

Metodologia della ricerca in psicologia

Scienze e Tecniche Psicologiche

Tiziana Lanciano
tiziana.lanciano@uniba.it



" When you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind"

(Lord Kelvin, 1883)



La misurazione – Cosa è

La misurazione è un modo per selezionare una specifica caratteristica e decidere **come osservarla**

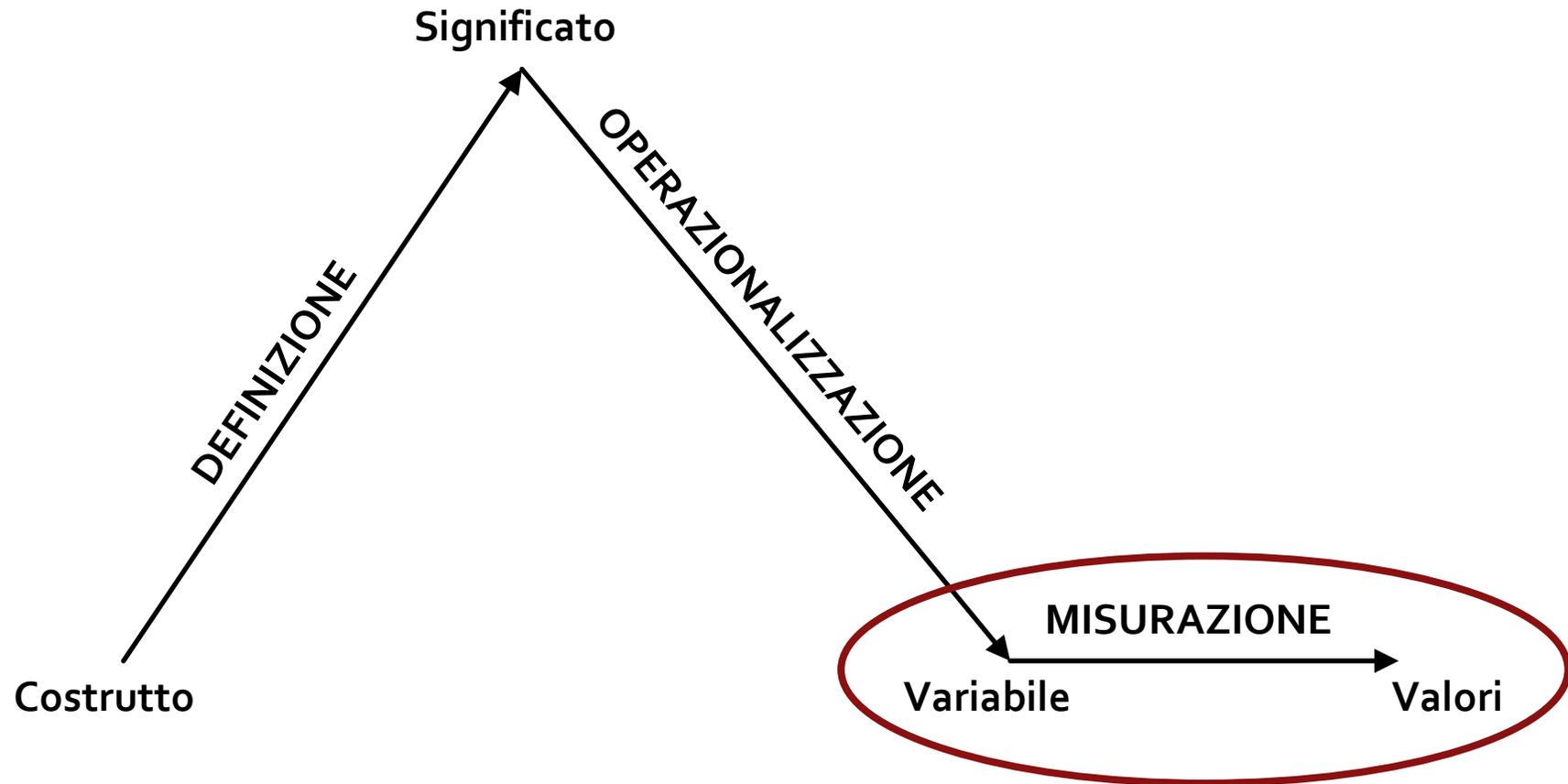
Bisogna comunque tener presente che l'assegnazione di numeri è un'operazione riduzionista della complessità psicologica

La misurazione – Cosa è

“La misurazione, nel senso più ampio, consiste nell’attribuzione di numeri a oggetti o eventi seguendo determinate regole. Il fatto che si possano assegnare dei numeri seguendo regole differenti porta a differenti tipi di scala e differenti tipi di misurazione”

(Stevens, 1946)

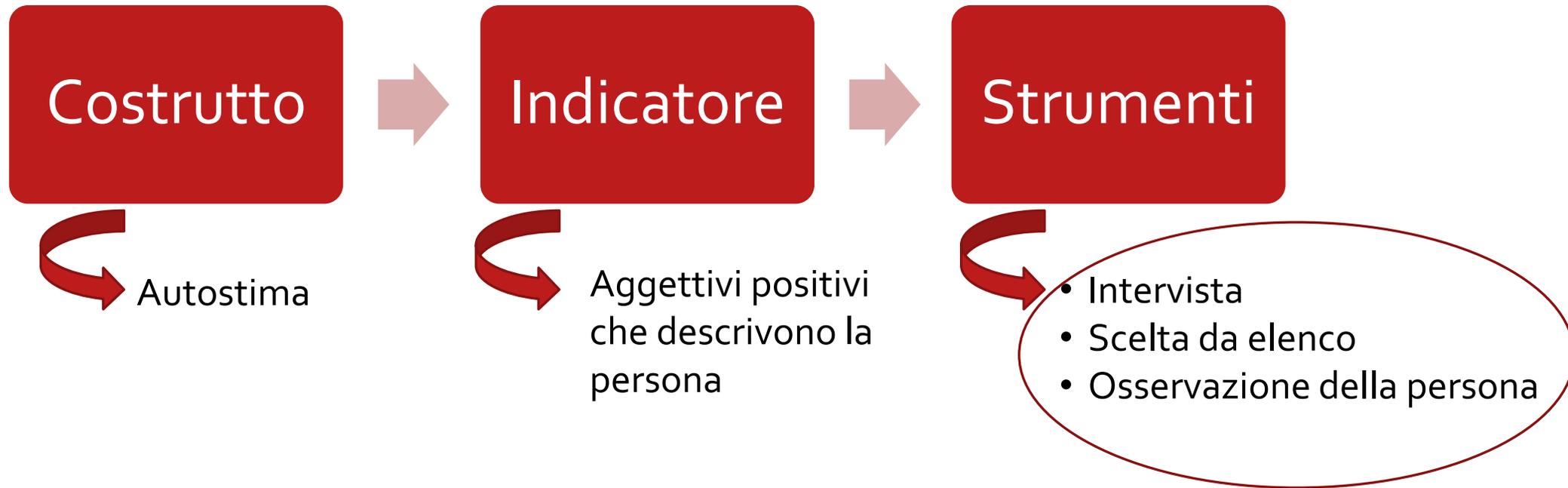
Processo che il ricercatore deve affrontare per rendere misurabile un costrutto psicologico



Come si misura?

La scelta degli strumenti è legata al tipo di indicatore scelto

Esempio. Misurare il livello di autostima:





Quale scala di misura?

Scale di misura

- Insieme delle diverse modalità adottate per descrivere le diverse modalità della variabile
- Esse sono modi di tradurre in numeri le caratteristiche che vengono osservate
- Il tipo di scala adottata per misurare una variabile definisce le elaborazioni statistiche più o meno appropriate

Scale di misura

- 1951 Stevens: classificazione delle possibili scale di misura
Quattro livelli di misurazione:

- **Nominale:** consiste nel classificare gli elementi sulla base della caratteristica misurata

Esempio: indichi la sua professione

1. Studente
2. Libero professionista
3. Impiegato
4. Disoccupato

Il valore 4 non significa che i disoccupati siano 4 volte *quel qualcosa* che sono gli studenti

Scale di misura

- 1951 Stevens: classificazione delle possibili scale di misura
Quattro livelli di misurazione:
 - **Ordinale:** consente non solo di classificare ma anche di attribuire un ordine

Esempio: Grado di istruzione

1. Licenza elementare
2. Licenza media
3. Diploma
4. Laurea

Le grandezze che intercorrono tra un livello e l'altro non sono omogenee

Scale di misura

- 1951 Stevens: classificazione delle possibili scale di misura
Quattro livelli di misurazione:
 - **Intervalli equivalenti:** consente di formare graduatorie avendo costante l'intervallo tra tutte le posizioni

Esempio: Temperatura Celsius

- ✓ 0°
- ✓ 10°
- ✓ 20°
- ✓ 30°

Esempio: Temperatura Farenheit

- ✓ 32°
- ✓ 50°
- ✓ 68°
- ✓ 86°

Non c'è uno zero assoluto ma relativo (scala tipicamente usata in psicologia)

Scale di misura

- 1951 Stevens: classificazione delle possibili scale di misura
Quattro livelli di misurazione:

- **Rapporti equivalenti:** consente di identificare una posizione anche rispetto all'assenza della caratteristica, cioè lo zero

Esempio: Grandezze fisiche

Caratteristiche delle scale di misura

Tipo di scala	Operazioni	Obiettivi
Nominale	Uguale/disuguale	Classificare gli stimoli in categorie definite (categorie discrete)
Ordinale	Maggiore di/minore di	Ordinare per rango gli stimoli su una singola dimensione
A intervalli equivalenti	Addizione/moltiplicazione/sottrazione/divisione	Specificare la distanza tra gli stimoli su una data dimensione
A rapporti equivalenti	Addizione/moltiplicazione Sottrazione/divisione Formazione dei rapporti di valore	Specificare la distanza tra gli stimoli su una data dimensione ed esprimere i rapporti dei valori della scala

Teorie sulla misurazione

Le teorie vengono utilizzate come guide per:

- La costruzione di test o strumenti self-report
- La loro interpretazione

Esse stabiliscono una corrispondenza empirica tra gli item di una scala e il costrutto psicologico che si ipotizza essi misurino

Teorie sulla misurazione

Due principali teorie sulla misurazione:

- Teoria classica dell'errore (Classical Test theory, C-T)
 - *Teoria classica dei test*
 - *Modello classico*
 - *Teoria dell'errore casuale*
- Teoria dei tratti latenti (Latent Trait theory, L-T)

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

- **Punteggio vero (V)**
 - valore reale ma teorico ("atteso") che ciascun oggetto possiede rispetto ad una caratteristica
 - quantità ipotetica non osservabile e non misurabile direttamente
 - ma può essere stimato attraverso la misurazione reale

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

- **Punteggio osservato (X)**
 - valore realmente osservato per l'oggetto rispetto alla caratteristica
 - ottenuto attraverso la procedura di misurazione
 - esso rappresenta una stima del punteggio vero

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

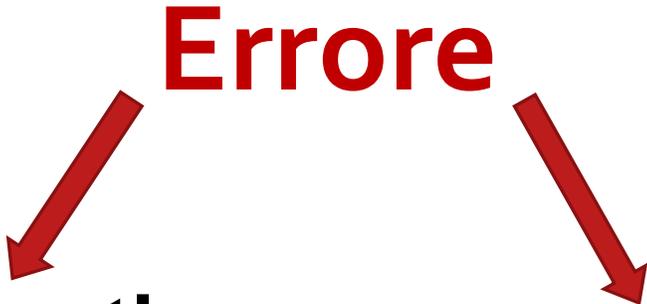
- **Errore (E)**
 - deviazione del punteggio osservato dal punteggio vero

punteggio osservato = punteggio vero + errore di misurazione

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

Errore



Errore sistematico

- ✓ quando influenza nella stessa direzione e con la stessa intensità tutti i punteggi raccolti ad un test;
- ✓ E' come aggiungere una costante a una variabile

Errore casuale

- ✓ Agisce in modo diverso per soggetti diversi, anche se sottoposti ad uno stesso test.

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

$$X = V + E \left\{ \begin{array}{l} X = \text{punteggio ottenuto dal soggetto al test} \\ V = \text{punteggio vero} \\ E = \text{fonte di errore} \end{array} \right.$$

✓ Es.

- ✓ Condizioni dei soggetti sottoposti al test
- ✓ Tentativi di indovinare
- ✓ Scarse istruzioni al test
- ✓ Forma non chiara per rispondere al test

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

Se si ripete la misurazione un numero infinito di volte:

- L'errore casuale tenderà a zero
- Il punteggio varierà attorno a un punteggio centrale che è considerato il punteggio vero

Teoria classica dell'errore

(Spearman, 1904; Novick, 1966)

- Il punteggio di un individuo ad un test ripetuto un numero infinito di volte non sarà sempre uguale ma varierà attorno al valore centrale che è considerato il punteggio vero
- La distribuzione di tali punteggi osservati assume la forma di una campana – *distribuzione normale o curva gaussiana* dove
 - la media corrisponde al punteggio vero
 - la deviazione standard (ds) è la variazione di errore
 - è chiamata errore standard di misura

La teoria dei tratti latenti

(Birnbaum, 1968)

Lo scopo è calcolare la probabilità che una persona risponda correttamente a un certo item, data da:

- Il livello di abilità/caratteristica del soggetto (tratto latente)
- Le caratteristiche della domanda o item, che sono di tre tipi:
 - Livello di difficoltà della domanda
 - Livello di discriminazione della domanda
 - Guessing, parametro che ci dice quanta probabilità ha il soggetto di rispondere correttamente se risponde a caso alla domanda



Come si valuta una misura?

VALIDITA'

della Ricerca

delle Misure

ATTENDIBILITA'

delle Misure



Come si valuta la misura

□ Validità delle misure

- Validità teorica o di costrutto
 - Convergente
 - Discriminante
- Validità di contenuto
- Validità empirica o di criterio
 - Predittiva
 - Concorrente
- Validità di facciata

□ Attendibilità delle misure



Validità della Misura

Si tratta di “un giudizio complessivo della misura in cui prove empiriche e principi teorici supportano **l’adeguatezza e l’appropriatezza delle conclusioni basate su punteggi al test**” (Messick, 1989: 13)

Validità teorica ($\rho_{\tau X}$)

Validità costrutto

Validità di contenuto

Validità empirica

Validità di facciata

- **Detta anche validità di costrutto**

- Si riferisce al livello in cui gli indicatori misurano accuratamente i costrutti teorici che interessa misurare

Validità convergente/discriminante

Validità costrutto

Validità di contenuto

Validità empirica

Validità di facciata

✓ La validità convergente di una misura è determinata dal grado di correlazione/dipendenza/convergenza con altre misure dello stesso costrutto o di costrutti affini (teoricamente molto legati)

- La possibilità di verificare la validità convergente dipende quindi dall'esistenza di costrutti, e relative misure, legati con quello misurato

Validità convergente/discriminante

Validità costrutto

Validità di contenuto

Validità empirica

Validità di facciata

✓ La validità discriminante è speculare alla validità convergente ed è data dal grado di indipendenza tra misure di costrutti diversi (teoricamente distinti)

Validità convergente/discriminante

Validità costrutto

Validità di contenuto

Validità empirica

Validità di facciata

- ✓ Multitrait-Multimethod Matrix (MTMM; Campbell e Fiske, 1959) consente di indagare simultaneamente la validità convergente e discriminante
- ✓ L'applicazione del *MTMM* richiede che vengano misurati almeno tre costrutti e che ciascuno di essi sia misurato con almeno tre metodi distinti

Esempio

- 3 Tratti (Competenze in Italiano, Matematica, Inglese)
- 3 metodi (prova standardizzata, valutazione studente, valutazione insegnante)

Validità convergente/discriminante

Validità costrutto

Validità di contenuto

Validità empirica

Validità di facciata

✓ Multitrait-Multimethod Matrix (MTMM; Campbell e Fiske, 1959)

1. Correlazione Massima: stesso tratto - stesso metodo
2. Correlazione Alta: stesso tratto – differente metodo
3. Correlazione Bassa differente tratto - stesso metodo
4. Correlazione Nulla: differente tratto – differente metodo

Validità convergente/discriminante

Validità costruito

Validità di contenuto

Validità empirica

Validità di facciata

LEGENDA MATRICE

- ✓ **Stesso tratto - stesso metodo (sfondo blu)**
- ✓ **Stesso tratto - differente metodo (sfondo verde)**
- ✓ **Differente tratto - stesso metodo (sfondo rosso)**
- ✓ **Differente tratto - differente metodo (sfondo fucsia)**

1 = MAX

AL = ALTA

BS = BASSA

0 = NULLA

Validità convergente/discriminante

Validità costrutto
 Validità di contenuto
 Validità empirica
 Validità di facciata

	Metodo	1			2			3		
Metodo	Tratti	A	B	C	A	B	C	A	B	C
1	A	1								
	B	BS	1							
	C	BS	BS	1						
2	A	AL	0	0	1					
	B	0	AL	0	BS	1				
	C	0	0	AL	BS	BS	1			
3	A	AL	0	0	AL	0	0	1		
	B	0	AL	0	0	AL	0	BS	1	
	C	0	0	AL	0	0	AL	BS	BS	1

Validità di contenuto

Validità costrutto
Validità di contenuto
Validità empirica
Validità di facciata

- ✓ Una misura ha validità di contenuto quando i suoi indicatori rappresentano in modo accurato l'universo di contenuto misurato

Validità empirica

Validità costrutto
Validità di contenuto
Validità empirica
Validità di facciata

Di criterio: si riferisce alla correlazione esistente tra lo strumento in esame ed altre variabili considerate come criterio

1. Predittiva: rappresenta la capacità di uno strumento di fare previsioni accurate. Intesa come correlazione tra il test e il risultato futuro, es. lavoro o profitto scolastico

- *Il punteggio verbale del SAT (Scholastic Aptitude Test e Scholastic Assessment Test) correla .40 con la media degli esami del primo anno alla fine del primo semestre*

Validità empirica

Validità costrutto

Validità di contenuto

Validità empirica

Validità di facciata

Di criterio: si riferisce alla correlazione esistente tra lo strumento in esame ed altre variabili considerate come criterio

2. Concorrente: è determinata osservando quanto lo strumento correla con altri strumenti che il ricercatore ritiene validi nel misurare la stessa caratteristica

- *“Il punteggio nella scala di autovalutazione dell’Aggressività da parte degli scolari correla .70 con la valutazione degli scolari fatta dagli insegnanti”*
- *“Il punteggio nella Scala di Disperazione di Beck (Beck Hopelessness Scale) correla .81 con il Beck Depression Inventory.”*

Validità di facciata

Validità costrutto
Validità di contenuto
Validità empirica
Validità di facciata

Uno strumento ha una buona validità di facciata se ha le sembianze di uno strumento che misura il costrutto che intende misurare

“ Quanto convincenti appaiono le prove che il soggetto deve affrontare; essa viene anche definita validità d'immagine e rappresenta il concetto che un test dovrebbe avere l'apparenza, per ogni persona, di quello che si suppone sia un test”

[McBurney, 1994]

NB: attenzione alla troppa validità di facciata però ...



Attendibilità delle misure

Attendibilità delle misure

- ✓ Il concetto di attendibilità rimanda al grado di stabilità e precisione di una procedura di misurazione. Consiste nella capacità dello strumento di ottenere misurazioni coerenti del fenomeno in analisi (vs. variabilità)
- ✓ Stabilità del punteggio nelle differenti somministrazioni
- ✓ Grado in cui la varianza osservata dei punteggi è sovrapponibile alla varianza vera
- ✓ Grado di accordo tra misurazioni indipendenti dello stesso costrutto (persone, tempi, luoghi diversi)
- ✓ Concetti alla sua base sono la **precisione** e la **stabilità**

Attendibilità delle misure

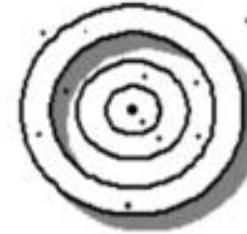
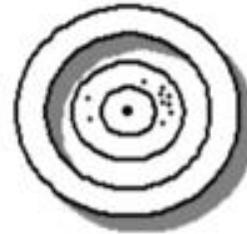
- ✓ La **precisione** è una misura di attendibilità interna alla misurazione
 - ✓ Grado di sistematicità o coerenza con cui eseguiamo una misurazione, cioè coerenza interna della misurazione, tra gli indicatori.
 - ✓ Si riferisce alla coerenza tra manifestazioni apparentemente diverse dello stesso costrutto all'interno dello stesso strumento di misura (correlazione tra gli indicatori in una scala di misura)
- ✓ La **stabilità** è una misura di attendibilità tra misurazioni
 - ✓ Grado di correlazione tra due o più occasioni di misurazione indipendenti dello stesso costrutto.
 - ✓ Si riferisce alla coerenza nel tempo di misurazioni dello stesso costrutto.
 - ✓ Es.: Due testimonianze di uno stesso testimone a distanza di molto tempo.
- **Accuratezza** = Capacità di cogliere con poca distorsione il valore del costrutto misurato È alla base della validità.

Metafora del tiro al bersaglio:

a) *preciso e accurato*
(misura attendibile
e ragionevolmente valida)

b) *inaccurato ma preciso*
(misura attendibile, ma poco valida)

c) *impreciso e perciò*
Inaccurato (né attendibile
Né validi)



Gli item correlano tra di loro e riflettono il reale costruito

Gli osservatori sono d'accordo tra di loro e codificano correttamente

Gli item correlano tra di loro ma non riflettono il reale costruito

Gli osservatori sono d'accordo tra di loro ma entrambi sbagliano la codifica nella stessa direzione

Gli item non correlano tra di loro e non possono riflettere il costruito

Gli osservatori non sono d'accordo tra di loro e non possono riflettere il costruito

Attendibilità delle misure

- 2 somministrazioni -



- **Test-retest:** si applica lo stesso strumento agli stessi soggetti in due situazioni successive e si calcola il coefficiente di correlazione tra le due serie di punteggi ottenuti
- Interpretazione attendibilità test-retest
 - Buoni coefficienti test-retest dovrebbero superare .80
 - 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

Attendibilità delle misure

- 2 somministrazioni -



- **Forme parallele:** richiede una doppia somministrazione agli stessi individui, gli item o i test saranno però forme in tutto equivalenti, diversi solo per il fraseggio
- Garantirsi che le due forme siano realmente parallele

Attendibilità delle misure

- 1 somministrazione -



- **Split-half:** Consiste nel dividere un test (o una scala multi-item) in due parti e nel calcolare la correlazione tra le due serie di punteggi così ottenute
 - Metà item
 - Pari e dispari
 - Altri criteri ...

- Risente della lunghezza della scala

Attendibilità delle misure

- consistenza interna -



- **Coefficiente α di Cronbach**: dipende dalla media delle inter-correlazioni tra tutti gli item del test, e dalla relazione di ogni item del test con il punteggio totale

$$\alpha = \frac{n}{(n-1)} \left(1 - \frac{\sum_{i=1}^n S_i^2}{S_t^2} \right)$$

n : numero item

S_i^2 : varianza di ciascun item

S_t^2 : varianza totale del test

Attendibilità delle misure - consistenza interna -



I valori di α di Cronbach (da 0 a 1)

Valori	Valutazione
$\alpha < .60$	inaccettabile
$.60 < \alpha < .65$	indesiderabile
$.65 < \alpha < .70$	appena accettabile
$.70 < \alpha < .80$	buono
$\alpha > .80$	ottimo

Attendibilità delle misure

- consistenza interna -

- Per l'osservazione di comportamenti:

Indici di concordanza-accordo (agreement)

1. Concordanza tra osservazioni di **diversi osservatori** prodotte nello **stesso tempo** (inter-rater agreement)
2. Concordanza tra osservazioni dello **stesso osservatore** prodotte in **tempi diversi**
3. Concordanza tra osservazioni di **diversi osservatori** prodotte in **tempi diversi**
4. Coerenza interna dell'osservazione dello **stesso osservatore** in un **stesso tempo** (*stabilità interna*)

Attendibilità delle misure

- consistenza interna -

- La misura più semplice di accordo tra codificatori è la percentuale di accordo definita con la seguente formula

$$P_a = \frac{N_a}{N} * 100$$

- N_a = numero di casi nei quali i codificatori si sono trovati d'accordo
- N = numero totale eventi osservati
- *Probabilità accordi dovuti al caso ??*

Attendibilità delle misure

- consistenza interna -

- Per osservazione di comportamenti:

- **K di Cohen**

- Si ha la compilazione di una griglia da parte di almeno due osservatori indipendenti
- Si procede a calcolare l'accordo fra giudici:
 - Matrice di accordo
 - Percentuale di accordo delle risposte degli osservatori indipendenti, che però è influenzata anche dall'accordo casuale

$$K = \frac{p_o - p_e}{1 - p_e}$$

p_o : proporzione di accordo osservato

p_e : proporzione di accordo atteso

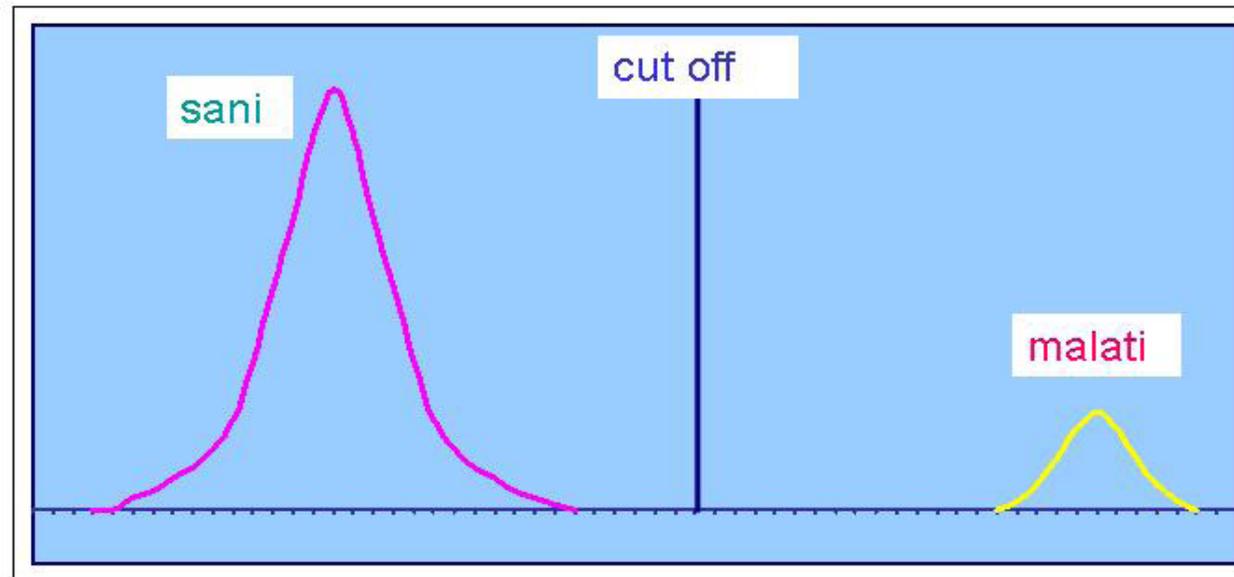
Test diagnostici

Falsi negativi e Falsi positivi

- In una situazione ideale ci si aspetta che un test sia in grado di discriminare perfettamente due popolazioni (es. sani e malati) non sovrapponibili (mutuamente esclusive), dove il 'cut off' rappresenta il valore soglia del test.

Falsi negativi e Falsi positivi

LA SITUAZIONE IDEALE

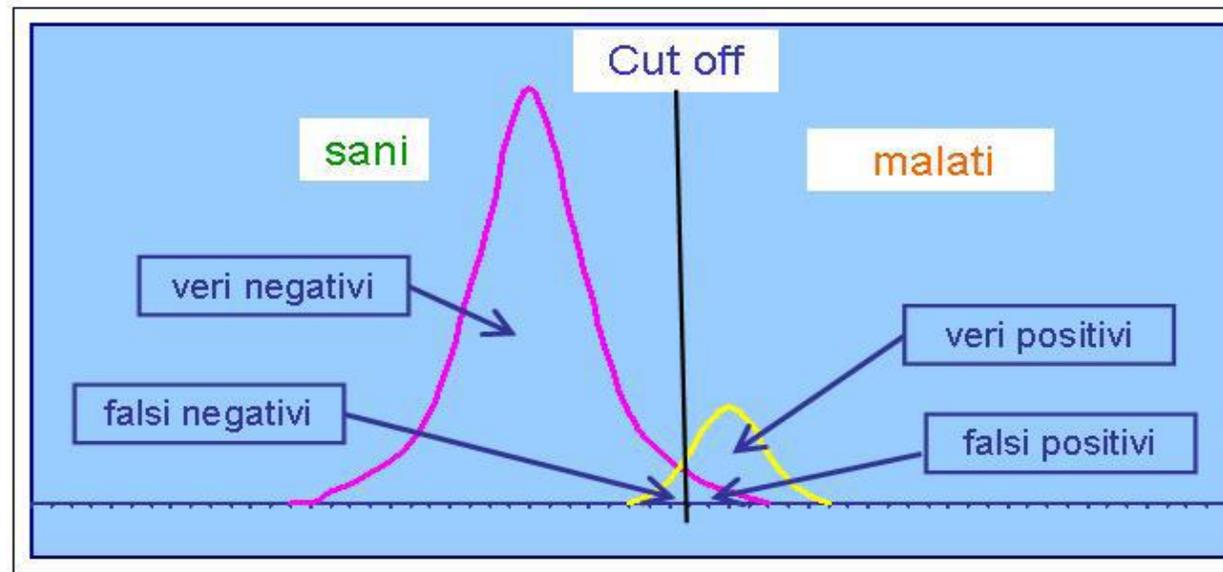


Falsi negativi e Falsi positivi

- In realtà quello che solitamente avviene è che le due popolazioni si sovrappongono in parte, ed il test necessariamente identificherà come positivi alcuni soggetti non malati (Falsi Positivi) e come negativi alcuni soggetti invece malati (Falsi Negativi).

Falsi negativi e Falsi positivi

LA SITUAZIONE REALE



Sensibilità e Specificità

- La **sensibilità** di un test che misura X è la probabilità che un soggetto portatore di X presenti un test positivo. [Un esame è cioè sensibile al 100% quando tutti i malati risultano positivi]
- La **specificità** di un test che misura X è la probabilità che un soggetto non portatore di X presenti un test negativo. [Un test è specifico al 100% quando tutti i sani risultano negativi]

Sensibilità e Specificità

Diagnosi di Patologia		Gold standard		Totale
		<i>Malato</i>	<i>Non malato</i>	
Test	<i>Si</i>	A – Veri positivi	B - Falsi positivi	A+B Totale positivi
	<i>No</i>	C – Falsi negativi	D – Veri negativi	C+D Totale negativi
Totale		A+C Totale Malati	B+D Totale Non malati	A+B+C+D Totale valutazioni

Sensibilità e Specificità

Diagnosi di Patologia		Gold standard		Totale
		<i>Malato</i>	<i>Non malato</i>	
Test	<i>Si</i>	a – Veri positivi	b - Falsi positivi	a+b Totale positivi
	<i>No</i>	c – Falsi negativi	d – Veri negativi	c+d Totale negativi
Totale		a+c Totale Malati	b+d Totale Non malati	a+b+c+d Totale valutazioni

Specificità (Veri Negativi) = $d/(b+d)$

Tasso Falsi Negativi = $c/(a+c)$

Sensibilità (Veri Positivi) = $a/(a+c)$

Tasso Falsi Positivi = $b/(b+d)$

Accuratezza = $a+d/(a+b+c+d)$

Sensibilità e Specificità

- La **Simulazione** tende ad incrementare il tasso di Falso Positivo riducendo la Specificità della procedura diagnostica.
- La **Dissimulazione** tende ad incrementare il tasso di Falso Negativo riducendo la Sensibilità della procedura diagnostica.

Utilizzo di più test

- il clinico può attendere i risultati di un primo test e farne un secondo in funzione dei risultati del primo (“test in serie”); oppure può somministrare una batteria di test e fare una diagnosi in funzione di tutti i risultati dei vari test (“test in parallelo”).
 - I “test in serie” incrementano la Specificità (% dei Veri Negativi), ma riducono la Sensibilità (% dei Veri Positivi), del protocollo diagnostico in quanto una diagnosi positiva alla fine viene posta solo per quei pazienti con un risultato positivo ai vari test somministrati in successione. Così facendo, infatti, viene ridotta la proporzione dei Falsi Positivi, ma anche quella dei Veri Positivi (Sensibilità).
 - I “test in parallelo”, al contrario, globalmente accrescono la Sensibilità, ma riducono la Specificità del protocollo diagnostico; in questo caso si favorisce l’identificazione dei Veri Positivi ma aumenta la quota dei Falsi Negativi.