Attendibilità

Misurazione

- I problemi di misurazione degli oggetti di studio sono comuni a *tutte* le discipline scientifiche.
- In psicologia il problema è solo più evidente, non più grave.
- I costrutti che intendiamo misurare sono infatti non direttamente osservabili.

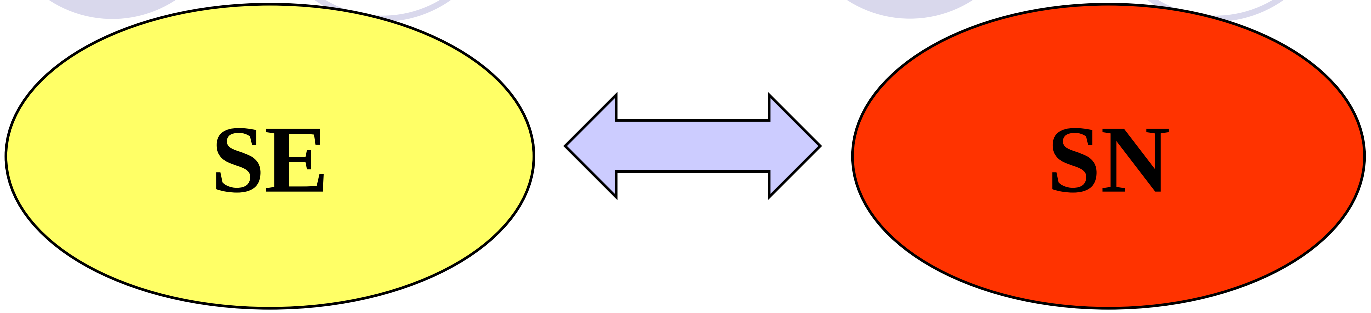
Il peso della teoria

- Spesso la definizione teorica di un fenomeno o costrutto varia da teoria a teoria.
- È stato così anche per molti concetti propri delle scienze mature.
- Spesso in psicologia, il disaccordo fra teorie rende problematica la misura del fenomeno.
- È la teoria, descrivendo le caratteristiche del fenomeno, a determinare le sue misure.

MISURAZIONE

Assegnazione di valori numerici ad oggetti o eventi secondo regole che ci consentono di rappresentare proprietà degli oggetti ed eventi tramite le proprietà del sistema numerico ⇨ utilizzo di uno strumento

MISURAZIONE



- **Sistema empirico (SE):** ciò che si vuole misurare; costituito da elementi legati tra loro da relazioni

- **Sistema numerico (SN):** ciò che serve per misurare; esprime mediante i numeri le relazioni esistenti nel sistema empirico

MISURAZIONE

- Occorre tenere presente che ad ogni misura è associato sempre un errore
- Il punteggio osservato si deve a:
  - Il livello del soggetto nella caratteristica (**V**)
  - Circostanze casuali, o errore casuale di misura (**E**)

$$\boxed{X = V + E}$$

X = misura rilevata

V = parte vera

E = errore: $\begin{cases} \text{fluttuazioni casuali (sempre)} \\ \text{costante o sistematico (talvolta).} \end{cases}$

MISURAZIONE - 1

- L’errore casuale di misurazione **E** è responsabile dell’imprecisione del punteggio **X**.
- Quando **E=0**, **X=V**, e la misura è perfettamente attendibile (precisa)
- È fondamentale sapere quanto è precisa una misurazione o un test. Dobbiamo conoscere quindi **V** e **E**
- **X** lo osserviamo, invece **V** ed **E** non li osserviamo direttamente: vanno stimati

MISURAZIONE – 2

- Facciamo attenzione a comprendere meglio il significato della porzione vera del punteggio.

$$V = V_A + V_B + \dots S$$

$$X = V_A + V_B + \dots S + E$$

V_A = parte vera del costrutto **A** (<u>quello che vogliamo misurare</u>)

V_B = parte vera del costrutto **B** (<u>che non volevamo misurare</u>)

S = altre fonti sistematiche, anche di errore

Attendibilità

- Affinché una misura sia valida, si richiede che essa misuri **sistematicamente** qualcosa.
- L'attendibilità rappresenta la precisione di una misura (ciò che nella misura non è errore).
- L'attendibilità è definibile come il rapporto fra la componente sistematica e la variabilità totale della misura.

$$\boxed{\rho = \frac{V}{V + E}}$$

Primo accenno alla validità

- Una misura è valida se coglie il concetto che essa tende rilevare.
- La validità è il limite superiore dell'attendibilità.

$$\frac{V_A}{V_A + V_B + S \ldots + E}$$

- Accertare la validità è più difficile che accertare l'attendibilità.

Assunzioni sull'attendibilità

- **L'errore di misurazione è una variabile aleatoria, distribuita normalmente.** Ciò significa che ci si aspetta tanti piccoli errori, vicini allo zero, e pochi errori di una certa entità.
- **La media degli errori di misurazione è uguale a 0.** Gli errori casuali tendono ad annullarsi all'aumentare del numero di misurazioni. Più misure facciamo, più precisa sarà la misurazione.
- **I punteggi veri e gli errori di misurazione sono tra loro indipendenti.** La varianza di errore è uguale ad ogni livello del punteggio vero: Facciamo lo stesso tipo di errore per punteggi veri bassi, e per punteggi veri alti.

Assunzioni sull’attendibilità

- **Gli errori di misurazione compiuti in due somministrazioni indipendenti sono fra loro indipendenti.**
- Se queste assunzioni sono vere, è possibile utilizzare diversi metodi per misurare la varianza vera e la varianza d’errore, e da queste calcolare l’attendibilità come rapporto:

$$\rho = \frac{V}{V + E}$$

Vari genere di attendibilità

1) **<u>Metodo del Test-Retest:</u>**

Si somministra il **test** al tempo **T1** ed al tempo **T2** e si calcola la correlazione tra i punteggi. Questo metodo non necessità di ulteriori specificazioni. Basta saper calcolare la **r di Pearson** tra due serie di punteggi.

2) **<u>Metodo delle Forme Parallele:</u>**

Si somministrano due versioni equivalenti del test (stessa media e stessa dev. St.) Quindi si calcola la correlazione tra i le due forme come stima dell'attendibilità test-retest.

3) **<u>Metodo dello Split-Half:</u>**

Si somministra il test in un unico tempo **T1**. Si divide il test a metà e si considerano le due metà come forme parallele (stessa media e stessa dev. St.) Quindi si calcola la correlazione tra le due metà come stima dell'attendibilità test-retest.

4) **<u>Metodo della Coerenza Interna:</u>**

Si somministra il test in un unico tempo **T1**. Ogni item viene considerato un test a se stante. Si stima (con apposite formule) la correlazione media tra tutti gli item. Questa è una stima dell'attendibilità

Test-retest

	PG T1	PG T2
ss1	11	12
ss2	15	14
ss3	17	14
ss4	20	19
ss5	20	21
ss6	25	27
ss7	22	18
ss8	21	24
ss9	34	31
ss10	38	36
ss11	40	37
Media	23,91	23,00
Dev St.	9,03	8,39

Svolgimento:

Cov (PG T1, PG T2) = 73.45

Dev.St.(PG T1) = 9.03

Dev.St.(PG T2) = 8.39

Rtt **=** 73.45/ 9.03 × 8.39 =

73.45 / 75.74 =**.96**

$$r_{tt} = \frac{Cov\ (PG\ T1,\ PG\ T2)}{Dev.St.(PG\ T1) \times Dev.St.(PG\ T2)}$$

Interpretazione attendibilità test-retest

- Buoni coefficienti test-retest dovrebbero superare .80 (livello piuttosto esigente).
- Il coefficiente test-retest cala all'aumentare del tempo trascorso fra le rilevazioni.
- Il coefficiente test-retest è interpretabile se si assume che il concetto misurato non si modifichi nel tempo

Forme parallele

- Il problema principale è quello di verificare che le due forme siano effettivamente parallele. Ciò significa verificare che le due forme abbiano la **stessa media** e la **stessa varianza**. Prendiamo i seguenti dati:

	T1 Forma A	T2 Forma B
ss1	11	12
ss2	15	14
ss3	17	14
ss4	20	19
ss5	20	21
ss6	25	27
ss7	22	18
ss8	21	24
ss9	34	31
ss10	38	36
ss11	40	37
Media	23,91	23,00
Dev St.	9,03	8,39

Svolgimento:

t-test sulle due medie

t_{sp}=1.3; *Gdl*=10; t_{cr}=2,23

Le medie sono uguali.

Fare Test F sulle due varianze

F_{sp}=1.15, *gdl* = 10,10
F_{cr}=2.97

Le varianze sono uguali.

Quindi r_{tt} = .96

Split-half

Prendiamo i seguenti dati di un test di 10 item somministrato a 11 soggetti:

Item / Sogg.	1	2	3	4	5	6	7	8	9	10
1	1	0	1	1	0	1	0	1	1	1
2	0	0	1	0	1	1	0	1	0	0
3	1	1	0	1	1	1	1	0	1	1
4	0	0	1	0	0	1	0	1	0	0
5	1	0	1	1	0	1	0	0	1	1
6	1	1	0	0	1	0	1	1	0	0
7	0	0	1	0	0	0	1	0	0	0
8	1	1	1	1	0	1	0	0	0	0
9	0	1	0	0	1	1	1	0	0	1
10	1	0	0	1	0	0	1	0	0	1
11	1	1	1	0	1	0	0	1	1	0

Sommiamo gli item pari e dispari

Sogg.	Item Dispari	Item Pari
1	3	4
2	2	2
3	4	4
4	1	2
5	3	3
6	3	2
7	2	0
8	2	3
9	2	3
10	2	2
11	4	2
Media	2,55	2,45
Dev.St.	0,89	1,08

Svolgimento:

Verificare che le due metà abbiano la stessa media e la stessa varianza. t-test=ns, F-test=ns: ***OK***

Calcolare la correlazione tra le due metà **=.40**

Interpretazione coefficiente split-half

- L'attendibilità dipende molto dalla lunghezza del test ➔ la correlazione split-half è una sottostima dell'attendibilità
  - Infatti la divisione del test a metà ne dimezza la lunghezza
- Ci sono delle formule che permettono di correggere tale sottostima.

Formula di Spearman-Brown

$$Rn_{xx} = \frac{nr_{tt}}{1+(n-1)r_{tt}}$$

Si chiama *formula profetica*

Rn_{xx} = attendibilità del test allungato.

n = il numero delle volte che il test viene allungato o abbreviato. Si ottiene come rapporto tra numero degli item finale su numero degli item iniziali.

r_{tt} = attendibilità iniziale.

Applicazione al problema precedente: Rn_{xx} = ,40; n = 10/5 = 2

$$Rn_{xx} = \frac{2 \times .40}{1+(2-1)\times .40} = \frac{.80}{1.40} = .58$$

Split-half

La formula di RULON

$$R_{xx} = 1 - \frac{\sigma^2_{differenza}}{\sigma^2_{totale}} \Rightarrow R_{xx} = 1 - \frac{1.17}{2.73} = 1 - .43 = .57$$

Sogg.	Item Dispari	Item Pari	Differenza	Totale Pari+Dispari
1	3	4	-1	7
2	2	2	0	4
3	4	4	0	8
4	1	2	-1	3
5	3	3	0	6
6	3	2	1	5
7	2	0	2	2
8	2	3	-1	5
9	2	3	-1	5
10	2	2	0	4
11	4	2	2	6
Media	*2,55*	*2,45*	*0,09*	*5,00*
Dev.St.	*0,89*	*1,08*	*1,08*	*1,65*
Varianza	*0,79*	*1,16*	*1,17*	*2,73*

Split-half

La formula di GUTTMAN

$$R_{xx} = 2\left(1 - \frac{\sigma^2{}_A + \sigma^2{}_B}{\sigma^2{}_{totale}}\right) \Rightarrow R_{xx} = 2\left(1 - \frac{0.79 + 1.16}{2.73}\right) = 2(1 - .71) = .58$$

Sogg.	Item Dispari (A)	Item Pari (B)	Differenza	Totale Pari+Dispari
1	3	4	-1	7
2	2	2	0	4
3	4	4	0	8
4	1	2	-1	3
5	3	3	0	6
6	3	2	1	5
7	2	0	2	2
8	2	3	-1	5
9	2	3	-1	5
10	2	2	0	4
11	4	2	2	6
Media	*2,55*	*2,45*	*0,09*	*5,00*
Dev.St.	*0,89*	*1,08*	*1,08*	*1,65*
Varianza	*0,79*	*1,16*	*1,17*	*2,73*

Coerenza interna

Si utilizza per item dicotomici o politomici

Il coefficiente α di Cronbach

$$\alpha = \frac{K}{K-1} \times \left(1 - \frac{\sum \sigma^2{}_i}{\sigma^2{}_{totale}}\right)$$

Si utilizza **solo** per **item dicotomici**

Il coefficiente KR-20

$$KR_{20} = \frac{K}{K-1} \times \left(1 - \frac{\sum p_i q_i}{\sigma^2{}_{totale}}\right)$$

Item	1	2	3	4	5	6	7	8	9	10	Totale
Sogg.											
1	*1*	*0*	*1*	*1*	*0*	*1*	*0*	*1*	*1*	*1*	7
2	*0*	*0*	*1*	*0*	*1*	*1*	*0*	*1*	*0*	*0*	4
3	*1*	*1*	*0*	*1*	*1*	*1*	*1*	*0*	*1*	*1*	8
4	*0*	*0*	*1*	*0*	*0*	*1*	*0*	*1*	*0*	*0*	3
5	*1*	*0*	*1*	*1*	*0*	*1*	*0*	*0*	*1*	*1*	6
6	*1*	*1*	*0*	*0*	*1*	*0*	*1*	*1*	*0*	*0*	5
7	*0*	*0*	*1*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	2
8	*1*	*1*	*1*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	5
9	*0*	*1*	*0*	*0*	*1*	*1*	*1*	*0*	*0*	*1*	5
10	*1*	*0*	*0*	*1*	*0*	*0*	*1*	*0*	*0*	*1*	4
11	*1*	*1*	*1*	*0*	*1*	*0*	*0*	*1*	*1*	*0*	6
Media	*0,64*	*0,45*	*0,64*	*0,45*	*0,45*	*0,64*	*0,45*	*0,45*	*0,36*	*0,45*	*5,00*
Dev.St.	*0,48*	*0,50*	*0,48*	*0,50*	*0,50*	*0,48*	*0,50*	*0,50*	*0,48*	*0,50*	*1,65*
Varianza	*0,23*	*0,25*	*0,23*	*0,25*	*0,25*	*0,23*	*0,25*	*0,25*	*0,23*	*0,25*	*2,73*

Item	1	2	3	4	5	6	7	8	9	10	Totale
Sogg.											
1	*1*	*0*	*1*	*1*	*0*	*1*	*0*	*1*	*1*	*1*	7
2	*0*	*0*	*1*	*0*	*1*	*1*	*0*	*1*	*0*	*0*	4
3	*1*	*1*	*0*	*1*	*1*	*1*	*1*	*0*	*1*	*1*	8
4	*0*	*0*	*1*	*0*	*0*	*1*	*0*	*1*	*0*	*0*	3
5	*1*	*0*	*1*	*1*	*0*	*1*	*0*	*0*	*1*	*1*	6
6	*1*	*1*	*0*	*0*	*1*	*0*	*1*	*1*	*0*	*0*	5
7	*0*	*0*	*1*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	2
8	*1*	*1*	*1*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	5
9	*0*	*1*	*0*	*0*	*1*	*1*	*1*	*0*	*0*	*1*	5
10	*1*	*0*	*0*	*1*	*0*	*0*	*1*	*0*	*0*	*1*	4
11	*1*	*1*	*1*	*0*	*1*	*0*	*0*	*1*	*1*	*0*	6
p(1)	*0,64*	*0,45*	*0,64*	*0,45*	*0,45*	*0,64*	*0,45*	*0,45*	*0,36*	*0,45*	*5,00*
q(0)	*0,36*	*0,55*	*0,36*	*0,55*	*0,55*	*0,36*	*0,55*	*0,55*	*0,64*	*0,55*	*1,65*
pXq	*0,23*	*0,25*	*0,23*	*0,25*	*0,25*	*0,23*	*0,25*	*0,25*	*0,23*	*0,25*	*2,73*

Utilizzando i dati precedenti...

Il coefficiente α di Cronbach

$$\alpha = \frac{K}{K-1} \times \left(1 - \frac{\sum \sigma^2{}_i}{\sigma^2{}_{totale}} \right) =$$

$$= \frac{10}{9} \times \left(1 - \frac{.23 \times 4 + .25 \times 6}{2.73} \right) =$$

$$= \frac{10}{9} \times \left(1 - \frac{2.42}{2.73} \right) = .1261$$

Il coefficiente KR-20

$$KR_{20} = \frac{K}{K-1} \times \left(1 - \frac{\sum p_i q_i}{\sigma^2{}_{totale}} \right) =$$

$$= \frac{10}{9} \times \left(1 - \frac{.23 \times 4 + .25 \times 6}{2.73} \right) =$$

$$= \frac{10}{9} \times \left(1 - \frac{2.42}{2.73} \right) = .1261$$

Per item politomici...

Item	1	2	3	4	5	6	7	8	9	10	Totale
Sogg.											
1	*4*	*6*	*2*	*5*	*7*	*6*	*4*	*2*	*4*	*1*	41
2	*1*	*7*	*6*	*4*	*7*	*4*	*6*	*1*	*1*	*1*	38
3	*2*	*7*	*6*	*5*	*8*	*5*	*5*	*3*	*2*	*4*	47
4	*4*	*8*	*8*	*8*	*9*	*6*	*6*	*7*	*9*	*4*	69
5	*1*	*9*	*8*	*6*	*8*	*5*	*5*	*1*	*1*	*1*	45
6	*3*	*3*	*2*	*3*	*6*	*4*	*5*	*2*	*2*	*1*	31
7	*4*	*7*	*5*	*7*	*9*	*9*	*8*	*6*	*6*	*6*	67
8	*1*	*8*	*6*	*6*	*9*	*6*	*7*	*1*	*1*	*1*	46
9	*6*	*8*	*8*	*8*	*10*	*7*	*10*	*8*	*8*	*6*	79
10	*1*	*5*	*2*	*3*	*6*	*3*	*3*	*3*	*1*	*1*	28
11	*1*	*8*	*5*	*6*	*9*	*4*	*8*	*1*	*2*	*1*	45
Media	*2,55*	*6,91*	*5,27*	*5,55*	*8,00*	*5,36*	*6,09*	*3,18*	*3,36*	*2,45*	*48,73*
Dev.St.	*1,75*	*1,70*	*2,37*	*1,75*	*1,34*	*1,69*	*2,02*	*2,60*	*2,98*	*2,11*	*16,18*
Varianza	*3,07*	*2,89*	*5,62*	*3,07*	*1,80*	*2,85*	*4,09*	*6,76*	*8,85*	*4,47*	*261,82*

$$\alpha = \frac{K}{K-1} \times \left(1 - \frac{\sum \sigma^2{}_i}{\sigma^2{}_{totale}} \right) =$$

$$= \frac{11}{10} \times \left(1 - \frac{43{,}49}{261{,}82} \right) = \frac{11}{10} \times (1 - {,}17) = .91$$

Interpretazione α

- Valore minimo accettabile: .70 (se abbiamo pochissimi item, può andar bene .60)
- È corretto calcolare α quando gli item della scala misurano tutti una caratteristica omogenea
- Allo stesso tempo, se gli item sono fra loro troppo simili, α sarà molto elevata, ma la validità della misura può risultare bassa
  - In questi casi rischiamo di avere una misura iper-specifica