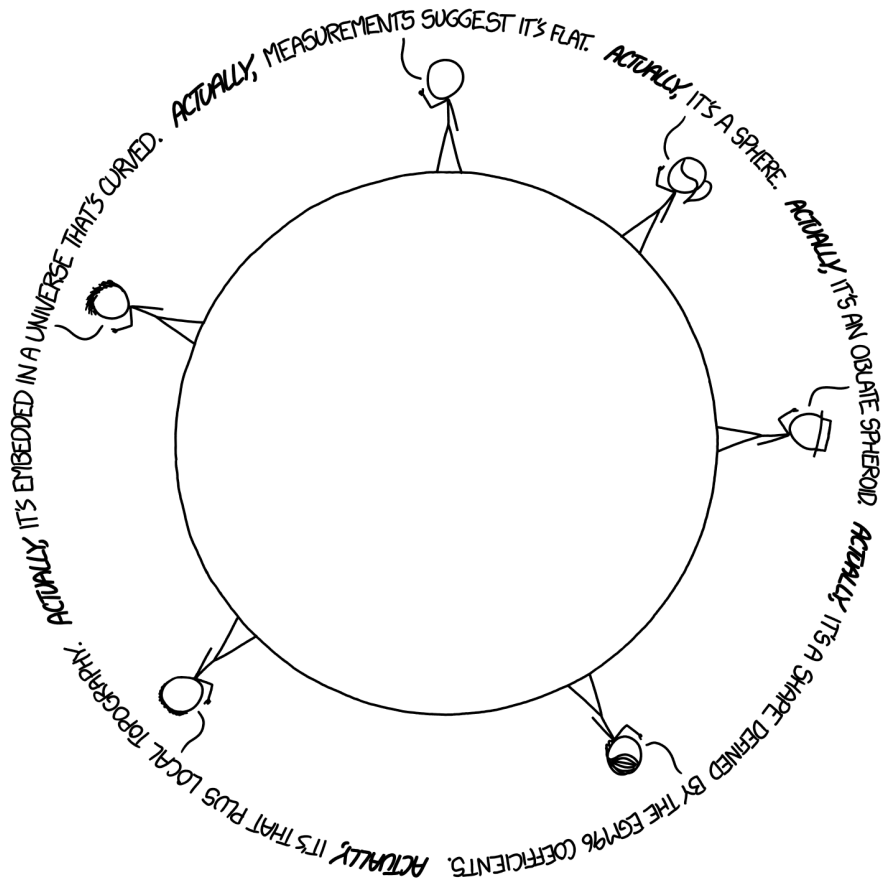


Corrado Caudek

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



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





Indice

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

2	Curve di crescita latente	23
2.1	Dati longitudinali	23
2.1.1	Un esempio concreto	24
2.1.2	Un secondo esempio	34
II	Appendici	1
	Appendice	3
A	Simbologia di base	3

Elenco delle figure

1.1	Simulazione della relazione tra punteggio osservato e punteggio vero per 100 individui in base alle assunzioni della CTT.	14
2.1	Modello di crescita latente.	28



Elenco delle tabelle



Prefazione

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

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

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

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

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

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

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

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

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

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

xiv

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

Prefazione

Corrado Caudek
Marzo 2022

Parte I

La teoria classica dei test



1

Fondamenti teorici

1.1 Valutazione psicometrica come ragionamento inferenziale

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

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

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

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

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

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

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

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

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

1.2 La Teoria Classica

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

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

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

quello di ottenere una stima di σ_E .

1.3 Le due componenti del punteggio osservato

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

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

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

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

1.3.1 Il punteggio vero

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

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

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

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

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

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

1.3.2 Somministrazioni ripetute

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

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

1.3.3 Le assunzioni sul punteggio ottenuto

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

$$\mu_X \equiv \mu_T,$$

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

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

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

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

1.4 L'errore standard della misurazione σ_E

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

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

1.5 Assiomi della Teoria Classica

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

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

e incorrelate con il punteggio vero,

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

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

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

La quantità σ_E è detta errore standard della misurazione.

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

1.6 L'attendibilità del test

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

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

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

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

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

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

1.6.1 La varianza del punteggio osservato

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

Dimostrazione. La varianza del punteggio osservato è uguale a

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

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

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

□

1.6.2 La covarianza tra punteggio osservato e punteggio vero

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

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

□

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

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

1.6.3 Definizione e significato dell'attendibilità

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

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

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

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

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

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

1.7 Attendibilità e modello di regressione lineare

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

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

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

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

1.7.1 Simulazione

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

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

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

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

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

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

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

Le prime 6 osservazioni così ottenute sono:

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

Un diagramma di dispersione è fornito nella figura seguente:

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

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

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

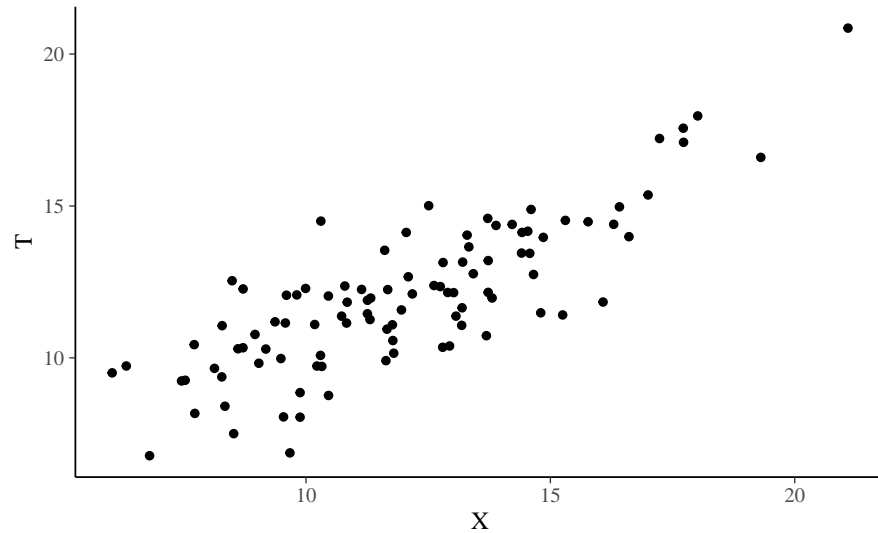


Figura 1.1: Simulazione della relazione tra punteggio osservato e punteggio vero per 100 individui in base alle assunzioni della CTT.

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

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

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

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

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

della varianza degli errori E :

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

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

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

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

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

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

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

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

$$X = a + bT + E$$

```

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

```

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

```

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

```

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

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

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

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

1.8 Misurazioni parallele e affidabilità

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

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

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

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

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

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

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

Dimostrazione. Per la misurazione parallela X abbiamo

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

per la misurazione parallela X' abbiamo

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

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

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

1.8.1 La correlazione tra misurazioni parallele

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

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

e

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

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

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

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

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

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

1.8.2 La correlazione tra due forme parallele del test

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

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

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

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

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

□

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

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

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

1.8.3 La correlazione tra punteggio osservato e punteggio vero

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

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

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

1.8.4 I fattori che influenzano l'attendibilità

Considerando le precedenti tre equazioni

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

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

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

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

1.9 Metodi alternativi per la stima del coefficiente di attendibilità

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

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

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

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

2

Curve di crescita latente

2.1 Dati longitudinali

Un importante classe di modelli a variabili latenti è quella dei modelli delle curve di crescita latente (*latent growth models*, LGM). I modelli delle curve di crescita latente vengono spesso utilizzati per analizzare dati longitudinali. In questo tipo di dati, una misura di esito viene ottenuta in diversi momenti del tempo e si vuole studiare il cambiamento nel tempo. In molti casi, la traiettoria nel tempo può essere modellata come una semplice funzione lineare o quadratica. Le differenze individuali vengono catturate dagli *effetti random* che sono convenientemente rappresentati da variabili latenti (continue), spesso chiamate *fattori di crescita* (*growth factors*).

Possiamo pensare ai modelli LGM come ad un'estensione del modello CFA dotato di `meanstructure`. L'inclusione della `meanstructure` significa che non possiamo usare in input la matrice di covarianza campionaria, ma dobbiamo invece utilizzare i dati grezzi (ovvero, le singole osservazioni per ciascun partecipante). Un'altro requisito degli LGM è che i dati devono essere forniti del formato `wide`, il che significa che ogni colonna rappresenta la variabile di esito in un diverso momento nel tempo. Si presume che ogni osservazione o riga sia indipendente dalle altre; le colonne mostrano invece una dipendenza temporale tra loro.

I modelli LGM sono un caso speciale di CFA e corrispondono un modello CFA a due fattori in cui le saturazioni fattoriali sono fissate a valori predefiniti. Nello specifico, le saturazioni sul fattore che specifica le intercette delle funzioni individuali di crescita sono tutte fissate a 1; le saturazioni sul fattore che specifica le pendenze delle funzioni individuali di crescita sono fissate in modo tale da corrispondere ai diversi momenti del tempo, laddove la saturazione in corrispondenza del momento t_0 è uguale a 0 e la progressione procede con incrementi di 1 (ovvero, $0, 1, \dots, t - 1$).

Si giunge così alla seguente specificazione del modello fattoriale:

$$y_{it} = \tau_i + (1)\xi_1 + (t)\xi_2 + \delta_t, \quad (2.1)$$

con $t = 0, 1, 2, 3$ nel caso di quattro misurazioni temporali. Si noti che, per disegno, $\tau_i = 0$.

2.1.1 Un esempio concreto

L'esempio che discuteremo utilizza i dati di un tutorial `lavaan`, ovvero un campione di dati artificiali chiamato `Demo.growth` in cui un punteggio viene misurato su quattro punti temporali. I dati sono i seguenti:

```
data(Demo.growth)
glimpse(Demo.growth)
#> Rows: 400
#> Columns: 10
#> $ t1 <dbl> 1.7256, -1.9842, 0.3195, 0.7769, 0.4489, ~
#> $ t2 <dbl> 2.1424, -4.4006, -1.2691, 3.5314, -0.7728, ~
#> $ t3 <dbl> 2.7732, -6.0166, 1.5600, 3.1382, -1.5035, ~
#> $ t4 <dbl> 2.51596, -7.02962, 2.86853, 5.36374, 0.078~
#> $ x1 <dbl> -1.1641, -1.7454, 0.9202, 2.3595, -1.0887, ~
#> $ x2 <dbl> 0.1742, -1.5769, -0.1418, 0.7080, -1.0100, ~
#> $ c1 <dbl> -0.02768, -2.03197, 0.05237, 0.01911, 0.65~
#> $ c2 <dbl> 0.55492, 0.12533, -1.25774, 0.64738, 0.730~
#> $ c3 <dbl> 0.25448, -1.56423, -1.80339, -0.43238, -0.~
#> $ c4 <dbl> -1.00640, 1.22927, -0.32726, -1.03240, -0.~
```

La variabile dipendente è rappresentata dalle quattro colonne chiamate `t1`, `t2`, `t3` e `t4` che corrispondono alla serie temporale con quattro misurazioni per ciascun soggetto.

Trasformiamo i dati in formato `long`:

```
demo_growth_long <- Demo.growth %>%
  dplyr::select(t1, t2, t3, t4) %>%
  pivot_longer(
    cols = starts_with("t"),
    names_to = "t",
```

```

    values_to = "y"
  ) %>%
  as.data.frame()
demo_growth_long$time <- rep(0:3, 400)
demo_growth_long$id <- rep(1:400, each = 4)

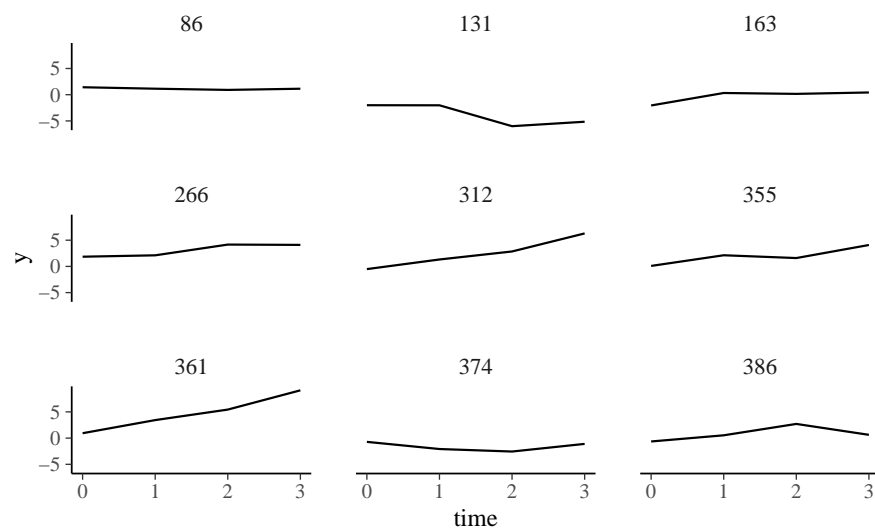
```

Esaminiamo un campione casuale di 9 soggetti. È presente una notevole variazione da soggetto a soggetto:

```

id_sel <- sample(1:400, 9)
d <- demo_growth_long[demo_growth_long$id %in% id_sel, ]
d %>%
  ggplot(aes(x = time, y = y)) +
  geom_line() +
  facet_wrap(~id)

```



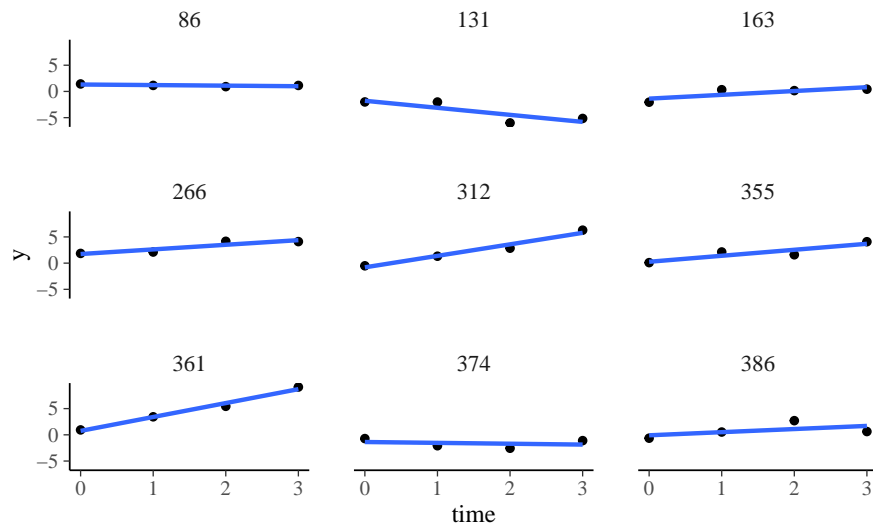
Notiamo che un semplice modello lineare è appropriato per rendere conto della variazione temporale della variabile risposta:

```

d %>%
  ggplot(aes(x = time, y = y)) +

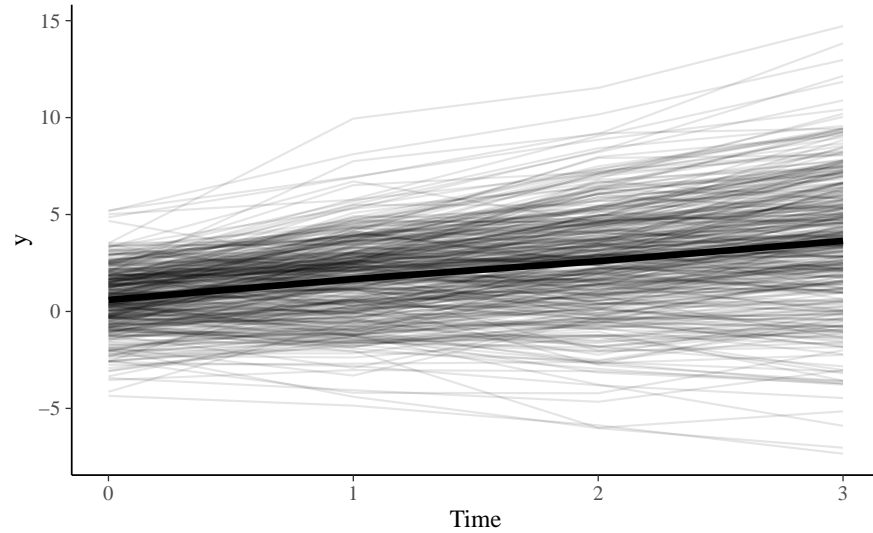
```

```
geom_point() +
stat_smooth(method = "lm", se = FALSE) +
facet_wrap(~id)
```



Complessivamente, i dati suggeriscono un andamento crescente della variabile risposta in funzione del tempo:

```
demo_growth_long %>%
  ggplot(aes(time, y, group = id)) +
  geom_line(alpha = 0.1) + # add individual line with transparency
  stat_summary( # add average line
    aes(group = 1),
    fun = mean,
    geom = "line",
    size = 1.5,
    color = "black"
  ) +
  labs(x = "Time", y = "y")
```

Per adattare un modello lineare di crescita a questi dati, specifichiamo il seguente modello a variabili latenti

$$y_j = \alpha_0 + \alpha_1 \lambda_j + \zeta_{00} + \zeta_{11} \lambda_j + \epsilon_j,$$

dove

- y_j è la variabile di interesse che cambia nel tempo, j .
- α_0 rappresenta l'intercetta della retta di regressione al tempo $t = 0$ (il punto di partenza della linea nera sopra).
- $\alpha_1 \lambda_j$ è il tasso medio di variazione nel tempo (la pendenza della linea nera nel grafico sopra). Qui λ_j è solo l'indice dei punti temporali considerati (0, 1, 2, 3).
- ζ_{00} è la varianza tra i soggetti nel punto $t = 0$.
- $\zeta_{11} \lambda_j$ è la varianza del tasso di variazione tra i soggetti.
- ϵ_j è la varianza di ciascun soggetto attorno alla sua retta di regressione.

Tali relazioni statistiche vengono rappresentate dal modello di equazioni strutturali della figura 2.1.

In altre parole, un lineare di crescita latente corrisponde ad un modello fattoriale con due variabili latenti: un fattore che corrisponde al “punteggio vero” delle intercette individuali e un fattore che corrisponde al “punteggio vero” delle pendenze individuali. Nella sintassi `lavaan` questo diventa:

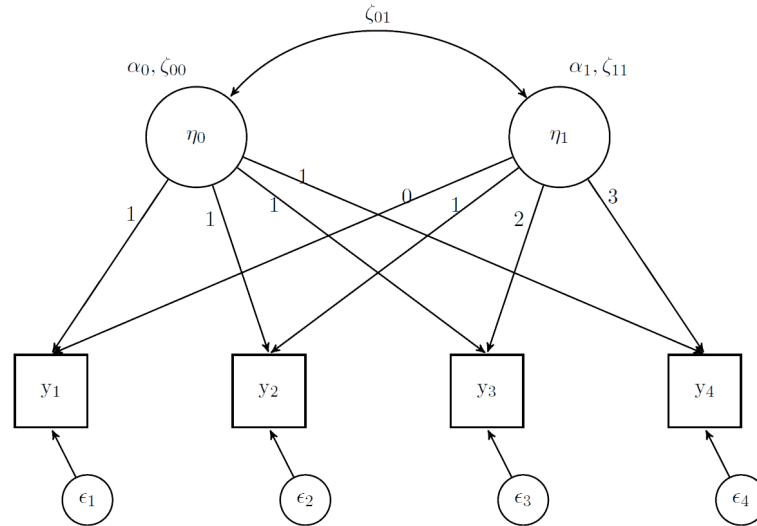


Figura 2.1: Modello di crescita latente.

```
model <- "
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4
"
```

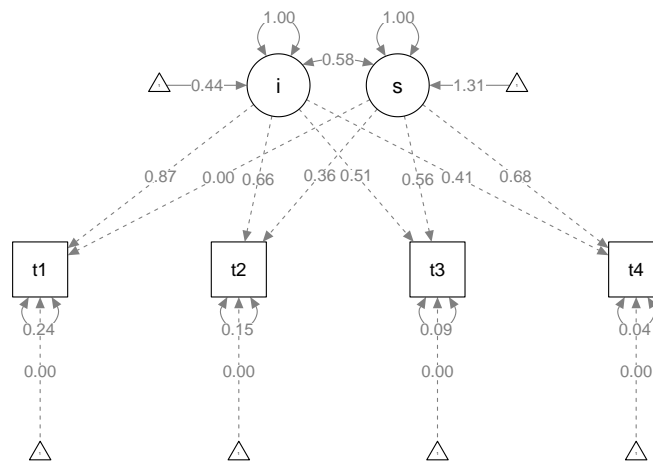
Si noti che, per il fattore η_0 (che rappresenta le intercette), i valori delle saturazioni fattoriali sono fissate a 1 – questo è il motivo per cui α_0 e ζ_{00} compaiono da soli nell’equazione precedente: in maniera esplicita sono $1 \cdot \alpha_0$ e $1 \cdot \zeta_{00}$. Le saturazioni per il fattore η_1 (che specifica le pendenze delle funzioni lineari) sono fissate ai valori dei diversi momenti nel tempo: qui i valori λ_j da 0 a 3. Abbiamo anche la correlazione tra η_0 e η_1 , rappresentata dalla doppia freccia ζ_{01} . Se questo parametro è positivo, questo significa che i partecipanti tendono a diventare sempre più diversi tra loro, al passare del tempo; un’interpretazione opposta si ha se il valore del parametro è negativo.

Adattiamo il modello ai dati:

```
fit <- growth(model, data = Demo.growth)
```

Esaminiamo il path diagram:

```
semPaths(
  fit, "std",
  posCol = c("black"),
  edge.label.cex = 0.9,
  sizeMan = 7,
  what = "path"
)
```



Ci sono 6 tipi di parametri di interesse:

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

	x
t1~~t1	0.5953
t2~~t2	0.6760
t3~~t3	0.6349
t4~~t4	0.5076
i~~i	1.9320
s~~s	0.5869
i~~s	0.6179
i~1	0.6147

	x
s~1	1.0063

- l'intercetta $i = 0.615$ è il valore atteso della variabile risposta al momento t_0 ;
- la pendenza $s = 1.006$ è il tasso di cambiamento medio della variabile risposta nel tempo. Ad ogni successivo momento temporale, il valore medio della variabile risposta aumenta in media di 1.006 punti;
- varianza $i = 1.932$ misura la variazione tra i soggetti al momento t_0 (ci dice quanto sono diverse le intercette delle rette di regressione tra i soggetti);
- varianza $s = 0.587$ misura la variazione del tasso di crescita tra i soggetti (ci dice quanto sono diverse le pendenze delle rette di regressione tra i soggetti);
- varianze **t1**, ..., **t4**: i valori da 0.595 a 0.508 ci dicono la variazione tra i soggetti all'interno di ciascun momento del tempo;
- la covarianza tra **i** e **s** = 0.618 ci dice che i valori della variabile risposta diventano via via più diversi nel tempo tra i rispondenti (un valore negativo avrebbe l'interpretazione opposta).

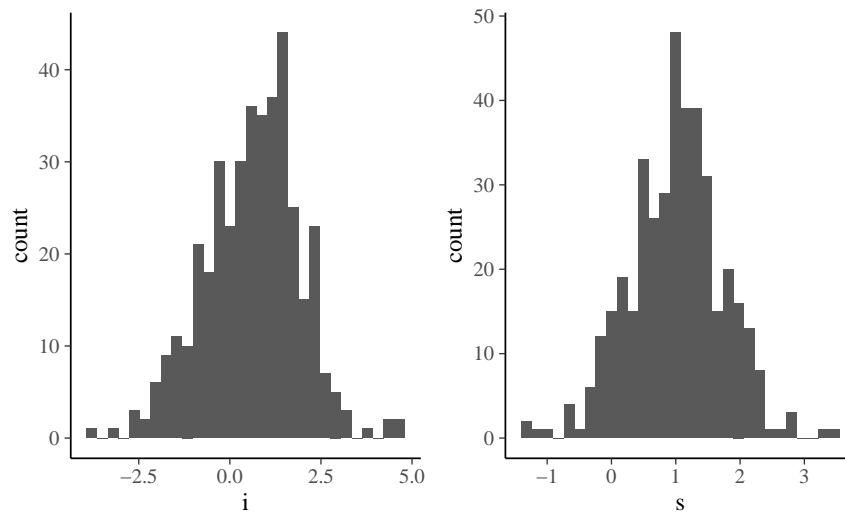
Le stime dell'intercetta e della pendenza della funzione di crescita per ciascun partecipante si ottengono nel modo seguente:

```
rand_eff <- as.data.frame(lavPredict(fit))
head(rand_eff)
#>      i      s
#> 1  1.2276  0.60278
#> 2 -2.6796 -1.38498
#> 3 -0.2955  0.88828
#> 4  1.1576  1.34051
#> 5 -0.4356  0.03193
#> 6 -1.3122  0.50259
```

Le distribuzioni delle stime individuali dell'intercetta e della pendenza della curva di crescita si ottengono nel modo seguente:

```
gi <- rand_eff %>%
  ggplot(aes(x = i)) +
```

```
geom_histogram()
gh <- rand_eff %>%
  ggplot(aes(x = s)) +
  geom_histogram()
gi + gh
```



Un buon modo per capire cosa fa il modello è quello di visualizzare i punteggi previsti. Useremo qui la funzione `predict()` per salvare un nuovo oggetto con i punteggi previsti a livello individuale per l'intercetta e la pendenza;

```
pred_lgm <- predict(fit)
head(pred_lgm)
#>           i           s
#> [1,]  1.2276  0.60278
#> [2,] -2.6796 -1.38498
#> [3,] -0.2955  0.88828
#> [4,]  1.1576  1.34051
#> [5,] -0.4356  0.03193
#> [6,] -1.3122  0.50259
```

Questi sono i valori previsti dal modello per ciascun partecipante. Se

calcoliamo la media di queste variabili otteniamo gli stessi risultati che sono stati riportati sopra:

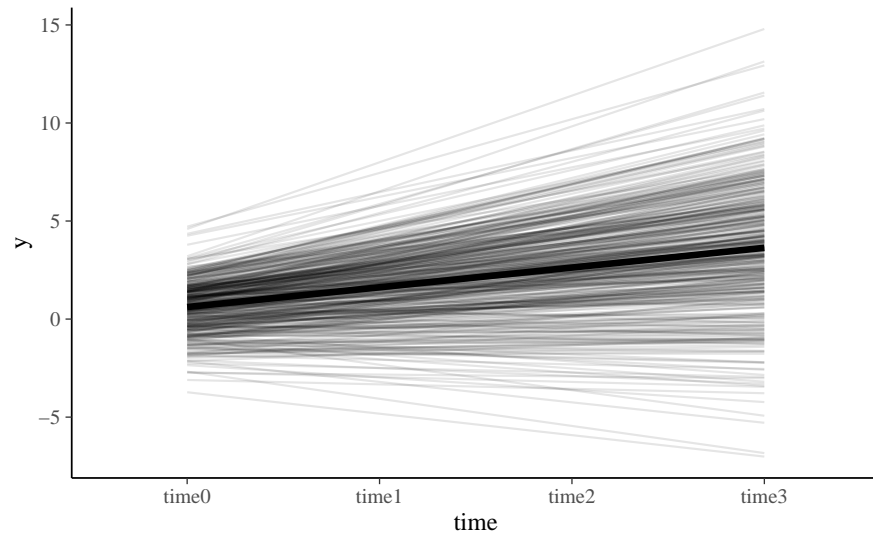
```
# average of the intercepts (first column)
mean(pred_lgm[, 1])
#> [1] 0.6147
# average of the slope (second column)
mean(pred_lgm[, 2])
#> [1] 1.006
```

Il cambiamento nel tempo previsto dal modello per il soggetto j -esimo è

$$y_j = \eta_0 + \lambda_j \eta_1.$$

Il cambiamento previsto per tutti i soggetti può essere visualizzato nel modo seguente:

```
# create long data for each individual
pred_lgm_long <- map(
  0:3, # loop over time
  function(x) pred_lgm[, 1] + x * pred_lgm[, 2]
) %>%
  reduce(cbind) %>% # bring together the wave predictions
  as.data.frame() %>% # make data frame
  setNames(str_c("time", 0:3)) %>% # give names to variables
  mutate(id = row_number()) %>% # make unique id
  gather(-id, key = time, value = pred) # make long format
# make graph
pred_lgm_long %>%
  ggplot(aes(time, pred, group = id)) + # what variables to plot?
  geom_line(alpha = 0.1) + # add a transparent line for each person
  stat_summary( # add average line
    aes(group = 1),
    fun = mean,
    geom = "line",
    size = 1.5,
    color = "black"
  ) +
  labs(y = "y", x = "time")
```



La linea nera più spessa rappresenta l'intercetta media e la pendenza media della curva di crescita del modello LGM. Ogni individuo ha una sua specifica intercetta e uno specifico tasso di cambiamento e questa diversità è catturata nelle componenti di varianza del modello

È anche possibile utilizzare le funzionalità di **lavaan** per verificare specifiche ipotesi di interesse relative ai dati longitudinali. Ad esempio, una possibile domanda riguarda l'uguaglianza degli errori nel tempo. Tale domanda può essere affrontata introducendo dei vincoli nella specificazione del modello LGM e, successivamente, confrontando la bontà dell'adattamento del modello vincolato e del modello generale.

Il modello vincolato è

```
model_eqerr <- "
  i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
  s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4
  t1 =~ a*t1
  t2 =~ a*t2
  t3 =~ a*t3
  t4 =~ a*t4
"
```

Adattiamo il modello vincolato:

```
fit_eqerr <- growth(model_eqerr, data = Demo.growth)
```

Confrontiamo il modello vincolato con il modello libero:

```
anova(fit, fit_eqerr)
#> Chi-Squared Difference Test
#>
#>           Df  AIC  BIC Chisq Chisq diff Df diff
#> fit           5 5528 5564  8.07
#> fit_eqerr     8 5524 5548  9.68      1.61      3
#>           Pr(>Chisq)
#> fit
#> fit_eqerr      0.66
```

Il test del rapporto di verosimiglianza (*likelihood ratio*) eseguito dalla funzione `anova()` produce un *p*-valore di 0.6573. Ciò significa che, nei dati esaminati, non si rileva una decremento della bontà dell'adattamento degna di nota nel passare dal modello libero al modello vincolato. Dunque, l'ipotesi dell'equaglianza della varianza degli errori nel tempo sembra ragionevole.

2.1.2 Un secondo esempio

Un modello leggermente più complesso aggiunge due regressori (*x1* e *x2*) che influenzano i fattori di crescita latenti. Inoltre, è stata aggiunta al modello una covariata *c* variabile nel tempo che influenza la misura del risultato nei quattro punti temporali.

```
model2 <- "
# intercept and slope with fixed coefficients
i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4
s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4
# regressions
i ~ x1 + x2
s ~ x1 + x2
# time-varying covariates
t1 ~ c1
t2 ~ c2
```



```

t3 ~ c3
t4 ~ c4
"

```

```
fit2 <- growth(model2, data = Demo.growth)
```

```
kable(coef(fit2), booktabs = TRUE, format = "markdown")
```

	x
i~x1	0.60839
i~x2	0.60411
s~x1	0.26224
s~x2	0.52173
t1~c1	0.14336
t2~c2	0.28900
t3~c3	0.32754
t4~c4	0.33049
t1~~t1	0.57982
t2~~t2	0.59559
t3~~t3	0.48141
t4~~t4	0.53521
i~~i	1.07946
s~~s	0.22376
i~~s	0.07475
i~1	0.58024
s~1	0.95758

I risultati mostrano che le due covariate x influenzano sia l'intercetta sia la pendenza della curva di crescita. Inoltre, vi sono evidenze di un effetto della covariata c .



Parte II

Appendici



A

Simbologia di base

Per una scrittura più sintetica possono essere utilizzati alcuni simboli matematici.

- $\log(x)$: il logaritmo naturale di x .
- L'operatore logico booleano \wedge significa “e” (congiunzione forte) mentre il connettivo di disgiunzione \vee significa “o” (oppure) (congiunzione debole).
- Il quantificatore esistenziale \exists vuol dire “esiste almeno un” e indica l'esistenza di almeno una istanza del concetto/oggetto indicato. Il quantificatore esistenziale di unicità $\exists!$ (“esiste soltanto un”) indica l'esistenza di esattamente una istanza del concetto/oggetto indicato. Il quantificatore esistenziale \nexists nega l'esistenza del concetto/oggetto indicato.
- Il quantificatore universale \forall vuol dire “per ogni.”
- \mathcal{A}, \mathcal{S} : insiemi.
- $x \in A$: x è un elemento dell'insieme A .
- L'implicazione logica “ \Rightarrow ” significa “implica” (se ...allora). $P \Rightarrow Q$ vuol dire che P è condizione sufficiente per la verità di Q e che Q è condizione necessaria per la verità di P .
- L'equivalenza matematica “ \Leftrightarrow ” significa “se e solo se” e indica una condizione necessaria e sufficiente, o corrispondenza biunivoca.
- Il simbolo $|$ si legge “tale che.”
- Il simbolo \triangleq (o $:=$) si legge “uguale per definizione.”
- Il simbolo Δ indica la differenza fra due valori della variabile scritta a destra del simbolo.
- Il simbolo \propto si legge “proporzionale a.”
- Il simbolo \approx si legge “circa.”
- Il simbolo \in della teoria degli insiemi vuol dire “appartiene” e indica l'appartenenza di un elemento ad un insieme. Il simbolo \notin vuol dire “non appartiene.”
- Il simbolo \subseteq si legge “è un sottoinsieme di” (può coincidere con l'insieme stesso). Il simbolo \subset si legge “è un sottoinsieme proprio di.”

- Il simbolo $\#$ indica la cardinalità di un insieme.
- Il simbolo \cap indica l'intersezione di due insiemi. Il simbolo \cup indica l'unione di due insiemi.
- Il simbolo \emptyset indica l'insieme vuoto o evento impossibile.
- In matematica, argmax identifica l'insieme dei punti per i quali una data funzione raggiunge il suo massimo. In altre parole, $\operatorname{argmax}_x f(x)$ è l'insieme dei valori di x per i quali $f(x)$ raggiunge il valore più alto.
- a, c, α, γ : scalari.
- x, y : vettori.
- X, Y : matrici.
- $X \sim p$: la variabile casuale X si distribuisce come p .
- $p(\cdot)$: distribuzione di massa o di densità di probabilità.
- $p(y \mid x)$: la probabilità o densità di y dato x , ovvero $p(y = Y \mid x = X)$.
- $f(x)$: una funzione arbitraria di x .
- $f(X; \theta, \gamma)$: f è una funzione di X con parametri θ, γ . Questa notazione indica che X sono i dati che vengono passati ad un modello di parametri θ, γ .
- $\mathcal{N}(\mu, \sigma^2)$: distribuzione gaussiana di media μ e varianza σ^2 .
- $\text{Beta}(\alpha, \beta)$: distribuzione Beta di parametri α e β .
- $\mathcal{U}(a, b)$: distribuzione uniforme con limite inferiore a e limite superiore b .
- $\text{Cauchy}(\alpha, \beta)$: distribuzione di Cauchy di parametri α (posizione: media) e β (scala: radice quadrata della varianza).
- $\mathcal{B}(p)$: distribuzione di Bernoulli di parametro p (probabilità di successo).
- $\text{Bin}(n, p)$: distribuzione binomiale di parametri n (numero di prove) e p (probabilità di successo).
- $\mathbb{KL}(p \parallel q)$: la divergenza di Kullback-Leibler da p a q .

Bibliografia

- Allen, M. J. and Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.
- Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334.
- Gulliksen, H. (1961). Measurement of learning and mental abilities. Technical report, Princeton University.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2):183–202.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4):443–477.
- Kline, P. (2013). *Handbook of psychological testing*. Routledge.
- Kuder, G. F. and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160.
- Lewis, C. (1986). Test theory and psychometrika: The past twenty-five years. *Psychometrika*, 51(1):11–22.
- Lord, F. M. and Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Nunnally, J. C. (1994). *Psychometric theory*. McGraw-Hill.

- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Samejima, F. (1983). Constant information model on the dichotomous response level. In *New horizons in testing*, pages 287–308. Elsevier.
- Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15:201–293.
- Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of the vectors of mind.