
VALIDITA' DELLE MISURE

Misurazione

- In psicologia si misurano spesso costrutti, formati da dimensioni (o fattori), tramite indicatori.
- Ogni atto di misurazione comporta degli errori
- Gli errori possono essere di due tipi:
 - ❖ Casuali
 - ❖ Sistematici
- Per tenere sotto controllo l'incidenza degli errori, si sono introdotti i concetti di *attendibilità* e *validità*

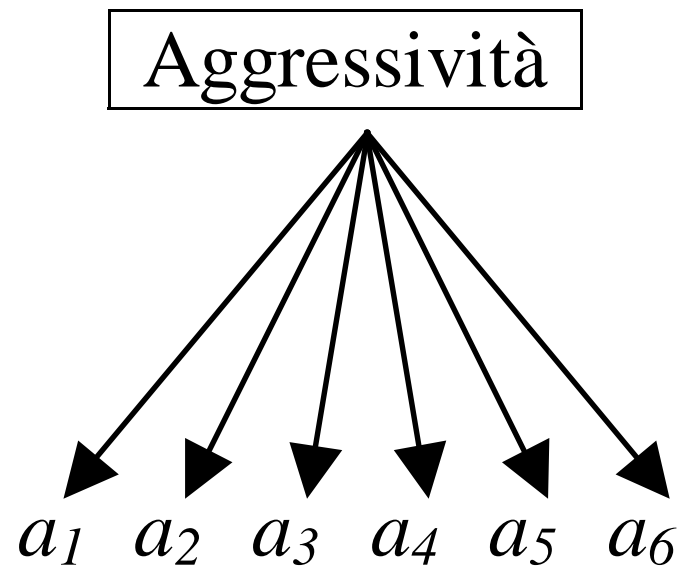
Cos'è un costrutto?

- Concetti astratti che indicano un aspetto della vita psichica del soggetto, non osservabili direttamente ma inferiti a partire da una serie di indicatori empirici osservabili.
- Il costrutto è “il prodotto di una fondata riflessione scientifica, un’idea sviluppata per permettere la categorizzazione e la descrizione di alcuni comportamenti direttamente osservabili” (Crocker & Algina 1986)
- Viene estratto da una teoria (es., narcisismo)
- Deve essere definito in modo operativo
- Può essere misurato da pochi o molti indicatori

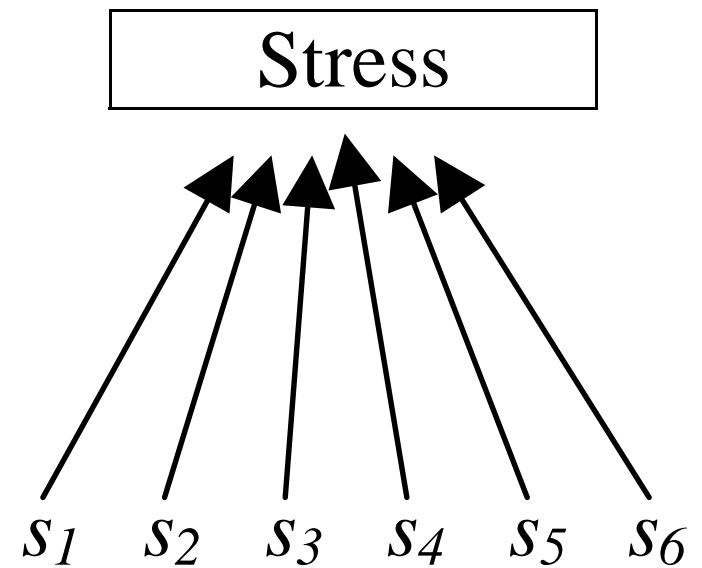
Cos'è un indicatore?

- È una variabile o una misura empirica osservabile che indica il costrutto non osservabile, tramite una regola di corrispondenza.
- Il rapporto tra costrutto e indicatori può essere di due tipi. Ciò dà nome agli indicatori:
 - ❖ **RIFLETTIVI**: gli indicatori riflettono il costrutto nel senso che sono una semplice manifestazione empirica del costrutto, un suo prolungamento osservabile, una conseguenza della presenza del costrutto stesso (aggressività e insulto);
 - ❖ **FORMATIVI**: gli indicatori non sono una manifestazione del costrutto, ma al contrario, formano, contribuiscono, determinano o addirittura causano il costrutto (es., stress; “rischio”).

Indicatori riflettivi e formativi



indicatori riflettivi



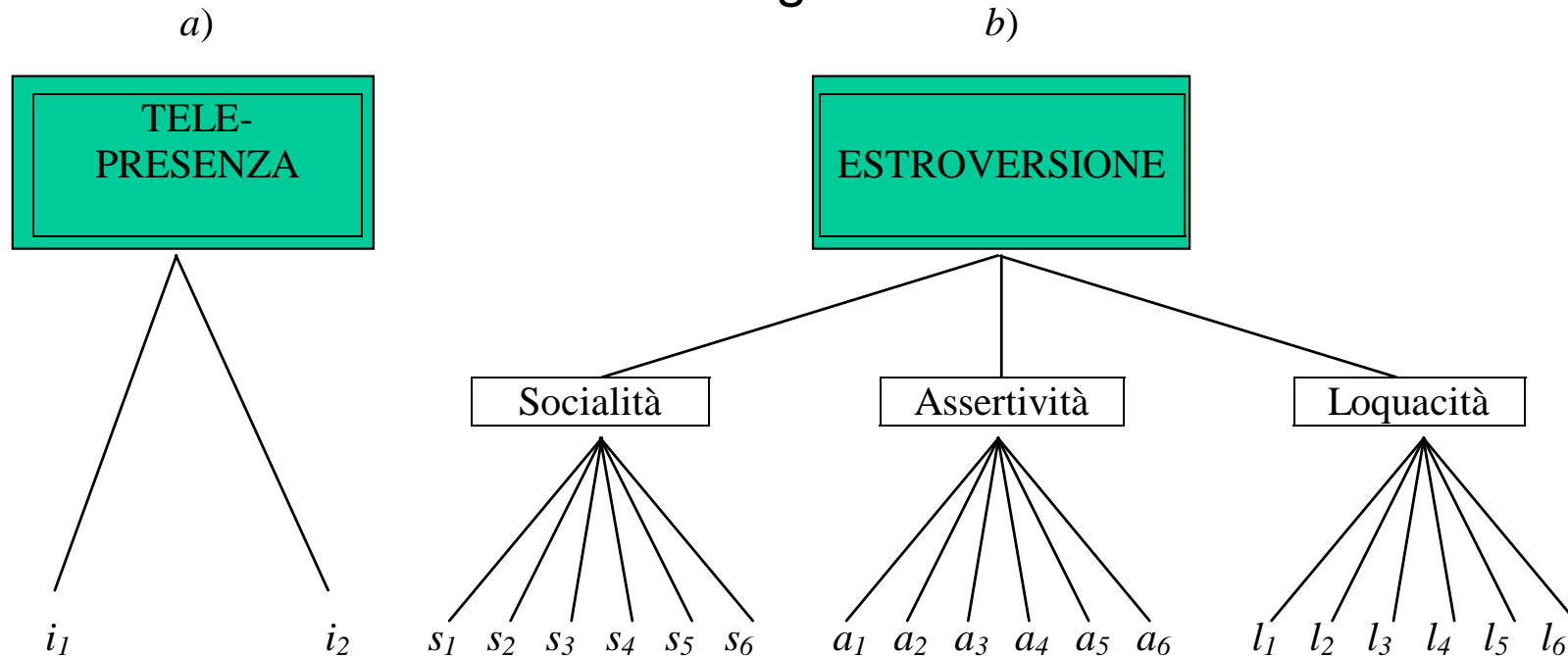
indicatori formativi

Cos'è un indicatore?

- Quando gli indicatori sono riflettivi, il costrutto è causa degli indicatori
- Quando sono formativi, essi stessi sono causa del costrutto
- Questo comporta una diversa considerazione degli errori, delle correlazioni tra item e delle tecniche statistiche da applicare:
 - ❖ **RIFLETTIVI**: devono correlare fra loro
 - ❖ **FORMATIVI**: possono non correlare fra loro.
- In psicologia spesso si usano item riflettivi.
- Per scale sintomatologiche, non è infrequente utilizzare item formativi.

Cos'è una dimensione?

- Esistono costrutti più complessi che implicano un'organizzazione gerarchica in *dimensioni* (o *fattori*) (caso *b*).
- Altri costrutti più concreti non necessitano spesso di una gerarchia concettuale (caso *a*)
- Le dimensioni sono aspetti psicologici organizzati non osservabili che rappresentano ognuno una parte del costrutto e che si collocano gerarchicamente tra i costrutti e gli indicatori.



Dimensionalità dei costrutti

- Lo *studio della dimensionalità* di un costrutto è lo studio del **numero** e delle **caratteristiche** delle dimensioni di un costrutto;
- Lo si esegue con tecniche di analisi fattoriale, che permettono di identificare le dimensioni latenti dei costrutti, raggruppando gli item che correlano più tra loro e scegliendo tra le possibili soluzioni la più semplice.
- Gli item all'interno della dimensione devono correlare molto tra loro.
- Le dimensioni o fattori devono invece essere indipendenti (non correlare).

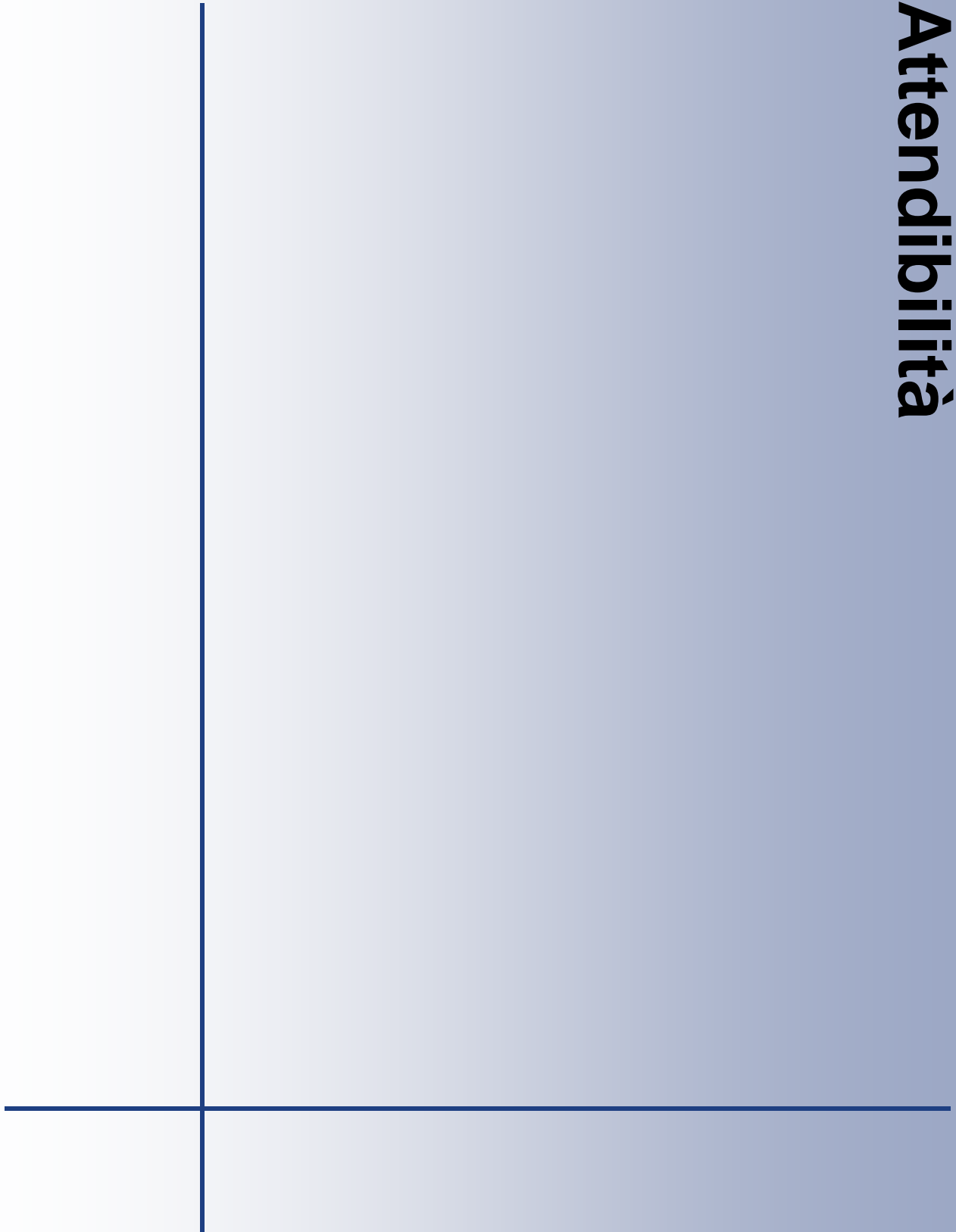
Il peso della teoria

- 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.

Strumenti psicometrici di analisi dei dati

Tecniche di ricerca e analisi dati

Attendibilità



Attendibilità e validità: introduzione

- Se nella misurazione del costrutto abbiamo commesso molti errori casuali, la misurazione non ha senso
- INFATTI:
 - ❖ Cosa sappiamo della lunghezza di un tavolo dopo numerose misurazioni che ci danno sempre risultati diversi?
 - ❖ Cosa sappiamo di un evento, oggetto di testimonianza giudiziale, se tutti i testimoni ci forniscono una versione diversa?
 - ❖ Come possiamo dire di aver misurato l'ansia se la nostra scala non discrimina tra soggetti effettivamente ansiosi e non ansiosi?

Attendibilità e validità: Definizioni

- **ATTENDIBILITÀ**: *grado di accordo o coerenza tra misurazioni indipendenti dello stesso costrutto.*
- **VALIDITÀ**: *grado in cui uno strumento misura ciò che dice di misurare.*
- I ricercatori affrontano le questioni relative alla attendibilità e validità tramite la chiarificazione di alcuni concetti di fondo, quali:
 - precisione, accuratezza e stabilità.

Attendibilità (*reliability*)

- **ATTENDIBILITÀ**: “il grado in cui i punteggi dei test sono liberi dagli errori di misura” (APA)
- grado di accordo o coerenza tra misurazioni indipendenti dello stesso costrutto.
- Se sono stati fatti molti errori, questa correlazione è bassa; se pochi, alta.
- Concetti alla sua base sono la precisione e la stabilità.

Precisione

- 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)
 - ❖ o nella stessa sessione di misurazione (l'osservatore codifica con A tutte le volte che un soggetto osservato mette in atto un determinato comportamento)
- Es.:
 - ❖ Coerenza tra le diverse affermazioni all'interno di una stessa testimonianza.
 - ❖ Coerenza di codifica dello stesso comportamento da parte di due osservatori
 - ❖ Coerenza nella risposta a item: “Sono ansioso”, e “Sono calmo”

Stabilità

- 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 costrutto

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

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

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 costrutto

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

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**)

$$X = V + E$$

X = misura rilevata

V = parte vera

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

MISURAZIONE - 1

- L'errore casuale di misurazione **E** è responsabile dell'imprecisione e della scarsa accuratezza 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** (quello che vogliamo misurare)

V_B = parte vera del costrutto **B** (che non volevamo misurare)

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.

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

Coefficiente di Attendibilità

- Da $X = V + E$, si ricava il *coefficiente di attendibilità*, che esprime la proporzione di varianza vera nella misurazione rispetto alla varianza totale.

$$r_{tt} = \frac{\sigma^2(V)}{\sigma^2(X)} = \frac{\sigma^2(X) - \sigma^2(E)}{\sigma^2(X)} = 1 - \frac{\sigma^2(E)}{\sigma^2(X)}$$

- Varia tra:
 - ❖ 0 → il punteggio è composto solo da errore
 - ❖ 1 → il punteggio osservato è vero
 - ❖ 0-1 → esprimono valori intermedi di attendibilità (di errore/vera).

Primo accenno alla validità

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

$$\rho = \frac{V_A}{V_A + V_B + S \dots + E}$$

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

Assunzioni sull'attendibilità - 1

- **La media degli errori (casuali) 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.
- **L'errore (casuale) 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.
- **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.
- **Gli errori di misura (casuali) compiuti in due occasioni diverse sono fra loro indipendenti**

Assunzioni sull'attendibilità - 2

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) Metodo Test-Retest (stabilità):

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

2) Metodo delle Forme Parallele (stabilità):

Si somministrano due versioni equivalenti del test (stessa media e stessa dev. St.) al tempo **T1** ed al tempo **T2**. Quindi si calcola la correlazione tra i le due forme come stima dell'attendibilità test-retest.

3) Metodo Split-Half:

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) Metodo della Coerenza Interna:

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, e da essa si deriva un coefficiente di 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:

$$Cov (PG T1, PG T2) = 73.45$$

$$Dev.St.(PG T1) = 9.03$$

$$Dev.St.(PG T2) = 8.39$$

$$R_{tt} = 73.45 / 9.03 \times 8.39 =$$

$$73.45 / 75.74 = \mathbf{.96}$$

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

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.

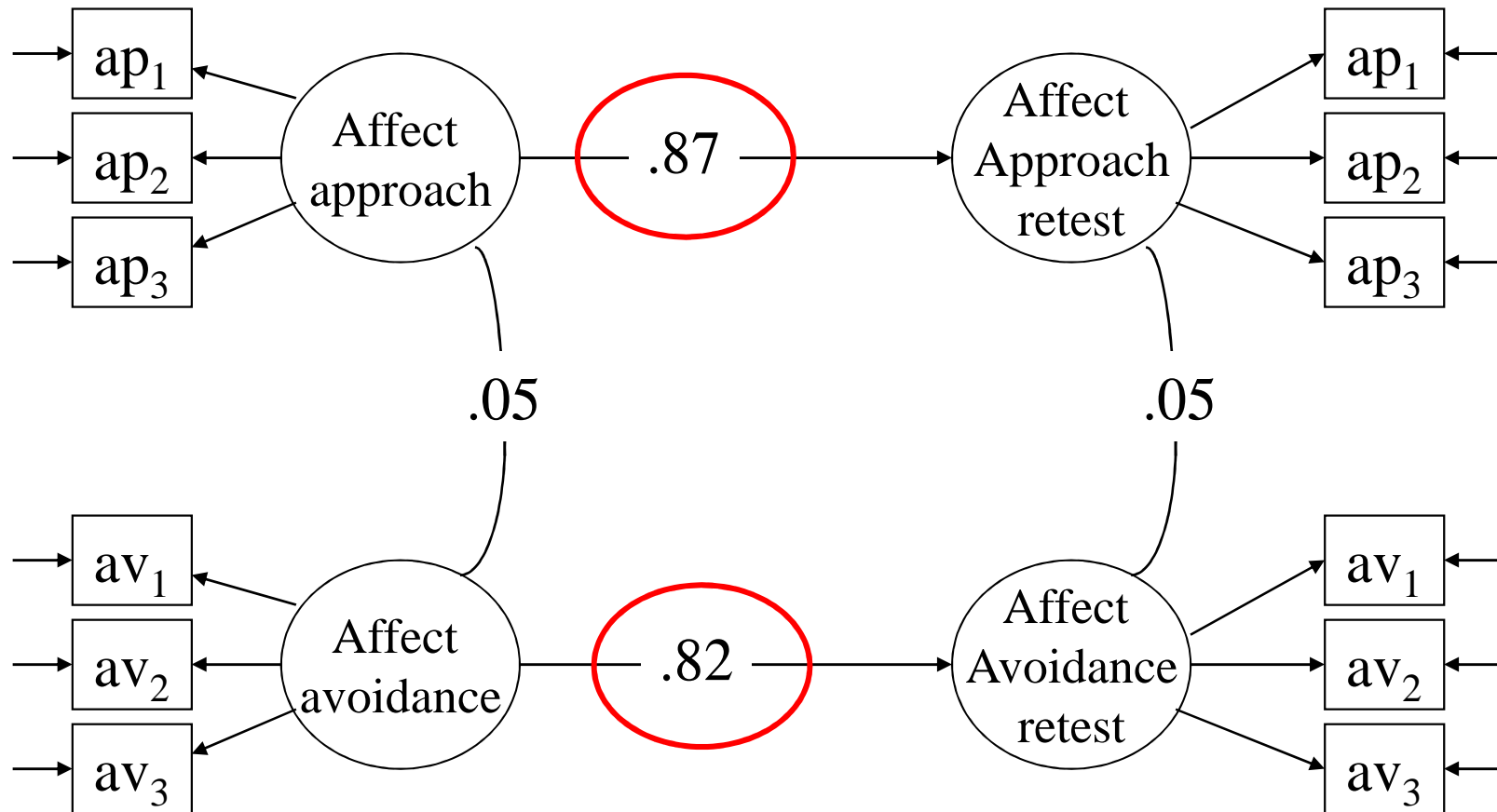
Fare Test F sulle due varianze

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

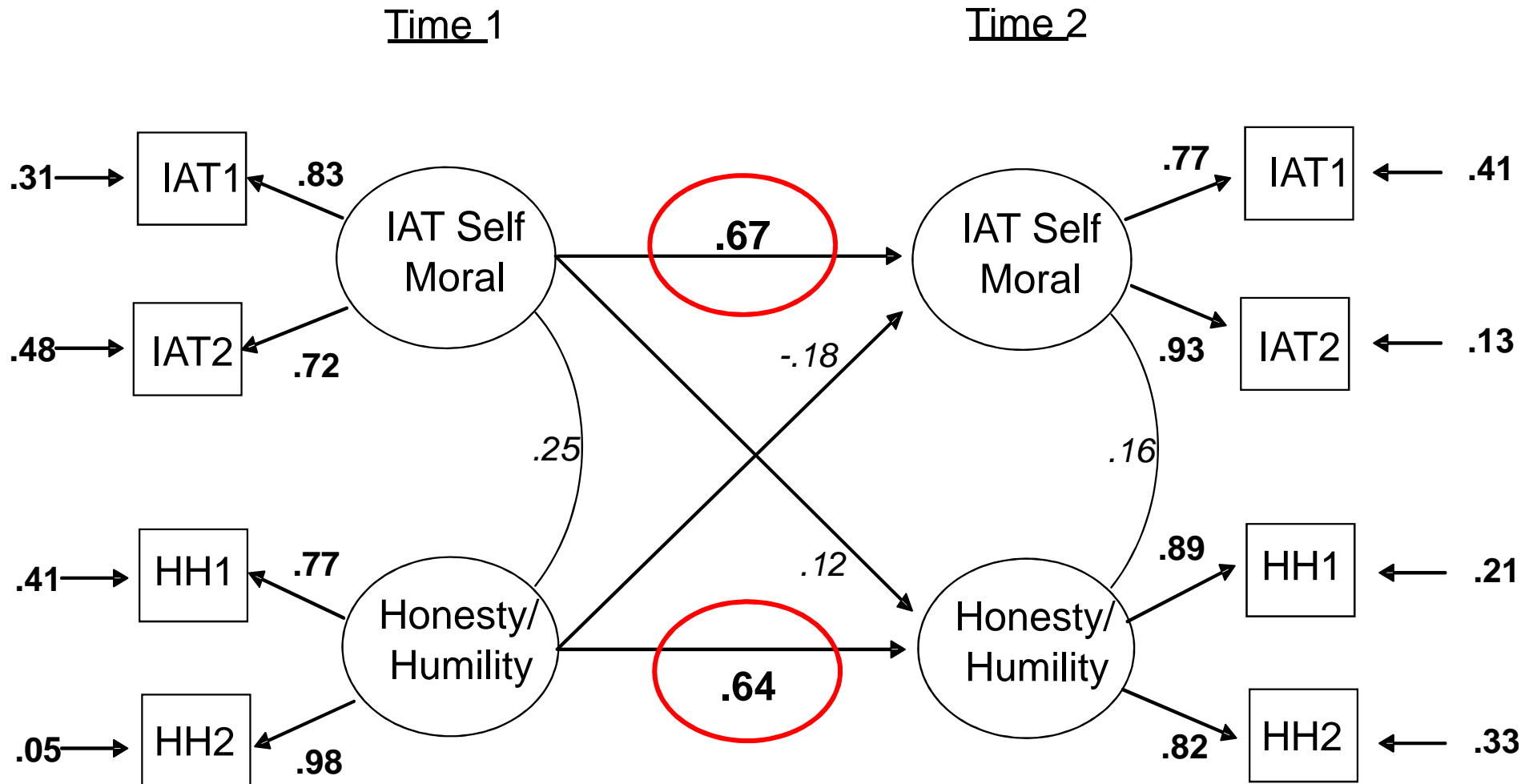
Le varianze sono uguali.

Quindi r. = 96

Esempio test-retest - 1



Esempio test-retest - 2



Split-half

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

Item	1	2	3	4	5	6	7	8	9	10
Sogg. 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***

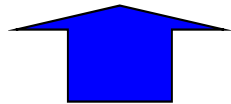
*Calcolare la correlazione tra le due metà = **.40***

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
- Viene spesso utilizzata una formula che permette 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: $Rn_{xx} = ,40$; $n = 10/5 = 2$

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

Coerenza interna: α **DI CRONBACH:**

Generalizzazione dell'attendibilità *split-half*.

SEMPLIFICANDO: Si procede come nello *split-half*, ma si considera ogni singolo item come una versione parallela della scala.

Perciò, l' α di Cronbach:

→ dipende dalla media delle intercorrelazioni tra tutti gli item del test, e dalla relazione di ogni item del test con il punteggio totale.

Nella prassi α di Cronbach si valuta così:

<.60 → problematico

.60-.70 → appena sufficiente

.70-.80 → discreto

>.90 → ottimo/eccellente

Formula α (alfa) di Cronbach

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

in cui:

k = numero di item

σ_i^2 = varianza di ciascun item

σ_X^2 = varianza totale del test

Esprime una misura del peso relativo della variabilità associata agli item rispetto alla variabilità associata alla loro somma.

Scala di Estroversione composta da 8 item (dati fittizi su 10 soggetti, SPSS)

11 : var00004

	var00002	var00003	var00004	var00005	var00006	var00007	var00008	var00009	var	var
1	1,00	1,00	5,00	1,00	2,00	1,00	2,00	1,00		
2	1,00	2,00	5,00	2,00	2,00	2,00	1,00	1,00		
3	5,00	4,00	4,00	4,00	5,00	4,00	5,00	5,00		
4	3,00	3,00	5,00	3,00	1,00	1,00	4,00	4,00		
5	1,00	2,00	5,00	2,00	1,00	2,00	1,00	2,00		
6	3,00	3,00	5,00	4,00	3,00	3,00	3,00	3,00		
7	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00		
8	2,00	3,00	5,00	3,00	2,00	2,00	2,00	3,00		
9	1,00	1,00	5,00	1,00	1,00	3,00	5,00	2,00		
10	1,00	1,00	5,00	1,00	1,00	1,00	1,00	1,00		
11										
12										
13										
14										
15										
16										
17										
18										

Visualizzazione dati / Visualizzazione variabili

SPSS Processore pronto

**Scala di Estroversione composta da 8 item
(dati fittizi su 10 soggetti, SPSS)**

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)				
N of Cases =		10,0		
Statistics for	Mean	Variance	Std Dev	N of
Scale	22,2000	73,9556	8,5997	Variables
Reliability Coefficients		8 items		
Alpha =	,9176	Standardized item alpha =	,8539	

Item-total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
VAR00002	19,9000	49,2111	,9611	,9873	,8845
VAR00003	19,8000	56,8444	,8889	,9895	,8958
VAR00004	17,3000	77,1222	-,5881	,8882	,9477
VAR00005	19,6000	53,3778	,8871	,9900	,8927
VAR00006	20,0000	55,5556	,7888	,9540	,9016
VAR00007	19,9000	58,9889	,7648	,8540	,9049
VAR00008	19,4000	56,4889	,6098	,8451	,9197
VAR00009	19,5000	50,7222	,9309	,9876	,8878

Considerazioni pratiche

- Per meglio valutare α , bisogna ricordare che essa dipende da due fattori:
 1. intercorrelazioni tra gli item
 2. lunghezza della scala (numero di item)
- Infatti, a parità di condizioni, all'aumentare del numero degli item, aumenta il valore del coefficiente di attendibilità.
- Perciò, per esempio, basterebbe aumentare il numero di item di una scala che ha un coefficiente sufficiente per ottenerne uno buono.
 - ❖ Ma bisogna pensare ai poveri rispondenti che si debbono sorbire una lista molto lunga di domande!

Specificità vs generalità

- α dipende anche dalle intercorrelazioni tra gli item e a volte l'intercorrelazione dipende dalla comune specificità degli item
- Se perciò per misurare l'estroversione utilizzo solo item relativi ad andare alle feste (“Mi piace andare alle feste”, “Alle feste mi diverto molto”, “Alle feste non mi diverto per niente”, “Preferisco andare ad una festa che andare al cinema”) sto di fatto misurando la stessa variabile con una forma leggermente diversa.
 - E' ovvio che, in queste condizioni, le loro correlazioni siano altissime e questo, di conseguenza, “gonfi” l'attendibilità.
- Coefficienti alti di queste scale non solo non sono una garanzia di attendibilità
 - anzi indicano un'estrema ristrettezza del contenuto del costrutto (enfaticano l'estroversione legata alla feste e non l'introversione in altri costrutti).

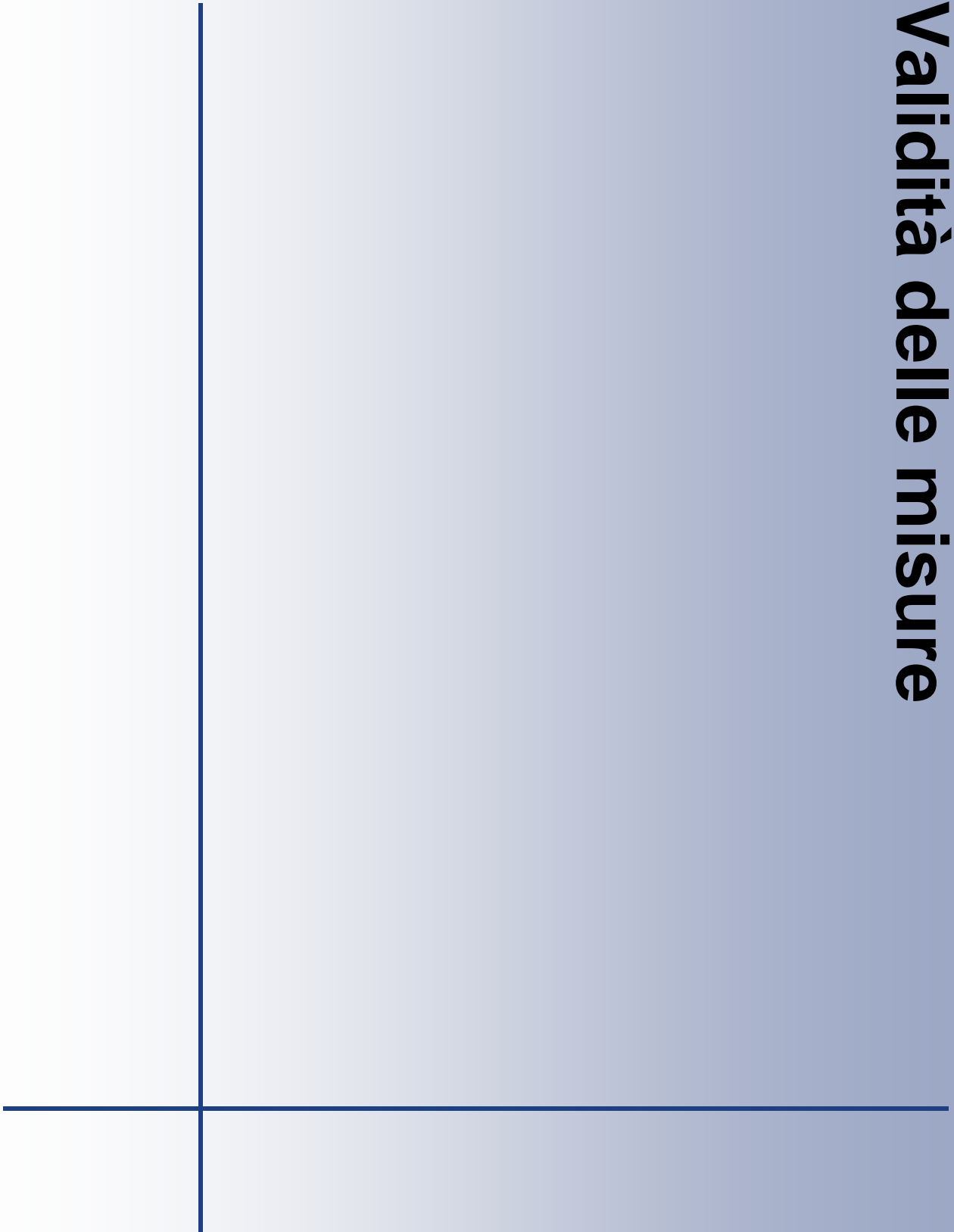
Questo comporta DUE PROBLEMI:

- 1) IL PARADOSSO DELL'ATTENUAZIONE:
 - incrementare oltre ogni limite la consistenza interna di un test non aumenta necessariamente la validità di costrutto (anzi la può limitare, come nel caso precedente)
- 2) DILEMMA DELLA LARGHEZZA DI BANDA:
 - l'eccessivo restringimento del contenuto della scala (es., piacere di andare alle feste), se anche permette di predire comportamenti specifici, come andare alle feste, probabilmente non correla con nient'altro di interesse psicologico;
 - l'eccessivo allargamento del contenuto (es., generico atteggiamento di apertura sociale) ha un grosso potere predittivo rispetto a costrutti psicologicamente rilevanti, a spese però di comportamenti specifici
 - Alta specificità, alta predittività di singoli comportamenti, ma scarsa predittività attraverso diverse situazioni
 - Bassa specificità, bassa predittività di singoli comportamenti, ma alta predittività attraverso diverse situazioni

Strumenti psicometrici di analisi dei dati

Tecniche di ricerca e analisi dati

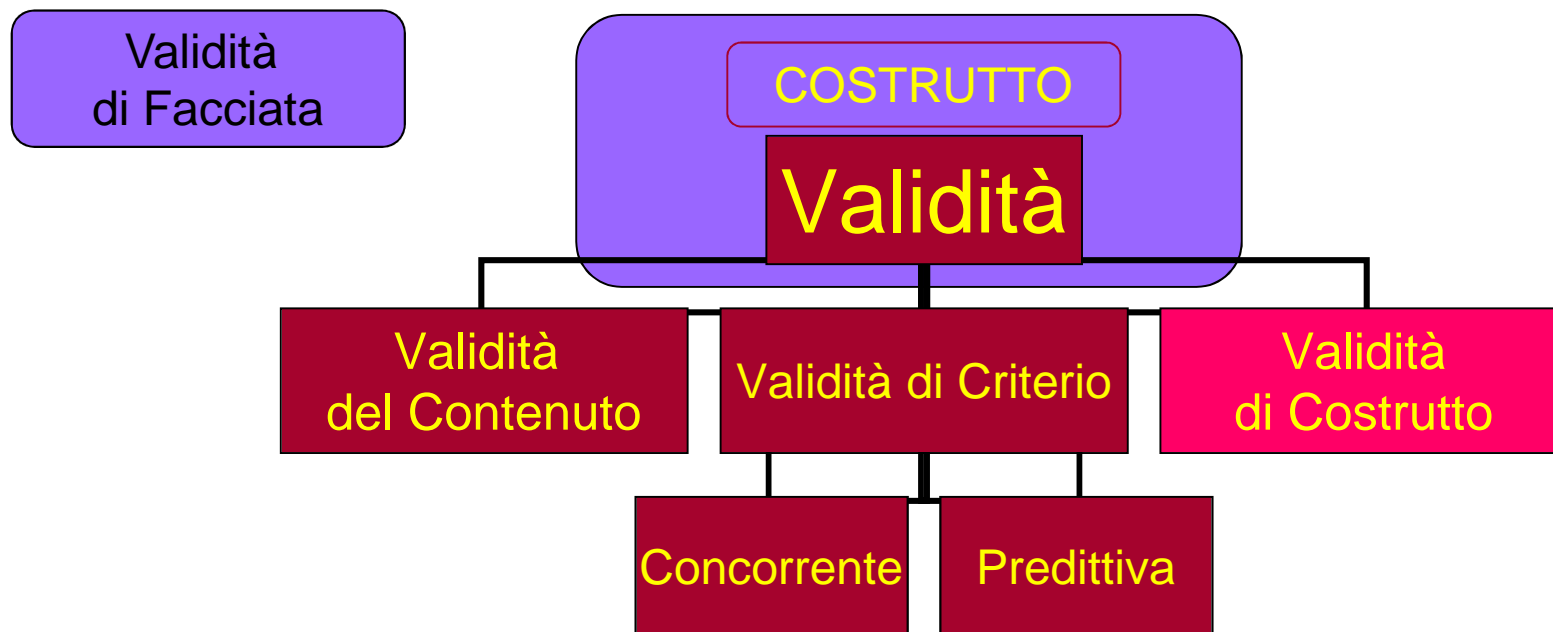
Validità delle misure



Che cosa è la Validità?

- Una misura è valida quando **misura ciò che intende misurare**.
- 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)
- La validità si articola in diverse sfumature, e in diverse modalità di giudicare raggiunte o meno differenti forme di validità

Tassonomia Tradizionale



Validità di facciata

- Il test sembra una misura sensata del costrutto.
 - ❖ Leggendo gli item, riesco a capire che costrutto voglio misurare.
- È importante nelle situazioni di selezione o di screening diagnostico
 - ❖ Una buona validità di facciata motiva i rispondenti (aspetto positivo)
 - ❖ Una buona validità di facciata permette ai rispondenti di aggiustare le proprie risposte (aspetto negativo)
- Non è particolarmente importante nei contesti di ricerca

Validità di Contenuto

- Rappresentatività del Contenuto del costrutto.
(Campionamento degli item)
 - ❖ Rilevanza del Contenuto
 - ❖ Gli item del test devono coprire l'intera definizione teorica del costrutto in esame, e cogliere gli aspetti più importanti del costrutto.
- **Senza validità del contenuto non ci sono altre forme di validità**
 - ❖ Se non abbiamo colto adeguatamente gli aspetti del costrutto, non può esservi alcuna forma di validità
- Questa forma di validità **non** viene valutata in modo quantitativo

Validità Rispetto al Criterio

- **Validità Concorrente** : il punteggio al test concorda con altre misure valide dello stesso costrutto.
- La validità concorrente si stabilisce, in generale, tramite la correlazione.
- Esempio
 - ❖ *“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.”*

Difficoltà nel verificare la validità concorrente

- In senso stretto, misurare la validità concorrente è possibile se esiste una misura valida che funga da paragone.
- Se non abbiamo una misura valida di confronto, la validità concorrente non può essere **verificata**. Può essere però **indagata** in modo indiretto.
 - ❖ È possibile indagare se esistono delle relazioni coerenti fra il nuovo test e misure (valide) di altri costrutti che debbono essere in vario modo collegate al costrutto per il quale si vuole sviluppare una misura

Validità Rispetto al Criterio

- Se una misura è usata per avere una stima del punteggio in un criterio differito nel tempo parliamo di **Validità Predittiva** della misura.
- Validità Predittiva - il punteggio nella misura è capace di predire accuratamente la prestazione nel dominio teorico cui essa appartiene.
- Esempi:
 - ❖ *Il punteggio verbale del SAT (una specie di test d'ingresso) correla .40 con la media degli esami del primo anno alla fine del primo semestre.*
 - ❖ *La scala delle Relazioni Spaziali correla .70 con il successo nella prova di disegno tecnico per studenti degli Istituti Tecnici.*
 - ❖ *Il punteggio di una misura di ansia generalizzata correla .50 con lo status diagnostico determinato da un'intervista clinica eseguita a tre mesi di distanza dallo screening*

Difficoltà di verificare la validità predittiva

- Il criterio deve essere valido concettualmente:
 - ❖ il criterio deve essere teoricamente connesso al costrutto che si vuole misurare.
- Il criterio deve essere misurato in modo valido:
 - ❖ Se un criterio concettualmente valido non è misurato in modo valido e attendibile, non posso interpretare la validità predittiva.
- La validità, in tutti i suoi aspetti, è un problema attinente sia alle variabili indipendenti, sia dipendenti

Validità Rispetto al Criterio: riassunto

Si differenzia in:

- Validità Concorrente

- ❖ il criterio è contemporaneo al test; a volte si tratta di altri strumenti simili al test in esame

- Validità Predittiva

- ❖ il criterio è successivo al test, spesso si tratta di esiti importanti per l'individuo

Si stima con:

- Coefficiente di Validità

- ❖ correlazione r

- Coefficiente di Determinazione

- ❖ $= r^2$

- Differenze fra gruppi-criterio

Esempio di validità concorrente

- Correlazioni fra Affect approach e Affect avoidance e una serie di misure concorrenti

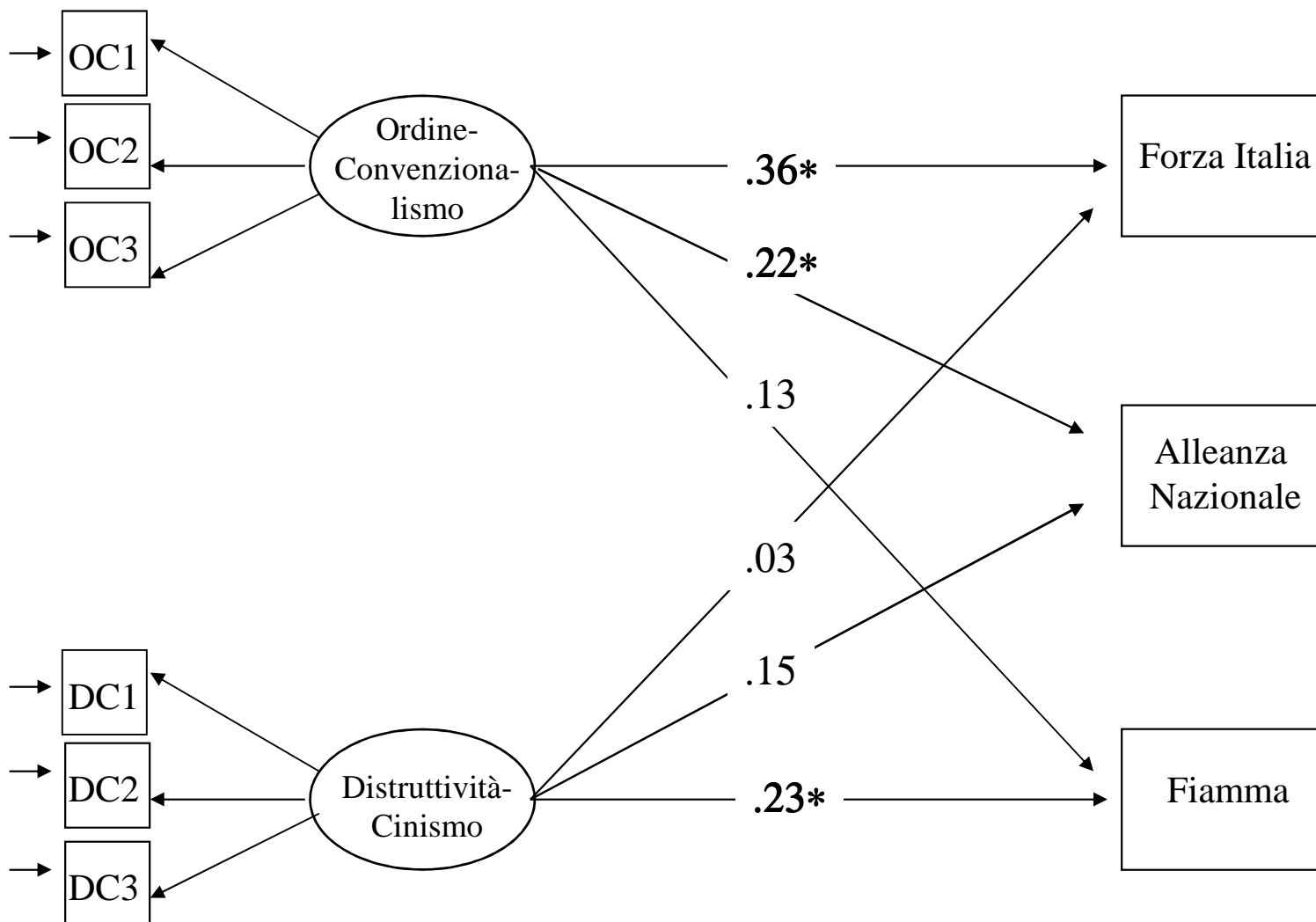
	Approach	Avoidance
<i>BIS/BAS</i> ^a		
BIS	.52**	.25**
BAS	.73**	.00
<i>Alexythimia total score</i> ^a	.11*	.63**
<i>CERQ</i> ^a		
Positive-Focused Cognitive Emotion Regulation	.43**	-.01
Negative-Focused Cognitive Emotion Regulation	.48**	.58**
<i>Emotional Expression</i> ^a		
Expression of Positive Emotions	.49**	-.04
Expression of Negative Emotions	.45**	.11*
<i>Cognitive Need for Closure</i> ^b	-.20**	.42**

Note. ^a Measure administered to the adolescent sample ($N = 608$). ^b Measure administered to the adult sample ($N = 272$).

* $p < .05$; ** $p < .01$.

Esempio di validità predittiva

- Sfaccettature di una misura di autoritarismo e previsione dell'intenzione di voto



Validità di Costrutto

- **Validità di Costrutto** - cerca di verificare se un punteggio al test misura il costrutto d'interesse.

- ❖ Poiché i costrutti sono astrazioni concettuali, la validità bisogna derivarla in maniera indiretta
- ❖ Il modo di procedere è quello di valutare un insieme di prove che **nel loro complesso** stabiliscono la validità di Costrutto

Validità di Costrutto

- La Validità Concorrente è di grande aiuto nel determinare la Validità di Costrutto di una misura.
- Nell'indagare la validità di costrutto si distingue la validità concorrente in due aspetti differenti:
 - ❖ **Validità Convergente** - Alta Correlazione con altre Misure dello stesso Costrutto.
 - ❖ **Validità Discriminante** - Bassa correlazione con altre Misure di altri Costrutti

Validità di Costrutto

- Approccio per “Gruppi Contrapposti”
- Matrici Multi-Tratto-Multi-Metodo MTMM
- Matrici Multi-Tratto-Multi-Item MTMI
- Analisi Fattoriale

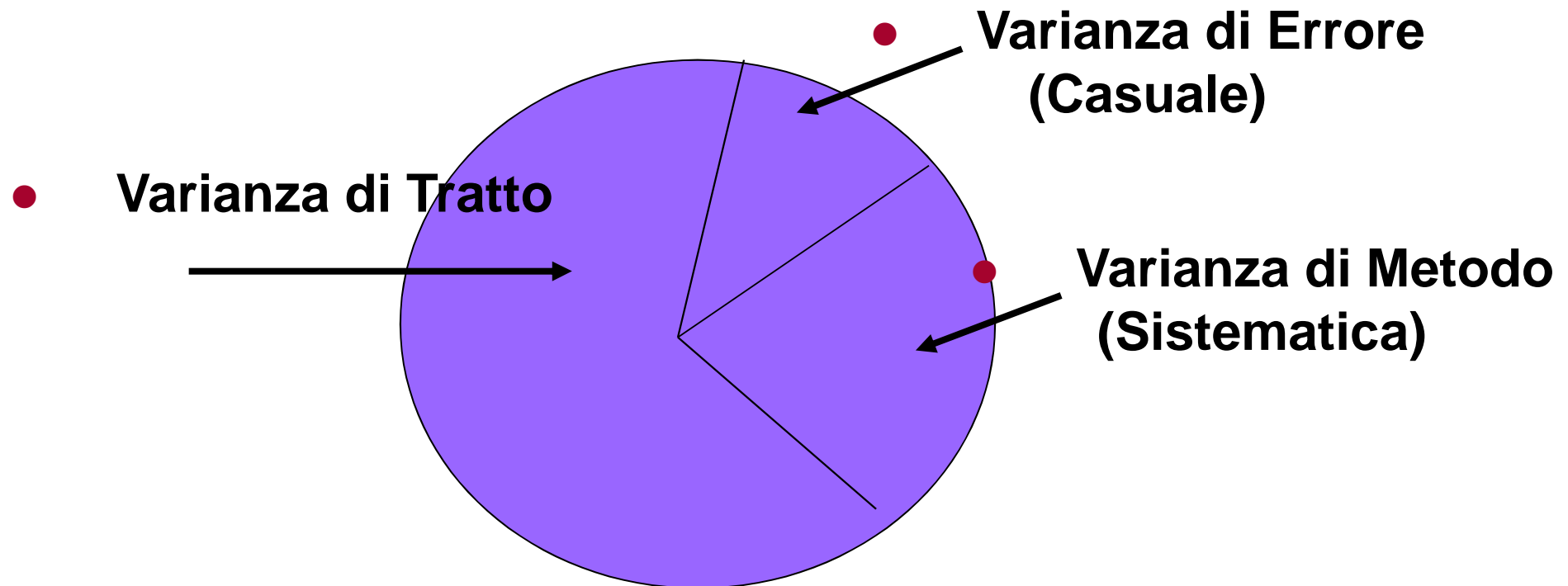
Approccio per Gruppi Contrapposti

- Se uno strumento misura ciò che dichiara di misurare allora soggetti con punteggi Alti e Bassi nella misura devono comportarsi in maniera differente entro il dominio concettuale del costrutto.
- Esempi banali:
 - ❖ **Alti Ansiosi vs. Bassi Ansiosi:** ci si aspetta che i due gruppi si comportino in maniera diversa mentre aspettano i risultati di un esame.
 - ❖ **Attenti vs. Disattenti:** ci si aspetta che i due gruppi si comportino in maniera diversa mentre l'insegnante parla a lezione.

Multi-Tratto Multi-Metodo (MTMM)

Campbell e Fiske, 1959

- Si basa sulla pratica di ottenere più misure degli **stessi** costrutti tramite metodi **diversi**.
- La teoria classica dei test può essere estesa sino ad affermare che: $\text{Punteggio} = \text{Tratto} + \text{Metodo} + E$
- $\text{Metodo} \neq \text{Errore Casuale}$



Multi-Tratto Multi-Metodo (MTMM)

SI DISTINGUONO

- correlazioni **monotratto-eterometodo**
 - ❖ Lo stesso tratto misurato da due o più metodi
 - ❖ Devono essere Alte
- correlazioni **eterotratto-monometodo**
 - ❖ Diversi tratti misurati dallo stesso metodo
 - ❖ Devono essere Basse o Nulle
- correlazioni **eterotratto-eterometodo**
 - ❖ Diversi tratti misurati da metodi diversi
 - ❖ Devono essere Nulle

Matrice Multi-Tratto-Multi-Metodo (MTMM)

Indagare la validità convergente e discriminante - Campbell e Fiske

		<u>Metodo 1</u>			<u>Metodo 2</u>			<u>Metodo 3</u>		
Tratti		A1	B1	C1	A2	B2	C2	A3	B3	C3
M1	A1	r_{A1A1}								
	B1	r_{B1A1}	r_{B1B1}							
	C1	r_{C1A1}	r_{C1B1}	r_{C1C1}						
M2	A2	r_{A2A1}	r_{A2B1}	r_{A2C1}	r_{A2A2}					
	B2	r_{B2A1}	r_{B2B1}	r_{B2C1}	r_{B2A2}	r_{B2B2}				
	C2	r_{C2A1}	r_{C2B1}	r_{C2C1}	r_{C2A2}	r_{C2B2}	r_{C2C2}			
M3	A3	r_{A3A1}	r_{A3B1}	r_{A3C1}	r_{A3A2}	r_{A3B2}	r_{A3C2}	r_{A3A3}		
	B3	r_{B3A1}	r_{B3B1}	r_{B3C1}	r_{B3A2}	r_{B3B2}	r_{B3C2}	r_{B3A3}	r_{B3B3}	
	C3	r_{C3A1}	r_{C3B1}	r_{C3C1}	r_{C3A2}	r_{C3B2}	r_{C3C2}	r_{C3A3}	r_{C3B3}	r_{C3C3}

- ❖ Coefficienti di validità elevati e significativi
- ❖ Coefficienti di validità superiori ai relativi coefficienti nei triangoli eterotratto-eterometodo
- ❖ Coefficienti di validità superiori ai relativi coefficienti nei triangoli eterotratto-monometodo
- ❖ Stesso ordinamento delle correlazioni in tutti i triangoli della matrice

Multi-Tratto Multi-Metodo (MTMM)

un esempio reale...

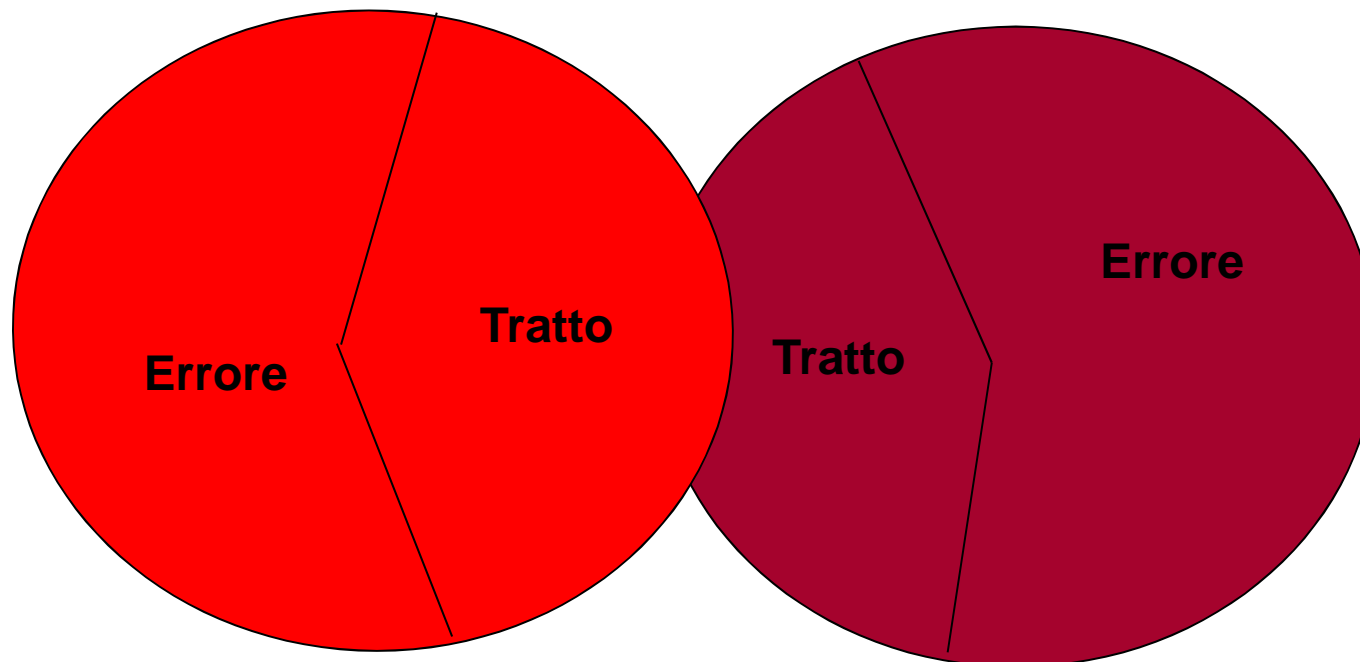
TABLE 3.7
A Synthetic Multitrait–Multimethod Matrix for Validating Measures of Anomie (A), Misanthropy (M), and Prejudice (P) Using Face-to-Face Interviews, Telephone Interviews, and Questionnaires^a

Constructs	Variable	Face-to-face interview			Telephone interview			Questionnaire		
		A	M	P	A	M	P	A	M	P
Face-to-face interview										
A	y ₁	—								
M	y ₂	.36	—							
P	y ₃	.26	.35	—						
Telephone interview										
A	y ₄	.65	.22	.31	—					
M	y ₅	.19	.61	.12	.40	—				
P	y ₆	.14	.15	.54	.32	.45	—			
Questionnaire										
A	y ₇	.50	.37	.42	.51	.33	.40	—		
M	y ₈	.32	.47	.39	.27	.42	.41	.51	—	
P	y ₉	.38	.35	.47	.35	.37	.44	.49	.57	—

^a The three validity diagonals are boldface. Each heterotrait–monomethod triangle is enclosed by solid lines and each heterotrait–heteromethod triangle is enclosed by broken lines.

Attenuazione

- I coefficienti di validità basati sulla correlazione sono distorti dall'attendibilità
- Attendibilità meno che perfette “*attenuano*” la stima della vera validità



Correzione per l'attenuazione

- Correzione del criterio

$$r'_{12} = \frac{r_{12}}{\sqrt{r_{22}}}$$

- Correzione del Test e del Criterio

$$r''_{12} = \frac{r_{12}}{\sqrt{r_{11} r_{22}}}$$

Relazione fra attendibilità e validità

- Una misura valida **deve** essere attendibile
 - Attendibilità condizione necessaria della validità
- Una misura attendibile non sempre è valida
 - Attendibilità condizione non sufficiente per la validità
- Come dimostra l'attenuazione, l'attendibilità rappresenta il limite superiore della validità.

$$Val = \frac{V_A}{V_A + V_B + \dots S + E}$$

$$\rho = \frac{V_A + V_B + \dots S}{V_A + V_B + \dots + E}$$