

Item Response Theory for beginners

Assuzioni e stima dei parametri

Dr. Ottavia M. Epifania

Corso IRT @ Università Libera di Bolzano, 16-18 Gennaio 2023

Bressanone

- ① Assunzioni
- ② Fit del modello
- ③ Stima dei parametri

Assunzioni

Assunzioni

Unidimensionalità

L'assunzione di **unidimensionalità** indica che un solo tratto latente è responsabile delle risposte agli item

Tale assunzione viene spesso valutata mediante modelli di **analisi fattoriale confermativa**

I modelli IRT multidimensionali consentono di gestire la presenza di più tratti latenti ma sono modelli più complessi e meno diffusi

In caso di violazione

Il modello IRT scelto può essere applicato, ma:

- si possono stime dei parametri distorte/non interpretabili
- anche se si ottenessero delle stime interpretabili, queste non hanno senso perché il modello non ha senso

Solitamente, si utilizza l'analisi fattoriale confermativa (CFA).

Se la soluzione ad un fattore presenta una buona fit, si suppone l'unidimensionalità

Indici di fit CFA

Comparative Fit Index (CFI) $> .90$

Standardized Root Mean Square Residual (SRMSR) $< .08$

Root Mean Square Error of Approximation (RMSEA) $< .08$

Assunzioni

Indipendenza locale

L'assunzione di **indipendenza locale** indica che non esiste alcuna relazione tra le risposte di un soggetto ad item diversi dopo aver controllato per il tratto latente

Le risposte di un soggetto a un insieme di item sono indipendenti quando la probabilità associata alle risposte fornite dal soggetto agli item è uguale al prodotto delle probabilità relative alle singole risposte

In caso di violazione

Si rischia di sovrastimare l'informatività (attendibilità) del test

Si correlano i **residui** (i.e., differenza tra la risposta data da un soggetto ad un item e il valore atteso per quella risposta)

La correlazione tra i residui si interpreta attraverso la statistica $Q3$ (Yen, 1984)

In genere, $Q3 \geq .20$ per una coppia di item è indicativo di dipendenza locale (sono disponibili anche altri cut-off)

Eliminare uno dei due item della coppia (solitamente quello con minore discriminatività o che ha una fit peggiore)

Se si utilizza una procedura adattica (e.g., Computerized Adaptive Testing) → vincolare la somministrazione di uno solo dei due item

Se possibile → combinare i due item in unico item



Assunzioni

Monotonicità

Monotonicità: La probabilità di rispondere correttamente aumenta all'aumentare del livello di tratto latente

Come si valuta

Si valuta per ogni item

Coefficiente H di Mokken (Originariamente di Loevinger).

$H \geq .3$: Item accettabile

$H \geq .5$: Item eccellente

$H < .3$: Item non accettabile

In caso di violazione

Effetti negativi sull'attendibilità e validità della scala

Fit del modello

Dopo aver verificato le assuzioni, si può procedere alla verifica della fit del modello mediante due statistiche principali:

- M^2 (Meyedeu-olivares & Joe, 2005): si basa sulla classificazione dei soggetti in base al loro pattern di risposta. Si basa sulla distribuzione χ^2 per cui tende ad essere significativa per campioni ampi (anche se il modello ha una buona fit)
- Root Mean Square Error of Approximation (RMSEA): Misura di quanto il modello si avvicina alla realtà:
 - $\leq .05$ Perfetto
 - $\leq .08$: Accettabile
 - $> .80$: No fit

Stima dei parametri

Stima dei parametri

Massima verosimiglianza

Massima verosimiglianza (Maximum Likelihood, ML) → trova i valori dei parametri che massimizzano la probabilità di ottenere i dati osservati (i.e., i valori che massimizzano la funzione di verosimiglianza dei dati osservati)

Due tipologie:

- ① **Massima verosimiglianza congiunta** (Joint Maximum Likelihood, JML): Permette di stimare per massima verosimiglianza **congiuntamente** i parametri degli item e delle persone
- ② **Massima verosimiglianza marginale** (Marginal Maximum Likelihood, MML): Permette di stimare per massima verosimiglianza i parametri degli item, i parametri delle persone sono stimati successivamente con procedure bayesiane

Vantaggi ML

All'aumentare dell'ampiezza campionaria \rightarrow le stime per ML convergono al valore vero (unbiased)

Svantaggi ML

Le stime di soggetti/item con punteggi estremi sono $+$ o $-\infty$

Per ovviare a questo problema:

\rightarrow Sottrarre .30 se il punteggio è massimo

\rightarrow Aggiungere .30 se il punteggio è minimo

Punteggi estremi:

Soggetti che hanno dato solo
risposte giuste o solo risposte errate

Item che hanno ricevuto solo
risposte giuste o solo risposte errate

Stima dei parametri

Approccio Bayesiano

Moltiplicando la funzione di verosimiglianza per una distribuzione a priori si ottiene la **distribuzione a posteriori**

Le stime dei parametri si ottengono dalla distribuzione a posteriori →
stime bayesiane

Due tipologie:

- **Maximum a Posteriori** (MAP): La stima del parametro è la moda della distribuzione a posteriori di quel parametro
- **Expected a Posteriori** (MAP): La stima del parametro è la media della distribuzione a posteriori di quel parametro

Vantaggi approcci bayesiani

Permettono la stima **finita** dei parametri dei soggetti e degli item anche in caso di punteggi estremi