

Affidabilità e Validità di una Misura

Master Avanzato in Economia e Politica Agraria

Giovanbattista Califano

Università degli Studi di Napoli Federico II, giovanbattista.califano@unina.it

Scale Attitudinali per lo Studio
delle Preferenze del Consumatore
19 Maggio 2025

Tabella di marcia

15/05	Hello, Psychometrics!	✓
19/05	Affidabilità e Validità di una Misura	○
22/05	Variabili Latenti: Riflessive o Formative?	○
29/05	Modelli ad Equazioni Strutturali	○

Riferimenti consigliati:

- ▶ Capitoli 4 e 7 da Jhangiani et al. (2019)
- ▶ Capitoli 7, 10 e 14 da Olivero and Russo (2022)
- ▶ Capitolo 12 da Mehmetoglu and Jakobsen (2022)

Jhangiani, R. S., Chiang, I. A., Cuttler, C., and Leighton, D. C. (2019). *Research Methods in Psychology*. Kwantlen Polytechnic University, 4th edition.

Mehmetoglu, M. and Jakobsen, T. G. (2022). *Applied Statistics Using Stata: A Guide for the Social Sciences*. SAGE Publications Ltd, 2nd edition.

Olivero, N. and Russo, V. (2022). *Psicologia dei Consumi*. McGraw-Hill Education, 3rd edition.

Nelle puntate precedenti...

Cosa intendiamo per misurazione?

Assegnazione *sistematica* di un punteggio a delle entità (individui, oggetti, eventi...), in modo tale che il punteggio rappresenti una caratteristica di quelle entità.

Nelle puntate precedenti...

Cosa intendiamo per misurazione?

Assegnazione *sistematica* di un punteggio a delle entità (individui, oggetti, eventi...), in modo tale che il punteggio rappresenti una caratteristica di quelle entità.

- ▶ Ma come facciamo a sapere che quel punteggio rappresenta **davvero** la caratteristica che vogliamo studiare?
- ▶ I ricercatori non lo danno per scontato...

Il test di Califano per l'intelligenza

Il test di Califano è stato messo a punto per misurare il quoziente intellettivo. Secondo il test, il QI di una persona è dato dal numero di lettere nel suo nome (n) moltiplicato per 20:

$$QI = n \cdot 20$$

Il test di Califano per l'intelligenza

Il test di Califano è stato messo a punto per misurare il quoziente intellettivo. Secondo il test, il QI di una persona è dato dal numero di lettere nel suo nome (n) moltiplicato per 20:

$$QI = n \cdot 20$$

- Questo test è più **affidabile** della maggior parte dei test già esistenti per misurare il QI :)

Il test di Califano per l'intelligenza

Il test di Califano è stato messo a punto per misurare il quoziente intellettivo. Secondo il test, il QI di una persona è dato dal numero di lettere nel suo nome (n) moltiplicato per 20:

$$QI = n \cdot 20$$

- ▶ Questo test è più **affidabile** della maggior parte dei test già esistenti per misurare il QI :)
- ▶ Potrebbe essere poco **valido** :(

Passiamo a...

Affidabilità (o *reliability*)

Validità (o *validity*)

Stata

L'affidabilità

si riferisce al grado di **coerenza** (o *consistency*) di una misura. La psicometria si concentra, in particolare, su tre tipologie di affidabilità:

1. Nel tempo (*test-retest*)
2. Tra ricercatori (*inter-rater*)
3. Tra item (*internal*)

L'affidabilità nel tempo

Test-Retest

Se il costrutto misurato è ipotizzato essere coerente nel tempo, anche i punteggi dovrebbero esserlo. L'affidabilità nel tempo consiste nello studio della correlazione fra due distribuzioni di misura ottenute somministrando due volte lo stesso test allo stesso gruppo di soggetti dopo un certo intervallo di tempo.

Esempio

Se una persona è molto intelligente oggi, dovrebbe esserlo anche tra una settimana...

L'affidabilità tra ricercatori

Inter-Rater

Se diversi ricercatori utilizzano lo stesso sistema per assegnare dei punteggi, questi punteggi dovrebbero essere positivamente correlati tra loro.

L'affidabilità tra item

Internal

L'affidabilità interna è generalmente la più rilevante quando si utilizzano strumenti self-report, e si riferisce alla coerenza delle risposte degli intervistati ad una misura multi-item.

L'affidabilità tra item

Food Technology Neophobia Scale (Cox & Evans, 2008)

La FTNS è uno strumento popolare per misurare la neofobia verso le tecnologie alimentari e si basa sulla scala di Likert.

FTNS 1: Non ho ragione di provare cibi altamente tecnologici perché quelli che mangio sono già abbastanza buoni

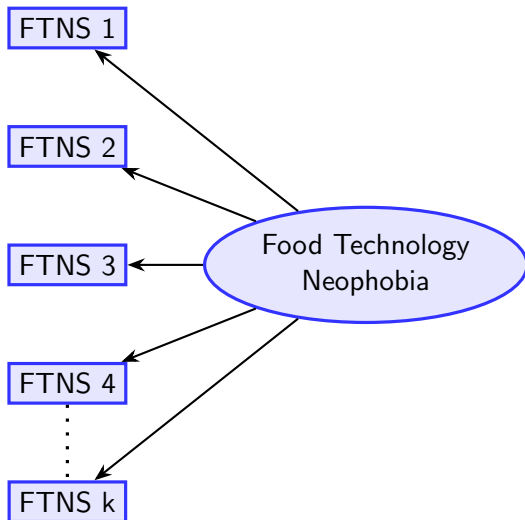
FTNS 2: Le nuove tecnologie alimentari sono qualcosa di cui sono incerto

FTNS 3: I nuovi prodotti alimentari non sono più salutari dei cibi tradizionali

(...)

<i>Fortemente in disaccordo</i>	<i>In disaccordo</i>	<i>Né d'accordo né in disaccordo</i>	<i>D'accordo</i>	<i>Fortemente d'accordo</i>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

L'affidabilità tra item



L'affidabilità tra item

La coerenza interna di una scala può essere misurata in diversi modi, ancora una volta basati sull'indice di correlazione.

L'approccio più intuitivo è la correlazione *split-half*:

1. Gli item della scala si dividono in due blocchi (es. pari e dispari);
2. Si aggregano i punteggi per i due blocchi, tramite somma o media;
3. Si calcola l'indice di correlazione tra i punteggi aggregati dei due blocchi.

Un coefficiente di correlazione split-half $r > .80$ è generalmente considerato un indicatore di buona coerenza interna.

L'affidabilità tra item

Un approccio più sofisticato e utilizzato è l'*alpha di Cronbach*, definito come:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{Y_i}^2}{\sigma_X^2} \right)$$

dove:

- ▶ k è il numero di item che compongono la scala,
- ▶ $\sigma_{Y_i}^2$ è la varianza del singolo item i ,
- ▶ σ_X^2 è la varianza del punteggio totale (dato dalla somma di tutti gli item).

Nella maggior parte dei casi, con $\alpha > .70$ siamo soddisfatti.

Passiamo a...

Affidabilità (o *reliability*)

Validità (o *validity*)

Stata

Validità di una misura

Abbiamo visto come il **Test di Califano** per il QI potrebbe godere di un buon grado di *affidabilità* (es. Test-Retest). Cosa possiamo dire riguardo la *validità*?

Validità di una misura

La validità ci dice quanto i punteggi di una misurazione rappresentano la caratteristica che dovrebbero rappresentare.

Anche la validità viene suddivisa in diverse tipologie:

- ▶ Validità di facciata
- ▶ Validità di contenuto
- ▶ Validità di criterio
- ▶ Validità discriminante

Validità di facciata e di contenuto

Sono le due tipologie di validità che non si stabiliscono con metodi quantitativi. La **validità di facciata** si riferisce al buon senso: ci vuole poco a dire che il numero di lettere contenute nel nome non ha niente a che fare con l'intelligenza.

La **validità di contenuto**, invece, si riferisce a quanto la misura *copre* il costrutto di interesse (es. componente di percezione del rischio e percezione di inutilità delle nuove tecnologie alimentari nella FTNS).

Validità di criterio

Un **criterio** può essere qualsiasi variabile che pensiamo debba essere correlata al costrutto che stiamo cercando di misurare:

- ▶ Validità concorrente: quando il criterio è misurato allo stesso tempo del costrutto;
- ▶ Validità predittiva: quando il criterio è misurato dopo il costrutto;
- ▶ Validità convergente: quando il criterio è una misura alternativa dello stesso costrutto.

Validità discriminante

Si tratta del grado in cui i punteggi di una misura *non* sono correlati con le misure di variabili che sono concettualmente distinte.

Validità di criterio e discriminante

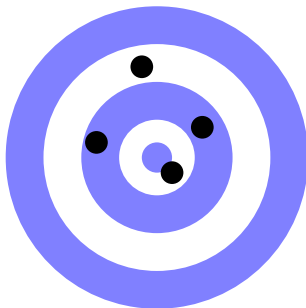
Facciamo un esempio

Immaginiamo di voler testare la validità di criterio e discriminante di una scala che misura l'autoefficacia nell'affrontare il master.

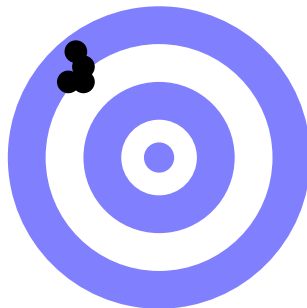
- ▶ **Validità concorrente:** l'autoefficacia nell'affrontare il master dovrebbe essere positivamente correlata con gli atteggiamenti verso Stata;
- ▶ **Validità predittiva:** l'autoefficacia nell'affrontare il master dovrebbe essere in grado di predire la media dei voti ottenuti durante il master;
- ▶ **Validità convergente:** se disponessimo di un'altra scala che misura la stessa autoefficacia, le due scale dovrebbero mostrare una correlazione positiva tra loro;
- ▶ **Validità discriminante:** l'autoefficacia si riferisce alla fiducia nelle proprie capacità di affrontare compiti o situazioni specifiche, mentre l'autostima è una valutazione generale di sé. Le due misure, quindi, non dovrebbero risultare troppo correlate.

Riassumendo

Affidabilità \equiv Precisione
Validità \equiv Accuratezza



Accurato, Non preciso



Non accurato, Preciso

- ▶ Misurare un costrutto psicologico richiede quattro passaggi fondamentali:
 1. **Definizione concettuale**: chiarire cosa si intende per il costrutto;
 2. **Definizione operativa**: tradurre il costrutto in qualcosa di osservabile e misurabile;
 3. **Implementazione della misura**: scegliere o costruire lo strumento adeguato;
 4. **Valutazione della misura**: testare lo strumento per verificarne validità e affidabilità.

1. Definizione concettuale

- ▶ Serve una definizione chiara del costrutto.
- ▶ Fondamentale consultare la letteratura scientifica.

2. Definizione operativa

- ▶ Precisa *come* il costrutto sarà misurato.
- ▶ Possibili più definizioni operative per lo stesso costrutto.

3. Implementare la misura

- ▶ Usare uno strumento esistente: più rapido, già validato, confrontabile con altri studi.
- ▶ Creare una nuova misura: solo se *necessario*; attenzione alla semplicità, chiarezza e numero di item.

4. Valutare la misura

- ▶ Test preliminare con alcuni partecipanti.
- ▶ Verificare: chiarezza delle istruzioni, tempi, comprensibilità.
- ▶ Raccogliere feedback per correggere eventuali problemi prima della raccolta dati.

Passiamo a...

Affidabilità (o *reliability*)

Validità (o *validity*)

Stata

clear all ;)

```
. webuse set data.califano.xyz  
. webuse masteristi
```