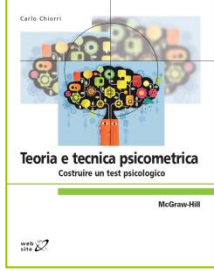
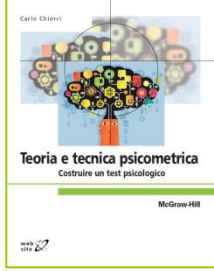
La valutazione preliminare degli item di un test



Valutazione preliminare

- Ruolo marginale della statistica
- Validità di contenuto e di facciata
- Analisi degli item e attendibilità
- Raffinamento del test

Copyright © 2011 The McGraw-Hill Companies srl



Validità di contenuto

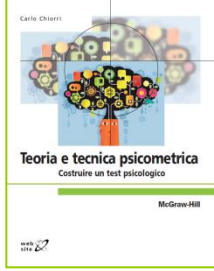
Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose

(Haynes et al., 1995, p. 238)

- Rilevanza (*relevance*): appropriatezza degli elementi del test elementi per la valutazione del costrutto

- Rappresentatività (*representativeness*): grado in cui gli elementi del test sono rappresentativi delle varie sfaccettature del costrutto target e riescono a coprire in modo adeguato il dominio di contenuto

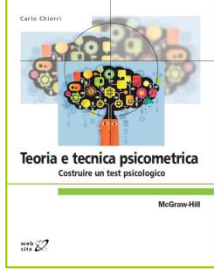
Copyright © 2011 The McGraw-Hill Companies srl



Fattori per la valutazione di rilevanza e rappresentatività

- **uso che si intende fare del test e le inferenze che si intende trarre dai dati ottenuti** (e.g., screening vs valutazione approfondita di un disturbo di personalità)
- **tipo di comportamento che si intende valutare** (e.g., comportamenti in contesti specifici o generalizzati, tipici o atipici, comuni o estremi, etc.)
- **parametro di interesse** (e.g. frequenza, intensità, valutazione)
- **popolazione target**
- **dominio concettuale** (come è definito il costrutto? In quale cornice di riferimento teorica?)
- **natura dinamica** (e.g. validità temporale)

Copyright © 2011 The McGraw-Hill Companies srl



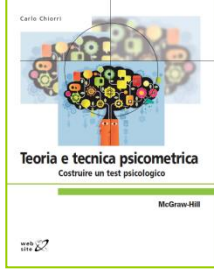
Corrispondenze

- **Metodo-modalità**: grado in cui un particolare metodo di valutazione è appropriato per il costrutto target (e.g., self-report è adatto per contenuti inconsapevoli?)

- **Metodo-funzione**: grado in cui un particolare metodo di assessment è appropriato per gli scopi della valutazione (e.g., self-report è adatto per fare diagnosi?)

- Necessità di un assessment multi-metodo

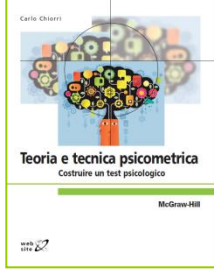
Copyright © 2011 The McGraw-Hill Companies srl



Componenti validità di contenuto

- **Campionamento del dominio**: grado in cui ogni *item* del test corrisponde ad un dominio di contenuto teoricamente definito e il grado in cui l'*insieme degli item* è in grado di rappresentare il dominio di contenuto di interesse
  - **Campionamento del contenuto**: quali contenuti per un test di profitto in psicometria?
  - **Campionamento del comportamento**: quali risposte si ottengono dai particolari item di un test di profitto in psicometria?

Copyright © 2011 The McGraw-Hill Companies srl



Checklist per la verifica della qualità degli item

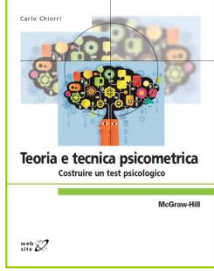
Test di prestazione tipica

Difetto	Item 1	Item 2	Item 3	...	Item *k*
Negazioni					
Forme passive					
Termini tecnici					
Termini ambigui					
Doppie domande					
Desiderabilità sociale					
Insinuazioni					
Carica emotiva					
Periodo di riferimento vago					
Assenza di suggerimenti per la memoria (dove, come quando)					
Vaghezza della domanda					
Assenza di riferimenti temporali					

Test di prestazione massima

Difetto	Item 1	Item 2	Item 3	...	Item *k*
Negazioni					
Incongruenze di genere o numero					
Alternative umoristiche					
Scarsa plausibilità dei distrattori					
Insieme intersezione					
Disomogeneità fra risposta corretta e distrattori					
Suggerimenti impliciti					
Doppie domande					

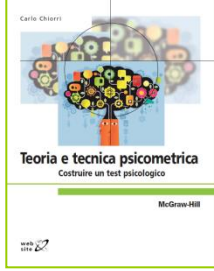
Copyright © 2011 The McGraw-Hill Companies srl



Linee guida per ottenere la validità di contenuto

- **Il dominio di contenuto deve riferirsi a comportamenti con un significato universalmente condiviso** (e.g., test idoneità alla guida vs test di personalità)
- **Definire accuratamente il dominio e le sfaccettature del costrutto e sottoporle a validazione di contenuto prima di sviluppare gli altri elementi dello strumento di valutazione** (e.g. che cosa è e non è il costrutto?)
- **Sottoporre tutti gli elementi di uno strumento di assessment a validazione di contenuto** (istruzioni, formato di risposta, etc.)
- **Per la generazione iniziale degli item e degli altri elementi considerare il parere della popolazione target e degli esperti**

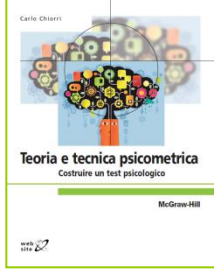
Copyright © 2011 The McGraw-Hill Companies srl



Linee guida per ottenere la validità di contenuto

- **Impiegare più giudici per la validità di contenuto e quantificare i giudizi utilizzando procedure quantitative**
- **Esaminare la rappresentazione proporzionale degli item**
- **Riportare i risultati della validazione di contenuto quando si pubblica un nuovo strumento di valutazione**
- **Utilizzare le successive analisi psicometriche per il raffinamento dello strumento di valutazione**

Copyright © 2011 The McGraw-Hill Companies srl

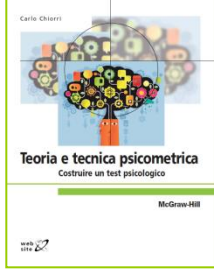


Esempio di metodo per la valutazione della validità di contenuto - esperti*

- Chiedere ad esperti (*Subject Matter Experts*) di valutare quanto è rilevante ogni item

	Quanto è rilevante l'item?		
Item del test	Per niente	Poco	Molto
Item del test	Per niente	Poco	Molto
Item del test	Per niente	Poco	Molto
[…]	[…]	[…]	[…]

Copyright © 2011 The McGraw-Hill Companies srl



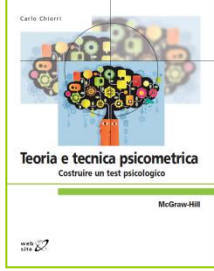
Strategia di Lawshe (1975) Content validity ratio (CVR)

a. l'item che riceve la valutazione massima da parte di più di metà dei giudici possiede un certo grado di validità di contenuto
b. maggiore è la proporzione di giudici oltre il 50% che assegnano valutazione massima all'item, maggiore è la validità di contenuto dell'item

$$CVR = \frac{n_{\max} - \frac{N}{2}}{\frac{N}{2}}$$

N. giudici	*CVR critico*	*N. giudici*	*CVR critico*
5	,99	13	,54
6	,99	14	,51
7	,99	15	,49
8	,75	20	,42
9	,78	25	,37
10	,62	30	,33
11	,59	35	,31
12	,56	40	,29

Copyright © 2011 The McGraw-Hill Companies srl

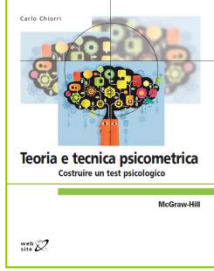


Esempio - Ortoressia

DEFINIZIONE DEL COSTRUTTO

- L' ortoressia nervosa (dal greco "orto" che significa sano e "orexis" che significa appetito) è un disturbo psicopatologico caratterizzato dalla maniacale ossessione per i cibi sani e "puri" (generalmente vegetali crudi, cereali e cibi macrobiotici e comunque alimenti privi di pesticidi, conservanti ed altre sostanze artificiali), enfatizzato dagli eccessivi allarmismi creati dai mass media intorno al cibo (mucca pazza, pesce al mercurio, influenza aviaria, cibi transgenici...).

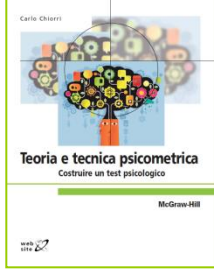
Copyright © 2011 The McGraw-Hill Companies srl



DEFINIZIONE DELLE SFACCETTATURE

1. *Paura della contaminazione/malattie*: questa scala deve permettere la valutazione della fobia relativa all'essere contaminati e al contrarre malattie attraverso il cibo. Soggetti con punteggi alti in questa scala sono persone con tratti ipocondriaci, preoccupati per la purezza e la qualità del cibo, ossessionati dal timore di riempire il proprio corpo di sostanze non naturali e potenzialmente tossiche, inconsapevoli dei rischi per la salute che il proprio regime alimentare comporta.
2. *Rapporto con gli altri*: questa scala deve permettere la valutazione delle modalità di interazione con gli altri, con particolare attenzione alle situazioni sociali collegate al cibo. Soggetti con punteggi alti in questa scala tendono all'isolamento e all'evitamento di situazioni sociali, a criticare lo stile alimentare altrui e a sentirsi incompresi per il proprio. (Esempio di item: Penso che la maggior parte dei miei amici mangi scorrettamente).
3. *Sfera emotiva*: questa scala deve permettere la valutazione delle emozioni e dei sentimenti relativi al cibo. Soggetti con punteggi alti in questa scala tendono ad utilizzare il cibo come meccanismo di coping e a regolare l'autostima in base all'adesione rigida a regole che se trasgredite producono senso di colpa
4. *Alimentazione limitata e particolare*: questa scala deve permettere la valutazione degli atteggiamenti, delle modalità e dei comportamenti relativamente alla ricerca, alla preparazione e al consumo del cibo. Soggetti con punteggi alti in questa scala sono persone con tratti ossessivi, rigidamente dipendenti dai propri rituali tramite l'adesione ai quali soddisfano il proprio bisogno di controllo.
5. *Paura di ingrassare* questa scala deve permettere la valutazione della paura di ingrassare e il desiderio di dimagrire che spesso si celano dietro la scelta di un regime alimentare salutista. Soggetti con punteggi alti in questa scala avranno un forte desiderio di essere magri e di perdere peso.

Copyright © 2011 The McGraw-Hill Companies srl



GRIGLIA DI VALUTAZIONE

Indichi per favore a quale sottoscala appartiene secondo Lei ogni item e la sua rilevanza per il costrutto di indagine, tenendo presente che:

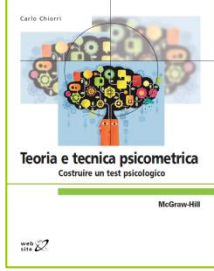
1 = Poco rilevante

2 = Abbastanza rilevante

3 = Molto rilevante

		Rilevanza			Sottoscala
1	Disinfetto frutta e verdura prima di mangiarle	1	2	3	
2	Preferisco pranzare/cenare a casa piuttosto che fuori per essere sicuro di ciò che mangio	1	2	3	
3	Se sgarro rispetto al mio regime alimentare sano la mia autostima diminuisce	1	2	3	
[…]	[…]	[…]	[…]	[…]	

Copyright © 2011 The McGraw-Hill Companies srl



Validità di facciata

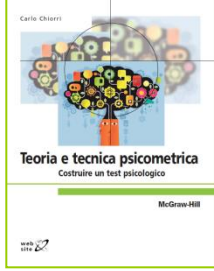
Content validity should not be confused with face validity. The latter is not validity in the technical sense; it refers not to what the test actually measures, but **to what it appears superficially to measure**. **Face validity pertains to whether the test 'looks valid'** to the examinees who take it, the administrative personnel who decide on its use, and other technically untrained observers

(Anastasi, 1988, p. 144)

- **Grado in cui gli item di un test *sembrano* misurare il costrutto che intendono misurare**

- Chi lo valuta?

Copyright © 2011 The McGraw-Hill Companies srl

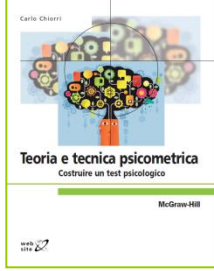


Validità di facciata

Un [giudice / valutatore / osservatore] che è un [potenziale soggetto / utilizzatore non professionale del test / membro dell'opinione pubblica] valuta [un item di un test / un test / una batteria di test] rispetto [alla sua rilevanza / alla sua adeguatezza / alla sua comprensibilità / etc.] mediante una tecnica [assoluta / relativa]

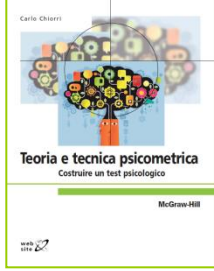
- Deve essere eseguita per tutte le parti del test
- **Tecniche**
  - **riformulazione**: parafrasi
  - **thinking aloud**: pensare ad alta voce
  - **intervista cognitiva**: analisi dei processi di pensiero che portano alla risposta
  - **intervista in profondità**: compromesso fra il colloquio clinico e l'intervista direttiva → utile per analisi più approfondite

Copyright © 2011 The McGraw-Hill Companies srl



Studio preliminare

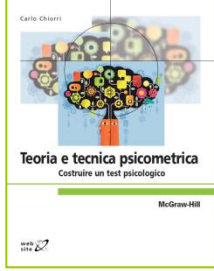
Copyright © 2011 The McGraw-Hill Companies srl



Raccogliere i dati

- Decidere la popolazione target
- Rappresentatività
- Quanti soggetti?
- Criteri di inclusione ed esclusione

Copyright © 2011 The McGraw-Hill Companies srl



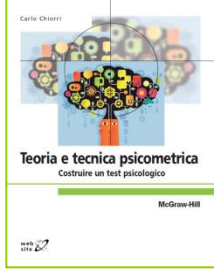
Criteri inclusione ed esclusione

Esempio per pazienti con disturbo della personalità

- Inclusione
  - Diagnosi da *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*
  - Remissione sintomi acuti Asse I

- Esclusione
  - QI $\leq$ 75
  - Titolo di studio inferiore alla licenza elementare
  - Diagnosi di Schizofrenia, Disturbo Schizoaffettivo, Schizofreniforme o Delirante
  - Demenza
  - Disturbo Mentale Organico

Copyright © 2011 The McGraw-Hill Companies srl



Scheda socio-demografica

Scheda socio-demografica

1. Sesso: ☐[0] Femmina ☐[1] Maschio

2. Età: _____

3. Titolo di studio

☐[0] Nessuno
☐[1] Licenza Elementare
☐[2] Licenza media
☐[3] Diploma Scuola Media Superiore
☐[4] Laurea Triennale/Diploma Universitario
☐[5] Laurea Specialistica/Magistrale/Vecchio Ordinamento
☐[6] Dottorato di ricerca/Master/Specializzazione

4. Stato civile

☐[1] Celibe/Nubile
☐[2] Convivente
☐[3] Coniugato
☐[4] Separato
☐[5] Divorziato
☐[6] Vedovo/a

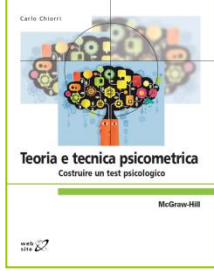
5. Professione

☐[1] Disoccupato/a
☐[2] Impiegato/a
☐[3] Libero/a Professionista/Imprenditore*
☐[4] Dirigente, Docente Universitario **
☐[5] Pensionato/a
☐[6] Studente/Studentessa
☐[7] Operaio industriale o agricolo
☐[8] Commerciante
☐[9] Casalinga
☐[10] Insegnante
☐[11] Artista
☐[12] Artigiano/a
☐[13] Commesso/a
☐[14] Infermiere/a
☐[15] Educatore/Educatrice
☐[16] Altro (specificare) _____

* (Avvocato, Notaio, Medico, Ingegnere, Architetto ecc., con studio proprio)
** (Magistrato, Medico, Ingegnere, Architetto, Ufficiale Forze Armate ecc., in istituzione o società)

6. Hai mai ricevuto una diagnosi psichiatrica? ☐[0] No ☐[1] Sì Specificare _____

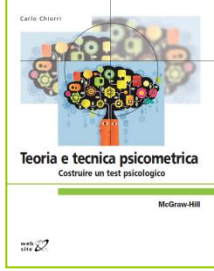
Copyright © 2011 The McGraw-Hill Companies srl



Analisi degli item

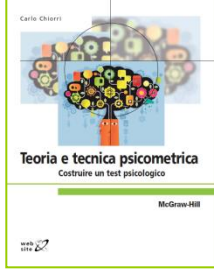
- Analisi statistica per valutare appropriatezza psicometrica degli item
- Non deve essere l'unico criterio per decidere se mantenere o meno l'item nell'item pool

Copyright © 2011 The McGraw-Hill Companies srl



Analisi degli item test di prestazione tipica

Copyright © 2011 The McGraw-Hill Companies srl

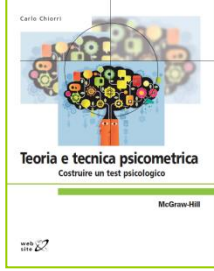


Numero di casi mancanti

- Differenza di significato della risposta omessa test di prestazione massima vs test di prestazione tipica

- Cause
  - Il soggetto rifiuta di rispondere, per cui intenzionalmente non ha indicato la sua risposta
  - Il soggetto non sa cosa rispondere e, nel dubbio, non indica alcuna risposta
  - Il soggetto ha semplicemente saltato l'item, per sbaglio
  - Il dato non è stata inserito per errore nel database

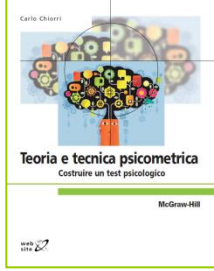
Copyright © 2011 The McGraw-Hill Companies srl



Tipi di missing

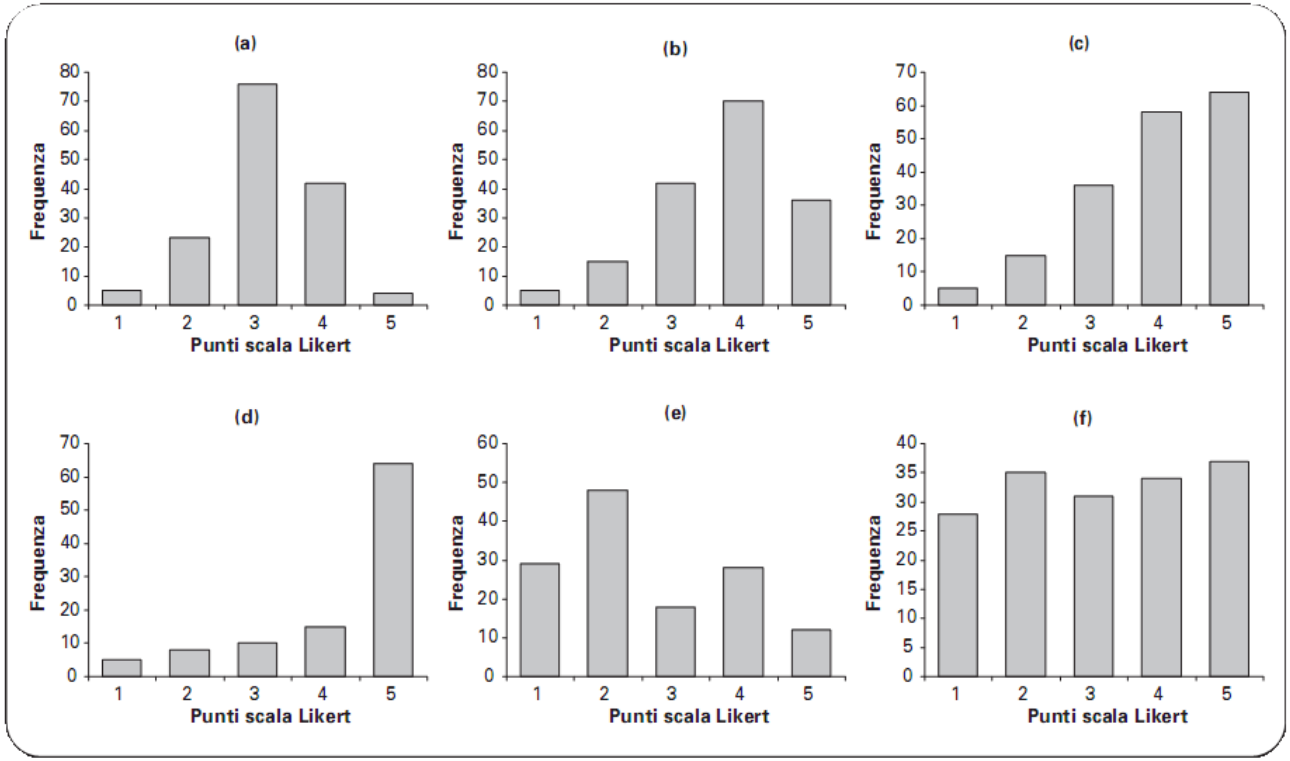
- Missing per definizione della sottopopolazione
- Missing completely at random
  - La probabilità che un dato sia mancante è completamente indipendente sia dalla caratteristica misurata dal test di cui l'item fa parte, sia da qualsiasi altra variabile considerata nell'analisi
  - Little's MCAR Test
- Missing at random
  - La probabilità che un dato sia mancante non dipende dalla caratteristica misurata dal test di cui l'item fa parte una volta controllato l'effetto di un altre variabili, dato che queste altre variabili innescano il meccanismo che determina i dati mancanti
- Missing not at random

Copyright © 2011 The McGraw-Hill Companies srl

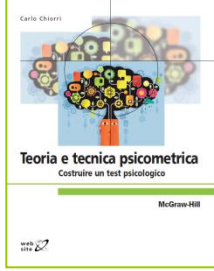


Forma della distribuzione

- Perché la distribuzione normale?



Copyright © 2011 The McGraw-Hill Companies srl

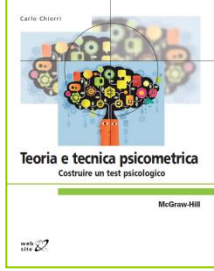


Punteggio minimo e punteggio massimo

- Adeguatezza scala di risposta di tipo Likert

- Cause
  - Item mal formulati
  - Scale di tipo Likert non adeguate per l'item (o viceversa)
  - Campioni troppo omogenei

Copyright © 2011 The McGraw-Hill Companies srl

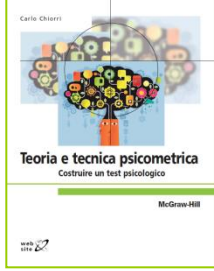


Indici di tendenza centrale e dispersione

- Valori indicativi (nessun supporto empirico)
- Possibile utilizzo indici ordinali (quartili)

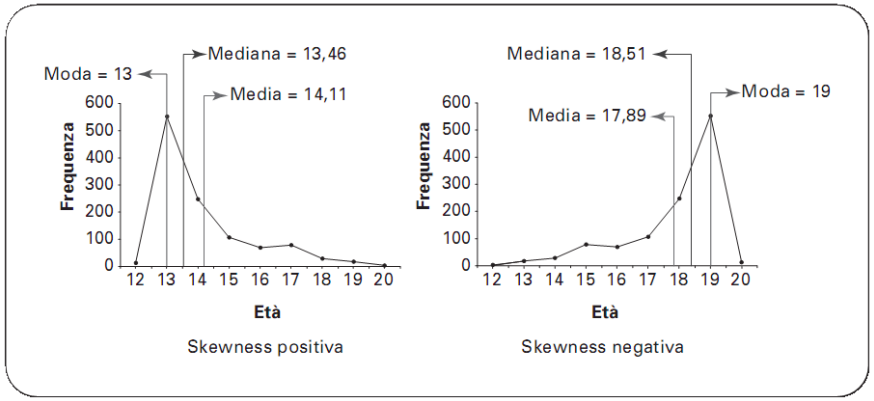
Punti della scala	Range media	Punti della scala	Range media	DS
1-2-3-4	1,75-3,25	0-1-2-3	0,75-2,25	0,6
1-2-3-4-5	2,00-4,00	0-1-2-3-4	1,00-3,00	0,8
1-2-3-4-5-6	2,25-4,75	0-1-2-3-4-5	1,25-3,75	1,0
1-2-3-4-5-6-7	2,50-5,50	0-1-2-3-4-5-6	1,50-4,50	1,2
1-2-3-4-5-6-7-8	2,75-6,25	0-1-2-3-4-5-6-7	1,75-5,25	1,4
1-2-3-4-5-6-7-8-9	3,00-7,00	0-1-2-3-4-5-6-7-8	2,00-6,00	1,6

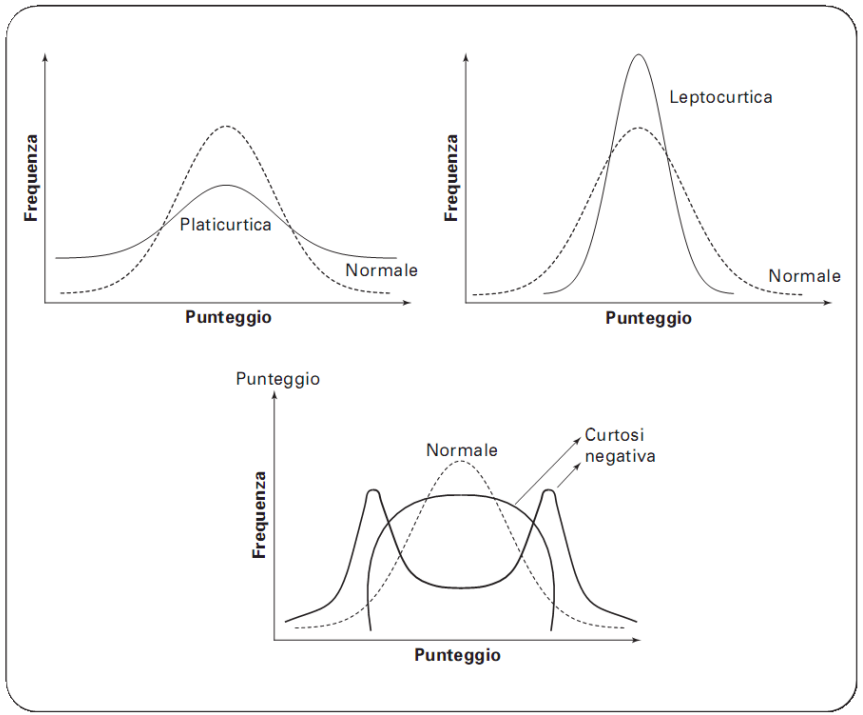
Copyright © 2011 The McGraw-Hill Companies srl



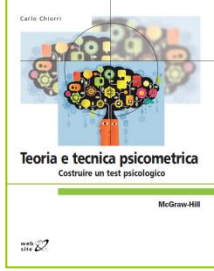
Indici di forma della distribuzione

- Skewness e curtosi
- Muthén e Kaplan (1985) e l'intervallo [-1;+1]





Copyright © 2011 The McGraw-Hill Companies srl



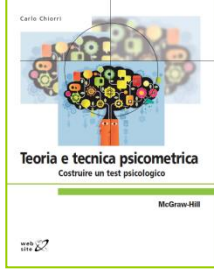
Discriminatività

- Confronto fra gruppi precostituiti (test orientati al criterio)
- Confronto fra "livelli alti" e "livelli bassi" al test (test orientati al costrutto)

- Test *t* per campioni indipendenti (info metrica)
- Test di Mann-Whitney (info ordinale)
- Dimensione dell'effetto almeno "moderata"

d	*r*	Effect size
< |0,20|	< |0,10|	Trascurabile
|0,20| - |0,50|	|0,10| - |0,30|	Piccola
|0,50| - |0,80|	|0,30| - |0,50|	Moderata
>|0,80|	>|0,50|	Grande

Copyright © 2011 The McGraw-Hill Companies srl



Ridondanza

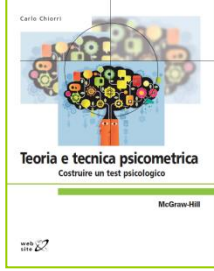
- Correlazioni maggiori di |,70|

Inter-Item Correlation Matrix

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
ITEM1	1.000	.716	.713	.694	.691	.694
ITEM2	.716	1.000	.679	.653	.697	.698
ITEM3	.713	.679	1.000	.595	.685	.749
ITEM4	.694	.653	.595	1.000	.595	.665
ITEM5	.691	.697	.685	.595	1.000	.700
ITEM6	.694	.698	.749	.665	.700	1.000

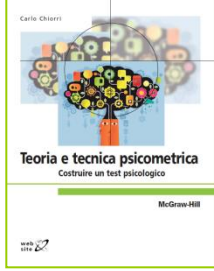
The covariance matrix is calculated and used in the analysis.

Copyright © 2011 The McGraw-Hill Companies srl



Attendibilità

Copyright © 2011 The McGraw-Hill Companies srl



Attendibilità (*reliability*)

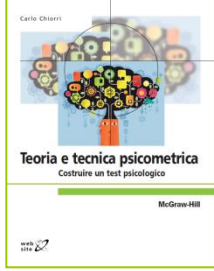
- **Proprietà psicometrica relativa all'accuratezza con cui un test o una scala misura una certa variabile psicologica**

- Perché è diversa dalla validità?

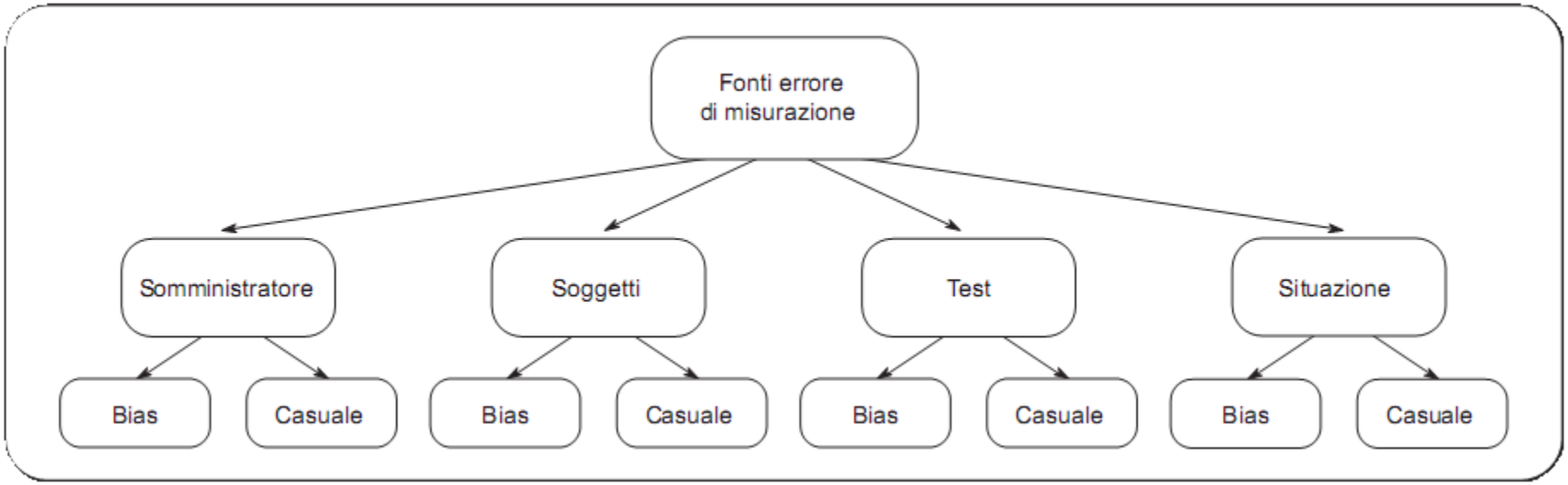
- Proporzione del punteggio misurato che non riflette l'errore di misurazione

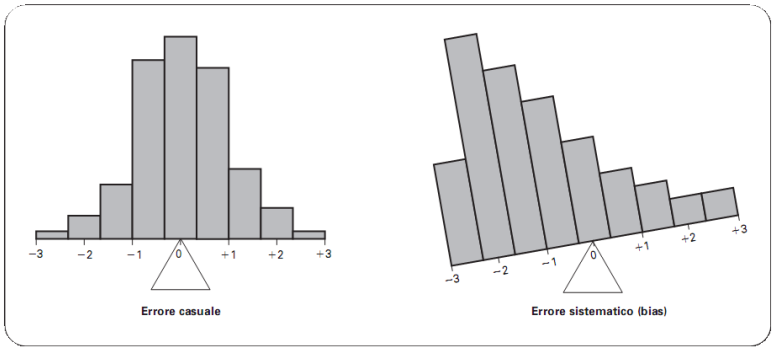
$$X = V + E$$

Copyright © 2011 The McGraw-Hill Companies srl

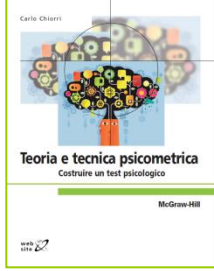


Fonti di errore nei test psicologici



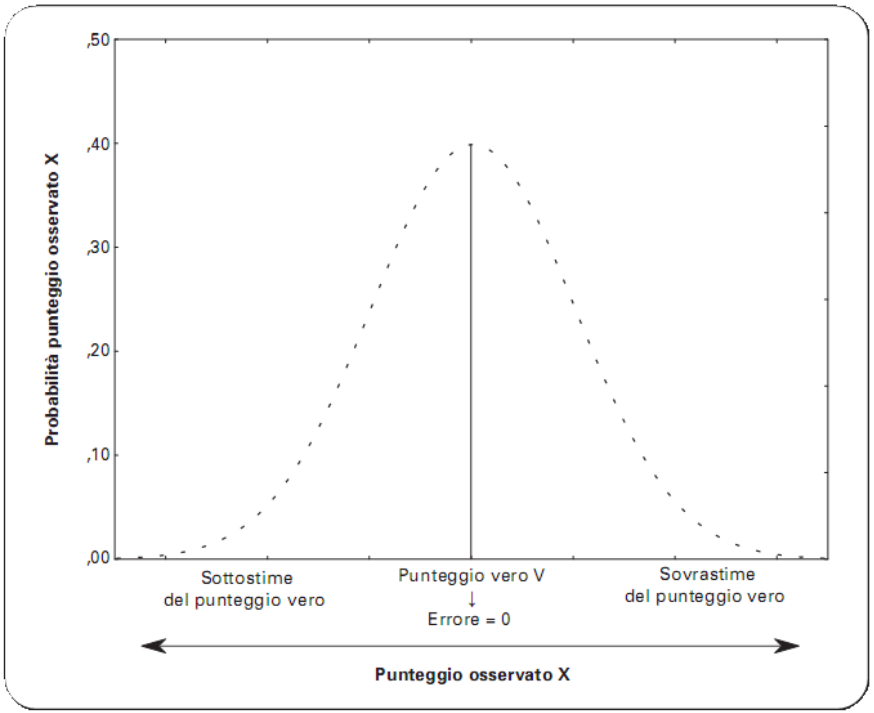


Copyright © 2011 The McGraw-Hill Companies srl

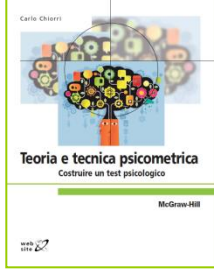


Errore casuale nella TCT

- Variabile aleatoria o stocastica



Copyright © 2011 The McGraw-Hill Companies srl



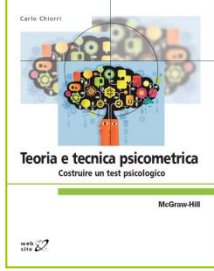
Errore casuale nella TCT

- Punteggio osservato in un dato soggetto in una data situazione è solo uno dei possibili punteggi osservabili

- Il punteggio vero *V*, quindi, può essere concepito come la **media dei punteggi osservati di infinite misurazioni**

- $E(X) = V$

Copyright © 2011 The McGraw-Hill Companies srl



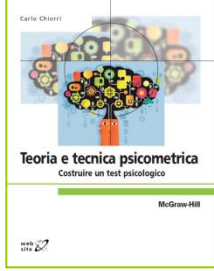
Assunzioni sull'errore di misurazione

1. L'errore di misurazione è casuale, è distribuito normalmente e ha valore atteso (media) uguale a zero: $E(E) = 0$.
   - Avremo quindi che $E(X) = E(V + E) = E(V) + E(E) = E(V) + 0 = E(V)$. Dato che V è costante, $E(X) = E(V) = V$

2. Punteggio vero ed errore di una misurazione sono indipendenti fra loro (= covarianza fra V ed E è uguale a zero)
   - L'entità dell'errore non dipende dalla quantità di punteggio vero, e quindi che l'errore di misurazione non dipende dalla quantità di costrutto posseduta dal soggetto

$$Cov(VE) = \frac{\sum (V - M_V)(E - M_E)}{n}$$

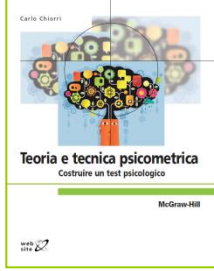
Copyright © 2011 The McGraw-Hill Companies srl



Assunzioni sull'errore di misurazione

3. Gli errori di misurazione di due misurazioni sugli stessi soggetti sono indipendenti fra loro (= la covarianza tra due distribuzioni di errori in due misurazioni diverse è uguale a zero)
   - L'errore di misurazione che distorce il punteggio in un item non è in relazione con quello che distorce il punteggio in un altro item

Copyright © 2011 The McGraw-Hill Companies srl



Stimare l'attendibilità

$$r_{tt} = \frac{V}{V+E} = \frac{V}{X}$$

$$s_X^2 = \frac{\sum (X - M_X)^2}{n} = \frac{\sum [(V+E) - (M_V + M_E)]^2}{n} = \frac{\sum [(V - M_V) + (E - M_E)]^2}{n} =$$

$$= \frac{\sum (V - M_V)^2}{n} + \frac{\sum (E - M_E)^2}{n} + \frac{2\sum [(V - M_V)(E - M_E)]}{n}$$

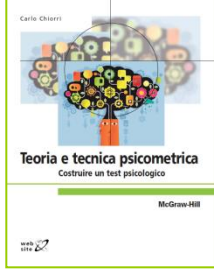
Varianza punteggio vero s^2_V

Varianza errore s^2_E

Covarianza errore-punteggio vero (= 0!)

$$s_X^2 = s_V^2 + s_E^2$$

Copyright © 2011 The McGraw-Hill Companies srl



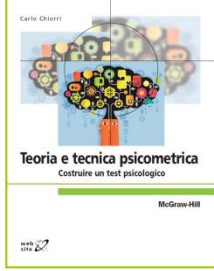
Test-retest

> The only meaningful estimate of reliability for a hetereogeneous test is the the retest variety
>
> (Guilford, 1965, p. 450)

- Stabilità dei punteggi nel tempo (→solo tratti)
- Due o più somministrazioni dello stesso test agli stessi soggetti

- Valore accettabile: ,70
- Valore ottimale: ,90

Copyright © 2011 The McGraw-Hill Companies srl

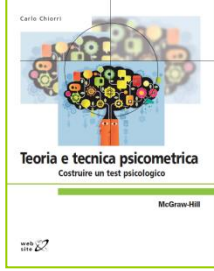


Quanto tempo?

- **Coefficiente di replicabilità** (dependability) = poche ore/giorni
- **Coefficiente di stabilità** (varie settimane/mesi)

- Solitamente 2-4 settimane

Copyright © 2011 The McGraw-Hill Companies srl



Esempio

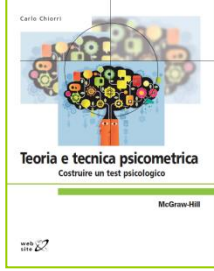
- Punteggi di 8 soggetti ad un test di estroversione somministrato due volte a distanza di tre settimane

	H	K
s01	55	51
s02	61	60
s03	43	38
s04	58	53
s05	38	37
s06	60	66
s07	48	46
s08	36	50

Correlazione = Attendibilità = ,782
Variazione: $t(7) = -0{,}11$, $p = .917$

Perché qui calcoliamo anche questo valore?

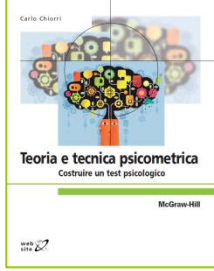
Copyright © 2011 The McGraw-Hill Companies srl



Fonti di errore due somministrazioni

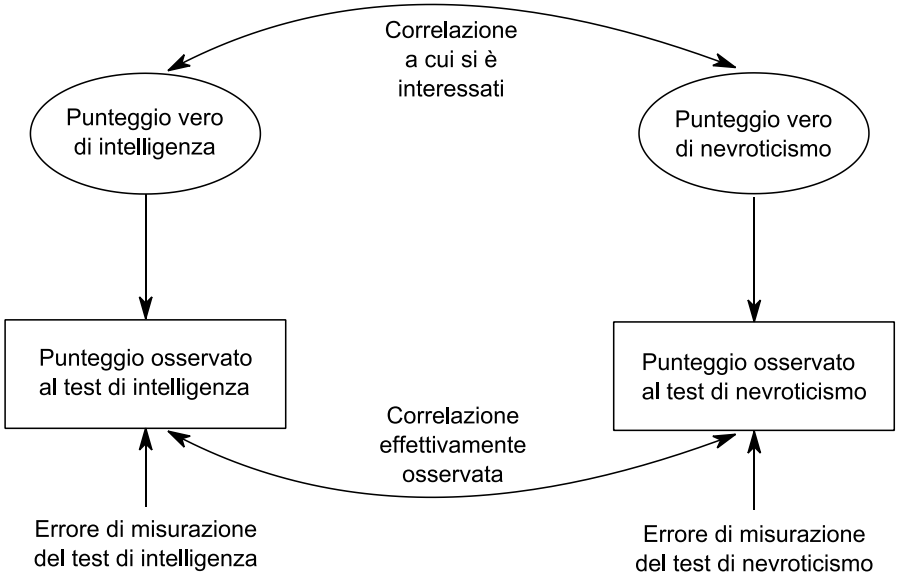
- Ricordo risposte fornite, strategie utilizzate, etc.
- Fluttuazioni nel livello del costrutto
- Variazione condizioni di somministrazione
- Interazione soggetto × momento somministrazione

Copyright © 2011 The McGraw-Hill Companies srl



Coerenza interna (*internal consistency*)

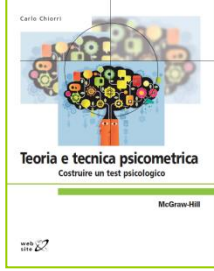
- Grado in cui gli item di un test o di una scala sono coerenti nel misurare il costrutto in oggetto
- Problema dell'attenuazione



Correzione per attenuazione

$$r_{V(X)V(Y)} = \frac{r_{XY}}{\sqrt{r_{ttX}\, r_{ttY}}}$$

Copyright © 2011 The McGraw-Hill Companies srl



Split-half

- Divisione a metà del test
  - Pari/dispari
  - Prima/Seconda metà
  - Etc.
- Correlazione fra i punteggi nelle due metà → non è attendibilità di un test con *k* item, ma di due test paralleli di *k*/2 item!
- Profezia di Spearman-Brown

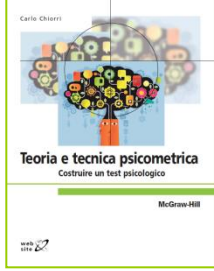
$$r_{ntt} = \frac{\left(\frac{k_{Modificata}}{k_{Esistente}}\right) r_{tt}}{1 + \left(\frac{k_{Modificata}}{k_{Esistente}} - 1\right) r_{tt}} \quad \Rightarrow \quad r_{SH} = \frac{2 r_{tt}}{1 + r_{tt}}$$

- Se due metà non parallele → coefficiente di Rulon

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

s^2_D è la varianza della differenza fra i punteggi nelle due metà

Copyright © 2011 The McGraw-Hill Companies srl



Esempio

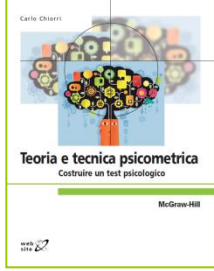
- Se abbiamo un test con 15 item e attendibilità ,85, se aggiungiamo 5 item otterremo un'attendibilità di:

$$r_{ntt} = \frac{\left(\frac{20}{15}\right),85}{1+\left(\frac{20}{15}-1\right),85} =,88$$

- Se togliamo 5 item l'attendibilità diventa:

$$r_{ntt} = \frac{\left(\frac{10}{15}\right),85}{1+\left(\frac{10}{15}-1\right),85} =,79$$

Copyright © 2011 The McGraw-Hill Companies srl



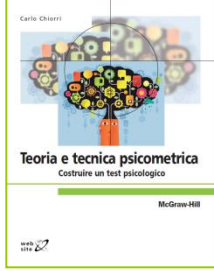
Quanti item aggiungere?

$$\text{numero di item paralleli da aggiungere} = \frac{(k_{Esistente}) \times r_{ntt}(1 - r_{tt})}{r_{tt}(1 - r_{ntt})} - k_{Esistente}$$

- Se abbiamo un test con 10 item e attendibilità ,75, e vogliamo ottenere un test con attendibilità ,80, dobbiamo aggiungere:

$$\frac{(10) \times ,80(1 - ,75)}{,75(1 - ,80)} - 10 = 3,33$$

Copyright © 2011 The McGraw-Hill Companies srl



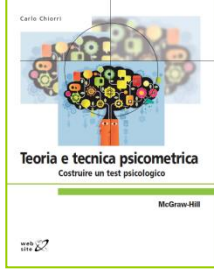
Esempio

	Item01	Item02	Item03	Item04	Item05	Item06
s01	4	4	2	4	4	2
s02	4	5	4	4	5	3
s03	2	1	2	2	2	2
s04	3	3	3	2	3	3
s05	4	4	5	4	4	4
s06	3	2	3	2	2	3
s07	1	3	2	3	3	1
s08	4	4	4	2	4	4

	Item01	Item03	Item04	**Totale 1**
s01	4	2	4	**10**
s02	4	4	4	**12**
s03	2	2	2	**6**
s04	3	3	2	**8**
s05	4	5	4	**13**
s06	3	3	2	**8**
s07	1	2	3	**6**
s08	4	4	2	**10**

	Item02	Item05	Item06	**Totale 2**
s01	4	4	2	**10**
s02	5	5	3	**13**
s03	1	2	2	**5**
s04	3	3	3	**9**
s05	4	4	4	**12**
s06	2	2	3	**7**
s07	3	3	1	**7**
s08	4	4	4	**12**

Copyright © 2011 The McGraw-Hill Companies srl



Esempio

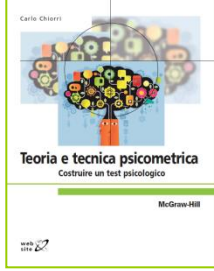
- Correlazione Totale 1 - Totale 2 = ,914

- Attendibilità Split-Half:

$$r_{SH} = \frac{2 \times ,914}{1 + ,914} = ,955$$

- Valore accettabile: ,60 (test prestazione tipica)
- Valore ottimale: ,80 (test prestazione massima)

Copyright © 2011 The McGraw-Hill Companies srl



KR-20 e KR-21

- Problemi con split-half
- Soluzione: indice di attendibilità che siae la media di tutte le possibilità attendibilità split-half

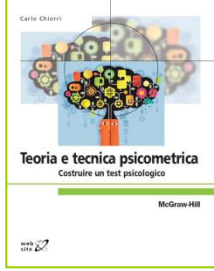
$$KR_{20} = \frac{k}{k-1}\left[1 - \frac{\sum_{j=1}^{k} P_j(1-P_j)}{s_X^2}\right]$$

Se item hanno la stessa difficoltà

$$KR_{21} = \frac{k}{k-1}\left[1 - \frac{kM_P M_{(1-P)}}{s_X^2}\right]$$

- k = numero di item del test, P_j = proporzione di risposte corrette all'item j, s^2_X = varianza del punteggio totale al test, M_P = media della proporzione di risposte corrette e $M_{(1-P)}$ = media della proporzione di risposte errate

Copyright © 2011 The McGraw-Hill Companies srl



Alpha (α) di Cronbach

- Estensione della KR-20 ad item non dicotomici

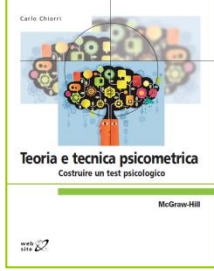
$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum_{j=1}^{k} s_j^2}{s_X^2}\right]$$

- k = numero di item, s^2_j = varianza del punteggio all'item j, s^2_X = varianza del punteggio totale al test

$$\alpha_z = \frac{kM_{rii}}{1 + M_{rii}(k-1)}$$

- k = numero di item, M_{rii} = correlazione media inter-item

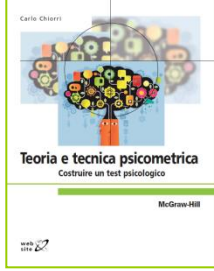
Copyright © 2011 The McGraw-Hill Companies srl



Linee guida

Valore di α o KR-20	*Interpretazione*
$\alpha \geq ,90$	Ottimo
$,80 \leq \alpha < ,90$	Buono
$,70 \leq \alpha < ,80$	Discreto
$,60 \leq \alpha < ,70$	Sufficiente
$\alpha < ,60$	Inadeguato

Copyright © 2011 The McGraw-Hill Companies srl



Esempio - KR-20

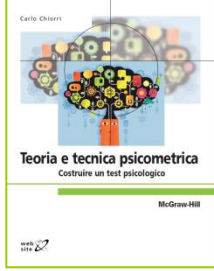
	Item01	Item02	Item03	Item04	Item05	Item06	**Totale**
s01	1	1	1	1	1	1	**6**
s02	1	0	0	0	1	1	**3**
s03	1	1	1	1	1	1	**6**
s04	1	1	1	1	0	1	**5**
s05	0	0	0	1	0	0	**1**
s06	1	1	1	1	0	1	**5**
s07	1	1	1	1	1	1	**6**
s08	1	0	1	1	1	1	**5**

Varianza = 2,734

P	.875	.625	.750	.875	.625	.875
1–P	.125	.375	.250	.125	.375	.125
P(1–P)	.109	.234	.188	.109	.234	.109

$$KR_{20} = \frac{6}{6-1}\left[1 - \frac{,109+,234+,188+,109+,234+,109}{2,734}\right] = ,768$$

Copyright © 2011 The McGraw-Hill Companies srl

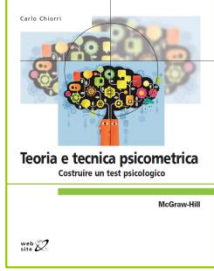


Esempio - α di Cronbach

	Item01	Item02	Item03	Item04	Item05	Item06	**Totale**
s01	4	4	2	4	4	2	**20**
s02	4	5	4	4	5	3	**25**
s03	2	1	2	2	2	2	**11**
s04	3	3	3	2	3	3	**17**
s05	4	4	5	4	4	4	**25**
s06	3	2	3	2	2	3	**15**
s07	1	3	2	3	3	1	**13**
s08	4	4.000	4	2	4	4	**22**
Varianza	1.11	1.44	1.11	0.86	0.98	0.94	**25.00**

$$\alpha = \frac{6}{6-1}\left[1 - \frac{1{,}11 + 1{,}44 + 1{,}11 + 0{,}86 + 0{,}98 + 0{,}94}{25}\right] = {,}891$$

Copyright © 2011 The McGraw-Hill Companies srl

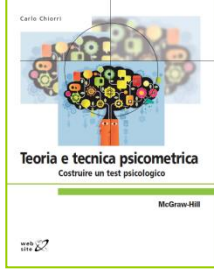


Perché più item = più attendibilità?

- Ogni volta che vengono aggiunti item si modificano la proporzione di varianza totale, quella di varianza vera e quella di errore del test finale

- La varianza del punteggio vero aumenta in progressione quadratica, quella del punteggio osservato in progressione lineare

Copyright © 2011 The McGraw-Hill Companies srl

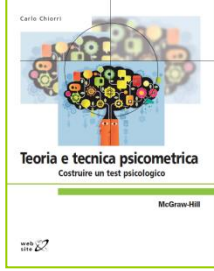


Coerenza interna ≠ omogeneità

- Alpha di Crobach alto = alta coerenza interna ma non necessariamente unidimensionalità
- Costrutti ampi (domini) e specifici (facet)

	Correlazione media inter-item						
	Costrutti ampi			Costrutti specifici			
n	,10	,20	,30	,40	,50	,60	,70
5	,36	,56	,68	,77	,83	,88	,92
10	,53	,71	,81	,87	,91	,94	,96
15	,63	,79	,87	,91	,94	,96	,97
20	,69	,83	,90	,93	,95	,97	,98
25	,74	,86	,91	,94	,96	,97	,98
30	,77	,88	,93	,95	,97	,98	,99
35	,80	,90	,94	,96	,97	,98	,99
40	,82	,91	,94	,96	,98	,98	,99

Copyright © 2011 The McGraw-Hill Companies srl



Coerenza interna ≠ omogeneità

(a)

	i01	i02	i03	i04	i05	i06
i01	-					
i02	,77	-				
i03	,81	,82	-			
i04	,25	,28	,30	-		
i05	,23	,32	,31	,78	-	
i06	,35	,37	,29	,81	,81	-

(b)

	i01	i02	i03	i04	i05	i06
i01	–					
i02	,48	–				
i03	,50	,51	–			
i04	,52	,50	,50	–		
i05	,47	,49	,50	,53	–	
i06	,53	,50	,47	,50	,50	–

Errore standard di alpha (Cortina, 1993)

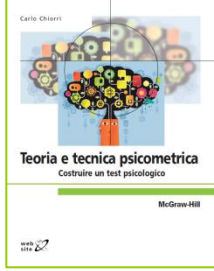
,0661

$$\alpha_{SE} = \frac{s_{rii}}{\sqrt{0{,}5k(k-1)-1}}$$

,0047

s_{rii} = deviazione standard dei coefficienti di correlazione della matrice, k = numero di item

Copyright © 2011 The McGraw-Hill Companies srl



Caratteristiche di alpha

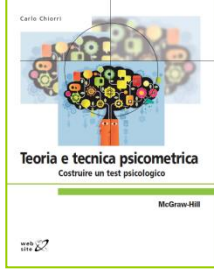
- La varianza del punteggio totale, ossia della somma degli item, è uguale alla somma delle varianze dei singoli item più le covarianze tra tutte le possibili coppie di item diversi

$$s_x^2 = \underbrace{\boxed{s_{i1}^2 + s_{i2}^2 + s_{i3}^2 + [\ldots] + s_{ik}^2}}_{\sum s_j^2} + \underbrace{\boxed{s_{i1i2} + s_{i1i3} + s_{i2i3} + [\ldots] + s_{i1ik} + s_{i2ik} + s_{i3ik}}}_{\sum s_{ij}}$$

- Mantenendo costante la varianza di ogni singolo item, più gli item sono correlati fra loro, maggiore sarà la varianza del punteggio totale rispetto alla somma delle varianze dei singoli item

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum s_j^2}{s_X^2}\right] = \frac{k}{k-1}\left[1 - \frac{\sum s_j^2}{\sum s_j^2 + \sum s_{ij}}\right]$$

Copyright © 2011 The McGraw-Hill Companies srl

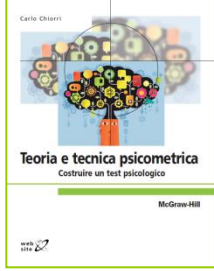


Correlazione multipla al quadrato

- Non è un indice specifico delle analisi psicometriche
- Proporzione di varianza condivisa fra una variabile con un insieme di altre variabili
- Valore accettabile: ,10

$$R^2_{X.YZ} = \frac{r^2_{XY} + r^2_{XZ} - 2r_{XY}r_{XZ}r_{YZ}}{1 - r^2_{YZ}}$$

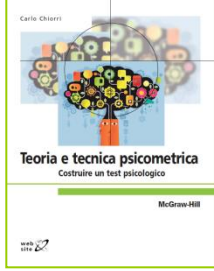
Copyright © 2011 The McGraw-Hill Companies srl



Alpha senza l'item

- Quale sarebbe il valore di α se l'item venisse escluso dal test?
- Se item tutti necessari, eliminazione = peggioramento α
- Variazione minima sostanziale: +,01

Copyright © 2011 The McGraw-Hill Companies srl

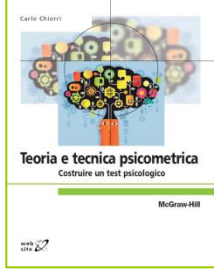


Conclusioni

- Coefficienti solitamente riportati: α di Cronbach e test-retest

- Evoluzioni: Teoria della generalizzabilità

Copyright © 2011 The McGraw-Hill Companies srl



Esempio

Test di prestazione massima

1. Quale tra i seguenti autori è uno dei massimi esponenti della psicoanalisi?
   a. Carl Rogers
   b. Burrhus Frederic Skinner
   c. Carl Gustav Jung *

2. Quale delle seguenti affermazioni non è corretta?
   a. L'etnia si riferisce ad una classificazione dell'uomo basata unicamente su tratti fisici e genetici *
   b. L'etnia può coincidere con una nazione quando l'identità nazionale è definita soprattutto da una origine comune
   c. L'etnia si basa sulla storia comune di una determinata popolazione

3. Jean Piaget ha studiato principalmente:
   a. Lo sviluppo cognitivo sui bambini *
   b. I meccanismi di difesa
   c. I riflessi condizionati

4. Il termine "comportamento prosociale" corrisponde a:
   a. disponibilità
   b. empatia
   c. altruismo *

5. Gianni Rodari è stato:
   a. Un importante scrittore per l'infanzia della seconda metà del '900 *
   b. Uno storico della scuola e dell'educazione
   c. Un pedagogista della prima metà del '900

6. Chi ha scritto Il Giornalino di Gian Burrasca?
   a. Italo Calvino
   b. Luigi Bertelli (Vamba)*
   c. Ignazio Silone

7. Nel sistema scolastico italiano attuale, la scuola secondaria inferiore ha durata:
   a. Quinquennale
   b. Biennale
   c. Triennale *

8. Nietzsche ritiene che la cultura occidentale:
   a. sia l'unica al mondo densa di valori positivi
   b. sia costituita dallo sviluppo dello spirito dionisiaco della tragedia greca
   c. si sia negativamente sviluppata tramite l'esaltazione dello spirito apollineo e la negazione di quello dionisiaco, entrambi presenti nella tragedia greca *

9. Cartesio concepisce il cogito come:
   a. prima regola del metodo conoscitivo
   b. fondamento di ogni conoscenza certa e criterio di verità *
   c. pensiero dell'essere

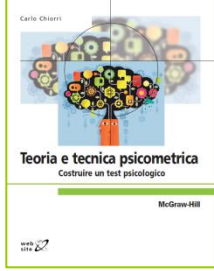
10. Qual è il nome della scuola materna, previsto dalla legge 53/2003?
    a. Scuola primaria
    b. Asilo
    c. Scuola dell'infanzia *

11. Rispetto ad Aristotele, Tommaso d'Aquino ritiene che l'ente:
    a. sia costituito dall'essenza e dall'atto d'essere *
    b. sia costituito da essenza ed esistenza
    c. sia costituito da materia e forma

12. Con il termine "persuasione" si intende:
    a. Il processo comunicativo che consente lo sviluppo intellettivo dell'individuo
    b. Il processo che consente di disorientare l'interlocutore
    c. Il processo che mediante atti di comunicazione porta alla formazione, al rafforzamento o alla modificazione di atteggiamenti *

13. Qual è il ruolo dell'identità nell'educazione interculturale?
    a. Ruolo statico, ma aperto alla tolleranza de "l'altro"
    b. Ruolo dinamico nel confronto con "l'altro" *
    c. Ruolo da occultare per accogliere "l'altro"

Copyright © 2011 The McGraw-Hill Companies srl

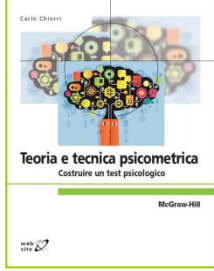


Difficoltà (*n* = 328)

	P(corretta)	Difficoltà [P(errata)]
d01	.430	.570
d02	.445	.555
d03	.720	.280
d04	.674	.326
d05	.747	.253
d06	.412	.588
d07	.866	.134
d08	.564	.436
d09	.162	**.838**
d10	.756	.244
d11	.159	**.841**
d12	.659	.341
d13	.567	.433

Formato dell'item	Difficoltà media ideale
Due alternative	,75-,85
Tre alternative	,66-,77
Quattro alternative	,62-,74
Cinque alternative	,60-,70
Completamento/risposta breve	,45-,55

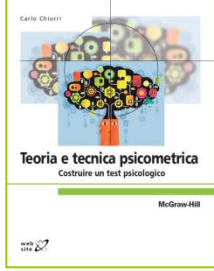
Copyright © 2011 The McGraw-Hill Companies srl



Distribuzione dei distrattori

	a	b	c
d01	.296	.274	**.430**
d02	**.445**	.310	.245
d03	**.720**	.004	.276
d04	.040	.286	**.674**
d05	**.747**	.166	.087
d06	.387	**.412**	.201
d07	.002	.132	**.866**
d08	.220	.216	**.564**
d09	.420	**.162**	.418
d10	.057	.187	**.756**
d11	**.159**	.651	.190
d12	.111	.230	**.659**
d13	.221	**.567**	.212

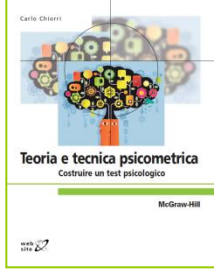
Copyright © 2011 The McGraw-Hill Companies srl



Discriminatività

				Significatività		Discriminatività		Effect size		Indice difficoltà Ebel	
Item	P (tot)	P(Q4)	P(Q1)	z	p	D		h		Id	
d01	.48	.75	.15	7.93	<.001	.60	Ottima	1.30	Alto	.48	Ottimale
d02	.47	.70	.18	6.93	<.001	.52	Ottima	1.12	Alto	.47	Ottimale
d03	.71	.89	.49	5.72	<.001	.39	Buona	0.90	Alto	.71	Ottimale
d04	.66	.85	.44	5.62	<.001	.40	Ottima	0.88	Alto	.66	Ottimale
d05	.72	.95	.44	7.44	<.001	.51	Ottima	1.23	Alto	.72	Ottimale
d06	.45	.70	.15	7.28	<.001	.55	Ottima	1.18	Alto	.45	Ottimale
d07	.86	.94	.76	3.38	<.001	.18	Insufficiente	0.52	Moderato	.86	Bassa
d08	.56	.86	.19	8.84	<.001	.67	Ottima	1.46	Alto	.56	Ottimale
d09	.20	.34	.03	5.21	<.001	.31	Buona	0.93	Alto	.20	Alta
d10	.78	.98	.53	7.11	<.001	.45	Ottima	1.22	Alto	.78	Bassa
d11	.19	.30	.06	3.94	<.001	.24	Sufficiente	0.65	Moderato	.19	Alta
d12	.67	.91	.38	7.40	<.001	.53	Ottima	1.19	Alto	.67	Ottimale
d13	.57	.87	.20	8.84	<.001	.66	Ottima	1.46	Alto	.57	Ottimale

Copyright © 2011 The McGraw-Hill Companies srl



Correlazioni e ridondanze

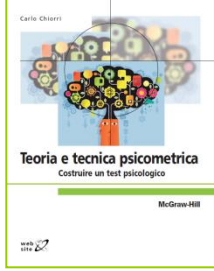
Inter-Item Correlation Matrix

	d01	d02	d03	d04	d05	d06	d07	d08	d09	d10	d11	d12	d13
d01	1.000	.102	.021	.131	.151	.150	.089	.205	.104	.049	.095	.184	.175
d02	.102	1.000	.191	.061	.154	.136	-.025	.119	.057	.123	.115	.063	.052
d03	.021	.191	1.000	.101	.074	.026	.073	.040	-.058	.135	-.026	.194	.002
d04	.131	.061	.101	1.000	.193	.080	-.064	-.022	.076	.105	.035	.034	.048
d05	.151	.154	.074	.193	1.000	.173	.018	.153	.103	.143	.003	.054	.185
d06	.150	.136	.026	.080	.173	1.000	.038	.236	.155	.028	.078	.106	.156
d07	.089	-.025	.073	-.064	.018	.038	1.000	.123	.003	.089	-.099	-.019	.089
d08	.205	.119	.040	-.022	.153	.236	.123	1.000	.236	.045	-.022	.158	.311
d09	.104	.057	-.058	.076	.103	.155	.003	.236	1.000	.095	.059	.054	.267
d10	.049	.123	.135	.105	.143	.028	.089	.045	.095	1.000	.149	.115	.106
d11	.095	.115	-.026	.035	.003	.078	-.099	-.022	.059	.149	1.000	.154	.042
d12	.184	.063	.194	.034	.054	.106	-.019	.158	.054	.115	.154	1.000	.136
d13	.175	.052	.002	.048	.185	.156	.089	.311	.267	.106	.042	.136	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.093	-.099	.311	.410	-3.159	.006	13

Copyright © 2011 The McGraw-Hill Companies srl



Correlazione item-totale corretta KR-20 senza l'item

Reliability Statistics

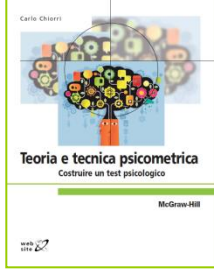
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.583	.572	13

Item-Total Statistics

	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
d01	.302	.548
d02	.234	.563
d03	.159	.578
d04	.157	.579
d05	.297	.551
d06	.282	.553
d07	.068	.589
d08	.331	.541
d09	.241	.563
d10	.231	.564
d11	.123	.582
d12	.255	.559
d13	.322	.544

?

Copyright © 2011 The McGraw-Hill Companies srl



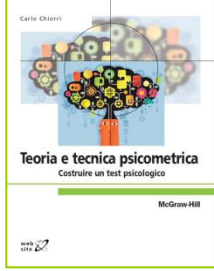
Esempio

Test di prestazione tipica

Indica quanto ti senti descritto da ognuno dei seguenti aggettivi

	Per niente	Poco	Mediamente	Abbastanza	Perfettamente
Loquace	1	2	3	4	5
Riservato	1	2	3	4	5
Pieno di energia	1	2	3	4	5
Scatenato	1	2	3	4	5
Taciturno	1	2	3	4	5
Assertivo	1	2	3	4	5
Timido	1	2	3	4	5
Socievole	1	2	3	4	5

Copyright © 2011 The McGraw-Hill Companies srl

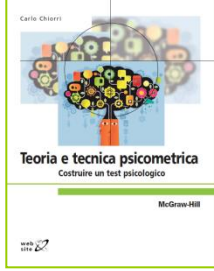


Statistiche descrittive

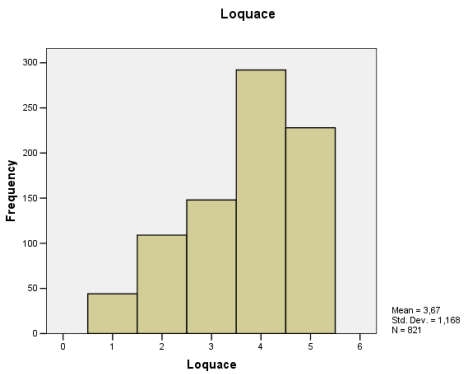
Descriptive Statistics

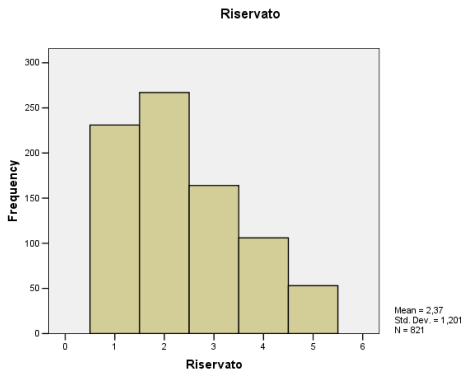
	N	Minimum	Maximum	Mean	Std.	Skewnes	Kurtosis
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
Loquace	821	1	5	3.67	1.168	-.649	-.482
Riservato	821	1	5	2.37	1.201	.603	-.577
Pieno di energia	821	1	5	3.73	1.058	-.543	-.382
Scatenato	821	1	5	3.47	.999	-.311	-.210
Taciturno	821	1	5	3.22	1.329	-.128	-1.157
Assertivo	821	1	5	3.68	1.043	-.522	-.303
Timido	821	1	5	2.52	1.228	.576	-.667
Socievole	821	1	5	3.78	1.111	-.675	-.339
Valid N (listwise)	821						

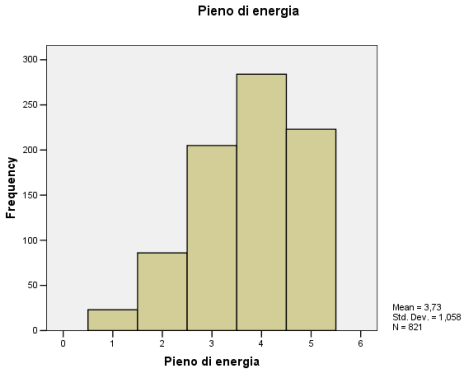
Copyright © 2011 The McGraw-Hill Companies srl

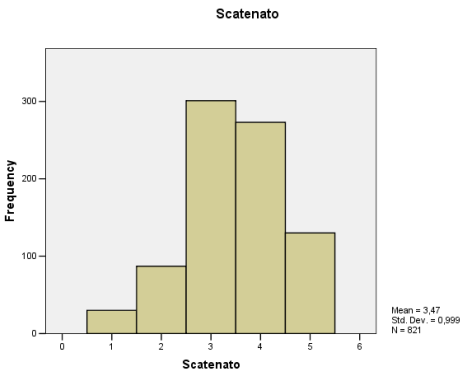


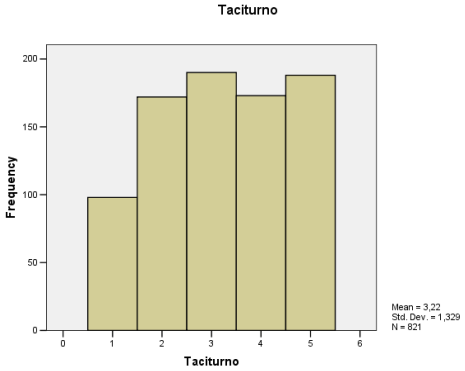
Distribuzioni

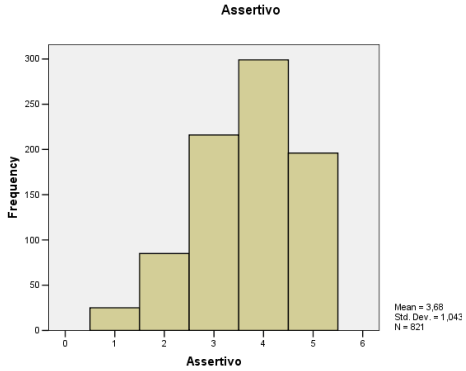


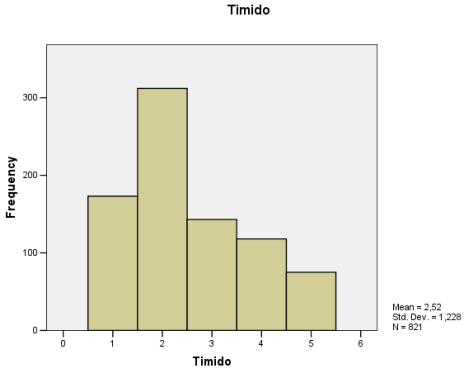


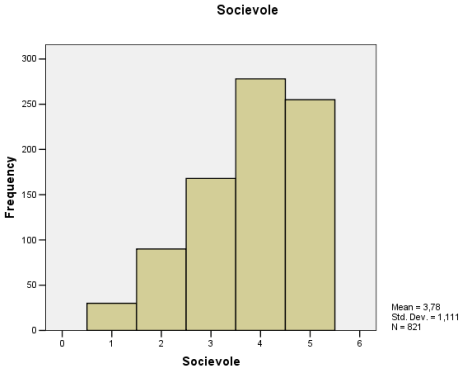




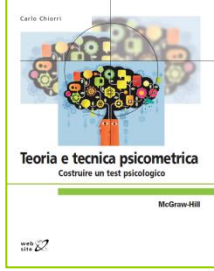








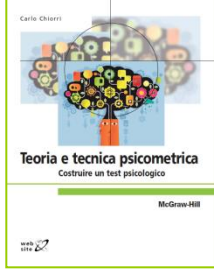
Copyright © 2011 The McGraw-Hill Companies srl



Discriminatività

	Q1		Q4		Significatività			Effect size	
	M	*DS*	*M*	*DS*	*t*	*gdl*	*p*	*d*	
Loquace	2.43	0.98	4.58	0.63	-26.88	354.753	<.001	2.61	Alto
Riservato	1.64	0.80	3.16	1.27	-14.84	362.993	<.001	1.43	Alto
Pieno di energia	2.90	0.97	4.50	0.72	-19.11	385.129	<.001	1.86	Alto
Scatenato	2.72	0.89	4.23	0.77	-18.64	411.712	<.001	1.81	Alto
Taciturno	1.89	0.89	4.41	0.83	-30.12	419.759	<.001	2.92	Alto
Assertivo	2.81	1.00	4.52	0.65	-20.80	359.272	<.001	2.02	Alto
Timido	1.72	0.82	3.56	1.16	-18.83	386.434	<.001	1.82	Alto
Socievole	2.53	0.89	4.78	0.42	-33.29	295.187	<.001	3.24	Alto

Copyright © 2011 The McGraw-Hill Companies srl



Ridondanza

Inter-Item Correlation Matrix

	Loquace	Riservato	Pieno di energia	Scatenato	Taciturno	Assertivo	Timido	Sociev ole
Loquace	1.000	.320	.260	.358	.564	.315	.324	.601
Riservato	.320	1.000	.061	.167	.350	.104	.292	.334
Pieno di energia	.260	.061	1.000	.454	.240	.660	.239	.342
Scatenato	.358	.167	.454	1.000	.253	.468	.239	.448
Taciturno	.564	.350	.240	.253	1.000	.270	.357	.544
Assertivo	.315	.104	.660	.468	.270	1.000	.269	.442
Timido	.324	.292	.239	.239	.357	.269	1.000	.339
Sociev ole	.601	.334	.342	.448	.544	.442	.339	1.000

The c ov ariance m atrix is calculated and used in the analysis.

Summary Item Statistics

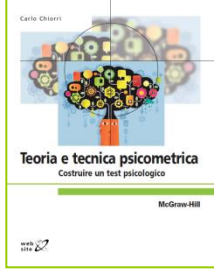
	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	.343	.061	.660	.599	10.820	.019	8

The covariance matrix is calculated and used in the analysis.

$$s_{rii} = \sqrt{.019} = .138$$

$$\alpha_{SE} = \frac{.138}{\sqrt{0.5 \times 8 \times (8-1) - 1}} = .027$$

Copyright © 2011 The McGraw-Hill Companies srl



Attendibilità

Reliability Statistics

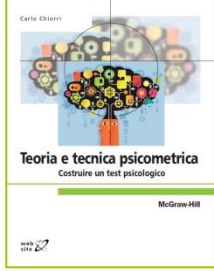
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.804	.807	8

Item-Total Statistics

	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Loquace	.611	.457	.767
Riservato	.352	.190	.807
Pieno di energia	.466	.468	.789
Scatenato	.504	.333	.785
Taciturno	.571	.419	.774
Assertivo	.532	.505	.780
Timido	.443	.206	.794
Socievole	.683	.513	.757

?

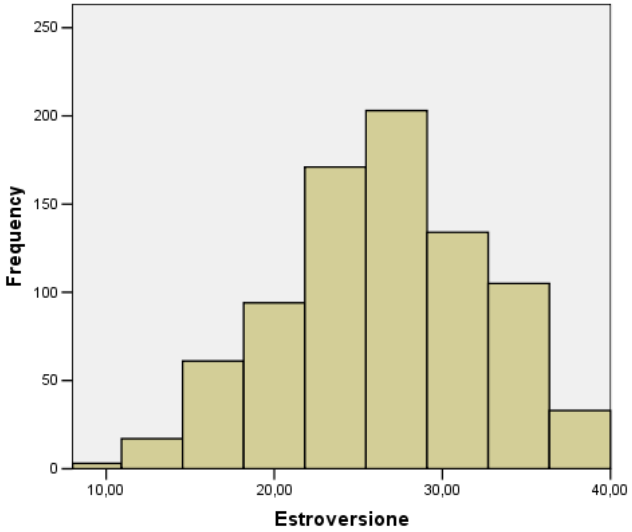
Copyright © 2011 The McGraw-Hill Companies srl



Punteggio totale

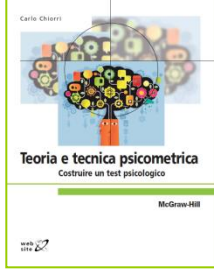
Descriptives

			Statistic	Std. Error
Estroversione	Mean		26.4397	.20796
	95% Confidence Interval for Mean	Lower Bound	26.0315	
		Upper Bound	26.8479	
	5% Trimmed Mean		26.5089	
	Median		27.0000	
	Variance		35.505	
	Std. Deviation		5.95862	
	Minimum		8.00	
	Maximum		40.00	
	Range		32.00	
	Interquartile Range		9.00	
	Skewness		-.183	.085
	Kurtosis		-.419	.170

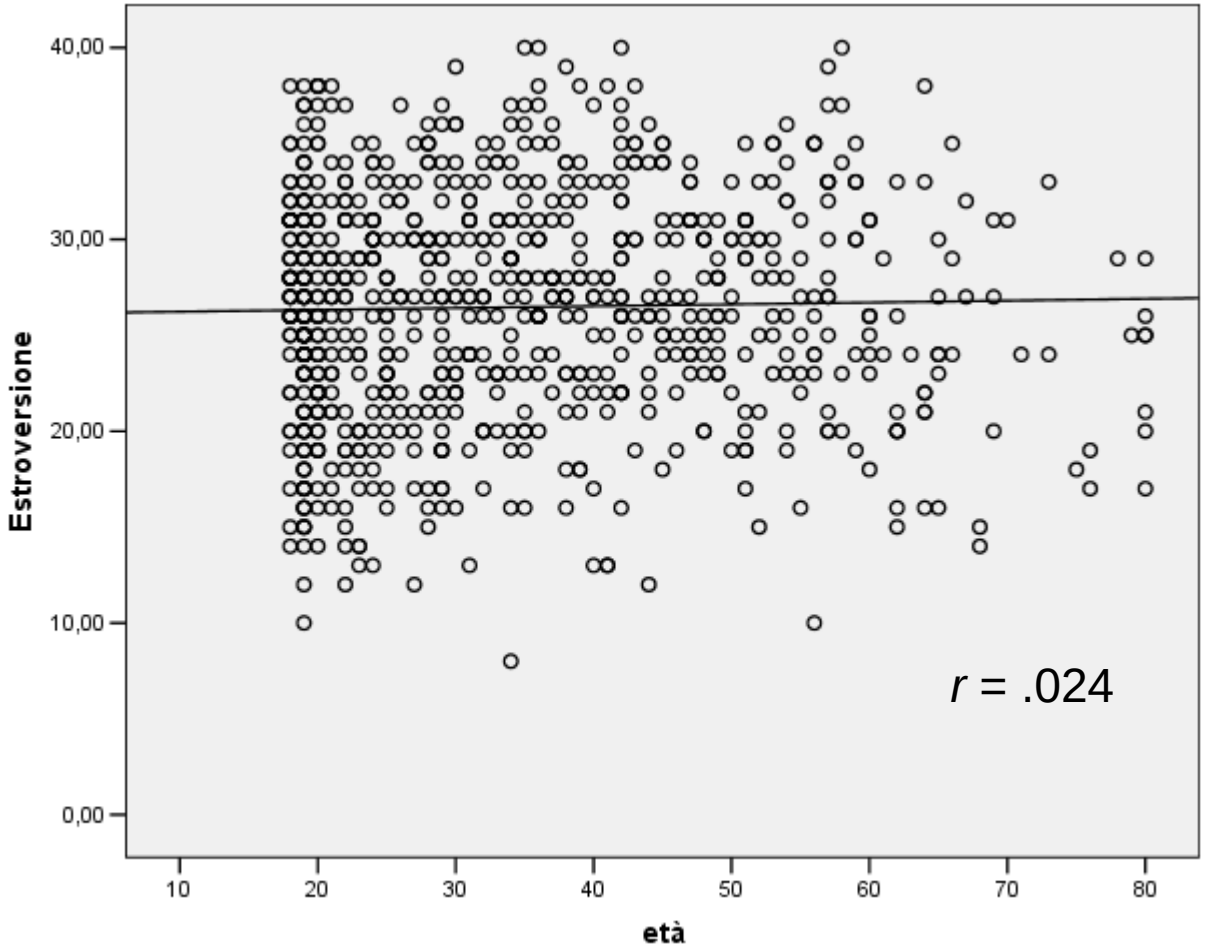


$K^2(2) = 15.507, p < .001$

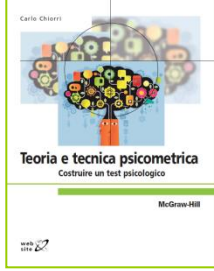
Copyright © 2011 The McGraw-Hill Companies srl



Relazione con l'età



Copyright © 2011 The McGraw-Hill Companies srl



Differenza di genere

Group Statistics

	sesso	N	Mean	Std. Deviation	Std. Error Mean
Estroversione	Femmina	473	26.4989	5.84624	.26881
	Maschio	347	26.3545	6.12382	.32874

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Estroversione	Equal variances assumed	2.679	.102	.343	818	.732	.14448	.42164	-.68314	.97209
	Equal variances not assumed			.340	725.581	.734	.14448	.42465	-.68922	.97817

Copyright © 2011 The McGraw-Hill Companies srl