Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

PARTE II-A: ELEMENTI DI TCT APPLICATA

Antonio Calcagnì
DPSS, Università di Padova

A.A. 2022/2023

Teoria Classica dei test

Diversi sono i modelli formali utilizzabili nello sviluppo, validazione ed analisi delle proprietà dei test psicologici. Il primo di questi ad essere sviluppato, ed ancora oggi utilizzato, è il c.d. **modello del punteggio vero** (o teoria classica dei test - TCT). Altri modelli ricadono all'interno della **teoria della risposta all'item** (*Item Response Theory* - IRT) e sono stati sviluppati per diverse tipologie di item (es.: dicotomici, politomici) e diverse tipologie di test (es.: test di personalità, cognitivi, di performance).

Ciascuno di tali modelli ha sviluppati propri metodi per l'analisi dei test e lo studio delle loro proprietà, tra cui l'*attendibilità*, la *validità*, lo *scoring*, il *bias*, l'*analisi degli item*.

In questo corso affronteremo alcuni aspetti importanti per l'analisi dei test utilizzando principalmente la **TCT** come modello formale di riferimento. Quest'ultima non verrà trattata in maniera esaustiva ma si farà riferimento solo ad alcuni argomenti di natura applicativa.

Per gli interessati, una sintesi introduttiva ai modelli IRT è disponibile in BN(3) mentre una comparazione tra TCT e IRT dal punto di vista applicativo è altresì disponibile in BN(5).

Teoria Classica dei Test

FONTI: BN(2.1,2.2)

Consideriamo un insieme di misure (o item) indicizzate tramite $j = 1, \ldots, p$ e un insieme di rispondenti (estratti casualmente da una popolazione di riferimento) indicizzati tramite $i = 1, \ldots, n$ su cui le misurazioni sono effettuate mediante un test.

La TCT stabilisce che

$$\boxed{Y_j = T_j + E_j}$$

dove:

- Y_j è la v.a. che governa la realizzazione della j-esima misura relativamente alla componente **osservabile**
- T_j è la v.a. che governa la realizzazione della j-esima misura relativamente alla componente **vera** (non osservabile)
- E_j è la v.a. che governa la realizzazione della j-esima misura relativamente alla componente di **errore**

Teoria Classica dei Test

FONTI: BN(2.1,2.2)

Consideriamo un insieme di misure (o item) indicizzate tramite $j = 1, \ldots, p$ e un insieme di rispondenti (estratti casualmente da una popolazione di riferimento) indicizzati tramite $i = 1, \ldots, n$ su cui le misurazioni sono effettuate mediante un test.

La TCT stabilisce che

$$\boxed{Y_j = T_j + E_j}$$

dove:

(i) $\mathbb{E}[E_j] = 0$ l'errore ha media nulla (errore non sistematico)

(ii) $\mathbb{C}\mathrm{or}[E_j, T_j] = 0$ l'errore e il valore vero della misura sono non correlati

(iii) $\mathbb{C}\mathrm{or}[E_j, E_{j'}] = 0$ due misurazioni differenti j e j' hanno errori non correlati

Attendibilità

Una misura importante che lega T a Y è la correlazione tra tali quantità, ossia ρ_{YT}, che esprime la forza con cui si esprimono le realizzazioni delle due v.a. Valori alti di ρ_{YT} indicano una notevole precisione nella quantificazione di T mediante Y. Nella realtà della rilevazione dei dati, solo Y è conoscibile mentre T (analogamente, E) sono stimabili solo utilizzando una coppia di *misure parallele*.

Teoria Classica dei Test

FONTI: BN(2.3)

Attendibilità

La definizione teorica di attendibilità è basata sull'utilizzo di *misure parallele*, aspetto raramente incontrato nella realtà della somministrazione di test.

Un modo per approssimare misure parallele è quello di agire sul disegno della somministrazione del test:

- somministrazione dello stesso test sullo stesso campione in occasioni differenti: in questo caso, ρ_{YT} misura la *stabilità temporale* di Y
- somministrazione di due forme differenti dello stesso test sullo stesso campione: in questo caso, ρ_{YT} misura l'*equivalenza del test alternativo*

Nella pratica della psicometria applicata, la quantità ρ_{YT} è spesso stimata utilizzando diversi procedimenti statistici.

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità

Dopo aver descritto il modello teorico per la formulazione dell'attendibilità, di seguito si riportano alcuni metodi per la stima dell'attendibilità in contesti reali di somministrazione di test.

Due approcci in generale:

1. Procedure che richiedono due somministrazioni del test
   - metodo delle forme parallele
   - metodo test-retest
2. Procedure che richiedono una sola somministrazione del test
   - metodo dello split-half
   - metodo basato sulla covarianza degli item (c.d. *coerenza interna*)

Teoria Classica dei Test

Major Error Source	Reliability Coefficient	Data-Collection Procedure	Statistical Treatment of Data
1. Change in examinee's overtime	1. Stability coefficient	1. Test, wait, retest	1. Compute Pearson product moment coefficient, $\hat{\rho}_{12}$
2. Content sampling from form to form	2. Equivalence coefficient	2. Give form 1, give form 2	2. Compute Pearson product moment coefficient, $\hat{\rho}_{12}$
3. Content sampling, or flawed items	3. Internal consistency coefficient	3. Give one form on one occasion	3a. Divide test into halves; correlate half-test, $\hat{\rho}_{AB}$; use Spearman Brown correction: $\hat{\rho}_{XX'} = \frac{2\hat{\rho}_{AB}}{1+\hat{\rho}_{AB}}$ b. Divide test into halves; use Guttman's or Rulon's formula: $\hat{\rho}_{XX'} = 1 - \frac{\hat{\sigma}^2_D}{\hat{\sigma}^2_X}$ c. Compute item variances; compute coefficient alpha: $\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\Sigma\hat{\sigma}^2_i}{\hat{\sigma}^2_X}\right)$

Approcci per la stima dell'attendibilità: Sintesi dei metodi per il calcolo empirico dell'attendibilità.
Fonte: Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt (Capitolo 7).

Teoria Classica dei Test

Ricordiamo la **struttura dei dati** a disposizione in questi casi: una matrice **Y** *soggetti (n) x item (p)* le cui righe rappresentano le unità statistiche (soggetti a cui è somministrato il test o questionario) mentre le colonne rappresentano gli item del test/questionario.

	item 1	...	item j	...	item p
1	y_{11}	...	y_{1j}	...	y_{1p}
$\vdots$	...	...	...	...	...
i	y_{i1}	...	y_{ij}	...	y_{ip}
$\vdots$	...	...	...	...	...
n	y_{n1}	...	y_{nj}	...	y_{np}

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Split-Half

Dopo aver somministrato il test Y al campione, la procedura si basa sulla suddivisione del test in due parti di medesima lunghezza (ad esempio, divisione tra item pari e dispari, assegnazione casuale degli item alle due forme) - Y_A e Y_B - che vengono considerate forme parallele del medesimo test.

Fatta l'assunzione che le misure composite Y_A e Y_B siano parallele, l'attendibilità è stimata utilizzando la formula di Sperman-Brown:

$$\hat{\rho}^2_Y = \frac{2\hat{\rho}^2_{AB}}{1 + \hat{\rho}^2_{AB}}$$

dove $\hat{\rho}^2_{AB}$ è la correlazione tra le due parti di Y calcolata tra i punteggi totali $\mathbf{y}_{\text{tot}_A}$ e $\mathbf{y}_{\text{tot}_B}$.

Procedure per la stima dell'attendibilità: Split-Half

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}_Y^2 = \frac{2\hat{\rho}_{AB}^2}{1+\hat{\rho}_{AB}^2} = \frac{2 \cdot 0.34}{1+0.34} = 0.507$$

Il risultato ci indica che l'attendibilità del test Y, calcolata usando la divisione a metà del test, non è particolarmente alta. Un modo, ad esempio, per aumentarne il valore è quello di aggiungere ulteriori item (paralleli e coerenti in contenuto) affinché le forme composite Y_A e Y_B risultino più lunghe.

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Split-Half

Un altro modo per calcolare l'attendibilità del test totale Y nel caso in cui Y_A e Y_B non siano parallele, è quello di utilizzare l'indice di Rulon come stima dell'attendibilità del test complessivo:

$$\hat{\rho}_Y^2 = 2\left(1 - \frac{\sigma^2_{\mathbf{y}_A} + \sigma^2_{\mathbf{y}_B}}{\sigma^2_{\mathbf{y}_{\text{tot}}}}\right)$$

dove con $\sigma^2_{\mathbf{y}_{(.)}}$ si intende in generale la varianza del vettore di dati $\mathbf{y}_{(.)}$.

Procedure per la stima dell'attendibilità: Split-Half

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}^2_Y = 2\left(1 - \frac{\sigma^2_{\mathbf{y}_A} + \sigma^2_{\mathbf{y}_B}}{\sigma^2_{\mathbf{y}_{\text{tot}}}}\right) = 2\left(1 - \frac{1.82 + 1.56}{4.54}\right) = 0.508$$

Il risultato ottenuto con l'indice di Rulon è pressoché simile a quello ottenuto con la formula di Sperman-Brown. Il rapporto tra le varianze delle due metà del test è pari a $\frac{1.82}{1.56} = 1.163$.

Procedure per la stima dell'attendibilità: Coerenza interna

Consideriamo una collezione di m misure $(Y_1, T_1, E_1), \ldots, (Y_m, T_m, E_m)$ che formano la *misura composta* $X = \sum_{j=1}^{m} Y_j$. Uno stimatore dell'attendibilità della misura composta ρ^2_X è noto come α di Cronbach:

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

dove maggiore è la componente di covarianza tra le misure componenti, maggiore sarà l'indice α finale. Quando invece la covarianza tra le misure è prossima allo zero, l'indice assumerà anch'esso valore prossimo allo zero.

(1) Nota: la sommatoria è sulle misure componenti che sono diverse. Infatti quando $j = h$ abbiamo $\mathbb{C}\text{ov}\left[Y_j, Y_h\right] = \mathbb{V}\text{ar}\left[Y_j, Y_h\right]$ già calcolata nella sommatoria precedente. Tale formulazione permette di operare sulle covarianze delle coppie di misure realmente differenti tra loro.

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Coerenza interna

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

La componente di covarianza tra le misure componenti $\mathbb{C}\text{ov}\,[Y_j, Y_h]$ fornisce un'indicazione sulla *coerenza* delle misure tra loro. Per tale ragione l'indice α si dice quantifichi la **coerenza interna** di una scala (o misura composita).

Nota: tale formulazione fornisce un'interpretazione dell'attendibilità in termini di *coerenza interna* della misura composita. A differenza di altre tecniche per valutare l'attendibilità di un test (ad esempio, *split-half*), l'α di Cronbach utilizza solo l'informazione derivante dalla covarianza dalle misure componenti (o item).

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Coerenza interna

Lo stimatore α:

- interpreta l'attendibilità di una misura composta in termini di coerenza interna rispetto alle misure componenti che la formano
- presuppone che le misure componenti *formino bene* la misura composta, nel senso dato dalla intercorrelazione tra le misure componenti
- è interpretato usando i seguenti valori di riferimento:

valore soglia	attendibilità
$\alpha > 0.9$	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

Teoria Classica dei Test
FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Coerenza interna

Un altro indice per la stima dell'attendibilità della scala Y quando le misure componenti (item) sono di tipo dicotomico è noto come indice di Kuder-Richardson (KR20):

$$\hat{\rho}^2_{XX'} = \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \pi_j(1-\pi_j)}{\sigma^2_{\mathbf{y}_{\text{tot}}}}\right)$$

L'indice si ottiene dalla sostituzione della generica $\mathbb{V}\text{ar}\,[Y]$ nella formula dell'α di Chronbach con la varianza della distribuzione binomiale.

L'interpretazione segue quella adottata per l'indice α di Chronbach.

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Coerenza interna

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3
$\boldsymbol{\pi}$	0.4	0.3	0.6	0.7	0.6	0.3			

$$\hat{\rho}^2_{XX'} = \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m} \pi_j(1-\pi_j)}{\sigma^2_{\mathbf{y}_{\text{tot}}}}\right) = \frac{6}{6-1}\left(1 - \frac{1.35}{4.54}\right) = 0.703$$

Il risultato differisce da quello ottenuto usando il metodo dello Split-Half: ciò dipende dalla particolare scelta della divisione a metà del test fatta con quest'ultimo metodo. Difatti, facendo variare le possibili suddivisioni a metà (es.: usando la suddivisione casuale) si ottengono diversi valori per $\hat{\rho}^2_{XX'}$.

Procedure per la stima dell'attendibilità: Correlazione intraclasse

Utilizzando i risultati della TCT, è possibile fornisce un calcolo dell'attendibilità come rapporto di varianze:

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

Tale formula è denominata coefficiente di correlazione intraclasse e rapporta la quantità di varianza della componente vera alla varianza complessiva.

Teoria Classica dei Test

FONTI: BN(2.5)

Procedure per la stima dell'attendibilità: Correlazione intraclasse

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}^2_{YT} = \frac{\hat{\sigma}^2_T}{\hat{\sigma}^2_T + \hat{\sigma}^2_E} = \hat{\rho}^2_{YT} = \frac{1.787^2}{1.787^2 + 1.161^2} = 0.703$$

Il risultato è analogo a quello ottenuto precedente ed evidenzia il calcolo della precisione del test Y come rapporto tra la varianza del misurando T e la varianza complessiva del test dovuta sia alla componente dell'errore E che a quella della componente latente T.

Teoria Classica dei Test

FONTI: BN(2.6)

Attendibilità e lunghezza del test

In generale, data una misura composita iniziale X (o test) avente attendibilità $\rho^2_{XX'}$ possiamo valutare come questa vari in funzione dell'aggiunta di m nuovi item.

Questo può essere utile, ad esempio, quando si vuole valutare quanti item occorra aggiungere per avere una desiderata attendibilità.

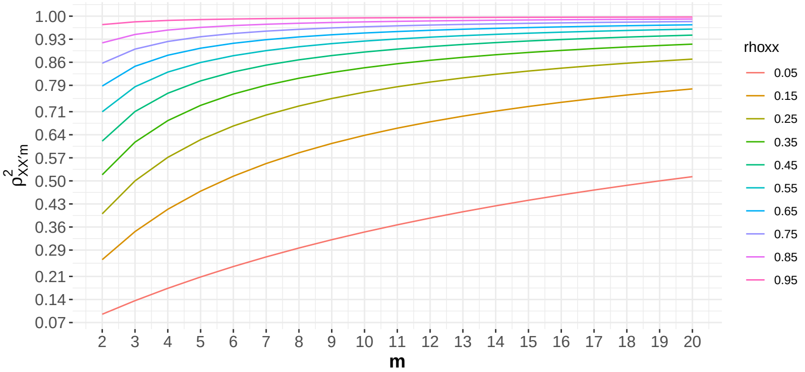
La risposta a tale quesito è nota come *formula profetica di Sperman-Brown*:

$$\rho^2_{XX'_m} = \frac{m\rho^2_{XX'}}{1 + (m-1)\rho^2_{XX'}}$$

dove $\rho^2_{XX'}$ indica l'attendibilità del test iniziale (o test di origine).

Teoria Classica dei Test

FONTI: BN(2.6)



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

Teoria Classica dei Test

FONTI: BN(2.10)

Fattori che influenzano l'attendibilità

- qualità del campione di individui su cui le misurazioni sono effettuate (rappresentatività)
- qualità delle misure componenti (o item)
- campionamento realmente casuale ed indipendenza delle osservazioni
- condizione di somministrazione del test
- lunghezza del test adeguata
- aspetti cognitivi degli individui testati: fatica, ricordo, bassa compliance

si approfondisca BN(2.10)

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9)

Stima dei punteggi veri ad un test

Dato il test (o misura composita) Y, abbiamo precedentemente calcolato $\mathbf{y}_{tot}$ come somma dei punteggi su p item. Il vettore $\mathbf{y}_{tot}$ contiene i punteggi osservati del campione di n individui testati. Per ottenere una stima dei punteggi veri $\boldsymbol{\tau}_{n\times 1}$ è possibile utilizzare lo stimatore lineare:

$$\tau_i = \hat{\rho}^2_{YT} y_{tot_i} + (1 - \hat{\rho}^2_{YT})\mu_x$$

dove $\mu_x = \frac{1}{n}\sum_{i=1}^{n} y_{tot_i}$ è la media complessiva delle misurazioni effettuate (stima della media dei punteggi veri nella popolazione). Lo stimatore utilizza una stima dell'attendibilità/precisione del test: quando questa è bassa, esso tende ad attribuire maggior peso alla componente μ_x rispetto alla misura osservata y_{tot_i}.

Nota: lo stimatore utilizzato per $\boldsymbol{\tau}$ potrebbe non essere adeguato quando la relazione tra Y e T è non lineare.

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9)

Stima dei punteggi veri ad un test

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$	$\boldsymbol{\tau}$
1	0	0	0	0	0	0	0	0	0	0.861
2	0	0	0	0	1	0	1	0	1	1.564
3	1	0	1	1	1	0	3	1	4	3.673
4	1	1	1	1	1	1	3	3	6	5.079
5	1	1	1	1	1	1	3	3	6	5.079
6	0	0	1	0	0	0	1	0	1	1.564
7	0	0	1	1	1	0	2	1	3	2.970
8	0	0	0	1	0	0	0	1	1	1.564
9	1	0	1	1	1	0	3	1	4	3.673
10	0	1	0	1	0	1	0	3	3	2.973

Caso con alta attendibilità: Per la stima di $\boldsymbol{\tau}$ si è utilizzata la stima di ρ^2_{YT} ottenuta tramite il KR20. Si noti come i punteggi veri siano stimati rispetto al tratto/costrutto latente continuo sotteso agli item dicotomici. Inoltre, il calcolo di $\boldsymbol{\tau}$ non richiede nessuna informazione aggiuntiva sul singolo individuo (es.: variazioni individuali dei punteggi) se non il suo punteggio osservato.

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9)

Stima dei punteggi veri ad un test

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$	$\boldsymbol{\tau}$
1	0	0	0	0	0	0	0	0	0	2.871
2	0	0	0	0	1	0	1	0	1	2.881
3	1	0	1	1	1	0	3	1	4	2.911
4	1	1	1	1	1	1	3	3	6	2.931
5	1	1	1	1	1	1	3	3	6	2.931
6	0	0	1	0	0	0	1	0	1	2.881
7	0	0	1	1	1	0	2	1	3	2.901
8	0	0	0	1	0	0	0	1	1	2.881
9	1	0	1	1	1	0	3	1	4	2.911
10	0	1	0	1	0	1	0	3	3	2.901

Caso con bassa attendibilità: In questo caso (estremo) si è ipotizzato un valore di precisione prossimo allo zero $\rho^2_{YT} = 0.01$. Si noti come i punteggi veri tendano ad assumere lo stesso valore della media della popolazione ignorando i punteggi osservati. Ciò è in linea con l'impostazione della TCT che formalizza il test (e le sue proprietà) come oggetto primario rispetto al comportamento di risposta dei partecipanti.

Teoria Classica dei Test

FONTI: BN(2.8.0,2.9)

Stima dei punteggi veri ad un test

Una volta calcolato $\hat{\tau}_i$ è possibile quantificare l'incertezza della stima mediante un intervallo di confidenza simmetrico $\mathcal{C}_i = (\tau_i^{\text{lb}}, \tau_i^{\text{ub}}) \in \mathbb{R}^2$ che utilizza il modello probabilistico della normale standardizzata come riferimento per $\boldsymbol{\tau}$:

$$\mathbb{P}(\tau_i \in \mathcal{C}_i) = 1 - \alpha$$
$$\mathbb{P}(\hat{\tau}_i - z_{1-\alpha}\sigma_E \leq \tau_i \leq \hat{\tau}_i + z_{1-\alpha}\sigma_E) = 1 - \alpha$$

dove $z_{1-\alpha}$ è il valore critico della distribuzione normale standardizzata corrispondente alla probabilità α scelta (es: $\alpha = 0.05$, $z_{1-\alpha} = 1.96$).

La componente $\sigma_E = \sqrt{\mathbb{V}\text{ar}[E]}$ e può essere calcolata utilizzando il risultato (vii) della TCT e la stima della precisione del test $\hat{\rho}^2_{YT}$ (calcolata opportunamente con uno dei procedimenti visti in precedenza):

$$\hat{\sigma}_E = \hat{\sigma}_{\mathbf{y}_{\text{tot}}}\sqrt{(1 - \hat{\rho}^2_{YT})}$$

Stima dei punteggi veri ad un test

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$\mathbf{y}_{\text{tot}_A}$	$\mathbf{y}_{\text{tot}_B}$	$\mathbf{y}_{\text{tot}}$	$\boldsymbol{\tau}^{\text{lb}}$	$\boldsymbol{\tau}$	$\boldsymbol{\tau}^{\text{ub}}$
1	0	0	0	0	0	0	0	0	0	-1.416	0.861	3.138
2	0	0	0	0	1	0	1	0	1	-0.713	1.564	3.841
3	1	0	1	1	1	0	3	1	4	1.396	3.673	5.950
4	1	1	1	1	1	1	3	3	6	2.802	5.079	7.356
5	1	1	1	1	1	1	3	3	6	2.802	5.079	7.356
6	0	0	1	0	0	0	1	0	1	-0.713	1.564	3.841
7	0	0	1	1	1	0	2	1	3	0.693	2.970	5.247
8	0	0	0	1	0	0	0	1	1	-0.713	1.564	3.841
9	1	0	1	1	1	0	3	1	4	1.396	3.673	5.950
10	0	1	0	1	0	1	0	3	3	0.693	2.970	5.247

Intervalli di confidenza per il punteggio vero: Per la stima di $\boldsymbol{\tau}$ si è utilizzata la stima di ρ^2_{YT} ottenuta tramite il KR20. Si noti come l'ampiezza degli intervalli di confidenza sia dipendente dall'attendibilità del test: maggiore è il valore di precisione ρ^2_{YT}, minore sarà l'ampiezza di $\mathcal{C}_i$.

Teoria Classica dei Test

FONTI: BN(6.1)

Validità

Quando il nostro interesse è rivolto a capire *cosa* un determinato test misuri siamo nell'ambito della validità di un test (o misura). Diversi sono i metodi per valutare la validità di un test e diverse sono anche le accezioni di riferimento:

- validità di contenuto (riferita al contenuto o semantica degli item/scale)
- validità fattoriale (riferita alla coerenza interna degli item, valutata mediante *analisi della dimensionalità*)
- validità di criterio (riferita alla coerenza di un test rispetto ad un criterio/test esterno)
- validità di costrutto (riferita alla convergenza o divergenza rispetto ad altre misure per un medesimo costrutto da quantificare)

Teoria Classica dei Test

FONTI: BN(6.1)

Validità

studio individuale: BN(6.1.1), BN(6.1.2), BN(6.1.4)

Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2)

Punteggi e norme

Quando una scala è somministrata ad un campione di n individui, le misure componenti $\mathbf{y}_{i1}, \ldots, \mathbf{y}_{ip}$ rilevate sul campione sono utilizzate per formare il *punteggio grezzo* $\mathbf{y}_{\text{tot}} = \sum_{j=1}^{p} y_{ij}$.

Il processo di calcolo del punteggio grezzo individuale usando gli item di un test è detto **scoring**.

L'interpretazione di $\mathbf{y}_{\text{tot}}$ viene effettuata per confronto con i **valori normativi** del test (solitamente riportati nel manuale associato al test/questionario). I valori normativi sono ottenuti durante la *fase di taratura del test* ed esprimono le caratteristiche di Y a livello di popolazione, queste ultime quantificate utilizzando i momenti di Y (solitamente media e varianza).

Taratura di un test: BN(2.12.3)

Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2)

Punteggi e norme

Per valutare/confrontare i punteggi ottenuti dagli individui ad una scala o tra più scale è opportuno trasformare i punteggi grezzi in *punteggi standardizzati*. Questi possono essere impiegati altresì per la costruzione dei *profili individuali*.

Diversi sono i modi per calcolare i punteggi standardizzati, ad esempio mediante:

(a) punteggio z

(b) punteggio T

Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2)

Punteggi e norme

Punteggio z

$$z_i = \frac{y_i - \overline{y}}{s_y}$$

dove:

- $\overline{y}$ è la media del campione
- s_y è lo scarto quadratico medio (standard deviation) del campione

I punteggi z hanno media nulla e varianza unitaria poiché utilizzano la Normale standardizzata come legge di riferimento.

I punteggi z si 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 $\pm 1, 2, \ldots, K$ deviazioni standard.

Punteggi e norme

Punteggio t

Si ottiene per trasformazione lineare dei punteggi z:

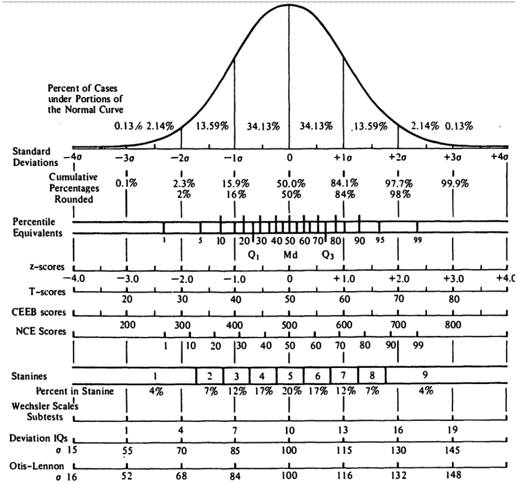
$$t_i = 50 + 10z_i$$

I punteggi t hanno media pari a 50 e deviazione standard pari a 10 e vengono utilizzati preferibilmente per rendere i punteggi finali tutti positivi (i punteggi z infatti possono assumere valori negativi).

Teoria Classica dei Test

FONTI: BN(2.12.1-2.12.2)

Punteggi e norme



Tipologie di punteggi standardizzati: Punteggi standardizzati rispetto al modello normale standard (con media e dev. standard). Fonte: Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt (Capitolo 19).

Teoria Classica dei Test

Note di sintesi

- La TCT decompone la variabilità della misurazione Y in una componente di errore E (accidentale) ed una componente attribuita al misurando T (latente)
- Obiettivo della TCT è di costruire un test di misura che abbia alta precisione/attendibilità, ossia quando $\mathbb{V}\text{ar}[T] > \mathbb{V}\text{ar}[E]$
- Le componenti di variabilità individuale (differenze individuali) non vengono modellate se non attraverso il campione/gruppo
- La TCT si focalizza maggiormente sulle caratteristiche del test più che su quelle degli item