

PSICOMETRIA

Curs 2024-2025

Autors:

Dra. Maite Barrios Cerrejón

Dra. M. Victòria Carreras Archs

Dra. Georgina Guilera Ferré



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement- No comercial- Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).

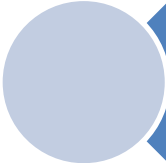
Tema	Número de diapositiva
Introducció a la Psicometria	3
Anàlisi d'ítems	25
Fiabilitat	77
Validesa	120
Baremació	211
Manual del Test	254



Introducció a la Psicometria



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement- No comercial- Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).



Lectura:

Muñiz, J. (2010). Las teorías de los tests: Teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31(1), 57-66.

Definició:

Etimològicament ➡ Psicometria = “Psykhē” – “Metrum”

Especialitat de la **Psicologia** que estudia els processos de mesura

Definició:

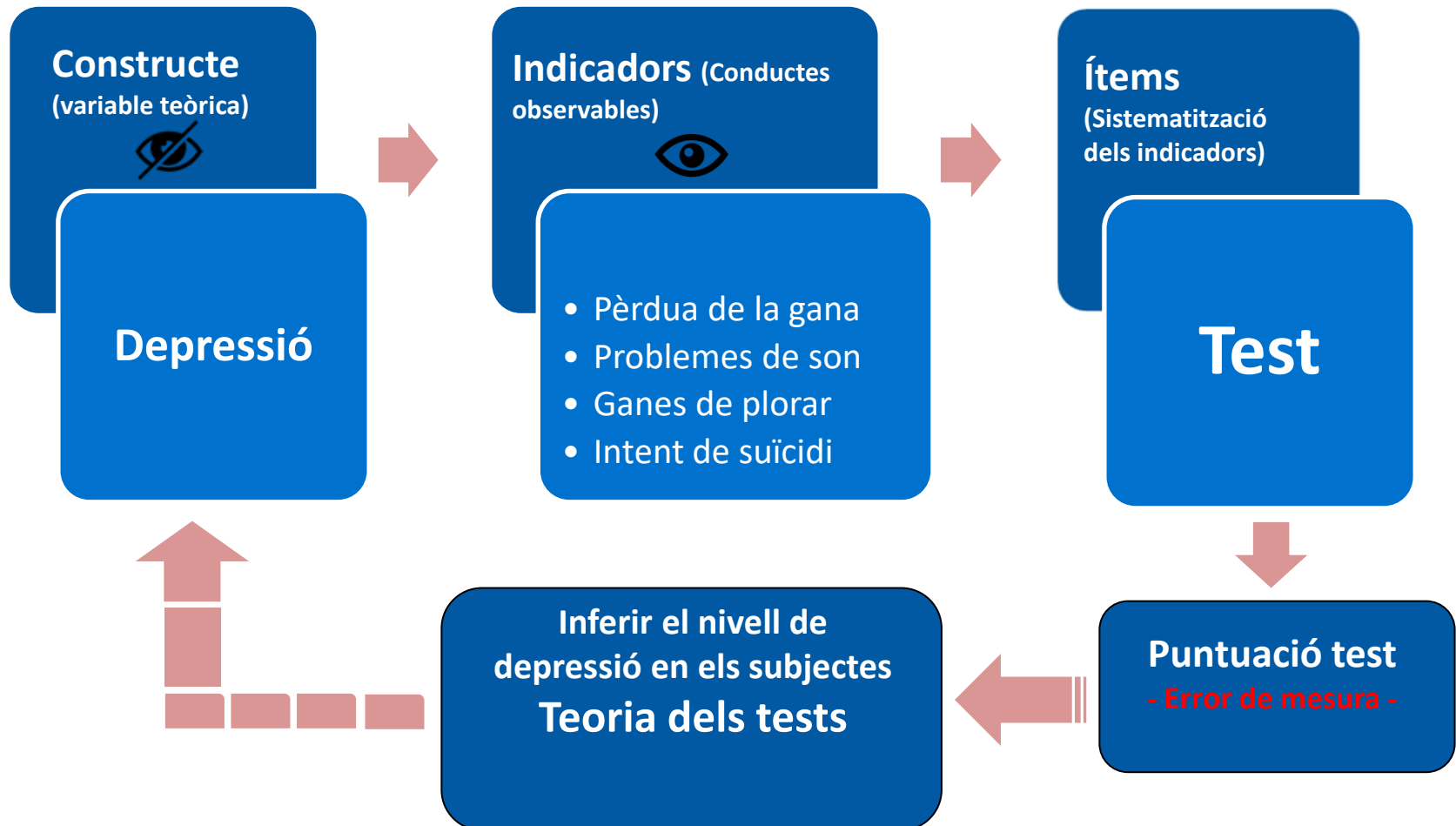
Psicometria: Terme que aglutina tot un **conjunt de models formals** que estableixen les bases per a que la mesura dels fenòmens psicològics es realitzi adequadament.

- Utilitza la formulació matemàtica per desenvolupar teories i models per a la mesura de variables psicològiques.
- Interès per promoure mètodes, tècniques i instruments per millorar la mesura de les variables psicològiques.

Com mesurem les característiques psicològiques?

- Les característiques psicològiques dels individus són inobservables, no es poden mesurar directament → **CONSTRUCTE**
- El constructe es pot manifestar a través d'una sèrie de conductes que sí són observables directament → **INDICADORS**
- Per avaluar el *nivell* dels subjectes en el *constructe* ens haurem de basar en els *indicadors*
- La sistematització d'un conjunt d'indicadors amb l'objectiu d'avaluar el nivell dels subjectes en un determinat constructe constitueix un **instrument de mesura o test**

Exemple: Depressió



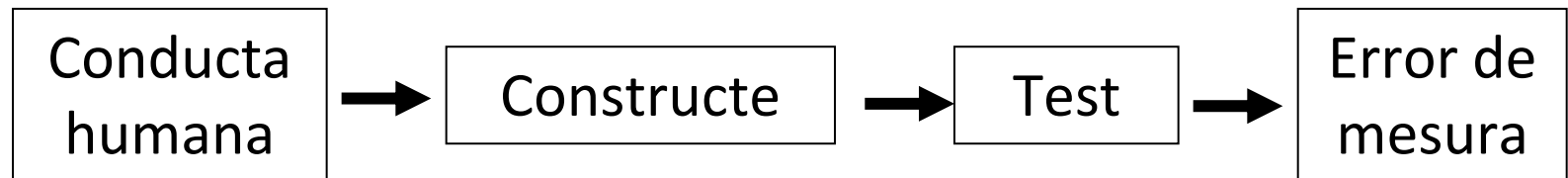
Instrument de mesura o test:

Sistematització d'un **conjunt d'indicadors** amb la finalitat d'avaluar diversos nivells d'un constructe (variable teòrica inobservable que només es pot mesurar mitjançant indicadors observables que el representin).

Elements del test:

Constitueixen l'operativització del constructe ➡ han de reunir una sèrie de requisits per a demostrar que compleixen la funció per la qual han estat creats.

Els instruments de mesura no són totalment precisos
contenen  **errors de mesura**



Continguts:

- **Teoria de la mesura:** fonamentació teòrica de la mesura
- **Escalament:** problemàtica de l'escalament dels estímuls psicològics i psicofísics
- **Teoria dels tests:** s'explicita la lògica i els models matemàtics subjacents a la construcció i ús dels tests
 - Teoria clàssica dels tests
 - Teoria de resposta a l'ítem
 - Tests referits al criteri

Teoria dels Tests

Teoria dels tests ➡ Funció matemàtica o model que estableix una relació entre:

- Variables observables (puntuacions en el test)
- Variables inobservables (puntuacions verdaderes o nivell dels subjectes en el tret mesurat pel test)

Questions:

- En quin grau les puntuacions del test indiquen el nivell de cada subjecte en el tret o característica mesurada?
- Com ens podem assegurar que la puntuació obtinguda per un subjecte és la que realment li correspon?
- En quin grau podem refiar-nos que, a partir de les puntuacions del test podem inferir el nivell del subjecte en el constructe mesurat?

Cal estimar els errors de mesura mitjançant la teoria dels tests

Teoria dels Tests

- Diferents models ➡ tots ells analitzen la incidència dels errors en la mesura dels constructes psicològics i el grau en què afecten la fiabilitat dels instruments de mesura.
- Teoria dels tests més estesa ➡ Teoria Clàssica dels Tests ➡ model lineal.

Teoria Clàssica dels Tests (TCT)

Model lineal de la puntuació veritable proposat per Spearman:

$$X = V + e$$

On:

X – puntuació en el test

V – puntuació verdadera, $E(X)$

e - error

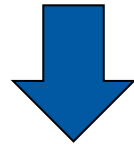
Teoria Clàssica dels Tests (TCT)

Model	$X = V + e$
Supòsits	$V = E(X)$ Error aleatori: $E(e) = 0$ $r_{ve} = 0$ $r_{e_j e_k} = 0$
Test paral·lels	$V_j = V_k \quad \text{i} \quad \sigma_{e_j}^2 = \sigma_{e_k}^2$

Limitacions TCT

- Absència d'invariància respecte a l'instrument utilitzat.
- Absència d'invariància respecte a la mostra utilitzada per calcular les propietats dels tests.
- S'assumeix que la precisió del test és la mateixa independentment del nivell del tret mesurat.
- No es disposa d'indicadors de bondat d'ajust que ens informin del grau en què el model s'ajusta a les dades.

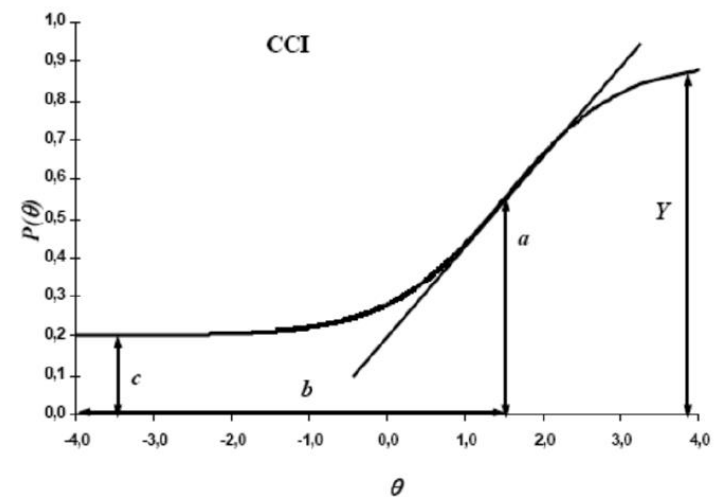
Les deficiències de la TCT han estat el factor desencadenant de models alternatius:



- La **Teoria de Resposta a l'Ítem (TRI)** considera l'ítem com la unitat d'anàlisi bàsica del test, enlloc de les puntuacions totals.
- En els **Tests Referits al Criteri (TRC)** l'objectiu és construir i avaluar tests que permetin interpretar les puntuacions de coneixements, habilitats i destreses en sentit absolut, sense referències a cap grup

Teoria de la Resposta a l'Ítem (TRI)

- Corba Característica de l'ítem (CCI)
- La forma concreta de la CCI ve donada per tres paràmetres:
 - a: l'índex de discriminació
 - b: la dificultat de l'ítem
 - c: la probabilitat d'encertar l'ítem per l'atzar
- Supòsits:
 - Unidimensionalitat
 - Independència local



Nivell subjectes en la capacitat (test)

Models TRI

- Model logístic d'un paràmetre (*o model de Rasch*)

$$P(\theta) = \frac{1}{1 + e^{-D(\theta - b_j)}}$$

- Model logístic de dos paràmetres

$$P(\theta) = \frac{1}{1 + e^{-Da_j(\theta - b_j)}}$$

- Model logístic de tres paràmetres

$$P(\theta) = c_j + (1 - c_j) \frac{1}{1 + e^{-Da_j(\theta - b_j)}}$$

On:

$P(\theta)$: Probabilitat d'encertar l'ítem j si el nivell del tret és θ

θ : nivell del tret o habilitat de la persona

b_j : Paràmetre de dificultat de l'ítem j

a_j : paràmetre de discriminació de l'ítem j

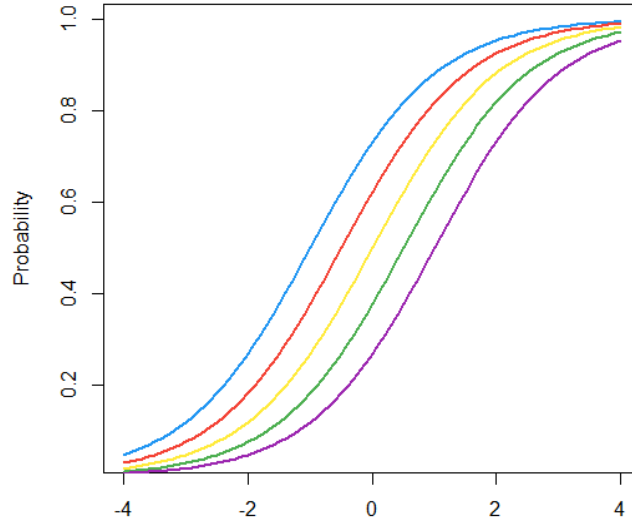
e : base dels logaritmes neperians (2,718)

D : Constant. $D=1,702$ (mètrica normal), $D=1$ (mètrica logística)

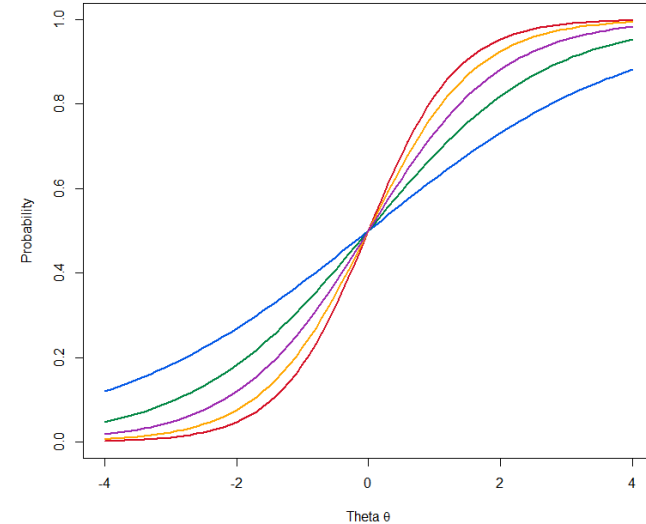
c_j : probabilitat d'encertar l'ítem j per persones que tenen un nivell en el tret molt baix.

Exemples de CCI

I.

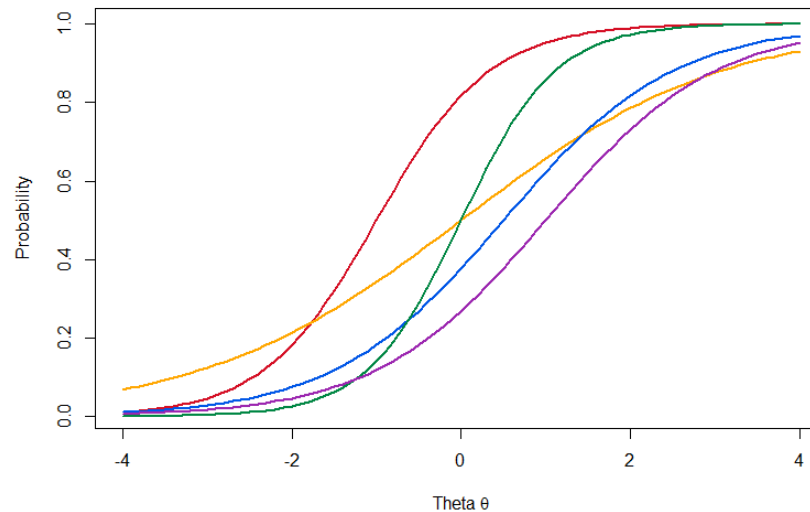


II.



Theta θ

III.



TCT vs TRI

	TCT	TRI
Model	Lineal	No lineal
Interpretació	Puntuació total del test	Ítem
Precisió	La mateixa per a tots els participants / puntuacions	Estimada per cada participant / puntuació
Puntuació	Depèn dels ítems	Independent dels ítems
Paràmetres dels ítems	Depèn de la mostra utilitzada	És independent de la mostra utilitzada

Test referits al criteri (TRC)

- Teoria desenvolupada a partir de la proposta de Glaser (1963).
- Gran influència en l'àmbit educatiu.
- L'objectiu és determinar si les persones dominen un criteri o camp de coneixement concret.
- Es pretén avaluar en quin grau els individus tenen uns determinats coneixements.

Què tractarem a l'assignatura?

- Anàlisi d'ítems
- Fiabilitat
- Validesa
- Baremació
- El manual del test

Anàlisi d'ítems



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement- No comercial- Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).

Lectures:

Abad, F. J., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis. [pàgines: 38-39, 43-46, 49-50, 54-57, 65-67]

Barbero García, M. I., Vila Abad, E. y Holgado Tello, F. P. (2010). *Psicometría*. Madrid: UNED. [pàgines: 72-80]

Muñiz, J. (2018). *Introducción a la Psicometría. Teoría clásica y TRI*. Madrid: Pirámide. [pàgines: 274 – 279]

Anàlisi d'ítems

- **Anàlisi qualitatiu:** judici d'experts sobre les característiques substancials, estilístiques, d'adequació i d'imparcialitat. Caldrà valorar:

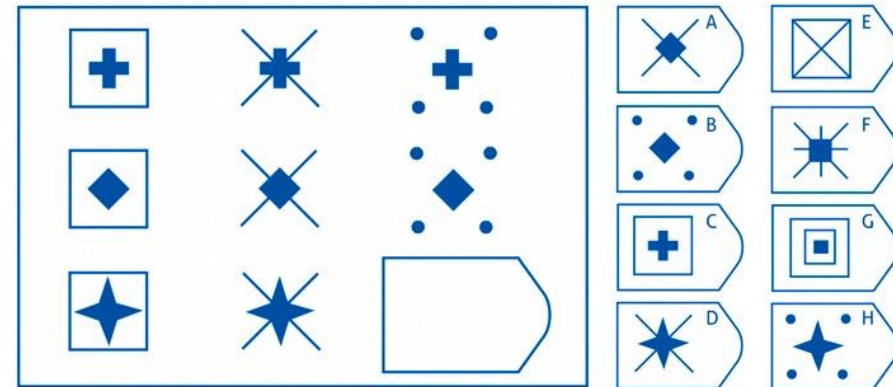
- Adequació del contingut
- Format dels ítems per a la població diana i finalitat del test
- Claredat en el redactat
- Correcció gramatical
- Seguiment de les regles bàsiques de redacció d'ítems

- **Anàlisi quantitatiu:** Càlcul d'indicadors per estudiar les característiques psicomètriques dels ítems a partir de les respostes obtingudes en mostres.

- Índex de dificultat (ID)
- Índex de discriminació (D)
- Anàlisi dels distractors
- Índex de fiabilitat
- Índex de validesa

Directrius per a la construcció dels ítems del test

- Definir l'**objectiu** de l'avaluació.
- **Context** en què s'utilitzaran els ítems (població objectiu, tipus d'administració i condicions en les que s'administrarà el test).
- **Domini** i context d'interès. Determinar els nivells del constructe i escollir els indicadors que el representen.



Elaboració dels ítems del test

Tancat o d'elecció

- Elecció binària
- Elecció múltiple
- Aparellament
- Format incomplet (*cloze*)
- Escales de classificació (*rating scales*)
- Llistats (*checklists*)

Obert o de construcció

- Resposta extensa
- Resposta curta

Recomanacions per a la construcció d'ítems de resposta múltiple I

- Contingut:
 - Cada ítem dirigit a avaluar un únic aspecte.
 - Parafrasejar material que ha de ser avaluat.
 - Cada ítem ha de ser independent de la resta d'ítems del test.
 - Avaluar contingut rellevant.
 - Opcions clares, unívokes i indiscutibles.
 - Evitar ítems “trampa”.
- Format:
 - Alternatives en format vertical.
- Estil:
 - Revisar aspectes gramaticals i ortogràfics.
 - Complexitat del llenguatge ajustada a les persones avaluades.
 - Brevetat.

Recomanacions per a la construcció d'ítems de resposta múltiple II

- Redacció de l'enunciat:
 - Idea central en l'enunciat i no en les alternatives.
 - Expressar l'enunciat en termes positius.
- Redacció de les alternatives:
 - Alternatives plausibles i discriminatives.
 - Una única alternativa és la resposta correcta.
 - Aleatoritzar la localització de la resposta correcta.
 - Ordenar les alternatives seguint un ordre lògic.
 - Alternatives independents entre sí.
 - No utilitzar com alternativa: totes les anteriors o cap de les anteriors.
 - Redacció en forma positiva.
 - Evitar donar pistes sobre la resposta correcta (*ex. longitud, us de determinats extrems, associació de termes, poca plausibilitat dels distractors, homogeneïtat*).
 - Distractors plausibles.
 - No és de l'humor.

Anàlisi d'ítems: passos

1

- Decidir les principals propietats de les puntuacions del test.

2

- Identificar les anàlisis d'ítems més rellevants per a aquestes propietats.

3

- Administrar els ítems a una mostra representativa de la població a la qual està dirigit el test.

4

- Estimar les anàlisis identificades en el pas 2 per cada un dels ítems.

5

- Establir un pla per a la selecció d'ítems. Identificar i revisar els que funcionin malament.

6

- Seleccionar el bloc final d'ítems.

Anàlisi d'ítems: indicadors clàssics

- Índex de dificultat (ID)
- Índex de discriminació (D)
- Anàlisi dels distractors
- Índex de fiabilitat de l'ítem
- Índex de validesa de l'ítem

Índex de dificultat

Sense correcció per atzar

$$ID = \frac{A}{N}$$

Correcció per atzar

$$ID = \frac{A - E/(k - 1)}{N}$$

Caldrà tenir en compte si es tracta d'un:

Test de potència

On N és el total d'individus que
han intentat contestar l'ítem

Test de velocitat

On N és el total d'individus
que **han contestat** l'ítem

Índex de dificultat sense corregir per atzar

$$ID = \frac{A}{N} \qquad ID_1 = \frac{4}{8} = 0,5$$

Participant	Ítems						Total
	1	2	3	4	5	6	
1	0	1	0	1	0	1	3
2	1	1	1	0	0	1	4
3	0	1	0	1	1	1	4
4	1	1	0	1	0	1	4
5	1	1	0	1	0	0	3
6	0	1	0	0	1	0	2
7	1	1	1	1	1	1	6
8	0	0	0	0	1	0	1
Total A	4	7	2	5	4	5	
ID	0,5	0,875	0,25	0,625	0,5	0,625	

Índex de dificultat sense corregir per atzar

$$ID = \frac{A}{N}$$

Participant	Ítems						Total
	1	2	3	4	5	6	
1	0	1	0	1	0	1	3
2	1	1	1	0	0	1	4
3	0	1	0	1	1	1	4
4	1	1	0	1	0	1	4
5	1	1	0	1	0	0	3
6	0	1	0	0	1	0	2
7	1	1	1	1	1	1	6
8	0	0	0	0	1	0	1
Total A	4	7	2	5	4	5	
ID	0,5	0,875	0,25	0,625	0,5	0,625	

$$ID_1 = \frac{4}{8} = 0,5$$

$$ID_2 = \frac{7}{8} = 0,875$$

$$ID_3 = \frac{2}{8} = 0,25$$

$$ID_4 = \frac{5}{8} = 0,625$$

$$ID_5 = \frac{4}{8} = 0,5$$

$$ID_6 = \frac{5}{8} = 0,625$$

Índex de dificultat sense corregir per atzar

Participant	Ítems						Total
	1	2	3	4	5	6	
1	0	1	0	1	0	1	3
2	1	1	1	0	0	1	4
3	0	1	0	1	1	1	4
4	1	1	0	1	0	1	4
5	1	1	0	1	0	0	3
6	0	1	0	0	1	0	2
7	1	1	1	1	1	1	6
8	0	0	0	0	1	0	1
Total A	4	7	2	5	4	5	
ID	0,5	0,875	0,25	0,625	0,5	0,625	

$$\bar{X} = \frac{\sum x_j}{N} = \sum ID_j$$

$$\bar{x} = \frac{\sum x_j}{N} = \frac{27}{8} = 3,38$$

$$\sum ID_j = 3,38$$

Exercici: Índex de dificultat sense corregir per atzar

<i>Participant</i>	<i>Ítems</i>							
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
<i>1</i>	0	0	0	1	1	0	1	1
<i>2</i>	1	0	0	1	0	0	1	1
<i>3</i>	1	1	0	1	1	0	1	1
<i>4</i>	1	0	1	1	1	1	1	0
<i>5</i>	0	1	0	1	1	1	1	0
<i>6</i>	1	1	1	1	0	0	1	1
<i>7</i>	1	0	0	0	0	1	1	0
<i>8</i>	1	0	0	1	1	0	1	1
<i>9</i>	1	1	1	1	1	1	1	1
<i>10</i>	0	1	0	1	0	0	1	1

Comproveu que la mitjana del test equival a la suma dels ID

Índex de dificultat corregit per atzar

$$ID = \frac{A - E / (k - 1)}{N}$$

Puntuació corregida del participant (Pc)

$$Pc = A - E / (k - 1)$$

Exemple: Índex de dificultat corregit per atzar

Test de potència

<i>P</i>	<i>Ítems (5 alternatives de resposta)</i>								<i>Totals</i>			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>A</i>	<i>E</i>	<i>NC</i>	<i>Pc</i>
1	1	1	0	1	X	1	1	X	5	1	2	4,75
2	0	1	0	1	1	0	X	1	4	3	1	3,25
3	1	1	1	1	X	1	1	0	6	1	1	5,75
4	1	0	0	1	1	0	X	0	3	4	1	2,00
5	1	1	1	0	1	1	1	1	7	1	0	6,75
6	1	0	1	1	1	X	0	0	4	3	1	3,25
7	1	1	1	1	1	X	1	0	6	1	1	5,75
8	0	0	1	1	1	1	X	1	5	2	1	4,50
9	1	0	1	1	X	1	X	1	5	1	2	4,75
10	1	X	1	X	1	1	1	1	6	0	2	6,00
A	8	5	7	8	7	6	5	5	1 – encert			
E	2	4	3	1	0	2	1	4	0 – error			
Cont.	10	9	10	9	7	8	6	9	X – no contestat			

 $\sum Pc=46,75$

Exemple: Índex de dificultat corregit per atzar

Test de potència

<i>P</i>	<i>Ítems (5 alternatives de resposta)</i>							
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
1	1	1	0	1	X	1	1	X
2	0	1	0	1	1	0	X	1
3	1	1	1	1	X	1	1	0
4	1	0	0	1	1	0	X	0
5	1	1	1	0	1	1	1	1
6	1	0	1	1	1	X	0	0
7	1	1	1	1	1	X	1	0
8	0	0	1	1	1	1	X	1
9	1	0	1	1	X	1	X	1
10	1	X	1	X	1	1	1	1
A	8	5	7	8	7	6	5	5
E	2	4	3	1	0	2	1	4
Cont.	10	9	10	9	7	8	6	9

$$ID = \frac{A - E/(k - 1)}{N}$$

$$ID_1 = \frac{8 - 2/4}{10} = 0,75$$

$$ID_2 = \frac{5 - 4/4}{10} = 0,40$$

Exemple: Índex de dificultat corregit per atzar

Test de potència

<i>P</i>	<i>Ítems (5 alternatives de resposta)</i>							
	1	2	3	4	5	6	7	8
1	1	1	0	1	X	1	1	X
2	0	1	0	1	1	0	X	1
3	1	1	1	1	X	1	1	0
4	1	0	0	1	1	0	X	0
5	1	1	1	0	1	1	1	1
6	1	0	1	1	1	X	0	0
7	1	1	1	1	1	X	1	0
8	0	0	1	1	1	1	X	1
9	1	0	1	1	X	1	X	1
10	1	X	1	X	1	1	1	1
A	8	5	7	8	7	6	5	5
E	2	4	3	1	0	2	1	4
Cont.	10	9	10	9	7	8	6	9

$$ID_1 = \frac{8 - 2/4}{10} = 0,75$$

$$ID_2 = \frac{5 - 4/4}{10} = 0,40$$

$$ID_3 = \frac{7 - 3/4}{10} = 0,625$$

$$ID_4 = \frac{8 - 1/4}{10} = 0,775$$

$$ID_5 = \frac{7 - 0/4}{10} = 0,70$$

$$ID_6 = \frac{6 - 2/4}{10} = 0,55$$

$$ID_7 = \frac{5 - 1/4}{10} = 0,475$$

$$ID_8 = \frac{5 - 4/4}{10} = 0,40$$

Exemple: Índex de dificultat corregit per atzar
Test de potència

P	Ítems (5 alternatives de resposta)								Totals			
	1	2	3	4	5	6	7	8	A	E	NC	Pc
1	1	1	0	1	X	1	1	X	5	1	2	4,75
2	0	1	0	1	1	0	X	1	4	3	1	3,25
3	1	1	1	1	X	1	1	0	6	1	1	5,75
4	1	0	0	1	1	0	X	0	3	4	1	2,00
5	1	1	1	0	1	1	1	1	7	1	0	6,75
6	1	0	1	1	1	X	0	0	4	3	1	3,25
7	1	1	1	1	1	X	1	0	6	1	1	5,75
8	0	0	1	1	1	1	X	1	5	2	1	4,50
9	1	0	1	1	X	1	X	1	5	1	2	4,75
10	1	X	1	X	1	1	1	1	6	0	2	6,00
A	8	5	7	8	7	6	5	5	1: encert 0: error X: no contestat ΣPc=46,75			
E	2	4	3	1	0	2	1	4				
Cont.	10	9	10	9	7	8	6	9				

$\bar{x} = \frac{46,75}{10} = 4,68$

$\sum ID = 4,68$

$Pc = A - E / (k - 1)$

Exemple: Índex de dificultat corregit per atzar

Test de velocitat

<i>P</i>	<i>Ítems (5 alternatives de resposta)</i>								<i>Totals</i>		
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>A</i>	<i>E</i>	<i>NC</i>
1	1	1	0	1	1	1	1	X	6	1	1
2	0	1	0	1	1	0	X	X	3	3	2
3	1	1	1	1	1	1	1	0	7	1	0
4	1	0	0	1	0	0	X	X	2	4	2
5	1	1	1	0	1	1	1	1	7	1	0
6	1	0	1	1	1	X	X	X	4	1	3
7	1	1	1	1	1	X	X	X	5	0	3
8	0	0	1	1	1	1	X	X	4	2	2
9	1	0	1	1	1	1	X	X	5	1	2
10	1	0	1	0	1	1	1	1	6	2	0
A	8	5	7	8	9	6	4	2	1: encert		
E	2	5	3	2	1	2	0	1	0 : error		
Cont.	10	10	10	10	10	8	4	3	X : no contestat		

$$ID = \frac{A - E/(k - 1)}{N}$$

$$ID_1 = \frac{8 - 2/4}{10} = 0,75$$

$$ID_6 = \frac{6 - 2/4}{8} = 0,69$$

Exemple: Índex de dificultat corregit per atzar

Test de velocitat

P	Ítems (5 alternatives de resposta)								Totals		
	1	2	3	4	5	6	7	8	A	E	NC
1	1	1	0	1	1	1	1	X	6	1	1
2	0	1	0	1	1	0	X	X	3	3	2
3	1	1	1	1	1	1	1	0	7	1	0
4	1	0	0	1	0	0	X	X	2	4	2
5	1	1	1	0	1	1	1	1	7	1	0
6	1	0	1	1	1	X	X	X	4	1	3
7	1	1	1	1	1	X	X	X	5	0	3
8	0	0	1	1	1	1	X	X	4	2	2
9	1	0	1	1	1	1	X	X	5	1	2
10	1	0	1	0	1	1	1	1	6	2	0
A	8	5	7	8	9	6	4	2	1: encert 0 : error X : no contestat		
E	2	5	3	2	1	2	0	1			
Cont.	10	10	10	10	10	8	4	3			

$$ID_1 = \frac{8 - 2/4}{10} = 0,75$$

$$ID_2 = \frac{5 - 5/4}{10} = 0,38$$

$$ID_3 = \frac{7 - 3/4}{10} = 0,63$$

$$ID_4 = \frac{8 - 2/4}{10} = 0,75$$

$$ID_5 = \frac{9 - 1/4}{10} = 0,88$$

$$ID_6 = \frac{6 - 2/4}{8} = 0,69$$

$$ID_7 = \frac{4 - 0/4}{4} = 1,00$$

$$ID_8 = \frac{2 - 1/4}{3} = 0,58$$

Exemple: Índex de dificultat corregit per atzar

Test de velocitat

P	Ítems (5 alternatives de resposta)								Totals				
	1	2	3	4	5	6	7	8	A	E	NC	Pc	
1	1	1	0	1	1	1	1	X	6	1	1	5,75	
2	0	1	0	1	1	0	X	X	3	3	2	2,25	
3	1	1	1	1	1	1	1	0	7	1	0	6,75	
4	1	0	0	1	0	0	X	X	2	4	2	1	
5	1	1	1	0	1	1	1	1	7	1	0	6,75	
6	1	0	1	1	1	X	X	X	4	1	3	3,75	
7	1	1	1	1	1	X	X	X	5	0	3	5	
8	0	0	1	1	1	1	X	X	4	2	2	3,5	
9	1	0	1	1	1	1	X	X	5	1	2	4,75	
10	1	0	1	0	1	1	1	1	6	2	0	5,5	
A	8	5	7	8	9	6	4	2	1: encert				ΣPc
E	2	5	3	2	1	2	0	1	0: error				
Cont.	10	10	10	10	10	8	4	3	X: no contestat				

$$\bar{x} = \frac{45,00}{10} = 4,5$$

$$\sum ID = 5,65$$

$$Pc = A - E/(k - 1)$$

Exercici: Índex de Dificultat corregit per atzar

Test de potència

<i>P</i>	<i>Ítems (5 alternatives de resposta)</i>							
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
1	1	1	X	1	X	0	0	1
2	1	1	1	1	1	X	1	1
3	0	1	1	0	1	1	1	1
4	0	1	1	0	0	0	0	1
5	0	0	0	1	0	0	0	1
6	1	1	0	1	0	0	0	0
7	1	1	0	X	0	1	1	1
8	X	1	1	X	1	X	1	1
9	1	0	1	1	1	X	0	1
10	1	0	1	0	0	1	0	1

Calculeu els ID dels
ítems i les Pc dels
participants

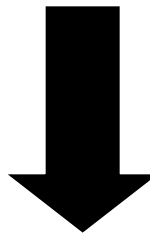
1 - Encerts

0 - Errors

X - No contesta

Anàlisi d'ítems: Índex de Discriminació

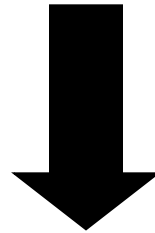
Avalua si els ítems són capaços de posar de manifest les diferències individuals entre les persones en la variable que estem mesurant.



Un ítem té capacitat discriminativa si els individus que tenen un nivell alt en la variable encerten més l'ítem o obtenen puntuacions més elevades que els individus que tenen un nivell baix en la variable.

Anàlisi d'ítems: Índex de Discriminació

Posa en relació l'execució de l'individu en l'ítem amb la seva execució global en el test



Càlcul:

- Índex basat en la diferència de les proporcions d'encert o de respostes positives a l'ítem
- Índexs basats en les correlacions ítem-test

Criteris per a la valoració de la discriminació de l'ítem (Ebel & Frisbie, 1991)

Índex de discriminació	Valoració de l'ítem
$\geq 0,40$	Ítem amb molt bona capacitat discriminativa
0,30 – 0,39	Ítem amb una capacitat discriminativa raonablement bona, però susceptible a ser millorat
0,20 – 0,29	Límit. Es requereix millorar l'ítem
$< 0,20$	Cal eliminar l'ítem o modificar-lo completament

Índex de Discriminació (D) basat en la diferència de les proporcions d'encert o respostes positives

$$D = \frac{n_s}{N_s} - \frac{n_i}{N_i}$$

On:

n_s – nombre persones que encerten ítem o donen una resposta positiva en el grup superior

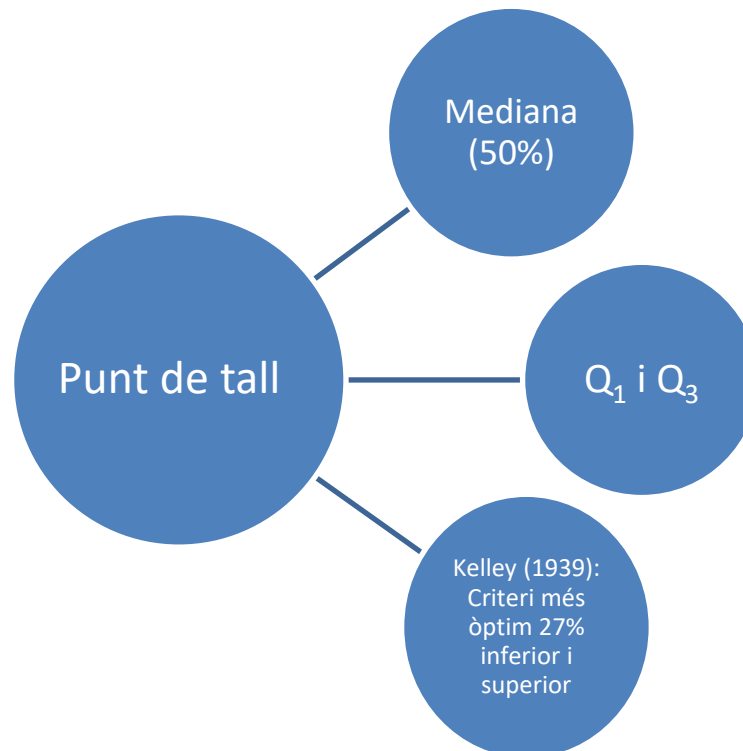
n_i – nombre persones que encerten ítem o donen una resposta positiva en el grup inferior

N_s – nombre persones que responen en el grup superior

N_i – nombre persones que responen en el grup inferior

Índex de Discriminació (D) basat en la diferència de les proporcions d'encert o respostes positives

Rang: $-1 \leq D \leq 1$, $D=1$: màxima capacitat discriminativa



Exemple: D basat en la diferència de les proporcions d'encert o respostes positives

Grups basats en la mediana (50%)

<i>P</i>	<i>Ítems</i>						<i>Punt. Total</i>	<i>Grup</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>		
1	0	1	0	1	0	1	3	Inf
2	1	1	1	0	0	1	4	Sup
3	0	1	0	1	1	1	4	Sup
4	1	1	0	1	0	1	4	Sup
5	1	1	0	1	0	0	3	Inf
6	0	1	0	0	1	0	2	Inf
7	1	1	1	1	1	1	6	Sup
8	0	0	0	0	1	0	1	Inf
<i>n_s</i>	3	4	2	3	2	4		
<i>n_i</i>	1	3	0	2	2	1		
<i>D</i>	0.5	0.25	0.5	0.25	0	0.75		

Exercici: D basat en la diferència de les proporcions d'encert o respostes positives

Índex de discriminació de l'ítem 1

Participant	ítem 1	ítem 2	ítem 3	ítem 4	ítem 5	ítem 6	ítem 7	ítem 8	Total
1	0	0	1	1	1	1	0	1	5
2	1	1	1	1	1	1	1	1	8
3	0	0	0	1	0	1	0	1	3
4	0	1	0	0	0	0	0	0	1
5	0	1	1	0	0	1	0	0	3
6	1	1	1	1	1	1	1	1	8
7	1	0	0	1	1	0	1	0	4
8	1	1	1	0	0	1	0	0	4
9	0	1	1	1	1	1	1	1	7
10	1	0	0	1	1	1	0	1	5

Índex de Discriminació (D) basats en la correlació ítem-test

Rang: $-1 \leq D \leq 1$, $D=1$: màxima capacitat discriminativa

<i>Tipus de Correlacions</i>		
<i>Ítem</i>	<i>Test</i>	<i>Correlació</i>
Quantitativa	Quantitativa	Pearson r
Dicotòmica	Quantitativa	Biserial-Puntual r_{bp}
Dicotomitzada	Quantitativa	Biserial r_b
Dicotòmica	Dicotòmica	Fi ϕ
Dicotomitzada	Dicotomitzada	Tetracòrica r_t

Restar la puntuació de l'ítem en la del test o
utilitzar la fórmula de correcció de la correlació

Coefficient de Correlació de Pearson

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

On:

S_{xy} Covariància test-ítem

S_x desviació típica del test

S_y desviació típica de l'ítem

Coefficient de Correlació Biserial-Puntual

$$r_{bp} = \frac{\bar{x}_p - \bar{x}_x}{S_x} \sqrt{\frac{p}{q}}$$

On:

$\bar{x}_p$ mitjana en el test dels subjectes que encerten o donen una resposta positiva a l'ítem

$\bar{x}_x$ mitjana del test

S_x desviació típica del test

p proporció de subjectes que encerten o donen una resposta positiva a l'ítem

q (1 - p)

Coefficient de Correlació Biserial

$$r_b = \frac{\bar{x}_p - \bar{x}_x}{S_x} \frac{p}{y}$$

On:

$\bar{x}_p$ mitjana en el test dels subjectes que encerten o donen una resposta positiva a l'ítem

$\bar{x}_x$ mitjana del test

S_x desviació típica del test

p proporció de subjectes que encerten o donen una resposta positiva a l'ítem

y ordenada que correspon en la distribució normal al valor que deixa per sota una àrea igual a p

Coefficient Fi ϕ

$$\Phi = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}}$$

On:

p_{xy} proporció d'encerts o respostes positives en x i y

p_x proporció d'encerts o respostes positives en x

p_y proporció d'encerts o respostes positives en y

$q_x = 1 - p_x$

$q_y = 1 - p_y$

Coefficient Tetracòric

		Test		
		0	1	
ítem	0	a	b	p
	1	c	d	q
		p'	q'	n

$$r_t = \sqrt{\frac{(ad) - (bc)}{n^2 p_y p'_y}}$$

On:

y ordenada de la corba normal que separa les dues categories del test (grup superior/inferior)

y' ordenada de la corba normal que separa les dues categories de l'ítem (encert/error)

a, b, c i d fi en la taula de contingència 2 x 2 entre el test i l'ítem

Exemple: Correlació biserial-puntual

P	Ítems						Totals	
	1	2	3	4	5	6	Total	Total-4
1	0	1	0	1	0	1	3	2
2	1	1	1	0	0	1	4	4
3	0	1	0	1	1	1	4	3
4	1	1	0	1	0	1	4	3
5	1	1	0	1	0	0	3	2
6	0	1	0	0	1	0	2	2
7	1	1	1	1	1	1	6	5
8	0	0	0	0	1	0	1	1
Total	4	7	2	5	4	5		

r_{bp} de l'ítem 4?

$$p = \frac{5}{8} = 0,625$$

$$q = \frac{3}{8} = 0,375$$

$$\bar{x} = \frac{2+4+3+3+2+2+5+1}{8} = 2,75$$

$$\bar{x}_p = \frac{2+3+3+2+5}{5} = 3$$

$$S_x^2 = \frac{2^2+4^2+3^2+3^2+2^2+2^2+5^2+1^2}{8} - 2,75^2 = 1,44 \Rightarrow S_x = 1,20$$

$$r_{bp} = \frac{3 - 2,75}{1,2} \sqrt{\frac{0,625}{0,375}} = 0,27$$

Exemple: Correlació biserial-puntual. Ítem 5

Item5	Total-5
0	3
0	4
1	3
0	4
0	3
1	1
1	5
1	0

$$p = 0,5$$

$$q = 0,5$$

$$\bar{X}_p = \frac{3+1+5+0}{4} = 2,25$$

$$\bar{X}_x = 2,88$$

$$S_x^2 = \frac{85}{8} - 2,88^2 = 2,36 \longrightarrow S_x = 1,54$$

$$r_{bp} = \frac{2,25 - 2,88}{1,54} \sqrt{\frac{0,5}{0,5}} = -0,41$$

Exercici: Correlació biserial-puntual. Ítem 6

<i>P</i>	<i>Ítems</i>						<i>Total</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	
<i>1</i>	0	1	0	1	0	1	3
<i>2</i>	1	1	1	0	0	1	4
<i>3</i>	0	1	0	1	1	1	4
<i>4</i>	1	1	0	1	0	1	4
<i>5</i>	1	1	0	1	0	0	3
<i>6</i>	0	1	0	0	1	0	2
<i>7</i>	1	1	1	1	1	1	6
<i>8</i>	0	0	0	0	1	0	1
<i>Total</i>	4	7	2	5	4	5	

Correcció de la correlació

En cas de no haver eliminat l'ítem per calcular la correlació, caldrà aplicar la correcció següent:

$$Cr = \frac{r_{jx} S_x - S_j}{\sqrt{S_x^2 + S_j^2 - 2r_{jx} S_x S_j}}$$

On:

r_{jx} Coeficient de correlació entre l'ítem i el test

S_x Desviació típica del test

S_j Desviació típica de l'ítem

S_x^2 Variància del test

S_j^2 Variància de l'ítem

Nota: Cal recordar que la variància per variables dicotòmiques:

$$S_j^2 = p \cdot q$$

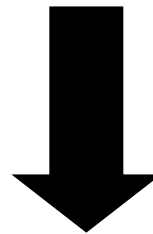
Exercici: Correlació biserial-puntual. Ítem 6 (sense restar puntuació de l'ítem – aplicant la correcció de la correlació)

Ítem 6	Total
1	3
1	4
1	4
1	4
0	3
0	2
1	6
0	1

Relacions entre indicadors: Dificultat – Variabilitat – Discriminació

Què ocorre si tothom encerta un ítem?

I si tothom el falla?



$$S_j^2 = 0$$

L'ítem no és útil per establir diferències entre persones.

No discrimina

Relacions entre indicadors: Dificultat – Variabilitat – Discriminació

En ítems dicotòmics (1 = encert i 0 = error)...

$$S_j^2 = p \cdot q$$

On:

p proporció d'encerts

q proporció d'errors (1-p)

S_j^2 màxima quan $p = 0,5$ i $q = 0,5 \rightarrow ID_j = 0,5$

S_x^2 s'aproxima al valor màxim quan els ID dels ítems que
composen el test són més propers a 0,5

Relacions entre indicadors: Dificultat – Variabilitat – Discriminació

Ítem pot discriminar si $S_j^2 > 0$

Mesura del poder discriminatiu d'un test $\rightarrow S_x^2$

$$S_x = \sum S_j \cdot r_{jx}$$

On:

S_x desviació típica del test

S_j desviació típica de l'ítem j

r_{jx} índex de discriminació (correlació ítem-test)

Relacions entre indicadors: Dificultat – Variabilitat – Discriminació

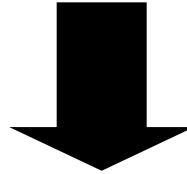
Els valors que pot prendre l'índex de discriminació d'un ítem depèn del valor de l'índex de dificultat.

Ítem fàcil → permet discriminar entre les persones amb nivells baixos en el tret mesurat (puntuacions baixes)

Ítem difícil → permet discriminar entre les persones amb nivells elevats en el tret mesurat (puntuacions altes)

Relacions entre indicadors: Dificultat – Variabilitat – Discriminació

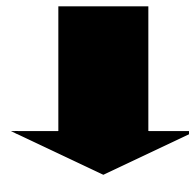
Major capacitat discriminativa del test quan conté:



- Majoria d'ítems de dificultat mitjana (en ítems dicotòmics ID al voltant de 0,5)
- Alguns ítems fàcils per discriminar entre persones amb menys nivell
- Alguns ítems difícils per discriminar entre persones amb més nivell

Anàlisi dels distractors. Respostes d'elecció múltiple

Els distractors han de ser igualment atractius per als individus que no coneguin la resposta



Tenir la mateixa probabilitat de ser escollits a l'atzar quan l'individu intenta endevinar la resposta correcta.

Exemple: Anàlisi dels distractors

		Alternatives					Total
		A	B	C	D	E	
Ítem	50% superior	5	7	47	7	4	70
	50% inferior	11	9	32	8	10	70
Percentatge		11,43%	11,43%	56,43%	10,71%	10,00%	

		Alternatives					Total
		A	B	C	D	E	
Ítem	50% superior	0	20	9	8	33	70
	50% inferior	2	15	14	18	21	70
Percentatge		1,43%	25,00%	16,43%	18,57%	38,57%	

Exercici: Anàlisi dels distractors

		<i>Alternatives</i>					<i>Total</i>
		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
<i>Ítem</i>	<i>50% superior</i>	6	93	6	41	4	150
	<i>50% inferior</i>	7	59	37	5	42	150

Exemple: Anàlisi dels distractors a partir de D

	Alternatives				
	A	B	C	D	Total
50% superior	14	55	18	13	100
50% inferior	20	35	26	19	100

Índex de Discriminació de l'alternativa A:

$$D = \frac{n_s}{N_s} - \frac{n_i}{N_i} = \frac{14}{100} - \frac{20}{100} = -0,06$$

Índex de Discriminació de l'alternativa C:

$$D = \frac{n_s}{N_s} - \frac{n_i}{N_i} = \frac{18}{100} - \frac{26}{100} = -0,08$$

Índex de Discriminació de l'alternativa D:

$$D = \frac{n_s}{N_s} - \frac{n_i}{N_i} = \frac{13}{100} - \frac{19}{100} = -0,06$$

Exercici: Anàlisi dels distractors a partir de D

	Alternatives				
	A	B	C	D	Total
50% Grup superior	12	3	40	45	100
50% Grup inferior	27	20	25	28	100

Índex de Fiabilitat de l'ítem

Estimació de la fiabilitat amb la que l'ítem mesura la característica que mesura el test en conjunt

$$IF = S_j D_j$$

On:

S_j desviació típica de les puntuacions en l'ítem j

D_j índex de discriminació de l'ítem j

Índex de Validesa de l'ítem

- Contribució de l'ítem en la predicció d'una variable externa (criteri)
- Correlació de l'ítem amb un criteri extern
- El tipus de correlació està en funció del tipus d'ítem i el tipus de criteri

<i>Tipus de Correlacions</i>		
<i>Ítem</i>	<i>Criteri</i>	<i>Correlació</i>
Quantitativa	Quantitativa	Pearson r
Dicotòmica	Quantitativa	Biserial-Puntual r_{bp}
Dicotomitzada	Quantitativa	Biserial r_b
Dicotòmica	Dicotòmica	Fi ϕ
Dicotomitzada	Dicotomitzada	Tetracòrica r_t

Exemple: Índex de Validesa de l'ítem

Ítem dicotòmic d'un qüestionari de depressió (0 = no depressió, 1 = depressió)
Criteri extern: puntuació total d'un test d'ansietat

Ítem depressió	Test d'ansietat
0	3
1	12
1	10
0	2
1	15
0	4
0	3
1	11
1	15
1	16
1	18
0	0

$$p = \frac{7}{12} = 0,583$$

$$q = \frac{5}{12} = 0,417$$

$$\bar{x}_p = \frac{12 + 10 + 15 + 11 + 15 + 16 + 18}{7} = 13,86 \quad \bar{x}_x = \frac{109}{12} = 9,08$$

$$S_x^2 = \frac{3^2 + 12^2 + 10^2 + 2^2 + 15^2 + 4^2 + 3^2 + 11^2 + 15^2 + 16^2 + 18^2 + 0^2}{12} - 9,08^2 = 36,97$$

$$S_x = 6,08$$

$$r_{bp} = \frac{13,86 - 9,08}{6,08} \sqrt{\frac{0,583}{0,416}} = 0,93$$

Fiabilitat



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement- No comercial- Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).



Lectura:

Prieto, G., y Delgado, A.R. (2010). Fiabilidad y validez. *Papeles del Psicólogo*, 31(1), 67-74.

Fiabilitat: Valora la **consistència** i la **precisió** de la mesura.

Teoria dels tests: Analitza la incidència dels errors en la mesura dels constructes psicològics i el grau en què afecten la **fiabilitat** dels instruments de mesura.

Tipus d'errors

- **Error conceptual:** ocorre quan mesurem una cosa pensant que mesurem una altra. Validesa.
- **Error sistemàtic:** aquell que afecta sempre de la mateixa manera a la mesura. Provoca el biaix en la mesura.
- **Error aleatori:** provocat per factors múltiples i incontrolats.

Quan les puntuacions d'un test seran totalment fiables?



Quan les mesures que s'obtinguin estiguin **lliures d'error de mesura**

Puntuació observada = Puntuació verdadera

Què és un error de mesura?



La diferència entre la puntuació observada i la puntuació verdadera. La causa d'aquesta diferència és l'error aleatori.

$$e = X - V$$

Totes les diferències en un procés de mesura es poden atribuir a l'error aleatori?



No, cal valorar l'estabilitat del constructe que es mesura (ex. capacitat visuoespacial vs temps de reacció)

Definició fiabilitat

Fiabilitat: Absència d'error de mesura (consistència o estabilitat de les mesures quan el procés de mesura es repeteix)

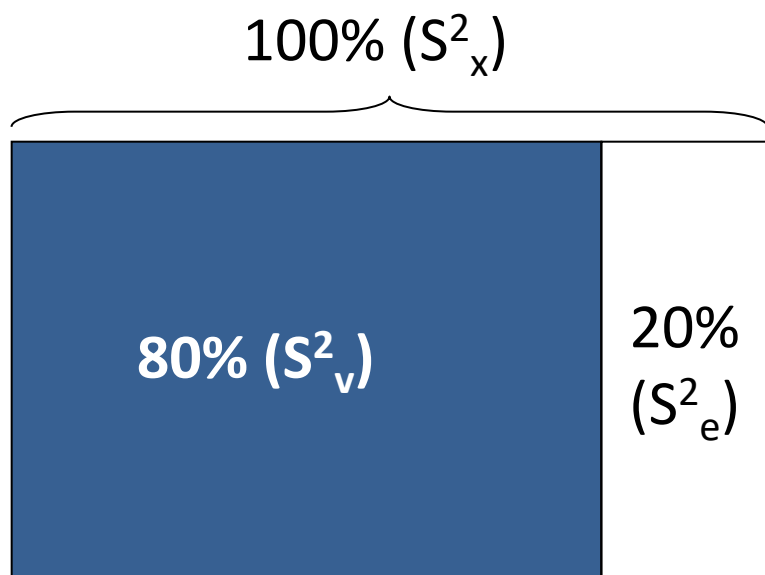
$$r_{xx} = \frac{S_V^2}{S_x^2}$$

On:

- variància vertadera del test
- variància obtinguda del test

Què indica un coeficient de fiabilitat?

$$r_{xx} = 0,80$$

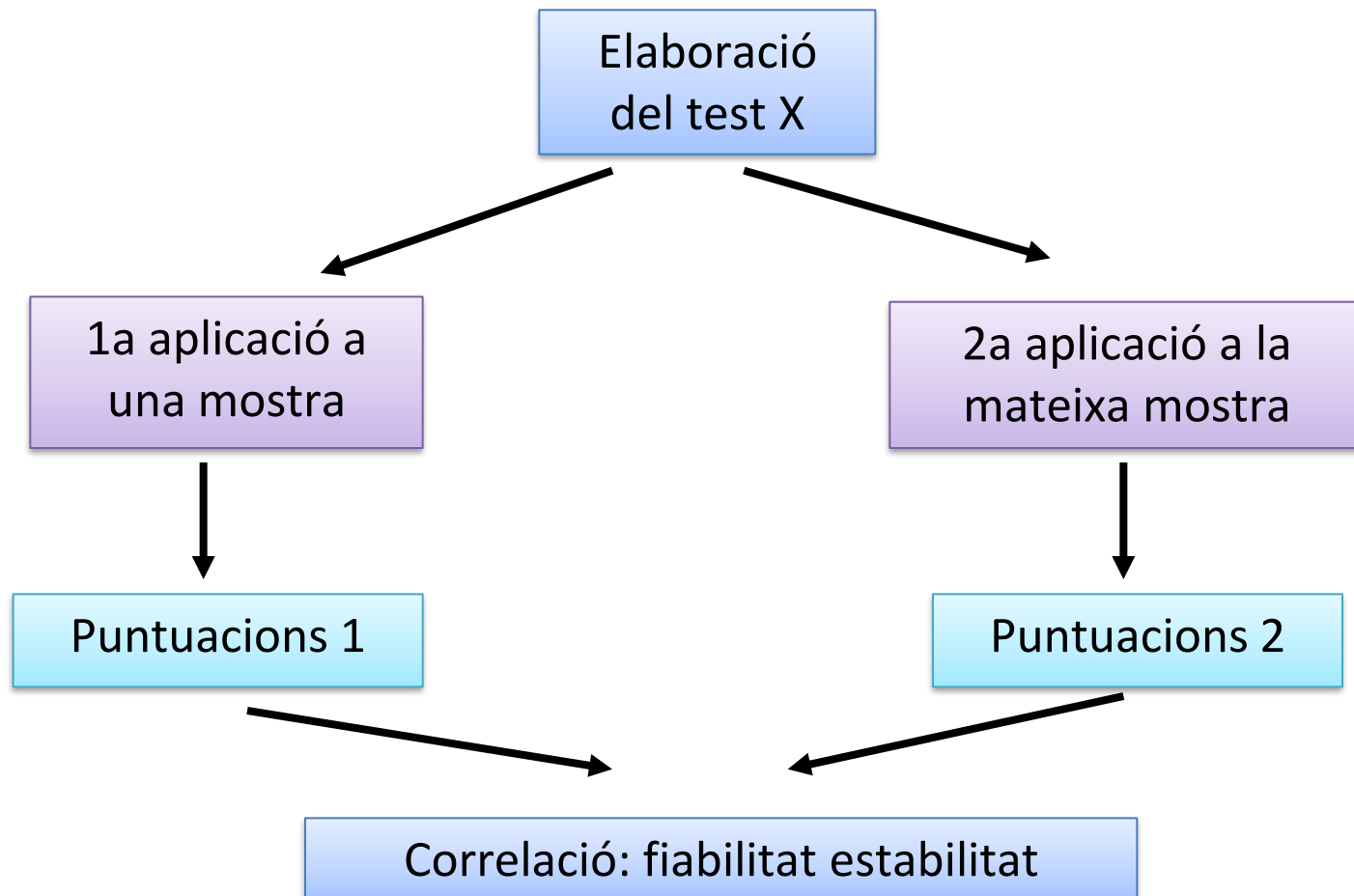


Del 100% de la variància obtinguda (S^2_x), el 80% és deguda a la variància vertadera (S^2_v) i per tant, el 20% restant és deguda a la variància d'error (S^2_e)

Procediments per estimar la fiabilitat

Estratègia	Valoració de la...	Procediment d'anàlisi
Test-retest	Estabilitat temporal de la mesura	• Coeficient de correlació de Pearson • Coeficient de correlació intraclass
Formes paral·leles	Equivalència de la mesura	• Coeficient de correlació de Pearson • Coeficient de correlació intraclass
Consistència interna	Consistència interna de la mesura	• Dues meitats - Correcció Spearman-Brown • Coeficient alfa de Cronbach (α) • Coeficient omega de McDonald (ω)

Fiabilitat test-retest



Coeficient de correlació de Pearson vs Coeficient de correlació intraclasse

Individus	T1	T2	T3
1	28	27	33
2	29	30	35
3	25	25	27
4	31	30	38
5	27	28	32
6	26	26	29
7	32	31	41
8	30	29	36
9	28	27	33
10	31	29	38

Matriu de correlacions (Pearson)

		T1	T2	T3
T1	Pearson's r	—		
T2	Pearson's r	0,910	—	
T3	Pearson's r	0,993	0,933	—

Matriu de correlacions (CCI)

		T1	T2	T3
T1	CCI	—		
T2	CCI	0,881	—	
T3	CCI	0,367	0,269	—

Criteris valoració fiabilitat test-retest basats en el coeficient de Correlació de Pearson (estabilitat mesura)

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.

Inadequada: $r < 0,55$

Adequada però amb algunes mancances: $0,55 \leq r < 0,65$

Adequada: $0,65 \leq r < 0,75$

Bona: $0,75 \leq r < 0,80$

Excel·lent: $r \geq 0,80$

Criteris valoració fiabilitat test-retest basats en el coeficient de Correlació Intraclass (estabilitat mesura)

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.

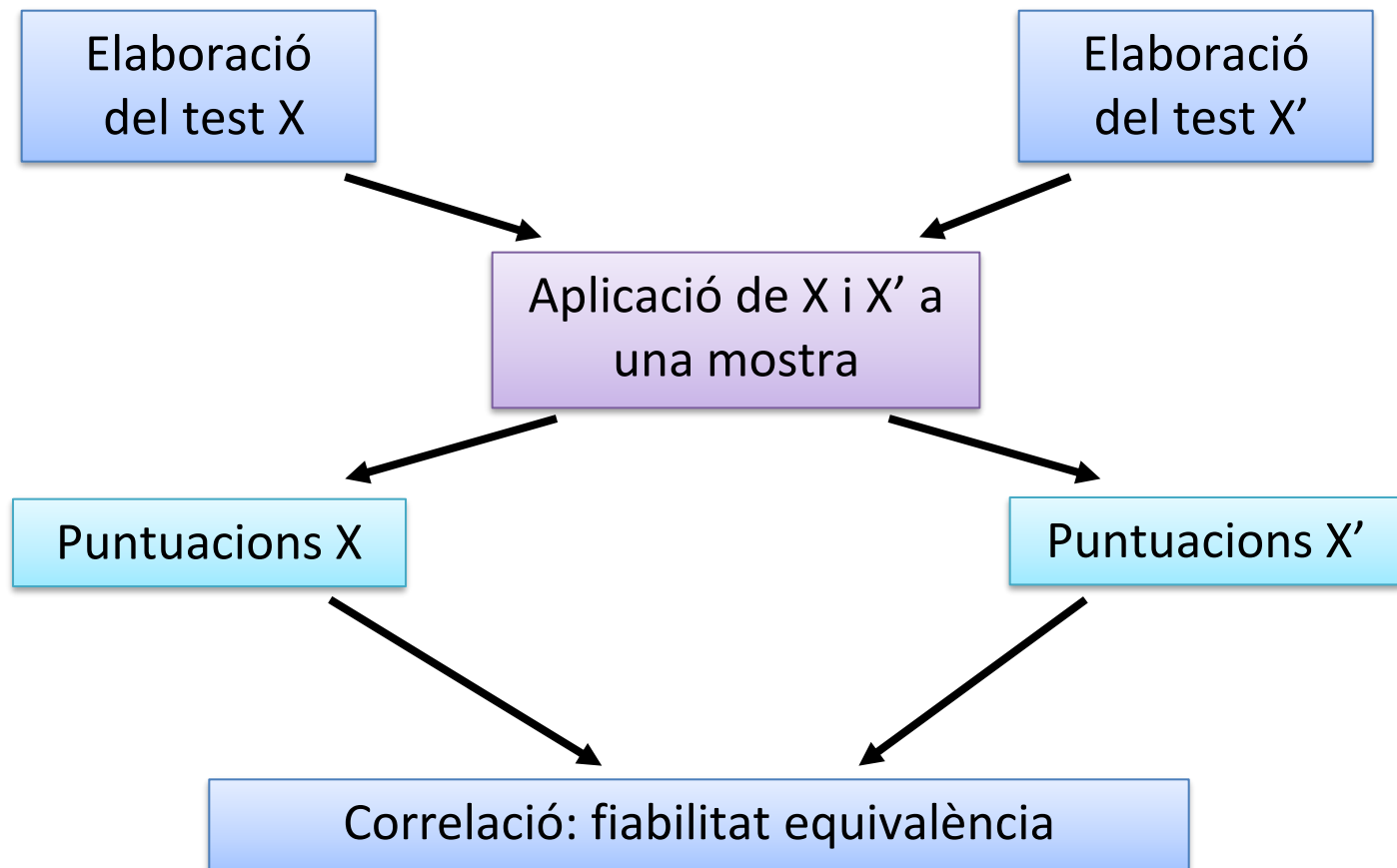
Inadequada: $CCI < 0,50$

Moderada: $0,50 \leq CCI < 0,75$

Bona: $0,75 \leq CCI < 0,90$

Excel·lent: $CCI \geq 0,90$

Fiabilitat formes paral·leles



Criteris valoració fiabilitat formes paral·leles (equivalència mesura)

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.

Inadequada: $r < 0,50$

Adequada però amb algunes mancances: $0,50 \leq r < 0,60$

Adequada: $0,60 \leq r < 0,70$

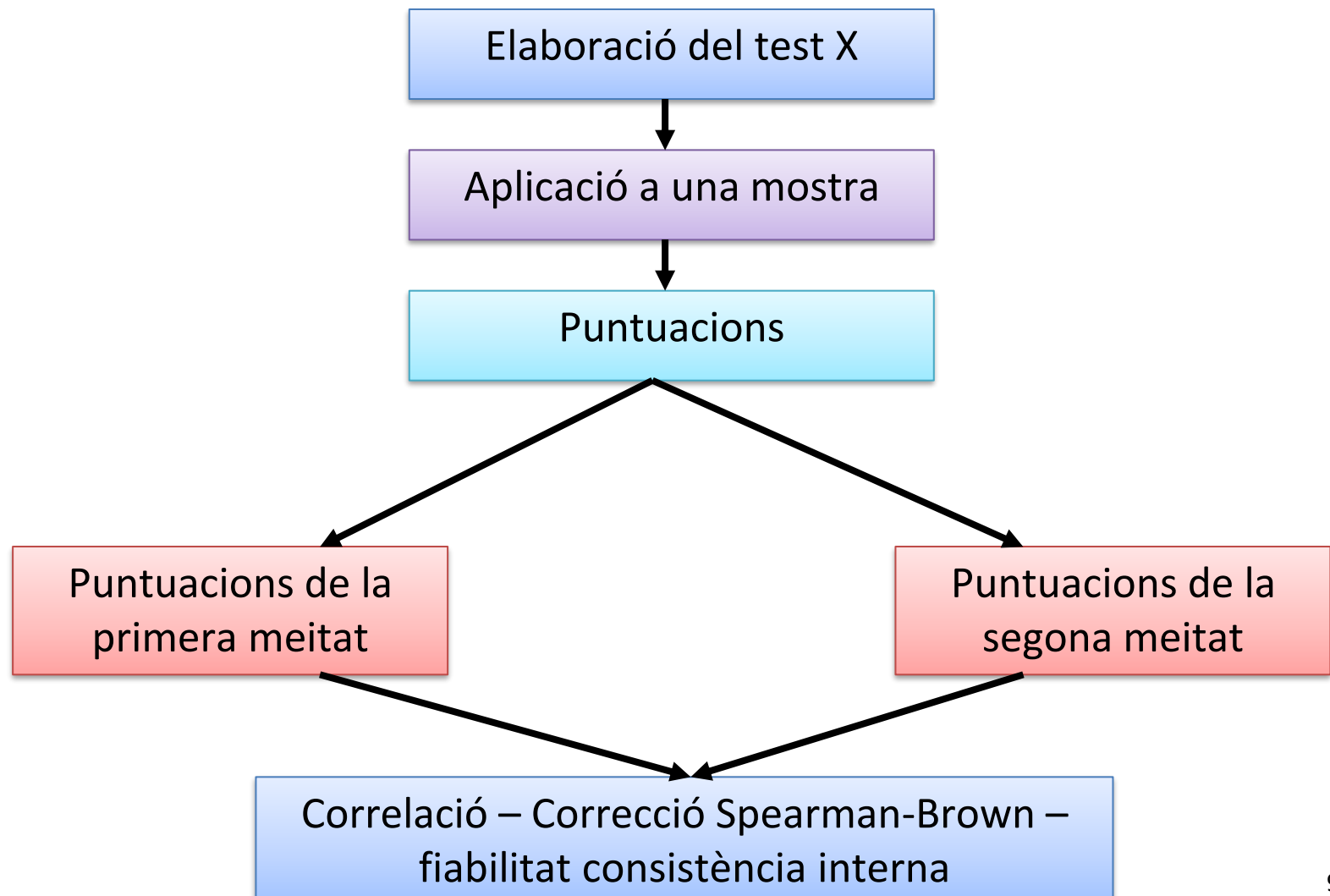
Bona: $0,70 \leq r < 0,80$

Excel·lent: $r \geq 0,80$

Consistència interna

- Fiabilitat dues meitats a partir del càlcul de la Correcció Spearman-Brown
- Coeficient alfa de Cronbach (α)
- Coeficient omega de McDonald (ω)

Fiabilitat dues meitats



Quins ítems han de formar part de cada una de les dues meitats?

Ítems meitat A	Ítems meitat B	Correlació entre les dues meitats
1, 2, 3	4, 5, 6	0,559
1, 2, 4	3, 5, 6	0,645
1, 2, 5	3, 4, 6	0,474
1, 2, 6	3, 4, 5	0,586
1, 3, 4	2, 5, 6	0,653
1, 4, 5	2, 3, 6	0,730
1, 5, 6	2, 3, 4	0,507
2, 3, 5	1, 4, 6	0,426
2, 4, 5	1, 3, 6	0,824
2, 4, 6	1, 3, 5	0,546



Correcció d'Spearman-Brown - dues meitats

$$r_{xx} = \frac{2 r_{x_1 x_2}}{1 + r_{x_1 x_2}}$$

On:

r_{xx} – Fiabilitat del test sencer

$r_{x_1 x_2}$ – Correlació del subtest 1 i el subtest 2

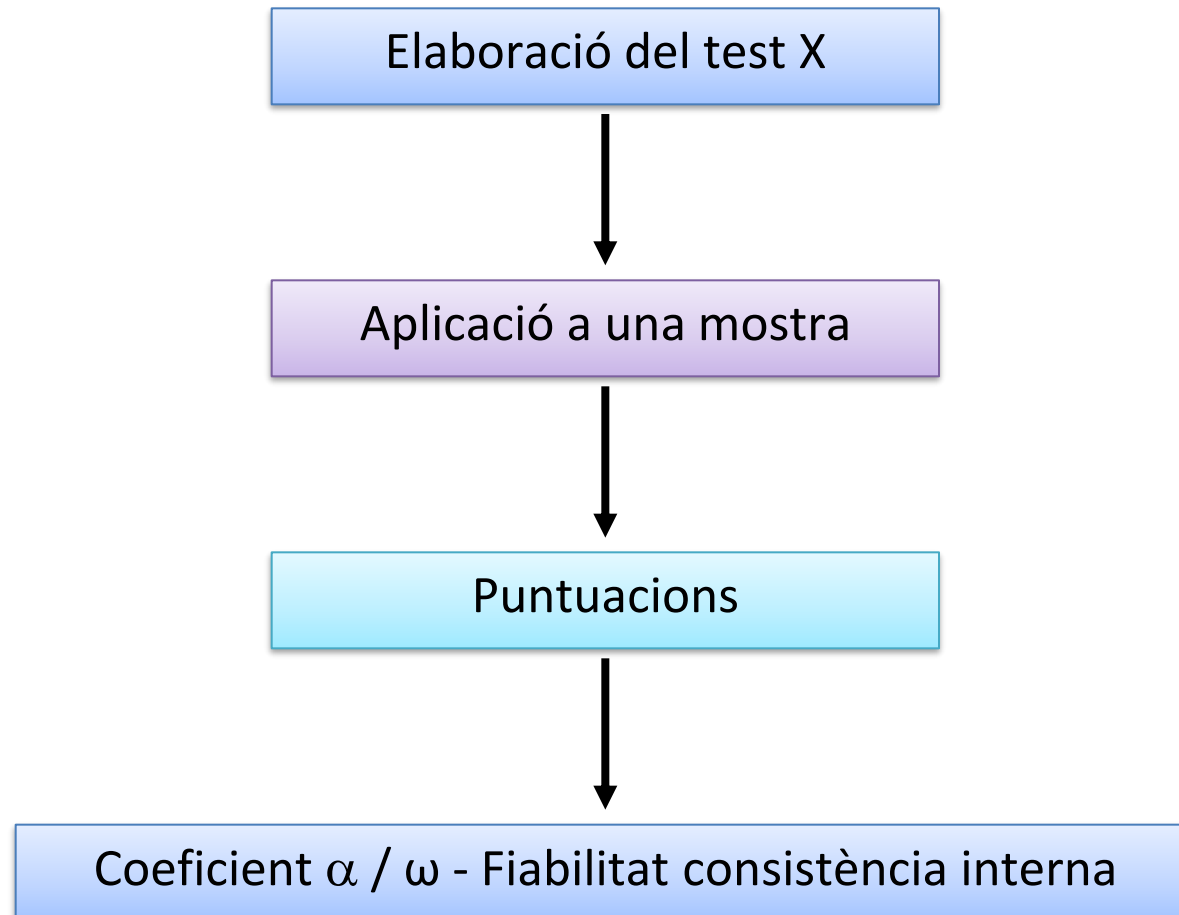
Exemple Correcció d'Spearman-Brown - dues meitats

Aplicació de la correcció d'Spearman-Brown a les dades de l'exemple anterior:

$$r_{x_1x_2} = 0,546$$

$$r_{xx} = \frac{2 \cdot 0,546}{1 + 0,546} = 0,706$$

Fiabilitat Alfa de Cronbach i Omega de McDonald



Coeficient alfa (Cronbach, 1951)

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum S_j^2}{S_x^2} \right)$$

On:

n nombre d'ítems del test

$\sum S_j^2$ suma de les variàncies dels n ítems

S_x^2 variància de la puntuació total del test

Exemple coeficient alfa

Quin serà el valor del coeficient alfa si $n = 40$,
 $\sum S_j^2 = 2$ i $S_x^2 = 6$?

$$\alpha = \frac{40}{40 - 1} \left(1 - \frac{2}{6} \right) = 0,68$$

Variacions del coeficient alfa: Kuder-Richardson (1937)

Aquest és el coeficient alfa per a ítems dicotòmics:

$$\alpha = KR20 = \frac{n}{n-1} \left(1 - \frac{\sum p_j \cdot q_j}{S_x^2} \right)$$

On:

n nombre d'ítems del test

p_j proporció de respostes positives en l'ítem

q_j proporció de respostes negatives en l'ítem

S_x^2 variància de la puntuació total del test

Coeficient Omega (McDonald, 1999)

Preferible quan els supòsits del coeficient alfa de Cronbach no es compleixen.

$$\omega = \frac{S_v^2}{S_x^2} = \frac{\left(\sum_{j=1}^J \lambda_j\right)^2}{S_x^2}$$

On:

S_v^2 variància de la puntuació vertadera del test

S_x^2 variància de la puntuació del test

λ_j pes o càrrega factorial, el qual representa la relació entre l'ítem i la variable latent

Exemple coeficient Omega

Quin serà el valor del coeficient Omega si $\sum_{j=1}^J \lambda_j = 4,547$ i $S^2_x = 25$?

Ítems	Càrregues factorials
Ítem 1	0,451
Ítem 2	0,364
Ítem 3	0,446
Ítem 4	0,325
Ítem 5	0,556
Ítem 6	0,734
Ítem 7	0,562
Ítem 8	0,513
Ítem 9	0,596
Total	4,547

$$\omega = \frac{(4,547)^2}{25} = 0,83$$

Criteris valoració fiabilitat consistència interna

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.

Inadequada: $r < 0,60$

Adequada però amb algunes mancances: $0,60 \leq r < 0,70$

Adequada: $0,70 \leq r < 0,80$

Bona: $0,80 \leq r < 0,85$

Excel·lent: $r \geq 0,85$

Contribució dels ítems a la consistència interna

- Cada ítem que forma part d'un instrument de mesura té una contribució particular a la consistència interna d'aquest
- Podem calcular el coeficient α per al conjunt d'ítems del test eliminant prèviament del càlcul les dades de l'ítem que volem avaluar

$$\alpha = 0,73$$

$$\omega = 0,74$$

	Alfa sense l'ítem	Omega sense l'ítem
Ítem 1	0,71	0,72
Ítem 2	0,68	0,69
Ítem 3	0,70	0,71
Ítem 4	0,76	0,77
Ítem 5	0,71	0,72

Factors que afecten la fiabilitat del test

Factors	Més fiable	Menys fiable
Longitud del test	Llarg	Curt
Variabilitat	Àmplia	Estreta
Tipus d'ítem	Puntuació objectiva	Puntuació subjectiva
Qualitat de l'ítem	Estadísticament fort	Estadísticament feble
Instruccions	Clares	Vagues
Límit de temps	Adaptat al nivell dels subjectes	Massa curt o llarg

Adaptat de: The National Council on Measurement in Education (NCME). Instructional Topics in Educational Measurement Series (ITEMS): Module 01: Reliability in Classical Test Theory.

Formes d'incrementar la fiabilitat del test

- Augmentar el nombre d'ítems – Fórmula d'Spearman-Brown
- Eliminar ítems problemàtics (baixes correlacions ítem-test)
- Millorar les condicions d'aplicació

Relació entre fiabilitat i longitud del test

Fórmula d'Spearman-Brown

$$R_{xx} = \frac{k r_{xx}}{1 + (k - 1) r_{xx}}$$

On:

R_{xx} fiabilitat del test allargat

r_{xx} fiabilitat del test abans d'allargar-lo

k nombre de vegades que hem allargat el test:

$$k = \frac{n_i + n_a}{n_i}$$

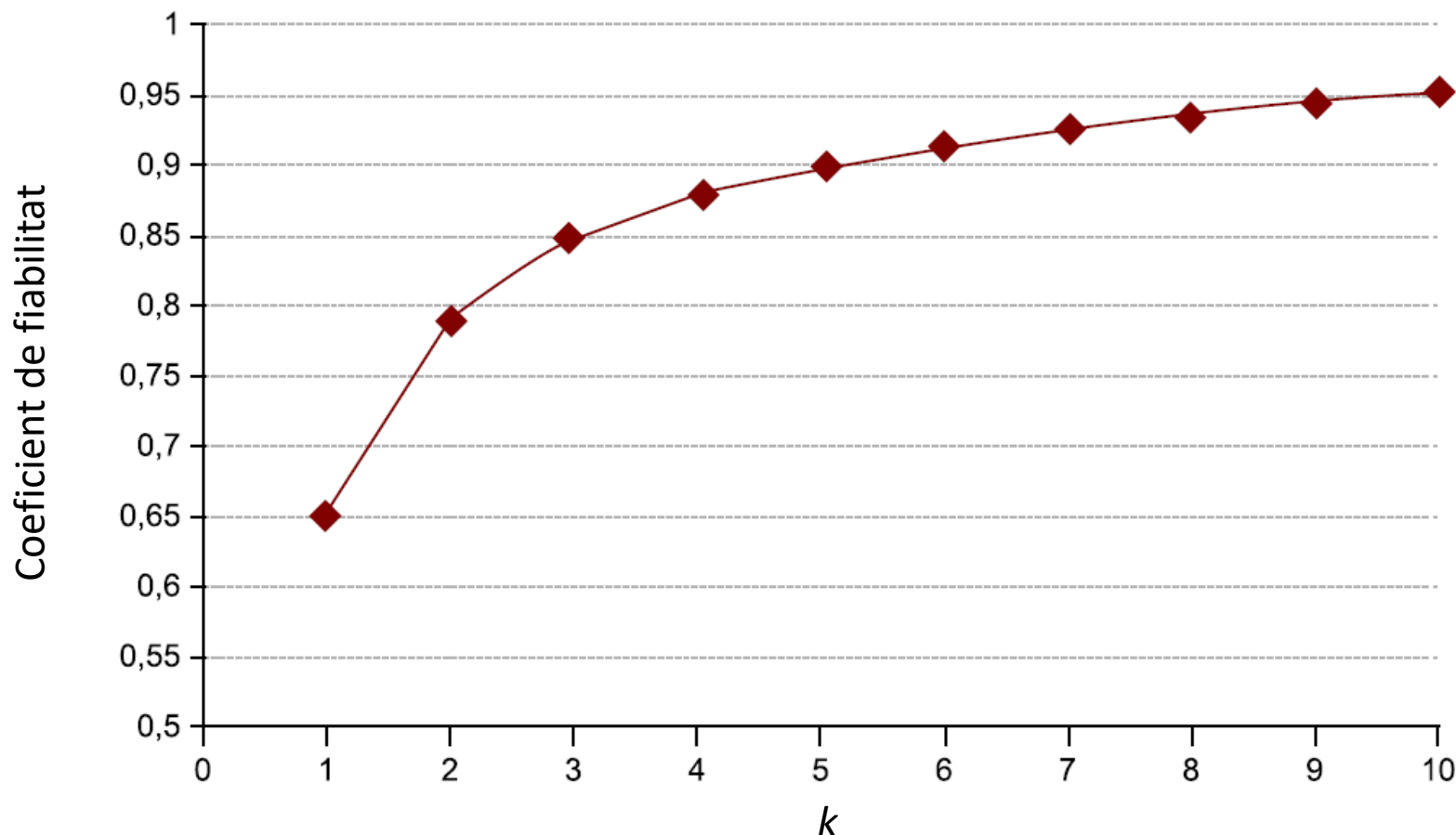
On:

n_i - nombre d'ítems inicials

n_a - nombre d'ítems afegits

Relació entre fiabilitat i longitud del test

Figura 1. Relación entre el coeficiente de fiabilidad y el aumento de ítems en un test



Exemple de relació entre fiabilitat i longitud del test

Suposem que la fiabilitat d'un test amb 20 ítems és de 0,70.
Quina seria la seva fiabilitat si afegim 40 ítems al test?

$$k = \frac{20 + 40}{20} = 3$$

$$R_{xx} = \frac{3 \cdot 0,7}{1 + (3 - 1) 0,7} = 0,88$$

Longitud del test per aconseguir una fiabilitat determinada

$$k = \frac{R_{xx} (1 - r_{xx})}{r_{xx} (1 - R_{xx})}$$

On:

R_{xx} fiabilitat que es pretén aconseguir

r_{xx} fiabilitat del test original

k nombre de vegades que hem allargat el test:

$$k = \frac{n_i + n_a}{n_i}$$

On:

n_i - Nombre d'ítems inicials

n_a - Nombre d'ítems afegits

Exemple de longitud del test per aconseguir una fiabilitat determinada

Suposem que un test amb 10 ítems té una fiabilitat de 0,62. Quants ítems caldrà afegir per aconseguir una fiabilitat de 0,75?

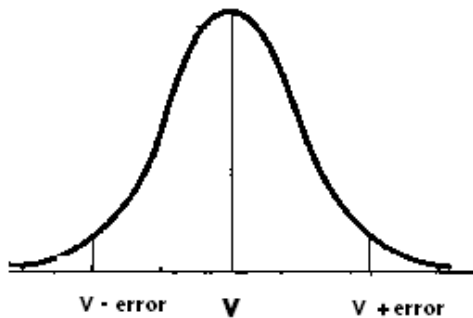
$$k = \frac{0,75 (1 - 0,62)}{0,62 (1 - 0,75)} = 1,84$$

$$1,84 = \frac{10 + n_a}{10} \rightarrow n_a = 1,84 \cdot 10 - 10 = 8,4 \approx 9 \text{ ítems}$$

Exercici: Longitud del test per aconseguir una fiabilitat determinada

Suposem que un test amb 50 ítems té una fiabilitat de 0,75. Quants ítems caldrà afegir per aconseguir una fiabilitat de 0,90?

Estimació per interval de la puntuació vertadera – Error típic de mesura



$$V_{\text{int}} = x_i \pm z \cdot S_e$$

On:

x_i – Puntuació obtinguda pel subjecte en el test

S_e – Error típic de mesura:

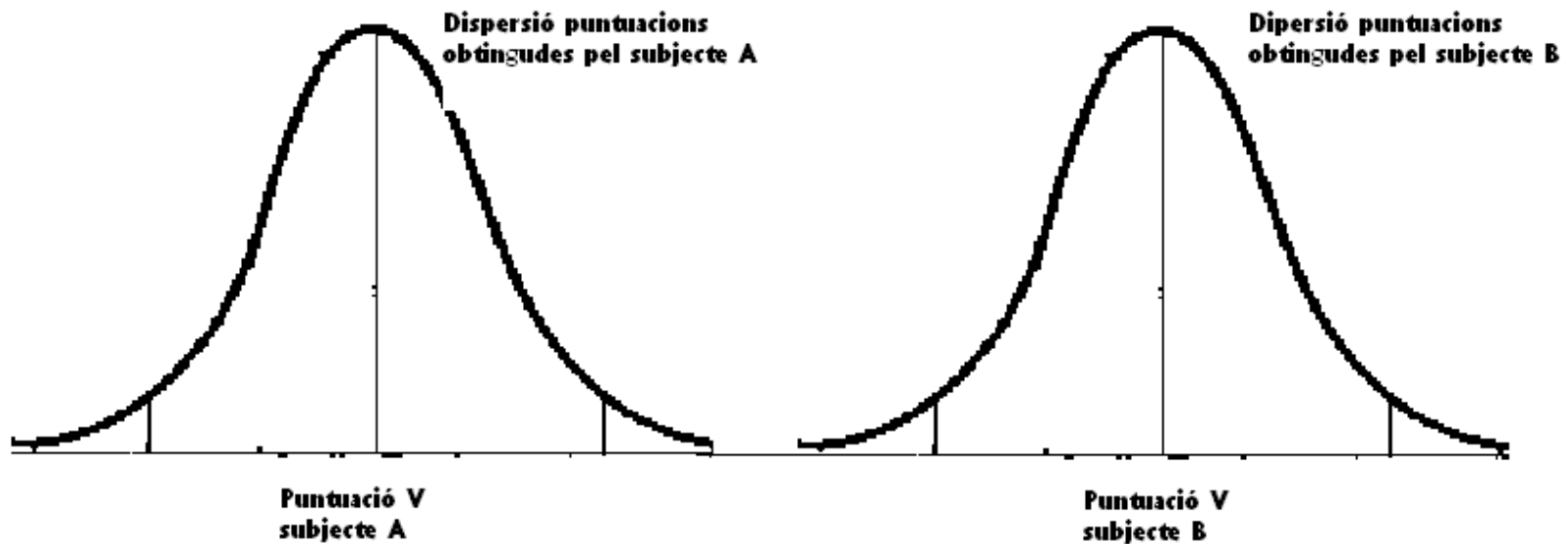
$$S_e = S_x \sqrt{1 - r_{xx}}$$

On:

S_x – desviació típica del test

r_{xx} – coeficient fiabilitat del test

Error típic de mesura - TCT



Desviació típica dels errors la mateixa o molt semblant per
totes les persones $\rightarrow S_e$

Aplicacions de l'error típic de mesura - TCT

1. Estimació per interval de la puntuació vertadera
2. Comparació de les puntuacions observades de dues persones
3. Valoració del canvi de la puntuació observada d'una persona després d'una intervenció

Exemple: Estimació per interval de la puntuació vertadera – Error típic de mesura

Suposem que el coeficient de fiabilitat d'un test és de 0,75, la seva mitjana 80 i la desviació típica 16. Quin és l'error típic de mesura? En quin interval se situa la puntuació vertadera d'un subjecte que té una puntuació observada de 86 a un nivell de confiança del 95% ($z = 1,96$)?

$$S_e = 16 \sqrt{1 - 0,75} = 8$$

$$V_{int} = 86 \pm 1,96 \cdot 8 \rightarrow 70,32 \leq V \leq 101,68$$

Exercici: Estimació per interval de la puntuació vertadera – Error típic de mesura

Suposem que el coeficient de fiabilitat d'un test és de 0,80, la seva mitjana 15 i la variància 9. Quin és l'error típic de mesura? En quin interval se situa la puntuació vertadera d'un subjecte que té una puntuació observada de 12, a un nivell de confiança del 95% ($z = 1,96$)?

Exercici: Comparar puntuacions observades de dues persones

En l'escala d'Extraversió d'una prova dues persones obtenen unes puntuacions observades de 12 i 14 punts, respectivament. Sabem que la desviació típica del test és de 2,7 i el coeficient de fiabilitat de 0,80.

Fins a quin punt podem concloure que ambdues persones difereixen en el seu nivell d'extraversió (NC = 95%; $z = 1,96$)?

Exercici: Fiabilitat

Un test de 32 ítems s'aplica a 200 subjectes i s'obté una mitjana de 23 i una desviació típica de 4. Al dividir el test en ítems parells i imparells, la correlació entre les dues meitats és de 0,91.

- 1) Quin és el coeficient de fiabilitat del test? Interpreta aquest coeficient i comenta si el valor és o no acceptable.
- 2) Quants ítems caldria afegir a aquest test per a elevar la seva fiabilitat fins a 0,99? Val la pena modificar el test per a augmentar la fiabilitat fins a aquest nivell?
- 3) Quant val el coeficient alfa d'aquest test, tenint en compte que la suma de les variàncies dels seus ítems és 1,25? Interpreta aquest coeficient i comenta si el valor és acceptable.
- 4) En el test original un subjecte ha arribat a una puntuació de 26. Entre quines puntuacions s'espera que es trobi la puntuació veritable del subjecte al nivell de confiança del 95%? Interpreta l'interval obtingut.

Validesa



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement - No comercial - Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).



Lectures:

Arias, A., y Sireci, S. G. (2021). Validez y validación para pruebas educativas y psicológicas. *Revista Iberoamericana de Psicología* , 14(1), 11-22.

Validesa – Les inferències o interpretacions que fem a partir del test s'adeqüen al propòsit del test.

Validació – Procés mitjançant el qual s'acumulen diverses facetes o evidències de validesa:

Històric:

- De contingut

- De criteri

- De constructe

Estàndards 1999 i 2014:

- Contingut

- Procés de resposta

- Estructura interna

- Relacions amb altres variables

- Conseqüencial

Facetes o fonts d'evidència de validesa:



Contingut – domini conductual



Procés de resposta – procés cognitiu al respondre



Estructura interna – relació entre ítems i dimensions



Relacions amb altres variables – relació amb variables externes



Conseqüencial – conseqüències de l'ús del test

Contingut

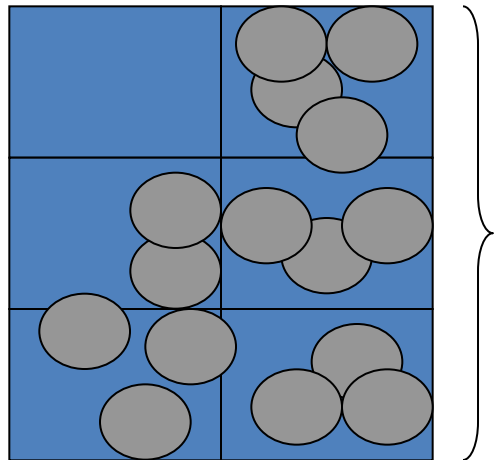
Rellevància

Contingut inclou informació directament relacionada amb el constructe

Representativitat

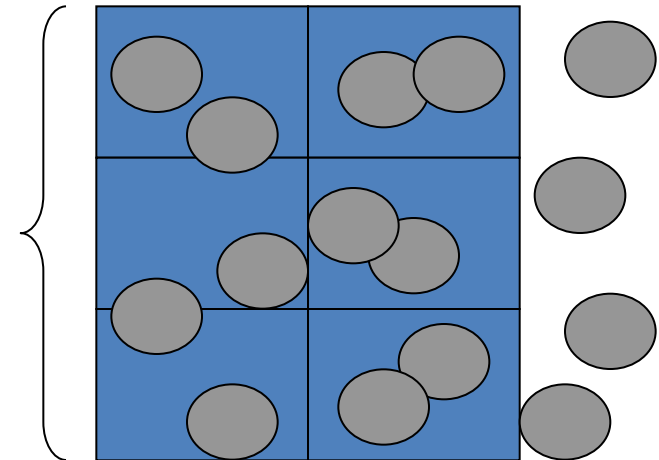
Informació complementària a l'aportada per la resta d'ítems

Falta de representativitat



Definició domini

Falta de rellevància



Contingut

Passos:

1. Definir i especificar el camp o domini del test.
2. Selecció de jutges experts.
3. Procés d'aparellament d'ítems amb els temes d'interès de mesura del test (representativitat).
4. Valoració de la rellevància.
5. Selecció dels ítems adequats.

Exemple: Contingut

Prova per avaluar els coneixements de Psicometria

Continguts a mostrejar: Teoria dels tests / Anàlisi d'ítems / Fiabilitat / Validesa / Baremació / Manual del test

Experts: professors de Psicometria de la UAB

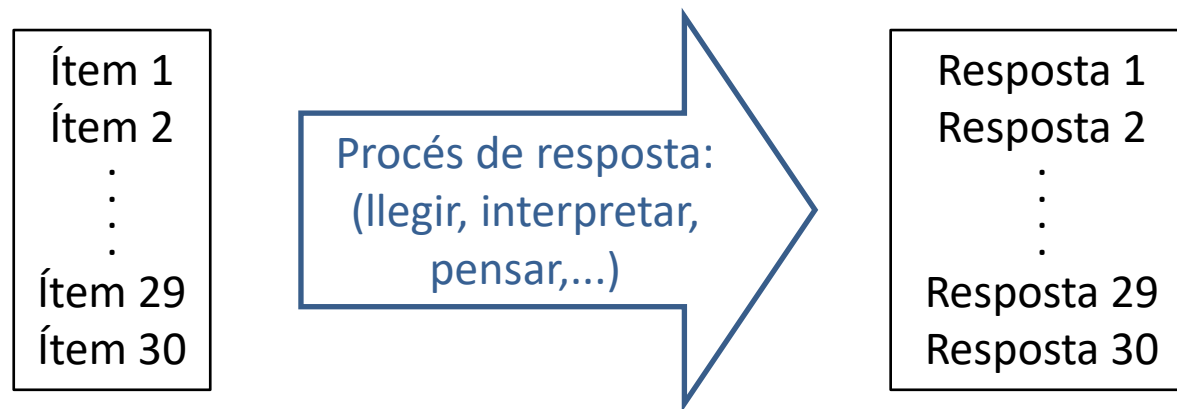
Aparellament: fer correspondre cada ítem amb cada un dels continguts (p.ex.: ítem 1 -> Baremació, ítem 2 -> Validesa,...) - Representativitat

Valoració: els experts valoren la rellevància dels ítems (p.ex.: puntuant-los de l'1 al 5)

Selecció d'ítems: seleccionar els ítems adequats (més ben valorats pels jutges experts) tenint en compte la importància relativa de cada contingut (p.ex.: 1 ítem de TCT, 5 d'Anàlisi d'ítems,...)

Procés de resposta

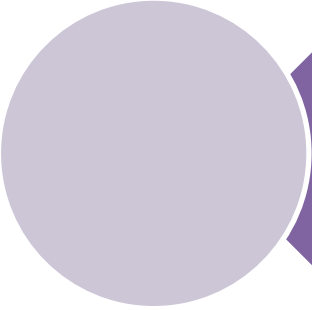
El procés que segueixen les persones per respondre el test és el previst.



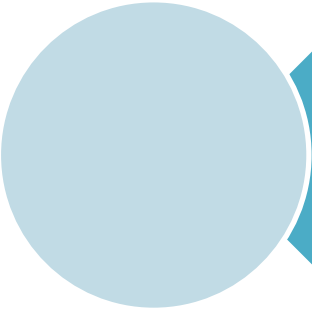
Estratègies: observació, entrevista cognitiva, *thinking aloud*, anàlisi dels patrons de resposta, etc.

Estructura interna

Les relacions dels ítems entre sí, dels ítems amb les dimensions i de les dimensions entre sí estan en consonància amb el model teòric previst per definir el constructe que es pretén mesurar.



Estructura interna unidimensional: el test ofereix una única puntuació perquè pretén mesurar un constructe unidimensional



Estructura interna multidimensional: el test ofereix més d'una puntuació perquè pretén mesurar un constructe multidimensional

Estructura interna

Passos:

1. Definir i formular hipòtesis sobre el constructe.

2. Recollir dades per comprovar les hipòtesis.

3. Consistència de les dades amb les hipòtesis formulades o altres alternatives.

Estructura interna: Anàlisi factorial

Tècnica de reducció de dades:

n variables $\rightarrow r$ factors ($r < n$)

MATRIU DE
DADES



MATRIU DE
CORRELACIONS



MATRIU
FACTORIAL

		Variables				
		1	2		n	
Subj.	1	X_{11}	X_{12}	$\cdot \cdot \cdot$	X_{1n}	
	2	X_{21}				
	$\cdot$					
	$\cdot$					
	$\cdot$					
	N	X_{N1}	$\cdot \cdot \cdot \cdot \cdot$	X_{Nn}		

		Variables				
		1	2		n	
Var.	1	r_{11}	r_{12}	$\cdot \cdot \cdot$	r_{1n}	
	2	r_{21}				
	$\cdot$	$\cdot$				
	$\cdot$	$\cdot$				
	$\cdot$	$\cdot$				
	n	r_{n1}	$\cdot \cdot \cdot \cdot \cdot$	r_{nn}		

		Factors				
		1	2		r	
Var.	1	a_{11}	a_{12}	$\cdot \cdot \cdot$	a_{1r}	
	2	a_{21}				
	$\cdot$	$\cdot$				
	$\cdot$	$\cdot$				
	$\cdot$	$\cdot$				
	n	a_{n1}	$\cdot \cdot \cdot \cdot \cdot$	a_{nr}		

Estructura interna: Anàlisi factorial

Suposem que construim un test per mesurar tres factors de manera que:

Factor 1: ítems 1 i 2 Factor 2: ítems 3, 4 i 5 Factor 3: ítems 6 i 7

Estructura matriu factorial:

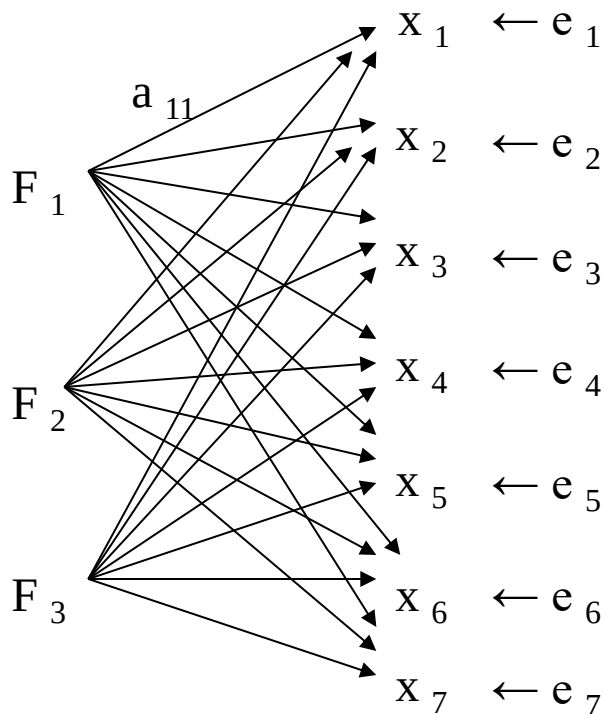
		AFE:		
		factors		
		1	2	3
ítems	1	a_{11}	a_{12}	a_{13}
	2	a_{21}	a_{22}	a_{23}
	3	a_{31}	a_{32}	a_{33}
	4	a_{41}	a_{42}	a_{43}
	5	a_{51}	a_{52}	a_{53}
	6	a_{61}	a_{62}	a_{63}
	7	a_{71}	a_{72}	a_{73}

		AFC:		
		factors		
		1	2	3
ítems	1	a_{11}	0	0
	2	a_{21}	0	0
	3	0	a_{32}	0
	4	0	a_{42}	0
	5	0	a_{52}	0
	6	0	0	a_{63}
	7	0	0	a_{73}

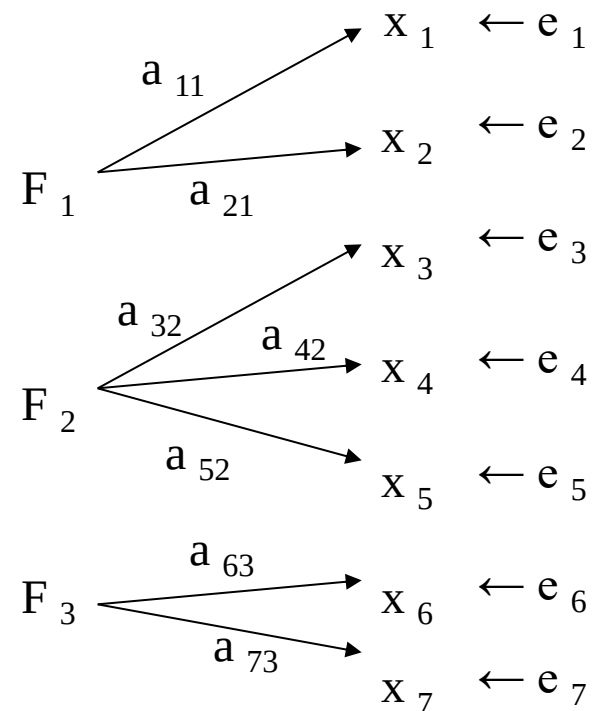
Estructura interna: Anàlisi factorial

Representació gràfica

AFE:



AFC:



Exemple: Anàlisi factorial exploratòria

“Se efectuó un análisis factorial exploratorio mediante extracción de ejes principales y rotación oblimin. Teniendo en cuenta el elevado número de ítems que se desarrollaron al inicio, se depuró la escala eliminando todos aquellos ítems que presentaran saturaciones inferiores a 0,35 o saturaciones complejas (superiores a 0,35 en más de un factor).”

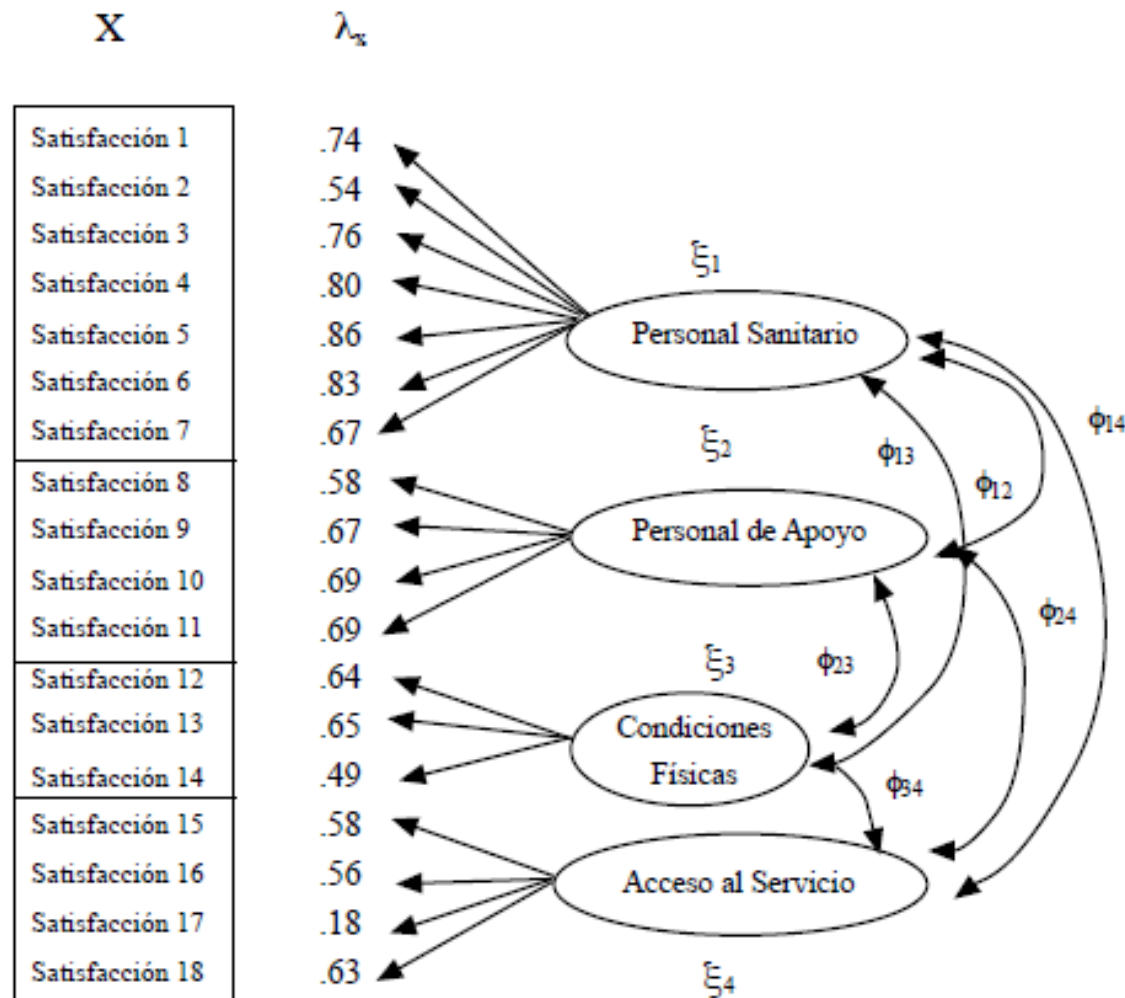
Exemple: Anàlisi factorial exploratòria

Matriz de saturaciones de los factores de la escala CONS-32				
Ítems	F1	F2	F3	F4
2. En el interior de la obra hay escaleras que permiten moverte de un modo seguro	0.60	0.00	0.20	0.06
3. Conozco las normas de seguridad de la obra	0.63	0.08	0.04	-0.07
5. Los andamios y las maderas están en mal estado	0.55	0.13	0.23	0.06
15. Se colocan los dispositivos para evitar la caída de personas	0.63	-0.17	0.03	0.08
16. Las zanjas están entibadas y protegidas	0.51	-0.17	0.08	-0.10
17. Los andamios tienen barandillas	0.71	-0.05	-0.21	0.08
22. Las escaleras tienen barandillas	0.64	0.02	0.01	0.13
30. Tenemos las medidas de seguridad a nuestro alcance (por ejemplo, redes para protegerse de la caída, arneses para sujetarnos, cascos, guantes, etc.)	0.58	-0.09	-0.02	0.08
4. Hago soldaduras con traje protector	0.19	-0.46	0.02	0.03
6. Al cortar madera uso guantes	-0.07	-0.74	0.06	-0.02
14. Al llegar a la obra por la mañana se revisa que la grúa, los andamios y el resto del material siga en perfecto estado	0.24	-0.42	0.09	0.15
8. Al cortar madera uso mascarilla de respiración	-0.07	-0.72	-0.08	0.02
25. Cuando trabajo con el martillo mecánico para taladrar el suelo, me protejo los oídos con cascos	0.21	-0.51	0.07	0.05
28. Uso gafas protectoras para cortar vidrio	-0.02	-0.58	0.20	0.03
31. Cuando corto baldosas utilizo guantes	0.02	-0.60	-0.02	-0.12
32. Entro en la obra con el casco puesto	0.33	-0.53	-0.07	0.02

Exemple: Anàlisi factorial confirmatòria

“El primero de los objetivos era comprobar si la estructura factorial de la escala era la esperada, con las 4 dimensiones descritas anteriormente. Para ello se llevó a cabo un análisis factorial confirmatorio mediante el programa AMOS 4. Los parámetros del modelo original fueron estimados siguiendo el criterio de Máxima Verosimilitud.”

Exemple: Anàlisi factorial confirmatòria



Exercici: AFE i AFC

Escala de Insomnio de Atenas

Edad:.....

Sexo:.....

Fecha:.....

Instrucciones: Esta escala pretende registrar su propia valoración acerca de cualquier dificultad que haya podido experimentar. Por favor, responda los siguientes ítems (marcando con un círculo el número apropiado), indicando las dificultades que le hayan ocurrido como mínimo tres veces a la semana durante el último mes.

1. Inducción del sueño (tiempo que tarda en dormirse después de apagar la luz)

0: Ningún problema 1: Ligeramente retrasado 2: Marcadamente retrasado 3: Muy retrasado o no durmió

2. Despertares durante la noche

0: Ningún problema 1: Problema menor 2: Problema considerable 3: Problema serio o no durmió

3. Despertar final antes de lo deseado

0: No fue antes 1: Un poco antes 2: Notablemente antes 3: Mucho antes o no durmió

4. Duración total del sueño

0: Suficiente 1: Ligeramente insuficiente 2: Notablemente insuficiente 3: Muy insuficiente o no durmió

5. Calidad general del sueño (no importa cuánto tiempo durmió)

0: Satisfactoria 1: Ligeramente insatisfactoria 2: Notablemente insatisfactoria 3: Muy insatisfactoria o no durmió

6. Sensación de bienestar durante el día

0: Normal 1: Ligeramente disminuida 2: Notablemente disminuida 3: Muy disminuida

7. Funcionamiento (físico y mental) durante el día

0: Normal 1: Ligeramente disminuido 2: Notablemente disminuido 3: Muy disminuido

8. Somnolencia durante el día

0: Ninguna 1: Moderada 2: Considerable 3: Intensa

El periodo de la auto-administración puede variar, dependiendo del diseño de cada estudio. Cuando la auto-administración se haya efectuado en un periodo distinto al del mes anterior, la segunda frase de las instrucciones debería corregirse consecuentemente.

Exercici: AFE i AFC

Tenint en compte que...

L'Escala d'Insomni d'Atenes pretén mesurar les dificultats en el dormir en dos moments diferents: a) aquells problemes que sorgeixen durant la nit, i b) aquells problemes que sorgeixen durant el dia

I analitzant l'enunciat dels ítems del test...

Proposar una hipotètica matriu factorial resultant d'una AFE i una altra resultant d'una AFC

Anàlisi factorial exploratòria (AFE)

	codi	edat	sexe	AIS_1	AIS_2	AIS_3	AIS_4	AIS_5	AIS_6	AIS_7	AIS_8
1	6373	19	Home	0	0	0	1	0	1	0	1
2	5592	26	Home	0	1	1	1	1	1	0	2
3	0801	19	Dona	2	0	0	2	2	2	1	3
4	9530	20	Dona	0	1	0	1	0	0	0	1
5	0889		Home	2	1	2	1	2	1	1	1
6	0001	20	Dona	0	0	0	0	0			
7	6359	19	Dona	0	0	1	2	2			
8	1587	35	Dona	0	1	2	2	2	1	1	0
9	8503	22	Dona	0	1	0	0	0	0	0	0
10	0002	20	Dona	1	0	0	0	0	0	0	1
11	4640	19	Dona	0	0	1	0	0	0	0	1
12	8905	23	Dona	1	1	0	1	1	1	1	2
13	3737	19	Home	0	1	0	1	1	0	0	1
14	2222	19	Dona	0	1	1	1	1	1	0	1
15	0822	22	Dona	0	0	1	0	0	0	0	0
16	2468	20	Home	2	1	0	2	2	1	0	3
17	5885	19	Dona	0	0	0	1	0	0	0	0
18	2874	20	Home	0	1	0	1	1	0	0	1
19	2205	18	Dona	0	0	0	0	0	0	0	0
20	1008	19	Dona	0	0	1	0	0	0	0	1
21	4781	19	Dona								
22	2506	19	Dona	0	1	1	1	0	0	0	0
23	2806	24	Home	0	0	1	0	0	0	0	0
24	2889	19	Dona	0	0	0	0	0	0	0	0
25	9862	19	Home	0	0	0	1	0	0	0	1
26	2206	20	Home	2	1	0	1	2	1	1	2

Correlation Matrix

Correlation Matrix

		AIS_1	AIS_2	AIS_3	AIS_4	AIS_5	AIS_6	AIS_7	AIS_8
AIS_1	Pearson's r	—							
	p-value	—							
AIS_2	Pearson's r	0.394	—						
	p-value	< .001	—						
AIS_3	Pearson's r	0.207	0.208	—					
	p-value	0.015	0.015	—					
AIS_4	Pearson's r	0.337	0.165	0.076	—				
	p-value	< .001	0.055	0.378	—				
AIS_5	Pearson's r	0.405	0.215	0.299	0.447	—			
	p-value	< .001	0.012	< .001	< .001	—			
AIS_6	Pearson's r	0.347	0.144	0.163	0.484	0.642	—		
	p-value	< .001	0.093	0.058	< .001	< .001	—		
AIS_7	Pearson's r	0.239	0.052	0.140	0.422	0.483	0.761	—	
	p-value	0.005	0.551	0.104	< .001	< .001	< .001	—	
AIS_8	Pearson's r	0.273	0.091	0.001	0.501	0.474	0.603	0.492	—
	p-value	0.001	0.290	0.987	< .001	< .001	< .001	< .001	—

Matriu de dades

Matriu factorial

	Factor	
	1	2
AIS_1	0.1434	0.60807
AIS_2	-0.1150	0.62502
AIS_3	0.0178	0.34687
AIS_4	0.5345	0.15482
AIS_5	0.5670	0.30395
AIS_6	0.9127	0.00210
AIS_7	0.8072	-0.09411
AIS_8	0.6987	-0.02891

Matriu de correlacions

Anàlisi factorial exploratòria (AFE)

Passos:

1. Comprovar les condicions d'aplicació: Kaiser-Meyer-Olkin (KMO) i Prova d'esfericitat de Barlett.
2. Seleccionar un mètode d'extracció de factors.
3. Escollir un mètode de rotació (ortogonal o obliqua).
4. Decidir el criteri per a determinar el nombre de factors.
5. Interpretar els resultats: comunalitats/unicitats, variància explicada, matriu factorial, etc.

Passos AFE

Condicions d'aplicació:

- Kaiser-Meyer-Olkin (KMO)
- Prova d'esfericitat de Bartlett

Mètode d'extracció:

- Eixos principals
- Mínims quadrats
- Màxima versemblança
- Etc.

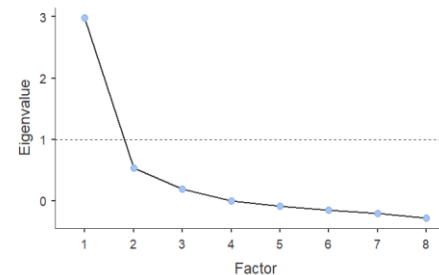
Rotació:

- Ortogonal (Varimax, Quartimax,...)
- Obliqua (Promax, Oblimin,...)

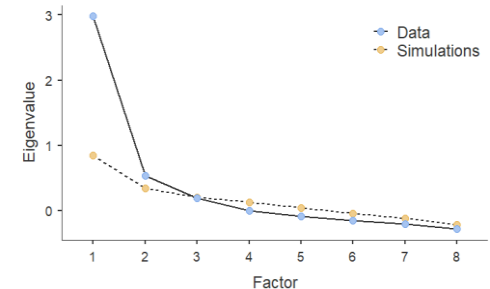
Nombre de factors:

- Criteri de Kaiser (valor propi > 1)
- Anàlisi paral·lela
- Gràfic de sedimentació

Scree Plot



Scree Plot



Interpretar resultats

Exemple AFE amb Jamovi (Athens Insomnia Scale)

Assumption Checks

Bartlett's Test of Sphericity

χ^2	df	p
372	28	< .001

Prova d'esfericitat de Bartlett: avalua l' H_0 de què les variables o ítems no estan relacionats entre ells.

KMO Measure of Sampling Adequacy

	MSA
Overall	0.803
AIS_1	0.819
AIS_2	0.692
AIS_3	0.666
AIS_4	0.891
AIS_5	0.850
AIS_6	0.753
AIS_7	0.768
AIS_8	0.861

KMO (Kaiser-Meyer-Olkin) indica fins a quin punt les associacions entre variables (ítems) poden ser explicades per un número de factors més reduït que el de les variables.

Exemple AFE amb Jamovi (Athens Insomnia Scale)

Exploratory Factor Analysis

Factor Loadings

	Factor		Uniqueness
	1	2	
AIS_1	0.1434	0.60807	0.540
AIS_2	-0.1150	0.62502	0.653
AIS_3	0.0178	0.34687	0.874
AIS_4	0.5345	0.15482	0.624
AIS_5	0.5670	0.30395	0.449
AIS_6	0.9127	0.00210	0.165
AIS_7	0.8072	-0.09411	0.400
AIS_8	0.6987	-0.02891	0.527

Note. 'Principal axis factoring' extraction method was used in combination with a 'oblimin' rotation

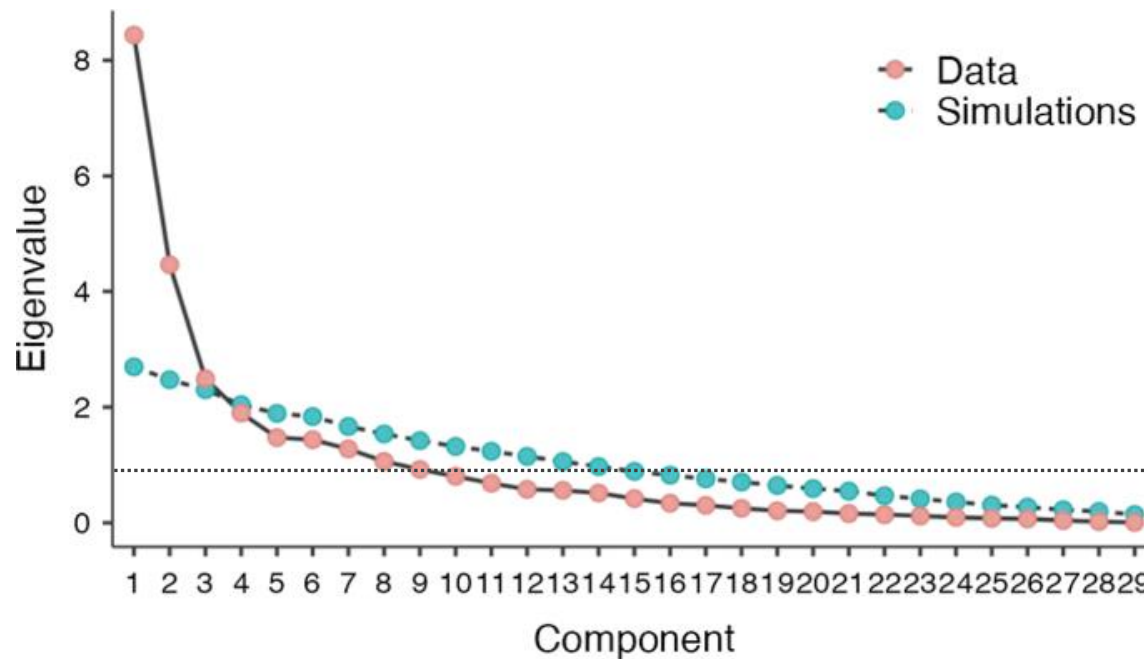
Unicitat (u^2): quantitat d'informació específica i única de l'ítem, que no comparteix amb els factors comuns.

Comunalitat (h^2): quantitat d'informació que l'ítem comparteix amb els factors comuns.

$$h^2 = 1 - u^2$$

AFE: Nombre de factors

El criteri de Kaiser (valor propi o autovalor > 1) tendeix a sobreestimar el nombre de factors a extreure



Criteri de Kaiser: 9 factors Anàlisi paral·lela: 3 factors

Exemple AFE amb Jamovi (Athens Insomnia Scale)

Factor Statistics

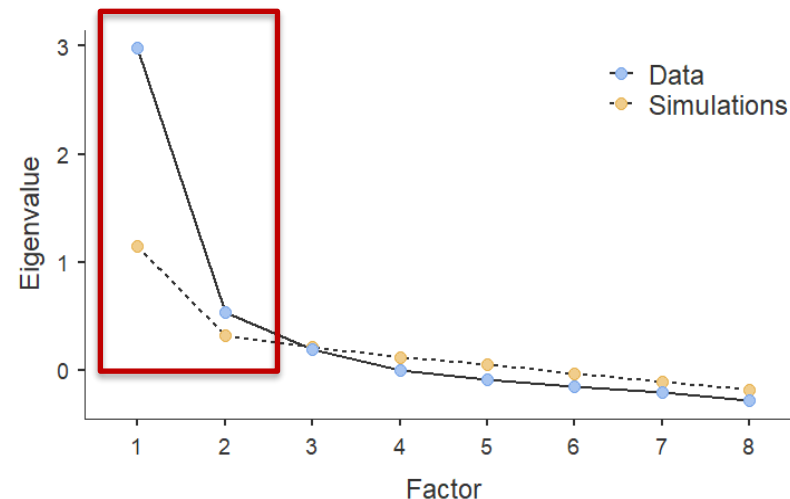
Nombre de factors basat en
l'anàlisi paral·lela i variància
explicada

Summary

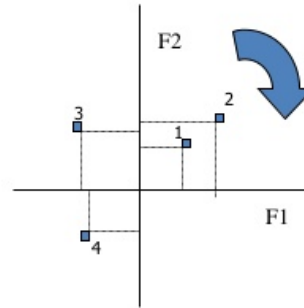
Factor	SS Loadings	% of Variance	Cumulative %
1	2.69	33.6	33.6
2	1.08	13.5	47.1

Gràfic de sedimentació

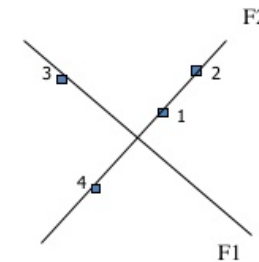
Scree Plot



AFE: Rotació

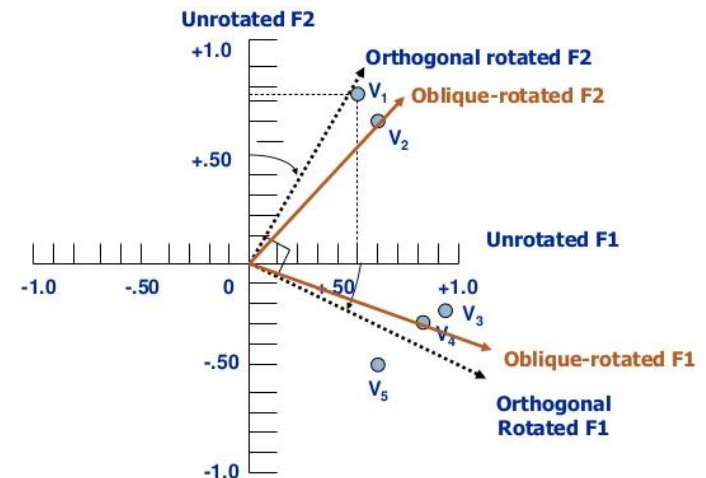
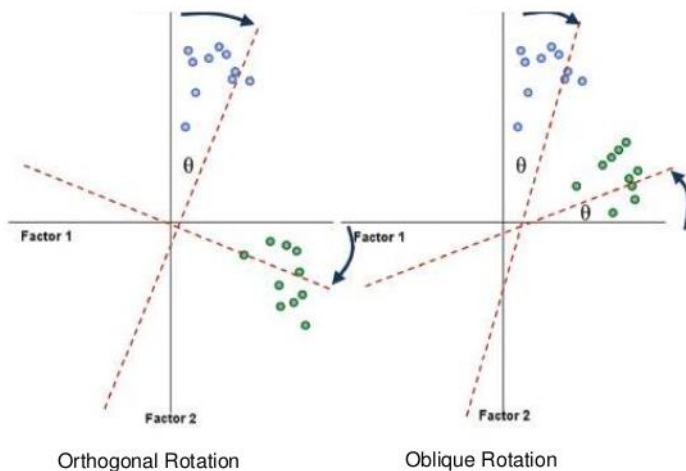


	Factor 1	Factor 2
x1	0.5	0.5
x2	0.8	0.8
x3	-0.7	0.7
x4	-0.5	-0.5



	Factor 1	Factor 2
x1	0	0.6
x2	0	0.9
x3	-0.9	0
x4	0	-0.9

Tipus de rotació: rotació ortogonal vs obliqua



Exemple AFE amb Jamovi (Athens Insomnia Scale)

Matriu rotada (rotació obliqua)

Correlació entre factors

Exploratory Factor Analysis

Factor Loadings

	Factor		Uniqueness
	1	2	
AIS_1	0.1434	0.60807	0.540
AIS_2	-0.1150	0.62502	0.653
AIS_3	0.0178	0.34687	0.874
AIS_4	0.5345	0.15482	0.624
AIS_5	0.5670	0.30395	0.449
AIS_6	0.9127	0.00210	0.165
AIS_7	0.8072	-0.09411	0.400
AIS_8	0.6987	-0.02891	0.527

Note. 'Principal axis factoring' extraction method was used in combination with a 'oblimin' rotation

Correlation Matrix

	1	2
1	—	0.516
2		—

Comunalitat AIS_1

$$h^2 = 1 - u^2 = 1 - 0,540 = 0,460$$

Autovalor Factor 1 (després de la rotació)

$$0,1434^2 + (-0,1150)^2 + \dots + 0,8072^2 + 0,6987^2 = 2,614$$

Interpretació AFE

Correlació ítem - factor

	Factor 1	
Ítems	Saturació/càrrega factorial	Comunalitat h^2
Ítem 1	0,767	0,588
Ítem 2	0,790	0,624
Ítem 3	0,504	0,254
Ítem 4	0,304	0,092
Autovalor		1,559
Percentatge de variància explicada		38,97%

Proporció de variància que el factor explica de l'ítem

El factor explica el 58,8% de la variància de l'ítem 1

Total de variància que el factor explica del conjunt d'ítems (Σ del quadrat de les saturacions)

El factor explica el 38,97% de la variància dels ítems $1,559/4 \cdot 100$

Anàlisi factorial confirmatòria (AFC)

Passos:

1. Especificar el model: nombre de factors, ítems que pertanyen a cada factor, relació entre factors i altres restriccions del model.
2. Seleccionar un mètode d'estimació.
3. Escollir els indicadors d'ajust i els corresponents criteris de valoració.
4. Interpretar els resultats: valorar l'ajust i els paràmetres (càrregues factorials, correlació entre factors, valors anòmals, etc.) del model.
5. En cas que l'ajust sigui pobre, explorar fonts de desajust (índexs de modificació) i reespecificar el model si és adient.

Passos AFC

Especificació model:

- Nombre de factors
- Ítems de cada factor
- Relació entre factors

Mètode d'estimació:

- Màxima versemblança
- Mínims quadrats ponderats
- Etc.

Indicadors d'ajust:

- CFI, TLI, NFI, etc.
- RMSEA, SRMR, etc.
- AIC, BIC, etc.

Criteris de valoració (Hooper et al., 2008; Hu i Bentler, 1999):

Índex d'ajust	Interpretació
Chi-quadrat/g.ll.	< 2 bo; < 5 acceptable
CFI i TLI	≥ 0,95 bo; ≥ 0,90 acceptable
RMSEA	≤ 0,05 bo; ≤ 0,08 acceptable
SRMR	≤ 0,05 bo; ≤ 0,06 acceptable

Interpretar resultats (i, si és adient, reespecificar el model)

Exemple AFC amb Jamovi (Athens Insomnia Scale)

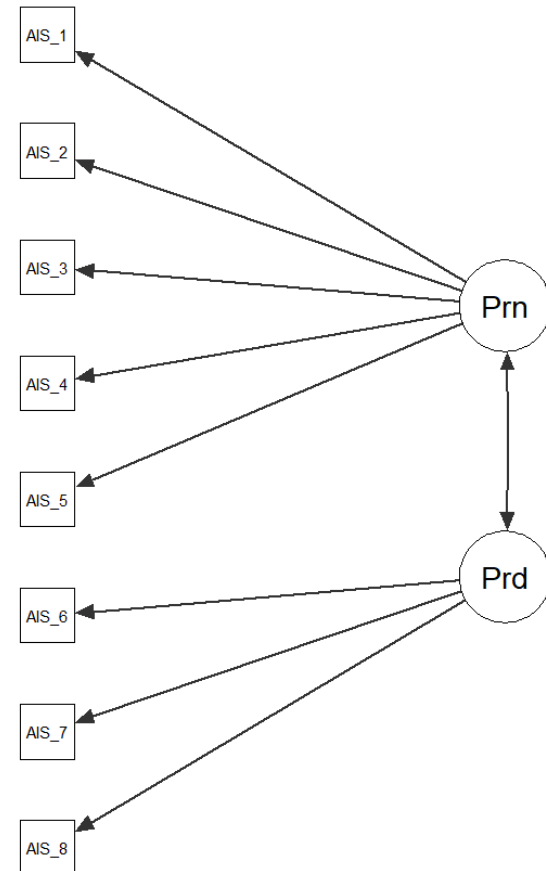
Especificació del model:

Factor 1 (Prn): Problemes durant la nit

Ítems: AIS_1, AIS_2, AIS_3, AIS_4, AIS_5

Factor 2 (Prd): Problemes durant el dia

Ítems: AIS_6, AIS_7, AIS_8



Exemple AFC amb Jamovi (Athens Insomnia Scale)

Indicadors d'ajust:

Model Fit

Test for Exact Fit

χ^2	df	p
43.5	19	0.001

Fit Measures

			RMSEA 90% CI	
CFI	TLI	RMSEA	Lower	Upper
0.932	0.900	0.0970	0.0589	0.135

Índex d'ajust	Valor	Valoració ajust
Chi-quadrat/graus de llibertat	43,5/19 = 2,29	Acceptable
CFI	0,932	Acceptable
TLI	0,900	Acceptable
RMSEA	0,097	Dolent

Exemple AFC amb Jamovi (Athens Insomnia Scale)

Paràmetres del model:

Factor Loadings

Factor	Indicator	Estimate	SE	Z	p	Stand. Estimate
Problemes nit	AIS_1	0.383	0.0663	5.78	< .001	0.513
	AIS_2	0.151	0.0495	3.06	0.002	0.287
	AIS_3	0.156	0.0494	3.16	0.002	0.295
	AIS_4	0.400	0.0593	6.75	< .001	0.590
	AIS_5	0.568	0.0562	10.10	< .001	0.817
Problemes dia	AIS_6	0.685	0.0484	14.16	< .001	0.962
	AIS_7	0.502	0.0472	10.65	< .001	0.787
	AIS_8	0.454	0.0574	7.91	< .001	0.634

Factor Covariances

		Estimate	SE	Z	p	Stand. Estimate
Problemes nit	Problemes nit	1.000 ^a				
	Problemes dia	0.795	0.0556	14.3	< .001	0.795
Problemes dia	Problemes dia	1.000 ^a				

^a fixed parameter

Exemple AFC amb Jamovi (Athens Insomnia Scale)

Índexs de modificació: Quan l'ajust sigui pobre, explorar fonts de desajust (índexs de modificació) i reespecificar el model

Modification Indices

Factor Loadings – Modification Indices

	Problemes nit	Problemes dia
AIS_1		2.27
AIS_2		4.15
AIS_3		2.84
AIS_4		3.53
AIS_5		2.91
AIS_6	0.0305	
AIS_7	1.9189	
AIS_8	3.5810	

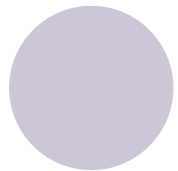
Residual Covariances – Modification Indices

	AIS_1	AIS_2	AIS_3	AIS_4	AIS_5	AIS_6	AIS_7	AIS_8
AIS_1		13.7	0.707	0.45503	0.335	0.244	0.7626	0.0869
AIS_2			2.594	0.00546	0.345	0.156	1.5834	0.1153
AIS_3				2.60050	3.016	0.182	0.0699	4.0665
AIS_4					2.868	0.948	0.6373	8.9311
AIS_5						1.545	0.9078	0.2303
AIS_6							3.5809	1.9189
AIS_7								0.0305
AIS_8								

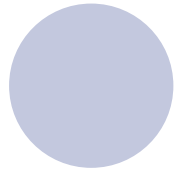
Passos: Identificar l'índex de modificació més elevat, incorporar l'estimació del paràmetre corresponent dins del model i reinterpretar els resultats

Relacions amb altres variables

Grau amb què les relacions de les puntuacions del test amb variables externes són les esperades.

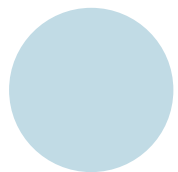


Validesa de criteri: predicció del criteri a partir de les puntuacions del test



Validesa convergent: relació del test amb mesures del mateix constructe o d'un constructe similar

Validesa discriminant: relació del test amb mesures d'un constructe diferent o d'un constructe poc relacionat



Validesa basada en la comparació de grups: comparació de les puntuacions del test entre grups definits

Validesa de criteri

Grau d'eficàcia amb què podrem diagnosticar o predir la variable *criteri* a partir de les puntuacions del test.

Tipus:

1. Coeficient de validesa (r_{xy})

Les puntuacions del test i del criteri són quantitatives – correlació entre el test i el criteri

2. Validesa de decisió

Les puntuacions del test i del criteri són dicotòmiques o dicotomitzaes – es compara la capacitat de classificació del test respecte a un diagnòstic

Validesa de criteri: Coeficient de validesa

Exemples de validesa de criteri - Coeficient de validesa (r_{xy})

Validem...	Criteri que volem predir...	Possible indicador del criteri...
Test d'intel·ligència	Rendiment acadèmic	Notes escolars
Sistema de selecció de personal en una empresa de ventes	Rendiment comercial	Número de ventes setmanals
Prova d'accés a la universitat	Rendiment acadèmic	Nota mitjana de la carrera
Test d'aptitud mecànica	Destresa mecànica	Temps (en hores) que es tarda en arreglar un cotxe amb la mateixa avaria

Dificultat de selecció dels indicadors del criteri

Validesa de criteri: Coeficient de validesa

Com decidim quin és l'indicador que cal escollir?

Requisits segons Thorndike i Hagen (1989):

Rellevants. Un indicador és rellevant en la mesura en què es correspon amb el criteri. Cal utilitzar el judici dels jutges experts.

Lliures de biaix. Les mesures del criteri han de representar la vertadera competència dels subjectes i no estar determinades per factors que actuïn diferencialment entre grups. Ex.: administratiu/va - físic

Fiables. Les mesures del criteri han de ser estables en el temps.

Accessibles. A igualtat de condicions d'adequació, escollirem aquell criteri que és més fàcil i pràctic aconseguir. Tindrem en compte aspectes econòmics, temporals, etc.

Validesa de criteri: Coeficient de validesa

Passos:

1. Definir el criteri que es pretén mesurar.
2. Identificar l'indicador o indicadors que s'utilitzaran per mesurar el criteri.
3. Selecció d'una mostra representativa de subjectes.
4. Administració del test i obtenció d'una puntuació per a cada persona.
5. Mesura de cada persona en el criteri: abans del test (retrospectiva), mateix temps que el test (concurrent) o després del test (predictiva).
6. Càlcul coeficient de validesa: determinar el grau de relació entre les puntuacions en el test i en la mesura del criteri.

Criteris valoració validesa concurrent

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.

- Inadequada: $|r| < 0,25$
- Suficient: $0,25 \leq |r| < 0,40$
- Bona: $0,40 \leq |r| < 0,50$
- Molt bona: $0,50 \leq |r| < 0,60$
- Excel·lent: $|r| \geq 0,60$

Criteris valoració validesa predictiva

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.

- Inadequada: $|r| < 0,20$
- Suficient: $0,20 \leq |r| < 0,35$
- Bona: $0,35 \leq |r| < 0,45$
- Molt bona: $0,45 \leq |r| < 0,55$
- Excel·lent: $|r| \geq 0,55$

Predicció del criteri a partir del test

$$y_i = r_{xy} \frac{S_y}{S_x} (x_i - \bar{x}) + \bar{y}$$

On:

$y_i \rightarrow$ Puntuació del subjecte en el criteri, es prediu a partir de la puntuació del subjecte en el test

$r_{xy} \rightarrow$ Coeficient de validesa del test

$S_y \rightarrow$ Desviació típica del criteri

$S_x \rightarrow$ Desviació típica del test

$x_i \rightarrow$ Puntuació del subjecte en el test

$\bar{x} \rightarrow$ Mitjana del test

$\bar{y} \rightarrow$ Mitjana del criteri

Interval de confiança del criteri

$$Y_{\text{int}} = y_i \pm z \cdot S_{xy}$$

← Error màxim →

On:

$y_i \rightarrow$ Puntuació predita del subjecte i

$z \rightarrow$ Valor de les taules llei normal (95%: $z=1,96$; 99%: $z=2,58$)

$S_{xy} \rightarrow$ Error típic d'estimació: $S_{xy} = S_y \sqrt{1 - r_{xy}^2}$

On:

$S_y \rightarrow$ Desviació típica criteri

$r_{xy} \rightarrow$ Coeficient de validesa del test

Exemple: Predicció del criteri

Suposem que tenim un test amb una fiabilitat de 0,75, una mitjana de 80 i una desviació típica de 16. Es correlaciona amb un criteri de mitjana 20 i una desviació típica de 8, obtenint un coeficient de correlació de 0,46. Un subjecte obté una puntuació de 86 en el test.

Quin és l'error típic d'estimació?

¿En quin interval se situa la puntuació del subjecte en el criteri a un nivell de confiança del 95% ($z = 1,96$)?

Exemple: Predicció del criteri

Error típic d'estimació:

$$S_{xy} = 8 \cdot \sqrt{1 - 0,46^2} = 7,10$$

Interval de confiança:

$$y_i = 0,46 \frac{8}{16} (86 - 80) + 20 = 21,38$$

$$Y_{\text{int}} \Rightarrow 21,38 \pm 1,96 \cdot 7,10 \Rightarrow 7,46 \leq Y_{\text{int}} \leq 35,30$$

Exercici: Predicció del criteri

Imaginem que tenim un test amb una fiabilitat de 0,80, una mitjana de 40 i una variància de 64. Es correlaciona amb un criteri de mitjana 60 i una desviació típica de 12, obtenint un coeficient de correlació de 0,52. Un subjecte presenta una puntuació de 35 en el test.

Quin és l'error típic d'estimació?

¿En quin interval se situa la puntuació del subjecte en el criteri a un nivell de confiança del 95% ($z = 1,96$)?

Factors que afecten al coeficient de validesa

- Validesa dels ítems
- Variabilitat
- Fiabilitat

$$\text{Si } r_{v_x v_y} = 1 \quad \rightarrow \quad r_{xy} = \sqrt{r_{xx}}$$

$$\text{Per tant, sempre es compleix que } r_{xy} \leq \sqrt{r_{xx}}$$

- Longitud del test

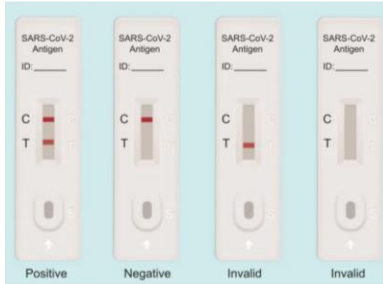
Validesa de criteri: Validesa de decisió

		Diagnòstic criteri		Total
		Positiu	Negatiu	
Resultat test	Positiu	Vertader positiu (f_{11})	Fals positiu (f_{12})	$f_{1.}$
	Negatiu	Fals negatiu (f_{21})	Vertader negatiu (f_{22})	$f_{2.}$
Total		$f_{.1}$	$f_{.2}$	N

*Positiu – Amb trastorn – Malalt – Apte – Tractament – Rehabilitació –
Teràpia – ...*

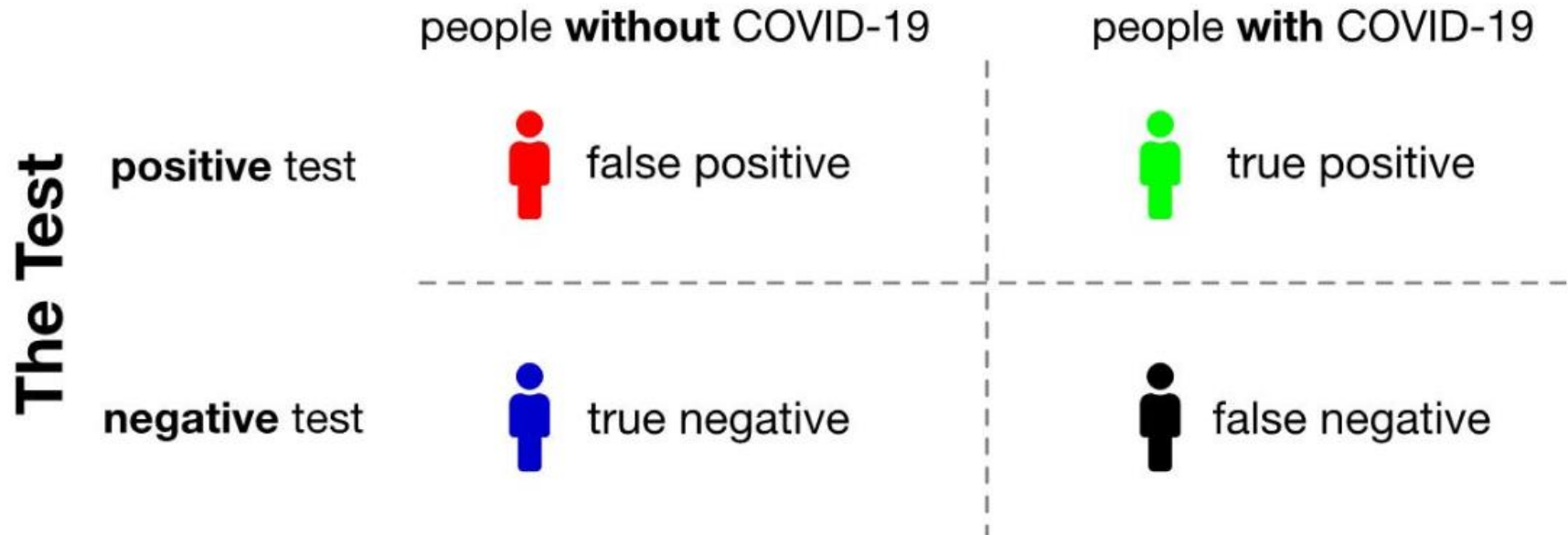
*Negatiu – Sense trastorn – Sa – No apte – No tractament – No rehabilitació
– No teràpia – ...*

Validesa de criteri: Validesa de decisió



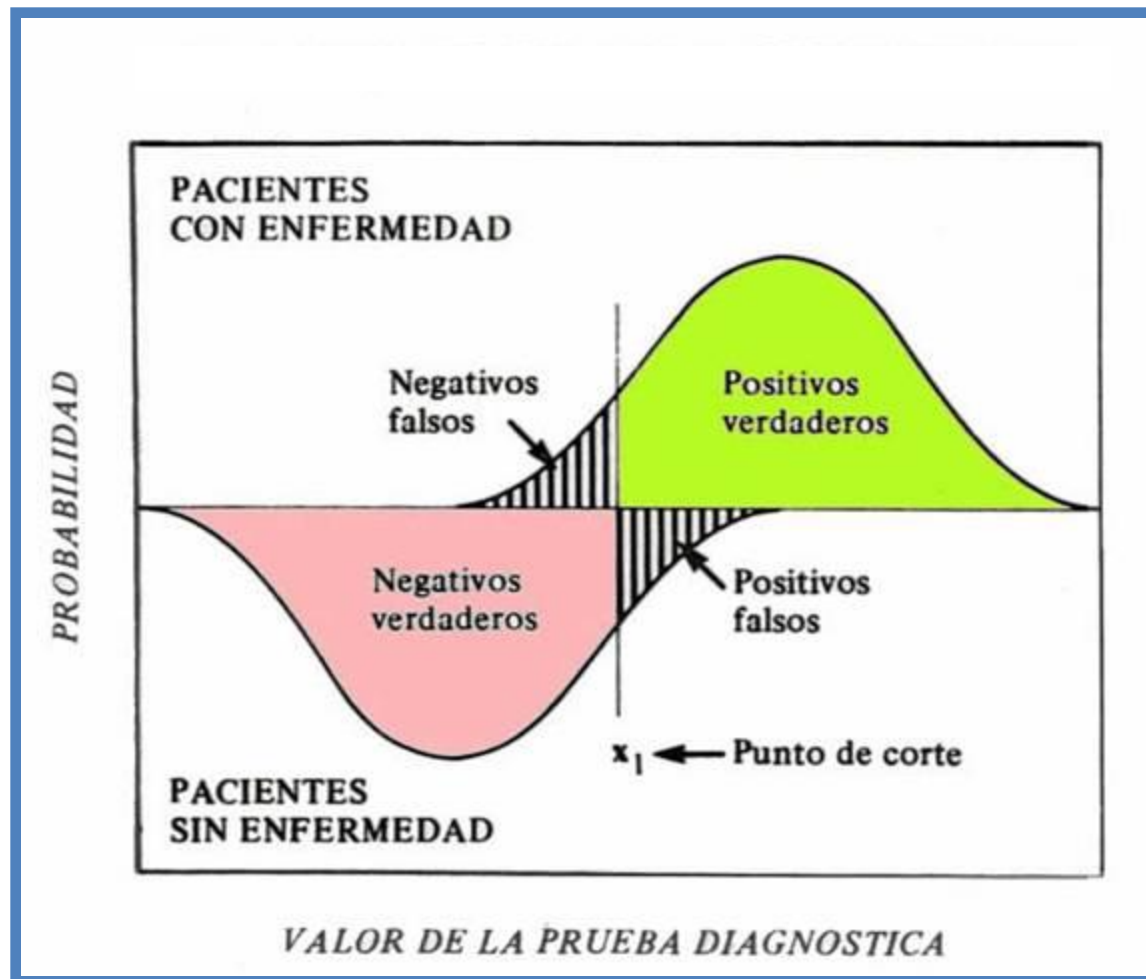
Imatge extreta de <https://www.statnews.com/>

The Truth



Imatge extreta de <https://www.statnews.com/>

Validesa de criteri: Validesa de decisió



Índexs de validesa de decisió: Proporció de classificacions correctes, sensibilitat i especificitat

$$P_c = \frac{\text{Vertaders positius } (f_{11}) + \text{Vertaders negatius } (f_{22})}{N}$$

$$\text{Sensibilitat} = \frac{\text{Vertaders positius } (f_{11})}{\text{Total persones amb trastorn } (f_{1.})}$$

Capacitat d'un test per a diagnosticar els casos positius. Indica la proporció de persones que tenen el trastorn que han estat classificades correctament. És màxima quan no hi ha FN.

$$\text{Especificitat} = \frac{\text{Vertaders negatius } (f_{22})}{\text{Total persones sanes } (f_{.2})}$$

Capacitat d'un test per a diagnosticar els casos negatius. Indica la proporció de persones que NO tenen el trastorn que han estat classificades correctament. És màxima quan no hi ha FP.

Índexs de validesa de decisió: Coeficient Kappa

$$K = \frac{F_c - F_a}{N - F_a}$$

F_c és la freqüència d'acords

F_a és la freqüència d'acords per atzar

On:

$$F_c = \text{Vertaders positius} (f_{11}) + \text{Vertaders negatius} (f_{22})$$

$$F_a = \frac{f_{1.} \cdot f_{.1} + f_{2.} \cdot f_{.2}}{N}$$

$$IC \Rightarrow K \pm z \cdot S_{e(K)}$$

$$S_{e(K)} = \sqrt{\frac{F_a}{N \cdot (N - F_a)}}$$

Valoració dels índexs de validesa de decisió

Valoració de Pc, S y E (Domènech i Granero, 2003)

Valors $\geq 0,80$: Acceptables

Valoració del coeficient Kappa (Altman, 1991)

Coeficient Kappa	Força de la concordança
< 0,20	Pobra
0,21 – 0,40	Feble
0,41 – 0,60	Moderada
0,61 – 0,80	Bona
0,81 – 1,00	Molt bona

Exemple: Validesa de decisió

		Depressió Post-part		Total
		Sí	No	
Escala de Hamilton	Sí	125	50	175
	No	25	300	325
	Total	150	350	500

$$P_c = \frac{\text{Vertaders positius } (f_{11}) + \text{Vertaders negatius } (f_{22})}{N}$$

$$P_c = \frac{125 + 300}{500} = 0,85$$

$$\text{Sensibilitat} = \frac{\text{Vertaders positius } (f_{11})}{\text{Total persones amb trastorn } (f_{.1})}$$

$$S = \frac{125}{150} = 0,83$$

$$\text{Especificitat} = \frac{\text{Vertaders negatius } (f_{22})}{\text{Total persones sanes } (f_{.2})}$$

$$E = \frac{300}{350} = 0,86$$

Exemple: Validesa de decisió

		Depressió Post-part		Total
		Sí	No	
Escala de Hamilton	Sí	125	50	175
	No	25	300	325
	Total	150	350	500

$$K = \frac{F_c - F_a}{N - F_a}$$

$$F_a = \frac{f_{1.} \cdot f_{.1} + f_{2.} \cdot f_{.2}}{N} \quad F_c = \text{Vertaders positius } (f_{11}) + \text{Vertaders negatius } (f_{22})$$

$$F_a = \frac{175 \cdot 150 + 325 \cdot 350}{500} = 280$$

$$F_c = 125 + 300 = 425$$

$$K = \frac{425 - 280}{500 - 280} = 0,66$$

Exemple: Validesa de decisió

		Depressió Post-part		Total
		Sí	No	
Escala de Hamilton	Sí	125	50	175
	No	25	300	325
	Total	150	350	500

$$K = 0,66$$

$$S_{e(K)} = \sqrt{\frac{F_a}{N \cdot (N - F_a)}}$$

$$IC \Rightarrow K \pm z \cdot S_{e(K)}$$

$$S_{e(K)} = \sqrt{\frac{280}{500 \cdot (500 - 280)}} = 0,05$$

$$IC(95\%) \Rightarrow 0,66 \pm 1,96 \cdot 0,05$$

$$0,56 \leq K \leq 0,76$$

Exercici: Validesa de decisió

		Diagnòstic clínic	
		No estrès	Estrès
Qüestionari d'estrès	Estrès	200	1235
	No estrès	2400	165

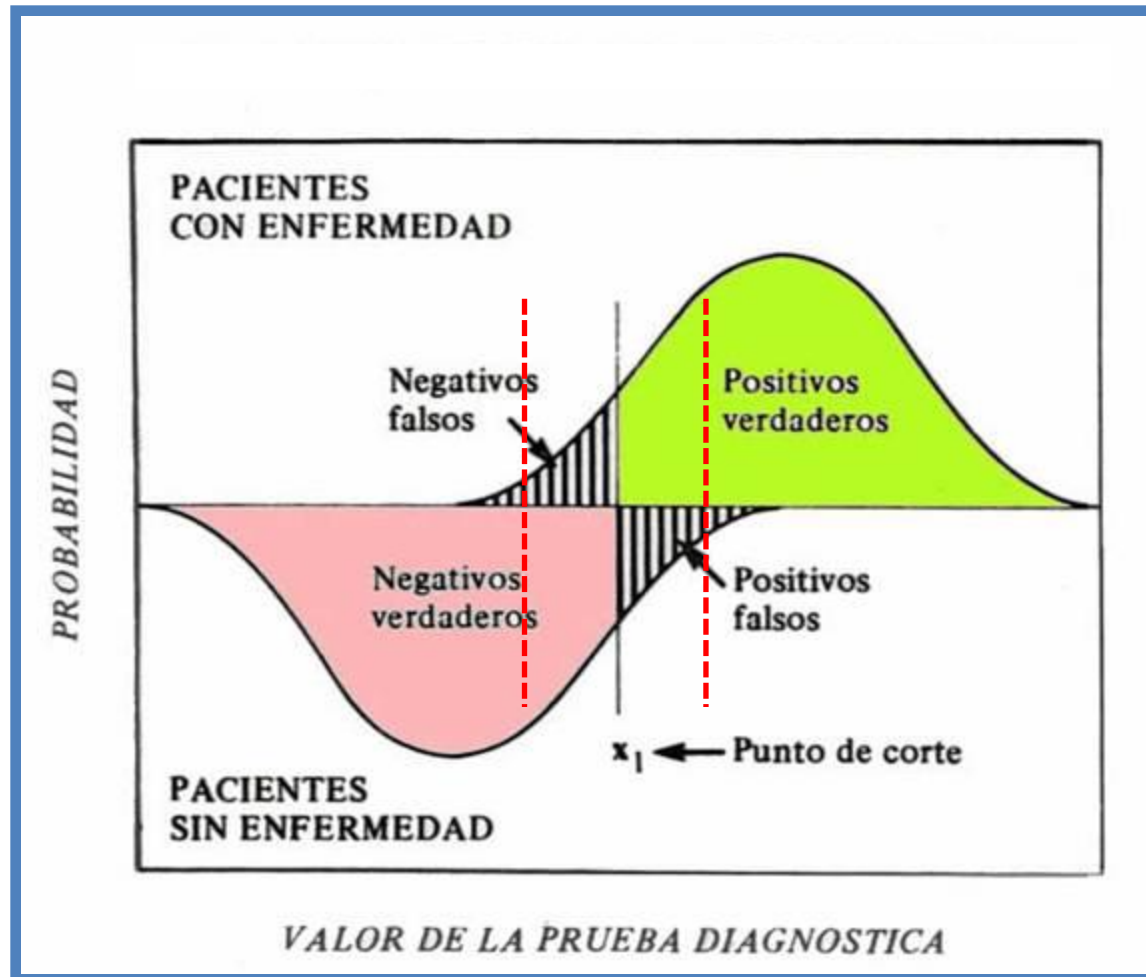
Obtenir:

- Proporció de classificacions correctes
- Sensibilitat
- Especificitat
- Coeficient Kappa i IC (95%; $z = 1,96$)

IMPORTANT!

L'ordre de les
columnes està
invertit

Validesa de decisió – Punt de tall



Validesa de decisió – Punt de tall

Passos:

1. Per cada possible punt de tall, calcular el número de FP, FN i el número total d'errors
2. Escollir el punt de tall que minimitzi el número total d'errors
3. En cas d'empat, valorar la implicació de cometre un FP o un FN

Exemple: Validesa de decisió – Punt de tall

Disposem de la decisió emesa per un psicòleg de la necessitat de fer teràpia (TE) o no fer-ne (NT) dirigides a 12 persones.

Se'ls aplica una escala les puntuacions de la qual oscil·len entre 1 i 10.

Com podem determinar aquell punt de tall que optimitzi les seves classificacions fetes per l'escala?

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 1

Persones	Escala	Diagnòstic	≥ 1	
A	6	NT	TE	FP
B	6	TE	TE	V _p
C	7	TE	TE	V _p
D	8	NT	TE	FP
E	5	NT	TE	FP
F	8	TE	TE	V _p
G	4	NT	TE	FP
H	9	TE	TE	V _p
I	3	NT	TE	FP
J	7	TE	TE	V _p
K	7	NT	TE	FP
L	10	TE	TE	V _p

6 FP

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 2

Persones	Escala	Diagnòstic	≥ 2	
A	6	NT	TE	FP
B	6	TE	TE	V _p
C	7	TE	TE	V _p
D	8	NT	TE	FP
E	5	NT	TE	FP
F	8	TE	TE	V _p
G	4	NT	TE	FP
H	9	TE	TE	V _p
I	3	NT	TE	FP
J	7	TE	TE	V _p
K	7	NT	TE	FP
L	10	TE	TE	V _p

6 FP

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 3

Persones	Escala	Diagnòstic	≥ 3	
A	6	NT	TE	FP
B	6	TE	TE	V _p
C	7	TE	TE	V _p
D	8	NT	TE	FP
E	5	NT	TE	FP
F	8	TE	TE	V _p
G	4	NT	TE	FP
H	9	TE	TE	V _p
I	3	NT	TE	FP
J	7	TE	TE	V _p
K	7	NT	TE	FP
L	10	TE	TE	V _p

6 FP

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 4

Persones	Escala	Diagnòstic	≥ 4	
A	6	NT	TE	FP
B	6	TE	TE	V _p
C	7	TE	TE	V _p
D	8	NT	TE	FP
E	5	NT	TE	FP
F	8	TE	TE	V _p
G	4	NT	TE	FP
H	9	TE	TE	V _p
I	3	NT	NT	V _N
J	7	TE	TE	V _p
K	7	NT	TE	FP
L	10	TE	TE	V _p

5 FP

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 5

Persones	Escala	Diagnòstic	≥ 5	
A	6	NT	TE	FP
B	6	TE	TE	V_p
C	7	TE	TE	V_p
D	8	NT	TE	FP
E	5	NT	TE	FP
F	8	TE	TE	V_p
G	4	NT	NT	V_N
H	9	TE	TE	V_p
I	3	NT	NT	V_N
J	7	TE	TE	V_p
K	7	NT	TE	FP
L	10	TE	TE	V_p

4 FP

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 6

Persones	Escala	Diagnòstic	≥ 6	
A	6	NT	TE	FP
B	6	TE	TE	V _p
C	7	TE	TE	V _p
D	8	NT	TE	FP
E	5	NT	NT	V _N
F	8	TE	TE	V _p
G	4	NT	NT	V _N
H	9	TE	TE	V _p
I	3	NT	NT	V _N
J	7	TE	TE	V _p
K	7	NT	TE	FP
L	10	TE	TE	V _p

3 FP

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 7

Persones	Escala	Diagnòstic	≥ 7	
A	6	NT	NT	V_N
B	6	TE	NT	FN
C	7	TE	TE	V_P
D	8	NT	TE	FP
E	5	NT	NT	V_N
F	8	TE	TE	V_P
G	4	NT	NT	V_N
H	9	TE	TE	V_P
I	3	NT	NT	V_N
J	7	TE	TE	V_P
K	7	NT	TE	FP
L	10	TE	TE	V_P

2 FP
1 FN

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 8

Persones	Escala	Diagnòstic	≥ 8	
A	6	NT	NT	V_N
B	6	TE	NT	FN
C	7	TE	NT	FN
D	8	NT	TE	FP
E	5	NT	NT	V_N
F	8	TE	TE	V_P
G	4	NT	NT	V_N
H	9	TE	TE	V_P
I	3	NT	NT	V_N
J	7	TE	NT	FN
K	7	NT	NT	V_N
L	10	TE	TE	V_P

1 FP
3 FN

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 9

Persones	Escala	Diagnòstic	≥ 9	
A	6	NT	NT	V_N
B	6	TE	NT	FN
C	7	TE	NT	FN
D	8	NT	NT	V_N
E	5	NT	NT	V_N
F	8	TE	NT	FN
G	4	NT	NT	V_N
H	9	TE	TE	V_P
I	3	NT	NT	V_N
J	7	TE	NT	FN
K	7	NT	NT	V_N
L	10	TE	TE	V_P

4 FN

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall 10

Persones	Escala	Diagnòstic	≥ 10	
A	6	NT	NT	V_N
B	6	TE	NT	FN
C	7	TE	NT	FN
D	8	NT	NT	V_N
E	5	NT	NT	V_N
F	8	TE	NT	FN
G	4	NT	NT	V_N
H	9	TE	NT	FN
I	3	NT	NT	V_N
J	7	TE	NT	FN
K	7	NT	NT	V_N
L	10	TE	TE	V_P

5 FN

Persones	Escala	Diagnòstic
A	6	NT
B	6	TE
C	7	TE
D	8	NT
E	5	NT
F	8	TE
G	4	NT
H	9	TE
I	3	NT
J	7	TE
K	7	NT
L	10	TE

Exemple: Validesa de decisió – Punt de tall

1. Per cada possible punt de tall, calcular el número de FP, FN i el número total d'errors

2. Escollir el punt de tall que minimitzi el número total d'errors

3. En cas d'empat, valorar la implicació de cometre un FP o un FN



	Punt de Tall	Fals Positiu	Fals Negatiu	Total errors
	1	6	0	6
	2	6	0	6
	3	6	0	6
	4	5	0	5
	5	4	0	4
≥	6	3	0	3
	7	2	1	3
	8	1	3	4
	9	0	4	4
	10	0	5	5

Exemple: Validesa de decisió – Punt de tall

Agafant com a punt de tall ≥ 6 , obtindríem...

		Psicòleg		Total
		TE	NT	
Escala	TE	6	3	9
	NT	0	3	3
	Total	6	6	12

I podríem calcular els índexs de validesa de decisió per aquest punt de tall concret

Exemple: Validesa de decisió – Punt de tall

$$P_c = \frac{6 + 3}{12} = 0,75$$

$$S = \frac{6}{6} = 1 \text{ (no hi han falsos negatius)}$$

$$E = \frac{3}{6} = 0,5 \text{ (hi han molts falsos positius)}$$

$$F_c = 6 + 3 = 9$$

$$F_a = \frac{9 \cdot 6 + 3 \cdot 6}{12} = 6$$

$$K = \frac{9 - 6}{12 - 6} = 0,5$$

Exercici: Validesa de decisió – Punt de tall

En una residència d'avis es posa a prova la validesa d'una escala d'observació dissenyada per a detectar la dependència funcional dels residents i assignar-los a un grup de rehabilitació (R) o no (NR).

Es disposa de les puntuacions obtingudes per 11 residents en l'escala d'observació, així com del diagnòstic emès per un especialista (R o NR).

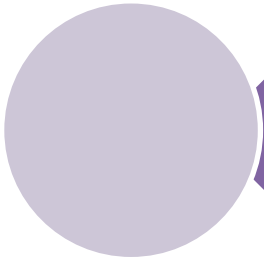
Determinar el punt de tall òptim en l'escala d'observació, tenint en compte que la rehabilitació no es considera nociva per l'individu. Construir la corresponent taula de classificacions.

Exercici: Validesa de decisió – Punt de tall

Residents	Escala observació	Diagnòstic especialista
1	7	NR
2	2	NR
3	1	NR
4	1	NR
5	5	NR
6	6	R
7	4	R
8	4	NR
9	3	NR
10	9	R
11	10	R

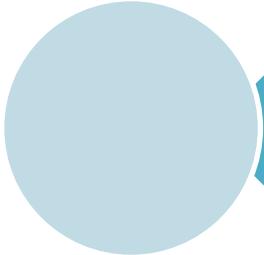
Validesa convergent i validesa discriminant

Grau amb què el test es relaciona, en la direcció i la magnitud esperades, amb variables externes que mesuren el mateix constructe (o similar) o un constructe diferent.



Coeficient de correlació

- Amb el mateix constructe o constructes relacionats (Validesa convergent)
- Amb constructes diferents o poc relacionats (Validesa discriminant)



Matriu multitret-multimètode (Campbell i Fiske, 1959)

- Mateix tret – diferent mètode (Validesa convergent)
- Diferent tret – mateix mètode (Validesa discriminant)

Criteris valoració validesa convergent

Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75.

- Inadequada: $|r| < 0,25$
- Adequada però amb algunes mancances: $0,25 \leq |r| < 0,40$
- Adequada: $0,40 \leq |r| < 0,50$
- Bona: $0,50 \leq |r| < 0,60$
- Excel·lent: $|r| \geq 0,60$

Exemple: Validesa convergent

Validez convergente del NEO-PI-R en relación a los adjetivos de Goldberg

Para evaluar la validez convergente se ha obtenido una matriz de correlaciones Producto-Momento de Pearson entre las cinco dimensiones del NEO-PI-R con las cinco dimensiones homónimas de los adjetivos de Goldberg (1992) para cada género. Como puede verse en la Tabla 2 cada dimensión del NEO-PI-R obtiene correlaciones positivas elevadas y significativas con las dimensiones de Goldberg, a excepción de las correlaciones entre Neuroticismo y Estabilidad Emocional que son inversas. Los resultados demuestran una buena validez convergente del NEO-PI-R frente a los adjetivos de Goldberg.

Tabla 2
Matriz de correlaciones por género entre el NEO-PI-R y los adjetivos de Goldberg.

	O	C	E	A	N	INT	AMA	RES	EST	EXT
O	--	-.04	.33	-.01	-.02	.41	.13	-.12	-.05	.18
C	-.13	--	.13	.23	-.34	.33	.29	.74	.31	.22
E	.41	-.00	--	.04	-.31	.34	.38	.02	.19	.70
A	-.08	.22	-.05	--	-.19	-.09	.46	.22	.16	.02
N	.02	-.39	-.23	-.22	--	-.16	-.33	-.24	-.70	-.35
INT	.45	.24	.34	-.09	-.13	--	.41	.39	.20	.39
AMA	.12	.20	.22	.53	-.23	.26	--	.45	.37	.47
RES	-.18	.70	-.12	.21	-.25	.28	.37	--	.32	.19
EST	-.11	.33	.11	.25	-.73	.14	.31	.33	--	.31
EXT	.32	.03	.73	-.07	-.26	.38	.28	.01	.19	--

Significación de r: 0 .09 < 0.01, 0 .05 < 0.05, 0.11 < 0.001

O= Apertura C= Responsabilidad E= Extraversión A= Amabilidad N= Neuroticismo;
INT= Intelecto AMA= Amabilidad RES= Responsabilidad EST= Estabilidad Emocional
EXT= Extraversión

Nota: En negrita las correlaciones iguales o superiores a .40

Exemple extret de Aluja, A., Blanch, A., Solé, D., Dolcet, J. M., & Gallart, S. (2008). Validez convergente y estructural del NEO-PI-R. Baremos orientativos. *Boletín de Psicología*, 92, 7-25.

Matriu multitret-multimètode

	A1	B1	C1	A2	B2	C2	A3	B3	C3
A1	(0.89)								
B1	0.51	(0.88)							
C1	0.38	0.37	(0.76)						
A2				(....)					
B2					(....)				
C2						(....)			
A3							(....)		
B3								(....)	
C3									(....)

A, B, C són trets
1, 2, 3 són mètodes

(...) Valors monotret-monomètode (FIABILITAT)

... Valors heterotret-monomètode (V. DISCRIMINANT)

... Valors monotret-heteromètode (V. CONVERGENT)

... Valors heterotret-heteromètode

Matriu multitret-multimètode

	A1	B1	C1	A2	B2	C2	A3	B3	C3
A1	0.89								
B1	0.03	0.88							
C1	0.28	0.17	0.76						
A2	0.73	0.14	0.22	0.93					
B2	0.15	0.69	0.03	0.18	0.79				
C2	0.12	0.04	0.81	0.03	0.15	0.85			
A3	0.77	0.11	0.09	0.77	0.21	0.16	0.83		
B3	0.21	0.75	0.18	0.21	0.68	0.03	0.15	0.92	
C3	0.19	0.05	0.78	0.22	0.09	0.72	0.14	0.09	0.89

A, B, C són trets
1, 2, 3 són mètodes

Exemple: Matriu multitret-multimètode

- En la següent matriu multitret-multimètode hi ha tres trets:
 - Extraversió
 - Lideratge
 - Intel·ligència social
- Cada tret està mesurat amb tres mètodes:
 - Autoinforme (AI)
 - Observació sistemàtica (OS)
 - Entrevista personal (EP)

Exemple: Matriu multitret-multimètode

		Extraversió			Lideratge			Intel·ligència social		
		AI	OS	EP	AI	OS	EP	AI	OS	EP
Extraversió	AI	0,80								
	OS	0,70	0,80							
	EP	0,60	0,70	0,90						
Lideratge	AI	0,20	0,21	0,10	0,80					
	OS	0,15	0,09	0,08	0,70	0,80				
	EP	0,10	0,12	0,10	0,60	0,60	0,80			
Intel·ligència social	AI	0,20	0,10	0,25	0,10	0,10	0,30	0,90		
	OS	0,09	0,10	0,10	0,10	0,20	0,10	0,70	0,80	
	EP	0,22	0,13	0,20	0,30	0,10	0,20	0,70	0,60	0,80

Exemple: Matriu multitret-multimètode

- A la diagonal principal apareixen les correlacions dels tests amb sí mateixos, això ens indica, els coeficients de **fiabilitat**.
- En l'anàlisi de la **validesa convergent**, és a dir en el cas que un mateix tret es mesura amb diferents mètodes, s'obtenen valors iguals o superiors a 0,60 en tots els casos.
- Les correlacions entre les mesures de diferents trets mesurades amb el mateix mètode ens informen de la **validesa discriminant**. En l'exemple la màxima correlació és de 0,20.

Exercici: Matriu multitret-multimètode

- Mesurem tres trets:
 - Raonament numèric (RN)
 - Factors espacial (FE)
 - Raonament abstracte (RA)
- Cada tret està mesurat amb tres mètodes:
 - Ítems V/F (VF)
 - Ítems d'elecció múltiple (EM)
 - Ítems de format incomplet (FI)

Exercici: Matriu multitret-multimètode

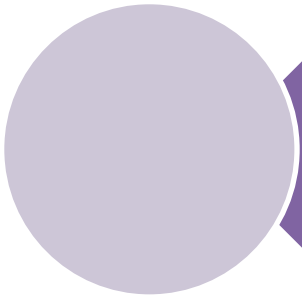
		Veritable /Fals			Elecció múltiple			Format incomplet		
		RN	FE	RA	RN	FE	RA	RN	FE	RA
Veritable/Fals	RN	0,95								
	FE	0,20	0,90							
	RA	0,30	0,28	0,92						
Elecció múltiple	RN	0,90	0,31	0,40	0,93					
	FE	0,26	0,87	0,33	0,37	0,94				
	RA	0,43	0,20	0,84	0,26	0,37	0,88			
Format incomplet	RN	0,79	0,27	0,31	0,77	0,15	0,23	0,89		
	FE	0,11	0,68	0,22	0,24	0,67	0,31	0,19	0,93	
	RA	0,19	0,18	0,50	0,19	0,33	0,72	0,41	0,30	0,64

Exercici: Matriu multitret-multimètode

- Identificar els coeficients de fiabilitat
- Identificar els valors monotret-heteromètode
- Identificar els valors multitret-monomètode
- Valorar la validesa convergent i la validesa discriminant

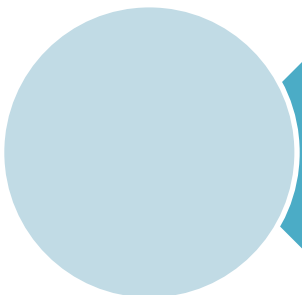
Validesa basada en la comparació de grups

Grau amb què al comparar les puntuacions del test entre grups definits els resultats estan en consonància amb el model teòric sobre el qual s'ha construït el test



Si el model teòric planteja que els grups difereixen en el constructe...

Les puntuacions del test hauran de reflectir aquestes diferències (comparació de grups amb resultats estadísticament significatius)

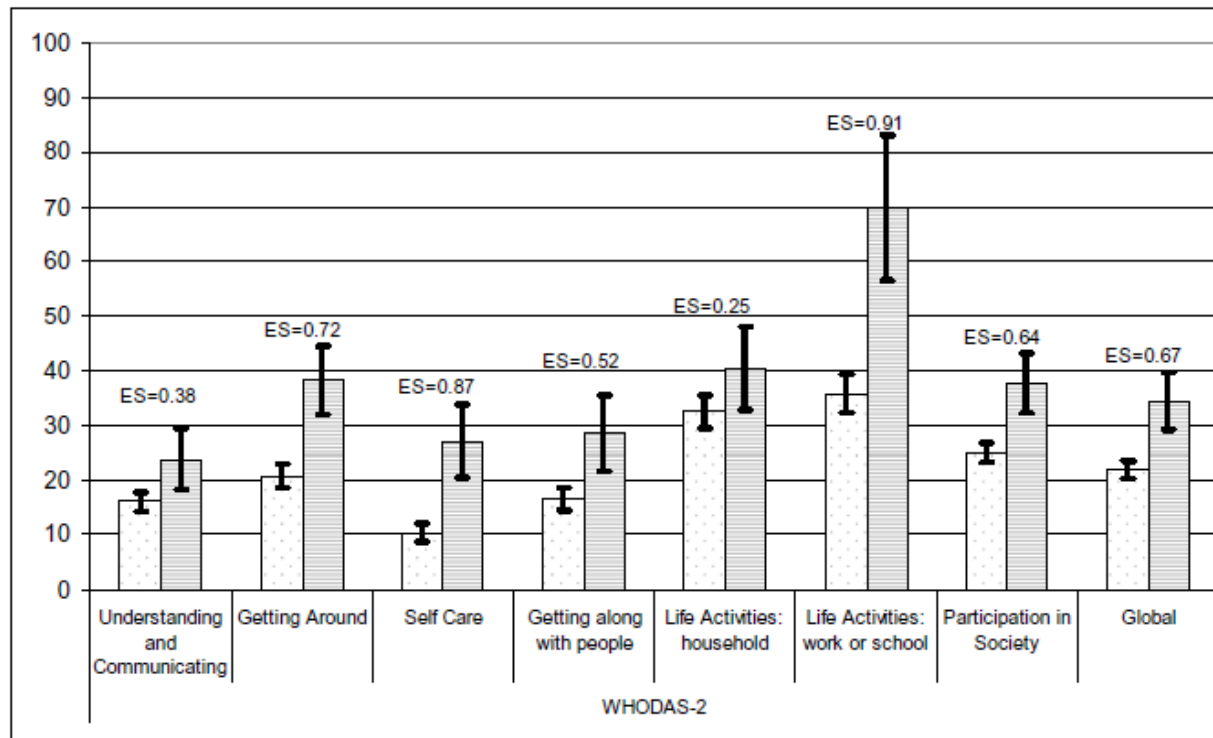


Si el model teòric planteja que els grups no difereixen en el constructe...

Les puntuacions del test hauran de ser similars (comparació de grups amb resultats no estadísticament significatius)

Exemple: Validesa basada en la comparació de grups

Comparació de les puntuacions de la WHODAS entre els pacients que, en el moment de l'estudi, estaven treballant i els que no

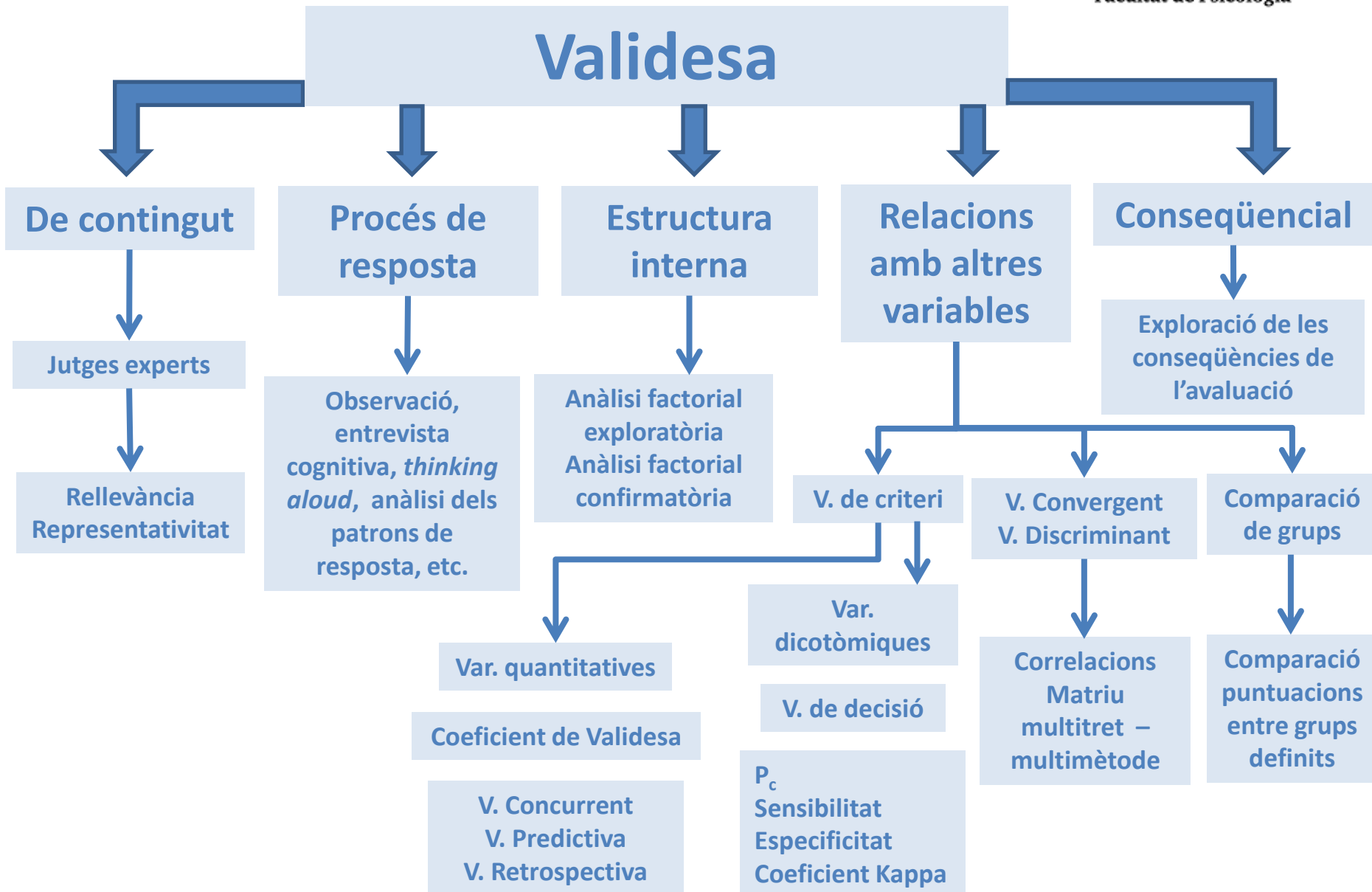


Exemple extret de Garin, O., Ayuso-Mateos, J. L., Almansa, J., Nieto, M., Chatterji, S., Vilagut, G., ... & Racca, V. (2010). Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases. *Health and Quality of Life Outcomes*, 8(1), 51.

Validesa conseqüencial

Les conseqüències positives i negatives d'aplicar el test coincideixen amb les que estan previstes.

1. Demostrar que les conseqüències de l'avaluació coincideixen amb les conseqüències que es pretenia assolir
2. Explorar l'existència de possibles conseqüències no previstes



Baremació



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement- No comercial- Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).



Lectures:

Gómez Benito, J. (1997). *Construcció d'instruments de mesura*. Barcelona: Edicions de la Universitat Oberta de Catalunya [únicament de la pàgina 34 a la 38]

Puntuacions directes

Puntuació Directa (PD): puntuació obtinguda directament de l'aplicació d'un test

El valor de PD dependran de:

- Com es puntuen les respostes a cada ítem (ex.: 0 i 1, de 1 a 5, de 1 a 8, etc.)
- Presència d'ítems inversos – Recodificació dels valors
- La fórmula per obtenir les PD (ex.: suma de les puntuacions de tots els ítems, ponderació d'ítems, etc.)
- Nombre d'ítems

Exemple: Ítems directes vs. ítems inversos

Qüestionari de confiança per a parlar en públic (Méndez, Inglés i Hidalgo, 1999)

- 30 ítems: 15 afirmacions redactades en positiu + 15 en negatiu
- Cada ítem es presenta en una escala de valoració d'1 (completament en desacord) a 6 (completament d'acord)
- Ítem **directe**: “Afronto amb plena confiança la perspectiva de fer una xerrada”
- Ítem **invers**: “Em sento aterrit(da) per la idea d’haver de parlar davant d’un grup de persones”



Abans de
calcular la
PD...

Valor original	1	2	3	4	5	6
Valor recodificat	6	5	4	3	2	1

Puntuacions directes

- Per sí soles, no tenen significat (Ex.: Què significa obtenir una PD de 25 en una prova de coneixements d'aritmètica? I obtenir una PD de 32 en un test que mesura ansietat?)
- No permeten comparar puntuacions obtingudes amb tests diferents
- Cal disposar de valors de referència per a poder-les interpretar, per a dotar-les de significat

En resum:

Les puntuacions que hem vist fins ara són les que es denominen **puntuacions directes** i, són les que obtenim directament del test o prova que apliquem.

La puntuació total d'un test depèn de diferents característiques com:

- la llargària ,
- la manera de puntuar cada ítem,
- la formula per obtenir la puntuació total, etc.

La puntuació directa del test és difícil d'interpretar si no disposem de cap altre informació. No ens aporta cap informació sobre el rendiment o nivell en el constructe del subjecte.

És necessari tenir valors de referència que ens permetin de situar les persones en una escala . **Grup normatiu**

Tipus d'interpretació

➤ Interpretació absoluta

- Puntuació Total mínima (PTm): suma de les respostes als ítems suposant que, en cada ítem, es respongui amb el valor mínim possible
- Puntuació Total Màxima (PTM): suma de les respostes als ítems suposant que, en cada ítem, es respongui amb el valor màxim possible
- Opció d'utilitzar punts de tal fixats prèviament (Ex.: < 5.0 Suspens, 5.0 – 6.9 Aprovat, 7.0 – 8.9 Notable, etc.)

➤ Interpretació relativa

Transformació de puntuacions (baremació) comparant la puntuació obtinguda per una persona amb el conjunt de puntuacions que han obtingut altres persones que han contestat el mateix test

Interpretació relativa - Transformació de les Puntuacions

Puntuació
Directa

- No aporta informació sobre el nivell del individu

TCT

- Interpretació: comparant la puntuació obtinguda amb les puntuacions de la població a la que pertany l'individu

Inferir

- Mostra representativa de la població

- Grup normatiu

Fases d'un Estudi Normatiu

1. Identificació de la població d'interès.
2. Selecció d'una mostra representativa → Grup normatiu.
3. Recollida de les dades → Aplicació del test.
4. Transformació de puntuacions directes en normatives.
5. Descripció de l'estudi normatiu en el manual del test.

Característiques de l'Estudi Normatiu

Té caducitat: canvis socioculturals amb el pas del temps

Cal recalcular els barems per adaptar el test a grups diferents

Informació que ha de contenir el manual d'un test sobre les normes d'interpretació

La mostra, el disseny mostral, les taxes de participació i els estadístics descriptius

Permetre, a l'usuari del test, d'avaluar l'adequació del procés i interpretar els resultats obtinguts

Si el test va dirigit a grups diferents, cal definir les normes per a cada grup

Tipus de Transformacions



Normes Cronològiques



Percentils



Puntuacions típiques



Puntuacions típiques normalitzades



Puntuacions típiques derivades

Normes Cronològiques

Determinar l'EM de cada EC en el test (Binet principis segle).

$$QI = \frac{EM}{EC} \cdot 100$$

- Només es poden utilitzar quan estudiem una característica que experimenta canvis progressius amb l'edat.
- Molt utilitzades en les primeres escales evolutives.

Fases Normes Cronològiques

1. Seleccionar mostres de nens corresponents a les diferents edats a les que el test va dirigit.
2. S'aplica el test als nens de cada edat, i es calcula la seva puntuació mitjana en el test per cada edat.
3. Es construeix una taula en la que a cada edat se li assigna la puntuació mitjana corresponent en el test.

Exemple: Normes Cronològiques

Tenint en compte la taula següent:

Edat cronològica (EC)	Mitjana de les PD	Edat mental (EM)
9	25	9
10	30	10
11	32	11

Si apliquem el test a un nen de 12 anys i obté una puntuació de 30, quina EM li correspon?

Resposta: Una EM de 10

En cas que l'edat d'aquest nen fos de 9, quin QI li correspon?

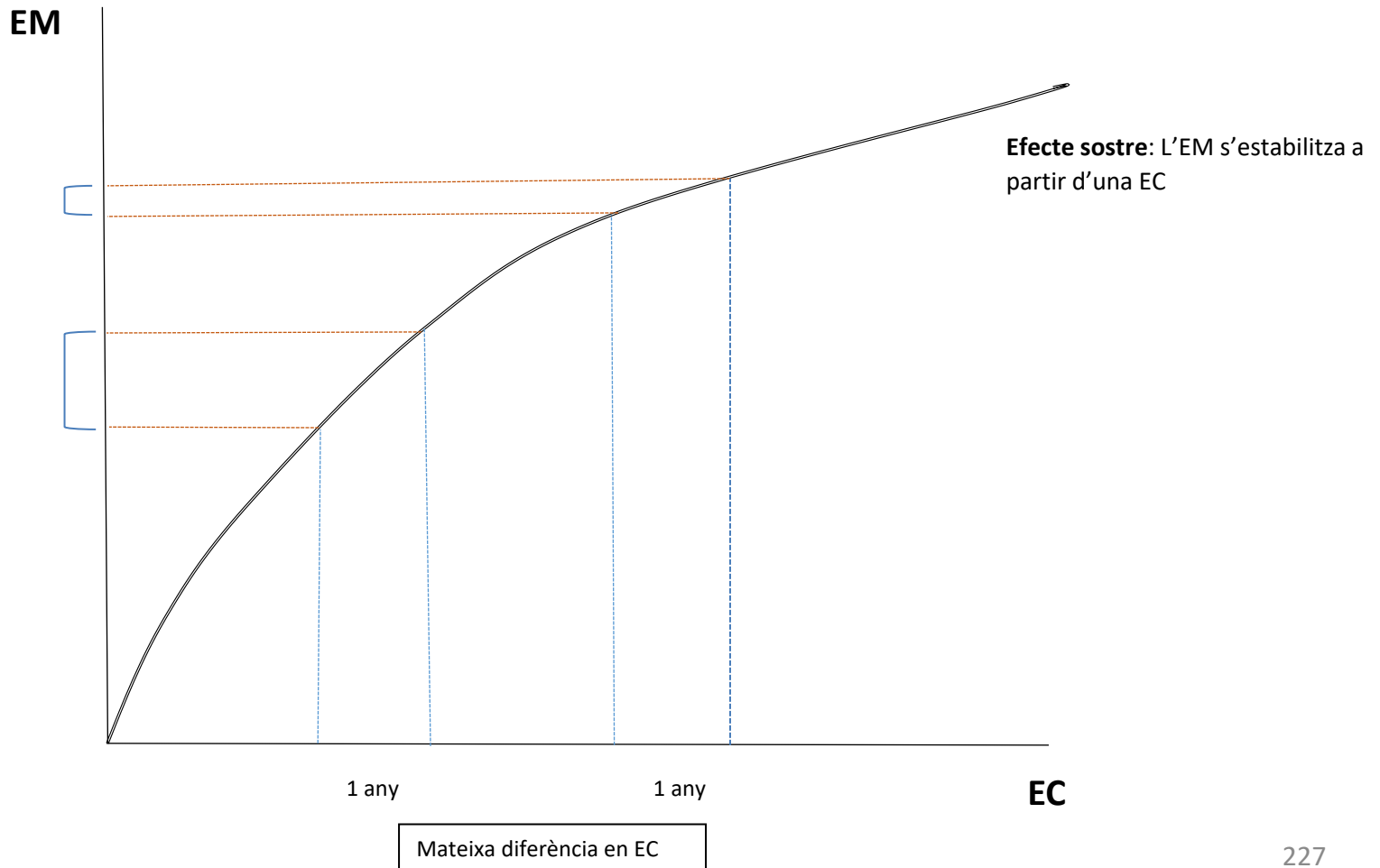
Resposta: Un QI de 111

$$QI = \frac{10}{9} \cdot 100 = 111$$

Normes Cronològiques: Inconvenients

- La relació entre EM i EC no és una unitat uniforme $\Rightarrow$
Les diferències obtingudes en EM no tenen el mateix significat per a totes les EC
- Les escales a les que s'apliquen tenen una edat màxima a partir de la qual la puntuació mitjana no augmenta amb l'edat

Normes Cronològiques: Inconvenients



Exercici 1: Normes cronològiques

En una prova de raonament abstracte s'han obtingut les següents puntuacions mitjanes per a edats diferents de nens/nenes que cursen educació primària:

Edat cronològica (EC)	Mitjana de les PD
7	25
8	30
9	35
10	40
11	45
12	50

1. Quina edat mental li correspondria a una puntuació de 45?
2. Quina puntuació hauria de treure un nen/a de 8 anys per obtenir un QI de 125?

Percentils

- Posició d'un individu en la variable que mesura el test respecte el seu grup de referència (grup normatiu)
- Punts o valors de la distribució que deixen un percentatge de casos per sota d'una PD i el percentatge complementari per sobre

Ex: Si a la PD = 47 li correspon el P_{30} voldrà dir que per sota d'aquest valor trobarem el 30% dels individus i per sobre tindrem el 70% dels casos restants

- El P_{50} es correspon a la mediana (M)
- Oscil·len entre 1 i 99
- L'escala percentil és ordinal (sense decimals)

Primers passos per obtenir Percentils

x	f_i
0	4
1	15
2	13
3	24
4	63
5	81
6	93
7	85
8	99
9	108
10	75
11	31
12	29
13	41
14	27
15	12
	$\Sigma f_i = 800$

X = puntuacions directes possibles i ordenades que es podien obtenir a la prova/test.
En el nostre exemple van de 0 a 15 punts

f_i = freqüències (numero de subjectes) que han obtingut la puntuació. P.e. 93 subjectes han obtingut una puntuació directa de 6

Σ (sumatori) de les f_i = numero total de subjectes del grup, que en el nostre cas és de $N = 800$.

Un cop disposem d'aquestes dues columnes de dades, les puntuació percentils es calculen mitjançant l'aplicació de la fórmula de la següent transparència.

Percentils

Transformació a percentils o centils: assignar a cada PD una puntuació percentual

$$P = \frac{F_{i-1} + 0.5 f_i}{N} \cdot 100$$

On:

- F_{i-1} - la freqüència acumulada per sota de la PD(x).
- f_i - la freqüència de persones que tenen la PD(x).
- N – nombre de persones del grup normatiu.

Percentils

- Avantatge: Facilitat d'interpretació
- Inconvenient: Escala ordinal (lloc $\leftrightarrow$ diferència)

	PD	P	
	1	1	
1 punt de diferència en PD	2	5	7 punts de diferència en P
	3	12	
2 punts de diferència en PD	4	15	7 punts de diferència en P
	5	19	

Diferències iguals entre percentils no impliquen diferències iguals entre puntuacions directes

Exemple: Percentils

x	f_i	F_i	P	P
0	4	4	0.25	1
1	15	19	1.4375	1
2	13	32	3.1875	3
3	24	56	5.5	6
4	63	119	10.9375	11
5	81	200	19.9375	20
6	93	293	30.8125	31
7	85	378	41.9375	42
8	99	477	53.4375	53
9	108	585	66.375	66
10	75	660	77.8125	78
11	31	691	84.4375	84
12	29	720	88.1875	88
13	41	761	92.5625	93
14	27	788	96.8125	97
15	12	800	99.25	99

$$P = \frac{F_{i-1} + 0.5 f_i}{N} \cdot 100$$

$$0 \rightarrow P = \frac{0 + 0.5 \cdot 4}{800} \cdot 100 = 0.25 \rightarrow 1$$

$$1 \rightarrow P = \frac{4 + 0.5 \cdot 15}{800} \cdot 100 = 1.44 \rightarrow 1$$

Exercici 2: Percentils

Es passa un test de 10 ítems a una mostra de 200 individus, obtenint-se la distribució de freqüències que es detalla:

x	f_i	F_i	P	P
0	4			
1	6			
2	10			
3	16			
4	40			
5	50			
6	46			
7	10			
8	12			
9	4			
10	2			

Calcular el percentil que li correspon a una puntuació directa de 4 i a una de 7

Puntuacions Típiques

Mesura quant se separa cada subj. de la mitjana → Distància en unitats de desviació tipus (unitat de mesura constant).

$$Z = \frac{X - \bar{X}}{S_x}$$

x - PD del subjecte

$\bar{x}$ - Mitjana del grup normatiu (GN)

S_x - desviació típica del GN

Si $z \sim N(0, 1) \Rightarrow -\infty \div \infty \Rightarrow -2,58 \div 2,58 \rightarrow 99\%$

Inconvenients:

- Valors negatius
- Valors decimals

Exemple: Puntuacions Típiques

Un test aplicat a un grup de 1500 subjectes es va distribuir normalment amb una mitjana de 25 i una variància de 16. Si un subjecte obté en aquest mateix test una puntuació directa de 31, quina puntuació típica li correspon?

$$z = \frac{31 - 25}{4} = 1,5$$

Exemple: Puntuacions Típiques

En una prova d'aritmètica (amb mitjana 20 i desviació típica 4) en Joan obté una puntuació de 24 i en Josep una de 28.

Aquestes dues puntuacions no ens indiquen directament el grau de coneixements aritmètics dels dos subjectes, ni ens permeten saber a quina distància estan entre sí.

Exemple: Puntuacions Típiques

Si calculem la puntuació típica respectiva (mitjançant la fórmula de z), veurem que en Joan obté una z d'1 i en Josep una de 2.

Joan $\longrightarrow x_i = 24$ $Z = 1$

Josep $\longrightarrow x_i = 28$ $Z = 2$

Sabem doncs que, ambdós subjectes estan per sobre de la mitjana, però en Josep es separa 2 desviacions tipus per sobre de la mitjana, mentre que en Joan tan sols es separa 1 desviació.

Exercici 3: Puntuacions Típiques

Si a més fan una altra prova de trigonometria amb $\bar{X} = 35$ i $S_X = 6$, i obtenen les següents puntuacions:

Joan $\longrightarrow x_i = 40$ $Z = 0,8\hat{3}$

Josep $\longrightarrow x_i = 38$ $Z = 0,5$

Globalment, quin dels dos té més coneixements de matemàtiques?

Exercici 4: Percentils i Puntuacions Típiques

Disposem de la següent taula de dades, amb la columna x de puntuacions obtingudes en una prova i la columna f_i (freqüència absoluta) amb el nombre de participants que han obtingut cadascuna de les possibles puntuacions:

x	f_i	F_i	P	z
0	4			
1	15			
2	13			
3	24			
4	63			
5	81			
6	93			
7	85			
8	99			
9	108			
10	75			
11	31			
12	29			
13	41			
14	27			
15	12			

Obtenir, per cadascuna de les puntuacions directes, els percentils i les puntuacions típiques, sabent que:

$$\bar{x} = 7,77$$

$$S_x = 3,15$$

Puntuacions Típiques Normalitzades

Assumim que les característiques mesurades per un test es distribueixen normalment en la població

Al treballar amb mostres, les dades generalment no són perfectament normals

Puntuacions típiques normalitzades (z_n)

PD ➡ Percentils ➡ Puntuacions Típiques Normalitzades

Si la distribució de les dades fos perfectament normal, aleshores

$$z = z_n$$

Exemple: Puntuacions típiques i puntuacions típiques normalitzades

x	f_i	P	z	P	p	z_n
0	4	0.25	-2.4656	1	0,01	-2.3263
1	15	1.4375	-2.1484	1	0,01	-2.3263
2	13	3.1875	-1.8311	3	0,03	-1.8808
3	24	5.5	-1.5138	6	0,06	-1.5548
4	63	10.9375	-1.1965	11	0,11	-1.2265
5	81	19.9375	-0.8792	20	0,20	-0.8416
6	93	30.8125	-0.5619	31	0,31	-0.4959
7	85	41.9375	-0.2447	42	0,42	-0.2019
8	99	53.4375	0.0726	53	0,53	0.0753
9	108	66.375	0.3898	66	0,66	0.4125
10	75	77.8125	0.7071	78	0,78	0.7722
11	31	84.4375	1.0244	84	0,84	0.9945
12	29	88.1875	1.3417	88	0,88	1.1750
13	41	92.5625	1.6589	93	0,93	1.4758
14	27	96.8125	1.9763	97	0,97	1.8808
15	12	99.25	2.2935	99	0,99	2.3263

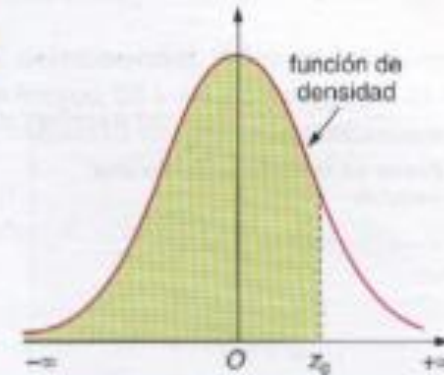
$$\bar{x} = 7,77$$

$$S_x = 3,15$$

veure taula distribució normal z

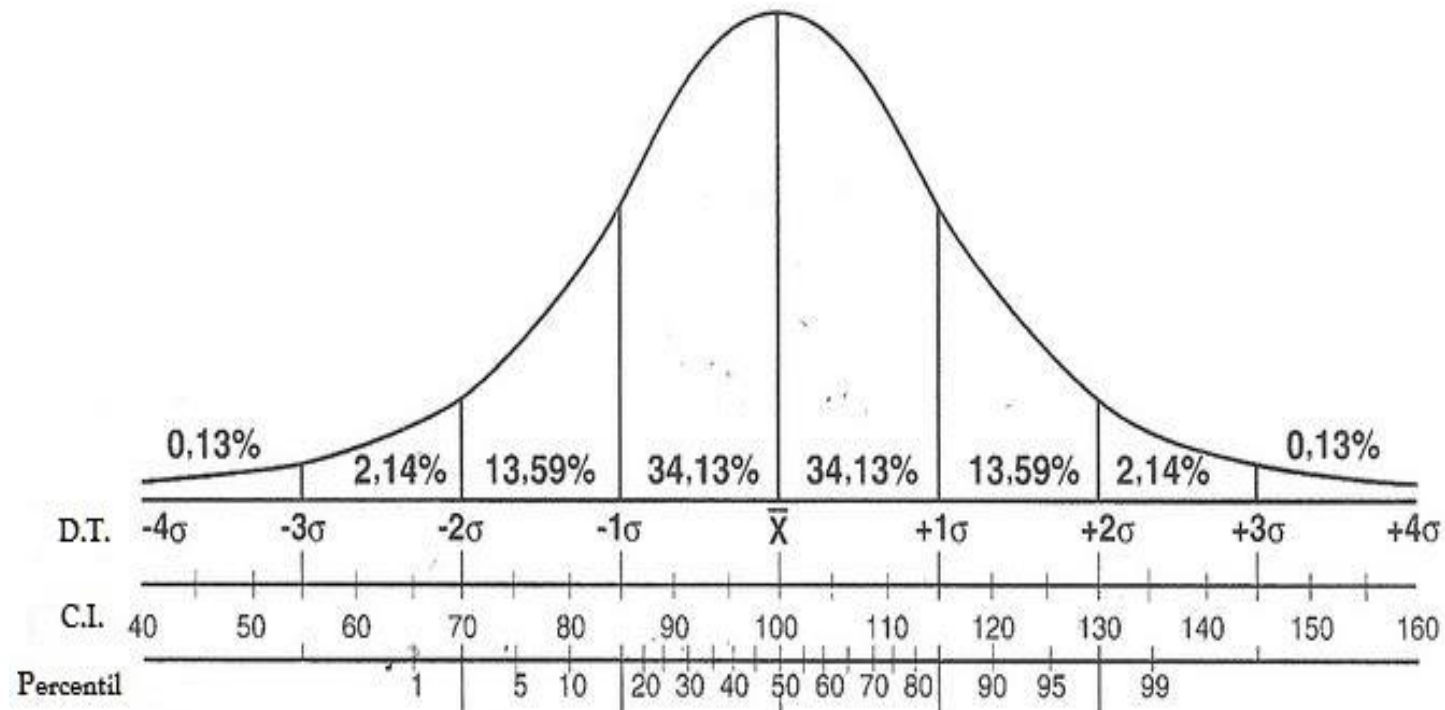
TABLA DE LA DISTRIBUCIÓN NORMAL $N(0, 1)$

$$P(Z \leq z_0) = \left\{ \begin{array}{l} \text{área del recinto} \\ \text{coloreado} \end{array} \right\}$$



z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9016
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319

Correspondència entre Puntuacions típiques i Percentils



Puntuacions Típiques Derivades

Totes les **Puntuacions típiques derivades** s'obtenen a partir de les puntuacions típiques originals (o primitives) z :

$$Z_d = \bar{X}_d + S_d \cdot Z_x$$

On: $\bar{X}_d$ - mitjana de les puntuacions típiques derivades

s_d - desviació de les puntuacions típiques derivades

z_x - puntuació típica original

Aplica una **transformació lineal** del tipus $y = a + bz$ i, per tant, permet obtenir qualsevol nova escala amb la mitjana i la desviació típica que es desitgi

Puntuacions Típiques Derivades Normalitzades

En cas que la puntuació derivada s'obtingui de la z_n , parlarem de
Puntuacions típiques derivades normalitzades

$$Z_{d_n} = \bar{X}_d + S_d \cdot Z_n$$

On: $\bar{X}_d$ - mitjana de les puntuacions típiques derivades.

s_d - desviació de les puntuacions típiques derivades.

z_n - puntuació típica normalitzada original

Tipus de Puntuacions Típiques Derivades

A cada z li correspon una puntuació derivada:

$$\text{Puntuacions } T \rightarrow T = 50 + 10 z$$

$$\text{Puntuacions } D \rightarrow D = 50 + 20 z$$

$$QI \rightarrow QI = 100 + 15 z$$

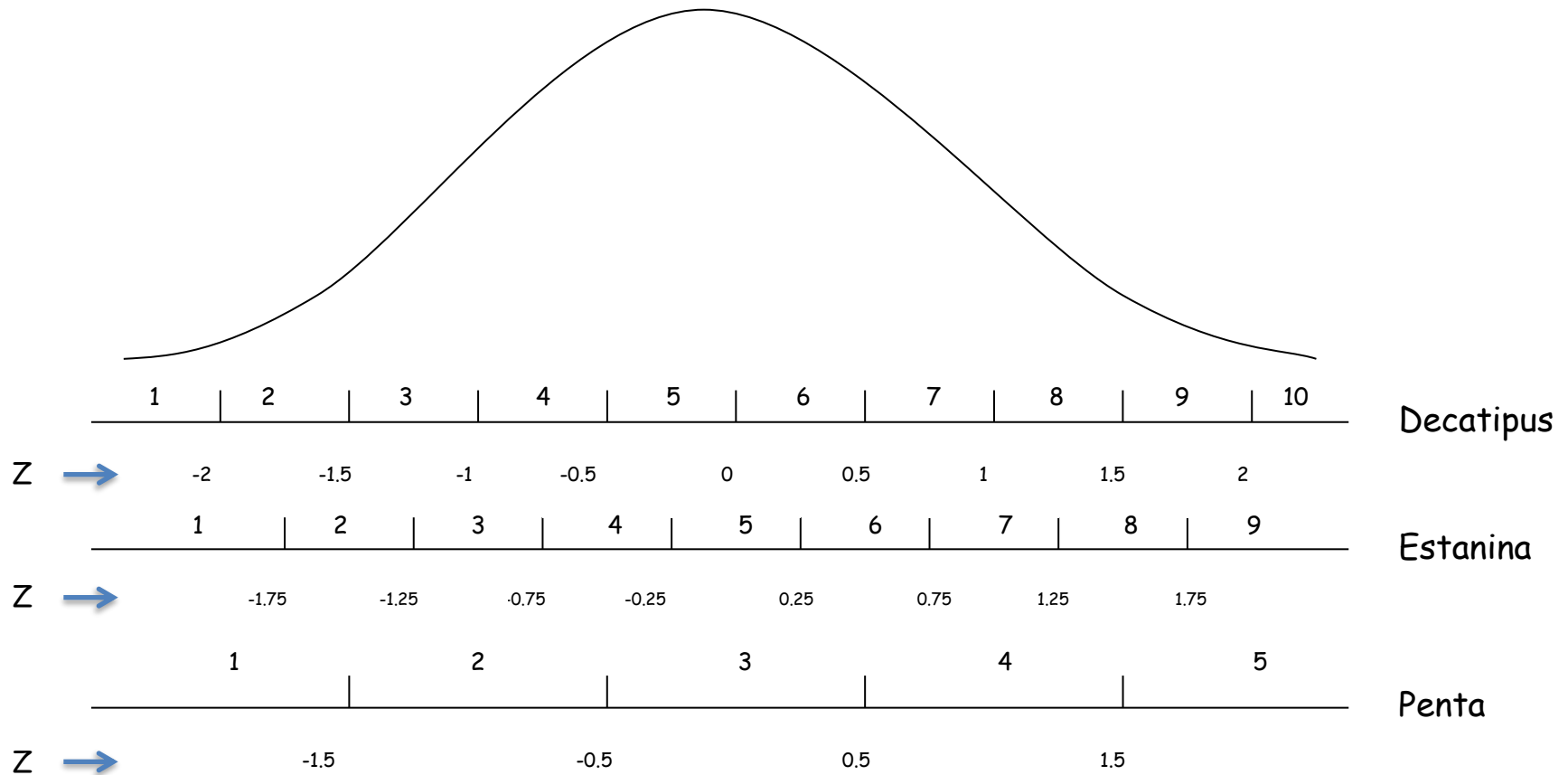
A diferents valors de z els correspon la mateixa puntuació derivada:

$$\text{Sten o decatipus (1-10)} \rightarrow \text{Sten} = 5,5 + 2 z$$

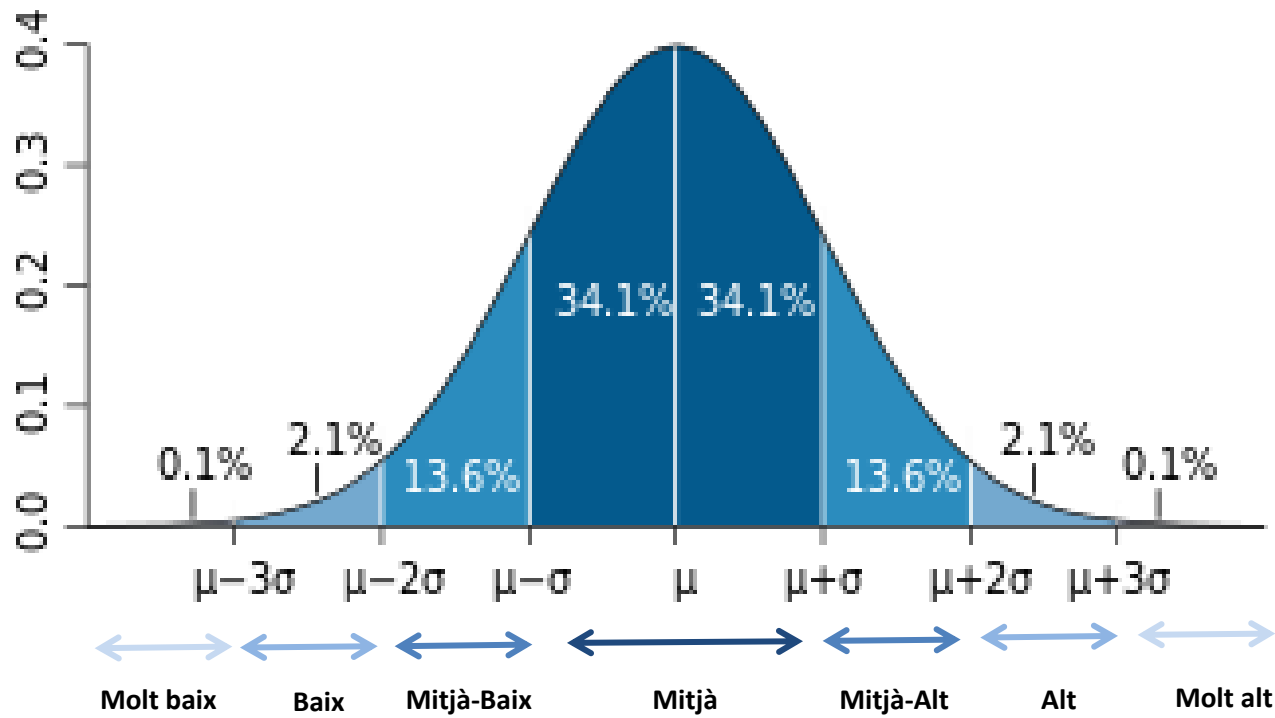
$$\text{Estanina (1-9)} \rightarrow e = 5 + 2 z$$

$$\text{Penta (1-5)} \rightarrow p = 3 + z$$

Puntuacions Típiques Derivades



Interpretació Puntuacions Típiques i Típiques Derivades



Exemple: Transformar la PD 9 de l'exemple anterior a PTDs

$$z = 0.3898$$

$$\text{Puntuacions T} \rightarrow T = 50 + 10 \cdot 0,3898 = 53,898 \rightarrow 54$$

$$\text{Puntuacions D} \rightarrow D = 50 + 20 \cdot 0,3898 = 57,796 \rightarrow 58$$

$$\text{QI} \rightarrow \text{QI} = 100 + 15 \cdot 0,3898 = 105,847 \rightarrow 106$$

$$\text{Decatipus (1-10)} \rightarrow \text{Sten} = 5,5 + 2 \cdot 0,3898 = 6.28 \rightarrow 6$$

$$\text{Estanina (1-9)} \rightarrow e = 5 + 2 \cdot 0,3898 = 5,7796 \rightarrow 6$$

$$\text{Penta (1-5)} \rightarrow p = 3 + 0,3898 = 3,3898 \rightarrow 3$$

x	z	T	D	QI	Decatipus	Estanina	Penta
9	0,3898	54	58	106	6	6	3

Exercici 5: Realitza la transformació de puntuacions de les següents dades

<i>x</i>	<i>f_i</i>	<i>F_i</i>	<i>P</i>	<i>z</i>	<i>T</i>	<i>D</i>	<i>QI</i>	<i>Sten</i>	<i>Estanina</i>	<i>Penta</i>
0	4									
1	15									
2	13									
3	24									
4	63									
5	81									
6	93									
7	85									
8	99									
9	108									
10	75									
11	31									
12	29									
13	41									
14	27									
15	12									

$\bar{x} = 7,77$

$S_x = 3,15$

Exercici 6: Baremació

Un individu obté una puntuació de 70 en una prova de mitjana 65, desviació típica 9 i una fiabilitat de 0,85.

- 1) Quina seria la seva puntuació en una escala T?
- 2) Quin seria l'interval (en puntuacions T) de la puntuació vertadera d'aquest individu (IC=95%, $z=1,96$)?
- 3) En quin nivell es trobaria l'individu?

Exercici 7: Baremació

Sabem que en una prova de coneixements la persona 1, en la escala T, obté una puntuació que coincideix amb la de la persona 2, en la escala D. Això implica que les dues persones han obtingut la mateixa puntuació a la prova? Raona la resposta. En cas afirmatiu, quina seria la puntuació obtinguda?

Manual de test



Aquesta obra està subjecta a una llicència de Creative Commons Reconeixement- No comercial- Compartir Igual 3.0 Espanya
(<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>).



Lectures:

Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G. y Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37(3), 192-197.

Muñiz, J., y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16.

El manual com a guia de l'usuari

El material d'ús de qualsevol instrument de mesura hauria de presentar necessàriament dues parts:

El **test** en qüestió: ítems i eina de registre de respostes

AV-I. Aprendizaje verbal inmediato

Instrucciones: En cada uno de los 3 ensayos, lea en voz alta la lista de 10 palabras con un intervalo de unos 3 segundos entre palabras. Después pida al evaluado que diga todas las palabras que recuerde, sin importar el orden. Registre cuántas palabras ha recordado en cada ensayo. Después de que el evaluado haya recordado todas las palabras que pueda, diga: "Trate de recordar estas palabras lo mejor que pueda porque voy a pedirle la lista otra vez dentro de algunos minutos".

	Tambor	Cortina	Campana	Café	Escuela	Padre	Luna	Jardín	Sombrero	Granjero	TOTAL
E1											
E2											
E3											

Intrusiones

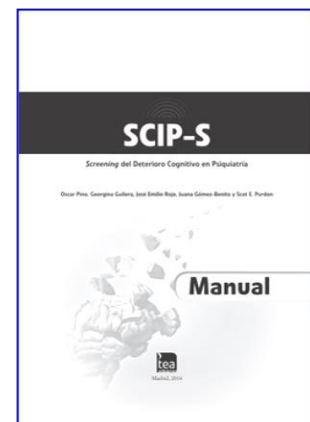
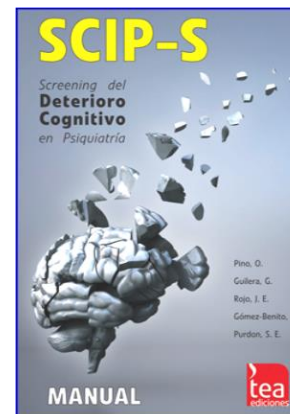
	E1	E2	E3
E1	☐	☐	☐
E2	☐	☐	☐
E3	☐	☐	☐

Perseveraciones

	E1	E2	E3
E1	☐	☐	☐
E2	☐	☐	☐
E3	☐	☐	☐

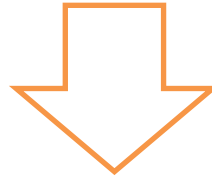
AV-I

El **manual del test**: informació sobre el test que permet conèixer les característiques del test i fer-ne un bon ús



El manual com a guia de l'usuari

L'elaboració del manual és la fase final de la construcció d'un instrument de mesura



Ha d'incloure informació relativa a totes les fases que s'han seguit en la construcció del test



Convé avaluar rigorosament i sistemàtica el test elaborat

Utilització dels tests

Estratègies de millora en l'ús de tests

1. Directrius sobre l'ús de tests:

European Federation of Psychological Assessment: <http://www.efpa.be>

Colegio Oficial de Psicólogos: <http://www.cop.es>

International Test Commission: <http://www.intestcom.org>

2. Formació dels usuaris, competències mínimes

3. Estàndards tècnics sobre l'ús de tests

4. Preparació i entrenament

5. Utilització de les dades, aspectes ètics

6. Avaluació de la qualitat dels tests

Directrius Internacionals per a l'ús dels tests

Colegio Oficial de Psicólogos i International Test Commission

<https://www.cop.es/index.php?page=directrices-internacionales>

Introducción

Directrices

Objetivo

Ámbito de aplicación

Destinatarios

Aspectos contextuales

Conocimientos

1. Uso ético de los tests

- 1.1 Actuar de forma ética y profesional .
- 1.2. Asegurarse de que son competentes para el uso de los tests.
- 1.3. Responsabilizarse del uso que hacen de los tests.
- 1.4. Asegurarse de que los materiales del test están seguros.
- 1.5. Asegurarse de que los resultados de los tests se tratan confidencialmente.

2. Utilización adecuada de los tests

- 2.1. Estimar la utilidad potencial de los tests en una situación evaluativa.
- 2.2. Elegir tests técnicamente correctos y adecuados a cada situación.
- 2.3. Prestar atención a los aspectos relacionados con el sesgo de los tests.
- 2.4. Hacer los preparativos necesarios para la aplicación del test.
- 2.5. Aplicar los tests adecuadamente.
- 2.6. Puntuar y analizar los resultados de los tests con precisión.
- 2.7. Interpretar los resultados adecuadamente.
- 2.8. Comunicar los resultados de forma clara y precisa.
- 2.9. Revisión de la adecuación del test y de su uso.

Deu passos en la construcció d'un test

1. Marc general
2. Definició de la variable mesurada
3. Especificacions
4. Construcció dels ítems
5. Edició
6. Estudis pilot
7. Selecció d'altres instruments de mesura
8. Aplicació del test
9. Propietats psicomètriques
10. Versió final del test

Deu passos en la construcció d'un test

1. Marc general.

Motiu per desenvolupar-lo i context en el qual s'aplicarà: població objecte de la mesura (pacients, empreses...), i circumstàncies d'aplicació (lloc, mitjans,...)

2. Definició de la variable mesurada.

Delimitar la variable objecte de mesura considerant totes les dimensions rellevants i identificar amb claredat els comportaments més representatius de la variable

3. Especificacions.

Requeriments d'aplicació de l'instrument:

- Ítems: tipus, número, longitud, contingut i distribució dels ítems, instruccions i aspectes de seguretat
- Aplicació: Suport, tipus, lloc, rang d'edat dels participants, adaptacions per a persones que no poden desenvolupar la tasca en igualtat de condicions, etc.

4. Construcció dels ítems.

Principis bàsics per a la construcció d'ítems: representativitat, rellevància, diversitat, claredat, senzillesa i comprensibilitat

5. Edició.

Composar i imprimir la primera versió d'ítems i construir la base de dades de correcció. Validesa aparent (aspecte físic de la prova)

Deu passos en la construcció d'un test

6. Estudis pilot.

Qualitatiu: Grups discussió per debatre aspectes de l'instrument de mesura.
Quantitatiu: Valora les propietats mètriques
L'objectiu és seleccionar els ítems amb majors garanties mètriques per maximitzar les propietats finals de l'instrument d'avaluació

7. Selecció d'altres instruments de mesura.

Les relacions amb variables externes permeten obtenir evidències de la validesa de l'instrument

8. Aplicació del test.

Inclou com s'ha seleccionat la mostra (tipus, grandària i procediment), l'aplicació de la prova als participants i el control de la qualitat i seguretat de la base de dades

9. Propietats psicomètriques.

Un cop aplicat el test a la mostra, cal estudiar totes les seves propietats psicomètriques.
Imprescindible el rigor metodològic

10. Versió final del test.

Elaborar la versió definitiva del test i el manual que permeti la seva utilització a altres persones o institucions