AFFIDABILITÀ

L'AFFIDABILITÀ

costituisce parte integrante delle verifiche empiriche sulla validità di uno strumento di misurazione: se esso non è affidabile, non ha senso verificarne altri aspetti

una misura è affidabile se produce risultati stabili o costanti

- da una misurazione all'altra della stessa variabile,
- nelle stesse persone,
- a parità di condizioni
- nonostante variazioni accidentali

il concetto di affidabilità è legato alla teoria classica della misurazione

AFFIDABILITÀ: TEORIA CLASSICA DELLA MISURAZIONE

Assunzioni alla base della teoria:

le prove cui un soggetto si sottopone sono indipendenti le une dalle altre, si tratti dei singoli item di una scala o di misurazioni ripetute dello stesso test

ogni osservazione è la somma di una componente vera e di una componente d'errore casuale

$$X_i = V_i + E_i$$

PUNTEGGIO VERO

È un punteggio teorico,

il valore medio di infinite misurazioni,

che non osservo mai direttamente

NON è un indice di una prestazione reale o valida e dunque non è un indice della validità di costrutto

punteggio vero ≠ prestazione reale

ma i due saranno tanto più correlati quanto maggiore è la validità di costrutto

AFFIDABILITÀ: TEORIA CLASSICA DELLA MISURAZIONE

Assunzioni alla base della teoria:

- da una prova all'altra la componente vera rimane costante: $V_1 = V_2$
- mentre l'errore varia casualmente: $r_{E1E2} = 0$
- e non c'è relazione tra le due componenti: $r_{VE1} = r_{VE2} = 0$

AFFIDABILITÀ:
TEORIA CLASSICA DELLA MISURAZIONE

Assunzioni alla base della teoria:

- su infinite misurazioni: $\sum E_i = 0$
- e il punteggio vero corrisponde alla media delle osservazioni:

$$\overline{X} = \overline{V} + \overline{E} = \overline{V} = 0$$

- scomponendo la varianza del punteggio osservato

$$s_X^2 = \frac{\sum (v_i + e_i)^2}{N} = \frac{\sum v_i^{\,2}}{N} + \frac{2\sum v_i e_i}{N} + \frac{\sum e_i^{\,2}}{N} = s_v^2 + s_e^2$$

- pertanto l'affidabilità corrisponde teoricamente a

$$r_{tt} = \frac{s_v^2}{s_X^2} = 1 - \frac{s_e^2}{s_X^2}$$

ERRORI CASUALI?

Si commettono

durante (*within occasion*) una singola prova di test, da un item all'altro, da uno stimolo all'altro (variazioni nei livelli di attenzione, di efficienza mentale, di umore della persona, nelle condizioni contestuali di somministrazione; errori non sistematici stimolo x rater)

più stimoli per rilevare la stessa variabile

ERRORI CASUALI?

. . . si commettono

da una prova all'altra

- *between occasions (test-retest): errori casuali legati a condizioni transitorie*
- *between raters (interrater): errori casuali legati a condizioni transitorie + errori non sistematici legati a caratteristiche proprie del rater e rater x target (leniency error, halo effect)*

⬇

più occasioni (tempo/rater) di misurazione della stessa variabile

COME STIMARE IL COEFFICIENTE DI AFFIDABILITÀ A PARTIRE DAI DATI OSSERVATI?

Within occasion

- Forme parallele / equivalenti
- Split-half
- Coerenza interna

Between occasions /raters

- Test-retest
  - coefficiente di affidabilità
  - coefficiente di stabilità
- Correlazione interrater

METODI DI STIMA DELL'AFFIDABILITÀ

Test-retest: correlazione tra punteggi aggregati osservati per lo stesso campione di persone in due occasioni di misurazione separate da un arco temporale

- Affidabilità: l'instabilità è dovuta solo ad errore casuale, il punteggio vero rimane invariato da un'occasione all'altra
- Stabilità: l'insabilità del punteggio dipende anche da cambiamento nel punteggio vero, rimane stabile quota di punteggio vero

Inter-rater o tra valutatori: correlazione tra più valutatori che valutano lo stesso target

METODI DI STIMA DELL'AFFIDABILITÀ

Split-half e la formula profetica di Spearman-Brown

$$r_{tt} = \frac{2\, r_{M1M2}}{1 + r_{M1M2}}$$

$$r_{tt} = N\, \bar{r} \,/\, \left[1 + (N - 1)\, \bar{r}\right]$$

$$n = r_{tt'}\, (1 - r_{tt}) /\, r_{tt}\, (1 - r_{tt'})$$

per calcolare di *quante volte* allungare il test e ottenere il coefficiente di affidabilità desiderato

Coerenza interna (Alpha di Cronbach) – valutata sulla coerenza delle risposte a più stimoli, unica sessione di valutazione

$$r_{tt} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum s_i^2}{s_X^2}\right)$$

Le forme di verifica dell'affidabilità NON sono intercambiabili

$$r_{tt} = N\,\bar{r} \,/\, \left[1 + \left(N - 1\right)\bar{r}\right]$$

$$10 * .30 / \left[1 + \left(9 * .30\right)\right] = .81$$

$$10 * .10 / \left[1 + \left(9 * .10\right)\right] = .52$$

$$30 * .10 / \left[1 + \left(29 * .10\right)\right] = .77$$

$$n = r_{tt'}\left(1 - r_{tt}\right) / \, r_{tt}\left(1 - r_{tt'}\right)$$

$$.80\left(1 - .52\right) / \, .52\left(1 - .80\right) = .384 / .104 = 3{,}7$$

$$N = 10$$

$$N = 10 * 3{,}7$$

$$37 * .10 / 1 + (36 * .10) = .80$$

LUNGHEZZA DEL TEST E AFFIDABILITÀ

In generale,se allungo il test, la varianza vera aumenta in progressione geometrica

$$s_{nV}^2 = n^2 s_V^2$$

mentre la varianza d'errore aumenta in progressione aritmetica

$$s_{nE}^2 = n s_E^2$$

AFFIDABILITÀ

Scala	Test-retest (M-F) USA	Alpha (M-F)
L (15)	77 - 81	75 - 65
F (60)	78 - 69	82 - 79
K (30)	84 - 81	85 - 83
Hs (32)	85 - 85	83 - 82
D (57)	75 -77	62 - 70
Hy (60)	72 - 76	62 - 62
Pd (50)	81 - 79	65 - 70
Mf (56)	82 - 73	57 - 34
Pa (40)	67 - 58	58 - 59
Pt (48)	89 - 88	91 - 90
Sc (78)	87 - 80	91 - 90
Ma (46)	83 - 68	57 - 63
Si (62)	92 - 91	84 - 84

DALL'AFFIDABILITÀ ALL' ERRORE STANDARD DELLA MISURA

Dispersione dei punteggi osservati intorno ai punteggi veri:

$$s_E = s_X \sqrt{1 - r_{tt}}$$

Grazie all'errore standard della misura e in base alle proprietà della distribuzione normale,
è possibile definire un intervallo di fiducia per il punteggio vero a partire dal punteggio osservato come segue:

$$V' = X \pm s_E \, z_{1-\alpha}$$

DALL'AFFIDABILITÀ ALL' ERRORE STANDARD DELLA MISURA

Dato il livello di probabilità che possiamo prefissare $\alpha =$ 05,

tale per cui intendiamo definire un intervallo di confidenza del 95% entro il quale ricade

probabilisticamente il punteggio vero,

supposto che il livello di affidabilità del test sia r_{tt}=80

la sua deviazione standard sia pari a 8

pertanto l'errore standard di misura

$$s_E = s_X \sqrt{1 - r_{tt}}$$

$$s_E = 8\sqrt{1-.80} = 3.56$$

per il punteggio osservato QI = 110,

si stima che il suo punteggio vero sia compreso tra

110 ± (356 * 1,96) con una probabilità pari a 95