quattro i livelli di equivalenza tra due misure, il primo che esprime l'equivalenza forte (misure parallele) e gli altri tre che esprimono invece una equivalenza debole (misure non parallele). Nel seguito, daremo le definizioni di tali livelli di equivalenza e i vincoli sul loro utilizzo.

Consideriamo una misurazione effettuata utilizzando due distinte osservabili X_j e $X_{j'}$ per un medesimo misurando (o costrutto) T^4 . Le osservabili X_j e $X_{j'}$ sono dette **parallele** (p_1) quando:

$$\begin{aligned} X_j &= T + E \\ X_{j'} &= T + E \end{aligned}$$

ossia X_j e $X_{j'}$ misurano la stessa quantità latente T con la stessa quantità di errore E . Si ha infatti che $\mathbb{E}[X_j] = \mathbb{E}[X_{j'}]$ e $\mathbb{V}\text{ar}[X_j] = \mathbb{V}\text{ar}[X_{j'}]$.

Le osservabili X_j e $X_{j'}$ sono dette **τ -equivalenti** (p_2) quando:

$$\begin{aligned} X_j &= T + E_j \\ X_{j'} &= T + E_{j'} \end{aligned}$$

ossia X_j e $X_{j'}$ misurano la stessa quantità latente T con diverse quantità di errore E_j e $E_{j'}$. Si ha inoltre che $\mathbb{E}[X_j] = \mathbb{E}[X_{j'}]$ e $\mathbb{V}\text{ar}[X_j] \neq \mathbb{V}\text{ar}[X_{j'}]$. È possibile comprendere (p_1) alla luce di (p_2): l'equivalenza forte richiede infatti che X_j e $X_{j'}$ abbiano stesso valore atteso (pari a T) e stesse varianze d'errore. Ad esempio, se $X_j \sim N(0, \sigma^2)$ e $X_{j'} \sim N(0, \sigma^2)$ abbiamo che entrambe le varianze d'errore devono essere uguali affinché le due misure siano parallele. Ciò non indica che gli errori osservati a livello campionario

⁴Usando la definizione (iv) della TCT ricordiamo che $\mathbb{E}[X_j] = \mathbb{E}[X_{j'}] = \mathbb{E}[T] = \tau$, il valore vero del misurando è il valore atteso delle misurazioni fatte mediante le osservabili X_j e $X_{j'}$

debbano necessariamente essere gli stessi ma che le *leggi probabilistiche generatrici* delle componenti di errore siano le stesse. Un primo indebolimento di tale richiesta è fatta in (p_2) dove si ammettono che le varianze di errore possano essere differenti $\sigma_j^2 \neq \sigma_{j'}^2$. Il risultato immediato di tale richiesta è che alle due misure non si richiede di essere precise allo stesso modo - una misura può infatti essere più variabile dell'altra - sebbene entrambe debbano quantificare lo stesso misurando T .

Le osservabili X_j e $X_{j'}$ sono dette **essenzialmente τ -equivalenti** (p_3) quando:

$$\begin{aligned} X_j &= \alpha + T + E_j \\ X_{j'} &= \alpha + T + E_{j'} \end{aligned}$$

ossia X_j e $X_{j'}$ quantificano la stessa quantità latente T a meno di una costante reale α con diversa precisione E_j e $E_{j'}$. Si ha infatti che $\mathbb{E}[X_j] = \alpha + \mathbb{E}[X_{j'}]$ e $\text{Var}[X_j] \neq \text{Var}[X_{j'}]$. Le proprietà richieste dalla (p_3) mostrano che il valore atteso di X_j e $X_{j'}$ non sarà esattamente il valore atteso di T , piuttosto una traslazione di quest'ultimo causata dalla costante α . La conseguenza di ciò risiede nel fatto che la variabile osservata non misurerà esattamente il misurando latente. L'indebolimento della (p_2) comporta inoltre che le medie dei punteggi osservati dei test siano differenti. Per comprendere al meglio quanto espresso all'interno della (p_3) , si pensi alla situazione descritta da [15] in cui X_j e $X_{j'}$ siano due item rispettivamente con il seguente contenuto semantico: «mi sento triste qualche volta» e «mi sento quasi sempre triste». Rispetto alla parte semantica, essi cercano di quantificare il medesimo costrutto. Tuttavia, poiché scritti in maniera non del tutto analoga, è possibile che i rispondenti ai due item li interpretino in maniera differente e, conseguentemente, rispondano in maniera altresì diversa. Da ciò deriva il fatto che il grado con cui essi sono associati al costrutto possa non essere lo stesso. Per cui, i due item sono costruiti per quantificare lo stesso misurando ma l'uno non coincide con l'altro: tale non coincidenza può essere rappresentata formalmente dall'introduzione di una costante α che fa sì che l'uno risulti *traslato* rispetto all'altro.

Infine, le osservabili X_j e $X_{j'}$ sono dette **congeneriche** (p_4) quando:

$$\begin{aligned} X_j &= \alpha + T\beta + E_j \\ X_{j'} &= \alpha + T + E_{j'} \end{aligned}$$

ossia X_j e $X_{j'}$ misurano la stessa quantità latente T con diverso grado di associazione, diverse scale e diverse quantità di errore. Si ha infatti che $\mathbb{E}[X_j] = \alpha + \mathbb{E}[X_{j'}]\beta$ e $\text{Var}[X_j] \neq \text{Var}[X_{j'}]$. Con (p_4) siamo di fronte ad un ulteriore indebolimento del modello delle misure parallele: ciò che ne consegue è un modello psicométrico con meno vincoli, in cui le varianze di errore e le medie dei punteggi osservati dei test sono differenti. La tabella 4.1 riporta in sintesi i quattro livelli di equivalenza sinora esposti unitamente all'interpretazione qualitativa che ne consegue.

Si noti che un modo per valutare empiricamente quale equivalenza sussiste tra test o misure differenti è quella di condurre un'analisi della dimensionalità dei due test mediante *analisi fattoriale confermativa* (CFA) e di valutare se i due modelli fattoriali sono *invarianti* o meno (si veda Capitolo 4).

misure	interpretazione
(p_1) parallele	i test o le misure hanno lo stesso punteggio vero per tutte le persone che è misurato allo stesso modo da entrambe le misure
(p_2) τ -equivalenti	i test o le misure hanno lo stesso punteggio vero per tutti i soggetti ma questo non è misurato allo stesso modo dalle due misure (varianze diverse)
(p_3) essenzialmente τ -equivalenti	i test o le misure hanno lo stesso punteggio vero per tutti i soggetti a meno di una costante α (traslazione), unitamente a varianze differenti
(p_4) congeneriche	i test o le misure sono in relazione lineare tra loro ed hanno varianze differenti

TABELLA 4.1: *Classificazione dell'equivalenza di due test o misure. Nota: le proprietà (p_1)-(p_4) sono ordinate dalla più forte (p_1) alla più debole (p_4)*

4.3 Determinare $\text{Var}[T]$ e $\text{Var}[E]$

Abbiamo affermato come le quantità ignote $\text{Var}[T]$ e $\text{Var}[E]$, alla base della decomposizione della variabilità osservata $\text{Var}[X]$, siano determinabili utilizzando misure parallele, definite nei termini di quanto esposto nel paragrafo precedente. Dunque, siano X_j e $X_{j'}$ due misure (p_1)-parallele e consideriamo la loro correlazione:

$$\begin{aligned}\text{Cor}[X_j, X_{j'}] &= \frac{\text{Cov}[X_j, X_{j'}]}{\sqrt{\text{Var}[E_j]}\sqrt{\text{Var}[E_{j'}]}} \\ &= \frac{\text{Var}[T]}{\text{Var}[X_j]}\end{aligned}$$

dove l'ultimo risultato, ottenuto sviluppando $\text{Cov}[X_j, X_{j'}]$ secondo la definizione data nel Capitolo 2 ed utilizzando gli assunti (ii)-(iii) della TCT, indica che la correlazione (o covarianza) fra le misure parallele è uguale al rapporto tra la varianza della parte vera e la varianza della misura osservata. Ri-arrangiando i termini rispetto a $\text{Var}[T]$ otteniamo il risultato cercato:

$$(vi) \quad \text{Var}[T] = \text{Cor}[X_j, X_{j'}] \text{Var}[X_j]$$

secondo cui la varianza del misurando T è uguale al prodotto tra la varianza della misura X_j (oppure $X_{j'}$ dal momento che le misure sono parallele) e la correlazione tra X_j e $X_{j'}$. La varianza ignota di T è ora espressa in funzione *esclusivamente* di ciò che osserviamo, vale a dire $X = (X_j, X_{j'})$. Quindi possiamo definire la varianza del misurando non osservato - che vogliamo quantificare attraverso X - come il prodotto tra la varianza vera e la correlazione tra le misure utilizzate come *proxy* per T .

Per determinare la quantità $\text{Var}[E_j]$, utilizziamo la definizione (vi) nella (v) come segue:

$$\begin{aligned}\text{Var}[X_j] &= \text{Var}[T_j] + \text{Var}[E_j] \\ &= \text{Cor}[X_j, X_{j'}] \text{Var}[X_j] + \text{Var}[E_j]\end{aligned}$$

e ri-arrangiando i termini:

$$\text{Var}[E_j] = \text{Var}[X_j] - \text{Cor}[X_j, X_{j'}] \text{Var}[X_j]$$

otteniamo infine:

$$(vii) \text{Var}[E_j] = \text{Var}[X_j] (1 - \text{Cor}[X_j, X_{j'}])$$

secondo cui la varianza dell'errore E_j è uguale al prodotto tra la varianza della misura X_j (oppure $X_{j'}$) e uno meno la correlazione tra la coppia X_j e $X_{j'}$. Come per il caso precedente, la varianza ignota di E_j è ora espressa in funzione *esclusivamente* di ciò che osserviamo.

Con (vii) abbiamo determinato completamente definizioni e proprietà della TCT per un test. La tabella 4.2 sintetizza i risultati notevoli finora raggiunti.

$X_j = T + E_j \quad (j = 1, \dots, p)$
(i) $\mathbb{E}[E_j] = 0$
(ii) $\text{Cor}[E_j, T] = 0$
(iii) $\text{Cor}[E_j, E_{j'}] = 0$
(iv) $\mathbb{E}[X_j] = \mathbb{E}[T]$
(v) $\text{Var}[X_j] = \text{Var}[T] + \text{Var}[E_j]$
(vi) $\text{Var}[T] = \text{Cor}[X_j, X_{j'}] \text{Var}[X_j]$
(vii) $\text{Var}[E_j] = \text{Var}[X_j] (1 - \text{Cor}[X_j, X_{j'}])$

TABELLA 4.2: *Sintesi della TCT per un test. Nota: la quantità $\text{Cor}[X_j, X_{j'}]$ è ottenuta utilizzando una coppia di misure (osservabili) parallele nel senso di (p_1) .*

4.4 Validità di una misura

Quando il nostro interesse è rivolto a capire cosa un determinato test misuri siamo nell'ambito della validità di un test (o misura). La validità di una misura ha a che fare quindi con la domanda «che cosa lo strumento o il test misura realmente?». Diverse sono le definizioni di validità di una misura. Tra queste, una definizione *operativa* è quella che definisce la validità rispetto ad un criterio (o riferimento). Di essa hanno detto bene Lord & Novick [20]:

«The validity coefficient of a measurement can therefore be stated only in relation to a second measurement; thus it makes no sense to speak of the validity coefficient of a measurement. When speaking of the validity coefficient of a measurement, we shall always make explicit the second measurement unless the identity of this measurement can be clearly inferred from the context» (p.59)

In questo senso, studiare la validità di X , ad esempio, significa confrontare quest'ultima con un'altra misura (o più misure) Y , detta appunto *criterio*. Per criterio s'intende, in senso lato, una qualsiasi variabile che si vuole spiegare e/o predire attraverso l'uso di un'altra variabile. La quantificazione della validità secondo il criterio Y avviene usando la correlazione tra X e Y , ossia $\text{Cor}[X, Y]$, che rappresenta un operatore lineare. Per tale motivo, si parla anche di **validità lineare**. In questo contesto, il coefficiente di validità esprime la precisione con cui è possibile prevedere il criterio partendo dal test. È possibile ulteriormente specificare la validità di criterio, distinguendo tra *validità predittiva* e *validità concorrente*. Ciò che differenzia queste ultime è la dimensione temporale di confronto con il criterio: infatti, mentre per validità predittiva s'intende la capacità del test di prevedere il criterio futuro, per validità concorrente s'intende la capacità di prevedere il criterio quando viene somministrato sincronicamente rispetto al test.

Consideriamo una misura X e un'altra misura Y (criterio). Per costruire un coefficiente per la validità possiamo allora semplicemente utilizzare il valore assoluto della correlazione tra misura e criterio:

$$|\text{Cor}[X, Y]| = \frac{|\text{Cov}[X, Y]|}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}$$

L'interpretazione che ne risulta è semplice: $|\text{Cor}[X, Y]| \approx 0$ indica scarsa concordanza tra X e il criterio Y (X non è valida secondo Y) mentre $|\text{Cor}[X, Y]| \approx 1$ indica perfetta concordanza (X è valida secondo Y). In altre parole, se la correlazione con valore assoluto tende a zero significa che la validità di X è nulla, ossia X non è Y -valida. Quando la correlazione con valore assoluto tende ad uno, invece, vuol dire che la X è valida secondo Y (oppure Y -valida). Da ciò si evince che la validità di criterio sia una *valutazione condizionale*.

Per le altre accezioni di validità il lettore interessato potrà consultare [2], in particolare il capitolo 6.

4.5 Dalla validità di una misura all'attendibilità

Consideriamo ancora il caso in cui X e Y siano (p_1)-parallele, nel senso che $X = T + E$ e $Y = T + E$. Il numeratore del coefficiente di validità, allora, si semplifica come segue:

$$\begin{aligned} \text{Cov}[X, Y] &= \text{Cov}[(T + E_X), (T + E_Y)] \\ &= \text{Cov}[T, T] + \cancel{\text{Cov}[T, E_X]}^0 + \cancel{\text{Cov}[T, E_Y]}^0 + \cancel{\text{Cov}[E_Y, E_X]}^0 \\ &= \text{Var}[T] \end{aligned}$$

Analogamente il denominatore:

$$\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]} = \text{Var}[X]$$

In questo caso, il coefficiente di validità $|\text{Cor}[X, Y]|$ si semplifica esprimendo l'**attendibilità** di un test o misura:

$$\rho_{XT}^2 = \frac{\text{Var}[T]}{\text{Var}[X]}, \quad \rho_{XT}^2 \in [0, 1]$$

che esprime, in altri termini, la *validità di una misura rispetto ad un criterio parallelo* Y . Alternativamente:

$$\rho_{XT}^2 = 1 - \frac{\text{Var}[E]}{\text{Var}[X]}$$

a causa della decomposizione di $\text{Var}[X]$ secondo la proprietà (v) della TCT. Si noti che nel caso limite in cui la varianza dell'errore sia pari a zero, l'attendibilità assume valore pari ad uno e, viceversa, se la varianza d'errore è uguale alla varianza dell'osservato, l'attendibilità assumerà valore necessariamente pari a zero. In sintesi:

$$\begin{aligned} \text{Var}[E] = 0 &\implies \rho_{XT}^2 = 1 \\ \text{Var}[E] = \text{Var}[X] &\implies \rho_{XT}^2 = 0 \end{aligned}$$

Da ciò si evince come l'attendibilità esprima informazioni rispetto alla precisione (imprecisione) della misura: infatti nella sua definizione è coinvolta la varianza e, in particolare, $\text{Var}[E]$ esprime l'imprecisione della misura. Da ciò, quando $\text{Var}[E] \rightarrow 0$ la precisione e dunque l'attendibilità dello strumento cresce. Varianza d'errore e attendibilità, infatti, sono inversamente proporzionali.

4.6 Misure composite

Finora abbiamo considerato misurazioni fatte usando più misure parallele *non gerarchiche*. Ad esempio, X_j e $X_{j'}$ sebbene distinte contribuiscono allo stesso modo a determinare il valore vero T . Ci sono casi tuttavia in cui le diverse misure sono utilizzate per *comporre* altre misure che sono sintesi delle precedenti come nel caso delle *misure indirette* viste nel Capitolo 1. In questo caso, le misure sono definite secondo un'organizzazione *gerarchica*. Diremo dunque che X è una **misura composta** quando $X = \phi(Y_1, \dots, Y_m)$, con $\phi(\cdot)$ funzione di legame nota, e chiameremo **misure componenti** le osservabili Y_1, \dots, Y_m .

Supponiamo di voler costruire un test complessivo X . Come visto in precedenza, questo è un insieme di scale Y_1, \dots, Y_m mentre le scale sono aggregati di item coerenti tra loro. Prendiamo in considerazione il caso semplice $X = Y_1 + Y_2$ in cui le misure componenti - o di livello inferiore - compongono additivamente la misura composta - o di livello superiore - con $m = 2$. Si noti che il test X è di lunghezza doppia e m indica la lunghezza del test. Scriviamo le due scale secondo la notazione della TCT:

$$(Y_1, T_1, E_1) \quad \text{e} \quad (Y_2, T_2, E_2)$$

e definiamo la misura composta (X, T, E) per somma:

$$\begin{aligned} X &= Y_1 + Y_2 \\ &= (T_1 + E_1) + (T_2 + E_2) \end{aligned}$$

La TCT introdotta per il caso di una singola scala (vedi Tabella 4.1) si generalizza come segue:

$$(viii) \quad \mathbb{E}[X] = \mathbb{E}[Y_1] + \mathbb{E}[Y_2]$$

$$(ix) \quad \mathbb{V}\text{ar}[X] = \mathbb{V}\text{ar}[Y_1] + \mathbb{V}\text{ar}[Y_2] + 2\text{Cov}[Y_1, Y_2]$$

$$(x) \quad \mathbb{V}\text{ar}[T] = \mathbb{V}\text{ar}[T_1] + \mathbb{V}\text{ar}[T_2] + 2\text{Cov}[T_1, T_2]$$

$$(xi) \quad \mathbb{V}\text{ar}[E] = \mathbb{V}\text{ar}[E_1] + \mathbb{V}\text{ar}[E_2] + 2\text{Cov}[E_1, E_2] \rightarrow 0$$

Possiamo notare che la varianza di X e T (ix)-(x) ha in sé ora una componente prima assente, vale a dire la covarianza tra le scale Y_1 e Y_2 : infatti se queste non co-variassero non avremmo una misura composta e la proprietà (ix) si ridurrebbe al caso (v). Si nota anche che la varianza del punteggio vero $\mathbb{V}\text{ar}[T]$ è uguale alle varianze dei punteggi veri delle due misurazioni sommato alla componente di covarianza. Infine possiamo notare ancora che nel caso Y_1 e Y_2 fossero (p_1)-parallele, le proprietà (viii)-(xi) si semplificherebbero ulteriormente poiché in quel caso si avrebbe che $\mathbb{E}[Y_1] = \mathbb{E}[Y_2]$ e $\mathbb{V}\text{ar}[Y_1] = \mathbb{V}\text{ar}[Y_2]$. La tabella 4.3 sintetizza i risultati della TCT generalizzati al caso di test composti.

$\underbrace{X}_{\text{misura composta}} = \underbrace{T_1 + E_1}_{\text{misura componente } Y_1} + \underbrace{T_2 + E_2}_{\text{misura componente } Y_2}$
(i) $\mathbb{E}[E_1] = \mathbb{E}[E_2] = 0$
(ii) $\text{Cor}[E_1, T_1] = \text{Cor}[E_2, T_2] = \text{Cor}[E_2, T_1] = \text{Cor}[E_1, T_2] = 0$
(iii) $\text{Cor}[E_1, E_2] = 0$
(viii) $\mathbb{E}[X] = \mathbb{E}[Y_1] + \mathbb{E}[Y_2]$
(ix) $\mathbb{V}\text{ar}[X] = \mathbb{V}\text{ar}[Y_1] + \mathbb{V}\text{ar}[Y_2] + 2\text{Cov}[Y_1, Y_2]$
(x) $\mathbb{V}\text{ar}[T] = \mathbb{V}\text{ar}[T_1] + \mathbb{V}\text{ar}[T_2] + 2\text{Cov}[T_1, T_2]$
(xi) $\mathbb{V}\text{ar}[E] = \mathbb{V}\text{ar}[E_1] + \mathbb{V}\text{ar}[E_2]$

TABELLA 4.3: Sintesi della TCT per un test composto di lunghezza doppia. Nota: gli assunti (i)-(iii) e (viii) sono gli stessi di quelli visti nel caso precedente.

4.7 Attendibilità di misure composite

È ragionevole estendere i risultati ottenuti nel caso dell'attendibilità di una singola scala al caso di misure composte. Siano Y_1, Y_2, Y_3, Y_4 misure componenti (p_1)-parallele e siano

$$X_1 = Y_1 + Y_2 \quad X_2 = Y_3 + Y_4$$

le rispettive misure composte. Se le misure componenti sono parallele, anche le misure composte X_1 e X_2 sono parallele (dunque entrambe sottendono unico T). L'attendibilità del test composito è dunque pari a:

$$\rho_{XT}^2 = \rho_{X_1 X_2}^2 = \frac{2\rho_{Y_1 Y_2}^2}{1 + \rho_{Y_1 Y_2}^2}$$

dove $\rho_{Y_1 Y_2}^2$ è l'attendibilità della misura componente Y_1 (rispetto alla misura parallela Y_2). Si noti che nel caso di misure parallele, l'attendibilità può essere scritta in generale per qualunque coppia di misure parallele X, X' e Y, Y' ottenendo:

$$\rho_{XX'}^2 = \frac{2\rho_{YY'}^2}{1 + \rho_{YY'}^2} \quad \text{con } \rho_{XX'}^2 \in [0, 1]$$

Quest'ultima formulazione è nota come **formula di Sperman-Brown** per un test di lunghezza doppia (poiché $m = 2$). È importante specificare che $\rho_{XX'}^2 > \rho_{YY'}^2$, ossia che l'attendibilità di un test composito $\rho_{XX'}^2$ è sempre maggiore di quella delle singole componenti $\rho_{YY'}^2$, che invece fanno riferimento ad un test di lunghezza unitaria. La figura 4.1 sintetizza graficamente tale risultato notevole.

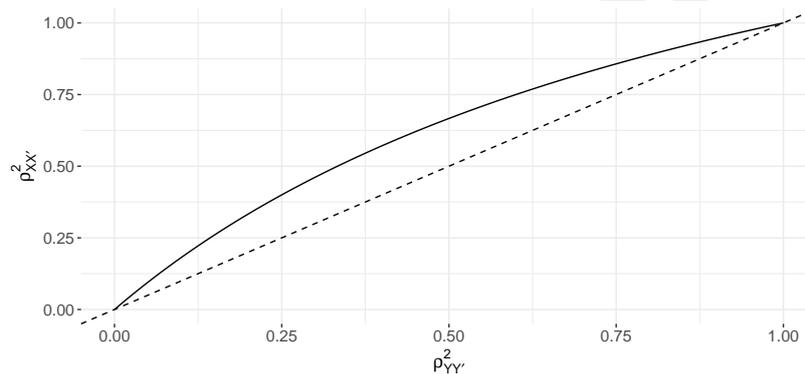


FIGURA 4.1: *Attendibilità di un test composito di lunghezza doppia: Curva di attendibilità calcolata secondo la formula di Sperman-Brown. In ordinata si ha un test di lunghezza doppia, mentre in ascissa un test di lunghezza unitaria. Un test composito è più preciso di un test unitario. Si noti $\rho_{XX'}^2 = \rho_{YY'}^2$, solo per valori estremi di attendibilità.*

4.7.1 Attendibilità e lunghezza del test

I risultati ottenuti per le misure composte di lunghezza due possono essere generalizzate quando $m > 2$. Ci sono due risultati importanti in questo caso che coinvolgono (i) la varianza del valore vero della misura composta $\text{Var}[T_m]$ e (ii) la varianza della componente di errore della misura composta $\text{Var}[E_m]$. Tali risultati supportano l'idea che all'aumentare della lunghezza di un test m , aumenti anche l'attribuzione della varianza delle misurazioni alla componente vera piuttosto che all'errore.⁵ Più in generale,

⁵In altri termini, della varianza complessiva si riesce a separare meglio la componente attribuita al misurando latente (costruito da quantificare) rispetto a quella dell'errore.

l'aggiunta di item ad un test - il c.d. **allungamento del test** - fa variare le componenti $\text{Var}[X]$, $\text{Var}[T]$, $\text{Var}[E]$ del modello di misura. In particolare, indicando con la solita simbologia della TCT:

$$(X_m, T_m, E_m)$$

la misura composta di lunghezza m (il test allungato), si hanno i seguenti risultati notevoli:

$$(xii) \quad \text{Var}[X_m] = m\text{Var}[Y] (1 + (m-1)\rho_{YY'}^2)$$

$$(xiii) \quad \text{Var}[T_m] = m^2\text{Var}[T]$$

$$(xiv) \quad \text{Var}[E_m] = m\text{Var}[E]$$

dove Y indica le misure componenti (p_1)-parallele, T rappresenta il valore vero sotteso a Y , E l'errore di queste ultime mentre $\rho_{YY'}$ indica l'attendibilità delle misure componenti. Le proprietà (xiii)-(xiv) rappresentano quei risultati a cui accennavamo all'inizio del paragrafo. Rispetto a questi ultimi, si nota come $\text{Var}[T_m]$ cresca in maniera *quadratica* al crescere di m mentre $\text{Var}[E_m]$ cresca in maniera *lineare* al crescere di m . Ciò implica necessariamente che all'aumento di lunghezza del test corrisponda $\text{Var}[T_m] > \text{Var}[E_m]$, ossia il fatto che la varianza del misurando cresca più velocemente della varianza d'errore. Il risultato è l'ottenimento di un test più preciso e con attendibilità più alta.⁶ La figura 4.2 rappresenta tale risultato graficamente.

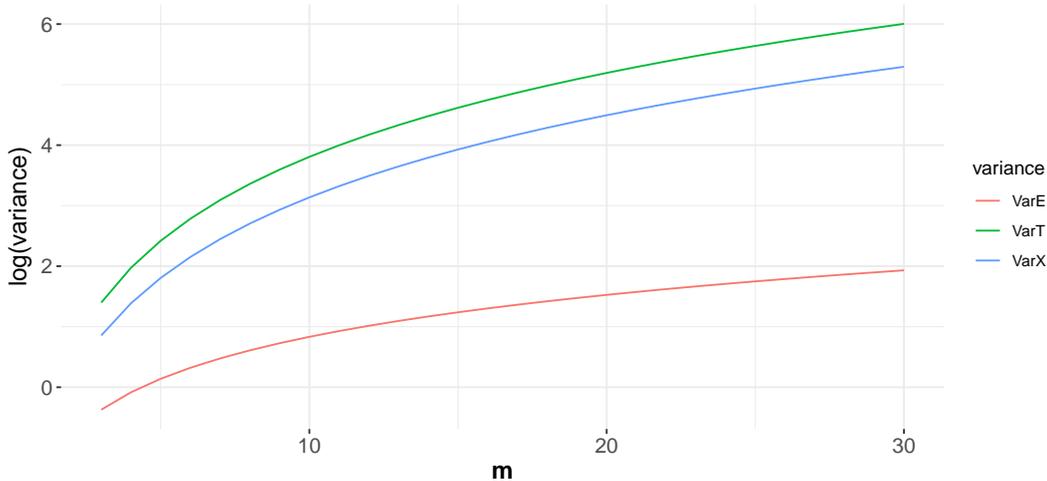


FIGURA 4.2: Componenti di varianza per una misura composta: varianza di errore $\text{Var}[E_m]$ (linea rossa), varianza della misurazione $\text{Var}[X_m]$ (linea blu) e varianza del valore vero $\text{Var}[T_m]$ (linea verde) per una misura composta (X_m, T_m, E_m) di lunghezza m prefissata. Nota: in ascissa sono riportate le lunghezze m della composizione (da 3 a 30) mentre in ordinata sono riportate le tre componenti di varianza rappresentate su scala logaritmica. Il grafico evidenzia come $\text{Var}[T_m] > \text{Var}[E_m]$ per ogni intero m fissato.

⁶Si noti che tale risultato segue dal fatto che nella definizione di attendibilità, ossia $\rho_{XT}^2 = \frac{\text{Var}[T]}{\text{Var}[X]}$, la quantità $\text{Var}[T]$ è posta al numeratore mentre $\text{Var}[X]$ rimane invariata.

La formula generale dell'attendibilità per un test di lunghezza iniziale m allungato di k item è la seguente:

$$\rho_{XX'}^2 = \frac{\frac{(m+k)}{m} \rho_{YY'}^2}{1 + \left(\frac{(m+k)}{m} - 1\right) \rho_{YY'}^2}$$

da dove si evince il fatto che l'incremento di attendibilità di un test non dipende solo dalla lunghezza $(m+k)$ ma anche dalla **fedeltà iniziale** $\rho_{YY'}^2$. Se quest'ultima è bassa, infatti, bisognerà creare nuovi item e/o un nuovo test che abbia una maggiore fedeltà iniziale poiché l'allungamento di un test con basso $\rho_{YY'}^2$ richiederebbe molti item in aggiunta. Ciò, in più, si scontra con varie difficoltà dal momento che un test più lungo comporta svantaggi (ad esempio stanchezza) che possono annullare i vantaggi dovuti all'incremento dell'attendibilità stessa del test. Infine, ri-arrangiando i termini della definizione di attendibilità possiamo ottenere la seguente formulazione:

$$k = m \left(\frac{\rho_{XX'}^2 (1 - \rho_{YY'}^2)}{\rho_{YY'}^2 (1 - \rho_{XX'}^2)} \right) - m$$

utile per determinare l'**allungamento necessario** k per ottenere un test con livello di attendibilità prescelto $\rho_{XX'}^2$ (in questo caso si fissa quest'ultima quantità al livello desiderato).

Di seguito un esempio di applicazione delle formule dell'attendibilità per l'allungamento di un test. Immaginiamo di disporre di un test inizialmente composto da $m = 13$ item avente attendibilità $\rho_{YY'}^2 = 0.78$. Chiediamoci a quanto ammonterebbe l'attendibilità del test se aggiungessimo $k = 8$ misure/item (p_1)-parallele. Applicando la formula profetica di Spearman-Brown otterremmo un'attendibilità pari a:

$$\rho_{XX'}^2 = \frac{\frac{(13+8)}{13} 0.78}{1 + \left(\frac{(13+8)}{13} - 1\right) 0.78} = 0.851$$

che corrisponde ad un incremento di 0.07 unità. Applicando invece la formula inversa, data l'attendibilità di arrivo, otteniamo invece:

$$k = 13 \cdot \frac{0.85(1 - 0.78)}{0.78(1 - 0.85)} - 13 \cong 8$$

come ottenuto dal calcolo precedente. Quindi per ottenere un test con un'attendibilità pari a 0.85 occorrerà avere un numero di item pari a 21 con un incremento di $(21 - 13) = 8$ item rispetto alla versione iniziale formata da 13 item.

Infine la figura 4.3 rappresenta diversi risultati ottenibili allungando il test iniziale fino a raggiungere la lunghezza complessiva m dati alcuni livelli differenti per l'attendibilità iniziale. Da questo risultato grafico si evince il fatto che, in presenza di scale iniziali con bassa attendibilità, possiamo aggiungere un certo numero di item per aumentarla. Nel caso opposto, tuttavia, se si hanno scale componenti molto buone (ad esempio, $\rho_{XX'} = 0.65$) diventa poco utile aggiungere ulteriori item poiché l'incremento di attendibilità avverrà più lentamente fino a raggiungere un plateau.

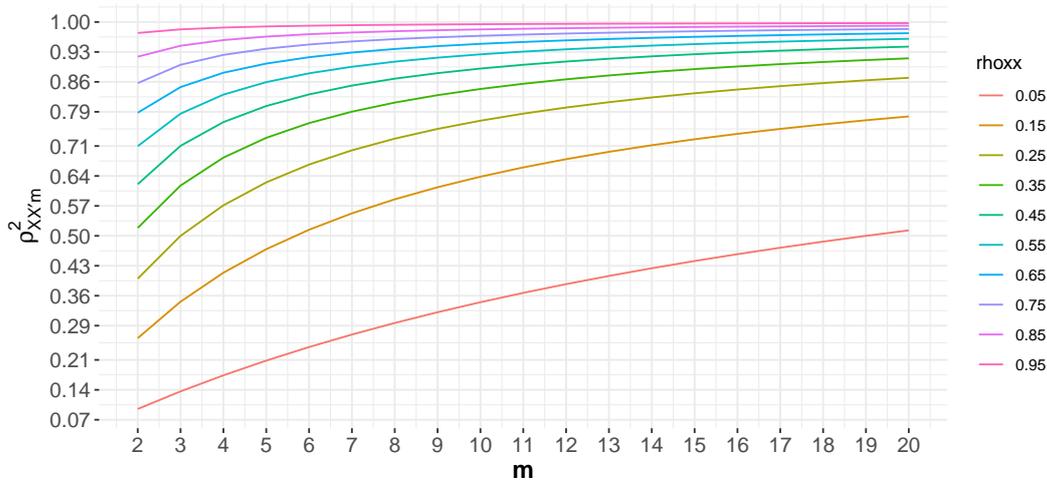


FIGURA 4.3: Attendibilità di una misura composta e lunghezza: Curve di attendibilità di una misura composta (o test) quando questa è allungata per un intero m . Le curve sono in funzione dell'attendibilità iniziale ρ_{XX}^2 , della misura composta (colore delle curve) e mostrano come questa, tenendo fisso tale valore, cambia in precisione $\rho_{XX'm}^2$ (ordinata) quando m nuove misure componenti (ascissa) sono aggiunte a formare il test.

4.8 Stime dell'attendibilità

4.8.1 α di Cronbach

Consideriamo m misure componenti $(Y_1, T_1, E_1), \dots, (Y_m, T_m, E_m)$ che formano la misura composta $X = \sum_{j=1}^m Y_j$. Una stima (approssimata) dell'attendibilità della misura composta $\rho_{XX'}^2$, è la seguente:

$$\hat{\rho}_{XX'}^2 = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \text{Var}[Y_j]}{\text{Var}[X]} \right)$$

nota anche come **α di Cronbach**. Tale formulazione fornisce un'interpretazione dell'attendibilità in termini di **coerenza interna** della misura composta. A differenza di altre tecniche per valutare l'attendibilità di un test (ad esempio, *split-half*), essa utilizza solo l'informazione derivante dalle misure/item componenti. Infatti il denominatore $\text{Var}[X]$, utilizzando il risultato (ix) della TCT, lo possiamo esprimere in funzione dei soli item come segue:

$$\text{Var}[X] = \sum_{j=1}^m \text{Var}[Y_j] + \sum_{j \neq h} \text{Cov}[Y_j, Y_h]$$

dove la sommatoria è sulle misure componenti che sono *diverse*: infatti quando $j = h$ abbiamo $\text{Cov}[Y_j, Y_h] = \text{Var}[Y_j, Y_h]$ già calcolata nella sommatoria precedente. Sostituendo quest'ultimo risultato nella formula dell' α di Cronbach otteniamo infine:

$$\hat{\rho}_{XX'}^2 = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m \text{Var}[Y_j]}{\sum_{j=1}^m \text{Var}[Y_j] + \sum_{j \neq h} \text{Cov}[Y_j, Y_h]} \right)$$

da cui si nota come la covarianza tra gli item fornisca informazione circa la coerenza interna. Infatti, maggiore è la covarianza, maggiore sarà l'indice α finale e, conseguentemente, la coerenza interna. Al contrario, quando la covarianza tra le misure è prossima allo zero, l'indice di Cronbach assumerà anch'esso valore prossimo allo zero con la conseguenza naturale che anche la coerenza interna sarà molto bassa.

In sintesi, lo stimatore α è non distorto per $\rho_{XX'}^2$, se le misure componenti sono p_1 -parallele o $p_2 - p_3$ -equivalenti, interpreta l'attendibilità di una misura composta in termini di coerenza interna rispetto alle misure componenti che la formano, assunto che queste formino bene la misura composta, ossia presentino una buona inter-correlazione tra loro. L'interpretazione dell'indice di Cronbach può essere fatta seguendo alcuni valori soglia tipicamente utilizzati nella ricerca psicologica [2]:

- $\alpha > 0.9$ *ottima*
- $0.8 \leq \alpha \leq 0.9$ *buona*
- $0.7 \leq \alpha < 0.8$ *discreta*
- $0.6 \leq \alpha < 0.7$ *sufficiente*
- $\alpha < 0.6$ *insufficiente*

Nella valutazione, tuttavia, i valori soglia devono comunque essere considerati anche in base al costrutto che viene misurato dal test. Ad esempio un test di performance spesso raggiunge coefficienti sensibilmente inferiori alla soglia $\alpha = 0.80$ che viene invece più facilmente raggiunta da test per abilità cognitive.

Sebbene ancora oggi risulti l'indice maggiormente utilizzato per la valutazione della coerenza interna di un test, l' α di Cronbach non è esente da criticità. Come dimostrato in un recente studio di simulazione [28], in cui l'indice α è stato confrontato con altri indici simili, l'indice di Cronbach funziona bene solo in particolari condizioni di ottimalità, spesso non riscontrabili nella pratica di ricerca. L'autore perciò suggerisce di utilizzare un approccio basato sull'analisi fattoriale confermativa per la valutazione della coerenza interna di un test, la quale fornisce altresì informazione più completa sulla struttura fattoriale sottostante unitamente ad indici (ad esempio, l'indice ω) che permettono una migliore valutazione dell'affidabilità anche in situazioni in cui gli item non rispettino le assunzioni ottimali richieste dall' α di Cronbach. Si rimanda il lettore al Capitolo 5 per una introduzione alla tecnica dell'analisi fattoriale confermativa allo studio della coerenza interna di un test.

4.8.2 Altre tecniche per la valutazione dell'attendibilità

Oltre all'indice di Cronbach e al metodo della coerenza interna, esistono altre procedure e metodi per la valutazione della precisione di un test, ognuno dei quali presuppone una particolare forma di attendibilità. Tra questi ricordiamo:

- metodo delle **forme parallele**: due forme parallele del test vengono somministrate a breve distanza di tempo agli stessi soggetti. L'attendibilità è qui intesa come interscambiabilità, equivalenza, delle due forme.

- metodo **test-retest**: lo stesso test viene somministrato agli stessi soggetti in momenti/tempi differenti. L'attendibilità è qui intesa come stabilità dei punteggi nel tempo (impermeabilità).
- metodo dello **split-half**: il test viene diviso a metà e si assume che queste siano parallele, ossia misurino ugualmente lo stesso costrutto. L'attendibilità è qui calcolata mediante l'*indice di Rulon*.

Infine, può essere utile richiamare l'attenzione sull'utilizzo degli indici rispetto alla natura degli item. In particolare, l' α di Cronbach si usa nel caso di politomici misurati al livello degli intervalli equivalenti mentre l'indice di Kuder-Richardson è usato agevolmente nel caso di item dicotomici. Una disamina più articolata su tale argomento è rintracciabile in [2], capitolo 2. Il lettore interessato potrà trovare in quella sede i riferimenti e i dettagli necessari.

4.9 Fattori che influenzano l'attendibilità

Diverse sono le fonti che possono minare l'attendibilità di un test o una scala. Tra queste ricordiamo: la qualità del campione di individui su cui le misurazioni sono effettuate (ampiezza, rappresentatività, eterogeneità), la qualità delle misure componenti o items (questi devono essere semanticamente chiari, scritti bene e con bassa desiderabilità sociale), le condizioni generali di somministrazione del test (la risposta deve essere causata dall'elicitazione della componente latente che si vuole indagare), stabilità delle condizioni di somministrazione (tipo di consegna del test, atteggiamento dell'esaminatore e del soggetto verso il test), un'adeguata lunghezza del test, l'eventuale limite di tempo della somministrazione, alcuni aspetti cognitivi degli individui testati come ad esempio la fatica, il ricordo e la bassa compliance.

4.10 Stima dei punteggi veri

Oltre alle quantità relative al test, vale a dire $\mathbb{E}[T]$, $\text{Var}[E]$, $\text{Var}[T]$ ed infine ρ_{XT}^2 , è possibile voler stimare i valori finali dei rispondenti al test, i c.d. **punteggi veri al test** (o *test scores*). Tali valori, che indicheremo con τ_1, \dots, τ_n , corrispondono ai valori degli individui sul costrutto latente quantificato, ossia i valori che questi avrebbero se fosse possibile misurare direttamente il misurando senza passare per le misure osservabili (*proxies*). L'approccio più utilizzato è quello di ottenere τ_1, \dots, τ_n attraverso un modello lineare che tenga conto delle qualità del test e delle misurazioni osservate come di seguito:

$$\hat{\tau}_i = \rho_{XT}^2 x_i + (1 - \rho_{XT}^2) \mu_x$$

dove $\mu_x = \frac{1}{NR} \sum_{i=1}^n \sum_{r=1}^R y_{ir}$ è la media complessiva delle misurazioni effettuate (stima della media dei punteggi veri nella popolazione) mentre ρ_{XT}^2 è l'attendibilità del test. Lo stimatore è una combinazione lineare convessa - si noti che $\rho_{XT}^2 + (1 - \rho_{XT}^2) = 1$ - che *pesa* in maniera differenziale il valore attribuito ai dati osservati x_i o alla sintesi complessiva μ_x ; in particolare, quanto più alta sarà l'attendibilità ρ_{XT}^2 , tanto più la componente individuale x_i peserà sui punteggi veri; al contrario, quanto più bassa sarà

l'attendibilità ρ_{XT}^2 maggiormente peserà la seconda componente della formula, ossia la media delle misurazioni. Si osservi come lo stimatore per τ_i potrebbe non essere adeguato quando la relazione tra X e T è *non lineare*. Esempi di applicazione della stima dei punteggi veri sono presenti nel laboratorio 5.

4.11 Esempio di applicazione della TCT

In questo paragrafo riportiamo un semplice esempio didattico per il calcolo delle caratteristiche di un test secondo la TCT. L'esempio è tratto interamente da [20] a cui si rimanda per ulteriori dettagli. Immaginiamo di aver misurato una certa quantità di interesse su un campione di $n = 10$ individui utilizzando un test con $p = 1$ item (indicatori). Ogni individuo è stato sottoposto a $R = 2$ misurazioni (repliche) *parallele*. In questo esempio, le misure parallele sono da intendersi sull'individuo e non sugli item. I dati raccolti sono riportati in tabella 4.4.

n	1	2	3	4	5	6	7	8	9	10
$r = 1$	125	119	109	104	101	98	97	94	90	81
$r = 2$	120	122	107	108	98	106	96	99	93	87

TABELLA 4.4: *Esempio di applicazione della TCT: dati misurati su $n = 10$ individui con $p = 1$ item e $R = 2$ misure parallele (repliche).*

Usando i dati osservati, ci interessa determinare le quantità ignote della TCT, vale a dire $\mathbb{E}[T]$, $\text{Var}[E]$, $\text{Var}[T]$ ed infine ρ_{XT}^2 . Nel caso di campioni molto elevati, il calcolo di tali quantità può essere fatto ricorrendo direttamente all'uso delle definizioni date fin'ora. In caso contrario, invece, è necessario correggere gli stimatori della TCT per ottenere stime non distorte delle quantità vere di popolazione come di seguito indicato:

- Stima di $\text{Var}[E]$:

$$\begin{aligned}\hat{\sigma}_E^2 &= \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{E_i}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{R-1} \sum_{r=1}^R (x_{ir} - \bar{x}_i)^2 \right)\end{aligned}$$

dove $\hat{\sigma}_E^2$ è la stima della *varianza di gruppo*, $\hat{\sigma}_{E_i}^2$ è la stima della *varianza individuale* mentre \bar{x}_i è la misurazione media per ciascun individuo.

- Stima di $\text{Var}[T]$:

$$\hat{\sigma}_T^2 = \frac{1}{nR^2} \left(\sum_{i=1}^n (a_i)^2 - \frac{\sum_{i=1}^n a_i}{n} \right) - \frac{1}{R} \hat{\sigma}_E^2$$

dove $\hat{\sigma}_E^2$ è la stima della *varianza dell'errore* e $a_i = \sum_{r=1}^R y_{ir}$ è la *somma delle repliche* per ciascun individuo.

- Stima di ρ_{XT}^2 :

$$\hat{\rho}_{XT}^2 = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_T^2 + \hat{\sigma}_E^2}$$

dove l'attendibilità è stimata rapportando la quantità di varianza della componente vera alla varianza complessiva, dovuta cioè alla varianza della componente d'errore e della componente vera stessa (**coefficiente di correlazione intraclass**).

Nell'esempio considerato le quantità stimate ammontano a $\hat{\sigma}_E^2 = 9.9$, $\hat{\sigma}_T^2 = 140.67$, $\rho_{XT}^2 = 0.934$. La stima dei punteggi veri può essere fatta ricorrendo al metodo regressivo come segue:

$$\begin{aligned}\hat{\tau}_1 &= 0.934 \left(\frac{125 + 120}{2} \right) + (1 - 0.934)102.7 \\ &\vdots \\ \hat{\tau}_{10} &= 0.934 \left(\frac{81 + 87}{2} \right) + (1 - 0.934)102.7\end{aligned}$$

ottenendo infine i risultati riportati in tabella 4.5.

n	1	2	3	4	5	6	7	8	9	10
$\hat{\tau}_i$	121.20	119.33	107.65	105.78	99.71	102.05	96.91	96.91	92.24	85.23

TABELLA 4.5: Esempio di applicazione della TCT: stima dei punteggi veri per = 10 individui rispondenti al test.

4.12 Selezione delle misure osservabili

Un test è formato dall'aggregazione di scale che raggruppano a loro volta un insieme *coerente* di misure osservabili (indicatori). L'analisi degli item ci permette di selezionare le domande che effettivamente permettono di misurare il costrutto latente, quelle cioè da inserire nella versione finale della scala. Un test formato da buoni item si caratterizza per un alto livello di precisione nel quantificare il misurando. Al contrario, test con item scarsi presentano anche bassi livelli di precisione. In questo senso si può parlare di attendibilità come grado di *impermeabilità* rispetto all'errore di misura [2]. Due sono gli scenari da considerare in questo senso:

- $\text{Var}[E] > \text{Var}[T]$ indica bassa precisione/attendibilità e ci informa sul fatto che la maggior parte della variabilità nel test è legata all'errore di misura
- $\text{Var}[T] > \text{Var}[E]$ indica alta precisione/attendibilità e ci informa che la variabilità del test è dovuta al misurando latente, dunque al tratto (o costrutto) che stiamo indagando

Da ciò discende naturalmente che per migliorare la precisione di un test è necessario avere buone misure osservabili, la cui inclusione nel test ne aumenti la precisione. La

procedura che permette di selezionare le osservabili tra quelle che aumentano l'attendibilità di un test è detta **analisi degli items**.

Considereremo due criteri per la selezione degli item:

- *difficoltà*: esprime quanto un item/indicatore risulta di difficile utilizzo da parte dei soggetti sottoposti al test. Se un item è troppo difficile, non riuscirà mai ad essere utilizzato in maniera ottimale per misurare il costrutto, ossia non permetterà una buona separazione tra $\text{Var}[T]$ (che desideriamo essere la più alta possibile) e $\text{Var}[E]$ (che invece desideriamo sia bassa).
- *capacità discriminativa*: esprime il grado di distinzione tra misure ottenute su soggetti diversi. Ovvero indica la capacità di un item/scala di discriminare tra soggetti che presentano bassi valori del misurando e soggetti che invece presentano alti valori del misurando (*potere discriminante* dell'item).

Gli item possono essere di diverso tipo, ad esempio *dicotomici* (tipici delle risposte «Sì»/«No») o *politomici* (tipici delle scale).

4.12.1 Misure osservabili dicotomiche

Una misura osservabile X_j è dicotomica quando si esprime con due sole categorie non ordinate, ad esempio $X_j \in \{0, 1\}$. Item di questa categoria hanno modalità di risposta del tipo «sì»/«no», «vero»/«falso», «0»/«1». Si assume che la variabile aleatoria (dicotomica) sottesa all'item segua in distribuzione la legge di Bernoulli $X_j \sim \text{Bern}(\pi)$ con $\pi \in (0, 1)$ che indica la proporzione di casi che rispondono correttamente all'item. Dalla teoria della probabilità ricordiamo che:

$$\mathbb{E}[X_j] = \pi \quad \text{Var}[X_j] = \pi(1 - \pi)$$

dove, per una misurazione con n prove indipendenti, il parametro è stimato come segue:

$$\hat{\pi} = \frac{\sum_{i=1}^n X_{ji}}{n}$$

Indice di difficoltà

Un indice di difficoltà può essere ottenuto dividendo il numero di risposte corrette date dai soggetti all'item rispetto al numero totale di risposte all'item. Un indice siffatto varia tra zero ed uno, dove l'estremo inferiore indica che l'item è completamente difficile (nessuna risposta corretta) mentre l'estremo superiore indica che l'item è completamente facile (nessuna risposta scorretta). Dalla caratterizzazione degli item dicotomici come variabili bernoulliane possiamo scrivere l'indice di difficoltà h per il j -esimo item come segue:

$$h_j = \hat{\pi}_j = \frac{\sum_{i=1}^n X_{ji}}{n}$$

mentre l'indice di difficoltà complessivo per una scala con p item è il seguente:

$$h_{\text{tot}} = \frac{\sum_{j=1}^p \hat{\pi}_j}{p}$$

Nel caso di item a risposta dicotomica con *formato a scelta multipla*, l'indice h_j deve essere corretto per il fattore di **guessing** (vale a dire il tentativo di indovinare la risposta a caso) come segue:

$$h_j = \hat{\pi}_j = \frac{(\sum_{i=1}^n X_{ji}) - g_j}{n}$$

dove

$$g_j = \frac{n - (\sum_{i=1}^n X_{ji})}{K - 1}$$

è il fattore di correzione per il *guessing* mentre K indica il numero di alternative di risposta all'item.⁷ La figura 4.4 mostra un esempio di item dicotomico (risposta corretta $X = 0$) con la corrispondente distribuzione di massa di probabilità per diversi valori di π (che invece rappresenta la probabilità della risposta $X = 1$); nel caso estremo rappresentato nel primo pannello di sinistra quasi tutti gli individui hanno risposto correttamente mentre nell'ultimo pannello di destra la quasi totalità degli individui ha risposto in maniera errata; nel pannello centrale solo il cinquanta per cento degli individui ha risposto correttamente all'item. Per ogni pannello è anche indicato l'indice di difficoltà h dell'item: valori vicini ad uno indicano che l'item è facile, viceversa valori prossimi a zero indicano che l'item è difficile. Per agevolare la lettura può essere utile considerare il complemento $(1 - h)$ dell'indice di difficoltà: in questo caso, valori prossimi a zero indicano che l'item è difficile mentre valori prossimi ad uno indicano che l'item risulta facile. Nell'esempio, g indica il fattore di *guessing* che risulta essere inversamente proporzionale alla difficoltà.

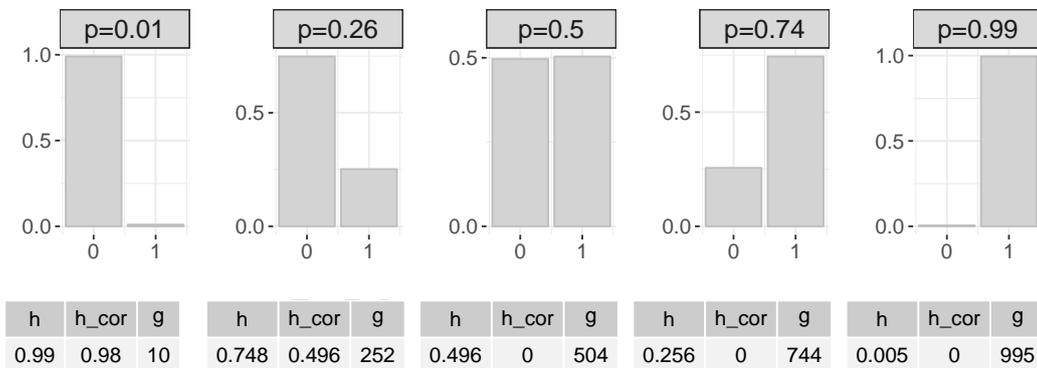


FIGURA 4.4: Difficoltà di un item dicotomico: Distribuzioni di massa di probabilità per un item dicotomico e calcolo degli indici h_j e h_j corretto (h_cor) per il fattore g . Nell'esempio, la risposta corretta è $X = 0$, $n = 1000$ e $K = 2$. Notiamo come all'aumentare di π (p) diminuisce la proporzione di risposte corrette e aumenta la difficoltà (h_j diminuisce progressivamente). Si nota come h_j corretto diminuisce più velocemente di h_j semplice e già per $\pi = 0.5$ indica che l'item è molto difficile ($h_cor=0$).

La figura 4.5 mostra invece il caso di un item con tre alternative di risposta di cui una sola (nell'esempio $X = 0$) è vera. L'interpretazione è del tutto analoga al caso

⁷Si noti come il *guessing*, in questo contesto, non risente delle caratteristiche dei rispondenti - come invece ci si aspetterebbe - ma è definito come una semplice correzione per il numero di categorie di risposte dell'item. Una formulazione del *guessing* che tenga conto invece delle caratteristiche del rispondente è ravvisabile nei modelli più moderni offerti dalla *teoria alla risposta all'item* (IRT). Per un approfondimento il lettore italiano potrà consultare [2], capitolo 3.

precedente. Si osservi che all'aumentare del numero di alternative di risposta, l'uso della legge Binomiale perde di significato e X_j può essere rappresentato (per approssimazione) usando la legge Normale. In questa circostanza, gli item possono essere scelti tra quelli proposti per il caso politomico.

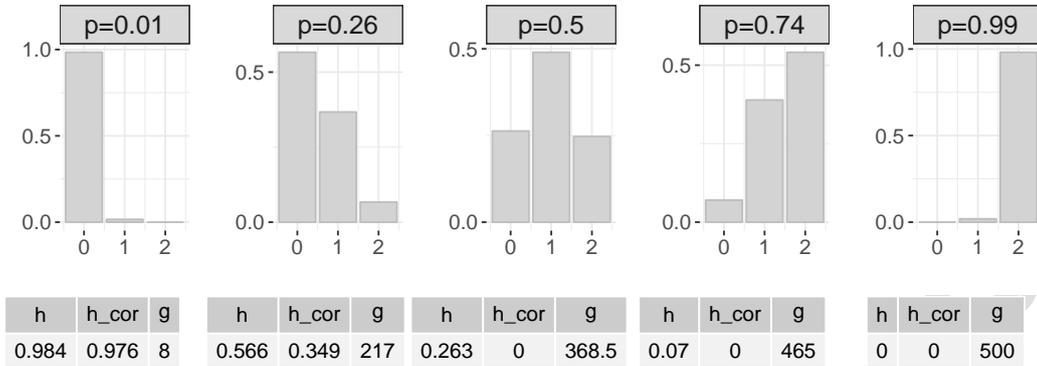


FIGURA 4.5: Difficoltà di un item dicotomico con più alternative: Distribuzioni di massa di probabilità per un item dicotomico a scelta multipla e calcolo degli indici h_j e h_j corretto (h_{cor}) per il fattore g . Nell'esempio, la risposta corretta è $X = 0$, $n = 1000$ e $K = 3$. Notiamo come all'aumentare di π (p) diminuisce la proporzione di risposte corrette e aumenta la difficoltà (h_j diminuisce progressivamente). Si nota come h_j corretto diminuisce più velocemente di h_j semplice e già per $\pi = 0.5$ indica che l'item è molto difficile ($h_{cor}=0$).

Capacità discriminativa

Uno degli obiettivi del testing è quello di fornire informazioni circa le differenze individuali sul costrutto misurato dal test. In tal senso, un buon test è tale se riesce a discriminare tra soggetti che presentano valori elevati sul costrutto e tra coloro che invece presentano valori bassi. Gli indici discriminativi informano su tale capacità di un item. Un primo indice a tale scopo è il seguente:

$$d_j = \frac{\sum_{i=1}^{n_a} x_{ji}^a}{n_a} - \frac{\sum_{i=1}^{n_b} x_{ji}^b}{n_b} \quad d_j \in [-1, 1]$$

dove $\mathbf{x}_j^a = \{x_{ji} : F(x_{ji}) \leq f_0\}$ e $\mathbf{x}_j^b = \{x_{ji} : F(x_{ji}) \geq f_1\}$ sono due partizioni del supporto del j -esimo item di numerosità n_a e n_b , ottenute rispetto ad una certa soglia individuata dalla funzione cumulata $F(x_{ji})$ della variabile item. Tali soglie sono solitamente scelte tra i percentili, ad esempio $f_0 = 0.25$ e $f_1 = 0.75$. In pratica, tale indice usa delle soglie, ad esempio quartili e percentili, per dividere la variabile osservata nei gruppi a e b , per ciascuno dei quali sono calcolate le proporzioni di risposte corrette che vengono infine sottratte tra loro. Più ampia è la differenza risultate d_j maggiore sarà il potere discriminante dell'item. I valori soglia solitamente considerati per l'interpretazione di d_j sono i seguenti [9]:

- $d_j \geq 0.30$: l'item discrimina bene
- $0.20 \leq d_j < 0.3$: l'item discrimina in modo sufficiente (sono richieste revisioni parziali)

- $d_j \leq 0.20$: l'item non discrimina per niente (sono richieste revisioni totali o la sua eliminazione)

4.12.2 Misure osservabili politomiche

Una misura osservabile X_j è politomica quando si esprime con più di due categorie ordinate $X_j \in (0, 1, 2, \dots, K)$. Esempio notevole di tale tipologia di item è la scala tipo Likert. In questo contesto l'item segue in distribuzione la legge Multinomiale (generalizzazione della legge Binomiale), $X_j \sim \text{Multinom}(\pi_1, \dots, \pi_K)$, con $\pi_k \in (0, 1)$ proporzioni di casi che rispondono alle K categorie ordinate sotto il vincolo $\sum_{k=1}^K \pi_k = 1$. Nella pratica dell'analisi degli item, tuttavia, si utilizza l'approssimazione $X_j \sim F(\theta)$ con F generica distribuzione di probabilità *simmetrica e centrata*. Solitamente, si utilizza l'approssimazione normale $F = N(\mu, \sigma^2)$ quando K è grande sebbene questa, in molti casi, non risulta ottimale e può distorcere i risultati dell'analisi. L'analisi degli item politomici avviene attraverso l'utilizzo di **analisi distribuzionale** degli item (ad esempio tramite istogramma o istogramma perequato), insieme agli indici di *tendenza centrale* (media, mediana), *forma* (simmetria, curtosi) e *dispersione*. Sono considerati buoni item quelli che presentano distribuzione simmetrica e centrata sul *valore teorico* della scala politomica⁸, aventi bassa varianza, media e mediane vicine. In generale, si preferiscono item che presentano una distribuzione tendenzialmente normale, la quale garantisce una dispersione elevata ed adeguata dei punteggi. Per esaminare la normalità della distribuzione, oltre alle tipiche rappresentazioni grafiche come il *qq-plot*, si possono analizzare i coefficienti di *curtosi* (spostamento verso l'alto o il basso della distribuzione) e *asimmetria* (spostamento verso destra o sinistra della distribuzione) che approssimano il valore zero quando la distribuzione è normale, mentre rimangono compresi tra meno uno ed uno se la deviazione dalla normalità non è marcata.

Indice di difficoltà

Negli item politomici gli indici di tendenza centrale come media e mediana svolgono la stessa funzione dell'indice di difficoltà degli item dicotomici. In tal senso, è possibile stabilire un intervallo intorno al valore teorico \bar{k} della scala che definisce il livello di difficoltà atteso ed esprimere un indice di difficoltà nel modo seguente:

$$h_j = |\text{median}(\mathbf{x}_j) - \bar{k}_j|$$

dove $|\cdot|$ indica il valore assoluto. L'item analizzato è considerato difficile per valori di $h_j \geq 1.5$. La figura 4.6 mostra le distribuzioni di $p = 4$ item politomici con $K = 10$. Possiamo osservare come gli item 2-3 risultano avere caratteristiche distribuzionali non in linea con le proprietà attese di simmetria e curtosi. Allo stesso modo, l'item 4 rappresenta una distribuzione bimodale, indice che l'item debba essere completamente rivisto o addirittura eliminato.

Capacità discriminativa

Solitamente si utilizza una versione dell'indice d_j approssimata, ottenuta applicando all'item politomico una riduzione di scala (in questo caso si fa diventare l'item dicotomico).

⁸Se l'item è espresso mediante $K = 5$ categorie ordinate, il valore teorico mediano della scala è $\bar{k} = 5$.

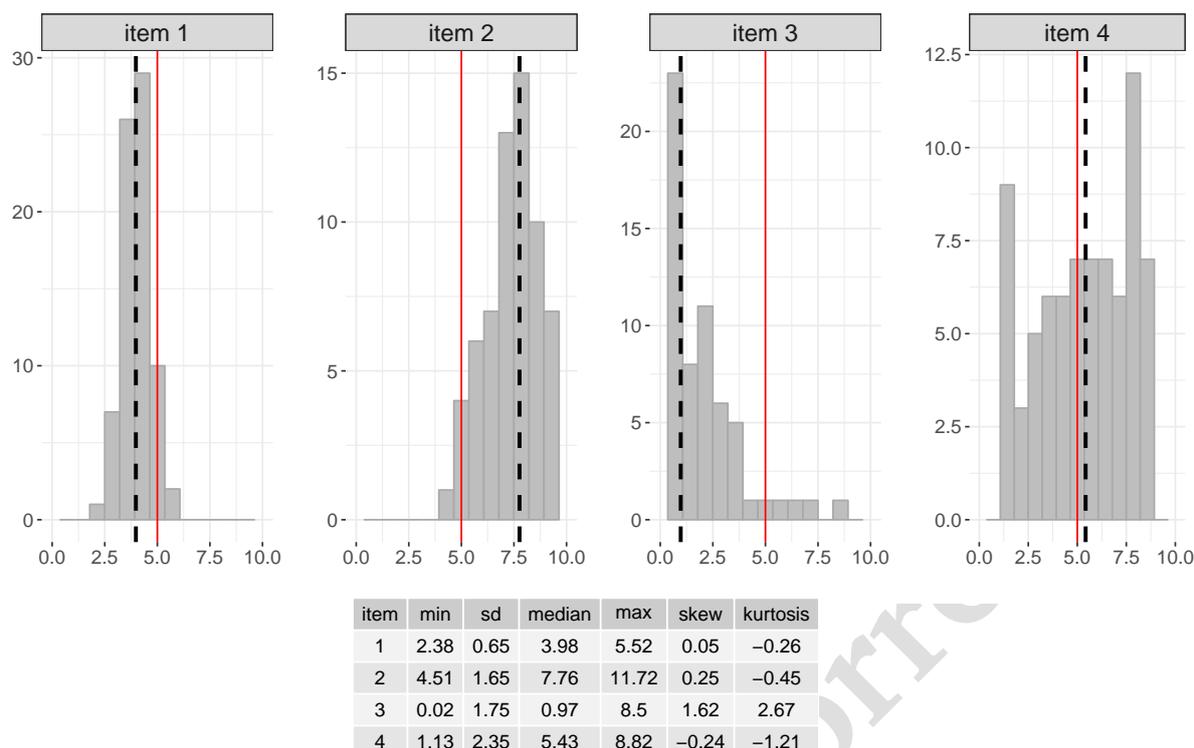


FIGURA 4.6: Distribuzioni di item politomici con $K = 10$ ottenute tramite istogramma. Nota: la linea rossa verticale indica il valore centrale teorico della scala.

In alternativa, si possono utilizzare indici di *effect size* (ad esempio, il d di Cohen) per quantificare la capacità discriminativa tra i gruppi a e b . Tali procedure sono utilizzate negli esempi descritti nel laboratorio 6.

Correlazione item-scala

Un ulteriore modo per valutare la bontà di un item è quello di ricorrere al concetto di coerenza interna. Abbiamo infatti appreso come una buona scala - un aggregato di item - affinché quantifichi bene il costrutto latente debba presentare una certa coerenza interna. Quest'ultima è sovente intesa nel senso dell'inter-correlazione tra gli item. Un indice circa la bontà di un item, che informi il ricercatore sul fatto di tenere o meno l'item nella scala che si sta valutando, è il c.d. **indice di correlazione item-totale** r_{XT} , ossia la correlazione tra il punteggio ottenuto sull'item X_j e il **punteggio totale** del test, ottenuto tramite integrazione degli item che compongono la scala (somma, media, media pesata). L'indice fornisce informazione circa l'attendibilità della misura osservata e, in particolare, coefficienti bassi suggeriscono che il costrutto non venga quantificato bene dall'item in questione. Nel caso di *item dicotomici*, ad esempio, l'indice da utilizzare è il coefficiente di **correlazione punto-biserial** mentre nel caso di item politomici l'indice può essere scelto tra i coefficienti di correlazione per variabili categoriali e/o ordinali (ad esempio, la **correlazione per ranghi**). Maggiori dettagli sono descritti in [2, 9].

4.13 Punteggi totali e standardizzati

I punteggi di un test (le misurazioni/quantificazioni finali ottenute per ciascun individuo sottoposto al test) vengono costruiti utilizzando le misure componenti di una scala. Date quindi Y_1, \dots, Y_p misure componenti, il punteggio totale di una scala (o misura composita) è solitamente calcolato come combinazione lineare delle misure componenti (se gli assunti della TCT per misure parallele sono riscontrati). Esempi combinazioni possono essere:

- Somma: $X = \sum_{j=1}^p Y_j$
- Media: $X = \frac{1}{p} \sum_{j=1}^p Y_j$
- Media pesata: $X = \frac{1}{p} \sum_{j=1}^p Y_j w_j$ (può essere utile richiedere che: $w_j \in (0, 1)$ e $\sum_{j=1}^p w_j = 1$)

Il risultato del calcolo è il c.d. **punteggio grezzo totale** mentre il calcolo di un **punteggio grezzo individuale** per ciascun individuo sulla scala è detto *scoring*. L'interpretazione del punteggio grezzo rilevato sul campione è solitamente effettuata per confronto con i **valori normativi** del test, ossia quei valori della popolazione di riferimento rispetto al costrutto ottenuti nella fase di taratura del test. I valori normativi sono quantificati utilizzando indici di sintesi della variabile aleatoria rappresentante il costrutto (solitamente *media* e *varianza*). Generalmente, per facilitare la lettura dei punteggi finali, le norme comprendono anche aspetti demografici della popolazione di taratura come genere, età e nazionalità. Per valutare/confrontare i punteggi ottenuti dagli individui ad una scala o tra più scale è opportuno trasformare i punteggi grezzi in **punteggi standardizzati** che possono facilmente essere impiegati altresì per la costruzione dei profili individuali. Tali punteggi infatti rendono tra loro uniformi, nel senso della confrontabilità, le scale di un test o i punteggi ottenuti da rispondenti aventi diverse caratteristiche demografiche (ad esempio, per età o genere). Di seguito indicheremo alcuni modi per effettuare il calcolo dei punteggi standardizzati:

- **Rango percentile:** si basa sulla posizione/rango a cui appartiene il punteggio grezzo del soggetto tra i percentili della popolazione

$$r_x = \frac{F_i - 0.5n_i}{n} \cdot 100$$

dove $F_i = F(X \leq x_i)$ indica la frequenza cumulata mentre n_i la frequenza assoluta di x_i nel campione di n unità. Il valore r_x rappresenta la percentuale di soggetti che hanno ottenuto un punteggio grezzo minore o uguale al punteggio x_i (esprime dunque un dato aggregato).

- **Punteggio z:**

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

dove \bar{x} è la media del campione mentre s_x è lo scarto quadratico medio del campione. I punteggi z hanno media nulla e varianza unitaria poiché utilizzano la Normale standardizzata come legge di riferimento. Essi esprimono su una scala centrata sullo zero e con unità di misura lo scarto quadratico medio: le distanze tra punteggi infatti sono espresse in termini di $\pm 1, 2, \dots, K$ deviazioni standard.

- **Punteggio t :** si ottiene per trasformazione lineare del punteggio z :

$$t_i = 50 + 10z_i$$

Essi hanno media pari a cinquanta e deviazione standard pari a dieci e vengono utilizzati preferibilmente per rendere i punteggi finali tutti positivi (i punteggi z infatti possono assumere valori negativi).

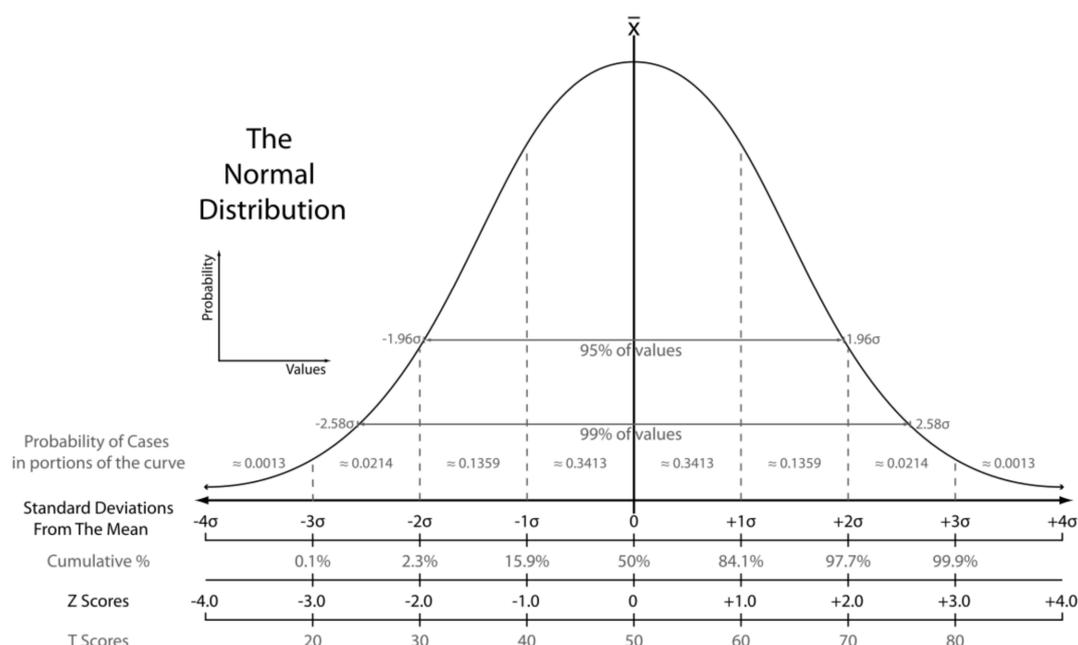


FIGURA 4.7: Tabella grafica per la conversione tra punteggi grezzi e standardizzati.

Fonte: https://en.wikipedia.org/wiki/Standard_score

La figura 4.7 mostra graficamente alcune regole di conversione tra punteggi grezzi e standardizzati avendo il modello Normale come riferimento distribuzionale della scala. Ulteriori dettagli sulle procedure di **standardizzazione** e di **taratura** che rendono possibile il calcolo dei valori normativi e dei punteggi grezzi, unitamente ad altre tecniche per la comparazione tra differenti test di misura che misurano lo stesso costrutto (c.d. **equalizzazione**), sono presenti in [2], capitolo 2, e più estensivamente in [9].

4.14 Note di sintesi

- (i) La TCT decompone la variabilità della misurazione X in una componente di errore E (accidentale) ed una componente attribuita al misurando T (latente)
- (ii) Obiettivo della TCT è di costruire un test di misura che abbia massima precisione/attendibilità, ossia quando $\text{Var}[T] > \text{Var}[E]$

- (iii) L'attendibilità viene costruita utilizzando misure parallele coerenti tra loro
- (iv) Le componenti di variabilità individuale (differenze individuali) non vengono modellate se non attraverso il campione/gruppo
- (v) La TCT si focalizza maggiormente sulle caratteristiche del test più che su quelle degli item

4.15 Laboratorio in R

I laboratori associati a questo capitolo sono i seguenti:

- o `lab5.R`: (A) Stima delle quantità della TCT, (B) Stime della TCT e coerenza interna degli items, (C) Aggiungere items ad una scala, (D) Utilizzo delle quantità stimate tau
- o `lab6.R`: (A) Analisi descrittiva degli items, (B) Coerenza delle scale, (C) Difficoltà degli items, (D) Capacità discriminativa degli items

Il loro contenuto può essere visualizzato e scaricato all'indirizzo <https://github.com/antcalcagni/psychological-testing/tree/master/labs>.

bozza non corretta

Analisi fattoriale confermativa

Nel capitolo precedente si è presa in esame la teoria classica dei test (TCT) che modella la generica misurazione osservata X_j in termini di una componente latente non osservata T ed una parte stocastica residuale E_j secondo la decomposizione lineare $X_j = T + E_j$. Tale modello sottende, dato un insieme di indicatori osservati $j = 1, \dots, p$, quello che spesso è definito come **assunto di unidimensionalità**: una scala di misurazione (*misura composita*), anche quando formata da più items (*misure componenti*), necessita di sottendere un unico misurando (*costrutto latente*) [2, 8]. Diversi sono i modi con cui approcciare un test, ad esempio mediante analisi della sua attendibilità (o precisione) oppure mediante **analisi della dimensionalità**. Quest'ultima viene adoperata, fra le altre cose, per individuare il numero K di componenti latenti (misurandi) *riflesso* dall'insieme $j = 1, \dots, p$ di misure componenti (o indicatori osservati). Specularmente, tale analisi permette di valutare l'unidimensionalità (o meno) di una data scala di misura. Tale analisi è di cruciale importanza nella costruzione di un test di misura: scale unidimensionali, ad esempio, garantiscono che un dato insieme di misure osservate siano coerenti/concordi nel quantificare un dato costrutto latente. Diverse sono le tecniche sovente utilizzate per l'analisi della dimensionalità: analisi fattoriale confermativa (CFA), analisi fattoriale esplorativa (EFA), analisi delle componenti generalizzata, analisi delle componenti basata sui minimi quadrati parziali (PLS). Recentemente, anche tecniche basate su grafi stocastici (c.d. *social network analysis*) sono state utilizzate per l'analisi della dimensionalità di test [10]. Nel presente capitolo verrà descritta la tecnica dell'analisi fattoriale confermativa che, a differenza di altre, consente al ricercatore di definire in maniera diretta la **struttura psicometrica** ipotizzata essere sottesa al test che si sta analizzando. L'approccio, ampiamente inscritto nella tecnica più generale dei modelli ad equazioni strutturali (SEMs) [5], è ampiamente utilizzato nella pratica di ricerca psicometrica e - sebbene non privo di limitazioni - fornisce al ricercatore ampia flessibilità nel definire e valutare la plausibilità empirica di modelli per la quantificazione di attributi psicologici. In questo modo, le ipotesi del ricercatore circa la struttura psicometrica di un determinato test si traducono in un modello statistico che è possibile valutare e

Hanno contribuito alla realizzazione del presente capitolo:

Chiara Fabian, Valeria Pamato, Tania Di Tillio, Giorgia Facchin, Federica Bertoldi, Nicholas Moro, Valeria Ziffoni, Daria Pascu, Giorgia Casano, Francesca Rostellato, Gianluca Mannella, Matteo Cattafi, Stefano Maraschin

analizzare mediante un campione casuale estratto da una popolazione di riferimento. In termini qualitativi, la CFA utilizza quale dato di partenza la matrice di correlazione delle variabili osservate $\text{Cov}[\mathbf{X}] = \text{Cov}[X_1, \dots, X_j, \dots, X_p]$ che viene opportunamente scomposta ed analizzata per determinare quanta parte di variabilità osservata è riconducibile ad una o più variabili latenti. Rispetto al modello di regressione lineare, ad esempio, mentre in quel contesto il dato di partenza è rappresentato da un vettore di dati osservati \mathbf{y} ed una matrice di predittori \mathbf{X} opportunamente posti in relazione secondo il modello lineare $\mathbf{y} = \boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, la CFA utilizza direttamente una *sintesi dei dati*, la statistica $\text{Cov}[\mathbf{X}]$ dove la variabile risposta \mathbf{y} da porre in relazione a \mathbf{X} è in questo contesto assente. Per tale motivo, nel linguaggio del *data mining*, spesso questo tipo di tecnica è anche definita come non-supervisionata (*unsupervised method*).

5.1 Definizione del modello

Nel seguito, per ragioni di chiarezza e semplicità espositiva, faremo riferimento al modello CFA unidimensionale e multidimensionale. Tali aggettivi, in questa particolare accezione, fanno riferimento al numero di variabili latenti e non come di solito accade al numero di variabili osservate che in maniera lapalissiana definiscono sempre un modello multidimensionale.

5.1.1 Modello ad una sola variabile latente

Dato un insieme di variabili osservate $(X_1, \dots, X_j, \dots, X_p)$ il modello fattoriale confermativo è un modello lineare definito come segue:

$$\begin{aligned} X_1 &= \eta\lambda_1 + \delta_1 \\ &\vdots \\ X_j &= \eta\lambda_j + \delta_j \\ &\vdots \\ X_p &= \eta\lambda_p + \delta_p \end{aligned}$$

dove η indica la variabile latente (fattore) che sottende le osservabili, $\delta_1, \dots, \delta_p$ indicano le variabili di errore della misurazione e $\lambda_1, \dots, \lambda_p$ rappresentano i coefficienti che legano le osservabili al misurando latente (c.d. *factor loadings*). In termini più generali, ogni variabile osservata dipende in media da una variabile non osservata η (il costrutto che si vuole misurare) e un coefficiente λ_j che quantifica il grado con cui la variabile osservata è associata al misurando, a cui si aggiunge una componente d'errore residuale δ_j . L'aggettivo unidimensionale è qui giustificato dal fatto che η non ha pedice ed è costante per $j = 1, \dots, p$. Si noti l'analogia tra il modello di regressione lineare semplice $Y = X_j\beta_j + \epsilon_j$ e il modello CFA: al posto di X_j vi è η , al posto del coefficiente angolare β si ha λ e al posto di ϵ_j si ha δ_j . Stressando infatti l'analogia si potrebbe sostenere, pur con qualche incautezza, che la CFA sia un modello di regressione lineare i cui predittori

non sono osservati. Considerando la j -esima variabile, notiamo quanto segue [8]:

$$\begin{aligned} X_j &= \underbrace{\eta\lambda_j}_{\text{parte comune } c} + \underbrace{\delta_j}_{\text{parte unica } u_j} \\ &= c + u_j \end{aligned}$$

dove c è la componente del modello *in comune* con le p osservabili mentre u_j è invece quella componente *unica* associata a ciascuna delle p osservabili. In ciò, il modello esprime il fatto che X_1, \dots, X_p siano linearmente dipendenti da una parte comune c (pesata secondo $\lambda_1, \dots, \lambda_p$) ed una parte unica/residuale u_j specifica per ciascuna di esse. In questa definizione, che ricalca quella della TCT quando $\boldsymbol{\lambda} = \mathbf{1}_p$, l'osservabile è in dipendenza lineare dal misurando e non viceversa: tale modello si dice **riflessivo**.¹ La figura 5.1 mostra una tipica rappresentazione grafica di un modello CFA riflessivo unidimensionale.

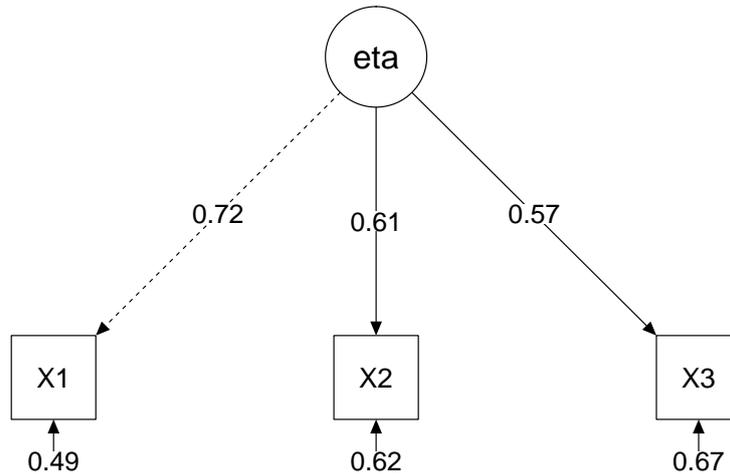


FIGURA 5.1: *Modello CFA unidimensionale: esempio di modello fattoriale (scala) unidimensionale con una sola variabile latente (misurando) e tre variabili osservate (indicatori). Nota: le variabili latenti sono rappresentate mediante cerchi mentre quelle osservate mediante quadrati, i numeri posti lungo le frecce orientate sono i coefficienti $\lambda_{X1}, \lambda_{X2}, \lambda_{X3}$ mentre i numeri posti lungo le frecce orientate sui quadrati indicano i residui del modello $\delta_{X1}, \delta_{X2}, \delta_{X3}$.*

Il modello fattoriale confermativo decompone la varianza osservata degli indicatori manifesti in due parti:

$$\text{Var}[X_j] = \text{Var}[c] + \text{Var}[\delta_j]$$

dove la frazione di varianza spiegata dalla variabile latente è data da $h^2 = \frac{\text{Var}[c]}{\text{Var}[X_j]}$ (comunalità) mentre la frazione di varianza rimanente dopo l'introduzione di η , vale a dire $u^2 = \text{Var}[\delta_j]$, indica la quota che necessita ancora di essere spiegata. Se le variabili

¹Sulla differenza tra *modelli riflessivi* e *modelli formativi* nel più generale contesto dei modelli ad equazioni strutturali (SEMs) si veda [36].

osservate sono standardizzate si ha che $h^2 + u^2 = 1$. Si osservi che il h^2 rivela quanto di X_j è spiegato da η : più tale rapporto è alto, maggiore sarà la variabilità spiegata da η (l'importanza di tale aspetto risulterà più chiara nel contesto dell'attendibilità di una scala).

5.1.2 Modello con più variabili latenti

Il modello fattoriale confermativo precedente è generalizzabile al caso con q variabili latenti (dove vale $q < p$ sempre):

$$\begin{aligned} X_1 &= \eta_1 \lambda_{11} + \dots + \eta_k \lambda_{k1} + \dots + \eta_q \lambda_{q1} + \delta_1 \\ &\vdots \\ X_j &= \eta_1 \lambda_{1j} + \dots + \eta_k \lambda_{kj} + \dots + \eta_q \lambda_{qj} + \delta_j \\ &\vdots \\ X_p &= \eta_1 \lambda_{p1} + \dots + \eta_k \lambda_{kp} + \dots + \eta_q \lambda_{qp} + \delta_p \end{aligned}$$

dove η_1, \dots, η_q sono i misurandi del modello. Il modello è multidimensionale nelle variabili latenti poiché vi sono più variabili latenti per ciascuna osservabile. Anche nel modello multidimensionale, vale la decomposizione della varianza in comunaltà ed unicità:

$$\begin{aligned} X_j &= \underbrace{\eta_1 \lambda_{1j} + \dots + \eta_k \lambda_{kj} + \dots + \eta_q \lambda_{qj}}_{\text{parte comune } c} + \underbrace{\delta_j}_{\text{parte unica } u_j} \\ &= c + u_j \end{aligned}$$

dove c in questo caso raccoglie la componente del modello lineare riferita ai fattori latenti η_1, \dots, η_q . Il modello con q misurandi può essere scritto in forma compatta usando l'algebra lineare:

$$\mathbf{x}_{p \times 1} = \mathbf{\Lambda}_{p \times q} \boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{p \times 1}$$

dove in questo caso la matrice $\mathbf{\Lambda}_{p \times q}$ contiene i coefficienti che congiungono le variabili manifeste (in riga) alle variabili latenti (in colonna). La figura 5.2 mostra una tipica rappresentazione grafica di un modello CFA riflessivo con $q = 2$ che è rappresentato utilizzando la rappresentazione algebrica che segue:

$$\mathbf{x}_{6 \times 1} = \mathbf{\Lambda}_{6 \times 2} \boldsymbol{\eta}_{2 \times 1} + \boldsymbol{\delta}_{6 \times 1}$$

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}}_{\mathbf{\Lambda}} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}}_{\boldsymbol{\delta}}$$

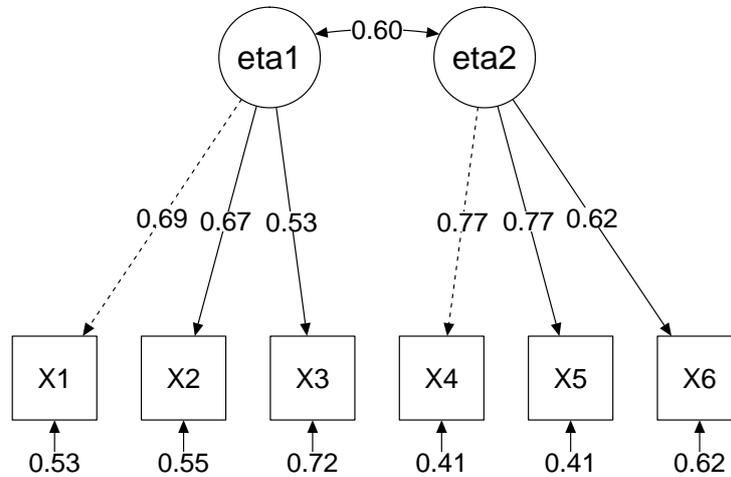


FIGURA 5.2: Modello CFA multidimensionale: esempio di modello fattoriale con due variabili latenti (misurandi) e sei variabili osservate (indicatori). Il grafico riporta due misurandi quantificati in modo unidimensionale: ciascuna delle due scale ha uno (ed uno solo) misurando. Nota: l'arco tra i due cerchi indicanti le variabili latenti rappresenta la correlazione tra i due misurandi.

L'esempio mostra un caso con $p = 6$ variabili osservate e $q = 2$ misurandi latenti. La matrice Λ è una matrice 6×2 dove in riga abbiamo le variabili osservate mentre in colonna le rispettive variabili latenti. All'interno della matrice il valore zero indica che la variabile osservata non ha un legame con la latente (*parametro fissato*); al contrario, i valori della matrice diversi da zero indicano che le osservate hanno un legame con le latenti (*parametro libero*). La figura 5.3, invece, mostra una variazione del modello precedente dove X_3 riflette contemporaneamente i misurandi η_1 e η_2 : la variabile osservata è *condivisa* da entrambi i misurandi, una violazione dell'assunzione della TCT di unidimensionalità della scala. La notazione algebrica del modello di Figura 5.3 è la stessa di quello precedente eccetto per la matrice Λ :

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}}_{\boldsymbol{\delta}}$$

Come notiamo dalla rappresentazione matriciale, il contributo dell'osservata X_3 rispetto alle variabili latenti è dato dalla combinazione lineare $\lambda_{31}\eta_1 + \lambda_{32}\eta_2$. Come conseguenza, parte della varianza di X_3 sarà attribuita sia a η_1 sia a η_2 . Si noti infine che un modello siffatto introduce un parametro libero in più (nello specifico, λ_{31}) rispetto a quello precedente.

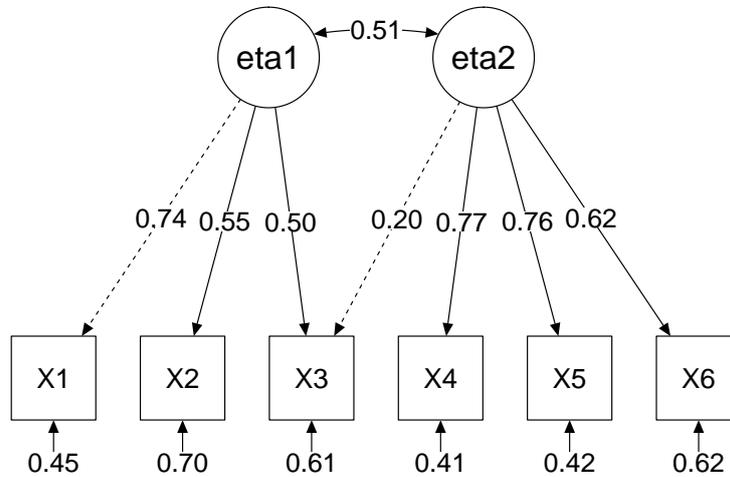


FIGURA 5.3: *Modello CFA multidimensionale: esempio di modello fattoriale con due variabili latenti (misurandi) e sei variabili osservate (indicatori). Il grafico riporta due misurandi quantificati in modo non unidimensionale: alcuni indicatori (in questo caso X_3) sono condivisi da più di un misurando e la scala non rispetta l'ipotesi della TCT di unidimensionalità. Ciò può essere dovuto, ad esempio, ad una non ottimale scelta degli indicatori (bassa coerenza interna) oppure al fatto che i due misurandi siano espressione di un unico misurando di secondo livello.*

In generale, l'azione del ricercatore sulla matrice dei coefficienti Λ permette di scrivere un modello di analisi fattoriale esplorativa (EFA) partendo dalla CFA: liberando infatti tutti i coefficienti di legame, il modello cessa di esprimere qualunque ipotesi strutturale (o ipotesi di misurazione) sulle variabili osservate. La rappresentazione algebrica che ne consegue è:

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}}_{\boldsymbol{\delta}}$$

graficamente rappresentata dal modello in Figura 5.4.

La rappresentazione algebrica vista sinora può essere agevolmente estesa ad includere il caso con p variabili manifeste e q variabili latenti, come segue:

$$\mathbf{x}_{p \times 1} = \Lambda_{p \times q} \boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{p \times 1}$$

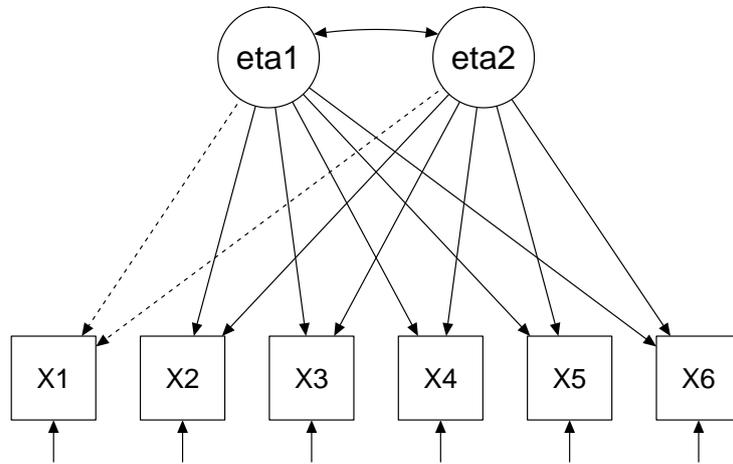


FIGURA 5.4: Modello EFA multidimensionale: generalizzazione del modello confermativo al modello fattoriale esplorativo tramite manipolazione della matrice Λ . Notiamo come il modello non esprima nessuna ipotesi specifica circa i legami tra le osservabili e i misurandi mentre, al contrario, tutte le osservate sono legate a tutte le variabili latenti. Tale modello, sebbene utile in alcuni contesti, non è un buon modello nel senso della TCT ed apporta scarso valore interpretativo.

$$\underbrace{\begin{bmatrix} X_1 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1k} & \dots & \lambda_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{j1} & \lambda_{j2} & \dots & \lambda_{jk} & \dots & \lambda_{jq} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pk} & \dots & \lambda_{pq} \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_k \\ \vdots \\ \eta_q \end{bmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_j \\ \vdots \\ \delta_p \end{bmatrix}}_{\boldsymbol{\delta}}$$

Come per i casi precedenti, la matrice Λ , altrimenti detta *factor pattern matrix*, è molto importante poiché esprime le ipotesi del ricercatore circa la formazione delle scale di misura (essa contiene difatti i coefficienti che legano le osservabili ai misurandi). Si noti che la decomposizione della varianza in comunaltà e unicità vale anche nel caso suddetto.

5.1.3 Assunzioni del modello

Il modello lineare dell'analisi fattoriale confermativa

$$\mathbf{x}_{p \times 1} = \Lambda_{p \times q} \boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{p \times 1}$$

assume quanto segue [8]:

- (i) $\mathbb{E}[X_j] = 0$

- (ii) $\mathbb{E}[\eta_k] = 0$
- (iii) $\mathbb{E}[\delta_j] = 0$
- (iv) $\text{Cov}[\eta_k, \delta_j] = 0$

dove (i)-(ii) indicano rispettivamente che i valori attesi delle v.a. manifeste e latenti sono pari a zero, (iii) traduce l'ipotesi che il valore atteso delle v.a. errori sia uguale a zero (si veda la *legge degli errori accidentali*), infine (iv) rappresenta il vincolo secondo cui le v.a. degli errori e dei misurandi siano incorrelati. La necessità di tali assunti sarà più chiara successivamente, allorché si discuterà della stima dei parametri del modello fattoriale confermativo. Per ora basti pensare che questi riflettono ipotesi del ricercatore circa la componente stocastica del modello, più nello specifico circa le leggi probabilistiche che governano il comportamento delle variabili osservate, delle variabili manifeste e degli errori.

5.1.4 Matrice di covarianza implicata dal modello

Nell'introduzione a questo capitolo si è fatto riferimento al fatto che la CFA utilizzi la matrice di covarianza $\text{Cov}[\mathbf{X}]$ quale dato di input al posto della matrice delle variabili osservate $\mathbf{X}_{n \times p}$. Data la decomposizione $\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta}$ alla base della CFA possiamo scrivere la matrice di covarianza *riprodotta* (o implicata) dal modello CFA $\boldsymbol{\Sigma}_X$, dove ciascun elemento di essa (vale a dire le varianze/covarianze) è espresso analiticamente in termini di parametri da determinare (*stimare*) mediante un metodo opportuno. Utilizzando la definizione nota di covarianza e quella del modello CFA:

$$\begin{aligned}\boldsymbol{\Sigma}_X &= \text{Cov}[(\mathbf{\Lambda}_{p \times q}\boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{q \times 1}), (\mathbf{\Lambda}_{p \times q}\boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{q \times 1})] \\ &= \mathbb{E}[(\mathbf{\Lambda}_{p \times q}\boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{q \times 1})(\mathbf{\Lambda}_{p \times q}\boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{q \times 1})^T]\end{aligned}$$

e usando gli assunti (i)-(iv) precedenti, dopo alcune manipolazioni algebriche otteniamo agevolmente $\boldsymbol{\Sigma}_X$ come segue:

$$\begin{aligned}\boldsymbol{\Sigma}_X &= \mathbf{\Lambda}_{p \times q} (\text{Cov}[\boldsymbol{\eta}_{q \times 1}, \boldsymbol{\eta}_{q \times 1}]) \mathbf{\Lambda}_{p \times q}^T + \text{Cov}[\boldsymbol{\delta}_{q \times 1}, \boldsymbol{\delta}_{q \times 1}] \\ &= \mathbf{\Lambda}_{p \times q} \boldsymbol{\Phi}_{q \times q} \mathbf{\Lambda}_{p \times q}^T + \boldsymbol{\Theta}_{\delta_{p \times p}}\end{aligned}$$

dove la matrice $\boldsymbol{\Phi}_{q \times q}$ esprime la correlazione tra i misurandi latenti, $\boldsymbol{\Theta}_{\delta_{p \times p}}$ indica la matrice di correlazione tra gli errori mentre l'esponente T indica l'operatore di trasposizione di una matrice.² Facendo riferimento nuovamente all'esempio riportato in Figura

²Si ricordi quanto segue: $\mathbf{x}_{p \times 1}$ è un vettore colonna costituito da p elementi, $\boldsymbol{\eta}_{q \times 1}$ è un vettore colonna contenente q elementi, $\mathbf{\Lambda}_{p \times q}$ è una matrice con p righe e q colonne, $\boldsymbol{\delta}_{p \times 1}$ è un vettore contenente p elementi. L'operazione $\boldsymbol{\eta}_{q \times 1} \mathbf{\Lambda}_{p \times q}$ è errata in quanto il prodotto tra matrici (prodotto righe \times colonne) richiede che il numero di righe della matrice che pre-moltiplica sia pari al numero di colonne del secondo termine del prodotto. In generale, il risultato dell'operazione $\mathbf{\Lambda}_{p \times q} \boldsymbol{\eta}_{q \times z}$ è una matrice di ordine $p \times z$. Per questo motivo nella definizione di $\boldsymbol{\Sigma}_X$ compare l'operatore di trasposizione, per rendere possibile i prodotti tra matrici.

5.2, il modello CFA riscritto in termini di covarianza attesa è il seguente:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.69 & 0 \\ 0.67 & 0 \\ 0.53 & 0 \\ 0 & 0.77 \\ 0 & 0.77 \\ 0 & 0.62 \end{bmatrix} \quad \mathbf{\Phi} = \begin{bmatrix} 1 & 0.60 \\ 0.60 & 1 \end{bmatrix} \quad \mathbf{\Theta}_{\delta} = \begin{bmatrix} 0.53 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.72 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.41 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.62 \end{bmatrix}$$

dove, come in precedenza, i valori nulli all'interno delle matrici indicano parametri fissi del modello, quelli ossia relativi alle ipotesi del ricercatore circa la parte strutturale del modello. In questo caso, inoltre, la matrice degli errori $\mathbf{\Theta}_{\delta_{p \times p}}$ è *diagonale* e possiede solo valori non nulli sulla diagonale principale (varianze delle componenti di errore) mentre i valori nulli fuori dalla diagonale indicano che gli errori sono incorrelati, in accordo all'assunto (iii) del modello.

Raccordo tra TCT e CFA

Le simmetrie tra CFA e TCT sono state ampiamente descritte dalla letteratura psicometria [22]. In generale, possiamo notare come le decomposizioni lineari alla loro base siano le stesse, come anche alcuni degli assunti (ad esempio quello concernente la correlazione tra gli errori). Sebbene non esprimano esattamente la stessa cosa, la CFA permette di costruire e valutare una scala di misura mediante un modello lineare che lega variabili manifeste a variabili latenti. In ciò, misurare secondo la CFA implica la definizione e stima di un modello confermativo come quelli riportati in precedenza.

TCT	CFA
$\mathbb{E}[E_j] = 0$	$\mathbb{E}[\delta_j] = 0$
$\text{Cor}[E_j, T] = 0$	$\text{Cor}[\eta_j, \delta_j] = 0$
$\text{Var}[X_j] = \text{Var}[T] + \text{Var}[E_j]$	$\text{Var}[X_j] = \lambda^2\phi + \theta_\delta$

TABELLA 5.1: Alcune corrispondenze tra CFA e TCT rispetto ai loro assunti di base. Nota: L'espressione $\text{Var}[X_j]$ è ottenuta dalla definizione generale $\text{Var}[X_j] = \text{Var}[\eta_j] + \text{Var}[\delta_j]$ notando che $\text{Var}[\delta_j] = \theta_\delta$ e $\text{Var}[\eta_j] = \lambda\phi\lambda = \lambda^2\phi$.

Un particolare punto di incontro tra CFA e TCT può essere rintracciato nella definizione di attendibilità di una scala. Mentre nella TCT si è opportunamente definita la misura di attendibilità utilizzando misure osservabili parallele, sfruttando dunque l'informazione presente nei dati, nella CFA l'attendibilità è invece definita utilizzando il modello stesso, come espresso dalla definizione che segue:

$$\rho_{XT}^2 = \frac{\lambda^2 \text{Var}[\eta]}{\text{Var}[X]} = \frac{\lambda^2 \phi}{\lambda^2 \phi + \theta_\delta}$$

da cui si può dedurre l'uguaglianza per quanto riguarda la componente vera della TCT $T = \lambda\phi$. Si noti come, nel caso $\lambda = 1$, la definizione dell'attendibilità secondo la CFA uguaglia quella fornita dalla TCT:

$$\rho_{XT}^2 = \frac{\text{Var}[\eta]}{\text{Var}[X]} = \frac{\phi}{\phi + \theta_\delta}$$

dove quest'ultima formulazione richiama quella della *correlazione intraclassa* utilizzata nel capitolo precedente. Dalla definizione di attendibilità deriva anche una formulazione operativa che ne permette agevolmente il calcolo, vale a dire l'indice ω definito come segue:

$$\omega_{XT} = \frac{\sum_{j=1}^p \lambda_j^2}{\sum_{j=1}^p \lambda_j^2 + \theta_{\delta_j}^2}$$

che, come si può notare, utilizza direttamente i parametri del modello CFA una volta che questi siano stati determinati numericamente. L'indice si interpreta in maniera analoga all' α di Cronbach (si veda § 4.8.1). In particolari casi, come ad esempio quello rappresentato dalle misure τ -equivalenti, questi indici si equivalgono, facendo meglio emergere la simmetria esistente tra TCT e CFA (per un maggiore dettaglio si veda [2]).

Derivazione di $\text{Var}[X_j]$ nella CFA (approfondimento)

Si consideri la Tabella 5.1 delle relazioni tra CFA e TCT. L'espressione $\text{Var}[X_j]$ è ottenuta come di seguito, dopo aver notato che $\text{Var}[\mathbf{X}] = \text{diag}(\text{Cov}[\mathbf{X}])$:

$$\begin{aligned} \text{diag}(\text{Cov}[\mathbf{X}]) &= \text{diag}(\mathbb{E}[(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})(\mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\delta})^T]) \quad \text{definizione di covarianza} \\ &\dots\text{svolgendo il prodotto:} \\ &= \text{diag}(\mathbb{E}[\mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{\Lambda} + \boldsymbol{\delta}\boldsymbol{\eta}^T\mathbf{\Lambda}^T + \mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\delta}^T + \boldsymbol{\delta}\boldsymbol{\delta}^T]) \\ &\dots\text{per la linearità del valore atteso } \mathbb{E}[\cdot]: \\ &= \text{diag}(\mathbb{E}[\mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{\Lambda}] + \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\eta}^T\mathbf{\Lambda}^T] + \mathbb{E}[\mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\delta}^T] + \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T]) \\ &\dots\text{riprendendo gli assunti della CFA:} \\ &= \text{diag}(\mathbb{E}[\mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{\Lambda}] + \cancel{\mathbb{E}[\boldsymbol{\delta}\boldsymbol{\eta}^T\mathbf{\Lambda}^T]}^0 + \cancel{\mathbb{E}[\mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\delta}^T]}^0 + \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T]) \\ &= \text{diag}(\mathbb{E}[\mathbf{\Lambda}\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{\Lambda}] + \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T]) \\ &\dots\text{poiché } \mathbf{\Lambda} \text{ è una costante/parametro e le sole componenti aleatorie sono } \boldsymbol{\eta} \text{ e } \boldsymbol{\delta}, \text{ portiamo fuori dal valore atteso la costante:} \\ &= \text{diag}(\mathbf{\Lambda}\mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^T]\mathbf{\Lambda} + \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T]) \\ &\dots\text{notiamo che } \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}^T] = \text{Cov}[\boldsymbol{\eta}] := \boldsymbol{\phi}: \\ &= \text{diag}(\mathbf{\Lambda}\boldsymbol{\phi}\mathbf{\Lambda} + \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T]) \\ &\dots\text{notiamo ancora che } \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^T] = \text{Cov}[\boldsymbol{\delta}] := \boldsymbol{\Theta}_{\delta}: \\ &= \text{diag}(\mathbf{\Lambda}\boldsymbol{\phi}\mathbf{\Lambda} + \boldsymbol{\Theta}_{\delta}) \\ &\dots\text{applicando l'operatore } \text{diag}(\cdot) \text{ prendiamo gli elementi } j = 1, \dots, p \text{ della diagonale della matrice:} \\ &= \lambda_j\phi_j\lambda_j + \theta_{\delta_j} \\ &= \lambda_j^2\phi_j + \theta_{\delta_j} \quad \text{con } j \in (1, \dots, p) \end{aligned}$$

Dal risultato ottenuto possiamo scrivere:

$$\text{Var}[X_j] = \text{diag}(\text{Cov}[\mathbf{X}])_j = \lambda_j^2\phi_j + \theta_{\delta_j}$$

dove gli addendi risultano rispettivamente:

$$\lambda_j^2\phi_j := \text{Var}[\eta_j] \quad \text{e} \quad \theta_{\delta_j} := \text{Var}[\delta_j]$$

5.1.5 CFA in sintesi

Si riporta in Tabella 5.2 una sintesi del modello fattoriale confermativo.

$\mathbf{x}_{p \times 1} = \mathbf{\Lambda}_{p \times q} \boldsymbol{\eta}_{q \times 1} + \boldsymbol{\delta}_{p \times 1}$
(i) $\mathbb{E}[X_j] = 0$
(ii) $\mathbb{E}[\eta_k] = 0$
(iii) $\mathbb{E}[\delta_j] = 0$
(iv) $\text{Cov}[\eta_k, \delta_j] = 0$
(v) $\text{Cov}[\mathbf{X}] = \boldsymbol{\Sigma}_X = \mathbf{\Lambda} \boldsymbol{\Phi} \mathbf{\Lambda}^T + \boldsymbol{\Theta}_\delta$
(vi) comunaltà: $\text{Var}[c_j] = \sum_{k=1}^q \lambda_{jk}^2$
(vii) unicità: $\text{Var}[\delta_j] = \theta_{\delta_j}$

TABELLA 5.2: Sintesi della modello CFA. Nota: la definizione di comunaltà qui riportata richiede che le variabili latenti siano incorrelate.

5.2 Stima dei parametri del modello

Il modello CFA è parametrizzato dalle seguenti componenti:

- la matrice $\mathbf{\Lambda}$ contiene i coefficienti che legano le osservabili alle latenti;
- la matrice $\boldsymbol{\Phi}$ contiene le varianze-covarianze delle variabili latenti quando $q > 1$;
- la matrice $\boldsymbol{\Theta}_\delta$ contiene le varianze-covarianze dei residui del modello.

L'approccio confermativo, in virtù del fatto che forza il ricercatore ha definire delle ipotesi sul modello di misura, non necessità di stimare le matrici $\mathbf{\Lambda}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}_\delta$ nella loro interezza ma, al contrario, richiede di determinare solo le componenti di queste che risultano diverse da zero (c.d. *parametri liberi*). Una volta stimati i parametri, rispetto ad un campione casuale ottenuto da una popolazione di riferimento, è possibile calcolare la matrice $\hat{\boldsymbol{\Sigma}}_X$, ovvero la matrice riprodotta dal modello (si veda § 5.1.4) che può essere studiata rispetto alla matrice di covarianze osservata \mathbf{S}_X (ad esempio, come si vedrà successivamente, per valutare la bontà di adattamento del modello ai dati). In generale, un buon modello confermativo è *parsimonioso* nel senso che presenta pochi parametri liberi rispetto al numero di parametri fissi: questo si realizza, appunto, nella definizione di un modello che sappia ben rappresentare le ipotesi del ricercatore circa la struttura di misurazione di un test [8].³ I **gradi di libertà** vengono utilizzati per quantificare

³La parsimoniosità, oltre ad essere un criterio estetico molto simile a quelli usati per valutare teorie scientifiche come la semplicità e l'eleganza [27], nel contesto statistico ha a che fare anche con la possi-

la parsimoniosità del modello, la sua semplicità strutturale. Nel caso della CFA si esprimono come di seguito:

$$df = \frac{p(p+1)}{2} - m \quad (5.1)$$

dove p indica il numero di osservabili del modello mentre m indica il corrispondente numero di parametri liberi. Questi ultimi - esprimono le ipotesi che il ricercatore pone sulle relazioni ammissibili tra le variabili manifeste e quelle latenti - possono essere fissati, vale a dire *vincolati*, almeno per due motivi: (i) rispecchiano appunto le ipotesi strutturali del ricercatore, (ii) servono per ridurre la quantità di parametri da stimare/-determinare numericamente: se $m = 0$ non ci sarebbero dati a sufficienza per stimare i parametri del modello e, conseguentemente, il modello diverrebbe *non identificabile*. Ad esempio, richiamando ancora il modello espresso nella Figura 5.2, abbiamo le seguenti matrici di parametri:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{12} \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} 1 & \phi_{12} \\ \phi_{21} & 1 \end{bmatrix}, \quad \mathbf{\Theta}_\delta = \begin{bmatrix} \theta_{\delta_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\delta_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{\delta_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{\delta_4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{\delta_5} & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_6} \end{bmatrix}$$

dove si evince chiaramente che $m = 13$ mentre $\frac{p(p+1)}{2} = 21$ con un totale di $df = 8$ gradi di libertà. Per facilitare l'identificabilità del modello - e dunque la possibilità di determinare univocamente i parametri ignoti del modello - si pone spesso il problema della c.d. scelta della **metrica delle variabili latenti** [8]. Due sono gli approcci usualmente utilizzati:

- approccio ULI (*unit loading identification*): si fissa ad 1 il primo dei coefficienti $\lambda_1, \dots, \lambda_p$ per ciascuna delle q variabili latenti. Nell'esempio di Figura 5.2, $\lambda_{11} = 1$ mentre $\lambda_{21}, \lambda_{31}$ si lasciano liberi di essere stimati;
- approccio UVI (*unit variance identification*): si fissa ad 1 la varianza delle variabili latenti ϕ_1, \dots, ϕ_q . Nell'esempio della Figura 5.2, $\phi_{\eta_1} = \phi_{\eta_2} = 1$.

Si noti che i due approcci sono mutualmente esclusivi. La stima dei parametri avviene attraverso una procedura iterativa che cerca di minimizzare la distanza tra la matrice osservata \mathbf{S}_X e la matrice di covarianza implicata dal modello $\mathbf{\Sigma}_X$: il modello riproduce meglio i dati osservati quando la distanza è minore. Questa procedura può essere attuata mediante diversi approcci, ad esempio quello della *massima verosimiglianza* (ML) o dei *minimi quadrati generalizzati* (GLS, DWLS). Si tralasciano in questa sede i dettagli tecnici della procedura di stima (per un approfondimento si rimanda, tra gli altri, a [8, 11]). Alla fine della procedura di stima i parametri stimati sono i seguenti:

- $\hat{\mathbf{\Lambda}}$ matrice dei coefficienti del modello fattoriale (c.d. *factor loadings*);
- $\hat{\mathbf{\Phi}}$ matrice delle varianze-covarianze tra i fattori latenti;

bilità/limite di determinare numericamente i parametri del modello dato un campione di dati, almeno nell'approccio basato sulla massima verosimiglianza [29].

- $\hat{\Theta}_\delta$ matrice delle varianze-covarianze degli errori del modello.

Spesso, per agevolare l'interpretazione dei risultati, tali stime vengono standardizzate in modo che siano espressi entro una scala interpretabile. Nel caso di $\hat{\lambda}_1, \dots, \hat{\lambda}_p$, ad esempio, questi vengono standardizzati su una scala compresa nell'intervallo $[-1, 1]$. Facendo ancora riferimento all'esempio di Figura 5.2, le stime dei parametri risultano i seguenti:

$$\hat{\Lambda} = \begin{bmatrix} 0.69 & 0 \\ 0.67 & 0 \\ 0.53 & 0 \\ 0 & 0.77 \\ 0 & 0.77 \\ 0 & 0.62 \end{bmatrix}, \quad \hat{\Phi} = \begin{bmatrix} 1 & 0.60 \\ 0.60 & 1 \end{bmatrix}, \quad \hat{\Theta}_\delta = \begin{bmatrix} 0.53 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.72 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.41 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.62 \end{bmatrix}$$

dove, nello specifico, le matrici \mathbf{S}_X e $\hat{\Sigma}_X$ sono:

$$\mathbf{S}_X = \begin{bmatrix} 1.00 & & & & & & \\ 0.43 & 1.00 & & & & & \\ 0.41 & 0.35 & 1.00 & & & & \\ 0.28 & 0.35 & 0.16 & 1.00 & & & \\ 0.32 & 0.32 & 0.19 & 0.59 & 1.00 & & \\ 0.24 & 0.32 & 0.18 & 0.47 & 0.46 & 1.00 & \end{bmatrix}, \quad \hat{\Sigma}_X = \begin{bmatrix} 1.00 & & & & & & \\ 0.46 & 1.00 & & & & & \\ 0.36 & 0.36 & 1.00 & & & & \\ 0.31 & 0.31 & 0.24 & 1.00 & & & \\ 0.31 & 0.31 & 0.24 & 0.59 & 1.00 & & \\ 0.25 & 0.25 & 0.19 & 0.47 & 0.47 & 1.00 & \end{bmatrix}$$

Si noti la prossimità numerica della matrice riprodotta dal modello alla matrice di covarianze osservata.

5.3 Valutazione del modello CFA

Dopo aver stimato i parametri del modello mediante un campione casuale estratto da una popolazione di riferimento, si pone il problema della *valutazione dell'adattamento* del modello ai dati. In breve, ciò esprime l'idea di quantificare quanto il modello stimato riesca a riprodurre la variabilità e le relazioni tra variabili presenti nei dati.⁴

5.3.1 Test del χ^2

Un modo per quantificare l'adattamento ai dati è dunque quello di studiare la differenza tra \mathbf{S}_X e Σ_X utilizzando il test del χ^2 dove una statistica opportuna $T = f(\mathbf{S}_X; \Sigma_X)$, funzione delle due matrici di covarianze, è studiata in distribuzione sotto la sua ipotesi nulla, ossia $\chi^2(df)$ dove df sono i gradi di libertà della legge Chi quadrato. L'ipotesi nulla codifica l'assenza di differenza tra \mathbf{S}_X e Σ_X . Ricordiamo che nel caso in cui $T_{oss} > \chi^2_{\alpha/2}$ (valore critico della statistica test) allora l'ipotesi nulla è rigettata e dunque il modello teorico implicato da Σ_X non si adatta bene ai dati osservati \mathbf{S}_X . Per affermare che il modello CFA si adatta bene ai dati, cercheremo di non rigettare l'ipotesi nulla. Si

⁴Per una trattazione più articolata dei temi riguardanti l'*adattamento ai dati* e la *capacità predittiva* di un modello, si veda [1]

tenga in debita considerazione che, come ogni decisione basata sulla statistica χ^2 , il test di ipotesi tende a rigettare spesso il modello teorico Σ_X (ossia $T_{oss} > \chi_{\alpha/2}^2$) quando la numerosità campionaria n è molto elevata. È utile dunque utilizzare altri indici per valutare l'adattamento del modello ai dati che sappiano tenere conto della parsimoniosità dei modelli che vengono di volta in volta testati.

5.3.2 RMSEA

È un indice che valuta in maniera assoluta il grado di non adattamento del modello ai dati:

$$\text{RMSEA} = \sqrt{\frac{T_n \text{df}}{\text{ndf}}}$$

e che permette di correggere la statistica test per la numerosità campionaria e per il numero dei gradi di libertà del modello. L'indice si esprime nell'intervallo $[0, \infty)$ con i seguenti valori soglia:

- $\text{RMSEA} < 0.05$ indica adattamento ottimo del modello ai dati;
- $0.05 \leq \text{RMSEA} < 0.1$ indica adattamento medio/scarso;
- $\text{RMSEA} > 0.10$ indica adattamento non adeguato del modello.

5.3.3 Altri indici

Altri indici vengono spesso utilizzati insieme e in maniera comparativa a RMSEA, tra cui CFI, TLI/NNFI, AGFI/GFI [8]. Questi ultimi classificano un modello come ottimo quando il loro valore supera la soglia convenzionalmente fissata a 0.95 mentre considerano un modello come accettabile quando il loro valore è invece superiore a 0.90. Un ulteriore modo per valutare l'adattamento del modello ai dati è quello di confrontare quest'ultimo con un modello di base (c.d. *null model*) e/o con più modelli annidati (c.d. *nested models*). L'indice per valutare la discrepanza tra modello teorico e modello nullo (o in generale modelli alternativi) è AIC (o BIC). Secondo questo criterio, il modello che meglio si adatta ai dati è quello che ottiene AIC/BIC più basso secondo la regola *the lower, the better*.

5.3.4 Indici di modifica

Gli indici di modifica valutano quanto è buona una soluzione di un modello CFA valutando come la statistica T_n cambierebbe se i parametri fissi (quelli non stimati) del modello venissero liberati (ossia stimati). Indici sensibilmente superiori al valore di quattro indicano che l'adattamento complessivo del modello migliorerebbe se quel parametro venisse incluso nel modello ipotizzato. La procedura si basa essenzialmente su una logica incrementale dove i parametri vengono liberati uno alla volta: il modello con il nuovo parametro incluso viene ri-stimato e gli indici di modifica vengono nuovamente calcolati e ciascuno di essi valutati secondo la regola precedente.

5.3.5 Analisi dei residui e correlazione tra gli errori

L'analisi dei residui si basa sull'osservazione della matrice $\mathbf{S}_X - \hat{\Sigma}_X$ alla ricerca di quei coefficienti Θ_δ che, rispetto a due variabili fissate, risultano elevati. Una volta individuato il residuo, si libera un parametro della matrice Θ_δ del modello e si ri-stima il modello con il nuovo parametro incluso. Ciò equivale a *correlare gli errori* delle variabili osservate. Ciò comporta un sensibile miglioramento dell'adattamento del modello ai dati. In generale, si consideri la matrice del modello Θ_δ che per errori non correlati si presenta come segue:

$$\Theta_\delta = \begin{bmatrix} \theta_{\delta_{11}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\delta_{11}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{\delta_{jj}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{pp}} \end{bmatrix}$$

dove gli unici parametri liberi sono quelli posti lungo la diagonale della matrice, vale a dire le varianze residue delle variabili osservate. Al contrario, liberare un parametro della matrice (ad esempio $\theta_{\delta_{2j}}$) equivale a correlare gli errori δ delle variabili osservate (nell'esempio, le variabili 2 e j):

$$\Theta_\delta = \begin{bmatrix} \theta_{\delta_{11}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\delta_{22}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & \theta_{\delta_{j2}} & 0 & \theta_{\delta_{jj}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{pp}} \end{bmatrix}$$

In termini qualitativi, ciò indica che le due osservabili esprimono una medesima informazione sul misurando η , convogliano quindi la stessa informazione. La Figura 5.5 mostra un esempio di modello confermativo con errori correlati, la cui matrice $\hat{\Theta}_\delta$ si presenta invece come di seguito:

$$\hat{\Theta}_\delta = \begin{bmatrix} 0.53 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & -0.03 & 0 & 0 & 0 \\ 0 & -0.03 & 0.72 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.41 & -0.10 \\ 0 & 0 & 0 & 0 & -0.10 & 0.62 \end{bmatrix}$$

Questo tipo di procedura, in generale, presenta due problemi. Il primo riguarda il fatto che è sempre possibile migliorare un modello correlando gli errori delle variabili osservate con il risultato di ottenere modelli con alto indice di adattamento ai dati (*fit* elevato) ma scarsa capacità predittiva. Il secondo problema riguarda, invece, il piano interpretativo: correlare gli errori tra variabili manifeste equivale ad ammettere espressamente un certo limite del modello così come è stato definito dal ricercatore: assumere errori correlati significa che le variabili utilizzate sono interscambiabili, cioè il test utilizzato è formato da indicatori ridondanti.

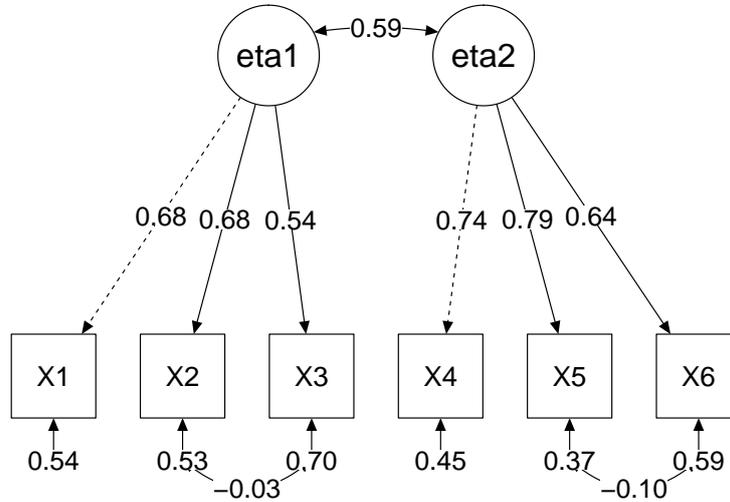


FIGURA 5.5: Esempio di modello confermativo con errori correlati: i parametri in oggetto sono $\theta_{\delta_{32}}$ e $\theta_{\delta_{56}}$.

5.4 Studio della dimensionalità della WISC-IV

Un'applicazione notevole per comprendere il modello CFA è quella che riguarda la quantificazione dell'*intelligenza* come effettuato dal test WISC-IV. Il modello teorico sotteso allo strumento comprende concettualizzazioni dell'intelligenza che la rappresentano in termini multifattoriali, con enfasi sul ragionamento fluido, memoria di lavoro e velocità di elaborazione degli stimoli. Da un punto di vista psicometrico, lo strumento quantifica il misurando intelligenza mediante la misura QI (quoziente intellettivo) totale mediante l'utilizzo di quattro scale: comprensione verbale (VCI), ragionamento percettivo (PRI), memoria di lavoro (WMI) e velocità di elaborazione (PSI), ciascuna delle quali è composta dai seguenti indicatori (o item):

- VCI: *similarities* (SO), *vocabulary* (VC), *comprehension* (CO)
- PRI: *block design* (DC), *picture concepts* (CI), *matrix reasoning* (RM)
- WMI: *digit span* (MC), *letter-number sequencing* (LN)
- PSI: *coding* (CR), *symbol search* (RS)

Le variabili osservate sono quantificate mediante svolgimento di compiti di performance cognitiva. Il nostro obiettivo, in questa sede, è quello di studiare lo strumento WISC-IV a partire dalla matrice di covarianza delle variabili osservate calcolata su un campione casuale di $n = 2200$ soggetti rappresentativo della popolazione normativa di riferimento per la taratura italiana:

$$\mathbf{S}_X = \begin{bmatrix} DC & SO & MC & CI & CR & VC & LN & RM & CO & RS \\ 1.00 & & & & & & & & & \\ 0.38 & 1.00 & & & & & & & & \\ 0.26 & 0.35 & 1.00 & & & & & & & \\ 0.34 & 0.43 & 0.28 & 1.00 & & & & & & \\ 0.25 & 0.14 & 0.15 & 0.11 & 1.00 & & & & & \\ 0.33 & 0.62 & 0.33 & 0.41 & 0.13 & 1.00 & & & & \\ 0.29 & 0.35 & 0.42 & 0.35 & 0.19 & 0.38 & 1.00 & & & \\ 0.42 & 0.41 & 0.29 & 0.43 & 0.20 & 0.40 & 0.35 & 1.00 & & \\ 0.27 & 0.51 & 0.24 & 0.35 & 0.15 & 0.59 & 0.30 & 0.30 & 1.00 & \\ 0.30 & 0.27 & 0.20 & 0.24 & 0.46 & 0.24 & 0.24 & 0.26 & 0.22 & 1.00 \end{bmatrix}$$

L'analisi della dimensionalità può giovare di analisi descrittive preliminari attraverso cui diventa possibile esplorare, visivamente e/o per mezzo di grafici, le relazioni che intercorrono tra le variabili osservate, alla ricerca di quelle che possono essere relazioni naturali di covariazione presente nei dati.⁵ Sebbene la struttura WISC-IV sia già nota ed ampiamente accettata in letteratura, di seguito ipotizzeremo diversi modelli di misurazione per la stessa matrice di covarianza \mathbf{S}_X e li valuteremo sulla base del loro adattamento ai dati. Ciò dovrebbe aiutare il lettore a comprendere meglio la logica che sottende la postulazione e la verifica di modelli di misurazione secondo l'ottica della CFA. L'adattamento dei diversi modelli ai dati campionari è effettuato mediante l'utilizzo della libreria `lavaan` (per maggiori dettagli si veda il laboratorio `lab9.R` associato a questo capitolo).

Modello 1

Il primo modello che analizzeremo è quello che ipotizza l'esistenza di un solo misurando, vale a dire una sola scala formata da tutte e dieci le variabili osservate presenti nella WISC-IV. La Figura 5.6 riporta il grafico del modello adattato mentre le matrici CFA stimate sono le seguenti:

$$\hat{\Lambda} = \begin{bmatrix} 1.00 \\ 1.01 \\ 0.85 \\ 0.70 \\ 0.79 \\ 0.78 \\ 0.64 \\ 0.72 \\ 0.37 \\ 0.54 \end{bmatrix} \quad \hat{\Phi} = [0.56] \quad \hat{\Theta}_\delta = \begin{bmatrix} 0.44 & & & & & & & & & & \\ 0.00 & 0.43 & & & & & & & & & \\ 0.00 & 0.00 & 0.59 & & & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.72 & & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.65 & & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.65 & & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.77 & & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.71 & & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.92 & & \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.84 & \end{bmatrix}$$

Si noti come i coefficienti rappresentati nel grafico sono standardizzati (variano nell'intervallo reale $[-1, 1]$) mentre quelli riportati nelle matrici sono invece assoluti (variano

⁵Quando la numerosità campionaria lo consente, è consigliato eseguire l'analisi mediante suddivisione del campione in due o più parti per evitare circolarità nel processo statistico decisionale [1].

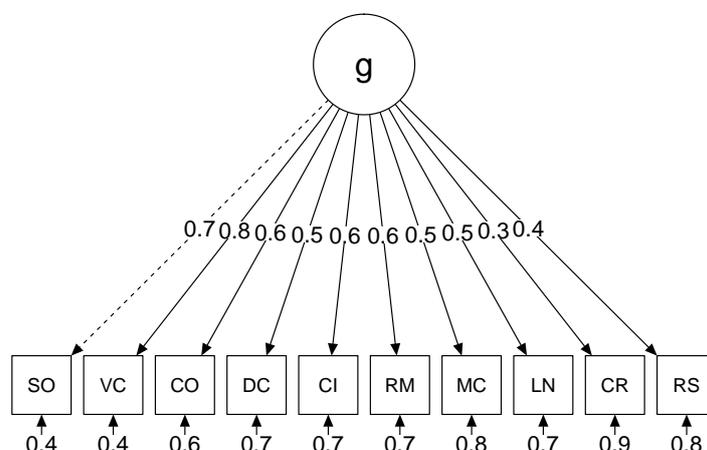


FIGURA 5.6: *WISC-IV: modello 1 con un solo fattore latente. Nota: la freccia tratteggiata indica che il parametro associato alla relazione tra variabile latente e manifesta è fissato ad uno secondo la metrica ULI. Indici di adattamento complessivi del modello: RMSEA = 0.11, CFI = 0.84, AIC = 57359.*

su tutto l'asse reale). La matrice Φ , in questo caso, contiene solo la varianza della variabile latente mentre Θ_{δ} è una matrice diagonale che riporta solo le varianze degli errori (in questo modello non abbiamo covarianze tra gli errori: fuori dalla diagonale infatti tutti i valori sono nulli). Per interpretare i risultati, si consideri a titolo esemplificativo, l'item CR: questi presenta una varianza d'errore pari a 0.9, un valore abbastanza alto. Ciò indica come tale indicatore presenti bassa comunaltà e alta unicità con il risultato che la variabile non concorre bene a quantificare il costrutto, nella terminologia della CFA *non lo riflette bene* (g non riesce difatti a cogliere bene la varianza di CR).⁶ Al contrario, le variabili SO e VC hanno alti valori di λ , per cui entrambi concorrono bene a quantificare g .⁷ Il modello adattato presenta i seguenti indici di adattamento: RMSEA = 0.11, CFI = 0.84, AIC = 57359. Complessivamente, considerando i valori degli indici di adattamento, il modello 1 non risulta soddisfacente.

Modello 2

Il secondo modello che analizzeremo è quello che ipotizza l'esistenza di tre misurandi, vale a dire una scala verbale (vr), una scala percettiva (pr) ed una riguardante l'elaborazione informazione (el). La Figura 5.7 riporta il grafico del modello adattato con le seguenti matrici CFA stimate:

⁶Si ricordi come il modello sia riflessivo: le frecce del grafico, difatti, sono orientate da g verso le osservabili e non viceversa.

⁷Si noti, a latere, che nel caso in cui si avesse una scala formata unicamente da tali misure componenti si avrebbe una misura composta con alta attendibilità.

a costruire la scala. Si noti, infine, l'alta correlazione tra le due scale ($\phi_{12} = 0.8$) ad indicare che le due scale non disambiguano bene il costrutto di intelligenza se usate separatamente. Per quanto riguarda gli indici di fit del modello osserviamo quanto segue rispetto al modello 1: RMSEA = 0.08, più basso del modello precedente e migliore, CFI è ora pari a 0.91, migliore del precedente sebbene sotto soglia mentre l'indice AIC è più basso di quello del modello precedente, a supporto del fatto che il modello corrente è migliore di quello precedente ad un solo fattore latente.

Modello 3

Il terzo modello specifica ulteriormente la scala dell'elaborazione cognitiva differenziandola in due scale differenti: memoria di lavoro (*wm*) e velocità dell'elaborazione (*psi*). Rispetto al modello precedente a tre fattori, gli indicatori *cr* e *rs* presentano valori di *factor loadings* più alti mentre si evidenzia una correlazione con la scala *psi* rispetto a quella del modello precedente *el*; la *varianza d'errore* risulta più bassa in valore nel caso di *cr* che sembra connettersi meglio alla scala *psi*. Gli indici di adattamento globale, in questo modello, migliorano notevolmente: RMSEA diminuisce ed arriva a 0.04, ottenendo quindi un valore sotto la soglia convenzionalmente fissata. L'indice CFI aumenta fino a 0.97, superando anch'esso il valore di soglia. L'indice AIC, rispetto al modello precedente, suggerisce di scegliere il modello corrente a quattro fattori rispetto a quello precedente. A completamento delle analisi di valutazione del modello, la Tabella 5.3 riporta gli indici di modifica per il modello calcolati mediante la libreria `lavaan` mentre la Tabella 5.4 mostra gli indici di attendibilità per le scale del modello.

lhs	op	rhs	mi	epc
PRI	=~	SO	44.66	0.53
VC	~~	CO	40.93	0.11
PSI	=~	DC	28.84	0.29
CI	~~	CR	21.42	-0.08
VCI	=~	CI	21.24	0.31
DC	~~	CR	19.66	0.07
PSI	=~	CI	13.60	-0.20
SO	~~	VC	13.40	-0.08
DC	~~	CI	13.13	-0.07
PRI	=~	VC	12.76	-0.30
...

TABELLA 5.3: *WISC-IV*: modello 3 a quattro fattori latenti. Estratto della tabella degli indici di modifica (*mi*) riportati secondo la sintassi `lavaan`: il simbolo =~ indica dipendenza mentre ~~ correlazione.

L'indice ω totale del test risulta alto, segnalando che il test con struttura a quattro fattori presenta alta precisione. Sebbene alcune scale, quando prese singolarmente, non non mostrano alta attendibilità, l'attendibilità totale risulta complessivamente più che accettabile. Dopo aver adattato il modello CFA ai dati, è possibile inoltre specializzare l'analisi e calcolare i valori dei misurandi $\hat{\eta}$ (c.d. *factor scores*) per ciascuna unità statistica su cui sono stati rilevati i dati. Nello specifico, è possibile calcolare i valori dei

	VCI	PRI	WMI	PSI	totale
α	0.80	0.66	0.59	0.63	0.82
ω	0.80	0.67	0.59	0.65	0.86

TABELLA 5.4: WISC-IV: modello 3 a quattro fattori latenti. Indici di attendibilità delle scale del modello complessivo.

costrutti $\hat{\eta}_{VCI}$, $\hat{\eta}_{PRI}$, $\hat{\eta}_{WMI}$, $\hat{\eta}_{PSI}$ che possono essere utilizzati in analisi successive, ad esempio per predire la performance scolastica dei soggetti sottoposti a WISC-IV come descritto dalla Figura 5.8.

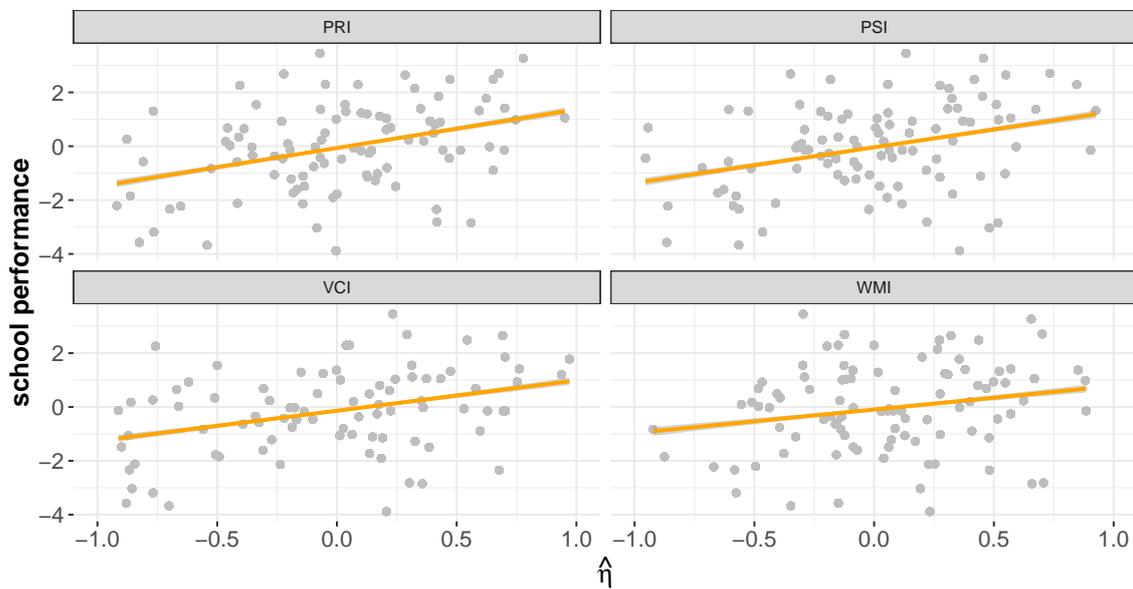


FIGURA 5.8: WISC-IV: modello 3 con quattro fattori latenti. Calcolo dei factor scores ed esempio del loro utilizzo nella previsione della performance scolastica.

Allo stesso modo, i *factor scores* possono essere impiegati per definire *profili* rispetto ad una variabile concomitante, ad esempio il successo scolastico (1=successo, 0=insuccesso) come rappresentato in Figura 5.9. Si noti come individui con alto successo scolastico presentino profili WISC-IV differenti da coloro che invece presentano basso successo scolastico.

Modello 4 (modello con fattori sovraordinati)

Il quarto modello introduce un fattore g (c.d. *intelligenza globale*) definito **fattore sovraordinato** che cattura parte covarianza espressa nella matrice Φ dei quattro fattori latenti del primo ordine.

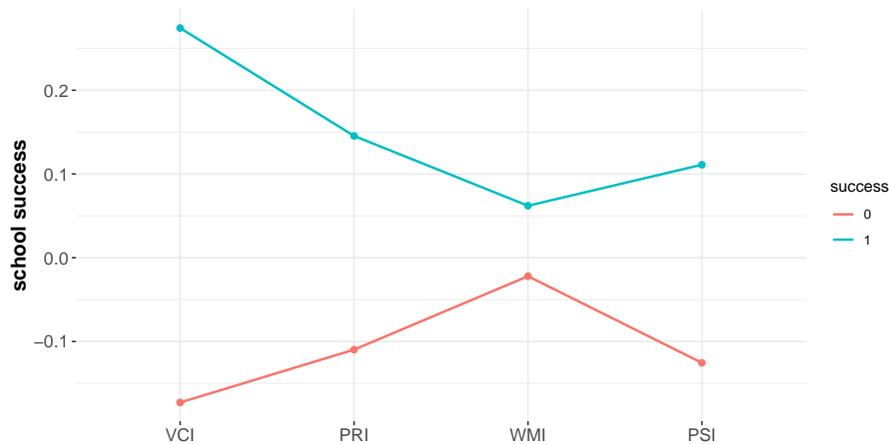


FIGURA 5.9: WISC-IV: modello 3 con quattro fattori latenti. Calcolo dei factor scores ed esempio del loro utilizzo per la creazione di profili di gruppo.

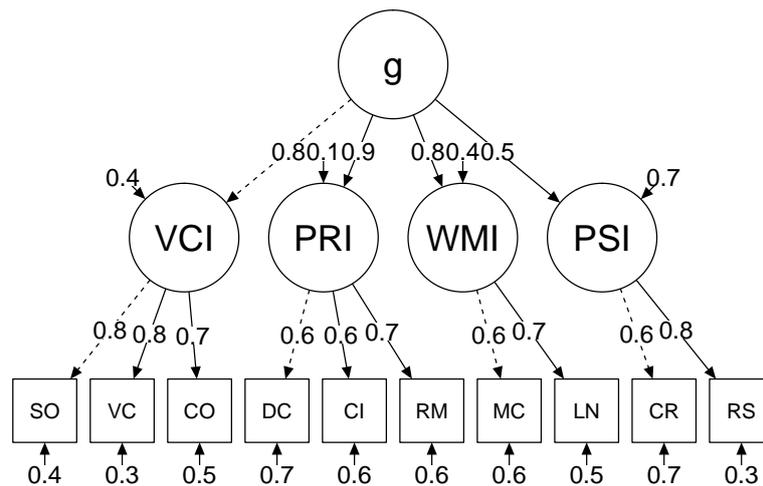


FIGURA 5.10: WISC-IV: modello 4 con fattori di primo e secondo livello. Nota: la freccia tratteggiata indica che il parametro associato alla relazione tra variabile latente e manifesta è fissato ad uno secondo la metrica ULI. Indici di adattamento complessivi del modello: RMSEA = 0.04, CFI = 0.98, AIC = 56553.

$$\hat{\Lambda} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.06 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.89 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.09 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.14 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.16 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.50 & 0.00 \end{bmatrix}$$

- o CFA e SEMs con lavaan:
https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_intro.pdf
- o CFA in a nutshell with lavaan:
https://www.yourpersonality.net/R/R_Notes_3.html
- o lavaan cheat sheet:
<https://github.com/jeromyanglim/lavaan-examples/blob/master/cheat-sheet-lavaan/cheat-sheet-lavaan.md>
- o lavaan google group (forum):
<https://groups.google.com/forum/#!forum/lavaan>

5.7 Laboratorio in R

I laboratori associati a questo capitolo sono i seguenti:

- o lab7.R: (A) Covarianza e correlazione, (B) Correlazione parziale, (C) Grafici per matrici di correlazione, (D) Comparare matrici di correlazione
- o lab8.R: (A) Sintassi di lavaan: modello e stime, (B) Sintassi lavaan: valutazione del modello
- o lab9.R: (A) WISC-IV: modelli e stime, (B) WISC-IV: confronto tra modelli
- o lab10.R: (A) BFI: dati e loro struttura, (B) BFI: esplorazione grafica struttura di correlazione, (C) BFI: contenuto degli items e soluzioni di clustering
- o lab11.R: (A) BFI: preparazione dati, (B) BFI: modelli e stime, (C) BFI: analisi dei profili

Il loro contenuto può essere visualizzato e scaricato all'indirizzo <https://github.com/antcalcagni/psychological-testing/tree/master/labs>.

6

Invarianza di misurazione

Bisogna sottolineare, però, che i confronti tra gruppi riguardo la media delle variabili misurate possono essere fatti solo se le intercette sono invarianti. In questo caso, le differenze tra i gruppi sono attribuibili unicamente alle differenze tra le medie dei fattori comuni.

Spesso nella ricerca applicata si è soliti confrontare strumenti di misura *tra gruppi*, ad esempio pazienti clinici vs. controllo, bambini con sviluppo atipico vs. bambini con sviluppo tipico, *o nel tempo*, ad esempio pazienti pre-trattamento vs. trattamento, pre-scolastico vs. post-scolastico, ecc. In tale contesto è legittimo chiedersi se e in quale modalità la **struttura fattoriale** di uno strumento rimane **invariante** tra gruppi o nel tempo. I test in psicologia sono formati da item costruiti per quantificare un fenomeno latente: per confrontare tra loro gruppi differenti, lo strumento deve misurare costrutti identici (stessa struttura fattoriale) tra popolazioni differenti. In questo caso, il test è detto invariante e i soggetti interpretano le domande (e il fattore latente) allo stesso modo. Quando c'è non invarianza, i gruppi rispondono in maniera diversa agli item e, conseguentemente, i misurandi non possono essere comparati. Joreskog [23] è stato il primo ad applicare l'analisi fattoriale confermativa (CFA) al problema dell'invarianza di misurazione. La CFA era già stata utilizzata come metodo per incorporare i vincoli e le ipotesi teoriche all'interno del modello del ricercatore, seguito poi dalla valutazione del suo adattamento complessivo ai dati. L'autore estese il metodo a modelli simultanei in gruppi multipli, dove diversi vincoli potevano essere posti ad ogni parametro da stimare.¹ Il primo passaggio da effettuarsi consisteva nell'adattamento ai dati di un modello privo di parametri fissati in modo tale da valutare la plausibilità del modello stesso. Verificato che l'adattamento della struttura fattoriale fosse accettabile, il passo successivo avrebbe incorporato i vincoli oggetto di interesse, tra i quali quelli inerenti i *factor loadings*. Questa *sequenza* può continuare fintantoché tutti i vincoli di invarianza sui parametri siano stati posti (ad eccezione dei *factor scores* che rimangono liberi di variare). In tal modo, dopo aver verificato l'esistenza di invarianza nelle strutture

Hanno contribuito alla realizzazione del presente capitolo:

Tania Di Tillio, Giorgia Facchin, Matteo Postal.

¹L'analisi dell'invarianza di misurazione è *de facto* un'analisi fattoriale confermativa multigruppo.

fattoriali in questione, sarebbe stato possibile effettuare i confronti tra gruppi rispetto alla media delle variabili osservate: le differenze tra i gruppi possono così essere attribuite unicamente alle differenze tra le medie dei fattori latenti. Da un punto di vista meramente tecnico, dunque, la valutazione dell'invarianza avviene attraverso una **procedura incrementale** dove un insieme di modelli CFA di tipo annidato (c.d. *nested models*) vengono valutati utilizzando gli indici di fit visti nel precedente capitolo, ossia χ^2 , RMSEA, CFI, AIC.² Tale procedura valuta quattro tipologie di invarianza, ordinate dalla più lasca alla più restrittiva, come segue:

1. configurale
2. debole
3. forte
4. esatta

Si noti che a ciascuna delle quattro tipologie di invarianza corrisponde uno specifico modello CFA. In termini più qualitativi, la procedura incrementale effettua una comparazione *a coppia* delle invarianze (vale a dire, debole vs. configurale, forte vs. debole, esatta vs. forte) e si arresta quando il confronto fallisce rispetto ai dati a disposizione. Ogni gruppo avrà il proprio modello fattoriale e ad ogni step viene introdotta una restrizione rispetto al modello fattoriale precedente. Questi stadi sequenziali corrispondono ai quattro livelli di invarianza precedentemente elencati. Un'ultimo aspetto prima di testare l'invarianza, in ogni caso, riguarda la qualità del campione di dati che è utilizzato: ad esempio, bisogna valutare se uno dei due gruppi presenta *outliers* rispetto all'altro. Gli studi di invarianza si basano sull'adattamento della matrice di covarianza osservata al modello del ricercatore, ragion per cui ogni bias presente nei gruppi confrontati avrà una certa influenza sui coefficienti fattoriali, le intercette e le varianze degli errori.

6.1 Schema generale di analisi dell'invarianza

Supponiamo, come in precedenza, di avere p variabili osservate su $g = 1, \dots, G$ gruppi distinti aventi n_g unità statistiche, con $q < p$ variabili latenti. Per semplicità di esposizione, fissiamo $G = 2$ e procediamo con due soli gruppi distinti. Consideriamo le $p \times p$ matrici di covarianza osservata per i due gruppi $\mathbf{S}_X^{(1)}$ e $\mathbf{S}_X^{(2)}$. Dati due generici e distinti livelli di invarianza, (1)-(2), l'analisi procede come segue:

- a) Si adattano due modelli CFA, \mathcal{M}_1 e \mathcal{M}_2 , per il primo livello di invarianza ottenendo i parametri CFA stimati per entrambi i modelli adattati
- b) Si calcola la statistica complessiva $\chi_0^2 = \chi_{\mathcal{M}_1}^2 + \chi_{\mathcal{M}_2}^2$ e i gradi di libertà df_0 (vedi slide 34)

²Un modello annidato è un modello che usa le stesse variabili di un altro modello ma specifica la stima di almeno un ulteriore parametro. Il modello con meno restrizioni e più parametri liberi e meno gradi di libertà, chiamato modello ridotto (*reduced model*), è annidato all'interno di un modello più ristretto, il cosiddetto modello completo (*full model*).

- c) Si adattano due modelli CFA, \mathcal{M}'_1 e \mathcal{M}'_2 , per il secondo livello di invarianza ottenendo i parametri CFA stimati per entrambi i modelli adattati
- d) Si calcola la statistica complessiva $\chi^2_1 = \chi^2_{\mathcal{M}'_1} + \chi^2_{\mathcal{M}'_2}$ e i gradi di libertà df_1
- e) Si valuta la statistica $(\chi^2_1 - \chi^2_0)$ mediante il test del χ^2 per modelli annidati (*nested chi-square test*)

La statistica utilizzata per il confronto è distribuita sotto l'ipotesi nulla $H_0 : \chi^2_1 - \chi^2_0 = 0$ come segue:

$$(\chi^2_1 - \chi^2_0) \stackrel{H_0}{\sim} \chi^2_{(df_1 - df_0)}$$

Si noti come H_0 formalizzi l'ipotesi che i due modelli siano invarianti (equivalenti). Per procedere nella valutazione dell'invarianza occorre dunque che H_0 non venga rigettata. Il rigetto dell'ipotesi nulla indica che i modelli sono *non-invarianti* (non equivalenti): l'obiettivo finale della valutazione dell'invarianza è quello di non rigettare H_0 . La presa di decisione dopo il test è effettuata come al solito: se il valore della statistica test $(\chi^2_1 - \chi^2_0)$ è maggiore del valore critico χ^2_{α} calcolato rispetto all' α scelto, il test rigetta l'ipotesi di equivalenza tra i due modelli.³ In generale:

- Se H_0 è rigettata la procedura di valutazione dell'invarianza si arresta: questo avviene quando la statistica test avrà un valore che cade nelle code della distribuzione, cioè maggiore del valore critico. In questo caso i due modelli sono differenti.
- Se H_0 non è rigettata la procedura di valutazione dell'invarianza continua con i modelli successivi più ristretti, dal momento che essa è di tipo incrementale. Questo accade quando la statistica test assume un valore prossimo allo zero: i due modelli sono invarianti rispetto al criterio valutato.

Si tenga presente che il test basato sul χ^2 , sebbene comunemente utilizzato per valutare l'adattamento ai dati di un modello statistico, dipende dalle dimensioni del campione: si osserva che esso tende a rigettare modelli accettabili se il campione è troppo grande in quanto anche piccole differenze di magnitudine nel fit risulteranno significative, mentre fallisce nel rigettare modelli non accettabili se il campione è di contro troppo piccolo. Per ovviare al problema delle dimensioni campionarie esistono altri indici di fit che possono essere utilizzati all'occorrenza, tra i quali l'indice AIC (Akaike Information Criterion) secondo cui il modello che meglio si adatta ai dati è il quello che presenta AIC minore.⁴

6.1.1 Invarianza configurale

L'invarianza configurale denota una condizione di invarianza rispetto alla struttura della matrice Λ in un modo tale che ogni gruppo abbia lo stesso numero di fattori comuni q

³Si tenga presente che una valutazione complessiva dell'invarianza di misurazione, tuttavia, non può non considerare l'utilizzo altresì di indici complessivi quali RMSEA, AIC e CFI unitamente al test del chi-quadrato.

⁴Ci sono vari modi per quantificare la magnitudine di una differenza tra modelli. Uno tra questi, ad esempio, si basa sulla seguente misura di *effect size* $\omega = \sqrt{(\chi^2_1 - \chi^2_0) / (N(df_1 - df_2))}$ dove le differenze tra χ^2_1 e χ^2_0 e gradi di libertà si riferiscono ai due modelli scelti. Valori di ω che approssimano 0.1 si riferiscono ad un effetto piccolo, uguali a 0.3 ad un effetto medio, mentre valori uguali o superiori a 0.5 indicano un effetto grande.

e le medesime posizioni degli zeri nella matrice dei *factor loadings*. Gli elementi diversi da zero sono parametri liberi (non vincolati). Più in dettaglio, tale tipo di invarianza implica che gli stessi indicatori definiscano un dato fattore in tutte le popolazioni: non sono presenti vincoli di invarianza, tranne che per il numero di coefficienti stimati e per le posizioni degli zeri nella matrice $\mathbf{\Lambda}$. Rigettare, ad esempio, tale tipologia di invarianza potrebbe suggerire la necessità di un diverso numero di fattori q nei gruppi o, più in generale, di diverse strutture fattoriali. Il primo passo per la valutazione dell'invarianza è quello di specificare un modello CFA che rifletta come il costrutto è stato operazionalizzato a livello teorico. Ogni gruppo ha il proprio modello adattato ai dati e il confronto viene svolto inizialmente per la prima tipologia di invarianza (che perciò funge da *baseline* per i confronti successivi) con l'obiettivo di valutare se lo stessa struttura fattoriale è valida in entrambi i gruppi. Formalmente, il modello CFA di invarianza configurale è il seguente:

$$\begin{aligned}\mathbf{X}^{(1)} &= \boldsymbol{\tau}^{(1)} + \mathbf{\Lambda}^{(1)}\boldsymbol{\eta}^{(1)} + \boldsymbol{\delta}^{(1)} && \text{(gruppo 1)} \\ \mathbf{X}^{(2)} &= \boldsymbol{\tau}^{(2)} + \mathbf{\Lambda}^{(2)}\boldsymbol{\eta}^{(2)} + \boldsymbol{\delta}^{(2)} && \text{(gruppo 2)}\end{aligned}$$

dove, come anticipato, ad ognuno dei due gruppi equivale un modello CFA sotto il vincolo che i parametri fissi e quelli liberi della matrice $\mathbf{\Lambda}$ siano gli stessi tra i due gruppi. Rispetto all'equazione di base della CFA, l'unica differenza è riscontrabile nell'aggiunta del termine $\tau_{p \times 1}$ che rappresenta le intercette del modello per ciascuna variabile osservata (medie degli item quando $\mathbf{\Lambda} = \mathbf{0}_{p \times q}$). In precedenza, infatti, la media delle variabili osservate era stata fissata a 0, secondo quanto previsto dagli assunti della CFA. La Figura 6.1 rappresenta un caso di modelli CFA adattati su due differenti gruppi nel caso dell'invarianza configurale mentre le matrici dei parametri adattati sono riportate di seguito. Il test del chi-quadrato è riportato in Tabella 6.1.

χ^2	df	m	AIC
15.46	16	38	1554.74

TABELLA 6.1: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza configurale.*

$$\hat{\mathbf{\Lambda}}^{(1)} = \begin{bmatrix} 1.00 & 0.00 \\ 0.99 & 0.00 \\ 0.84 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 0.97 \\ 0.00 & 0.73 \end{bmatrix} \quad \hat{\mathbf{\Lambda}}^{(2)} = \begin{bmatrix} 1.00 & 0.00 \\ 1.63 & 0.00 \\ 1.17 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.40 \\ 0.00 & 1.11 \end{bmatrix}$$

$$\hat{\boldsymbol{\Phi}}^{(1)} = \begin{bmatrix} 0.49 & \\ 0.30 & 0.43 \end{bmatrix} \quad \hat{\boldsymbol{\Phi}}^{(2)} = \begin{bmatrix} 0.38 & \\ 0.18 & 0.22 \end{bmatrix}$$

$$\hat{\boldsymbol{\tau}}^{(1)} = [0.26 \ 0.17 \ 0.08 \ -0.05 \ -0.12 \ 0.01]$$

$$\hat{\boldsymbol{\tau}}^{(2)} = [-0.32 \ -0.24 \ -0.13 \ 0.05 \ -0.18 \ -0.25]$$

$$\text{diag}(\hat{\boldsymbol{\Theta}}_{\boldsymbol{\delta}}^{(1)}) = [0.36 \ 0.30 \ 0.29 \ 0.64 \ 0.61 \ 0.63]$$

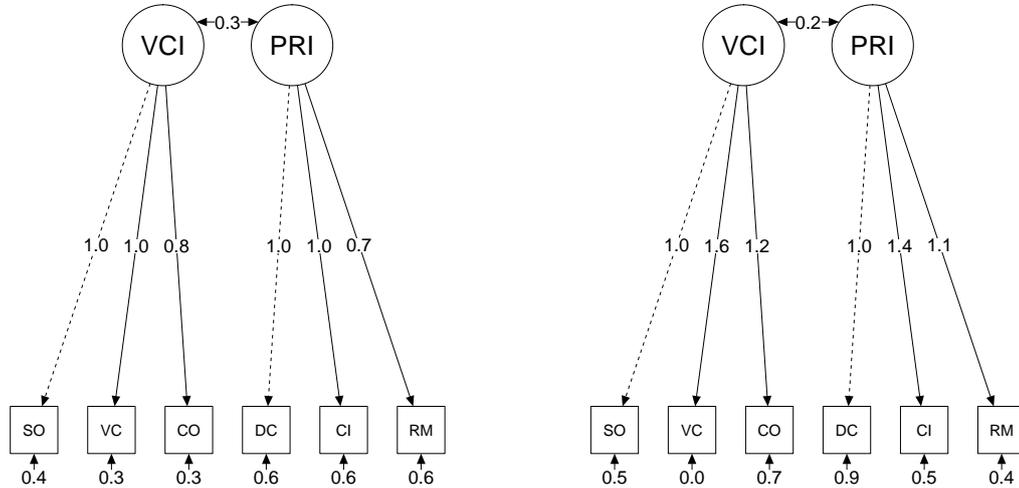


FIGURA 6.1: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza configurale.*

$$\text{diag}(\widehat{\Theta}_\delta^{(2)}) = [0.46 \ 0.03 \ 0.66 \ 0.94 \ 0.52 \ 0.41]$$

Si può notare come ogni gruppo abbia un modello CFA proprio ma strutturalmente identico con due fattori latenti legati ognuno alle stesse variabili osservate. Ciò significa che in entrambi i modelli: SO, VC, CO riflettono la variabile latente VCI mentre DC, CI, RM riflettono la variabile latente PRI (anche se con diversi *factor loadings*). Come per le altre tipologie di invarianza, il primo passaggio consiste qui nel calcolare i modelli CFA per i due gruppi \mathcal{M}_1 e \mathcal{M}_2 e nell'adattarli ai dati ottenendo le matrici di covarianza $\mathbf{S}_X^{(1)}$ e $\mathbf{S}_X^{(2)}$. Successivamente viene calcolata la statistica complessiva $\chi_{\mathcal{M}_1}^2 + \chi_{\mathcal{M}_2}^2$ con i relativi gradi di libertà, rispettivamente 15.46 e 16. Dal momento che per operare il confronto si usa la statistica differenza dei due χ^2 , si necessita di un secondo modello che corrisponde ad un livello di invarianza più ristretto, ossia quella debole.

6.1.2 Invarianza debole

Se l'invarianza configurale non viene rigettata, il passaggio successivo consiste nel vincolare come invarianti tutti i coefficienti della matrice $\mathbf{\Lambda}$: tale condizione è nota come invarianza debole o invarianza metrica. Essa implica che, tra i gruppi, le differenze nelle covarianze delle variabili misurate siano dovute ai fattori comuni. Se l'invarianza debole viene rigettata, significa che almeno alcuni coefficienti sono non invarianti tra i gruppi. Per quanto riguarda il secondo livello di invarianza, quella debole si ripete la stessa operazione della configurale ma in questo caso i coefficienti fattoriali sono uguali nei due

gruppi (stessi *factor loadings*) mentre le intercette si lasciano differire tra loro. In tal modo si cerca di capire se i soggetti dei due gruppi attribuiscono lo stesso significato al costrutto latente oggetto di studio. Il modello CFA di invarianza debole è il seguente:

$$\mathbf{X}^{(1)} = \boldsymbol{\tau}^{(1)} + \mathbf{\Lambda}\boldsymbol{\eta}^{(1)} + \boldsymbol{\delta}^{(1)} \quad (\text{gruppo 1})$$

$$\mathbf{X}^{(2)} = \boldsymbol{\tau}^{(2)} + \mathbf{\Lambda}\boldsymbol{\eta}^{(2)} + \boldsymbol{\delta}^{(2)} \quad (\text{gruppo 2})$$

dove, a differenza dell'invarianza configurale, nelle equazioni si nota l'assenza dell'apice su $\mathbf{\Lambda}$, ad indicare il fatto che la matrice è la stessa nei due gruppi. I due modelli non sono solo strutturalmente uguali ma hanno anche i medesimi valori stimati di $\mathbf{\Lambda}$. Si noti che ciascun gruppo ha un proprio modello fattoriale sotto il vincolo:

$$\mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)}$$

che traduce l'ipotesi che i due gruppi abbiano la stessa struttura fattoriale. Dal momento che ogni livello di invarianza include quello precedente, nel modello debole i due gruppi hanno la medesima struttura della matrice $\mathbf{\Lambda}$ come in quello configurale, sebbene il primo sia più vincolato rispetto a quest'ultimo (si ha difatti un'unica matrice $\mathbf{\Lambda}$). La Figura 6.2 rappresenta un caso di modelli CFA adattati su due differenti gruppi nel caso dell'invarianza debole mentre le matrici dei parametri adattati sono riportate di seguito. Il test del chi-quadrato è riportato in Tabella 6.2.

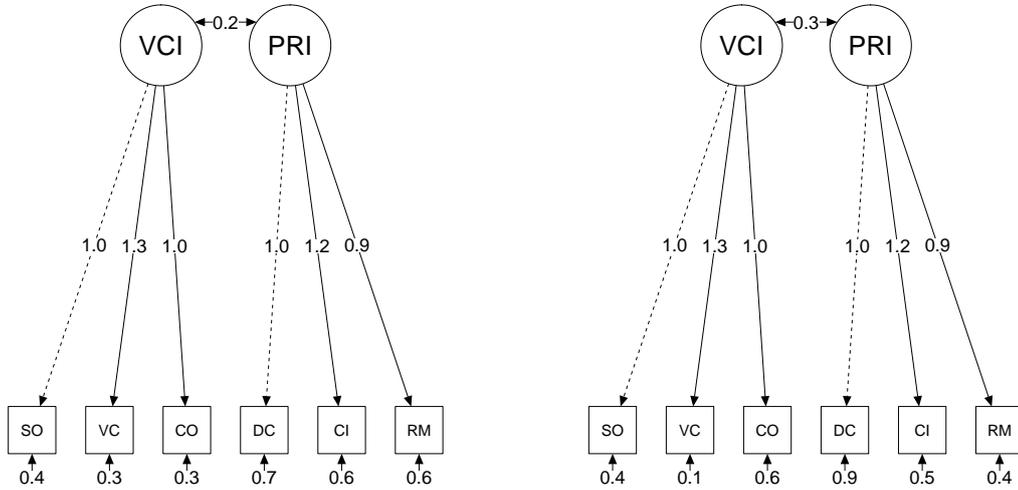


FIGURA 6.2: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza configurale.*

$$\hat{\mathbf{\Lambda}}^{(1)} = \hat{\mathbf{\Lambda}}^{(2)} \begin{bmatrix} 1.00 & 0.00 \\ 1.31 & 0.00 \\ 1.02 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.15 \\ 0.00 & 0.88 \end{bmatrix}$$

	Df	AIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
configurale	16	1554.74	15.46			
debole	20	1550.20	18.91	3.46	4	0.48

TABELLA 6.2: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza debole.*

$$\hat{\Phi}^{(1)} = \begin{bmatrix} 0.34 & \\ 0.21 & 0.35 \end{bmatrix} \quad \hat{\Phi}^{(2)} = \begin{bmatrix} 0.52 & \\ 0.26 & 0.31 \end{bmatrix}$$

$$\hat{\tau}^{(1)} = [0.26 \ 0.17 \ 0.08 \ -0.05 \ -0.12 \ 0.01]$$

$$\hat{\tau}^{(2)} = [-0.32 \ -0.24 \ -0.13 \ 0.05 \ -0.18 \ -0.25]$$

$$\text{diag}(\hat{\Theta}_\delta^{(1)}) = [0.42 \ 0.25 \ 0.29 \ 0.69 \ 0.57 \ 0.61]$$

$$\text{diag}(\hat{\Theta}_\delta^{(2)}) = [0.42 \ 0.13 \ 0.63 \ 0.91 \ 0.54 \ 0.42]$$

Anche in questo caso altri due modelli CFA vengono adattati ai dati, \mathcal{M}'_1 su $\mathbf{S}_X^{(1)}$ e \mathcal{M}'_2 su $\mathbf{S}_X^{(2)}$ attraverso i quali è successivamente calcolata la statistica complessiva data dalla somma dei chi-quadrato dei due modelli (che in questo caso vale $\chi_0^2 = 18.91$) con i rispettivi gradi di libertà. Per valutare se proseguire nel processo di valutazione dell'invarianza, si utilizza la statistica differenza tra i chi-quadrato riferiti ai due livelli. In tal modo avviene il confronto tra il modello CFA configurale e quello debole, mediante il quale si può valutare se la differenza tra i due è statisticamente significativa. In questo caso, il valore della statistica test d'interesse è pari a $18.91 - 15.46 = 3.46$, la differenza tra i gradi di libertà dei due modelli è $20 - 16 = 4$ mentre la probabilità associata alla statistica test è 0.48. Si può osservare come il modello debole ha più gradi di libertà di quello configurale: il numero di parametri che il modello può al più esprimere non cambia (essendo pari a $\frac{p \times (p+1)}{2}$) mentre m , ossia il numero di parametri da stimare, cambia. La conclusione, in questo esempio, è che il test del χ^2 non consente di rigettare l'ipotesi che i due modelli siano equivalenti: i due gruppi sono invarianti rispetto al criterio debole e condividono la medesima struttura fattoriale Λ . Se H_0 fosse invece rigettata a livello debole, avremmo potuto affermare che la struttura fattoriale di VCI e PRI è diversa nelle due popolazioni. Se questa tipologia di invarianza non viene rigettata, come in questo caso, le intercette possono essere successivamente vincolate: è questo il caso dell'invarianza forte (o scalare).

6.1.3 Invarianza forte

Nell'invarianza forte i due gruppi condividono non solo gli stessi coefficienti fattoriali (*factor loadings*) ma anche le stesse intercette. Il vincolo traduce l'ipotesi che i due gruppi condividano la stessa struttura fattoriale e allo stesso tempo interpretino gli item nello stesso modo. In un modello con invarianza forte non c'è presenza di eventuale *bias*

dovuto alle differenze tra gli item usati nei due gruppi. Il modello CFA di invarianza forte è rappresentato formalmente come segue:

$$\mathbf{X}^{(1)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(1)} + \boldsymbol{\delta}^{(1)} \quad (\text{gruppo 1})$$

$$\mathbf{X}^{(2)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(2)} + \boldsymbol{\delta}^{(2)} \quad (\text{gruppo 2})$$

dove ciascun gruppo ha un proprio modello fattoriale sotto i vincoli:

$$\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$$

$$\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$$

La Figura 6.3 rappresenta un caso di modelli CFA adattati su due differenti gruppi nel caso dell'invarianza forte mentre le matrici dei parametri adattati sono riportate di seguito. Il test del chi-quadrato è riportato in Tabella 6.3.

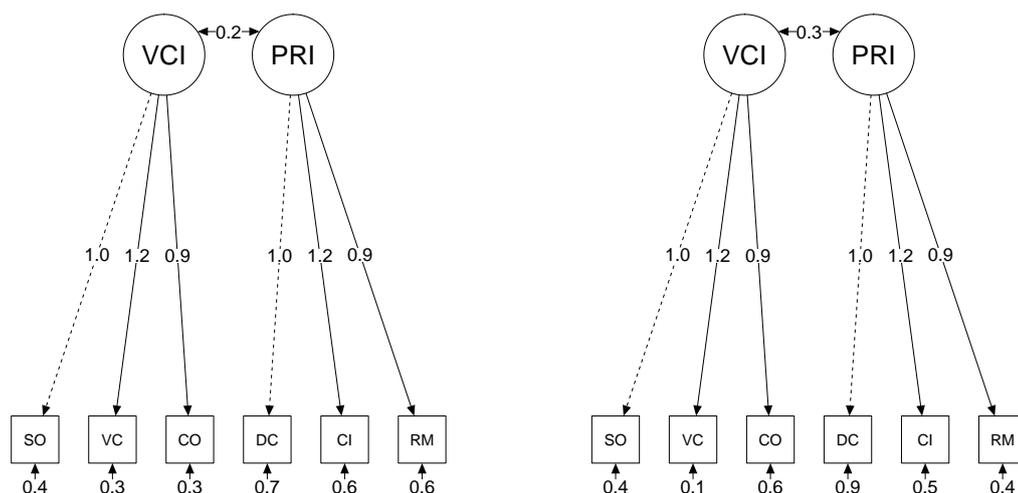


FIGURA 6.3: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza forte.*

	Df	AIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
debole	20	1550.20	18.91			
forte	24	1549.00	25.72	6.80	4	0.15

TABELLA 6.3: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza forte.*

Dai risultati riportati in Tabella 6.3 notiamo come i due modelli sono invarianti rispetto al criterio forte, hanno cioè la stessa struttura fattoriale mentre i soggetti di entrambi i

gruppi tendono a rispondere allo stesso modo agli item. Una volta accertata la presenza di invarianza forte, si può procedere con lo studio dell'ultima tipologia di invarianza, quella esatta.

6.1.4 Invarianza esatta

Quest'ultimo stadio rappresenta una situazione ideale, molto spesso difficile da realizzare nelle applicazioni reali. In generale, una struttura fattoriale è invariante in senso esatto quando i due gruppi condividono non solo gli stessi coefficienti fattoriali e le stesse intercette ma anche gli stessi errori di misura. Formalmente, il modello è il seguente:

$$\mathbf{X}^{(1)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(1)} + \boldsymbol{\delta} \quad (\text{gruppo 1})$$

$$\mathbf{X}^{(2)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(2)} + \boldsymbol{\delta} \quad (\text{gruppo 2})$$

con i seguenti vincoli di struttura:

$$\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$$

$$\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$$

$$\boldsymbol{\delta}^{(1)} = \boldsymbol{\delta}^{(2)}$$

che traducono l'ipotesi che i due gruppi condividono la stessa struttura fattoriale, gli item non producono *bias* di risposta nei gruppi e questi ultimi possono essere confrontati in termini di differenze tra costrutti latenti. La Figura 6.4 rappresenta un caso di modelli CFA adattati su due differenti gruppi nel caso dell'invarianza forte mentre le matrici dei parametri adattati sono riportate di seguito. Il test del chi-quadrato è riportato in Tabella 6.4.

	Df	AIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
forte	24	1549.00	25.72			
esatta	30	1544.60	33.32	7.60	6	0.27

TABELLA 6.4: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza esatta.*

Dai risultati riportati in Tabella 6.4 notiamo come i due modelli siano invarianti in senso esatto: i due gruppi condividono la medesima struttura fattoriale $\boldsymbol{\Lambda}$, non vi è *bias* indotto dalle differenze indotte dagli item $\boldsymbol{\tau}$ ed inoltre i residui $\boldsymbol{\delta}$ sono gli stessi in entrambi i gruppi. Conseguentemente, i due gruppi possono essere confrontati tra loro in termini di punteggi fattoriali $\hat{\boldsymbol{\eta}}^{(1)}$ e $\hat{\boldsymbol{\eta}}^{(2)}$.

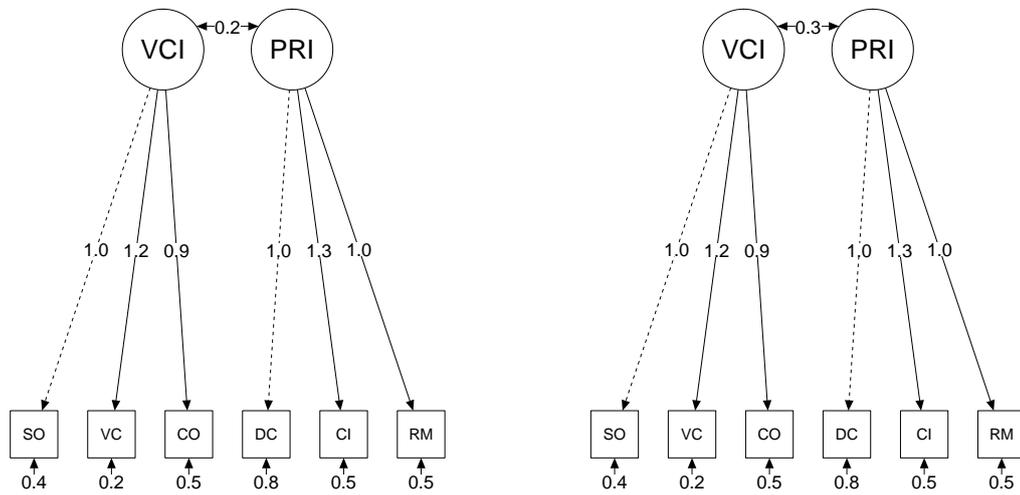


FIGURA 6.4: *Invarianza configurale: esempio di CFA multigruppo per la valutazione dell'invarianza esatta.*

6.2 Note di sintesi

Per concludere il capitolo sull'invarianza di misurazione, si riportano di seguito i quattro modelli CFA di invarianza (si noti che i vincoli sono rappresentati dalla presenza/assenza dell'indice dei gruppi posto in apice):

per $g = 1, \dots, G$:

1. $\mathbf{X}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ invarianza configurale
2. $\mathbf{X}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ invarianza debole
3. $\mathbf{X}^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)}$ invarianza forte
4. $\mathbf{X}^{(g)} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}$ invarianza esatta

6.3 Alcune Risorse utili

- o Tutorial sull'invarianza con lavaan:
<https://tinyurl.com/yy3rutpz>
- o Breve talk sull'invarianza con lavaan:
<https://users.ugent.be/~yrosseel/lavaan/multiplegroup6Dec2012.pdf>
- o Raccolta di articoli sull'invarianza pubblicato su *Frontiers in Psychology*:
<https://www.frontiersin.org/research-topics/1695/measurement-invariance>

6.4 Laboratorio in R

I laboratori associati a questo capitolo sono i seguenti:

- o lab12.R: (A) WISC-IV: dati, (B) WISC-IV: valutazione invarianza

Il loro contenuto può essere visualizzato e scaricato all'indirizzo <https://github.com/antcalcagni/psychological-testing/tree/master/labs>.

bozza non corretta

A

Elementi essenziali di Algebra Lineare

In questa appendice saranno descritti alcuni degli elementi base dell'algebra lineare essenziali alla comprensione di alcune parti della presente dispensa. La trattazione sarà sistematica ma succinta: molti degli argomenti di base verranno perciò omessi mentre ad altri sarà dato maggiore risalto.

A.1 Matrici, vettori, scalari

Una *matrice* è una organizzazione di oggetti disposti secondo due riferimenti, attraverso le *righe* e le *colonne*:

$$\begin{pmatrix} x_{1,1} & \dots & x_{1,m} \\ x_{2,1} & \dots & x_{2,m} \\ \dots & x_{i,j} & \dots \\ x_{n,1} & \dots & x_{n,m} \end{pmatrix}$$

dove n è il numero di oggetti disposti per riga mentre m il numero di colonne. Gli indici i e j assolvono alla funzione di contatore e permettono di utilizzare dei valori numerici per rappresentare gli oggetti di una matrice: ad esempio ponendo $i = (1, 2, 3, \dots, n)$ e $j = (1, 2, 3, \dots, m)$ possiamo indicare con l'oggetto $x_{i,j}$ il generico elemento posto sulla i -esima riga e sulla j -colonna (nell'esempio appena fatto l'elemento centrale) mentre $x_{1,2}$ indica l'oggetto della matrice posto all'incrocio della prima riga e della prima colonna e così di seguito. L'utilizzo del linguaggio matriciale nella statistica permette la manipolazione di grandi quantità di dati e permette la risoluzione di taluni problemi in modo elegante attraverso gli strumenti tipici che l'algebra lineare fornisce. Sebbene quest'ultima abbia una immediata interpretazione nel campo della fisica ciò non è lo stesso in statistica dove l'algebra assolve invece ad un compito di tipo meramente computazionale.

Un secondo tipo di organizzazione è il *vettore*, una matrice avente solo una colonna ($m = 1$):

$$\begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \dots \\ x_{i,1} \\ \dots \\ x_{n,1} \end{pmatrix}$$

o una sola riga ($n = 1$):

$$(x_{1,1} \quad x_{2,1} \quad \dots \quad x_{i,1} \quad \dots \quad x_{n,1})$$

Nel primo caso il vettore sarà denominato *vettore colonna* mentre in quest'ultimo caso *vettore riga*.

Un terzo tipo di organizzazione è lo *scalare*, ossia un singolo elemento della matrice (o del vettore). Matrici, vettori e scalari possono essere utilizzati assieme per manipolare differenti oggetti in diversi contesti. Nel seguito dettaglieremo in modo più approfondito alcuni di questi.¹ Rispetto alla notazione,

¹In questa breve sezione abbiamo preferito parlare di *oggetti* piuttosto che *numeri* per evidenziare come questi ultimi siano solo alcuni dei possibili oggetti manipolabili. In algebra lineare non è scontato che una matrice sia sempre composta da oggetti numerici e l'uso di essi (ad esempio i naturali, i reali)

è utile ricordare come una matrice è sempre indicata con una lettera maiuscola grassetto, un vettore con una lettera minuscola grassetto mentre uno scalare con una lettera minuscola non posta in grassetto; la dimensionalità è posta al pedice della lettera che la indica (ad esempio, $\mathbf{X}_{n \times m}$ indica la matrice \mathbf{X} avente n righe ed m colonne). Una matrice avente righe e colonne uguali ($n = m$) si dice *quadrata* (ad esempio, $\mathbf{X}_{2 \times 2}$ è una matrice quadrata) diversamente si dice *rettangolare* (ad esempio, $\mathbf{X}_{2 \times 3}$ è una matrice rettangolare). Una matrice che ha componenti sulla diagonale principale diversi da zero e tutti zero nelle altre posizioni si dice *diagonale*. Una matrice che ha componenti diverse da zero solo al di sotto della diagonale principale si dice *triangolare inferiore* mentre il caso speculare (matrice con elementi non nulli solo al di sopra della diagonale principale) si dice *triangolare superiore*. La matrice che ha il triangolo superiore uguale a quello inferiore è detta *simmetrica* (sempre quadrata). Un altro particolare caso di matrice è la matrice *identità* (denotata genericamente con $\mathbf{I}_{n \times m}$) una matrice diagonale quadrata i cui elementi sono tutti uno e la matrice *nulla* composta da tutti zero in ogni sua posizione $\mathbf{0}_{n \times m}$.

A.2 Alcune operazioni di algebra lineare

Il linguaggio dell'algebra lineare permette di manipolare matrici, vettori e scalari mediante peculiari operazioni a cui in questa sezione sarà dato breve ed essenziale spazio. Evidenzieremo solo quanto sarà necessario per la comprensione dei modelli e metodi esposti in questa dispensa.

La prima operazione da considerare è la *somma* (o *differenza*) che si indica con "+" (o "-"):

$$\begin{pmatrix} x_{1,1} & \dots & x_{1,m} \\ x_{2,1} & \dots & x_{2,m} \\ \dots & x_{i,j} & \dots \\ x_{n,1} & \dots & x_{n,m} \end{pmatrix} + \begin{pmatrix} y_{1,1} & \dots & y_{1,m} \\ y_{2,1} & \dots & y_{2,m} \\ \dots & y_{i,j} & \dots \\ y_{n,1} & \dots & y_{n,m} \end{pmatrix} = \begin{pmatrix} x_{1,1} + y_{1,1} & \dots & x_{1,m} + y_{1,m} \\ x_{2,1} + y_{2,1} & \dots & x_{2,m} + y_{2,m} \\ \dots & x_{i,j} + y_{i,j} & \dots \\ x_{n,1} + y_{n,1} & \dots & x_{n,m} + y_{n,m} \end{pmatrix}$$

L'operazione è eseguita elemento per elemento e le matrici devono avere la stessa dimensione. Per dare un esempio, consideriamo il seguente:

$$\mathbf{A}_{4 \times 3} = \begin{pmatrix} 8 & 2 & 4 \\ 1 & 2 & 1 \\ 2 & 3 & 5 \\ 0 & 1 & 5 \end{pmatrix} \quad \mathbf{B}_{4 \times 3} = \begin{pmatrix} 9 & 8 & 1 \\ 12 & 8 & 0 \\ 1 & 4 & 7 \\ 1 & 2 & 1 \end{pmatrix}$$

poiché le due matrici hanno la stessa dimensionalità (entrambe hanno dimensione 4×3 ossia tre quattro e tre colonne) ha senso scrivere $\mathbf{A}_{4 \times 3} + \mathbf{B}_{4 \times 3} = \mathbf{C}_{4 \times 3}$ e possono dunque essere sommate dando luogo ad una terza matrice avente ancora lo stesso numero di righe e di colonne:

$$\begin{pmatrix} 8 & 2 & 4 \\ 1 & 2 & 1 \\ 2 & 3 & 5 \\ 0 & 1 & 5 \end{pmatrix} + \begin{pmatrix} 9 & 8 & 1 \\ 12 & 8 & 0 \\ 1 & 4 & 7 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 17 & 10 & 5 \\ 13 & 10 & 1 \\ 3 & 7 & 12 \\ 1 & 3 & 6 \end{pmatrix}$$

la matrice \mathbf{C} è la matrice di destinazione dell'operazione somma e contiene elemento per elemento la somma degli elementi delle matrici \mathbf{A} e \mathbf{B} . Le seguenti valgono per la somma:

1. $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$
4. $\omega \cdot (\mathbf{A} + \mathbf{B}) = (\mathbf{A}\omega + \mathbf{B}\omega)$

Allo stesso modo è possibile sommare due vettori:

$$\mathbf{a}_{3 \times 1} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \quad \mathbf{b}_{3 \times 1} = \begin{pmatrix} 1 \\ 8 \\ 2 \end{pmatrix}$$

è sempre tale all'interno di un preciso contesto o ambiente di riferimento (ad esempio l'ambiente \mathbb{N} , \mathbb{R} senza la dichiarazione del quale non avrebbe senso scrivere una matrice con oggetti numerici. Tuttavia, per la trattazione di questa dispensa e rispetto a coloro i quali si rivolge, utilizzeremo direttamente i numeri per indicare gli oggetti, non ponendo il problema appena delineato.

e poiché \mathbf{a} e \mathbf{b} hanno la stessa dimensione ha senso scrivere che $\mathbf{a}_{3 \times 1} + \mathbf{b}_{3 \times 1} = \mathbf{c}_{3 \times 1}$ dove il vettore risultante è ottenuto sommando elemento per elemento come segue

$$\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 8 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 11 \\ 3 \end{pmatrix}$$

Questo vale anche nel caso in cui \mathbf{a} e \mathbf{b} siano vettori riga (e dunque $\mathbf{a}_{1 \times 3}$ e $\mathbf{b}_{1 \times 3}$).

Possiamo ora considerare l'operazione *prodotto*. Vi sono diversi tipi di operazione prodotto (prodotto tra scalari, tra uno scalare ed una matrice o vettore, tra matrici, prodotto elemento per elemento, prodotto di Kronecker) ma in questa breve trattazione saranno considerate solo le prime tre di queste e le denoteremo genericamente col simbolo di " \cdot ". L'operazione di prodotto può avvenire dunque tra scalari, tra vettori e tra matrici come tra uno scalare ed un vettore, uno scalare ed una matrice, un vettore ed una matrice. Vediamoli in ordine.

Il primo caso (prodotto tra scalari) si riduce al noto prodotto aritmetico e restituisce ancora uno scalare (ad esempio: $2 \cdot 3 = 6$).

Il secondo caso, il prodotto tra uno scalare ed un vettore o matrice, restituisce ancora un vettore o una matrice, come nell'esempio che segue:

$$3 \cdot \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 3 \cdot 1 & 3 \cdot 2 \\ 3 \cdot 2 & 3 \cdot 0 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 0 \end{pmatrix}$$

La stessa operazione può essere fatta tra uno scalare ed un vettore. Questa operazione gode delle seguenti proprietà:

1. $a \cdot (b \cdot \mathbf{A}) = (a \cdot b) \cdot \mathbf{A}$
2. $(a + b) \cdot \mathbf{A} = \mathbf{A} \cdot a + \mathbf{B} \cdot b$
3. $a \cdot (\mathbf{A} + \mathbf{B}) = \mathbf{A} \cdot a + \mathbf{B} \cdot b$

Il terzo caso, prodotto tra vettori o matrici, può restituire una matrice, un vettore o uno scalare e l'operazione può essere effettuata solo nel caso in cui la seconda dimensione (il numero di colonne) del vettore o matrice che pre-moltiplica è uguale alla prima dimensione (il numero delle righe) del vettore o matrice che post-moltiplica (prodotto *righe \times colonne*); consideriamo ad esempio:

$$\mathbf{a}_{1 \times 3} = (2 \quad 4 \quad 9) \quad \mathbf{b}_{3 \times 1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

poiché il numero delle colonne del vettore che pre-moltiplica $\mathbf{a}_{1 \times 3}$ è uguale al numero di righe del vettore che post-moltiplica $\mathbf{b}_{3 \times 1}$ ha senso scrivere che $\mathbf{a}_{1 \times 3} \cdot \mathbf{b}_{3 \times 1} = \mathbf{c}_{3 \times 1}$ dove la matrice risultante (in particolare uno scalare) ha quale numero di righe quelle del vettore che pre-moltiplica e per numero di colonne quello del vettore che post-moltiplica ed è ottenuta sommando il prodotto elemento per elemento, come segue:

$$(2 \quad 4 \quad 9) \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = (2 \cdot 1 + 4 \cdot 0 + 9 \cdot 1) = (11)$$

Considerando ancora ad esempio quanto segue:

$$\mathbf{A}_{4 \times 3} = \begin{pmatrix} 9 & 8 & 1 \\ 12 & 8 & 0 \\ 1 & 4 & 7 \\ 1 & 2 & 1 \end{pmatrix} \quad \mathbf{b}_{3 \times 1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

poiché la condizione sulla dimensionalità prima esposta sussiste anche in questo caso, ha senso scrivere che $\mathbf{A}_{4 \times 3} \cdot \mathbf{b}_{3 \times 1} = \mathbf{c}_{4 \times 1}$ dove il vettore risultante è ottenuto moltiplicando una riga alla volta della matrice che pre-moltiplica con il singolo elemento del vettore che post-moltiplica, come segue:

$$\begin{pmatrix} 9 & 8 & 1 \\ 12 & 8 & 0 \\ 1 & 4 & 7 \\ 1 & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} (9 \cdot 1 + 8 \cdot 0 + 1 \cdot 1) \\ (12 \cdot 1 + 8 \cdot 0 + 0 \cdot 1) \\ (1 \cdot 1 + 4 \cdot 0 + 7 \cdot 1) \\ (1 \cdot 1 + 2 \cdot 0 + 1 \cdot 1) \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \\ 8 \\ 2 \end{pmatrix}$$

lo stesso risultato può essere ottenuto considerando l'operazione appena vista in un altro modo, ossia rendendo il prodotto una combinazione lineare del vettore con le colonne della matrice, come segue:

$$1 \cdot \begin{pmatrix} 9 \\ 12 \\ 1 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 8 \\ 8 \\ 4 \\ 2 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 7 \\ 1 \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \\ 8 \\ 2 \end{pmatrix}$$

Il prodotto tra matrici è una estensione del caso tra vettori appena descritto. Poniamo le seguenti matrici:

$$\mathbf{A}_{3 \times 3} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 7 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad \mathbf{B}_{3 \times 4} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 8 & 8 & 0 & 1 \\ 7 & 9 & 1 & 1 \end{pmatrix}$$

poiché è rispettata la regola precedente che rende consistente l'operazione prodotto applicata su questi dati (ricordiamo, il numero delle colonne della matrice che pre-moltiplica deve essere uguale al numero delle righe della matrice che post-moltiplica) ha senso considerare la seguente applicazione:

$$\begin{aligned} & \begin{pmatrix} 1 & 2 & 1 \\ 2 & 7 & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 & 0 \\ 8 & 8 & 0 & 1 \\ 7 & 9 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} (1 \cdot 1 + 2 \cdot 8 + 1 \cdot 7) & (1 \cdot 1 + 2 \cdot 8 + 1 \cdot 9) & (1 \cdot 0 + 2 \cdot 0 + 1 \cdot 1) & (1 \cdot 0 + 2 \cdot 1 + 1 \cdot 1) \\ (2 \cdot 1 + 7 \cdot 8 + 1 \cdot 7) & (2 \cdot 1 + 7 \cdot 8 + 1 \cdot 9) & (2 \cdot 0 + 7 \cdot 0 + 1 \cdot 1) & (2 \cdot 0 + 7 \cdot 1 + 1 \cdot 1) \\ (1 \cdot 1 + 1 \cdot 8 + 0 \cdot 7) & (1 \cdot 1 + 1 \cdot 8 + 0 \cdot 9) & (1 \cdot 0 + 1 \cdot 0 + 0 \cdot 1) & (1 \cdot 0 + 1 \cdot 1 + 0 \cdot 1) \end{pmatrix} \\ &= \begin{pmatrix} 24 & 26 & 1 & 3 \\ 65 & 67 & 1 & 8 \\ 9 & 9 & 0 & 1 \end{pmatrix} \end{aligned}$$

L'operazione prodotto gode delle seguenti proprietà:

1. $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B}) \cdot \mathbf{C} = (\mathbf{A} \cdot \mathbf{C}) + (\mathbf{B} \cdot \mathbf{C})$
3. $(\mathbf{A} \cdot \mathbf{B}) \cdot \omega = (\omega \cdot \mathbf{A}) \cdot \mathbf{B}$
4. $\mathbf{A} \cdot \mathbf{I} = \mathbf{A}$
5. $\mathbf{A} \cdot \mathbf{0} = \mathbf{0}$
6. $\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A}$ (non sempre)

Un'altra operazione nota è quella della *trasposizione* che data una generica matrice \mathbf{A} consiste nel trasporre le righe in colonne e le colonne in righe e la denotiamo col simbolo \mathbf{A}^T . Ad esempio:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 0 \end{pmatrix}$$

Le proprietà dell'operazione trasposizione sono le seguenti:

1. $(\mathbf{A} \cdot \mathbf{B})^T = \mathbf{B}^T \cdot \mathbf{A}^T$
2. $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
3. $((\mathbf{A})^T)^T = \mathbf{A}$
4. $(\mathbf{I})^T = \mathbf{I}$

L'ultima operazione a cui faremo accenno è quella dell'*inversione*. In generale, una matrice \mathbf{A} quadrata può ammettere al massimo una ed una sola matrice \mathbf{A}^{-1} che è chiamata inversa della matrice di partenza. La matrice inversa è tale che se moltiplicata per la matrice originaria, tale prodotto è la matrice identità, ossia $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. In generale, dato $\mathbf{A}\mathbf{x} = \mathbf{b}$ è possibile affermare che una matrice \mathbf{A} è non invertibile se il prodotto $\mathbf{A}\mathbf{x} = \mathbf{0}$ (con \mathbf{x} non nullo) poiché una matrice moltiplicata per un vettore il cui risultato è nullo non può restituire un vettore non nullo (in questo caso \mathbf{b}). Il calcolo dell'inversione di una matrice non verrà qui affrontato ma basti sapere che il significato dell'operazione inversione è connesso al concetto di *determinante* di una matrice e di *autovalori* di una matrice (aspetti questi che verranno

ripresi successivamente). A titolo di definizione, una matrice è invertibile se è *non singolare*, ossia se il suo determinante è diverso da zero o se i suoi autovalori sono diversi da zero.² Alcune proprietà dell'operazione inversione:

1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
2. $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$

A.3 Note finali

Per terminare questa succinta esposizione dell'algebra delle matrici è opportuno fare riferimento ad altre operazioni che caratterizzano l'algebra delle matrici, come ad esempio il determinante, il rango, la norma, gli autovalori e gli autovettori. Rispetto allo scopo di questa dispensa, queste non verranno trattate. Inoltre, non verranno qui presi in dettaglio ulteriori argomenti base come quello di spazio vettoriale e sottospazio, di indipendenza lineare, di trasformazione lineare, di ortogonalità e proiezioni ortogonali, decomposizioni di matrici, fattorizzazioni di matrici. Per tutti questi aspetti si rimanda il lettore ad un testo specializzato di Algebra 1 (ad esempio: <http://joshua.smcvt.edu/linearalgebra/>).

²A titolo introduttivo, le operazioni a cui abbiamo appena accennato (autovalori, determinanti, inversione) sono connesse al più generale problema della risoluzione di sistemi lineari mediante l'algebra lineare. A tal fine sono stati sviluppati diversi metodi sebbene il più noto sia quello di Gauss-Jordan, sotto il nome di *algoritmo di Gauss-Jordan*. In breve, esso permette di scrivere un sistema lineare $\mathbf{Ax} = \mathbf{b}$ mediante procedimenti di fattorizzazione della matrice dei coefficienti \mathbf{A} che permettono di semplificare il problema mediante la trasformazione della matrice dei coefficienti in una matrice triangolare alta i cui elementi sulla diagonale prendono il nome di *pivot*. Tale processo di fattorizzazione della matrice \mathbf{A} si arresta se non è possibile garantire la non negatività dei pivot e poiché il determinante di tale matrice è pari alla somma dei pivot, tale matrice ha determinante nullo. Una matrice, infine, con pivot nulli è una matrice non invertibile (il processo di inversione è altresì basato sull'algoritmo di Gauss-Jordan).

bozza non corretta

B

Elementi essenziali del linguaggio R

Questa breve appendice riporta alcuni elementi introduttivi all'ambiente di calcolo R e al suo linguaggio. L'organizzazione dei contenuti è di livello base e procede mostrando alcune semplici applicazioni per la risoluzione di problemi di calcolo e rappresentazione dell'informazione. Per facilitare l'apprendimento di taluni contenuti sono forniti anche una serie di esercizi di crescente difficoltà le cui soluzioni sono disponibili all'indirizzo https://github.com/antcalcagni/psychological-testing/blob/master/appendice_C_soluzioni.R.

B.1 Scrittura di espressioni

R è un'ottima calcolatrice.

Esercizi

Calcola il risultato delle seguenti espressioni utilizzando R.

$$1) \frac{(45 + 21)^3 + \frac{3}{4}}{\sqrt{32 - \frac{12}{17}}}$$

$$2) \frac{\sqrt{7 - \pi}}{3(45 - 34)}$$

$$3) \sqrt[3]{12 - e^2} + \ln(10\pi)$$

$$4) \frac{\sin(\frac{3}{4}\pi)^2 + \cos(\frac{3}{2}\pi)}{\log_7 e^{\frac{3}{2}}}$$

$$5) \frac{\sum_{n=1}^{10} n}{10}$$

Note:

- In R la radice quadrata si ottiene con la funzione `sqrt()` mentre per radici di indici diversi si utilizza la notazione esponenziale ($\sqrt[3]{x}$ è dato da $x^{(1/3)}$).

A cura di: *Claudio Zandonella Callegher*

<code>x + y</code>	Addizione	<code>> 5 + 3</code> <code>[1] 8</code>
<code>x - y</code>	Sottrazione	<code>> 7 - 2</code> <code>[1] 5</code>
<code>x * y</code>	Moltiplicazione	<code>> 4 * 3</code> <code>[1] 12</code>
<code>x / y</code>	Divisione	<code>> 8 / 3</code> <code>[1] 2.666667</code>
<code>x %% y</code>	Resto della divisione	<code>> 7 %% 5</code> <code>[1] 2</code>
<code>x %/% y</code>	Divisione intera	<code>> 7 %/% 5</code> <code>[1] 1</code>
<code>abs (x)</code>	Valore assoluto	<code>> abs(3-5^2)</code> <code>[1] 22</code>
<code>sign (x)</code>	Segno di un'espressione	<code>> sign(-8)</code> <code>[1] -1</code>
<code>sqrt (x)</code>	Radice quadrata	<code>> sqrt(25)</code> <code>[1] 5</code>
<code>log (x)</code>	Logaritmo naturale	<code>> log(10)</code> <code>[1] 2.302585</code>
<code>exp (x)</code>	Esponenziale	<code>> exp(1)</code> <code>[1] 2.718282</code>
<code>sin (x)</code> <code>cos (x)</code> <code>tan (x)</code> <code>asin (x)</code> <code>acos (x)</code> <code>atan (x)</code>	Funzioni trigonometriche	<code>> sin(pi/2)</code> <code>[1] 1</code> <code>> cos(pi/2)</code> <code>[1] 6.123234e-17</code>
<code>factorial (x)</code>	Fattoriale	<code>> factorial(6)</code> <code>[1] 720</code>
<code>choose (n, k)</code>	Coefficiente combinatorio	<code>> choose(5,3)</code> <code>[1] 10</code>

TABELLA B.1: *Principali funzioni matematiche in R*

- Il valore di π si ottiene con `pi`.
- Il valore di e si ottiene con `exp(1)`.
- In R per i logaritmi si usa la funzione `log(x, base=a)`, di base viene considerato il logaritmo naturale.

B.2 Vettori

B.2.1 Creazione di vettori

In R per definire un vettore si utilizza il comando `nome_vettore <- c(oggetti)`. Ricorda che gli elementi devono essere separati da una virgola.

Esercizi

1. Crea il vettore `x` contenente i numeri 4, 6, 12, 34, 8.
2. Crea il vettore `y` contenente tutti i numeri pari compresi tra 1 e 25 (`?seq()`).
3. Crea il vettore `z` contenente tutti i primi 10 multipli di 7 partendo da 13 (`?seq()`).
4. Crea il vettore `s` in cui le lettere "A", "B" e "C" vengono ripetute nel medesimo ordine 4 volte (`?rep()`).
5. Crea il vettore `t` in cui le lettere "A", "B" e "C" vengono ripetute ognuna 4 volte (`?rep()`).

B.2.2 Selezione elementi di un vettore

In R per selezionare gli elementi di un vettore si deve indicare all'interno delle parentesi quadre la **posizione degli elementi** da selezionare, non il valore dell'elemento stesso
`nome_vettore[indice_posizione]`

In alternativa si può definire la condizione logica che gli elementi che si vogliono selezionare devono rispettare.

<code>x == y</code>	Uguale	<code>> 5==3</code> [1] FALSE
<code>x != y</code>	Diverso	<code>> 5!=3</code> [1] TRUE
<code>x >= y</code>	Maggiore o uguale	<code>> 5>=3</code> [1] TRUE
<code>x > y</code>	Maggiore	<code>> 13>7</code> [1] TRUE
<code>x <= y</code>	Minore o uguale	<code>> 5<=3</code> [1] FALSE
<code>x < y</code>	Minore	<code>> 13<7</code> [1] FALSE

TABELLA B.2: Operatori logici in R

Per concatenare più operazioni logiche si possono usare la congiunzione logica "e" (`&`) o la disgiunzione inclusiva "o" (`|`). Per **eliminare degli elementi** da un vettore si utilizza all'interno delle parentesi quadre l'operatore "-" insieme agli indici di posizione degli elementi da eliminare (esempio: `x[c(-2,-4)]` oppure `x[-c(2,4)]`).

Esercizi

1. Del vettore `x` seleziona il 2°, 3° e 5° elemento.
2. Del vettore `y` seleziona tutti i valori minori di 13 o maggiori di 19.
3. Del vettore `z` seleziona tutti i valori compresi tra 24 e 50.
4. Elimina dal vettore `z` i valori 28 e 42.
5. Del vettore `s` seleziona tutti gli elementi uguali ad "A".
6. Del vettore `t` seleziona tutti gli elementi diversi da "B".

B.2.3 Funzioni ed operazioni tra vettori

Per compiere operazioni tra vettori è necessario che essi abbiano identica lunghezza.

<code>nuovo_vettore <- c(vettore1,vettore2)</code>	Per unire più vettori in un unico vettore
<code>length(nome_vettore)</code>	Per valutare il numero di elementi contenuti in un vettore
<code>vettore1 + vettore2</code> <code>vettore1 - vettore2</code>	Somma di due vettori Differenza tra due vettori
<code>vettore1 * vettore2</code> <code>vettore1 / vettore2</code>	Prodotto tra due vettori Rapporto tra due vettori

TABELLA B.3: Operazioni con vettori

Nota: In R il prodotto e rapporto tra vettori sono eseguiti elemento per elemento (al contrario di molti altri software).

Esercizi

1. Crea il vettore j unendo i vettori x ed z .
2. Elimina gli ultimi tre elementi del vettore j e controlla che i vettori j e y abbiano la stessa lunghezza.
3. Calcola la somma tra i vettori j e y .
4. Moltiplica il vettore z per una costante $k=3$.
5. Calcola il prodotto tra i primi 10 elementi del vettore y ed il vettore z .

B.3 Matrici

B.3.1 Creazione di matrici

In R per definire una matrice di n righe e s colonne si utilizza il comando
`nome_matrice <- matrix(data, nrow=n, ncol=s, byrow=FALSE)`

Nota: Di default R riempie la matrice per colonne, impostando `byrow = TRUE` si riempie per righe.

Esercizi

1. Crea la matrice A così definita:

2	34	12	7
46	93	27	99
23	38	7	04
2. Crea la matrice B contenente tutti i primi 12 numeri dispari disposti su 4 righe e 3 colonne.
3. Crea la matrice C contenente i primi 12 multipli di 9 disposti su 3 righe e 4 colonne.
4. Crea la matrice D formata da 3 colonne in cui le lettere "A", "B" e "C" vengano ripetute 4 volte ciascuna rispettivamente nella prima, seconda e terza colonna.
5. Crea la matrice E formata da 3 righe in cui le lettere "A", "B" e "C" vengano ripetute 4 volte ciascuna rispettivamente nella prima, seconda e terza riga.

B.3.2 Selezione di elementi di una matrice

In R per selezionare gli elementi di matrice si deve indicare all'interno delle parentesi quadre l'indice di riga e l'indice di colonna (**separati da virgola**) degli elementi da selezionare oppure la condizione logica che devono rispettare.

```
nome_matrice[indice_riga , indice_colonna]
```

Nota: per selezionare tutti gli elementi di una data riga o di una data colonna basta lasciare vuoto rispettivamente l'indice di riga o l'indice di colonna.

Esercizi

1. Utilizzando gli indici di riga e di colonna seleziona il numero 27 della matrice A
2. Seleziona gli elementi compresi tra la seconda e quarta riga, seconda e terza colonna della matrice B.
3. Seleziona solo gli elementi pari della matrice A (Nota: utilizza l'operazione resto %%).
4. Elimina dalla matrice C la terza riga e la terza colonna.
5. Seleziona tutti gli elementi della seconda e terza riga della matrice B.
6. Seleziona tutti gli elementi diversi da "B" appartenenti alla matrice D.

B.3.3 Funzioni ed operazioni tra matrici

Note:

- Per il significato di determinante di una matrice considera: <https://it.wikipedia.org/wiki/Determinante>.
- Per il significato di matrice inversa considera: https://it.wikipedia.org/wiki/Matrice_invertibile.
- Per compiere operazioni elemento per elemento tra due matrici, esse devono avere la stessa dimensione.
- Per compiere il prodotto matriciale il numero di colonne della prima matrice deve essere uguale al numero di righe della seconda matrice (vedi https://it.wikipedia.org/wiki/Moltiplicazione_di_matrici).
- É possibile assegnare nomi alle colonne e righe di una matrice rispettivamente attraverso i comandi:

```
colnames(nome_matrice)<-c("nome_1",..., "nome_s")
rownames(nome_matrice)<-c("nome_1",..., "nome_n")
```

Esercizi

1. Crea la matrice G unendo alla matrice A le prime due colonne della matrice C.
2. Crea la matrice H unendo alla matrice C le prime due righe della matrice trasposta di B.
3. Ridefinisci la matrice A eliminando la seconda colonna. Ridefinisci la matrice B eliminando la prima riga. Verifica che le matrici così ottenute abbiano la stessa dimensione.
4. Commenta i differenti risultati che otteniamo nelle operazioni $A*B$, $B*A$, $A\%*B$ e $B\%*A$.
5. Assegna i seguenti nomi alle colonne e alle righe della matrice C: "col_1", "col_2", "col_3", "col_4", "row_1", "row_2", "row_3".

<code>nuova_matrice <- cbind(matrice1,matrice2)</code>	Per unire due matrici creando nuove colonne (le matrici devono avere lo stesso numero di righe)
<code>nuova_matrice <- rbind(matrice1,matrice2)</code>	Per unire due matrici creando nuove righe (le matrici devono avere lo stesso numero di colonne)
<code>nrow(nome_matrice)</code>	Per valutare il numero di righe della matrice
<code>ncol(nome_matrice)</code>	Per valutare il numero di colonne della matrice
<code>dim(nome_matrice)</code>	Per valutare la dimensione della matrice (righe e colonne)
<code>t(nome_matrice)</code>	Per ottenere la trasposta della matrice
<code>diag(nome_matrice)</code>	Ottenere un vettore con gli elementi della diagonale della matrice
<code>det(nome_matrice)</code>	Ottenere il determinante della matrice (la matrice deve essere quadrata)
<code>solve(nome_matrice)</code>	Ottenere l'inversa della matrice (la matrice deve essere quadrata)
<code>colnames(nome_matrice)</code> <code>rownames(nome_matrice)</code>	Nomi delle colonne della matrice Nomi delle righe della matrice
<code>matrice1 + matrice2</code>	Somma elemento per elemento di due matrici
<code>matrice1 - matrice2</code>	Differenza elemento per elemento tra due matrici
<code>matrice1 * matrice2</code>	Prodotto elemento per elemento tra due matrici
<code>matrice1 / matrice2</code>	Rapporto elemento per elemento tra due matrici
<code>matrice1 %*% matrice2</code>	Prodotto matriciale

TABELLA B.4: Operazioni con matrici

B.4 DataFrames

B.4.1 Creazione di DataFrames

Uno degli oggetti più utilizzati in R sono i DataFrames. I DataFrames permettono di raccogliere all'interno di uno stesso oggetto vettori di diverso tipo (i.e., vettori numerici, logici, fattori o stringhe di caratteri). Per questo motivo, i DataFrames sono utili per riportare tutti i dati riguardanti le diverse variabili misurate in un esperimento.

In genere ogni riga di un DataFrames rappresenta una singola osservazione e nelle colonne sono riportate i vari valori delle variabili misurate.

Esistono due formati principali di DataFrames:

- **Wide:** ogni singola riga rappresenta un soggetto e ogni sua risposta o variabile misurata sarà riportata in una diversa colonna.

	Id	age	sex	item_1	item_2	item_3
1	subj_1	21	F	2	0	2
2	subj_2	23	M	1	2	0
3	subj_3	19	F	1	1	1

- **Long:** ogni singola riga rappresenta una singola osservazione. Quindi i dati di ogni soggetto saranno riportati su più righe e le variabili che non cambiano tra le osservazioni saranno ripetute.

	Id	age	sex	item	response
1	subj_1	21	F	1	2
2	subj_1	21	F	2	1
3	subj_1	21	F	3	1
4	subj_2	23	M	1	0
5	subj_2	23	M	2	2
6	subj_2	23	M	3	1
7	subj_3	19	F	1	2
8	subj_3	19	F	2	0
9	subj_3	19	F	3	1

In R per definire un DataFrame si utilizza il comando:

```
nome_DataFrame <- data.frame(variabile_1=c(...), ..., variabile_s=c(...))
```

All'interno vanno riportate le variabili che si vogliono inserire separate da virgole. Ogni variabile deve avere la **stessa lunghezza**.

Nota: di default R considera una variabile stringa all'interno di un DataFrame come una variabile categoriale. È possibile cambiare questa opzione specificando `stringsAsFactors=FALSE`.

Esercizi

Esercizi:

1. Crea il dataframe `data_wide` riportato precedentemente.
2. Crea il dataframe `data_long` riportato precedentemente.

B.4.2 Selezione di elementi di un DataFrame

In R per selezionare gli elementi di un DataFrame si può, analogamente alle matrici, indicare all'interno delle parentesi quadre l'indice di riga e l'indice di colonna (**separati da virgola**).

```
nome_DataFrame[indice_riga , indice_colonna]
```

Per accedere ad una specifica variabile del DataFrame è possibile utilizzare l'operatore "\$":

```
nome_DataFrame$nome_variabile
```

Per quanto riguarda l'indice di riga è possibile definire una condizione logica rispetto ad una variabile, mentre per l'indice di colonna si può indicare il nome delle variabili:

```
nome_DataFrame[condizione_logica , c("variabile_1", ..., "variabile_s")]
```

Nota: per selezionare tutti gli elementi di una data riga basta lasciare vuoto l'indice di colonna.

Esempio: `data_wide[data_wide$sex=="F", c("Id","age")]`

Esercizi

1. Utilizzando gli **indici numerici** di riga e di colonna seleziona i dati del soggetto `subj_2` riguardanti le variabili `item` e `response` dal DataFrame `data_long`.
2. Compi la stessa selezione dell'esercizio precedente usando però questa volta una condizione logica per gli indici di riga e indicando direttamente il nome delle variabili per gli indici di colonna.
3. Considerando il DataFrame `data_wide` seleziona le variabili `Id` e `sex` dei soggetti che hanno risposto 1 alla variabile `item_1`.
4. Considerando il DataFrame `data_long` seleziona solamente i dati riguardanti le ragazze con età superiore ai 20 anni.
5. Elimina dal DataFrame `data_long` le osservazioni riguardanti il soggetto `subj_2` e la variabile `"sex"`.

B.4.3 Funzioni con DataFrames

<code>nome_DataFrame <- cbind(nome_DataFrame, nuova_variabile) nome_DataFrame\$nome_variabile <- dati</code>	Per aggiungere una nuova variabile al DataFrame (deve avere lo stesso numero di righe)
<code>nome_DataFrame <- rbind(nome_DataFrame, nuovi_dati)</code>	Per aggiungere delle osservazioni (i nuovi dati devono essere coerenti con la struttura del DataFrame)
<code>nrow(nome_DataFrame)</code>	Per valutare il numero di osservazioni del DataFrame
<code>ncol(nome_DataFrame)</code>	Per valutare il numero di variabili del DataFrame
<code>colnames(nome_DataFrame) names(nome_DataFrame)</code>	Nomi delle colonne del DataFrame
<code>rownames(nome_DataFrame)</code>	Nomi delle righe del DataFrame

TABELLA B.5: Operazioni con DataFrames

Nota: È possibile assegnare nomi alle colonne e righe di un DataFrame allo stesso modo delle matrici, attraverso i comandi

```
colnames(nome_DataFrame) <- c("nome_1", ..., "nome_s")
names(nome_DataFrame) <- c("nome_1", ..., "nome_s")
rownames(nome_DataFrame) <- c("nome_1", ..., "nome_n")
```

Esercizi

1. Aggiungi sia al DataFrame `data_wide` che `data_long` la variabile numerica `"memory_pre"`.

```

      Id memory_pre
1 subj_1          3
2 subj_2          2
3 subj_3          1

```

2. Aggiungi sia al DataFrame `data_wide` che `data_long` la variabile categoriale "gruppo".

```

      Id      gruppo
1 subj_1 trattamento
2 subj_2 trattamento
3 subj_3 controllo

```

3. Aggiungi al DataFrame `data_wide` i dati del soggetto `subj_4` e `subj_5`.

```

      Id age sex item_1 item_2 item_3 memory_pre      gruppo
1 subj_4  25  F      1      0      2          1 trattamento
2 subj_5  22  M      1      1      0          3 controllo

```

4. Considerando il DataFrame `data_wide` calcola la variabile "memory_post" data dalla somma degli item.
5. Considerando il DataFrame `data_wide` cambia i nomi delle variabili `item_1`, `item_2` e `item_3` rispettivamente in `problem_1`, `problem_2` e `problem_3`.

B.5 Liste

B.5.1 Creazione di Liste

Le liste sono degli speciali oggetti in R che permettono di contenere al loro interno altri oggetti indipendentemente dalla loro tipologia. Possiamo quindi avere nella stessa lista sia vettori, sia matrici sia DataFrames.

In R per definire una lista si utilizza il comando:

```
nome_Lista <- list(nome_oggetto_1 = oggetto_1, ..., nome_oggetto_n = oggetto_n)
```

All'interno si possono riportare vari oggetti se vogliono inserire con i relativi nomi, separati da virgole.

Esercizi

1. Crea la lista `esperimento_1` contenente:
 - il DataFrame `data_wide`
 - la matrice `A`
 - il vettore `x`
 - la variabile `info = "Hello world!"`
2. Crea la lista `esperimento_2` contenente:
 - il DataFrame `data_long`
 - la matrice `C`
 - il vettore `y`
 - la variabile `info = "Prima raccolta dati"`

B.5.2 Selezione di elementi di una lista

In R per selezionare gli elementi di una lista si possono usare le doppie parentesi quadre indicando l'indice della posizione dell'oggetto che si vuole selezionare:

```
nome_lista[[indice_posizione]]
```

In alternativa, se i nomi degli oggetti sono stati specificati, è possibile utilizzare l'operatore "\$" e il nome dell'oggetto da selezionare all'interno della lista:

```
nome_lista$nome_oggetto
```

In seguito per accedere a specifici elementi all'interno degli oggetti si utilizzano le stesse norme precedentemente presentate a seconda del tipo di oggetto.

Esempio: `esperimento_1[[2]][,2]`

```
esperimento_1$data_wide$age
```

Nota: per definire o cambiare i nomi degli oggetti contenuti in una lista è possibile utilizzare la funzione: `names(nome_lista) <- c(nome_oggetto_1, ..., nome_oggetto_n)`

Esercizi

1. Utilizzando gli **indici numerici** di posizione seleziona i dati dei soggetti `subj_1` e `subj_4` riguardanti le variabili `age`, `sex` e `gruppo` dal DataFrame `data_wide` contenuto nella lista `esperimento_1`.
2. Compila la stessa selezione dell'esercizio precedente usando però questa volta il nome dell'oggetto per selezionare il DataFrame dalla lista.
3. Considerando la lista `esperimento_2` seleziona gli oggetti `data_long`, `y` e `info`.
4. Cambia i nomi degli oggetti contenuti nella lista `esperimento_2` rispettivamente in `"dati_esperimento"`, `"matrice_VCV"`, `"codici_Id"` e `"note"`

B.6 Tipi di vettori

In R ci sono 4 tipi differenti di vettori: numerici, logici, caratteri e fattori.

B.6.1 Vettori numerici

I vettori numerici sono utilizzati per compiere operazioni aritmetiche, in R sono indicati come `num`. In R ci sono è possibile specificare se i numeri contenuti nel vettore sono numeri interi, avremmo quindi un vettore di valori interi (indicato in R come `int`). Per fare ciò è possibile aggiungere `L` ad ogni valore numerico nel definire il vettore oppure usare la funzione `as.integer()` per trasformare un vettore numerico in un vettore intero.

Esempio:

```
> x <- c(4L, 6L, 12L, 34L, 8L)
>
> x <- as.integer(c(4, 6, 12, 34, 8))
```

Nota: per trasformare un vettore intero in un vettore numerico è possibile usare la funzione `as.numeric()`.

B.6.2 Vettori logici

I vettori logici sono formati dai valori `TRUE` e `FALSE`, che possono essere abbreviati rispettivamente in `T` e `F`. In R i vettori logici sono indicati come `logic`. In genere, i vettori logici sono il risultato delle operazioni in cui viene chiesto ad R di valutare la condizione logica di una proposizione.

```
> x > 10
```

```
[1] FALSE FALSE TRUE TRUE FALSE
```

Nota: in R, come in molti altri software di programmazione, `TRUE` assume il valore numerico 1 e `FALSE` assume il valore 0.

```
> sum(x>10)
```

```
[1] 2
```

È possibile trasformare un vettore numerico in un vettore logico attraverso la funzione `as.logical()`, gli 0 assumeranno il valore `FALSE` mentre qualsiasi altro numero assumerà il valore `TRUE`.

```
> as.logical(c(1,0,.034,-1,0,8))
```

```
[1] TRUE FALSE TRUE TRUE FALSE TRUE
```

B.6.3 Vettori di caratteri

I vettori di caratteri contengono stringhe di caratteri e sono indicati in R con `chr`. Non è possibile eseguire operazioni aritmetiche con vettori di caratteri ma solo valutare se due stringhe sono uguali o differenti.

```
> j<-c("Hello","World","hello","world")
> j=="hello"
```

```
[1] FALSE FALSE TRUE FALSE
```

Per trasformare un vettore qualsiasi in un vettore di caratteri è possibile usare la funzione `as.character()`.

```
> as.character(x)
```

```
[1] "4" "6" "12" "34" "8"
```

```
> as.character(x>10)
```

```
[1] "FALSE" "FALSE" "TRUE" "TRUE" "FALSE"
```

B.6.4 Fattori

I fattori sono utilizzati per definire delle variabili categoriali, sono indicati in R con `Factor`. Per creare una variabile categoriale in R si utilizza la funzione:

```
nome_variabile<-factor(c(..., data, ...), levels=c(...))
```

L'opzione `levels=c(...)` è usata per specificare quali sono i possibili livelli della variabile categoriale. È possibile modificare o aggiungere nuovi livelli della variabile anche in un secondo momento utilizzando la funzione:

```
levels(nome_fattore)<- c(..., nuovi_livelli, ...)
```

Nota: nel creare un fattore R associa ad ogni livello un valore in ordine crescente e assegna agli elementi del vettore il loro valore numerico a seconda del proprio livello. Pertanto se un fattore è trasformato in un vettore numerico vengono restituiti tali valori numerici e non i livelli anche nel caso fossero dei numeri. Prendiamo per esempio la variabile `anni_istruzione`

```
> anni_istruzione<-factor(c(11,8,4,8,11,4,11,8))
```

```
> anni_istruzione
```

```
[1] 11 8 4 8 11 4 11 8
Levels: 4 8 11
```

```
> as.numeric(anni_istruzione)
```

```
[1] 3 2 1 2 3 1 3 2
```

Per riottenere gli esatti valori numerici è necessario eseguire

```
> as.numeric(as.character(anni_istruzione))
```

```
[1] 11 8 4 8 11 4 11 8
```

Esercizi

1. Crea la variabile categoriale **sex** così definita:

```
[1] M F M F M F F F M
Levels: F M
```

2. Rinomina i livelli della variabile **sex** rispettivamente in "donne" e "uomini".
3. Crea la variabile categoriale **intervento** così definita:

```
[1] CBT          Psicanalisi CBT          Psicanalisi CBT          Psicanalisi
[7] Controllo   Controllo   CBT
Levels: CBT Controllo Psicanalisi
```

4. Correggi nella variabile **intervento** la 7° e 8° osservazione con la voce **Farmaci**.
5. Aggiungi alla variabile **intervento** le seguenti nuove osservazioni:

```
[1] "Farmaci" "Controllo" "Farmaci"
```

bozza II

C

Licenza GNU GPL

GNU GENERAL PUBLIC LICENSE (Version 3, 29 June 2007)

Copyright © 2007 Free Software Foundation, Inc. <https://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Abstract

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program—to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS

0. Definitions.

“This License” refers to version 3 of the GNU General Public License.

“Copyright” also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

“The Program” refers to any copyrightable work licensed under this License. Each licensee is addressed as “you”. “Licensees” and “recipients” may be individuals or organizations.

To “modify” a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a “modified version” of the earlier work or a work “based on” the earlier work.

A “covered work” means either the unmodified Program or a work based on the Program.

To “propagate” a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To “convey” a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays “Appropriate Legal Notices” to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The “source code” for a work means the preferred form of the work for making modifications to it. “Object code” means any non-source form of a work.

A “Standard Interface” means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The “System Libraries” of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A “Major Component”, in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The “Corresponding Source” for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work’s System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- (a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- (b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".
- (c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- (d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate"

if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

- (a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.
- (b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- (c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- (d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.
- (e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section

must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

7. Additional Terms.

“Additional permissions” are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- (a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or
- (b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or
- (c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or
- (d) Limiting the use for publicity purposes of names of licensors or authors of the material; or
- (e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or
- (f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered “further restrictions” within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An “entity transaction” is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party’s predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

11. Patents.

A “contributor” is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor’s “contributor version”.

A contributor’s “essential patent claims” are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, “control” includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor’s essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a “patent license” is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To “grant” such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. “Knowingly relying” means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient’s use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is “discriminatory” if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others’ Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License “or any later version” applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version

number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.>

Copyright (C) <textyear> <name of author>

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful,

but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see [<https://www.gnu.org/licenses/>](https://www.gnu.org/licenses/).

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

```
<program> Copyright (C) <year> <name of author>
```

This program comes with ABSOLUTELY NO WARRANTY; for details type 'show w'. This is free software, and you are welcome to redistribute it under certain conditions; type 'show c' for details.

The hypothetical commands `show w` and `show c` should show the appropriate parts of the General Public License. Of course, your program's commands might be different; for a GUI interface, you would use an "about box".

You should also get your employer (if you work as a programmer) or school, if any, to sign a "copyright disclaimer" for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <https://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <https://www.gnu.org/licenses/why-not-lgpl.html>.

bozza non corretta

Bibliografia

- [1] Adelchi Azzalini and Bruno Scarpa. *Data analysis and data mining: An introduction*. Oxford University Press, 2012.
- [2] Claudio Barbaranelli and Elena Natali. *I test psicologici: teorie e modelli psicometrici*. Carocci, 2011.
- [3] Birgitta Berglund, Giovanni B Rossi, James T Townsend, and Leslie R Pendrill. *Measurement with persons: theory, methods, and implementation areas*. Psychology Press, 2012.
- [4] IFCC ILAC IUPAC IUPAP ISO OIML (2012) BIPM, IEC. *The international vocabulary of metrology—basic and general concepts and associated terms (VIM)*. JCGM 200:2012, <http://www.bipm.org/vim>, 2012.
- [5] Kenneth A. Bollen. *Structural Equations with Latent Variables*. Wiley, 1989.
- [6] Francesco Caravenna and Paolo Dai Pra. *Probabilità: Un'introduzione attraverso modelli e applicazioni*, volume 67. Springer Science & Business Media, 2013.
- [7] Giuseppe Cicchitelli. *Statistica. Principi e metodi*. Pearson, 2008.
- [8] Piergiorgio Corbetta. *Metodi di analisi multivariata per le scienze sociali*. il Mulino, 1992.
- [9] Linda Crocker and James Algina. *Introduction to classical and modern test theory*. ERIC, 1986.
- [10] Sacha Epskamp, Mijke Rhemtulla, and Denny Borsboom. Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, 82:904–927, 2017.
- [11] Luigi Fabbris. *Statistica Multivariata*. McGraw-Hill, 1997.
- [12] Ian Farrance and Robert Frenkel. Uncertainty of measurement: a review of the rules for calculating uncertainty components through functional relationships. *The Clinical Biochemist Reviews*, 33(2):49, 2012.
- [13] Alessandro Ferrero, Dario Petri, Paolo Carbone, and Marcantonio Catelani. *Modern measurements: fundamentals and applications*. John Wiley & Sons, 2015.
- [14] Roberto Festa. *Cambiare opinione. Temi e problemi di epistemologia bayesiana*. CLUEB, 1996.
- [15] James M Graham. Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and psychological measurement*, 66(6):930–944, 2006.
- [16] SV Gupta. *Measurement uncertainties: physical parameters and calibration of instruments*. Springer Science & Business Media, 2012.
- [17] Ifan Hughes and Thomas Hase. *Measurements and their uncertainties: a practical guide to modern error analysis*. Oxford University Press, 2010.
- [18] Raghu N Kacker. Probability distributions and coverage probability in gum, jcgM documents, and statistical inference. *Measurement*, 65:61–70, 2015.
- [19] George J Klir and Mark J Wierman. *Uncertainty-based information: elements of generalized information theory*, volume 15. Physica, 2013.
- [20] Frederic M Lord and Melvin R Novick. *Statistical theories of mental test scores*. IAP, 2008.
- [21] Luca Mari, Valentina Lazzarotti, and Raffaella Manzini. Measurement in soft systems: epistemological framework and a case study. *Measurement*, 42(2):241–253, 2009.
- [22] Roderick P. McDonald. *Test theory: A unified approach*. Lawrence Erlbaum Associates, 1999.

- [23] William G. Meredith. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58:525–543, 1993.
- [24] Joel Michell. Quantitative science and the definition of measurement in psychology. *British journal of Psychology*, 88(3):355–383, 1997.
- [25] Joel Michell. *Measurement in psychology: A critical history of a methodological concept*, volume 53. Cambridge University Press, 1999.
- [26] Joel Michell. Constructs, inferences, and mental measurement. *New Ideas in Psychology*, 31(1):13–21, 2013.
- [27] Samir Okasha. *Il primo libro di filosofia della scienza*. Einaudi, 2006.
- [28] Massimiliano Pastore. Tra alfa e omega c'è di mezzo la cfa? *Giornale Italiano di Psicologia*, 3:761–782, 2017.
- [29] Domenico Piccolo. *Statistica per le decisioni: la conoscenza umana sostenuta dall'evidenza empirica*. Il mulino, 2010.
- [30] Calyampudi Radhakrishna Rao and Sandip Sinharay. Handbook of statistics 26: Psychometrics. 2006.
- [31] Sidney Resnick. *A probability path*. Birkhauser Verlag AG, 2003.
- [32] Sheldon M Ross. *Calcolo delle probabilità*. Maggioli Editore, 2013.
- [33] Simona Salicone. *Measurement Uncertainty: an approach via the mathematical theory of evidence*. Springer Science & Business Media, 2007.
- [34] Stanley Smith Stevens et al. On the theory of scales of measurement. 1946.
- [35] Ross E Traub. Classical test theory in historical perspective. *Educational Measurement*, 16:8–13, 1997.
- [36] Vincenzo Esposito Vinzi, Laura Trinchera, and Silvano Amato. PIs path modeling: From foundations to recent developments and open issues for model assessment and improvement. 2010.