

Psicometria: Recursos Pràctics



Irene Checa Esquiva
Begoña Espejo Tort

Esta publicación no puede ser reproducida, ni total ni parcialmente, ni registrada en, o transmitida por, un sistema de recuperación de información, de ninguna forma ni por ningún medio, sea fotomecánico, fotoquímico, electrónico, por fotocopia o por cualquier otro, sin el permiso de la editorial. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

PSICOMETRIA: RECURSOS PRÀCTICS

Autor: Irene Checa Esquiva
Begoña Espejo Tort

I.S.B.N.: 978-84-946437-6-7

Editorial: Ángeles Carrillo Baeza

Formato: Digital: descarga y online. PDF

BLOC 1. ESCALAMENT	7
B1. PRÀCTICA 1. ESCALAMENT DE THURSTONE.....	8
PRÀCTICA 2. ESCALAMENT DE GUTTMAN	17
B1. PRÀCTICA 3. LIKERT. ANÀLISI D'ÍTEMS DE TIPUS AFECTIU.....	27
B1 PRÀCTICA 4. ANÀLISI D'ÍTEMS DE TIPUS COGNITIU	50
BLOC 2. TEORIA CLÀSICA DE TESTS (TCT).....	61
B2. PRÀCTICA 1. SUPÒSITS TEORIA CLÀSSICA DE TESTS (TCT)	63
B2. PRÀCTICA 2. FORMES D'ESTIMAR LA FIABILITAT.....	67
B2. PRÀCTICA 3. FACTORS QUE AFECTEN LA FIABILITAT	74
B2. PRÀCTICA 4. ESTIMACIÓ DE PUNTUACIONS DES DE LA TCT.....	79
BLOC 3. VALIDESA.....	83
PRÀCTICA 1. EVIDÈNCIES DE VALIDESA EXTERNA.....	85
B3. PRÀCTICA 2. EVIDÈNCIES DE VALIDESA INTERNA	93
B3. PRÀCTICA 3. FACTORS QUE AFECTEN LA VALIDESA	109
BLOC 4: TEORIA DE RESPOSTA A L'ÍTEM (TRI)	117
B4. PRÀCTICA 1. PARÀMETRES DE LA TRI.....	119
B4. PRÀCTICA 2. MODELS DE LA TRI I CÀLCULS DE PROBABILITATS	127

Bloc 1. Escalament

B1. Pràctica 1. ESCALAMENT DE THURSTONE

Ara posarem en pràctica el procediment ideat per Thurstone de les comparacions aparellades per a escalar objectes. Es demana als subjectes que valoren dos a dos els estímuls i que indiquen quin ítem del parell té més quantitat de la característica en qüestió. Per a fer-ho, els participants escriuen en un full en blanc tots els parells possibles en una matriu i es demana, per a cada parell, que assignen un 1 a l'ítem preferit (o que tinga la característica en més mesura) i 0 a l'altre ítem del parell.

Considerem un exemple en el qual heu de valorar la **imparcialitat en els informatius** de les cadenes de televisió més importants a escala estatal. Les cadenes que s'han de valorar són les següents: TV1, La 2, Antena 3, Cuatro, Tele 5 i La Sexta. Per a fer-ho, escriueu en un paper una matriu com la següent:

	TVE1	La 2	Antena 3	Cuatro	Tele 5	La Sexta
TVE1	--					
La 2		--				
Antena 3			--			
Cuatro				--		
Tele 5					--	
La Sexta						--

A continuació, heu d'indicar parell a parell la vostra valoració de cada cadena respecte a les altres. **Es comparen sempre els estímuls de la columna respecte els de cada fila.** Per exemple, si comparem TVE1 amb La 2, i jo considere que TVE1 té uns informatius més imparcials que La 2, assignaré un 1 a la casella corresponent. Si compare TVE1 amb Antena 3 i considere que Antena 3 té informatius més imparcials que TVE1, assignaré un 0 a la casella corresponent:

	TVE1	La 2	Antena 3	Cuatro	Tele 5	La Sexta
TVE1	--					
La 2	1	--				
Antena 3	0		--			
Cuatro				--		
Tele 5					--	
La Sexta						--

Omplirem així la matriu per davall de la diagonal i tindrem finalment una matriu d'1 i 0. La part de la taula per damunt de la diagonal té els valors oposats, de manera que, si s'assigna un 1 a TVE1, perquè es considera més imparcial que La 2, s'ha d'assignar un 0 a la part superior de la diagonal quan compare La 2 amb TVE1, i assignaré un 1 a la cel·la en la qual compare Antena 3 i TVE1, així:

	TVE1	La 2	Antena 3	Cuatro	Tele 5	La Sexta
TVE1	--	0	1			
La 2	1	--				
Antena 3	0		--			
Cuatro				--		
Tele 5					--	
La Sexta						--

Una vegada que cada persona té tots els parells puntuats amb 1 i 0, se sumen tots els 1 assignats a cada comparació i es crea la matriu de freqüències amb el nombre d'1 de cada cel·la. Suposem que cent persones han emès els seus judicis. Tenim totes les taules individuals i sumem els 1 que hi ha a cada cel·la. Al final, tindrem una matriu com aquesta:

MATRIU DE FREQUÈNCIES OBSERVADES

	TVE1	La 2	Antena 3	Cuatro	Tele 5	La Sexta
TVE1	--	20	50	70	60	70
La 2	80	--	50	30	20	60
Antena 3	50	50	--	10	20	30
Cuatro	30	70	90	--	40	70
Tele 5	40	80	80	60	--	70
La Sexta	30	40	70	30	30	--

Ja tenim la matriu de comparacions dels participants i podem començar a calcular les preferències del grup. Per a fer-ho, hem de seguir DIVERSOS PASSOS.

1) **Sumem totes les freqüències per columnes**

MATRIU DE FREQUÈNCIES OBSERVADES AMB TOTALS

	TVE1	La 2	Antena 3	Cuatro	Tele 5	La Sexta
TVE1	--	20	50	70	60	70
La 2	80	--	50	30	20	60
Antena 3	50	50	--	10	20	30
Cuatro	30	70	90	--	40	70
Tele 5	40	80	80	60	--	70
La Sexta	30	40	70	30	30	--
Σ	230	260	340	200	170	300

2) **S'ordenen les columnes** de la matriu en funció del sumatori de freqüències que acabem de fer, **de menor a major**.

MATRIU DE FREQUÈNCIES OBSERVADES ORDENADA

	Tele 5	Cuatro	TVE1	La 2	La Sexta	Antena 3
TVE1	60	70	--	20	70	50
La 2	20	30	80	--	60	50
Antena 3	20	10	50	50	30	--
Cuatro	40	--	30	70	70	90
Tele 5	--	60	40	80	70	80
La Sexta	30	30	30	40	--	70
Σ	170	200	230	260	300	340

Ja tenim l'ordre d'imparcialitat dels informatius de les diferents cadenes, segons que ha opinat el grup, **de menys a més imparcialitat**: Tele 5, Cuatro, TVE-1, La 2, La Sexta i Antena 3.

Ara cal calcular la **distància percebuda** entre les cadenes de televisió. Perquè una cadena es pot percebre molt més imparcial que una altra, però igual d'imparcial que una tercera.

- 3) **Es calculen les proporcions** de cada cel·la. Com han contestat cent persones, cal dividir cada freqüència per 100.

MATRIU DE PROPORCIONS

	Tele 5	Cuatro	TVE1	La 2	La Sexta	Antena 3
TVE1	0.60	0.70	--	0.20	0.70	0.50
La 2	0.20	0.30	0.80	--	0.60	0.50
Antena 3	0.20	0.10	0.50	0.50	0.30	--
Cuatro	0.40	--	0.30	0.70	0.70	0.90
Tele 5	--	0.60	0.40	0.80	0.70	0.80
La Sexta	0.30	0.30	0.30	0.40	--	0.70

- 4) **Les proporcions es transformen en puntuacions típiques.** Per a fer-ho utilitzarem la taula de puntuacions típiques i buscarem les proporcions dins de la taula, per veure quina puntuació típica correspon a cada proporció.

MATRIU DE PUNTUACIONS TÍPIQUES

	Tele 5	Cuatro	TVE1	La 2	La Sexta	Antena 3
TVE1	0.26	0.53	--	-0.84	0.53	0
La 2	-0.84	-0.52	0.84	--	0.26	0
Antena 3	-0.84	-1.28	-0.52	0	-0.52	--
Cuatro	-0.25	--	-0.52	0.53	0.53	1.28
Tele 5	--	0.26	-0.25	0.84	0.53	0.84
La Sexta	-0.52	-0.52	-0.52	-0.25	--	0.53

IMPORTANT!!!

Les puntuacions típiques indiquen la **DISTÀNCIA PERCEBUDA ENTRE DOS ESTÍMULS!** A més, els **valors positius** indiquen **preferència per l'estímul de la columna** enfront del de la fila, mentre que els valors negatius indiquen preferència per l'estímul de la fila enfront del de la columna.

Per exemple, la puntuació típica **0,26** indica que es prefereix o es considera més imparcial la cadena Tele 5 que TVE1. No obstant això, la puntuació típica **-0,84** indica que es considera menys imparcial Tele 5 que La 2 (es prefereix La 2). I, a més, **la distància percebuda entre Tele 5 i TVE1 és menor** (en termes absoluts) **que la distància percebuda entre Tele 5 i La 2** (encara que en aquest cas siga en sentit contrari). Un altre exemple més: la distància percebuda entre Antena 3 i Cuatro és la més gran de totes (**1,28**) i indica clarament que Antena 3 es percep molt més imparcial que Cuatro, ja que el valor de la puntuació típica és positiu.

- 5) **Es calcula la mitjana aritmètica de les puntuacions típiques per a cada columna.** Aquest valor indicarà la **distància mitjana** de cada cadena respecte a totes les altres en conjunt.

IMPORTANT!!!

- Per a fer la suma CAL MANTENIR EL SIGNE.
- Per a calcular la mitjana CAL DIVIDIR PEL NOMBRE TOTAL D'ESTÍMULS (en aquest cas, sis cadenes de televisió).

MATRIU DE PUNTUACIONS TÍPIQUES

	Tele 5	Cuatro	TVE1	La 2	La Sexta	Antena 3
TVE1	0.26	0.53	--	-0.84	0.53	0
La 2	-0.84	-0.52	0.84	--	0.26	0
Antena 3	-0.84	-1.28	-0.52	0	-0.52	--
Cuatro	-0.25	--	-0.52	0.53	0.53	1.28
Tele 5	--	0.26	-0.25	0.84	0.53	0.84
La Sexta	-0.52	-0.52	-0.52	-0.25	--	0.53
Σ	-2.19	-1.53	-0.97	0.28	1.33	2.65
$\Sigma/n=\Sigma/6$	-0.37	-0.26	-0.16	0.05	0.22	0.44

La mitjana **-0,37** indica que Tele 5 es considera menys imparcial (valor negatiu) que les altres cadenes de televisió considerades conjuntament, mentre que la mitjana **0,44** indica que Antena 3 es percep com més imparcial (valor positiu) que les altres cadenes globalment.

6) Finalment, es calcula el valor d'escala de cada estímul.

En realitat, ja hem calculat les distàncies percebudes entre estímuls, però per a eliminar valors negatius i facilitar la percepció de les distàncies, restem A TOTES LES MITJANES DE PUNTUACIONS TÍPIQUES el menor VALOR DE TOTS. D'aquesta manera, els valors escalars començaran en 0 i s'hi podran observar més fàcilment les distàncies. En aquest cas, el menor valor de tots és el primer, -0,37. Per tant, restem a totes les mitjanes aquest valor.

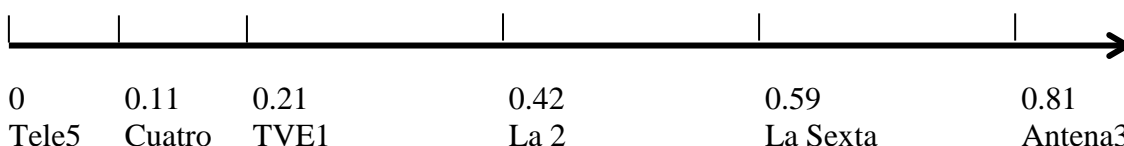
IMPORTANT!!!

Per aquesta raó, **TOTS ELS VALORS ESCALARS SERAN POSITIUS**, i el menys preferit (cadena percebuda com menys imparcial) serà 0.

MATRIU DE PUNTUACIONS TÍPIQUES

	Tele 5	Cuatro	TVE1	La 2	La Sexta	Antena 3
TVE1	0.26	0.53	--	-0.84	0.53	0
La 2	-0.84	-0.52	0.84	--	0.26	0
Antena 3	-0.84	-1.28	-0.52	0	-0.52	--
Cuatro	-0.25	--	-0.52	0.53	0.53	1.28
Tele 5	--	0.26	-0.25	0.84	0.53	0.84
La Sexta	-0.52	-0.52	-0.52	-0.25	--	0.53
Σ	-2.19	-1.53	-0.97	0.28	1.33	2.65
$\Sigma/n = \Sigma/6$	-0.37	-0.26	-0.16	0.05	0.22	0.44
$(\Sigma/n) - (-0,37)$	0	0.11	0.21	0.42	0.59	0.81

Si dibuixem una línia i col·loquem damunt les distàncies, començant per 0, es podrà apreciar visualment la distància que hi ha entre la imparcialitat percebuda de les cadenes per als participants:



S'observa que la cadena percebuda com a menys imparcial és Tele 5, i darrere, de prop, Cuatro i després TVE1. A continuació, i a certa distància, hi ha La 2, seguida també a distància per La Sexta. La cadena percebuda com a menys imparcial clarament és Antena 3, a certa distància de la següent, que és La Sexta.

Ara farem el mateix exercici, però **amb altres dades**. Heu d'escriure en el paper la matriu de freqüències d'1 i 0 amb la vostra opinió sobre la **PREFERÈNCIA PER PROGRAMES D'ENTRETENIMENT** de les cadenes de televisió generalistes. Després farem recompte d'1 i 0 en cada cel·la i anotarem el nombre de participants (N) per formar la matriu de freqüències observades.

MATRIU DE FREQUÈNCIES OBSERVADES N=

	TVE1	La 2	Antena 3	Cuatro	Tele 5	La Sexta
TVE1	--					
La 2		--				
Antena 3			--			
Cuatro				--		
Tele 5					--	
La Sexta						--
Σ						

MATRIU DE FREQUÈNCIES OBSERVADES ORDENADA

TVE1						
La 2						
Antena 3						
Cuatro						
Tele 5						
La Sexta						

MATRIU DE PROPORCIIONS

TVE1						
La 2						
Antena 3						
Cuatro						
Tele 5						
La Sexta						

MATRIU DE PUNTUACIONS TÍPIQUES

TVE1						
La 2						
Antena 3						
Cuatro						
Tele 5						
La Sexta						
Σ						
$\Sigma/n = \Sigma/6$						

VALORS ESCALARS (restem a tots els valors la puntuació típica mitjana menor)

Valor escalar						

Posiciona les televisions per ordre de preferència respecte a la programació d'entreteniment:



1. Quin és l'ordre de preferència pel que fa a la programació d'entreteniment? Quina cadena preferiu vosaltres? I la que menys preferiu?

2. Quina és la distància percebuda entre Antena 3 i La Sexta? Quina cadena es prefereix entre totes dues?

3. I entre La Sexta i Cuatro? Quina cadena es prefereix entre l'una i l'altra?

4. Són iguals les distàncies? A què creus que pot ser degut?

Pràctica 2. ESCALAMENT DE GUTTMAN

Ara posarem en pràctica l'escalament de Guttman amb dades de sis ítems i deu subjectes. Cal comprovar si l'escala de sis ítems compleix els criteris establerts per Guttman i es pot considerar un bon escalograma. Aquest tipus d'escalament està pensat per a ítems d'APTITUD i, per tant, tenen cert grau de dificultat. La idea de Guttman és poder escalar a ítems i subjectes al mateix temps, de manera que es pretén ORDENAR ELS ÍTEMS DE MENOR A MAJOR DIFICULTAT i també ORDENAR els SUBJECTES DE MENOR A MAJOR PUNTUACIÓ.

Així, si en aquesta escala una persona aconsegueix un 5, és perquè ha encertat ELS CINQ PRIMERS ÍTEMS, i una altra persona que ha obtingut una puntuació de 3 només hi haurà encertat ELS TRES PRIMERS ÍTEMS. A més, el primer subjecte tindrà més «quantitat» del constructe que es mesura que el segon subjecte. Aquestes són les dades que hem recollit:

Subjectes	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6
1	0	0	1	0	0	1
2	1	0	0	0	1	0
3	1	1	1	0	0	0
4	1	1	0	0	0	0
5	1	1	0	1	1	1
6	1	0	0	0	0	0
7	1	1	1	0	1	1
8	0	1	1	0	0	1
9	1	0	0	0	0	0
10	1	0	1	0	0	0

Recordeu que són ítems d'aptitud i que, per tant, hi ha una resposta correcta que es puntua amb 1. Les respostes incorrectes es puntuen amb 0. Per a comprovar si aquests sis ítems es poden considerar un bon escalograma cal seguir els passos que hi ha tot seguit.

- 1) **Calculem la suma total per columna**, que ens donarà el nombre de vegades que s'ha encertat cada ítem.

Subjectes	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6
1	0	0	1	0	0	1
2	1	0	0	0	1	0
3	1	1	1	0	0	0
4	1	1	0	0	0	0
5	1	1	0	1	1	1
6	1	0	0	0	0	0
7	1	1	1	0	1	1
8	0	1	1	0	0	1
9	1	0	0	0	0	0
10	1	0	1	0	0	0
Σencerts	8	5	5	1	3	4

En veure els totals, podem comprovar que els ítems tal com s'han presentat als subjectes NO estan ordenats de menor a major dificultat. Per tant, hem de reordenar per estudiar posteriorment les respostes totals dels subjectes.

2) **Ordenem la matriu anterior per columnes de menor a major puntuació en el sumatori (Σ).**

Subjectes	Ítem 1	Ítem 2	Ítem 3	Ítem 6	Ítem 5	Ítem 4
1	0	0	1	1	0	0
2	1	0	0	0	1	0
3	1	1	1	0	0	0
4	1	1	0	0	0	0
5	1	1	0	1	1	1
6	1	0	0	0	0	0
7	1	1	1	1	1	0
8	0	1	1	1	0	0

9	1	0	0	0	0	0
10	1	0	1	0	0	0
Σencerts	8	5	5	4	3	1

Ara els ítems sí que estan ordenats en funció de la dificultat: ítem 1, ítem 2, ítem 3, ítem 6, ítem 5, ítem 4.

- 3) Calculem la **puntuació total de cada subjecte** (suma d'1). A partir d'aquesta puntuació, **marquem i comptem els errors** (*) que conté cada patró de resposta en relació amb l'escala tipus que hi correspondria.

IMPORTANT!!!

ESCALA TIPUS: a un encert només pot precedir un encert, a un error només pot seguir un altre error.

Per exemple, el subjecte maig ha obtingut una puntuació de 5. L'escala tipus o patró de respostes esperat, si els ítems formen un bon escalograma, hauria de ser **111110**. Els errors seran els 1 i els 0 que no figuren en les posicions d'aquest patró de resposta. En l'exemple veiem que hi ha dos errors, ja que en l'ítem 3 hauria d'aparèixer un 1 i en el seu lloc apareix un 0. Així mateix, en l'ítem 6 hauria d'aparèixer un 0, però apareix un 1. Aquests són els dos errors.

D'altra banda, el subjecte 1 té una puntuació total de 2. El patró de resposta hauria de ser 110000. No obstant això, en els ítems 1 i 2 apareix un 0 quan la puntuació hauria de ser un 1. En l'ítem 6 apareix un 1, quan la puntuació hauria de ser 0. Aquests són els tres errors. Així mateix, el subjecte 2 i el subjecte 10 tenen cadascun una puntuació total de 2. Però també presenten dos errors en el patró de resposta. I el mateix passa amb el subjecte 8, que presenta una puntuació total de 3 i dos errors en el patró.

Subjectes	Ítem 1	Ítem 2	Ítem 3	Ítem 6	Ítem 5	Ítem 4	Punt.	Errors
1	0*	0	1*	1*	0	0	2	3
2	1	0*	0	0	1*	0	2	2
3	1	1	1	0	0	0	3	0
4	1	1	0	0	0	0	2	0
5	1 1	1 1	0* 1	1 1	1 1	1* 0	5	2
6	1	0	0	0	0	0	1	0
7	1	1	1	1	1	0	5	0
8	0*	1	1	1*	0	0	3	2
9	1	0	0	0	0	0	1	0
10	1	0*	1*	0	0	0	2	2
Σcerts	8	5	5	4	3	1		
Σerrors	2	2	3	2	1	1		11

4) A continuació, estimem quant s'aproxima l'escala a un escalograma perfecte mitjançant el **càlcul del coeficient de reproductibilitat (CR)**.

El CR dóna una valoració GLOBAL de l'escala com a escalograma. Si tots els ítems presenten el patró de resposta esperat en funció de la puntuació total del subjecte, llavors l'escalograma és perfecte (CR = 1). Però això és molt difícil o quasi impossible. Segons Guttman, per a considerar un conjunt d'ítems com a un bon escalograma **almenys el 90% de les respostes** han de correspondre a les respostes esperades segons l'escala tipus corresponent, és a dir, cal obtenir un **CR ≥ 0.90**. El CR es calcula de la manera següent:

$$CR = 1 - \frac{\sum \text{errors}}{N \cdot k} \quad N = \text{nombre de subjectes}; k = \text{nombre d'ítems}$$

$$CR = 1 - \frac{11}{10 \cdot 6} = 1 - \frac{11}{60} = 1 - 0,183 = 0,817$$

El **CR = 0,817 < 0,90** fet pel qual no queda un escalograma perfecte.

5) Per a millorar l'escalograma podem eliminar l'ítem que més errors tinga. Per això, **comptem per columna el nombre d'errors que aporta cada ítem.**

Subjectes	Ítem 1	Ítem 2	Ítem 3	Ítem 6	Ítem 5	Ítem 4	Punt.	Errors
1	0*	0	1*	1*	0	0	2	3
2	1	0*	0	0	1*	0	2	2
3	1	1	1	0	0	0	3	0
4	1	1	0	0	0	0	2	0
5	1	1	0*	1	1	1*	5	2
6	1	0	0	0	0	0	1	0
7	1	1	1	1	1	0	5	0
8	0*	1	1	1*	0	0	3	2
9	1	0	0	0	0	0	1	0
10	1	0*	1*	0	0	0	2	2
Σcerts	8	5	5	4	3	1		
Σerrors	2	2	3	2	1	1		11

Això ens porta a **eliminar la columna de l'ítem 3**, el que més errors aporta.

6) En eliminar una columna, cal **tornar a calcular les puntuacions totals i el nombre d'errors:**

Subjectes	Ítem 1	Ítem 2	Ítem 6	Ítem 5	Ítem 4	Punt.	Errors
1	0*	0	1*	0	0	1	2
2	1	0*	0	1*	0	2	2
3	1	1	0	0	0	2	0
4	1	1	0	0	0	2	0
5	1	1	1	1	1	5	0
6	1	0	0	0	0	1	0

7	1	1	1	1	0	4	0
8	0*	1	1*	0	0	2	2
9	1	0	0	0	0	1	0
10	1	0	0	0	0	1	0
Σencerts	8	5	4	3	1		
Σerrors	2	1	2	1	0		6

7) **Tornem a calcular** el coeficient de reproductibilitat (CR):

$$CR = 1 - \frac{6}{10 \cdot 5} = 1 - \frac{6}{50} = 1 - 0,12 = 0,88$$

El **CR = 0,88 < 0,90** per la qual cosa millora, però encara no satisfà el criteri.

8) Es pot **calcular el coeficient de reproductibilitat de l'ítem (CR_i)**

IMPORTANT!!!

El CR_i **no pot ser inferior** a la proporció màxima de respostes possibles per l'ítem.

$$CR_i = 1 - \frac{\text{Nre. errors de l'ítem}}{\text{Nre. errors possibles}}$$

Per a obtenir la proporció màxima de respostes possibles per l'ítem cal procedir de la manera següent:

- a) Calcular la proporció d'encerts (p, proporció d'1) de cada ítem.
- b) Calcular la proporció de la proporció d'errors (q, proporció de 0) de cada ítem.
- c) Una vegada calculades les proporcions, marquem la major de les dues, és a dir, la major de p o de q.
- d) Comparem el CR de l'ítem amb la proporció major que hem marcat anteriorment.

Subjectes	Ítem 1	Ítem 2	Ítem 6	Ítem 5	Ítem 4	Punt.	Errors
1	0*	0	1*	0	0	1	2
2	1	0*	0	1*	0	2	2
3	1	1	0	0	0	2	0
4	1	1	0	0	0	2	0
5	1	1	1	1	1	5	0
6	1	0	0	0	0	1	0
7	1	1	1	1	0	4	0
8	0*	1	1*	0	0	2	2
9	1	0	0	0	0	1	0
10	1	0	0	0	0	1	0
Σencerts	8	5	4	3	1		
Σerrors	2	1	2	1	0		6
p	8/10 = 0.80	5/10 = 0.50	4/10 = 0.40	3/10 = 0.30	1/10 = 0.10		
q	2/10 = 0.20	5/10 = 0.50	6/10 = 0.60	7/10 = 0.70	9/10 = 0.90		
CR_i	1-2/10 = 0.80	1-1/10 = 0.90	1-2/10 = 0.80	1-1/10 = 0.90	1-0/10 = 1		

En aquest cas, **tots els ítems superen o igualen el criteri**. S'observa que el CR dels ítems és sempre igual o superior a la proporció màxima de respostes possibles per l'ítem:

Ítem 1: $0.80 = 0.80$

Ítem 2: $0.90 > 0.50$

Ítem 3: $0.80 > 0.60$

Ítem 4: $0.90 > 0.70$

Ítem 5: $1 > 0.90$

- 9) Es pot **calcular el coeficient de reproductibilitat dels subjectes (CR_s)**. Aquest criteri serveix per a detectar subjectes amb un patró de resposta molt inusual, fet que ens pot portar a qüestionar-ne la representativitat:

$$CR_s = 1 - \frac{\text{Nre. errors del subjecte}}{\text{Nre. errors possibles}}$$

Subjectes	Item1	Item2	Item6	Item5	Item4	Punt.	Erroros	CR _s
1	0*	0	1*	0	0	1	2	1-2/6 = 1-0.333 = 0.667
2	1	0*	0	1*	0	2	2	0.667
3	1	1	0	0	0	2	0	1-6/6 = 1-0 = 1
4	1	1	0	0	0	2	0	1
5	1	1	1	1	1	5	0	1
6	1	0	0	0	0	1	0	1
7	1	1	1	1	0	4	0	1
8	0*	1	1*	0	0	2	2	0.667
9	1	0	0	0	0	1	0	1
10	1	0	0	0	0	1	0	1

No s'observen subjectes amb un patró d'errors massa gran, com per a qüestionar-los.

- 10) Finalment, es pot calcular el **coeficient d'escalabilitat (CS)**, el qual **ha de ser més gran que 0,60**. Encara que si és molt elevat, pot indicar que el CR s'ha unflat artificialment en usar els elements extrems.

$$CS = \frac{PM}{1 - RMM} = \frac{CR - RMM}{1 - RMM}$$

CR: coeficient de reproductibilitat (en l'apartat 7 es va fer el càlcul: 0,88)

RMM: reproductibilitat marginal mínima (mitjana de les proporcions màximes)

Subjectes	Ítem 1	Ítem 2	Ítem 6	Ítem 5	Ítem 4	Punt.	Errors
1	0*	0	1*	0	0	1	2
2	1	0*	0	1*	0	2	2
3	1	1	0	0	0	2	0
4	1	1	0	0	0	2	0
5	1	1	1	1	1	5	0
6	1	0	0	0	0	1	0
7	1	1	1	1	0	4	0
8	0*	1	1*	0	0	2	2
9	1	0	0	0	0	1	0
10	1	0	0	0	0	1	0
Σencerts	8	5	4	3	1		
Σerrors	2	1	2	1	0		6
p	8/10 = 0.80	5/10 = 0.50	4/10 = 0.40	3/10 = 0.30	1/10 = 0.10		
q	2/10 = 0.20	5/10 = 0.50	6/10 = 0.60	7/10 = 0.70	9/10 = 0.90		
Prop. màx.	0.80	0.50	0.60	0.70	0.90		

Calculem RMM:

$$RMM = (0,80 + 0,50 + 0,60 + 0,70 + 0,90) / 5 = 0,70$$

Calculem CS (CR era 0,88, vegeu l'apartat 7):

$$CS = \frac{CR - RMM}{1 - RMM} = \frac{0,88 - 0,70}{1 - 0,70} = \frac{0,18}{0,30} = 0,60$$

S'observa que el **coeficient d'escalabilitat (CS) compleix el criteri**, ja que és igual al valor mínim de 0,60.

Com a **conclusió sobre aquesta escala** podríem dir que, encara que el CR no arribava el criteri per molt poc (CR = 0,88), el coeficient d'escalabilitat CS = 0,60) sí que el compleix. A més, quan hem calculat els CR dels ítems i dels subjectes, no hem pogut trobar cap ítem a eliminar, ni tampoc cap subjecte amb un patró de respostes massa extrem com perquè siga raonable eliminar-lo.

B1. Pràctica 3. LIKERT. ANÀLISI D'ÍTEMS DE TIPUS AFECTIU

L'anàlisi d'un test continua amb l'anàlisi dels elements, és a dir, dels ítems que el componen. Es pot fer l'anàlisi des del model clàssic (TCT) o des de la teoria de resposta als ítems (TRI), que veurem més avant. L'anàlisi d'ítems pot tenir dos objectius: seleccionar d'entre un gran conjunt d'ítems els que siguen els millors **per a construir un test**, o comprovar el funcionament correcte dels ítems que componen un **test ja construït**, bé perquè s'haja adaptat d'una altra llengua o cultura, bé perquè es prova en una altra població. D'altra banda, segons el constructe que mesura el test, podem parlar d'anàlisi d'ítems de tipus cognitiu (rendiment, aptituds, intel·ligència, coneixements) i d'**ítems de tipus afectiu** (personalitat, preferències, actituds).

En qualsevol cas, analitzar els ítems implica calcular una sèrie d'índexs o indicadors de qualitat que permetran comprovar si s'aconsegueixen els nivells exigits. En el cas d'ítems de tipus cognitiu, haurem d'estudiar, a més, la dificultat dels ítems.

En aquesta pràctica, treballarem amb el **Cuestionario de Ansiedad Laboral (CAL)**. El qüestionari està format per **vint-i-sis ítems** i es va elaborar per a avaluar l'ansietat en contextos laborals en mostra espanyola.¹ Els ítems avaluen tres dimensions de l'ansietat o factors: ansietat cognitiva, ansietat fisiològica i ansietat motora. S'utilitza una escala de resposta tipus Likert de set punts que va des de «1. Nunca» fins a «7. Siempre». El factor d'ansietat cognitiva el componen dotze ítems, l'ansietat fisiològica huit ítems, i l'ansietat motora, sis ítems. En la taula següent es presenten tots els ítems numerats. Cada número va seguit de la inicial de la dimensió que avalua, per identificar-lo més fàcilment. També els distingim amb colors. Cal destacar que en la dimensió cognitiva l'ítem 6 (Cal6c) és un ítem INVERTIT. Fixeu-vos en la redacció de l'enunciat:

1. Nunca	2. Rara vez
3. Algunas veces	4. La mitad de las veces
5. Bastantes veces	6. Casi siempre
7. Siempre	

Cal1c	Me preocupo fácilmente por las cosas relacionadas con mi trabajo.
Cal2m	Mientras trabajo realizo movimientos repetitivos tales como rascarme, golpear con los dedos alguna parte de mi cuerpo o un objeto, mover las piernas rítmicamente, etc.
Cal3c	Pienso que desempeño mi trabajo peor que mis compañeros.

¹ El qüestionari original es va desenvolupar i validar en castellà, per això els ítems i l'escala de resposta estan en castellà.

Cal4f	Cuando estoy trabajando siento que tengo temblores.
Cal5m	Mientras trabajo mis movimientos son menos precisos de lo habitual.
Cal6c	Cuando desempeño mi trabajo no me siento seguro de que lo hago correctamente.
Cal7f	Cuando estoy trabajando tengo mareos.
Cal8c	Tengo problemas para concentrarme en las tareas de mi trabajo.
Cal9m	Cuando desempeño mi trabajo tengo problemas para expresarme.
Cal10c	Me preocupa que mis compañeros se den cuenta de mis dificultades en el trabajo.
Cal11f	Me sudan las manos mientras trabajo.
Cal12c	Me preocupa que algo negativo que afecte a mi trabajo pueda ocurrir.
Cal13f	Mientras trabajo siento que mis músculos están tensos.
Cal14c	Me preocupa no ser aceptado por mis compañeros de trabajo.
Cal15f	Mi trabajo me fatiga.
Cal16c	Pensar en la jornada de trabajo me produce desasosiego.
Cal17c	Doy demasiadas vueltas a las cosas relacionadas con mi trabajo.
Cal18c	Siento temor de que pueda fallar en mi trabajo.
Cal19c	Cuando acabo de trabajar me siguen preocupando los problemas de trabajo.
Cal20c	Me cuesta mucho tomar una decisión relacionada con mi trabajo.
Cal21m	A causa de mi trabajo fumo, como o bebo demasiado.
Cal22f	Mi trabajo me produce molestias estomacales.
Cal23f	Mi trabajo me produce dolores de cabeza.
Cal24f	Durante la jornada de trabajo mi corazón sufre alteraciones de ritmo.
Cal25m	Evito tener contactos con mis compañeros de trabajo.
Cal26m	Durante la jornada de trabajo trato de hacer muchas cosas a la vez.

Ara farem l'exercici amb la dimensió o factor **ansietat cognitiva**, amb dades en una mostra de treballadors espanyols de diferents sectors. La dimensió d'ansietat cognitiva fa referència als pensaments i sentiments de preocupació, por, inseguretats, etc., que tenen les persones sobre

qüestions relacionades amb el seu treball. Aquests pensaments i sentiments poden portar l'individu a un estat d'alerta, desassossec o tensió, i generar falta de concentració, de decisió, i la idea que les persones de l'entorn laboral observen la seua conducta i s'adonen dels seus problemes i malapteses.

Ja que el constructe que s'avalua NO és una aptitud, sinó que s'avalua el grau d'ansietat percebut pels subjectes, els ítems són de tipus afectiu. No hi ha respostes correctes ni respostes incorrectes. El format de resposta és tipus Likert i, per tant, el nivell de mesura és ordinal (encara que en la pràctica es puga considerar d'interval), de manera que es poden ordenar les set respostes possibles a cada ítem en funció del que mesura el test, de manera que a major puntuació major quantitat del constructe avaluat, en aquest cas, major grau d'ansietat cognitiva. Al final, es pot obtenir la puntuació total de la persona en el test en sumar el valor numèric assignat a les opcions de resposta.

Utilitzarem el programa SPSS. Seguirem l'enfocament clàssic per a fer l'anàlisi dels dotze ítems. Segons aquest enfocament, els índexs que ens interessin són: discriminació i impacte dels ítems sobre la fiabilitat del test.

Accedeix a l'arxiu de dades [cal_alumnes.sav](#)

PAS 1. Abans de començar amb l'anàlisi dels ítems cal **INVERTIR** els ítems redactats en sentit invers. És un PAS NECESSARI previ, ja que, en cas contrari, podem obtenir resultats contradictoris i estranys, inclosa la possibilitat d'obtenir: un valor negatiu en la consistència interna de l'escala!

En aquesta escala, només cal invertir l'ítem 6. Recordeu que per a invertir un nou ítem cal restar el valor màxim en l'escala de resposta (7) més 1 a tots els valors. En aquest cas:

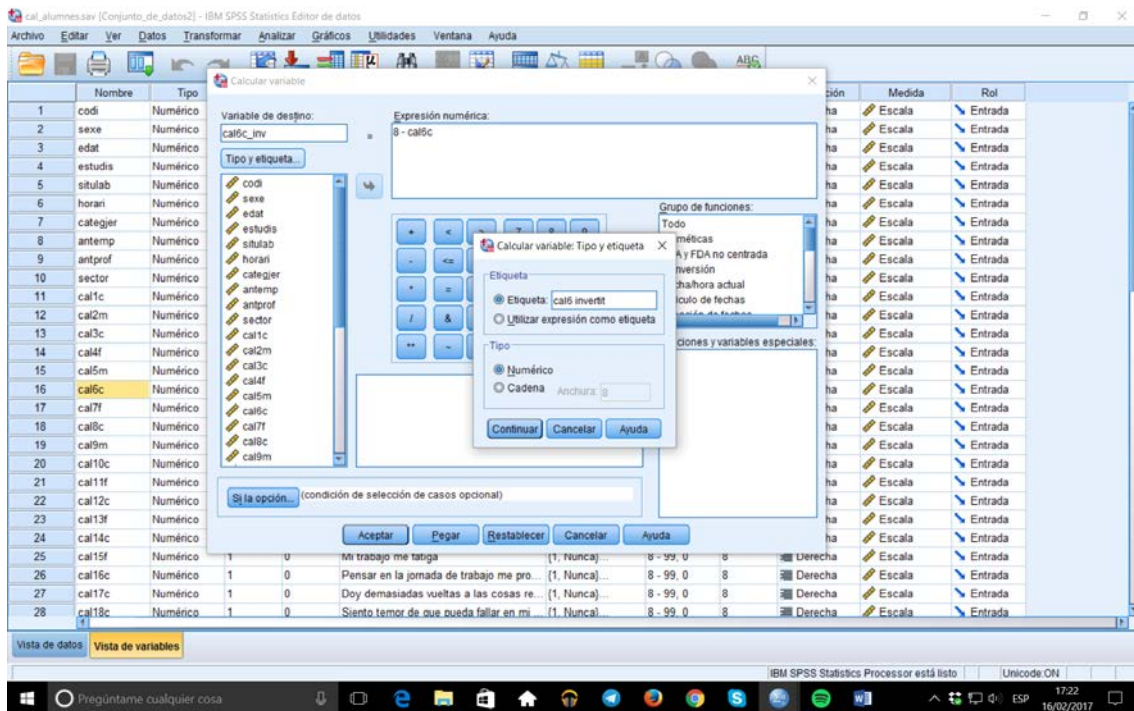
$$\text{cal6c_inv} = 8 - \text{cal6c}$$

Per a fer-ho amb SPSS hem de CALCULAR una puntuació de nou basada en l'original. Seguirem els passos següents:

Menú Transformar

Calcular variable

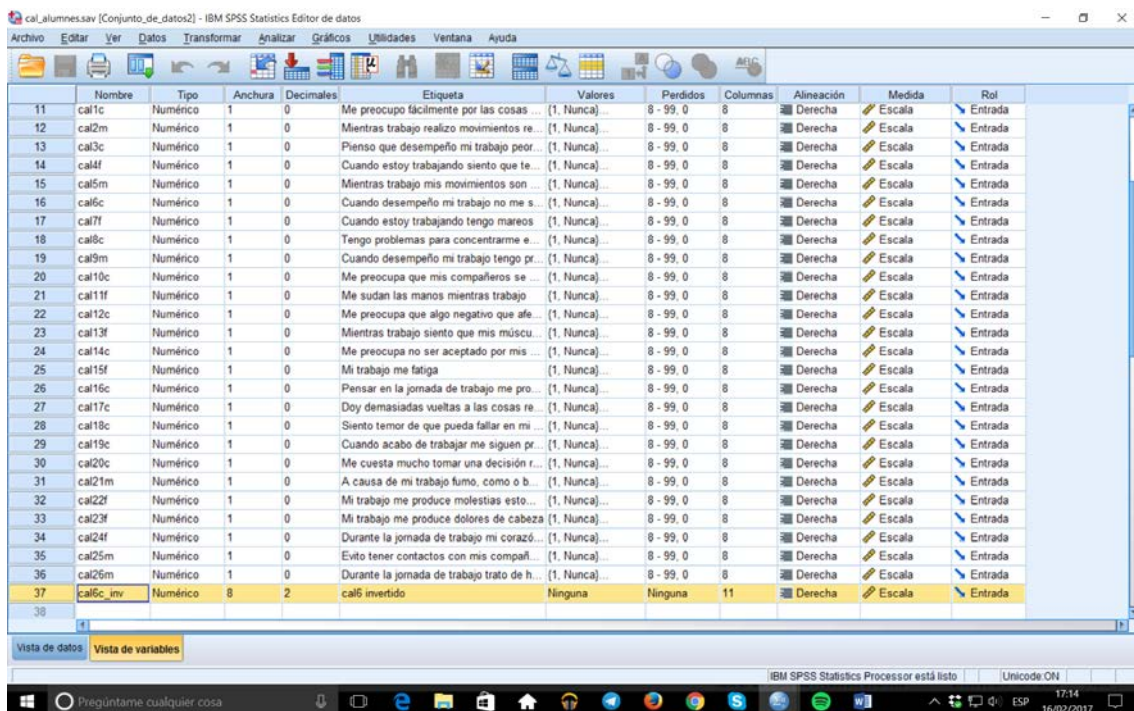
Fent clic al botó que hi ha davall del nom de la nova variable (**Tipus i etiqueta**) podem escriure una etiqueta per a aquesta nova variable.



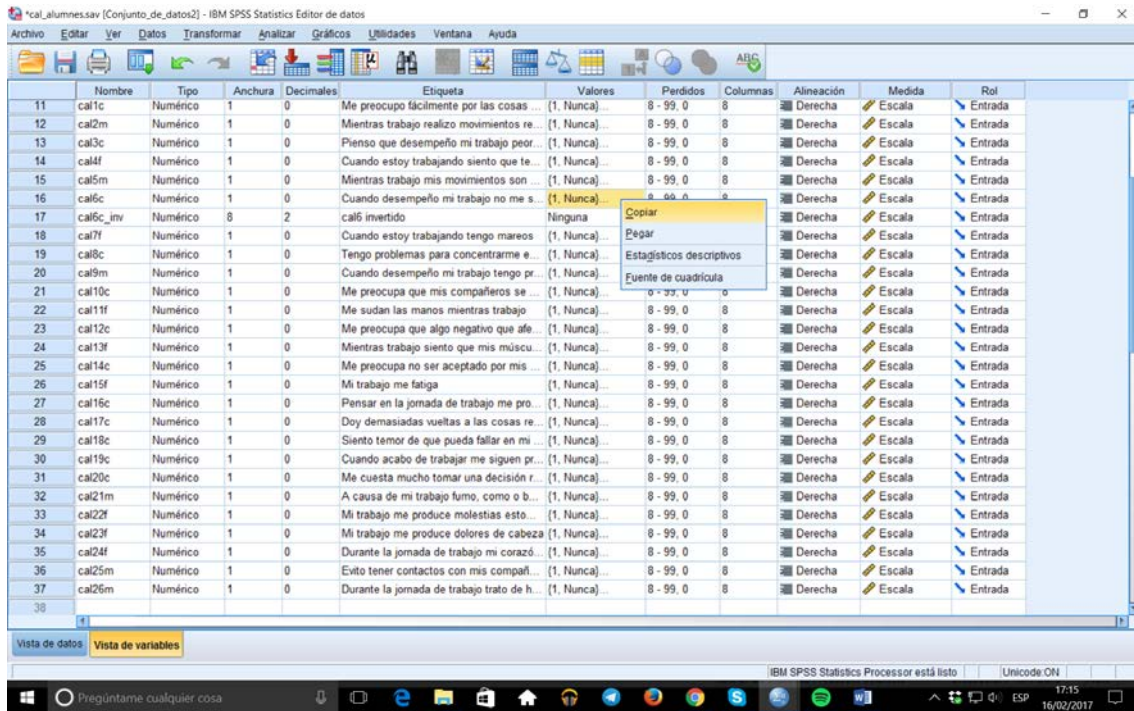
La nova variable apareixerà AL FINAL DE L'ARXIU DE DADES. Una vegada tenim TOTS ELS ÍTEMS DE L'ESCALA que analitzarem en el mateix sentit, passem a estudiar les característiques dels ítems.

Per tal de no oblidar que cal utilitzar ítem invertit, podem arrossegat-lo des de la llista de variables i posar-lo al costat de l'ítem original. Per a fer-ho, seguim els passos següents:

- a) Anem a la pestanya «Vista de variables», seleccionem el nou ítem (cal6c_inv) punxant a l'esquerra de l'ítem, damunt del nombre de fila (en aquest cas 37) i soltem el ratolí.



- b) Amb la variable seleccionada, punxem de nou amb el ratolí damunt del número de fila (37) i subjectant arrosseguem fins on la volem col·locar, en aquest cas, al costat de la variable original (cal6c).
- c) Ja al lloc, i per comoditat, podem llevar els dos decimals que SPSS posa per defecte i copiar les etiquetes dels valors d'un altre ítem, i enganxar-la a l'ítem cal6c_inv.



PAS 2. Ja amb l'arxiu preparat i amb tots els ítems en el mateix sentit, de manera que a major resposta a l'ítem tots expressen major nivell d'ansietat, comprovarem si es compleixen els supòsits que han de seguir els ítems que es responen amb una escala Likert.

Supòsit 1. Tots els ítems han de posar de manifest diferències en la manera de respondre dels subjectes.

Supòsit 2. Tots els ítems han de seguir una distribució aproximadament normal.

Aquests supòsits fan referència a les qüestions següents:

- a) Els ítems de l'escala posen de manifest diferències en la forma de respondre dels subjectes? Si hi haguera algun ítem al qual tots els subjectes respongueren igual, llavors aquest ítem no serviria per a posar de manifest diferències individuals entre els subjectes, que és just l'objectiu de l'escala, i caldria eliminar-lo.

Estadístic que resol la pregunta: la **desviació típica de cada ítem**, que és una mesura de la dispersió de les respostes dels subjectes. COM MÉS VARIABILITAT O DISPERSIÓ en les respostes, millor.

- b) Els ítems de l'escala han de seguir una distribució normal, segons Likert. Perquè si no és així, tindríem els elements massa esbiaixats, és a dir, amb puntuacions molt extremes. Això indicaria que molts dels subjectes han respost amb puntuacions altes en ansietat o molts amb puntuacions baixes, de manera que no podríem distingir entre persones amb nivells ALTS en ansietat i persones amb nivells BAIXOS en ansietat.

Estadístic que resol la pregunta: també la **desviació típica de cada ítem** i altres mesures de dispersió, ja que si tots els subjectes tenen puntuacions extremes tindrem poca variabilitat. Per a tenir una idea de la forma de la distribució demanem la **curtosi** i la **asimetria**, i podem veure els gràfics amb els histogrames. La curtosi i la asimetria són indicadors, ja que com més proper a 0 la asimetria, menys asimetria, i com més proper a 0 el valor de la curtosi menor apuntament. Valors d'asimetria i curtosi iguals a zero indiquen corba normal.

Per a comprovar si **estadísticament** podem considerar que la distribució de les respostes de cada ítem és normal, demanarem la prova de Kolmogorov-Smirnov.

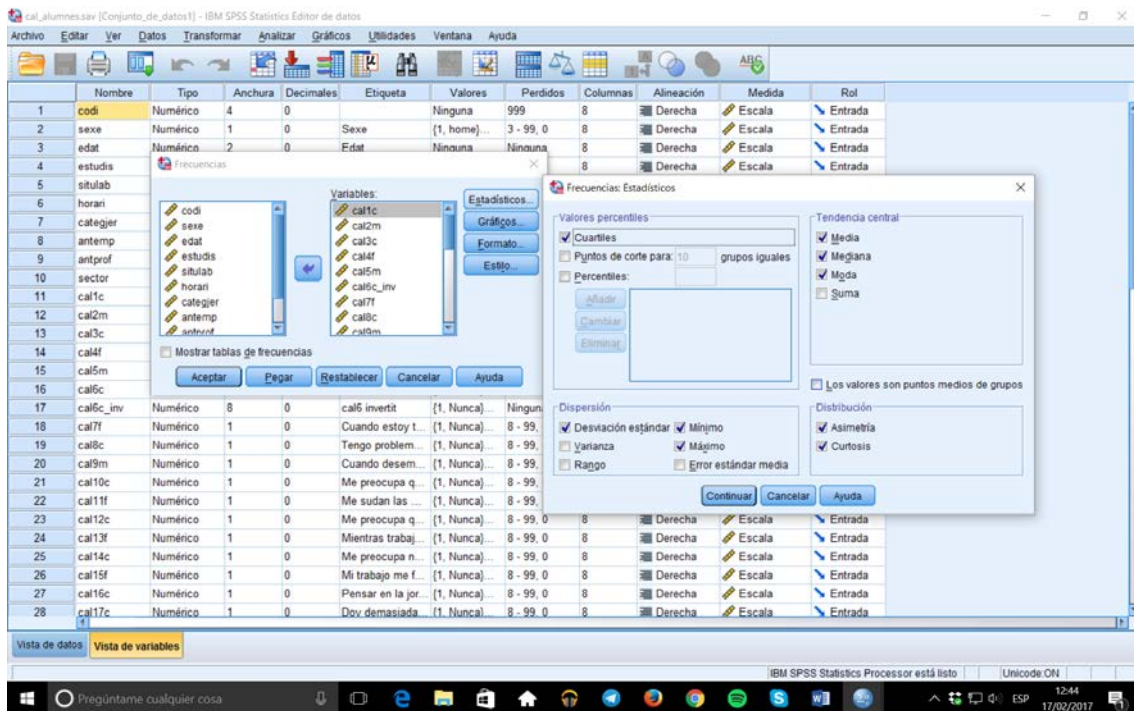
Seguim els passos següents en SPSS i primer demanem tots els estadístics descriptius, per fer una primera ullada a la distribució de les respostes:

Menú Analitzar

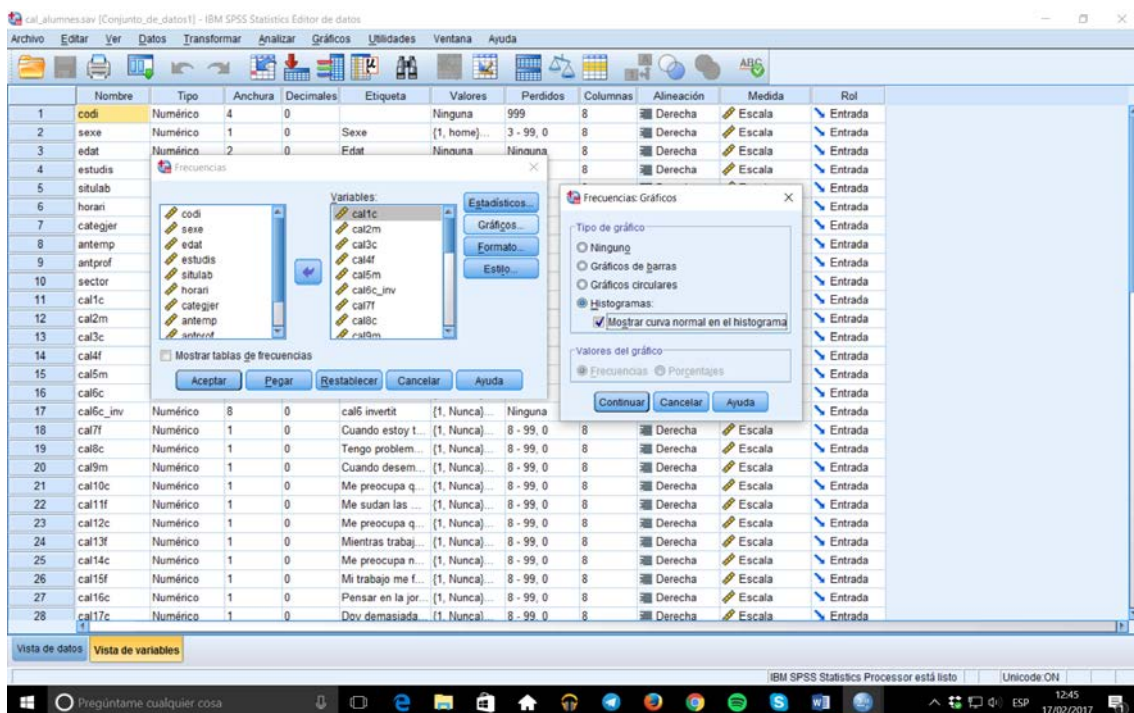
Estadístics descriptius

Freqüències

- SELECCIONEM ELS ÍTEMS de l'escala cognitiva i els passem a la dreta.
- DESMARQUEM L'OPCIÓ de mostrar taules de freqüències, a la part inferior esquerra, davall de la llista de variables (no necessitem les taules de freqüències ara mateix).
- Punxem en el botó d'opcions i seleccionem els descriptius que ens permeten observar la distribució de respostes dels ítems: quartils, moda, mitjana i mediana, desviació típica, mínim, màxim, curtosi i asimetria:



Acceptem i a continuació cliquem en el botó de gràfics i demanem histogrames amb corba normal.



Aquests són els resultats que ofereix la taula d'estadístics:

Estadístics

		cal1c	cal3c	cal6c inv	cal8c	cal10c	cal12c	cal14c	cal16c	cal17c	cal18c	cal19c	cal20c
N	Válidos	500	499	500	500	500	498	500	500	500	499	500	500
	Perdidos	0	1	0	0	0	2	0	0	0	1	0	0
Media		5,19	1,87	2,42	2,23	2,10	3,86	2,92	2,44	3,32	2,97	3,35	2,37
Mediana		6,00	2,00	2,00	2,00	2,00	3,00	2,00	2,00	3,00	3,00	3,00	2,00
Moda		7	1	2	2	1	3	1	2	3	3	3	2
Desv. típ.		1,668	1,007	1,322	,922	1,421	1,900	1,932	1,366	1,598	1,550	1,634	,891
Asimetria		-,687	1,558	1,671	,974	1,604	,377	,898	1,258	,565	1,116	,634	,767
Error típ. de asimetria		,109	,109	,109	,109	,109	,109	,109	,109	,109	,109	,109	,109
Curtosis		-,535	3,620	2,808	2,601	2,290	-1,085	-,376	1,561	-,521	,846	-,420	2,215
Error típ. de curtosis		,218	,218	,218	,218	,218	,218	,218	,218	,218	,218	,218	,218
Mínimo		1	1	1	1	1	1	1	1	1	1	1	1
Máximo		7	7	7	7	7	7	7	7	7	7	7	7
Percentils	25	4,00	1,00	2,00	2,00	1,00	2,00	1,00	1,00	2,00	2,00	2,00	2,00
	50	6,00	2,00	2,00	2,00	2,00	3,00	2,00	2,00	3,00	3,00	3,00	2,00
	75	7,00	2,00	3,00	3,00	3,00	5,00	4,00	3,00	5,00	3,00	5,00	3,00

Observem que tots els ítems tenen respostes que van des del mínim (1) fins al màxim (6) en l'escala de resposta.

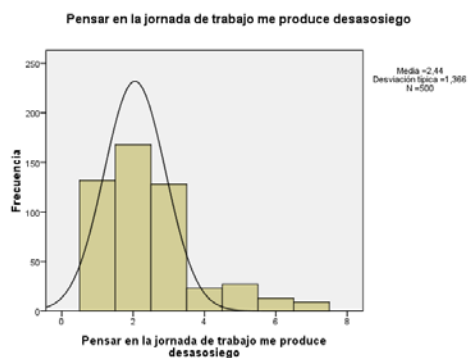
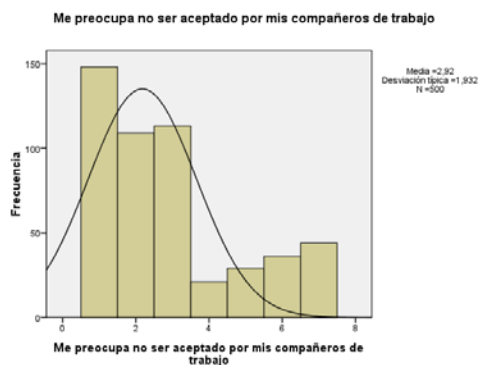
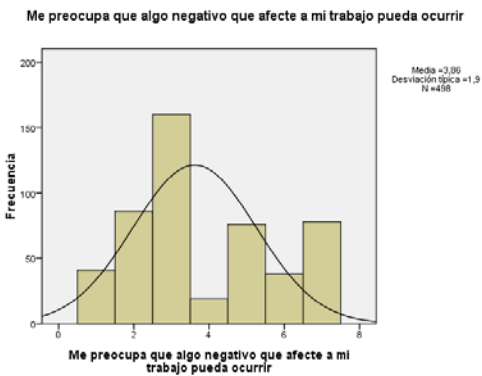
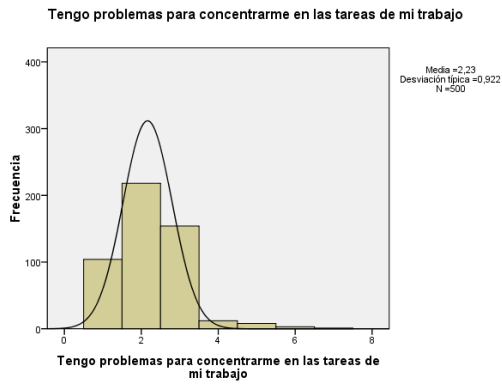
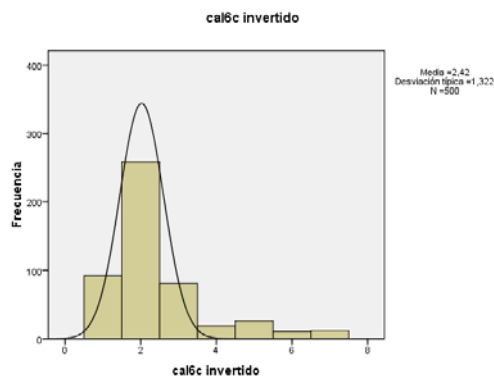
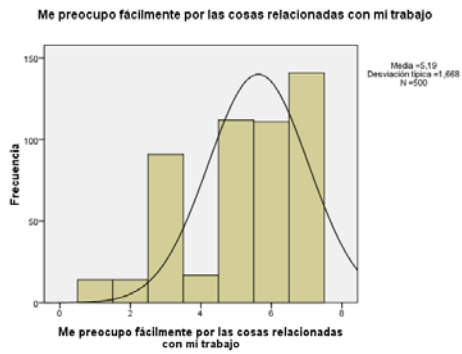
VARIABILITAT

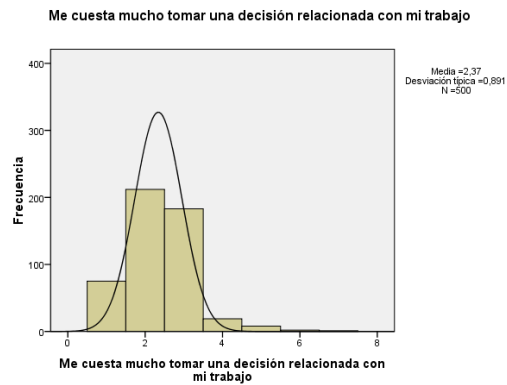
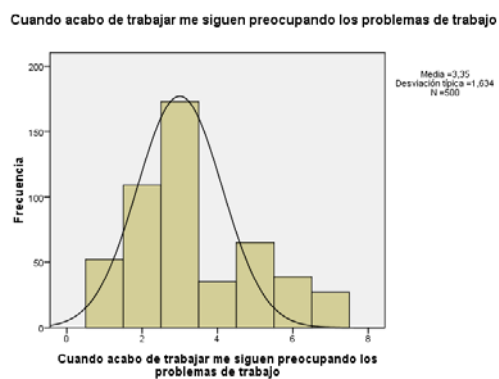
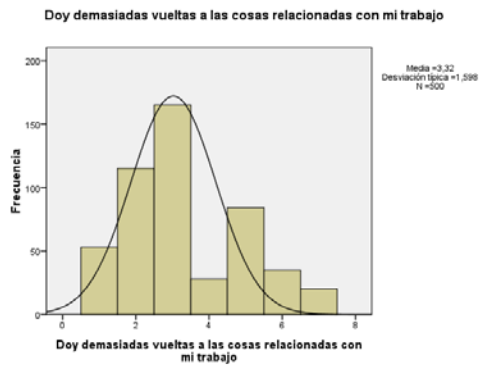
Els ítems amb **major** variabilitat són **el 12 i el 14**, mentre que els elements amb **menor** variabilitat són **el 3, el 8 i el 20**. És previsible que els ítems amb més variabilitat funcionen millor en l'escala.

FORMA DE LA DISTRIBUCIÓ

En groc s'han marcat els ítems amb major asimetria i curtosis. Els ítems **3, 6 i 10** presenten valors molt alts en tots dos i són els que més s'allunyen de la normalitat.

En els histogrames amb corba normal veiem de manera gràfica com es distribueixen les respostes en cada ítem:





Gràficament no s'observa molta normalitat, excepte potser un poquet en els últims cinc ítems.

Demaneu a SPSS les proves de Kolmogorov-Smirnov per tal de confirmar estadísticament si les respostes dels ítems es distribueixen normalment:

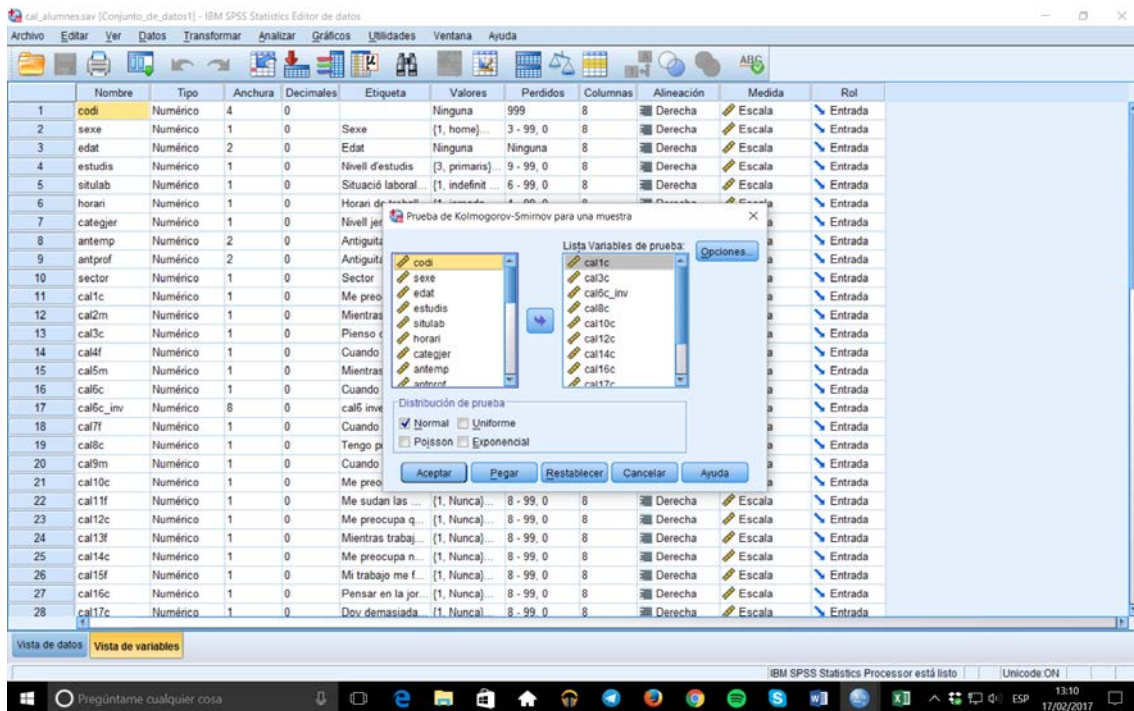
Menú Analitzar

Proves no paramètriques

Quadres de diàleg antics

K-S d'1 mostra

Seleccionem els ítems de l'escala cognitiva (inclòs el 6 invertit, NO el 6 original) i per defecte ens oferirà la prova de distribució normal, que ja ve marcada:



Aquests són els resultats de la prova per a tots els ítems:

Prueba de Kolmogorov-Smirnov para una muestra													
		cal1c	cal3c	cal6c_inv	cal8c	cal10c	cal12c	cal14c	cal16c	cal17c	cal18c	cal19c	cal20c
N		500	499	500	500	500	498	500	500	500	499	500	500
Parámetros normales a,b	Media	5.19	1.87	2.42	2.23	2.10	3.86	2.92	2.44	3.32	2.97	3.35	2.37
	Desviación estándar	1.67	1.01	1.32	0.92	1.42	1.90	1.93	1.37	1.60	1.55	1.63	0.89
Máximas diferencias extremas	Absoluta	0.19	0.24	0.33	0.24	0.24	0.25	0.22	0.23	0.25	0.29	0.25	0.23
	Positivo	0.14	0.24	0.33	0.24	0.24	0.25	0.22	0.23	0.25	0.29	0.25	0.23
	Negativo	-0.19	-0.19	-0.19	-0.19	-0.22	-0.11	-0.16	-0.15	-0.13	-0.13	-0.11	-0.19
Estadístico de prueba		0.19	0.24	0.33	0.24	0.24	0.25	0.22	0.23	0.25	0.29	0.25	0.23
Sig. asintótica (bilateral)		,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c	,00 ^c
a. La distribución de prueba es normal.													
b. Se calcula a partir de datos.													
c. Corrección de significación de Lilliefors.													

Totes les proves de normalitat indiquen que les respostes **NO es distribueixen normalment** en cap dels ítems.

Això pot ser degut al mateix constructe avaluat, ansietat, i a la grandària de la mostra. Perquè la distribució fóra normal hauria d'haver-hi molt poques respostes baixes i altes, i més respostes de nivells mitjans. Potser depenent del tipus de treball, hi ha més gent del normal que mostre nivells alts o baixos d'ansietat...

Per això, seguirem amb l'anàlisi dels ítems.

- c) **Seguim en el PAS 2.** Ja hem vist si es compleixen o no els supòsits 1 (els ítems han de posar de manifest diferències en la forma de respondre dels subjectes) i 2 (els ítems han de seguir una distribució aproximadament normal). Ho hem comprovat veient la variabilitat de les respostes i la forma de la distribució, en les respostes dels ítems de l'escala.

Ara seguim amb la qüestió següent: **tenim prou variabilitat en les variables sociodemogràfiques que hem recollit?** En aquest cas les variables són: sexe, edat, nivell d'estudis, situació laboral, horari, nivell jeràrquic, antiguitat a l'empresa, antiguitat en la professió i sector.

Per exemple, si gran part dels subjectes de la mostra foren homes no podríem inferir res sobre les dones. O si majoritàriament treballen en el sector públic, tampoc podríem estudiar als treballadors del sector privat.

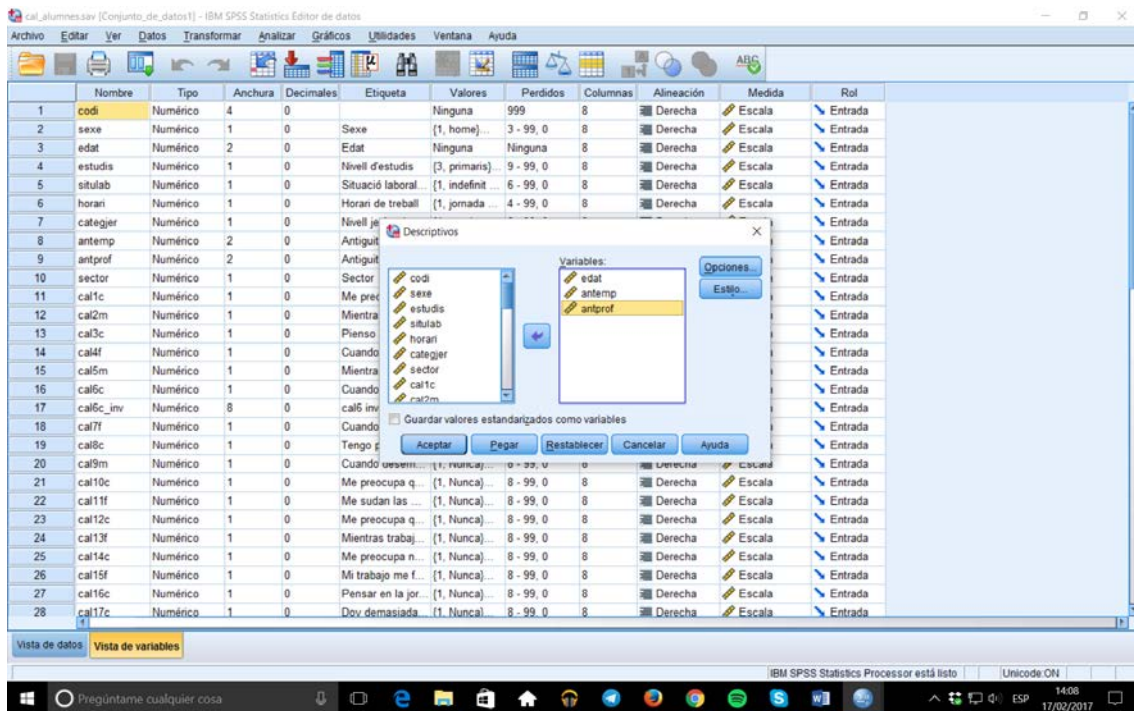
Estadístic que resol la pregunta: per a les variables quantitatives (edat, antiguitat a l'empresa i antiguitat en la professió) cal utilitzar la desviació típica. Per a les variables qualitatives, el percentatge de casos (sexe, nivell d'estudis, situació laboral, horari i nivell jeràrquic).

Primer treballarem amb els ítems demogràfics QUANTITATIUS i per anar més ràpid entrem per l'opció d'estadístics descriptius, que per defecte ja ofereix els descriptius que ens interessin sense taules de freqüències:

Menú Analitzar

Estadístics descriptius

Descriptius



Aquests són els resultats:

Estadístics descriptius

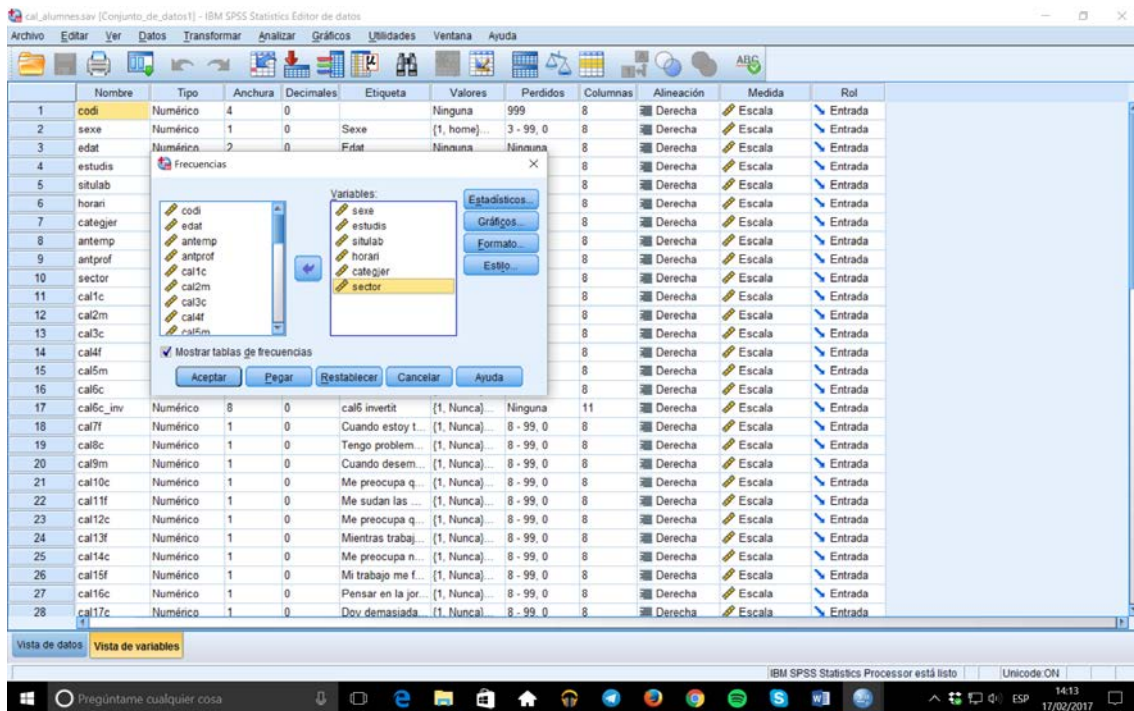
	N	Mínimo	Máximo	Media	Desv. típ.
edat	500	18	64	36,07	9,560
antemp	490	0	42	9,82	8,582
antprof	498	0	40	10,45	8,547
N vàlid (segon llista)	490				

Ara seleccionem els ítems demogràfics QUALITATIUS i demanem només la taula de freqüències (ve marcat per defecte):

Menú Analitzar

Estadístics descriptius

Freqüències



Els resultats de les anàlisis són els següents:

sexe

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos home	266	53,2	53,2	53,2
dona	234	46,8	46,8	100,0
Total	500	100,0	100,0	

situlab

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos indefinit o fixe	399	79,8	80,3	80,3
temporal o en pràctiques	98	19,6	19,7	100,0
Total	497	99,4	100,0	
Perduts Sistema	3	,6		
Total	500	100,0		

estudis

		Freqüència	Percentatge	Percentatge vàlid	Percentatge acumulat
Vàlids	primaris	82	16,4	16,5	16,5
	FP o CF	73	14,6	14,7	31,2
	BUP o batxiller	107	21,4	21,5	52,7
	titulació universitària mitjana	122	24,4	24,5	77,3
	llicenciats, tècnics sup. o doctors	113	22,6	22,7	100,0
	Total	497	99,4	100,0	
Perduts	Sistema	3	,6		
Total		500	100,0		

horari

		Freqüència	Percentatge	Percentatge vàlid	Percentatge acumulat
Vàlids	jornada partida fixa	194	38,8	39,0	39,0
	jornada intensiva fixa	191	38,2	38,4	77,3
	horari flexible i/o irregular	113	22,6	22,7	100,0
	Total	498	99,6	100,0	
Perduts	Sistema	2	,4		
Total		500	100,0		

sector

		Freqüència	Percentatge	Percentatge vàlid	Percentatge acumulat
Vàlids	Públic	238	47,6	51,0	51,0
	Privat	229	45,8	49,0	100,0
	Total	467	93,4	100,0	
Perduts	Sistema	33	6,6		
Total		500	100,0		

categjer

		Freqüència	Percentatge	Percentatge vàlid	Percentatge acumulat
Vàlids	empleat o treballador	261	52,2	52,2	52,2
	supervisor o capatàs	51	10,2	10,2	62,4
	comandament intermedi	92	18,4	18,4	80,8
	directiu	77	15,4	15,4	96,2
	alt directiu	19	3,8	3,8	100,0
	Total	500	100,0	100,0	

Veiem que totes les variables presenten molta variabilitat. L'edat inclou treballadors majors d'edat fins als 64 anys, amb gent que té molt poca antiguitat a l'empresa i en la professió (menys d'un any) fins a gent que té 40 anys o més d'experiència o antiguitat. També veiem que hi ha més o menys la mateixa quantitat de dones que d'homes, que hi ha una bona distribució pel que fa al nivell d'estudis, ja que hi ha treballadors de tots els nivells.

Pel que fa a la situació laboral hi ha molts més treballadors fixos o indefinits, però també n'hi ha de temporals. Els tres tipus d'horaris de treball es distribueixen de manera molt semblant entre tots els treballadors, igual que ocorre amb el sector de l'empresa. Pel que fa al nivell jeràrquic, hi ha treballadors en tots els nivells, si bé hi ha molts més treballadors de base i menys alts directius, la qual cosa és normal.

PAS 3. Avaluació dels ítems.

Supòsit 3. Tots els ítems han de mesurar la mateixa dimensió:
DISCRIMINACIÓ

- a) Els elements de l'escala es comporten com si mesuraren la mateixa cosa que l'escala en el seu conjunt? Els ítems mesuren el mateix? Quan tots els ítems d'una escala mesuren el mateix constructe correlacionen alt ($> 0,40$).

Estadístic que resol el problema: **matriu de correlacions**

- b) Les respostes dels subjectes a cada ítem són consistents amb la puntuació total? Quan un ítem mesura el mateix que l'escala en la qual s'inclou s'espera que les puntuacions dels subjectes en aquest ítem siguin consistents amb les puntuacions dels subjectes en l'escala total, és a dir, que els subjectes amb puntuacions altes en l'ítem tinguen puntuacions també altes en l'escala, i que els subjectes amb puntuacions baixes en l'ítem tinguen puntuacions baixes en l'escala.

Estadístic que resol el problema: **índex d'homogeneïtat corregit**. A més, amb aquesta anàlisi podem veure també l'IMPACTE DE CADA ÍTEM SOBRE LA FIABILITAT DEL TEST (tercer criteri, el comentarem després).

Anem a SPSS i demanem tots els estadístics al mateix temps, ja que aquesta opció permet obtenir la matriu de correlacions i l'índex d'homogeneïtat corregit:

Menú Analitzar

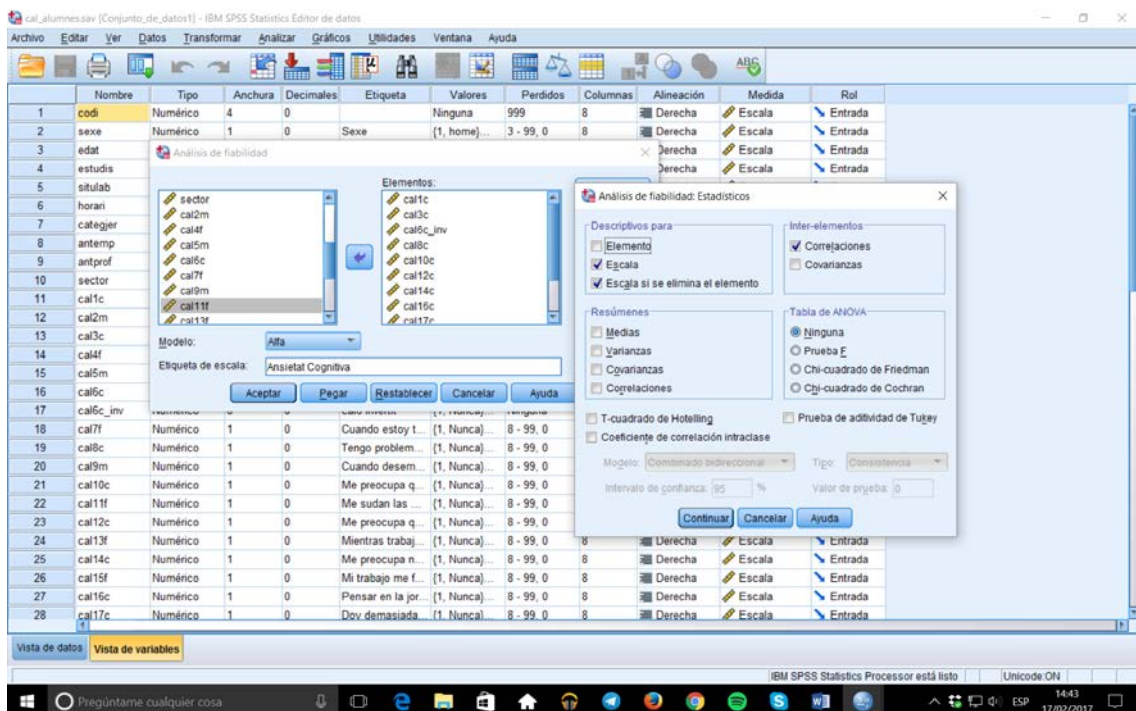
Escala

Anàlisi de fiabilitat

En el botó d'estadístics cal demanar també descriptius per a: «Escala» i «Escala si s'elimina element», i «Matriu de correlacions». També podem donar una etiqueta a l'escala i anomenar-la pel nom: ansietat cognitiva.

MOLT IMPORTANT!!!

CAL TRIAR L'ÍTEM 6 INVERTIT (cal6_inv) en lloc de l'ítem original O TINDREM RESULTATS ESTRANYS!



Aquests són els resultats:

Escala: ansietat cognitiva

Resum del processament dels casos

		N	%
Casos	Vàlids	496	99,2
	Exclusos ^a	4	,8
	Total	500	100,0

a. Eliminació per llista basada en totes les variables del procediment.

Estadístics de fiabilitat

Alfa de Cronbach	Alfa de Cronbach basada en els elements tipificats	Nre. d'elements
,750	,750	12

Matriu de correlacions entre elements

	cal1c	cal3c	cal6c_inv	cal8c	cal10c	cal12c	cal14c	cal16c	cal17c	cal18c	cal19c	cal20c
cal1c	1,000	,007	-,145	-,063	,054	,330	,222	-,023	,313	,182	,432	,097
cal3c	,007	1,000	,313	,332	,248	,021	,133	,140	,109	,246	,130	,217
cal6c_inv	-,145	,313	1,000	,234	,254	-,044	,049	,107	,095	,161	-,014	,122
cal8c	-,063	,332	,234	1,000	,206	-,025	,100	,238	,139	,175	,025	,314
cal10c	,054	,248	,254	,206	1,000	,267	,334	,184	,191	,358	,191	,291
cal12c	,330	,021	-,044	-,025	,267	1,000	,406	,110	,344	,391	,420	,172
cal14c	,222	,133	,049	,100	,334	,406	1,000	,121	,247	,401	,229	,179
cal16c	-,023	,140	,107	,238	,184	,110	,121	1,000	,266	,188	,152	,272
cal17c	,313	,109	,095	,139	,191	,344	,247	,266	1,000	,387	,537	,339
cal18c	,182	,246	,161	,175	,358	,391	,401	,188	,387	1,000	,325	,296
cal19c	,432	,130	-,014	,025	,191	,420	,229	,152	,537	,325	1,000	,191
cal20c	,097	,217	,122	,314	,291	,172	,179	,272	,339	,296	,191	1,000

No hi ha moltes correlacions superiors a ,40

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Índice de homogeneidad corregido (IH). Indica el GRADO DE DISCRIMINACIÓN de los ítems	Alfa de
cal1c			,287		
cal3c			,288	,220	,743
cal6c_inv			,147	,182	,758
cal8c			,239	,219	,747
cal10c			,440	,265	,727
cal12c			,471	,347	,722
cal14c			,448	,273	,726
cal16c			,275	,149	,745
cal17c	31,72	66,272	,543	,408	,712
cal18c	32,07	66,212	,567	,347	,709
cal19c	31,70	67,011	,502	,415	,718
cal20c	32,67	75,741	,413	,247	,735

L'IH és la correlació entre l'ítem i la puntuació total del test feta amb tots els ítems EXCEPTE aquest ítem. Es considera un valor **adequat** si és **SUPERIOR A 0,200**.
Hi ha autors que prenen punts de tall superiors a 0,300
 En aquest cas, tots els ítems superen el criteri mínim excepte l'ítem 6.
 Els millors són els ítems 17, 18 i 19.

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
cal1c	29,84	72,138			,447
cal3c	33,16	76,855			
cal6c_inv	32,63	77,938			
cal8c	32,80	78,207			
cal10c	32,93	70,418			
cal12c	31,17	64,958			
cal14c	32,11	65,203			
cal16c	32,60	74,657			
cal17c	31,72	66,272			
cal18c	32,07	66,212	,567	,347	,709
cal19c	31,70	67,011	,502	,415	,718
cal20c	32,67	75,741	,413	,247	,735

Si mirem la segona columna i comparem els valors amb els de la taula posterior (Estadísticos de l'escala), podem veure quin ítem APORTA MÉS VARIÀNCIA a l'escala, i per tant és més important.


L'ÍTEM 12 és el que MÉS variància aporta perquè SI S'ELIMINA es perd molta variància! Després hi ha els ítems 14, 17, 18 i 19.

L'ítem 8 és el que MENYS variància aporta, seguit pels ítems 6 i 3.

Estadísticos de la escala

Media	Varianza	Desviación típica	N de elementos
35,04	82,968	9,109	12

Estadístics total-elemento



	Media de la escala si se elimina el elemento			Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
cal1c	29,84	<p>La correlació múltiple al quadrat indica el grau en què es pot predir la puntuació en l'ítem a partir de la puntuació en els restants ítems. Els ítems 6 i 16 són els que presenten els valors més baixos.</p>		,260	,747
cal3c	33,16			,220	,743
cal6c_inv	32,63			,182	,758
cal8c	32,80			,219	,747
cal10c	32,93	70,418	,440	,265	,727
cal12c	31,17	64,958	,471	,347	,722
cal14c	32,11	<u>65,203</u>	,448	,273	,726
cal16c	32,60	74,657	,275	,149	,745
cal17c	31,72	<u>66,272</u>	,543	,408	,712
cal18c	32,07	<u>66,212</u>	,567	,347	,709
cal19c	31,70	<u>67,011</u>	,502	,415	,718
cal20c	32,67	75,741	,413	,247	,735

PAS 4. Tercer criteri de qualitat: L'IMPACTE DE L'ÍTEM sobre la consistència interna del test

L'impacte de l'ítem sobre el test s'analitza comparant la consistència interna del test complet (Alfa = 0,750, vegeu la taula «Estadístiques de fiabilitat» més avant) amb la qual s'obtidria després d'eliminar cada un dels ítems. Quan la consistència interna del test (Alfa) puja després d'eliminar un ítem és perquè l'ítem empitjora el test i per tant és inadequat. En aquest cas, si observem la taula a continuació, veiem que **eliminant l'ítem 6 (cal6_inv) Alfa puja** a 0,758. Amb la resta dels ítems no passa el mateix, ja que en eliminar-ne qualsevol la consistència interna (Alfa) disminueix.

Estadísticos de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N elementos
,750	,750	12

Estadísticas de total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
cal1c	29,84	72,138		,60	,747
cal3c	33,16	76,855			,743
cal6c_inv	32,63	77,938			,758
cal8c	32,80	78,207			,747
cal10c	32,93	70,418			,727
cal12c	31,17	64,958		,7	,722
cal14c	32,11	65,203	,448	,273	,726
cal16c	32,60	74,657	,275	,149	,745
cal17c	31,72	66,272	,543	,408	,712
cal18c	32,07	66,212	,567	,347	,709
cal19c	31,70	67,011	,502	,415	,718
cal20c	32,67	75,741	,413	,247	,735

Observant els valors de la darrera columna s'aprecia que només hi ha un ítem l'**eliminació del qual** pot millorar la consistència interna de l'escala, i és de nou l'**ítem 6**.

En **conclusió**, podem dir que tots els ítems són prou discriminadors, excepte l'ítem 6 (IH = 0,147). La correlació múltiple al quadrat és bona en tots els casos excepte per l'ítem 6, de nou, i per l'ítem 16 (en els dos casos per davall de 0,200). Finalment, l'eliminació de l'ítem 6 també pareix millorar la fiabilitat total de l'escala. Per tant, es recomanaria seguir prestant atenció a aquest ítem (cal6_inv) en altres mostres per observar si es segueix comportant igual. I si és així, se'n recomanaria l'eliminació de l'escala.

Ara, **per a practicar**, repetiu les anàlisis anteriors amb el mateix arxiu de dades, però segmentant per sexe. Primer seleccioneu els homes i feu les anàlisis anteriors només amb aquest grup. Després, seleccioneu les dones i torneu a analitzar l'escala.

Per a seleccionar un grup de subjectes (per exemple, els homes, als quals hem posat etiqueta numèrica igual a 1) cal fer el pas següent:

Menú Dades

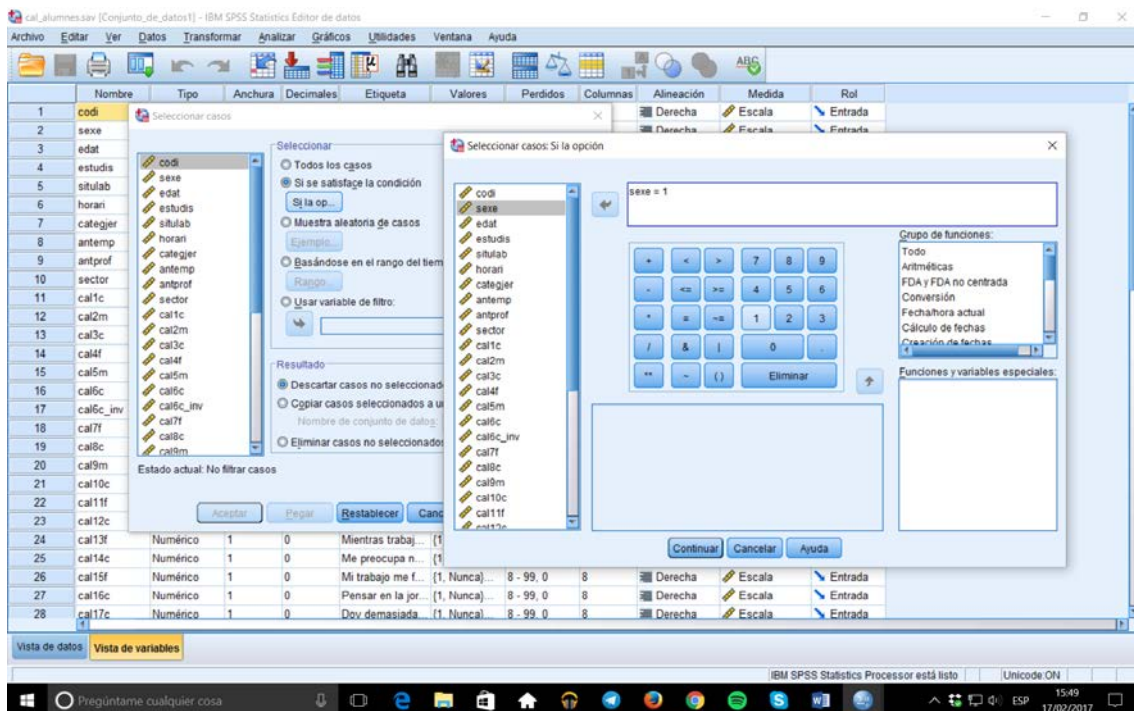
Seleccionar casos

Seleccioneu «Si se satisfà la condició»

Cliqueu en el botó «Si l'opció»

Introduïu la condició seleccionant la variable Sexe: sexe = 1

(Quan es tracte de les dones, la condició serà sexe = 2)



B1 Pràctica 4. ANÀLISI D'ITEMS DE TIPUS COGNITIU

L'anàlisi d'un test continua amb l'anàlisi dels elements, és a dir, dels ítems que el componen. Es pot fer l'anàlisi des del model clàssic (TCT) o des de la teoria de resposta als ítems (TRI), que veurem més avant.

L'anàlisi d'ítems pot tenir dos objectius: seleccionar d'entre un gran conjunt d'ítems els que siguin els millors **per a construir un test**, o comprovar el funcionament correcte dels ítems que componen un **test ja construït**, bé perquè s'haja adaptat d'una altra llengua o cultura, bé perquè es prova en una altra població.

D'altra banda, segons el constructe que mesura el test podem parlar d'anàlisi d'**ítems de tipus cognitiu** (rendiment, aptituds, intel·ligència, coneixements) i d'ítems de tipus afectiu (personalitat, preferències, actituds).

En qualsevol cas, analitzar els ítems implica calcular una sèrie d'índexs o indicadors de qualitat que permetran comprovar si s'aconsegueixen els nivells exigits. En el cas d'ítems de tipus cognitiu haurem d'estudiar, a més, la **dificultat** dels ítems.

En aquesta pràctica, analitzarem els ítems del **test de aptituds diferenciales DAT-5**, que és una bateria de huit tests que avaluen diferents aptituds. Va ser dissenyada per a mesurar la capacitat dels estudiants d'aprendre o actuar eficaçment en un cert nombre d'àrees, així com per a avaluar el potencial d'un estudiant o candidat a un lloc de treball. Presenta l'avantatge de poder ser aplicat en forma total o parcial. Ajuda a avaluar set aptituds bàsiques: raonament verbal, raonament numèric, raonament abstracte, raonament mecànic, relacions espacials, ortografia, rapidesa i exactitud perceptiva.

Nosaltres estudiarem únicament el **raonament verbal del DAT-5** amb dades en una mostra d'adults. El test està format per **quaranta ítems** amb preguntes plantejades com analogies, i mesura la capacitat d'entendre conceptes formulats amb paraules. Utilitzarem el programa SPSS, com fins ara.

Seguirem l'enfocament clàssic per a fer l'anàlisi dels quaranta ítems. Segons aquest enfocament, els índexs que ens interessin són: dificultat, discriminació i impacte dels ítems sobre la fiabilitat del test.

Ja que el constructe que s'avalua és una aptitud, els ítems són de tipus cognitiu. El format de resposta és dicotòmic (1,0) amb un nivell de mesura ordinal, de manera que un encert (1) indica més aptitud que una fallada (0), per la qual cosa es poden ordenar les dues respostes possibles en funció del que mesura el test. Al final, es pot obtenir la puntuació total de la persona en el test sumant el nombre d'encerts.

Accedeix a l'arxiu de dades [DAT5_RV.sav](#)

PAS 1. Primer criteri de qualitat: la DIFICULTAT dels ítems

La dificultat d'un ítem es calcula com la proporció d'encerts en aquest ítem. En el cas d'**ítems dicotòmics**, com els del DAT-5, **la dificultat coincideix amb la mitjana de l'ítem**.

Un test d'aptitud ha de tenir ítems de diferents nivells de dificultat, ordenats de manera creixent. És recomanable que el test s'iniciï amb ítems fàcils perquè les persones avaluades prenguen confiança i es motiven, però ha d'acabar amb alguns ítems molt difícils de manera que ningú o quasi ningú pugui contestar-los correctament. Un subjecte que encerta tot el test és un subjecte que no pot ser adequadament classificat pel test, perquè no podem saber en quin nivell de dificultat té el límit.

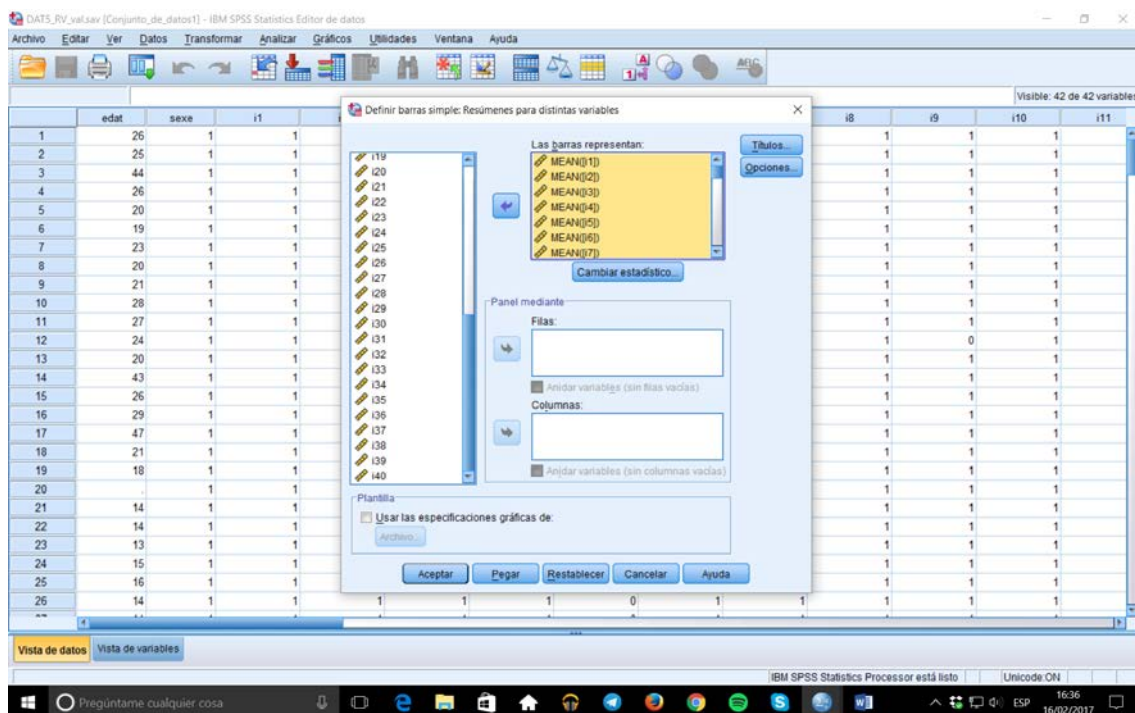
El gràfic de les mitjanes de cada ítem (índexs de dificultat) ens permet veure la distribució de la dificultat dels ítems al llarg del test, i per tant ens permet comprovar si la progressió de la dificultat és l'adequada.

Menú Gràfics

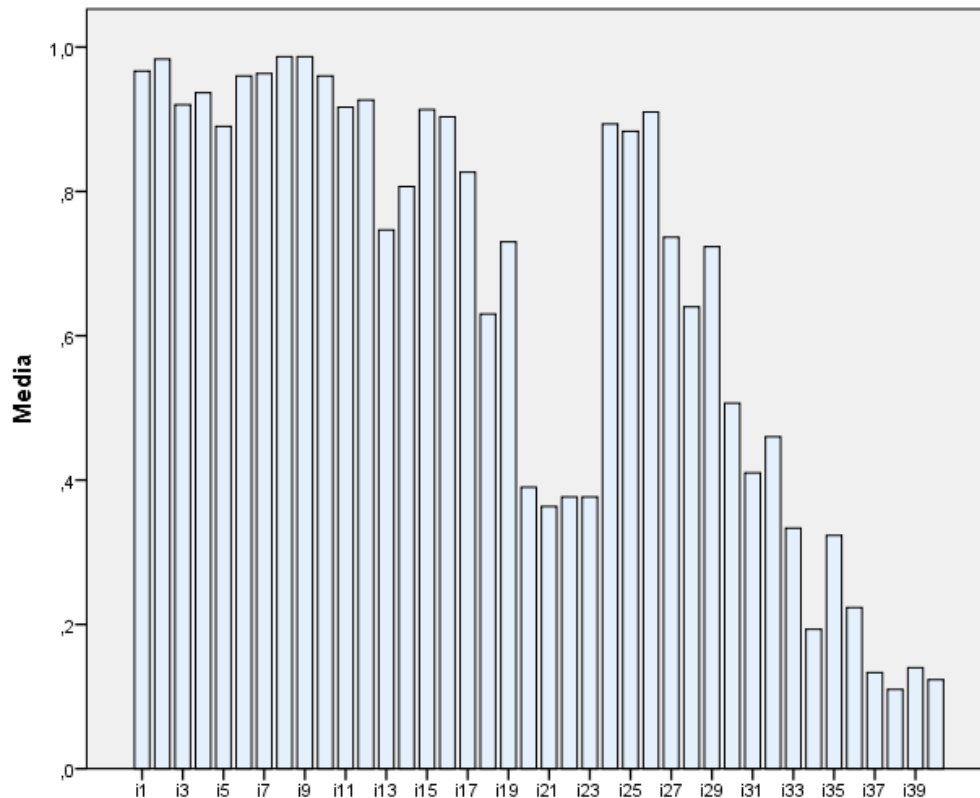
Quadres de diàleg antics

Barres (simple)

Resums per a distintes variables



En el gràfic es pot observar que la progressió de la dificultat dels ítems sembla adequada, encara que els ítems del 20 al 23 són massa difícils per al lloc que ocupen. També sembla que hi ha massa ítems molt fàcils (quasi fins a l'ítem 17) per a quaranta ítems que hi ha en total.



També podem veure la taula de valors de les mitjanes dels ítems, que ens indicaran numèricament el valor de la dificultat.

Menú Analitzar

Estadístics descriptius

Descriptius

The screenshot shows the IBM SPSS Statistics interface. The main window displays a data table with the following structure:

	edat	sexe	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11
1	26	1	1	1	1	1	1	1	1	1	1	1	1
2	25	1	1	1	1	1	0	1	1	1	1	1	1
3	44	1	1	1	1	1	1	1	1	1	1	1	1
4	26	1	1	1	1	1	1	1	0	1	1	1	1
5	20	1	1	1	1	1	1	1	1	1	1	1	1
6	19	1	1	1	1	1	1	1	1	1	1	1	1
7	23	1	1	1	1	1	1	1	1	1	1	1	1
8	20	1	1	1	1	1	1	1	1	1	1	1	1
9	21	1	1	1	1	1	1	1	1	1	1	1	1
10	28	1	1	1	1	1	1	1	1	1	1	1	1
11	27	1	1	1	1	1	1	1	1	1	1	1	1
12	24	1	1	1	1	1	1	1	1	1	0	1	1
13	20	1	1	1	1	1	1	1	1	1	1	1	1
14	43	1	1	1	1	1	1	1	1	1	1	1	1
15	26	1	1	1	1	1	1	1	1	1	1	1	1
16	29	1	1	1	1	1	1	1	1	1	1	1	1
17	47	1	1	1	1	1	1	1	1	1	1	1	1
18	21	1	1	1	1	1	1	1	1	1	1	1	1
19	18	1	1	1	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1	1	1	1
21	14	1	1	1	1	1	1	1	1	1	1	1	1
22	14	1	1	1	1	1	1	1	1	1	1	1	1
23	13	1	1	1	1	1	1	1	1	1	1	1	1
24	15	1	1	1	1	1	1	1	1	1	1	1	1
25	16	1	1	1	0	1	1	1	1	1	1	1	1
26	14	1	1	1	1	1	0	1	1	1	1	1	1

The 'Descriptives' dialog box is open, showing the following variables selected for analysis:

- edat
- sexe
- i1
- i2
- i3
- i4
- i5
- i6
- i7
- i8
- i9
- i10
- i11

The dialog box also includes options for 'Guardar valores estandarizados como variables' (Save standardized values as variables) and buttons for 'Aceptar', 'Pegar', 'Restablecer', 'Cancelar', and 'Ayuda'.

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desviación estándar
i1	300	0	1	,97	,180
i2	300	0	1	,98	,128
i3	300	0	1	,92	,272
i4	300	0	1	,94	,244
i5	300	0	1	,89	,313
i6	300	0	1	,96	,196
i7	300	0	1	,96	,188
i8	300	0	1	,99	,115
i9	300	0	1	,99	,115
i10	300	0	1	,96	,196
i11	300	0	1	,92	,277
i12	300	0	1	,93	,261
i13	300	0	1	,75	,436
i14	300	0	1	,81	,396
i15	300	0	1	,91	,282
i16	300	0	1	,90	,296
i17	300	0	1	,83	,379
i18	300	0	1	,63	,484
i19	300	0	1	,73	,445
i20	300	0	1	,39	,489
i21	300	0	1	,36	,482
i22	300	0	1	,38	,485
i23	300	0	1	,38	,485
i24	300	0	1	,89	,309

Es pot observar com els deu primers ítems presenten mitjanes molt altes.

RECORDA que l'índex de dificultat (ID) d'un ítem és la PROPORCIÓ

D'ENCERTS (en ítems cognitius és igual a la mitjana), de manera que **COM MÉS GRAN ÉS LA PROPORCIÓ**

D'ENCERTS (MITJANA) MENOR ÉS L'ID, és a dir, **MÉS FÀCIL ÉS L'ÍTEM!**

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

i25	300	0	1	,88	,322
i26	300	0	1	,91	,287
i27	300	0	1	,74	,441
i28	300	0	1	,64	,481
i29	300	0	1	,72	,448
i30	300	0	1	,51	,501
i31	300	0	1	,41	,493
i32	300	0	1	,46	,499
i33	300	0	1	,33	,472
i34	300	0	1	,19	,396
i35	300	0	1	,32	,469
i36	300	0	1	,22	,417
i37	300	0	1	,13	,341
i38	300	0	1	,11	,313
i39	300	0	1	,14	,348
i40	300	0	1	,12	,329
N vàlid (per llista)	300				

Ací es pot observar com els quatre últims ítems presenten mitjanes molt baixes.

COM MENOR ÉS LA PROPORCIÓ D'ENCERTS (MITJANA), MAJOR ÉS L'ID, és a dir, MÉS DIFÍCIL ÉS L'ÍTEM!

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

PAS 2. Segon criteri de qualitat: la DISCRIMINACIÓ dels ítems

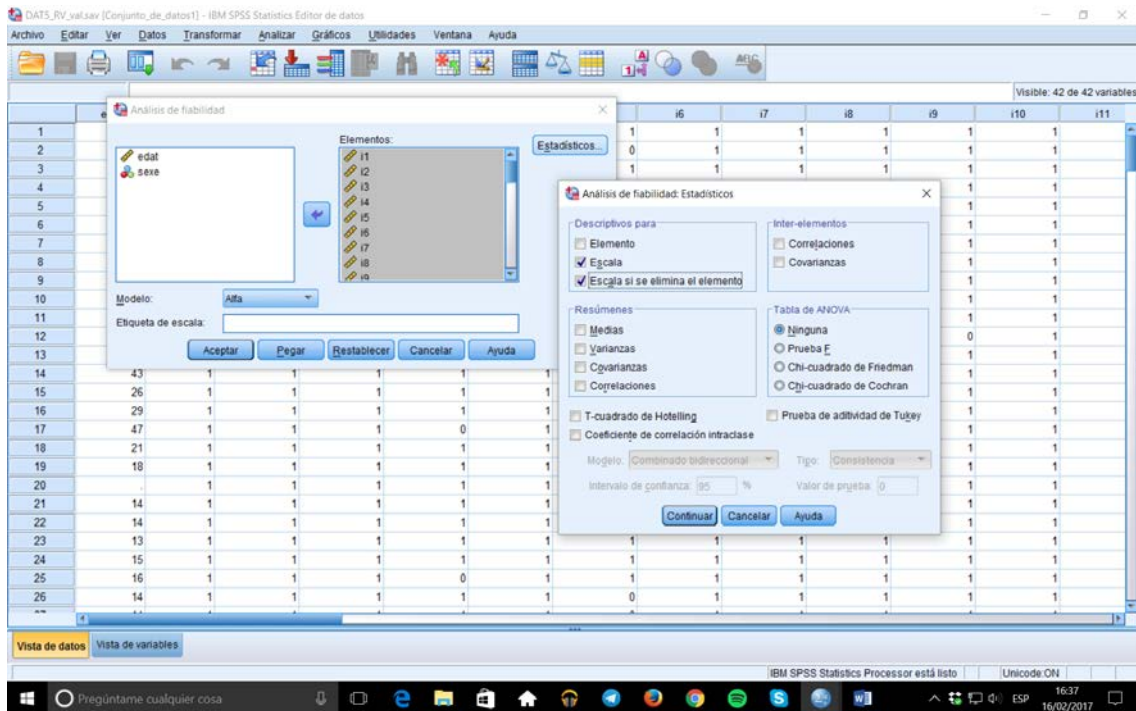
A més de la dificultat, és important que els ítems DISCRIMINEN entre persones amb puntuacions baixes o altes en el test. Per a comprovar-ho, utilitzarem l'ÍNDIX D'HOMOGENEÏTAT CORREGIT. A més, amb aquesta anàlisi podrem veure també l'IMPACTE DE CADA ÍTEM SOBRE LA FIABILITAT DEL TEST (tercer criteri, el comentarem més avant).

Menú Analitzar

Escala

Anàlisi de fiabilitat

Al botó d'estadístics cal demanar també descriptius per a: «Escala» i «Escala si s'elimina element»



Estadísticas de total-elemento

	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	A C e h
i1	25,24	26,957	,327	,820
i2	25,22	27,229	,265	
i3	25,28	27,006	,184	
i4	25,27	26,852	,273	
i5	25,31	26,751	,232	
i6	25,24	27,101	,226	
i7	25,24	27,079	,248	
i8	25,22	27,307	,233	
i9	25,22	27,307	,233	
i10	25,24	26,974	,288	

Índex d'homogeneïtat corregit (IH)

L'IH és la correlació entre l'ítem i la puntuació total del test feta amb tots els ítems EXCEPTE aquest ítem. Es considera un valor adequat si és SUPERIOR a 0,200. En aquest cas tots els ítems superen el criteri excepte l'ítem 3.

Hi ha autors que prenen punts de tall més estrictes (superiors a 0,300).

i11	25,29	26,827	,243	,820
i12	25,28	26,896	,235	,821
i13	25,46	25,506	,433	,814
i14	25,40	25,845	,397	,816
i15	25,29	26,447	,370	,817
i16	25,30	26,699	,266	,820
i17	25,38	26,075	,357	,817
i18	25,57	25,811	,317	,818
i19	25,47	26,270	,248	,821
i20	25,81	26,186	,235	,822
i21	25,84	25,861	,308	,819
i22	25,83	25,729	,332	,818
i23	25,83	25,729	,332	,818
i24	25,31	26,455	,330	,818
i25	25,32	26,178	,401	,816
i26	25,29	26,342	,400	,817
i27	25,47	26,076	,295	,819
i28	25,56	25,631	,357	,817
i29	25,48	26,431	,210	,822
i30	25,70	25,503	,365	,817
i31	25,79	25,750	,322	,818
i32	25,74	25,723	,322	,818
i33	25,87	25,752	,339	,818
i34	26,01	26,043	,347	,817
i35	25,88	25,557	,385	,816
i36	25,98	26,160	,297	,819
i37	26,07	26,721	,217	,821

i38	26,09	26,600	,279	,819
i39	26,06	26,447	,289	,819
i40	26,08	26,468	,302	,819

PAS 3. Tercer criteri de qualitat: IMPACTE DE L'ÍTEM sobre la consistència interna del test

L'impacte de l'ítem sobre el test s'analitza comparant la consistència interna del test complet (Alfa = 0,823, vegeu la taula d'estadístiques de fiabilitat a continuació) amb la qual s'obtidria després d'eliminar cada un dels ítems. Tornant a la taula d'ítems anterior, si s'elimina l'ítem 40 veiem que alfa baixa a 0,819, i si eliminem l'ítem 36, alfa baixa a 0,821. I així amb tots els ítems. Quan la consistència interna del test (Alfa) puja després d'eliminar un ítem és perquè aquest ítem empitjora el test i per tant és inadequat. En aquest cas no succeeix amb cap ítem (vegeu la darrera columna de la taula d'estadístiques de total d'element a la pàgina següent), ja que eliminant qualsevol dels quaranta ítems només aconseguiríem disminuir la consistència interna de l'escala.

Estadísticos de fiabilidad

Alfa de Cronbach	N elements
,823	40

Estadísticos de total d'element

	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	Alfa de Cronbach si el elemento se ha suprimido
i1	25,24			,820
i2	25,22			,821
i3	25,28			,822
i4	25,27			,820
i5	25,31			,821
i6	25,24			,821

En observar els valors de la darrera columna es comprova que NO hi ha cap ítem que després d'eliminar-lo faça millorar la consistència interna de l'escala.

i7	25,24	27,079	,248	,821
i8	25,22	27,307	,233	,822
i9	25,22	27,307	,233	,822
i10	25,24	26,974	,288	,820
i11	25,29	26,827	,243	,820
i12	25,28	26,896	,235	,821
i13	25,46	25,506	,433	,814
i14	25,40	25,845	,397	,816
i15	25,29	26,447	,370	,817
i16	25,30	26,699	,266	,820
i17	25,38	26,075	,357	,817
i18	25,57	25,811	,317	,818
i19	25,47	26,270	,248	,821
i20	25,81	26,186	,235	,822
i21	25,84	25,861	,308	,819
i22	25,83	25,729	,332	,818
i23	25,83	25,729	,332	,818
i24	25,31	26,455	,330	,818
i25	25,32	26,178	,401	,816
i26	25,29	26,342	,400	,817
i27	25,47	26,076	,295	,819
i28	25,56	25,631	,357	,817
i29	25,48	26,431	,210	,822
i30	25,70	25,503	,365	,817
i31	25,79	25,750	,322	,818
i32	25,74	25,723	,322	,818
i33	25,87	25,752	,339	,818

i34	26,01	26,043	,347	,817
i35	25,88	25,557	,385	,816
i36	25,98	26,160	,297	,819
i37	26,07	26,721	,217	,821
i38	26,09	26,600	,279	,819
i39	26,06	26,447	,289	,819
i40	26,08	26,468	,302	,819

En **conclusió**, en general els ítems són discriminadors, de dificultat variada, encara que possiblement hi ha massa ítems molt fàcils (tot i incloure en els tests per motivar els subjectes). L'ítem 3 no contribueix a discriminar entre els subjectes amb diferents nivells d'intel·ligència (IH = 0,183), si bé l'eliminació no sembla afectar la fiabilitat total de l'escala.

Ara, **per a practicar**, repetiu les anàlisis anteriors amb el mateix arxiu de dades, però **SEGMENTANT PER SEXE**. Primer seleccioneu els homes i feu les anàlisis anteriors només amb aquest grup. Després, seleccioneu les dones i torneu a analitzar l'escala.

Per a seleccionar un grup de subjectes (per exemple, els homes, als quals hem posat etiqueta numèrica igual a 1) cal fer el pas següent:

Menú Dades

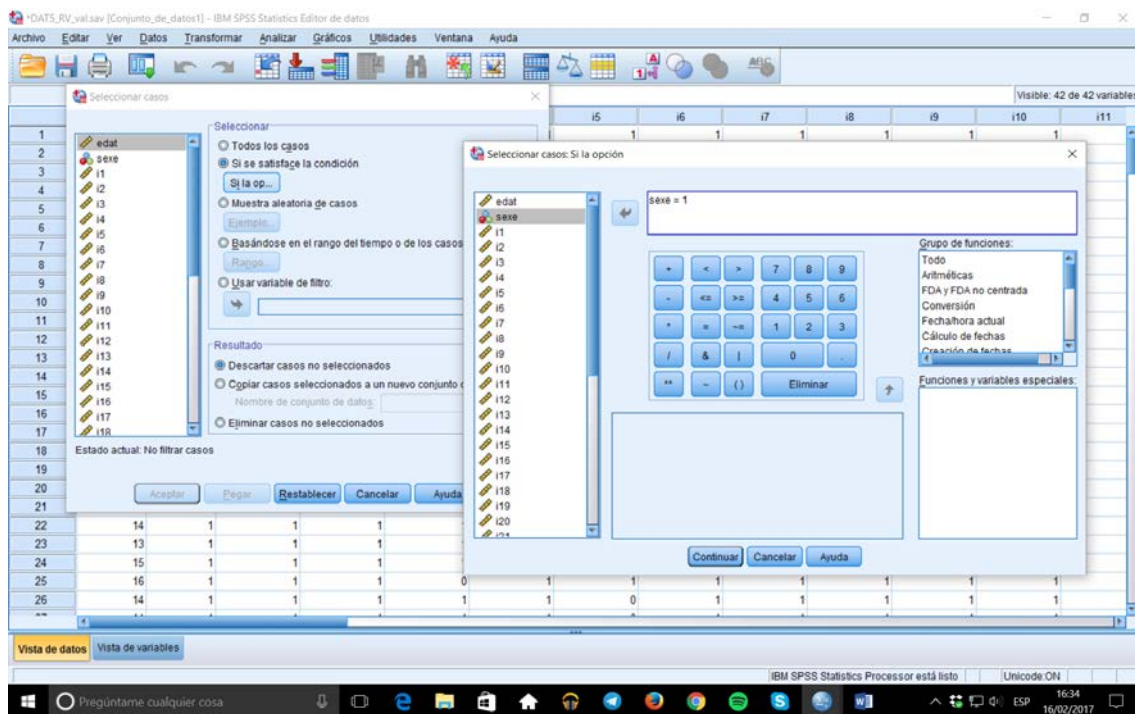
Seleccioneu casos

Seleccioneu «Si se satisfà la condició»

Cliqueu en el botó «Si l'opció»

Introduïu la condició seleccionant la variable Sexe: sexe = 1

(Quan es tracte de les dones, la condició serà sexe = 2)



Bloc 2. Teoria Clàssica de Tests (TCT)

B2. Pràctica 1. SUPÒSITS TEORIA CLÀSSICA DE TESTS (TCT)

Recordant el procés d'inferència psicomètrica és evident que la teoria del test és una part important que inclou tres aspectes: puntuació observada, puntuació vertadera i la relació que té amb el constructe que s'avalua. Per tant, les teories dels test inclouen els dos aspectes més estudiats dels instruments de mesura: la fiabilitat i la validesa. Per a què necessitem aquestes teories del test? Perquè els instruments contenen error de mesura i perquè els trets que habitualment mesurem en psicologia no són observables i han de ser mesurats indirectament.

La TCT (teoria clàssica de test) i la TRI (teoria de resposta a l'ítem) han sigut les dues teories de test més desenvolupades i provenen del desenvolupament de l'escalament de Likert i Guttman, respectivament.

La TCT es basa en el **model lineal de Spearman** i en el compliment del seus supòsits. La clau del model es resumeix en:

$$X = V + E$$

X = puntuació observada en el test

V = puntuació vertadera

E = error de mesura aleatori

Suposem que poguérem aplicar infinites vegades un test a la mateixa persona, la distribució de les puntuacions observades coincidiria amb la puntuació vertadera, ja que l'error de mesura és aleatori. Però, oh! Gran dificultat! No és possible aplicar infinites vegades el mateix instrument i, a més, suposar que mai canviaran els resultats.

Aleshores què podem fer? Partir dels supòsits de paral·lisme (que no es poden comprovar) i comprovar-ne les conseqüències:

Supòsit	Conseqüència derivada
$X = V + E; X' = V' + E'$	$X = X'$
Els tests paral·lels tenen la mateixa puntuació vertadera.	La mitjana de les puntuacions observades serà igual.
$S_e^2 = S_{e'}$	$S_x^2 = S_{x'}$
El tests paral·lels tenen la mateixa variància d'error.	La variància de les puntuacions observades serà igual.

Un indicador de la fiabilitat pot ser la correlació entre les puntuacions observades i les puntuacions vertaderes, al qual denominem **índex de fiabilitat** (r_{xv}). I, si elevem al quadrat

l'índex, obtindrem informació de la variància observada que podem atribuir a la variància vertadera. Açò és el que denominem **coeficient de fiabilitat** ($r_{xx'}$), que es pot estimar a partir de dues formes paral·leles d'un instrument. I aquest coeficient oscil·la entre 0 i 1!

Si ja hem comprés la importància de la relació entre les puntuacions observades, les vertaderes i l'error, podem entendre que:

$$r_{xx'} = \frac{S_v^2}{S_x^2} = 1 - \frac{S_e^2}{S_x^2}$$

I, per tant, tenim que

$$S_v^2 = S_x^2 \cdot r_{xx'}$$

I que

$$S_e^2 = S_x^2(1 - r_{xx'})$$

Tot coneixent aquestes qüestions, respon a les preguntes següents:

1. En un test d'assertivitat, tenim dues mesures paral·leles A i A'. La variància de les puntuacions observades és 24 i el coeficient de fiabilitat és 0,91 (escriu en el buit que hi ha a continuació les dades que et donem en l'exercici i respon en acabant a les preguntes).

a) Quines dues condicions de paral·lisme s'hi han de complir?

b) Pot ser la variància de les puntuacions vertaderes 26? Per què?

c) Quin és l'**índex** de fiabilitat? Què indica?

d) Quina es la proporció de variància observada atribuïble a la variància vertadera?

- e) Quina és la proporció de variància observada que és atribuïble a la variància de l'error?
- f) Quant val la variància de l'error del test? I l'error típic?
2. Si la variància de les puntuacions vertaderes és igual a la variància de les puntuacions observades:
- a) Quant val la variància de l'error?
- b) I el coeficient de fiabilitat?
3. Si la variància de les puntuacions observades és igual a la variància de l'error:
- a) Quant val la variància de les puntuacions vertaderes?
- b) I el coeficient de fiabilitat?
4. El coeficient de fiabilitat d'un test d'atenció en una mostra de 200 nens en segon de Primària fou 0,89. La mitjana de les puntuacions fou 30,99 i la variància observada, 36 (escriu en el buit que hi ha a continuació les dades que et donem en l'exercici i respon en acabant a les preguntes).
- a) Quant val la variància de l'error del test?
- b) I l'error típic o grandària mitjana de l'error?
- c) Quin valor té la proporció de variància vertadera continguda a la variància observada?

SOLUCIONS

1.

- a) Que tinguem la mateixa puntuació vertadera i la mateixa variància de l'error. Però no es comproven les condicions, sinó les conseqüències: tenir la mateixa mitjana i variabilitat.
- b) No, ja que $X = V + E$, per tant la variància de les puntuacions observades sempre serà superior, i en aquest cas és 24. Per tant, la variància de les puntuacions vertaderes mai podrà ser 26.
- c) L'arrel quadrat de $0,91 = 0,95$
- d) El 91%, ja que el coeficient de fiabilitat és 0,91.
- e) 9%, ja que és exactament la variància de la puntuació observada que **no** és explicada per la vertadera.
- f) Si $S_e^2 = S_x^2(1 - r_{xx'})$, també podem calcular-ne l'arrel quadrada.

Dues maneres de calcular-la:

$$24(1 - 0,91) = 2,16$$

O:

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

$$\text{Com que: } S_x = \sqrt{24} = 4,9$$

$$\text{Tenim que: } S_e = 4,9 \sqrt{1 - 0,91} = 1,47$$

2.

- a) Si pensem que $X = V + E$, i que $X = V$, llavors E ha de ser 0.
- b) El coeficient de fiabilitat ha de ser 1, perfecte, ja que l'instrument no conté error. Evidentment, açò es un supòsit teòric que, desgraciadament, no es donarà en la realitat mai.

3.

- a) Si pensem que $X = V + E$, i que $X = E$, llavors V ha de ser 0.
- b) El coeficient de fiabilitat seria 0, ja que l'instrument únicament mesura error, però res de puntuació vertadera.

4.

$$\begin{aligned} \text{a) } S_e &= S_x \sqrt{1 - r_{xx'}} \\ S_e &= 6 \sqrt{1 - 0,89} = 1,98 = 1,99^2 = 3,96 \end{aligned}$$

- b) 1,98
- c) 89%, ja que el coeficient de fiabilitat és 0,89.

B2. Pràctica 2. Formes d'estimar la fiabilitat

La fiabilitat és un concepte a estimar i hi ha diferents maneres de fer-ho. En repassarem tres mètodes:

1. Mètode de formes paral·leles: consisteix en la correlació lineal entre dues sèries de punts en dues formes paral·leles d'un test. S'entén la fiabilitat com a equivalència entre mesures. Crear formes paral·leles d'un test amb dificultat similar és complicat, i fatigós per a la persona que el contesta.
2. Mètode test-retest: és necessari administrar el test en dos moments diferents i calcular la correlació entre les puntuacions obtingudes en els dos moments. S'entén la fiabilitat com a estabilitat. Presenta desavantatges com els canvis que es poden produir en les persones i la variabilitat del lapse de temps triat.
3. Mètode de dues meitats: subdividim el test en dues parts i les correlacionem. S'entén la fiabilitat com a consistència interna i s'ha d'aplicar la correcció Spearman-Brown per test de longitud doble.

Tots els procediments requereixen el compliment dels supòsits de paral·lelisme clàssic, es a dir, igualtat de puntuacions vertaderes i de variàncies d'error. O en tot cas, les conseqüències: igualtat de mitjanes i de variàncies de les puntuacions observades. Per tant, hem de comprovar els supòsits abans d'estimar la fiabilitat.

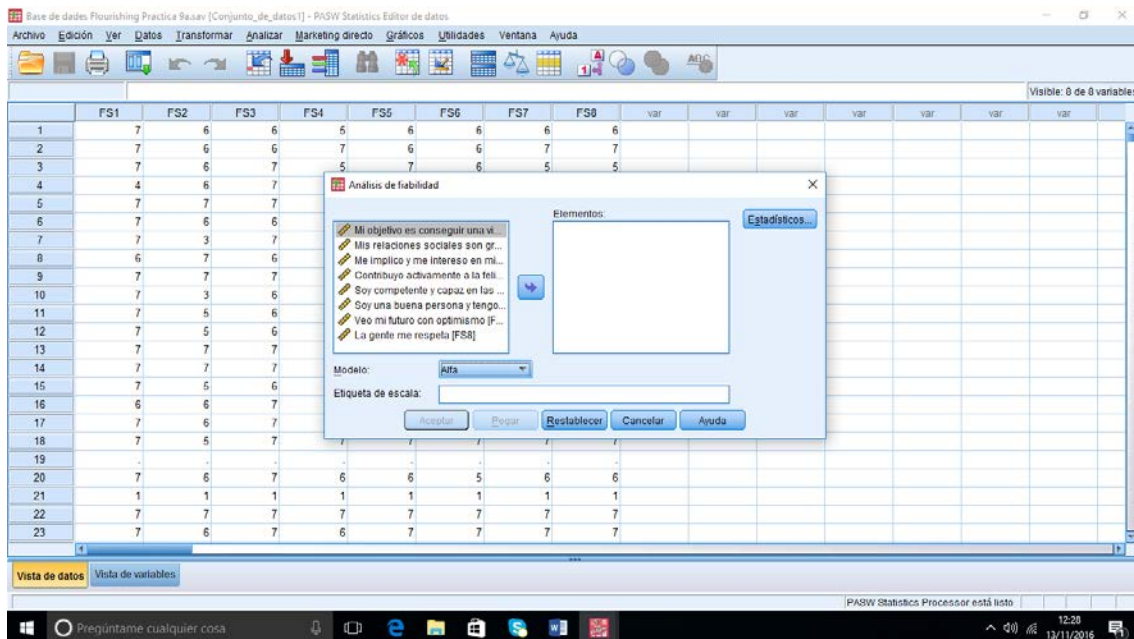
La pràctica té l'objectiu de calcular la fiabilitat pel mètode de dues meitats i es segueix amb la base de dades de *Flourishing*, disponible en [aquest enllaç](#).

Per a observar el coeficients de fiabilitat que podem obtenir seguim els passos:

Menú Analitzar

Escales

Anàlisi de fiabilitat

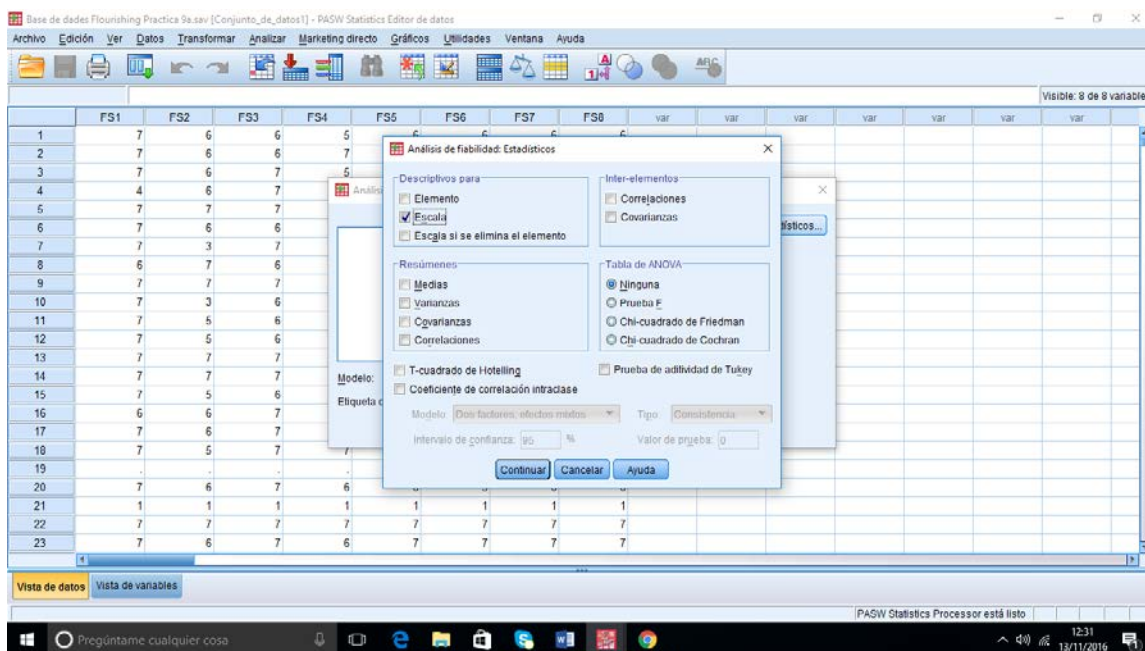


- Disposem de diferents models. Seleccionem:

Model - Dues meitats

- Per a comprovar que la subdivisió de dues meitats que va fer SPSS és bona, és a dir, que es compleixen les conseqüències dels supòsits, seleccionem:

Estadístics – Descriptius per a – Escala



- Obtindrem els estadístics de fiabilitat i d'escala següents:

Estadísticos de fiabilidad

Alfa de Cronbach	Parte 1	Valor	,764
		N de elementos	4 ^a
	Parte 2	Valor	,797
		N de elementos	4 ^b
		N total de elementos	8
Correlación entre formas			,743
Coeficiente de Spearman-Brown	Longitud igual		,853
	Longitud desigual		,853
Dos mitades de Guttman			,849

a. Los elementos son: FS1, FS2, FS3, FS4

b. Los elementos son: FS5, FS6, FS7, FS8

Estadísticos de la escala

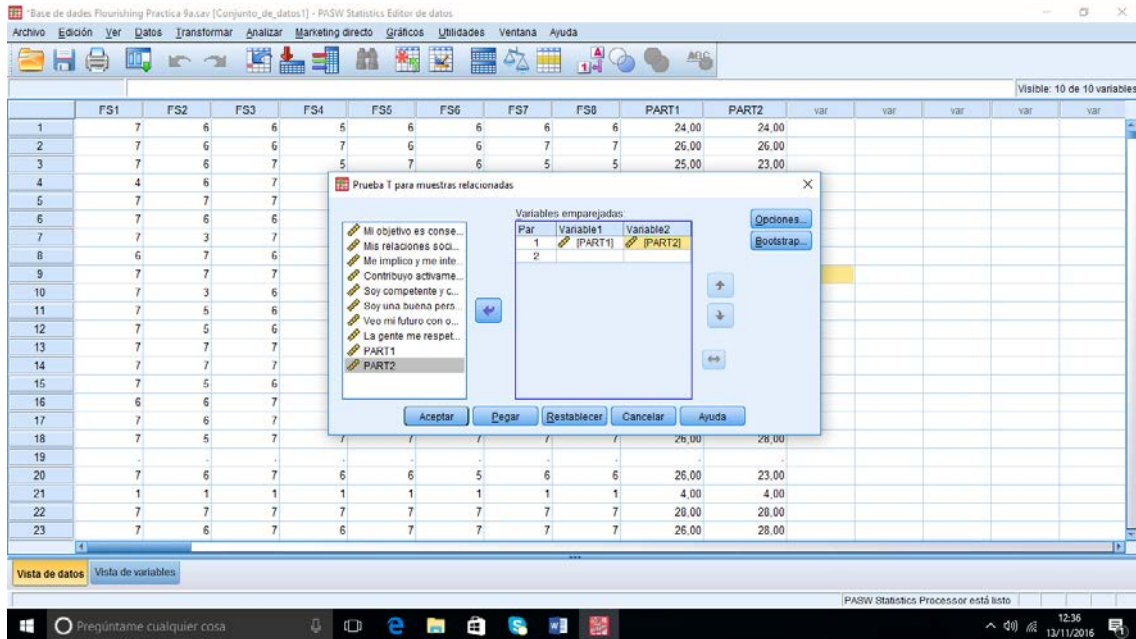
	Media	Varianza	Desviación típica	N de elementos
Part 1	23,84	11,038	3,322	4 ^a
Part 2	23,31	13,914	3,730	4 ^b
Ambdues parts	47,15	43,366	6,585	8

a. Els elements són: FS1, FS2, FS3, FS4

b. Els elements són: FS5, FS6, FS7, FS8

- En primer lloc, s'ha de comprovar que les mitjanes són iguals, per això calculem les puntuacions totals per a les dues meitats i fem un contrast d'hipòtesi per a comprovar si les mitjanes de les dues meitats són prou properes entre si.
- Es recorda que es poden calcular puntuacions totals per a les parts, en el menú **Transformar - Calcular nova variable**. Es crearan dues variables: una amb el sumatori dels ítems 1-4 i una altra dels ítems 5-8.
- Per a fer el contrast d'hipòtesi, es demana:

Menú Analitzar – Comparar mitjanes - Prova t per a mostres relacionades ja que són les mateixes persones les que responen a ambdues parts.



- Els resultats que obtenim són:

Estadísticas de muestras emparejadas

	Media	N	Desviación estándar	Media de error estándar
Par 1 PART 1	5,9600	550	,83059	,03542
PART 2	5,8286	550	,93252	,03976

Prueba de muestras emparejadas

	Diferencias emparejadas					t	gl	Sig. (bilateral)
	Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
				Inferior	Superior			
Par 1 PART1 - PART2	,13136	,63922	,02726	,07782	,18490	4,820	549	,000

Es pot comprovar que encara que descriptivament les mitjanes pareixen molt properes (5,96 i 5,82), la diferència entre elles és significativa i, per tant, no es pot confirmar la primera conseqüència dels supòsits de paral·lelisme, la igualtat de mitjanes.

També s'ha de comprovar la segona conseqüència, la igualtat de variàncies. Això no ho fa directament el SPSS i, per tant, s'ha de fer el càlcul a mà.

La fórmula per al càlcul:

$$t_{\text{emp}} = \frac{(F - 1)\sqrt{n - 2}}{2\sqrt{F(1 - r^2)}}$$

que es distribueix amb n-2 graus de llibertat

F és $F = S_1^2/S_2^2$

n és la grandària de la mostra i

r és la correlació entre ambdues parts.

Calculamos:

$$F = 13.91/11.04 = 1.26$$

$$t_{\text{emp}} = \frac{(F - 1)\sqrt{n - 2}}{2\sqrt{F(1 - r^2)}} = \frac{(1.26 - 1)\sqrt{550 - 2}}{2\sqrt{1.26(1 - 0.74^2)}} = 4.057$$

Aquest valor és a la zona de rebuig de la H_0 , perquè supera el valor teòric per a $t_{975,548} = 1,96$. Significa el mateix que dir que la probabilitat d'acceptar H_0 es $p < 0,05$. Per tant, rebutgem també la H_0 d'igualtat de variàncies.

EXERCICI

Ara, estima la fiabilitat amb el mètode de dues meitats, comprova els supòsits de paral·lelisme, en la base de dades disponible en [aquest enllaç](#) amb dades del mateix qüestionari de *Flourishing* en una altra mostra.

- a) Es compleixen els supòsits de paral·lelisme?
- b) Quin és el coeficient de fiabilitat que hauríem d'utilitzar en aquest cas?

SOLUCIONS

a)

Estadísticos de fiabilidad

Alfa de Cronbach	Parte 1	Valor	,768
		N de elementos	4 ^a
	Parte 2	Valor	,795
		N de elementos	4 ^b
		N total de elementos	8
Correlación entre formas			,748
Coeficiente de Spearman-Brown	Longitud igual		,856
	Longitud desigual		,856
Dos mitades de Guttman			,853

a. Los elementos son: FS1, FS2, FS3, FS4

b. Los elementos son: FS5, FS6, FS7, FS8

Estadísticos de la escala

	Media	Varianza	Desviación típica	N de elementos
Parte 1	23,70	11,729	3,425	4 ^a
Parte 2	23,23	14,480	3,805	4 ^b
Ambas partes	46,94	45,705	6,761	8

a. Los elementos son: FS1, FS2, FS3, FS4

b. Los elementos son: FS5, FS6, FS7, FS8

- Comprovem els supòsits:

Igualtat de mitjanes:

Estadísticas de muestras emparejadas

		Media	N	Desviación estándar	Media de error estándar
Par 1	PART1	5,9261	531	,85621	,03716
	PART2	5,8079	531	,95132	,04128

Prueba de muestras emparejadas

	Diferencias emparejadas					t	gl	Sig. (bilateral)
	Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
				Inferior	Superior			
Par 1 PART1 - PART2	,11817	,64778	,02811	,06295	,17340	4,204	530	,000

No hi ha igualtat de mitjanes, el primer supòsit no es compleix.

Igualtat de variàncies:

$$F=14.48/11.73 = 1.23$$

$$t_{\text{emp}} = \frac{(F - 1)\sqrt{n - 2}}{2\sqrt{F(1 - r^2)}} = \frac{(1.23 - 1)\sqrt{531 - 2}}{2\sqrt{1.23(1 - 0.75^2)}} = 3.6$$

Aquest valor és a la zona de rebuig de la H_0 , perquè supera el valor teòric per a $t_{975,529} = 1,96$. Significa el mateix que dir que la probabilitat d'acceptar H_0 és $p < 0,05$. Per tant, rebutgem també la H_0 d'igualtat de variàncies.

B2. Pràctica 3. Factors que afecten la fiabilitat

Ja coneixem que la fiabilitat és la propietat que parla de l'exactitud en la mesura d'un constructe psicològic, és a dir, una puntuació d'un test és més fiable si conté poc error de mesura. La qüestió és: la fiabilitat és una propietat inherent al test? S'ha comprovat que no. El factors que poden influir en el coeficient de fiabilitat són la longitud del test i la variabilitat de la mostra.

- Longitud del test: com més ítems componguen un test, la mostra de conductes representades és més gran i, per tant, cometem menys errors en estimar la puntuació vertadera.
- Variabilitat de la mostra: el coeficient de fiabilitat depèn de la proporció en què les diferències que detecta el test (variància de les puntuacions observades) són degudes a les diferències que realment hi ha entre les persones (variància de les puntuacions vertaderes). Però, com més grans són les diferències que hi ha entre les persones, menys es noten les distorsions que introdueix l'error i, per tant, més gran serà la fiabilitat.

Resumint, la fiabilitat és més gran com més variable siga la nostra mostra i com més ítems formen el test.

Si volem augmentar o disminuir la fiabilitat d'un instrument, disposem de les fórmules de Spearman-Brown per a calcular la nova fiabilitat.

FÓRMULES DE SPEARMAN-BROWN

Cas general:

$$r_{cc'} = \frac{k * r_{xx'}}{1 + (k-1)r_{xx'}}$$

$r_{xx'}$ = coeficient de fiabilitat

k = longitud final / longitud inicial

Cas de longitud doble:

$$r_{cc'} = \frac{2r_{xx'}}{1+r_{xx'}}$$

Ítems per a arribar a una fiabilitat:

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

$r_{cc'}$ = nova fiabilitat

FÓRMULA DE FIABILITAT PER A UNA DETERMINADA VARIABILITAT DE LA MOSTRA

$$r_{XX_1} = 1 - \frac{S^2(1-r_{XX'})}{S^2''}$$

S^2 = variància de la mostra original

S^2'' = variància de la mostra proposada

EXERCICIS

1. El coeficient de fiabilitat de l'escala Flourishing (8 ítems) en una mostra de 950 persones va ser de 0,67. Si afegim quatre ítems més:
 - a) Quant val el coeficient de fiabilitat?
 - b) Si augmentem la longitud del test fins a 24 ítems, quina és la fiabilitat?
 - c) Si volem arribar a una fiabilitat de 0,95, quants ítems cal afegir?
2. Apliquem un test de comprensió verbal format per vint ítems a dos-cents estudiants de cinquè de Primària. La puntuació mitjana en el test és de 15, la variància de les puntuacions és de 25 i el coeficient de fiabilitat, 0,89. Apliquem un altre test de comprensió verbal a dos-cents estudiants amb altes capacitats. La puntuació mitjana en el test és de 24, la desviació típica de les puntuacions és de 4 i el coeficient de fiabilitat, 0,72.
 - a) Es poden comparar directament els coeficients de fiabilitat? Raona la resposta.

- b) Quin seria el coeficient de fiabilitat del test en la mostra de estudiants d'altas capacitats en condicions d'igualtat de variabilitat? En quina mostra seria més fiable?
3. Estudiant un test d'aptitud verbal en una mostra de professors, la variància dels errors de mesura va ser igual a 5 i representava el 8% de la variància total del test. Aquest test d'aptitud verbal va obtenir una mitjana de 25 i estava format per huitanta ítems.
- a) Quin és el coeficient de fiabilitat del test?
- b) Quina és la variància de les puntuacions vertaderes?
- c) Els creadors del test s'han adonat que el test és molt llarg i volen comprovar si amb la meitat d'ítems seria suficientment fiable. Seria correcte?
- d) Finalment, es replantegen que volen un coeficient de fiabilitat de 0,80. Quants ítems tindria el test en aquest cas?

SOLUCIONS

1.

a) $r_{xx'} = 0,67$; + 4 ítems

$$r_{cc'} = \frac{k \cdot r_{xx'}}{1 + (k-1)r_{xx'}}$$

$k = 12/8 = 1,5$. La longitud inicial era 8 i n'afegim 4 més, per tant la longitud final és 12.

$$r_{cc'} = \frac{1,5 \cdot 0,67}{1 + (1,5-1)0,67} = 0,75$$

a) $r_{xx'} = 0,67$; fins a 24 ítems

$$r_{cc'} = \frac{k \cdot r_{xx'}}{1 + (k-1)r_{xx'}}$$

$k = 24/8 = 3$.

$$r_{cc'} = \frac{3 \cdot 0,67}{1 + (3-1)0,67} = 0,85$$

b)

$$k = \frac{r_{cc'}(1-r_{xx'})}{r_{xx'}(1-r_{cc'})} = \frac{0,95(1-0,67)}{0,67(1-0,95)} = 10,33$$

Ja tenim la k. Ara substituïm: $10,33 = x/8$ i, per tant, $x = 8 * 10,33 = 82,66$

83, aproximadament, seran els ítems finals, però em demana els ítems que hem d'afegir-hi. Restem 83 menys els ítems inicials,

$83 - 8 = 75$ ítems més que ha de afegir-s'hi, si volem arribar a 0,95 de fiabilitat.

2.

a) No es poden comparar els coeficients de fiabilitat perquè la interpretació d'aquest coeficient depèn de la variabilitat de les mostres en les quals el test és aplicat. En aquest cas, la variabilitat de la mostra d'alumnes d'altres capacitats és més reduïda i, per tant, per a comparar-les hem d'igualar les condicions de variabilitat.

b) $r_{xx_1} = 1 - \frac{16(1-0,72)}{25} = 0,82$

Quan la variabilitat del segon test en la mostra d'altres capacitats és igualada a la variància de la primera mostra, el coeficient de fiabilitat és 0,82. Encara així, és menor que la fiabilitat obtinguda amb la primera mostra.

3.

- a) Si la variància atribuïda a l'error és el 8%, significa que l'atribuïda a les puntuacions vertaderes és 92%. Això significa que el coeficient de fiabilitat és 0,92.
- b) Coneixent la variància dels errors i el coeficient de fiabilitat, podem conèixer la variància de les puntuacions observades, la qual la necessitem per al càlcul de la variància de les vertaderes. Per tant,

$$S^2_e = S^2_x(1-r_{xx'})$$

$$5 = S^2_x(1-0,92)$$

$$S^2_x = 5/0,08 = 62,5 \text{ (variància de les puntuacions observades)}$$

$$\text{Si } r_{xx'} = \frac{S^2_v}{S^2_x}$$

0,92 * 62,5 = 57,5 és la variància de les puntuacions vertaderes.

I comprovem que

$$62,5 = 57,5 + 5$$

- c) Per al càlcul d'una nova fiabilitat, si el test tinguera quaranta ítems, calculem primer la k.

k = long. final / long. inicial

$$k = 40/80 = 0,5$$

$r_{cc'} = \frac{k * r_{xx'}}{1 + (k-1)r_{xx'}} = \frac{0,5 * 0,92}{1 + (0,5-1)0,92} = 0,85$, per tant, sí que valdria la pena deixar el test en la meitat d'ítems perquè la fiabilitat encara estaria en valors acceptables.

- d) Per al càlcul dels ítems necessaris, calculem primer la k.

$$k = \frac{r_{cc'}(1-r_{xx'})}{r_{xx'}(1-r_{cc'})} = \frac{0,80(1-0,92)}{0,92(1-0,80)} = 0,35$$

Però ens demana els ítems finals, per tant

k = long. final / long. inicial

$$0,35 = \text{long. final} / 80$$

Long. final = 0,35 * 80 = 28 ítems.

Amb 28 ítems, el test presentaria un coeficient de fiabilitat de 0,80.

B2.Pràctica 4. Estimació de puntuacions des de la TCT

Ja hem vist que podem conèixer la fiabilitat d'un instrument i com poden afectar-hi qüestions com la variabilitat de la mostra o el nombre d'ítems. Però el que vertaderament ens interessa és: en quina mesura la puntuació observada ens serveix per a estimar la puntuació vertadera?

Tenim tres formes d'obtenir una estimació de la puntuació vertadera:

1. Estimació puntual: la X es comporta com a indicador de la V, però no coneixem quin error introduïm.
2. Estimació amb equació de regressió: com que la relació és lineal s'utilitza una equació de regressió lineal. $Y' = A + BX$
3. Estimació per interval: com que coneixem $X = V + E$, tenim les dades de X i coneixem que els errors es distribueixen normalment, podem estimar la probabilitat associada a qualsevol quantitat d'error.

Practiquem amb les dues estimacions que sí que consideren l'error de mesura.

Estimació amb equació de regressió

$$Y' = A + BX$$

A = valor esperat en Y quan X = 0

B = canvi esperat en Y quan X augmenta una unitat

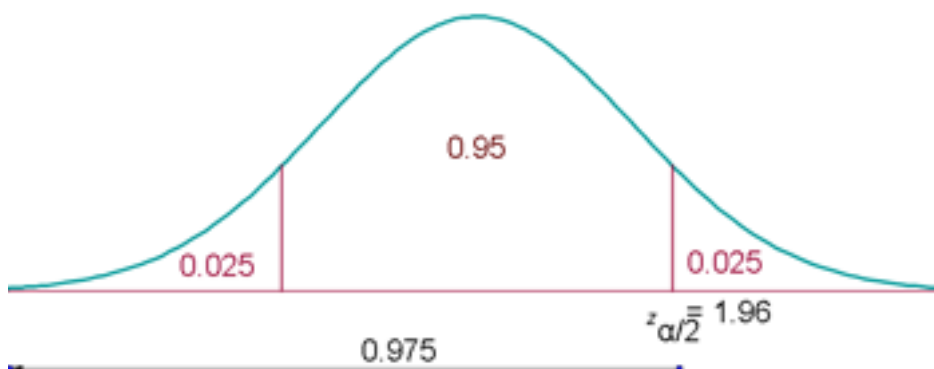
El càlcul d'estimació es fa així:

$$V' = \bar{X} \cdot (1 - r_{xx}) + r_{xx} \cdot X$$

Estimació amb interval de confiança

Passos:

1. Establim el nivell de confiança $(1-\alpha)$ desitjat: 0,95 o 0,99. Calculem l'interval en què es troba V amb una confiança de 99% o 95%.



2. Determinem la puntuació típica de l'error que delimita aquest interval:

$$- \alpha = 0,05 \quad z_{0,025} = 1,96$$

$$- \alpha = 0,01 \quad z_{0,005} = 2,58$$

3. Calculem l'error màxim:

$$E_{\text{màx}} = Z_{\alpha/2} * S_e$$

4. Obtenim els límits de confiança:

$$L_i = X - E_{\text{màx}}$$

$$L_s = X + E_{\text{màx}}$$

EXERCICIS

1. Hem aplicat un test de cohesió. La mitjana de les puntuacions observades ha sigut 45 i la consistència interna del test en la mostra és de 0,91. La variància dels errors és 4.
 - a. Quina puntuació vertadera estimaríem a una persona que ha obtingut una puntuació de 48 en el test?
 - b. Entre quins valors es troba la puntuació vertadera d'una persona que ha obtingut una puntuació de 36 amb un nivell de confiança del 0,95?
 - c.

2. En una mostra de tres-centes persones, hem aplicat un test d'autoconcepte. La mitjana de les puntuacions observades ha sigut 16 i l'índex de fiabilitat del test en la mostra és de 0,89. La desviació típica dels errors en la mostra és d'1,5.
 - a. Quina puntuació vertadera estimaríem a una persona que ha obtingut una puntuació de 12 en el test?
 - b. Entre quins valors es troba la puntuació vertadera d'una persona que ha obtingut una puntuació de 18 amb un nivell de risc de 0,01?

SOLUCIONS

1.

$$a) V' = \bar{X} \cdot (1 - r_{xx'}) + r_{xx'} \cdot X$$

$$V' = 45 (1-0,91) + 0,91 \cdot 48$$

$$V' = 4,05 + 43,68 = 47,73$$

La puntuació vertadera estimada seria 47,73.

$$b) 95\% \alpha = 0,05 \quad z_{0,025} = 1,96$$

$$S_e = \sqrt{4} = 2$$

$$E_{m\grave{a}x} = Z_{\alpha/2} \cdot S_e$$

$$E_{m\grave{a}x} = 1,96 \cdot 2 = 3,92$$

$$L_i = X - E_{m\grave{a}x} = 36 - 3,92 = 32,08$$

$$L_s = X + E_{m\grave{a}x} = 36 + 3,92 = 39,92$$

La puntuació vertadera estimada deu estar entre 32,08 i 39,92 amb un nivell de confiança de 95%.

2.

$$a) V' = \bar{X} \cdot (1 - r_{xx'}) + r_{xx'} \cdot X$$

Em donen l'índex de fiabilitat, per tant $r_{xx'} = \sqrt{0,89} = 0,94$

$$V' = 16 (1-0,94) + 0,94 \cdot 12$$

$$V' = 0,96 + 11,28 = 12,24$$

La puntuació vertadera estimada seria 12,24.

$$b) 99\% \alpha = 0,01 \quad z_{0,005} = 2,58$$

$$E_{m\grave{a}x} = Z_{\alpha/2} \cdot S_e$$

$$E_{m\grave{a}x} = 2,58 \cdot 1,5 = 3,87$$

$$L_i = X - E_{m\grave{a}x} = 18 - 3,87 = 14,13$$

$$L_s = X + E_{m\grave{a}x} = 18 + 3,87 = 21,87$$

La puntuació vertadera estimada deu estar entre 14,13 i 21,87 amb un nivell de confiança del 99%.

Bloc 3. Validesa

Pràctica 1. Evidències de validesa externa

Ja s'ha assenyalat que les dues qualitats d'un instrument d'avaluació són la fiabilitat i la validesa. Des dels anys 80, s'ha reconegut que la validesa no és una propietat inherent al test i per tant avaluem les evidències de validesa en una mostra determinada. Des de les últimes normes APA, les evidències de validesa han reforçat el concepte unitari i la necessitat de presentar diferents fonts d'evidència i no una només.

Entre les evidències de validesa es troben les que agrupen les relacions del test amb altres variables: evidència convergent, discriminant i criterial. Les característiques que tenen són les següents:

- Evidència de validesa convergent: amb un test que mesure el mateix tret, s'obtenen correlacions altes i positives. Habitualment s'utilitza coeficient de correlació.
- Evidència de validesa discriminant: amb un test que mesure un tret, que en principi no té relació amb el constructe que s'ha de validar; no s'obtenen correlacions. S'utilitza coeficient de correlació.
- Evidència de validesa criterial: capacitat predictiva del test sobre mesures criteri. Habitualment s'utilitza regressió lineal simple o múltiple.

Practiquem amb els *outputs* d'una base de dades que conté els resultats dels instruments següents: Flourishing Scale (l'escala que volem validar), SWLS (*Satisfaction with life scale*), WISC verbal i manipulatiu (intel·ligència), LOT-R (optimisme-pessimisme).

Esperem que:

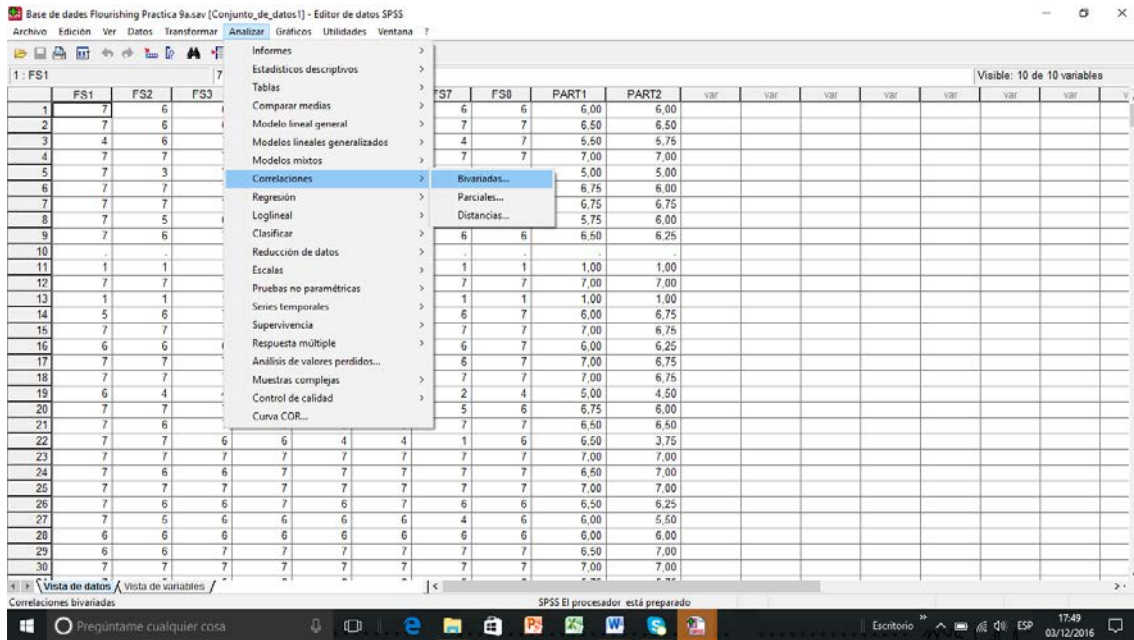
- FS presente correlacions positives i altes amb SWLS, ja que mesuren el mateix, benestar subjectiu.
- FS no presente correlació amb les puntuacions d'intel·ligència verbal i manipulativa del WISC, ja que no esperem relació entre el benestar i la intel·ligència.
- FS serà variable explicativa de l'optimisme mesurat amb el LOT-R.

En primer lloc, demanem correlació entre la FS, SWLS i WISC verbal i manipulatiu.

Menú Analitzar

Correlacions

Bivariades



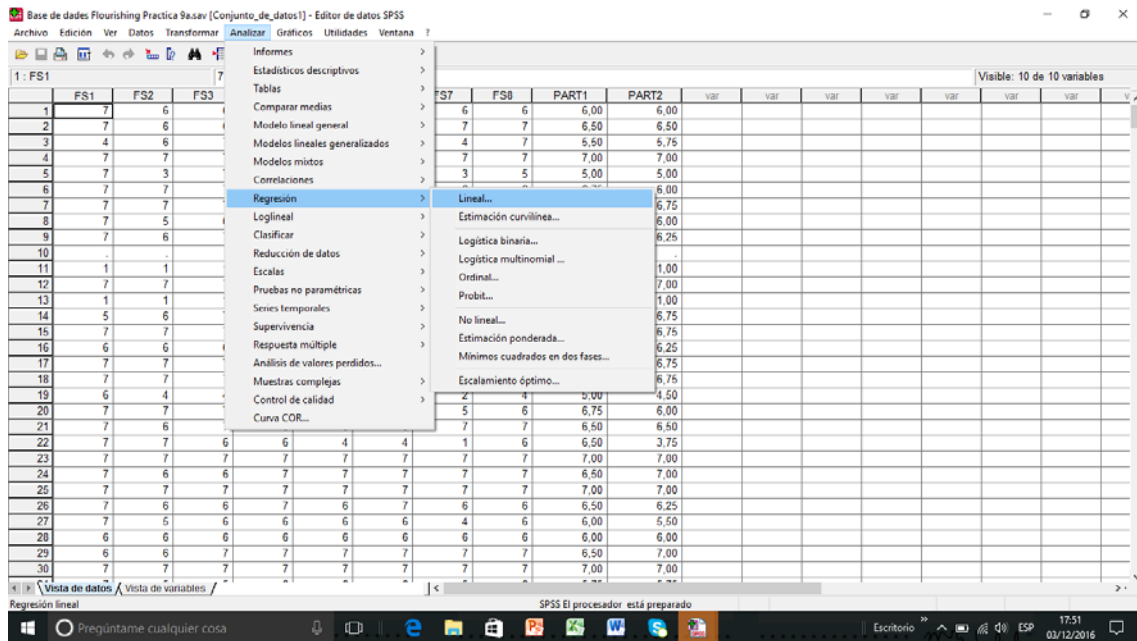
I afegim les quatre variables: FS, SLWS, WISC-V i WISC-M.

		FS	SWLS	WISC-V	WISC-M
FS	Correlación de Pearson	1	,555(**)	,120	,035
	Sig. (bilateral)		,000	,651	,842
	N	530	521	525	525
SWLS	Correlación de Pearson	,555(**)	1	,058	,098
	Sig. (bilateral)	,000		,769	,720
	N	521	521	521	521
WISC-V	Correlación de Pearson	,120	,058	1	,447(**)
	Sig. (bilateral)	,651	,769		,000
	N	525	521	525	525
WISC-M	Correlación de Pearson	,035	,098	,447(**)	1
	Sig. (bilateral)	,842	,720	,000	
	N	525	521	525	525

** La correlación es significativa al nivel 0,01 (bilateral).

Com s'observa, la correlació entre FS i SWLS, que les dues mesuren benestar subjectiu, és alta i significativa ($r = ,555$; $p < ,001$). Per tant, el FS presenta evidències de validesa convergent. D'altra banda, el FS no presenta correlacions amb els dos factors d'intel·ligència: verbal ($r = ,120$; $p = ,651$) i manipulativa ($r = ,035$; $p = ,842$), és a dir, no presenta relació amb mesures que no ha de relacionar-se.

Per a avaluar les evidències de validesa criterial, executem una anàlisi de regressió simple, on la variable independent (FS) ha de predir la experiència d'emocions positives.



Menú Analitzar

Regressió

Lineal

I afegim PANAS-P com a variable dependent i FS com a variable independent.

Aquests són els resultats:

Resum del model

Model	R	R quadrat	R quadrat corregida	Error típic. de l'estimació
1	,376(a)	,141	,140	6,36260

a Variables predictores: (constant), FS

ANOVA(b)

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regressió	3381,724	1	3381,724	83,535	,000(a)
	Residual	20565,180	508	40,483		
	Total	23946,904	509			

a Variables predictores: (constant), FS

b Variable dependent: PANASPOS

Coefficients(a)

Model		Coeficientes estandarizados		t	Sig.
		B	Error típ.		
1	(Constant)	14,636	2,211	6,619	,000
	FS	,426	,047	9,140	,000

Es pot observar que el FS es comporta com a predictor significatiu de les emocions positives ($t = 9,140$; $p < ,001$) amb una valor de beta de ,426. Això significa que cada canvi en FS suposa un augment de 0,426 en les emocions positives. El percentatge de variància de les emocions positives que explica la FS es el 14% ($r^2 = ,14$)

EXERCICIS

1. Es vol validar un nou instrument d'ansietat competitiva denominant ANS-D. Per això s'avaluen les evidències de validesa externa utilitzant les mesures SCAT (ansietat) i Lider (lideratge). Observa l'*output* següent i respon a les preguntes.

		Ans-D	SCAT	Lider
Ans-D	Correlación de Pearson	1	,615(**)	,014
	Sig. (bilateral)		,000	,851
	N	240	240	240
SCAT	Correlación de Pearson	,615(**)	1	,058
	Sig. (bilateral)	,000		,769
	N	240	240	240
Lider	Correlación de Pearson	,014	,058	1
	Sig. (bilateral)	,851	,769	
	N	240	240	240

- a) Podem assegurar que el test ANS-D presenta evidències de validesa convergent? Justifica la resposta.

- b) ANS-D presenta evidències de la validesa discriminant? Justifica la resposta.

2. En l'intent d'obtenir evidència test-criteri del mateix test ANS-D utilitzem una mesura dels pensaments negatius durant la competició per a avaluar-ne la capacitat predictiva. Observa els outputs i respon a les preguntes següents:

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,524(a)	,354	,358	7,32560

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regressió	3381,724	1	3381,724	95,254	,000(a)
	Residual	20565,180	238	40,483		
	Total	74246,904	239			

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta	B	Error típ.
1	(Constant)	16,874	2,254		7,635	,000
	ANS-D	,546	,051	,547	11,250	,000

- a) Creus que el test ANS-D es comporta com un predictor significatiu dels pensaments negatius durant la competició? Justifica la resposta.
- b) Si augmentem una unitat la puntuació de ANS-D, quin és el canvi que suposa en els pensaments negatius?
- c) Quin percentatge del pensaments negatius explica ANS-D?
3. D'altra banda, volem validar un qüestionari de capacitat d'atenció sostinguda per a nens de set a dotze anys (ATENT). Per a avaluar les evidències de validesa test-criteri, hem utilitzat una mesura de conductes disruptives a l'aula. Observa els outputs i respon a les preguntes:

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,140(a)	,084	,091	7,32560

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regressió	381,724	1	481,724	1,254	,094
	Residual	2165,180	140	18,483		
	Total	7466,904	141			

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta	B	Error típ.
1	(Constant)	6,874	1,254		2,635	,000
	ATENT	,080	,021	,011	1,254	,094

a) Creus que el test ATENT es comporta com un predictor significatiu de les conductes disruptives a l'aula? Justifica la resposta.

b) Si augmentem una unitat la puntuació d'ATENT, quin és el canvi que suposa en les conductes disruptives a l'aula?

c) Quin percentatge de les conductes disruptives a l'aula explica el test ATENT?

SOLUCIONS

1. a) Hi ha evidències de validesa convergent perquè ANS-D presenta una correlació significativa i elevada amb SCAT ($r = ,615$; $p < ,000$), ja que les dues avaluen ansietat.

b) Sí que presenta evidències de validesa discriminant, ja que la correlació entre ANS-D, que mesura ansietat, i Lider que avalua lideratge, presenten una correlació nul·la ($r = ,014$; $p = ,851$).
2. a) Es demostra que ANS-D es comporta com a predictor del pensament negatiu en competició perquè la prova F de l'anàlisi de regressió és significativa ($F = 95.254$; $p < ,000$).

b) Si augmentem una unitat la puntuació del ANS-D, suposa augment en els pensaments negatius de ,546.

c) ANS-D explica un 35,8% del pensaments negatius perquè la r quadrat corregida és ,358.
3. a) El test ATENT no és un predictor significatiu de les conductes disruptives a l'aula perquè la prova F de l'anàlisi de regressió no és significativa ($F = 1.254$; $p = ,094$).

b) Si augmentem una unitat la puntuació d'ATENT, suposa augment únicament de ,080 en les conductes disruptives a l'aula.

c) ATENT únicament explica el 9,1% de les conductes disruptives a l'aula.

B3. Pràctica 2. Evidències de validesa interna

En el procés de validació d'un nou instrument, s'ha de comprovar en quina mesura les relacions entre els ítems reproduïen l'estructura hipotetitzada. Per a fer-ho, habitualment s'utilitza la tècnica estadística multivariada de reducció de dades: l'anàlisi factorial.

L'anàlisi factorial té com a objectiu trobar el nombre mínim de dimensions que expliquen les diferents respostes de les persones o allò que els ítems tenen en comú. Aquesta anàlisi té diferents fases:

1. Com s'extrauen els factors i a partir de quina informació?
2. Quants factors obtenim i quina importància tenen?
3. Quins ítems integren cada factor?
4. Com s'interpreten aquests factors? Quin significat teòric tenen?
5. Com s'obtenen les puntuacions dels individus en els factors?

Però què és un factor? És un grup homogeni d'elements que tenen correlacions elevades entre si per a compartir alguna característica subjacent.

Per tant, els passos que hem de decidir són:

- Tipus d'anàlisi:
 - o Exploratòria: si no coneixem la distribució dels factors o és la primera aproximació a les dades de la mostra.
 - o Confirmatòria: si volem confirmar l'estructura proposada.
- Mètode d'extracció de factors:
 - o Anàlisi de components principals: analitzem tota la variància, la comuna i la no comuna, col·locant 1 a la diagonal de la matriu de correlacions. Sol sobreestimar els pesos factorials i el percentatge de variància explicada.
 - o Anàlisi d'eixos principals: s'analitza només la variància compartida, es calcula la comunalitat en la diagonal. És una anàlisi més conservadora, ja que té en compte l'error o variància no compartida de les variables. A més, s'utilitza aquest mètode d'extracció quan les distribucions són considerades no normals.
- Mètode de rotació:
 - o Ortogonal: quan considerem que els factors són independents.
 - o Obliqui: quan considerem que els factors estan relacionats.

Comencem amb l'anàlisi factorial amb el SPSS.

Treballarem amb el qüestionari d'autoconcepte social (AUSO). Està format per huit ítems (com que no està validat en valencià deixem els ítems en l'idioma original):

1. *Me siento implicado/a con la sociedad.*
2. *Me siendo aceptado/a cuando estoy en un grupo.*
3. *Tengo mucho éxito en mis relaciones con la gente.*
4. *Colaboro para mejorar la sociedad.*
5. *Caigo bien a la gente.*
6. *Me siento útil en la sociedad.*
7. *Cumplo adecuadamente las obligaciones que tengo como ciudadano.*
8. *Tengo muchos amigos.*

En principi, l'autoconcepte social presenta dos factors: responsabilitat social (1, 4, 6 i 7) i acceptació social (2, 3, 5 i 8). Comprovem com funciona en aquesta mostra de 278 alumnes de Psicologia.

Menú Analitzar

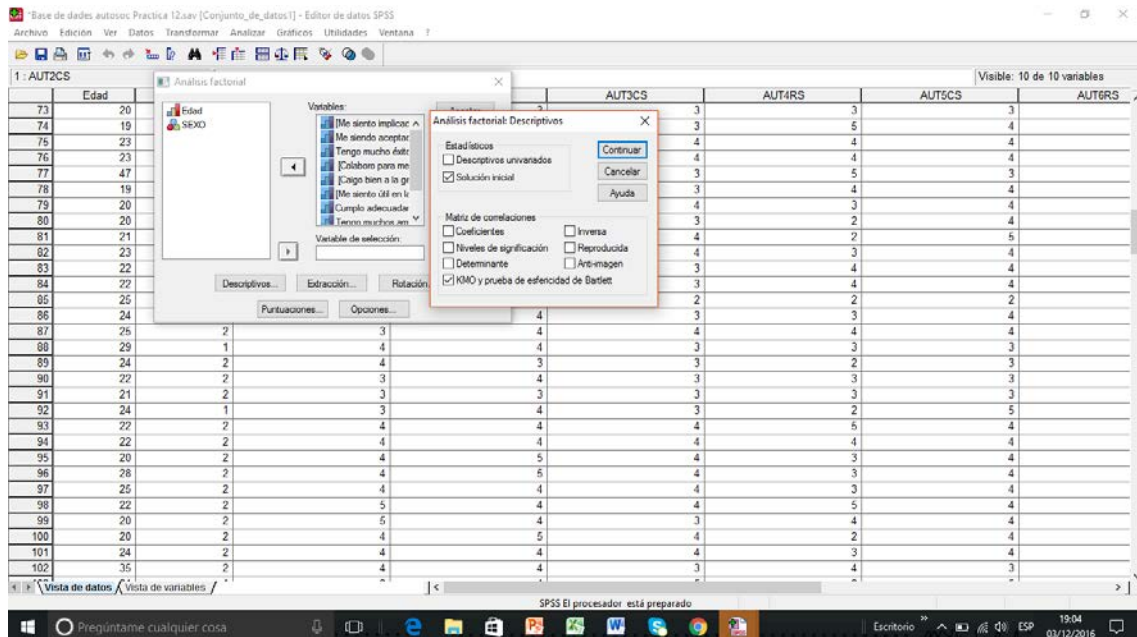
Reducció de dades

Anàlisi factorial

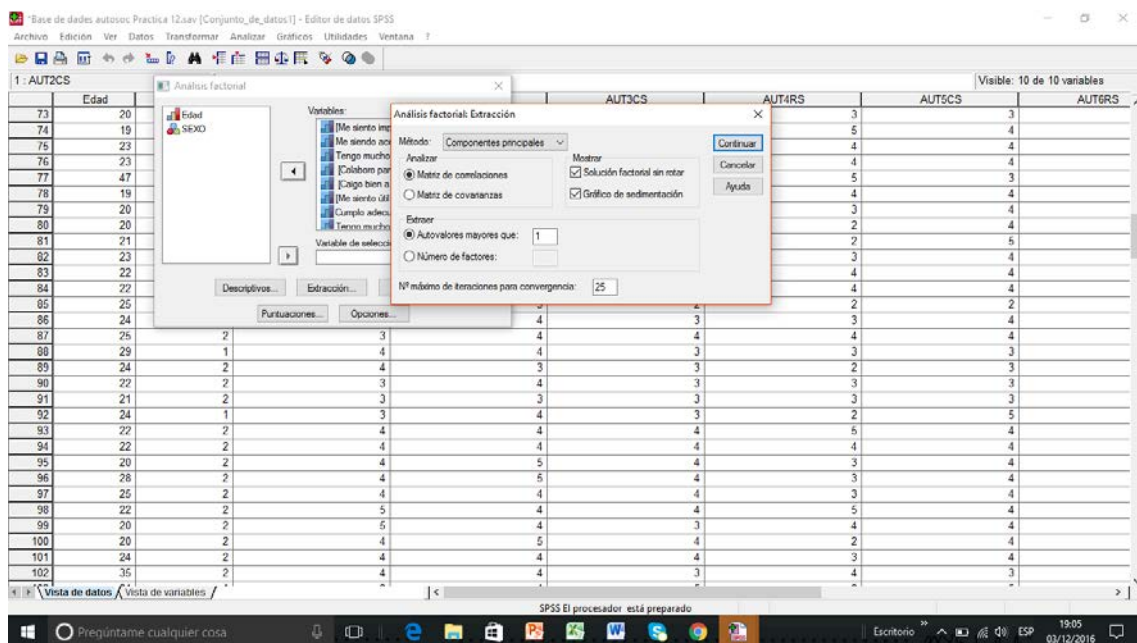
The screenshot shows the SPSS 'Análisis factorial...' dialog box. The 'Reducción de datos' menu is open, and 'Análisis factorial...' is selected. The background data table is as follows:

	Edad	SEXO	AUT2CS	AUT3CS	AUT4RS	AUT5CS	AUT6RS
73	20	:	3	3	3	3	3
74	19	:	4	3	5	4	4
75	23	:	4	4	4	4	4
76	23	:	4	4	4	4	4
77	47	:	2	3	5	3	3
78	19	:	4	3	4	4	4
79	20	:	4	4	3	4	4
80	20	:	4	3	2	4	4
81	21	:	5	4	2	5	5
82	23	:	4	4	3	4	4
83	22	:	4	3	4	4	4
84	22	:	4	3	4	4	4
85	25	:	2	2	2	2	2
86	24	:	4	3	3	4	4
87	25	:	4	4	4	4	4
88	29	:	4	3	3	3	3
89	24	:	3	3	2	3	3
90	22	:	4	3	3	3	3
91	21	:	3	3	3	3	3
92	24	:	4	3	2	5	5
93	22	:	4	4	5	4	4
94	22	2	4	4	4	4	4
95	20	2	4	5	4	3	4
96	28	2	4	5	4	3	4
97	25	2	4	4	4	3	4
98	22	2	5	4	4	5	4
99	20	2	5	4	3	4	4
100	20	2	4	5	4	2	4
101	24	2	4	4	4	3	4
102	35	2	4	4	3	4	3

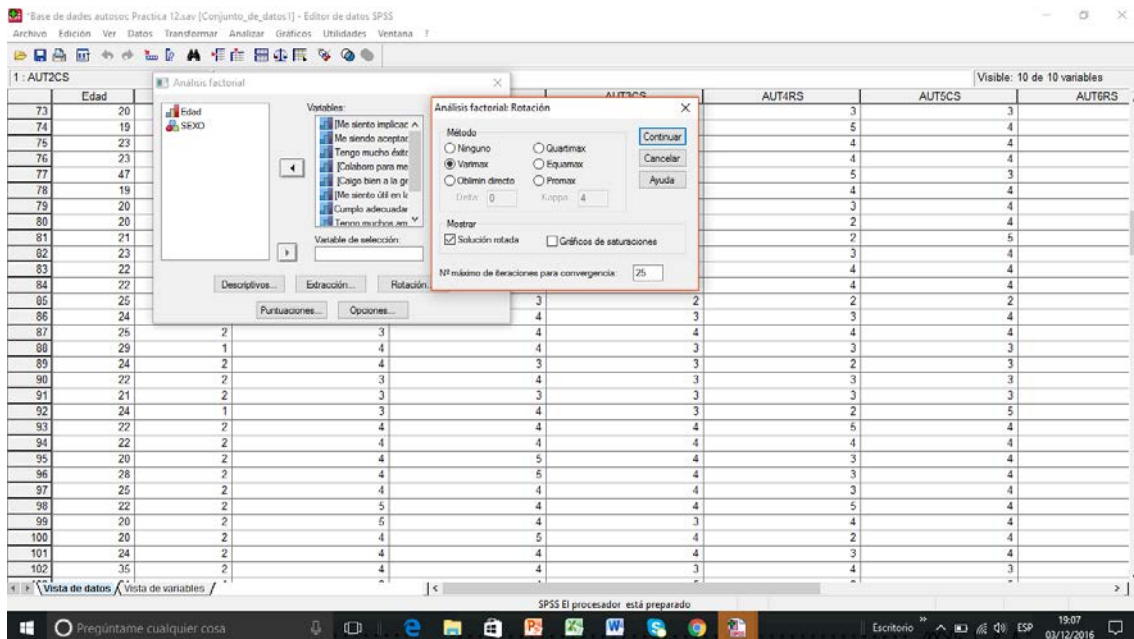
Incloem els huit ítems com a variables i demanem els estadístics d'adequació mostral KMO i prova d'esfericitat de Bartlett.



En Extracció demanem el mètode de components principals, analitzem la matriu de correlacions, demanem el gràfic de sedimentació i extraiem els autovalors més grans que 1, seguint la regla de Kaiser.



En la casella de Rotació, demanem mètode Varimax perquè pensem que els factors són independents.



Els resultats que obtenim de l'anàlisi factorial són:

KMO i prova de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin		,846
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	726,478
	gl	28
	Sig.	,000

KMO és superior a ,700 i Bartlett és significatiu, per tant, les dades d'adequació mostral són bones per a continuar.

La prova de significació de Bartlett confirma que la matriu de correlacions és significativament diferent de la matriu identitat.

Comunalidades

	Inicial	Extracció
[Me siento implicado/a con la sociedad]	1,000	,541
Me siendo aceptado/a cuando estoy en un grupo]	1,000	,708
Tengo mucho éxito en mis relaciones con la gente]	1,000	,669
[Colaboro para mejorar la sociedad]	1,000	,683
[Caigo bien a la gente]	1,000	,638
[Me siento útil en la sociedad]	1,000	,635
Cumplo adecuadamente las obligaciones que tengo como ciudadano]	1,000	,439
Tengo muchos amigos]	1,000	,604



Les comunalitats són totes superiors a ,200 i cap és molt diferent a la resta.

Mètode d'extracció: Anàlisi de components principals.

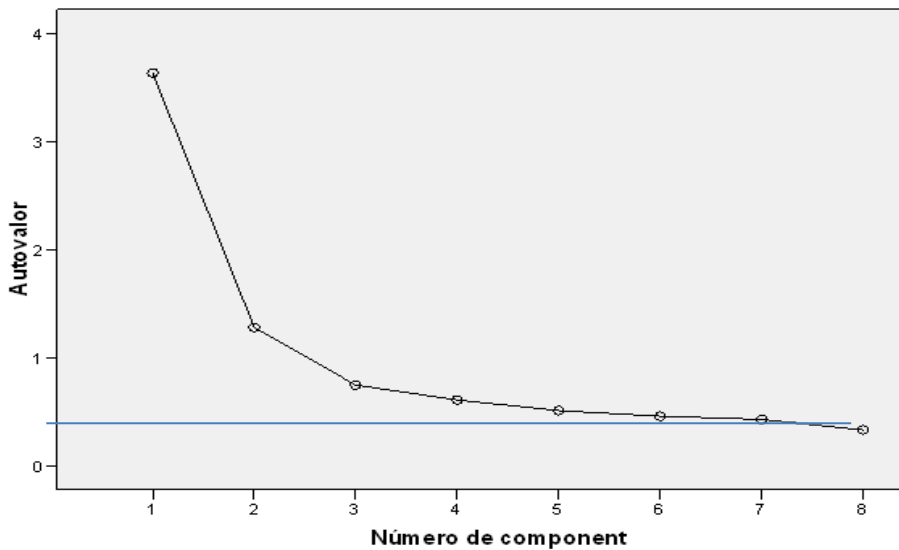
Clarament, els resultats mostren dos factors amb autovalors majors que 1 (criteri de Kaiser). Després de la rotació, el primer factor explica el 32.699% i el segon 28.786% (total = 61.4565).

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,636	45,453	45,453	3,636	45,453	45,453	2,614	32,669	32,669
2	1,280	16,003	61,456	1,280	16,003	61,456	2,303	28,786	61,456
3	,747	9,336	70,792						
4	,609	7,609	78,400						
5	,510	6,374	84,775						
6	,458	5,730	90,504						
7	,428	5,352	95,856						
8	,332	4,144	100,000						

Método de extracción: Análisis de Componentes principales.

Gràfic de sedimentació



El gràfic de sedimentació confirma l'extracció de dos factors: responsabilitat social (1, 4, 6 i 7) i acceptació social (2, 3, 5 i 8), com indiquen les saturacions factorials.

Matriz de componentes rotados(a)

	Componente	
	1	2
Me siento implicado/a con la sociedad	,190	,710
Me siendo aceptado/a cuando estoy en un grupo	,788	,294
Tengo mucho éxito en mis relaciones con la gente	,755	,315
Colaboro para mejorar la sociedad	,096	,821
Caigo bien a la gente	,797	,050
Me siento útil en la sociedad	,381	,699
Cumplo adecuadamente las obligaciones que tengo como ciudadano	,110	,654
Tengo muchos amigos	,765	,139

Les saturacions factorials són totes per damunt de ,300 i s'agrupen clarament en els dos factors pressuposats.

Mètode d'extracció: anàlisi de components principals.
Mètode de rotació: normalització Varimax amb Kaiser.

Per tant, el qüestionari AUSO en aquesta mostra presenta unes dades d'adequació mostrals bones, la variància explicada total és de 61,456% amb dos factors.

EXERCICIS

Ara pots practicar tu amb les dades del qüestionari [LOT-R](#). Es tracta d'un qüestionari de deu ítems, però quatre són únicament de control. Per tant, els sis ítems que han de analitzar-se són:

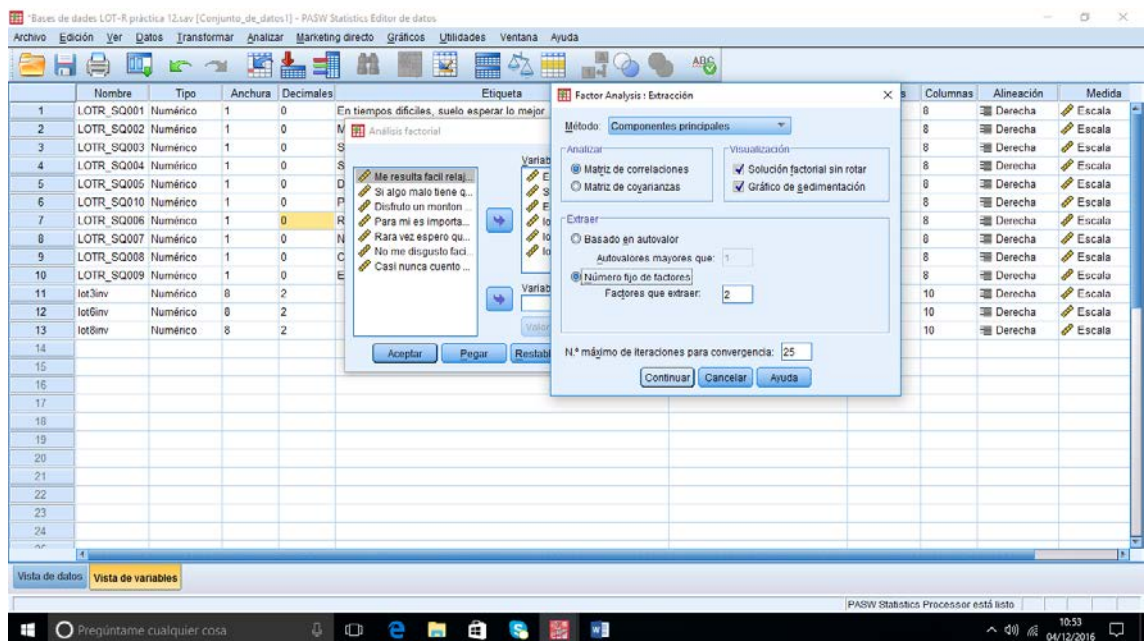
1. *En tiempos difíciles, suelo esperar lo mejor.*
3. *Si algo malo tiene que pasar, estoy seguro que me pasará.*
4. *Siempre soy optimista respecto al futuro*
6. *Rara vez espero que las cosas salgan a mi manera.*
8. *Casi nunca espero que me sucedan cosas buenas.*
9. *En general, espero que me ocurran más cosas buenas que malas.*

Es considera, en principi, que aquest instrument conté dos factors: optimisme (1, 4 i 9) i pessimisme (2, 6 i 8).

Ara respon a les preguntes següents:

1. S'ha d'invertir algun ítem abans de començar l'anàlisi factorial? Quins?
2. Les dades tenen bons estadístics d'adequació mostral? Comenta'ls.
3. Les comunalitats entre els ítems són bones?
4. Seguint la regla de Kaiser, s'extrauen els dos factors hipotetitzats? Què podem concloure de les dades de la mostra?
5. Què podem dir de les saturacions factorials? Podem extraure la matriu rotada?

Ara, prova de demanar l'extracció en funció dels dos factors que esperàvem. Per a fer-ho, observa aquesta pantalla. S'hi veu com assenjala: Extraure - Nombre fix de factors - Factors a extraure - 2.



6. Les dades d'adequació mostral són iguals? Per què penses que és així?

7. Com són ara els factors extrets? I les saturacions factorials?

8. Finalment, quina creus que és la solució factorial més adequada per a aquestes dades?

SOLUCIONS

1. Sí, han d'invertir-se els ítems 2, 6 i 8, ja que avaluen pessimisme.
2. Sí, ja que tant KMO és superior a ,700 com la prova de Bartlett és significativa.

KMO i prova de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,808
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	829,855
	gl	15
	Sig.	,000

3. Sí, són totes superiors a ,200 i prou similars entre elles.

Comunalidades

	Inicial	Extracción
En tiempos difíciles, suelo esperar lo mejor	1,000	,443
Siempre soy optimista en cuanto al futuro	1,000	,597
En general, espero que me ocurran más cosas buenas que malas	1,000	,596
Si algo malo tiene que pasar, estoy seguro que me pasará	1,000	,351
Rara vez espero que las cosas salgan a mi manera	1,000	,441
Casi nunca espero que me sucedan cosas buenas	1,000	,579

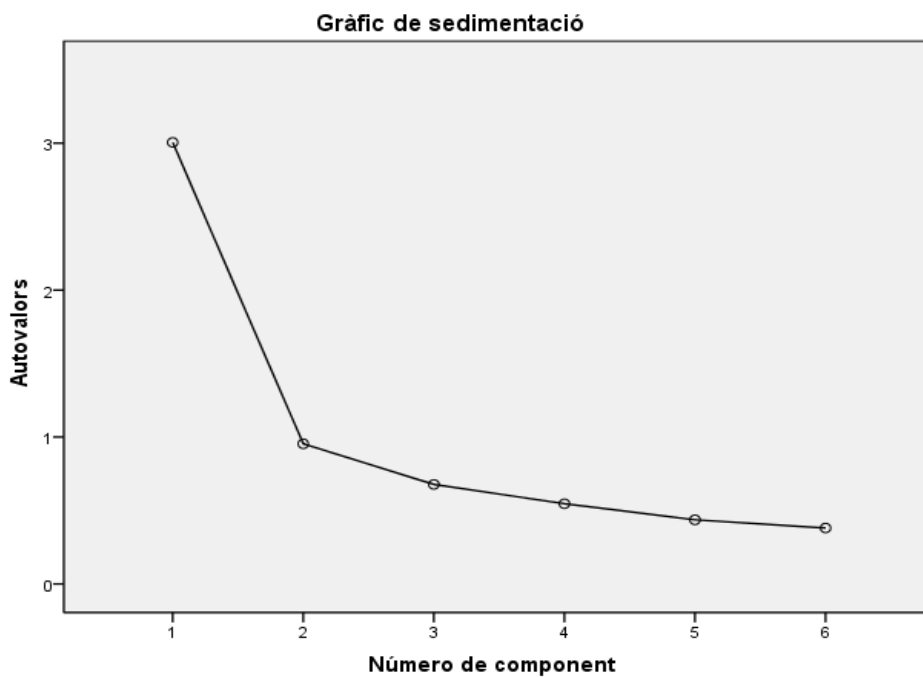
Mètode d'extracció: anàlisi de components principals.

4. Seguint la regla de Kaiser d'autovalors majors que 1, aquestes dades presenten un únic factor que explica el 50.085% de la variància. S'observa en els autovalors i en el gràfic de sedimentació que el segon factor està molt prop d'arribar al valor 1.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,005	50,085	50,085	3,005	50,085	50,085
2	,954	15,898	65,983			
3	,677	11,291	77,274			
4	,546	9,104	86,378			
5	,437	7,278	93,656			
6	,381	6,344	100,000			

Método de extracción: Análisis de Componentes principales.



5. Les saturacions factorials són totes superiors a ,300, però no podem extraure la matriu rotada, ja que únicament s'extrau un factor amb autovalor major que 1.

Matriz de componentes^a

	Componente
	1
En tiempos difíciles, suelo esperar lo mejor	,665
Siempre soy optimista en cuanto al futuro	,773
En general, espero que me ocurran más cosas buenas que malas	,772
Si algo malo tiene que pasar, estoy seguro que me pasará	,592
Rara vez espero que las cosas salgan a mi manera	,664
Casi nunca espero que me sucedan cosas buenas	,761

Método de extracción: Análisis de componentes principales.

1 componentes extraídos

6. Després de demanar extraure dos factors, els resultats de KMO i Bartlett són iguals, ja que estan basats en la matriu de correlacions i és igual. Les comunalitats sí que són diferents, ja que estan basades en la correlació de cada ítem amb el seu factor. S'hi pot veure que són més elevades que en l'extracció d'un únic factor.

Comunalidades

	Inicial	Extracción
En tiempos difíciles, suelo esperar lo mejor	1,000	,664
Siempre soy optimista en cuanto al futuro	1,000	,751
En general, espero que me ocurran más cosas buenas que malas	1,000	,642
Si algo malo tiene que pasar, estoy seguro que me pasará	1,000	,637
Rara vez espero que las cosas salgan a mi manera	1,000	,636
Casi nunca espero que me sucedan cosas buenas	1,000	,629

Método de extracción: Análisis de Componentes principales.

7. Es pot observar que, en aquest cas, s'obtenen dos factors que expliquen 65,983% de la variància i les saturacions factorials són totes adequades.

Variància total explicada

Component	Autovalors inicials			Sumes de les saturacions al quadrat de l'extracció			Sumes de les saturacions al quadrat de la rotació			
	Total	% de la variància	% acumulat	Total	% de la variància	% acumulat	Total	% de la variància	% acumulat	
dimension0	1	3,005	50,085	50,085	3,005	50,085	50,085	2,101	35,020	35,020
	2	,954	15,898	65,983	,954	15,898	65,983	1,858	30,963	65,983
	3	,677	11,291	77,274						
	4	,546	9,104	86,378						
	5	,437	7,278	93,656						
	6	,381	6,344	100,000						

Mètode d'extracció: anàlisi de components principals.

Matriz de componentes rotados^a

	Componente	
	1	2
En tiempos difíciles, suelo esperar lo mejor	,810	,090
Siempre soy optimista en cuanto al futuro	,838	,220
En general, espero que me ocurran más cosas buenas que malas	,720	,352
Si algo malo tiene que pasar, estoy seguro que me pasará	,088	,794
Rara vez espero que las cosas salgan a mi manera	,203	,772
Casi nunca espero que me sucedan cosas buenas	,420	,673

8. En observar les dades globals, podem concloure que triem l'estructura de dos factors pels motius següents:
- Per a continuar l'estructura de la proposta original de dos factors.
 - Les comunalitats són millors en la proposta de dos factors.
 - S'explica un 15% més de la variància amb un factor més únicament.
 - El segon factor presenta un autovalor de 0,954, molt pròxim al valor 1 de referència.

B3. Pràctica 3. Factors que afecten la validesa

Si considerem la validesa com la qualitat del test de mesurar el que vertaderament vol mesurar, no podem oblidar que, igual que la fiabilitat, és una qualitat que es veu afectada per diferents factors.

Recordem que podem calcular diferents fonts d'evidència de la validesa:

1. De contingut: demostra que els ítems del test són una mostra representativa del tret que vol mesurar-se.
2. De constructe: l'adequació de les inferències fetes a partir del test per a mesurar un constructe determinat.
3. Relativa a un criteri: grau en què el test és capaç de pronosticar un criteri.

En aquesta última evidència de validesa, s'utilitza el càlcul del **coeficient de validesa** o correlació entre test (X) i criteri (Y):

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{S_x S_y}$$

I, per tant, el coeficient de validesa està en funció de dues variables:

- Relació entre test i criteri
- La fiabilitat del test i la fiabilitat del criteri

Normalment, tant test com criteri tenen error de mesura, i per això disposem de la **fórmula d'atenuació** que estima el coeficient de validesa tenint en compte tant l'error de mesura del test i el criteri, com el grau de relació que hi ha entre ambdós.

Fórmula d'atenuació del coeficient de validesa

$$\rho_{xy} = \frac{r_{xy}}{\sqrt{r_{xx'}} \cdot \sqrt{r_{yy'}}$$

on:

r_{xy} el coeficient de validesa (correlació entre test y criteri)

$r_{xx'}$ la fiabilitat del test

$r_{yy'}$ la fiabilitat del criteri

A més dels errors de mesura, la validesa es pot veure influïda per la mateixa variable que influïa la fiabilitat: **la longitud del test**. Quan un instrument canvia de longitud podem calcular-ne la nova validesa.

Fórmula de nova validesa quan canvia de longitud

$$R_{xy} = \frac{r_{xy}}{\sqrt{\frac{1 - r_{xx'}}{k} + r_{xx'}}$$

on:

$$k = \frac{\text{longitud inicial}}{\text{longitud final}}$$

R_{xy} el nou coeficient de validesa

r_{xy} el coeficient de validesa anterior

$r_{xx'}$ el coeficient de fiabilitat del test

Una altra variable que pot influir en el coeficient de validesa és la variabilitat de la mostra. Si a partir de les puntuacions en el test, seleccionem una mostra amb unes característiques més homogènies, això influirà en la disminució de la variància del test i indirectament disminuirà el coeficient de validesa. Sempre és millor l'heterogeneïtat! Tenim un rang més ample de respostes!

Finalment, amb el coeficient de validesa podem estimar la puntuació en el criteri a partir de la puntuació obtinguda en el test. Tenim a l'abast dos tipus d'estimació: puntual i l'interval de confiança.

Càlcul del pronòstic lineal simple

$$y' = r_{xy} \frac{S_y}{S_x} (X - \bar{X}) + \bar{Y}$$

On:

r_{xy} el coeficient de validesa (correlació entre test y criteri)

S_y la d.t. del criteri

S_x la d.t. del test

X la puntuació en el test a partir de la qual es vol fer el pronòstic en el criteri (y')

\bar{X} la mitjana de las puntuacions en el test

\bar{Y} la mitjana de las puntuacions en el criteri

Càlcul del interval de confiança per a les puntuacions estimades

Passos:

1. Establim el nivell de confiança $(1-\alpha)$ desitjat: 0,95 o 0,99. Calculem l'interval en què es troba Y' amb una confiança de 99% o 95%.

2. Determinem la puntuació típica de l'error que delimita aquest interval:

$$- \alpha = 0,05 \quad z_{0,025} = 1,96$$

$$- \alpha = 0,01 \quad z_{0,005} = 2,58$$

3. Calculem l'error típic d'estimació:

$$S_e = S_{y.x} = S_y \sqrt{1 - r_{xy}^2}$$

4. Obtenim els límits de confiança:

$$y' \pm z_{\alpha/2} \cdot S_e$$

EXERCICIS

1. Apliquem un test d'intel·ligència a una mostra i obtenim un coeficient de fiabilitat de 0,81. Triem com a criteri per a obtenir evidència de validesa una prova de creativitat, i el coeficient de fiabilitat és 0,56. El coeficient de correlació entre ambdues proves és 0,20.
 - a) Quin serà el coeficient de validesa?
 - b) Calcula el coeficient de validesa atenuat.
 - c) Calcula el coeficient de validesa atenuat, si la fiabilitat del test fóra 0,95.
 - d) Calcula el coeficient de validesa atenuat, si la fiabilitat del criteri fóra 0,90.
 - e) Creus que té sentit que el coeficient de validesa atenuat siga menor quan la fiabilitat és major? Justifica la resposta.
 - f) Calcula el coeficient de validesa, si la fiabilitat del test i del criteri foren perfectes.

- g) A partir dels resultats, creus que s'ha fet una bona tria del criteri? Encara que el test i el criteri foren perfectament fiables, podríem fer bones prediccions de la creativitat a partir del test d'intel·ligència? Pareix adequat usar les puntuacions del nostre test per a predir la creativitat?
2. Continuant amb el mateix test d'intel·ligència amb una fiabilitat de 0,81, s'aplica a una mostra de estudiants de Primària. La variància del test és 8 i del criteri (prova de aptitud verbal) és 6. La mitjana de les puntuacions del test és 30 i la del criteri 40.
- a) Quina seria la puntuació que pronosticarem en la prova d'aptitud verbal per a un nen que ha obtingut en el test d'intel·ligència una puntuació de 35?
- b) Calcula l'interval de confiança amb 5% de risc sobre la puntuació estimada en el criteri.
- c) Com seria l'interval de confiança si fóra al 99%? Més ample o més estret? Justifica la resposta.
- d) Suposem que seleccionem els deu nens que presenten el valor més alt en el test d'intel·ligència i volem fer amb ells el pronòstic de la prova d'aptitud verbal. Per a aquest grup, com seria el coeficient de validesa? Major o menor? Raona la resposta.

SOLUCIONS

1.

a) El coeficient de validesa serà de 0,20, la correlació entre les dos proves.

b) Apliquem la fórmula:

$$\frac{0.20}{\sqrt{0.89}\sqrt{0.56}} = \frac{0.20}{0.94*0.75} = \frac{0.20}{0.705} = 0.28$$

c) Apliquem la fórmula canviant la dada de la fiabilitat del test:

$$\frac{0.20}{\sqrt{0.95}\sqrt{0.56}} = \frac{0.20}{0.97*0.75} = \frac{0.20}{0.73} = 0.27$$

d) Apliquem la fórmula canviant la dada de la fiabilitat del criteri:

$$\frac{0.20}{\sqrt{0.89}\sqrt{0.90}} = \frac{0.20}{0.94*0.95} = \frac{0.20}{0.89} = 0.22$$

e) Sí, té sentit, el coeficient de validesa atenuat serveix precisament per a corregir o atenuar els errors de mesura. Si les fiabilitats de test i criteri són més elevades, significa que hi ha menys quantitat d'error a atenuar.

f) Encara que podem resoldre la pregunta amb reflexió teòrica, apliquem la fórmula:

$$\frac{0.20}{\sqrt{1}\sqrt{1}} = \frac{0.20}{1*1} = \frac{0.20}{1} = 0,20$$

Si no hi ha error a atenuar perquè el test i el criteri tenen una fiabilitat. Per tant, no pot canviar el coeficient de validesa inicial.

g) Amb una correlació entre el test i el criteri de 0,20, no pot considerar-se la intel·ligència un bon predictor de la creativitat. Encara que els dos foren totalment precisos, no hi ha relació entre ells.

2.

a) Apliquem la fórmula. Per això necessitem les desviacions típiques de test i criteri i el problema em dona les variàncies.

$$DT \text{ test: } \sqrt{8} = 2,83$$

$$DT \text{ criteri: } \sqrt{6} = 2,45$$

$$Y' = 0,20 \frac{2,45}{2,83} (30 - 35) + 40 = 0,20 * 0,86 * 5 + 40 = 40,86$$

Un nen ha obtingut una puntuació de 35 en la prova d'intel·ligència, li pronosticarem una puntuació de 40,86 en la prova d'aptitud verbal.

b) Calculem primer l'error típic:

$$S_y * \sqrt{1 - r_{xy}^2} = 6 * \sqrt{1 - 0.81^2} = 3.52$$

I coneixem que $\alpha = 0,05$ $z_{0,025} = 1,96$

Calculem intervals de confiança:

$$\text{Límit inferior: } 40.86 - 1,96 * 3,52 = 33,96$$

Límit inferior: $40.86 + 1,96 * 3,52 = 47,76$

- c) El càlcul d'un interval de confiança més conservador, com el de 99%, suposa que l'interval és més ample perquè ha d'incloure més possibilitats. Seria més difícil fer una estimació errada, però no et dóna una estimació tan precisa.
- d) Si seleccionem una mostra que presenta menor variabilitat, el coeficient de validesa serà menor, perquè en ser una mostra més homogènia hi ha més probabilitat d'error.

Bloc 4: Teoria de Resposta a l'Ítem (TRI)

B4. Pràctica 1. Paràmetres de la TRI

Les teories dels tests més importants en el desenvolupament de la psicometria han sigut la teoria clàssica del test i la teoria de resposta a l'ítem. Aquesta última teoria va nàixer com a conseqüència de les crítiques cap a la TCT, algunes de les quals eren:

- a. El resultat de la mesura depèn de l'instrument: mesure el mateix tret amb dos instruments diferents i puc obtenir diferents resultats.
- b. La fiabilitat i la validesa d'un test ho són per a la mostra, no son inherents al test.
- c. La TCT demana que es compleix el supòsit d'homoscedasticitat de l'error, és a dir, la variància del errors és constant en totes les observacions.

La teoria de resposta a l'ítem és «un conjunt de models matemàtics que intenten definir, a partir d'una funció matemàtica (corba característica de l'ítem), la probabilitat que una persona done una resposta específica, considerant el nivell de tret i les propietats de la tasca».

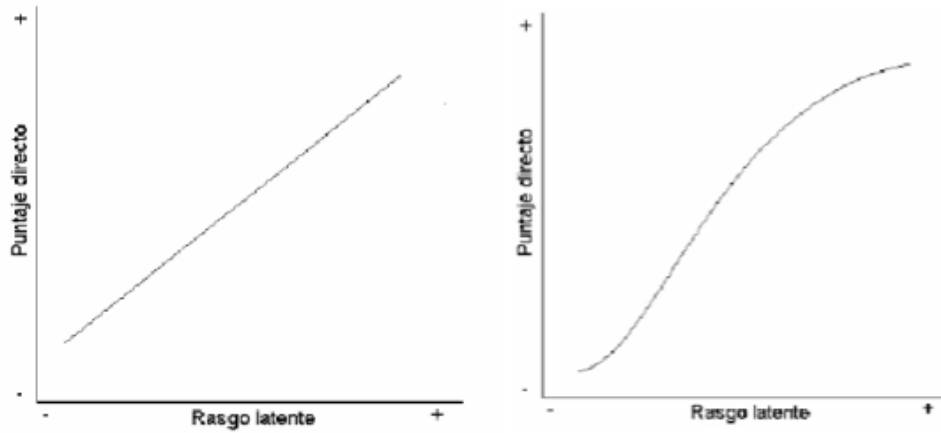
Podríem resumir que les tres característiques més importants de la TRI són:

- La unitat d'anàlisi és l'ítem, no la puntuació en el test.
- Les relacions entre el tret i la puntuació directa no tenen un comportament lineal, sinó que es modela una probabilitat.
- Les característiques de l'ítem i les característiques de les persones es col·loquen en la mateixa recta amb la mateixa unitat de mesura: el continu del tret.

En la figura es pot observar la diferència en el comportament de les respostes. Mentre que en la TCT o en la teoria de la generalitzabilitat, la relació és lineal, en la TRI la corba representa una probabilitat.

TCT y TG

TRI



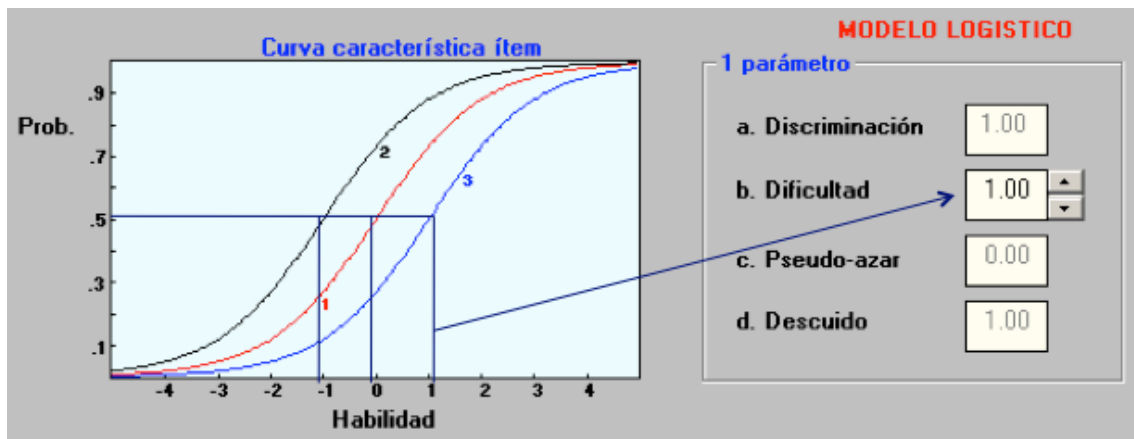
Considerant els ítems com d'encert o errada, descriurem diferents models en funció de la quantitat de paràmetres que consideren:

1. Model d'un paràmetre: model de Rasch. Considera únicament la dificultat de l'ítem.
2. Model de dos paràmetres: considera la dificultat de l'ítem i la seua discriminació.
3. Model de tres paràmetres: considera dificultat, discriminació i diviniació.

Paràmetres

Dificultat (b)

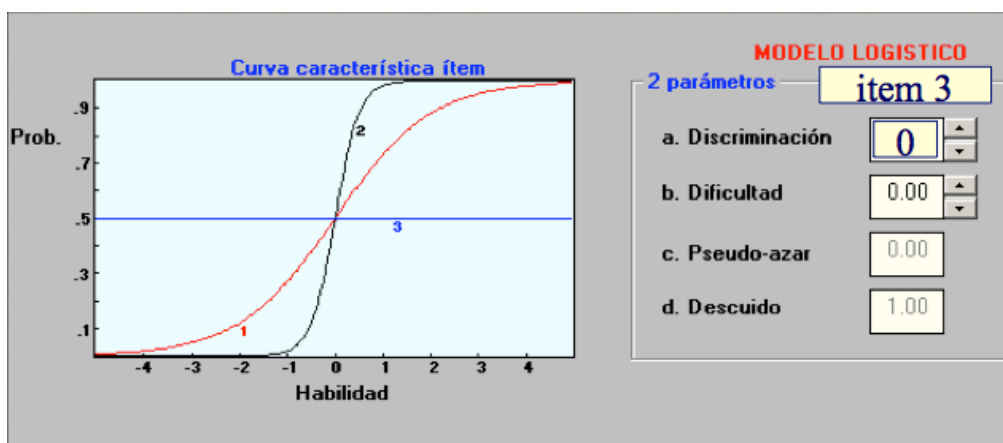
- Indica la quantitat de tret requerida per a encertar l'ítem.
- El valor està determinat pel punt de màxim pendent en la corba característica de l'ítem.
- És un índex de posició de l'ítem respecte del tret.
- Pot anar des de $-\infty < b < \infty$



Es pot observar que l'ítem 2 demana menys habilitat per a obtenir una probabilitat de 0,5 d'encertar-lo, i per tant, és el més fàcil. Per contra, l'ítem 3 és el més difícil dels tres presentats.

Discriminació (a)

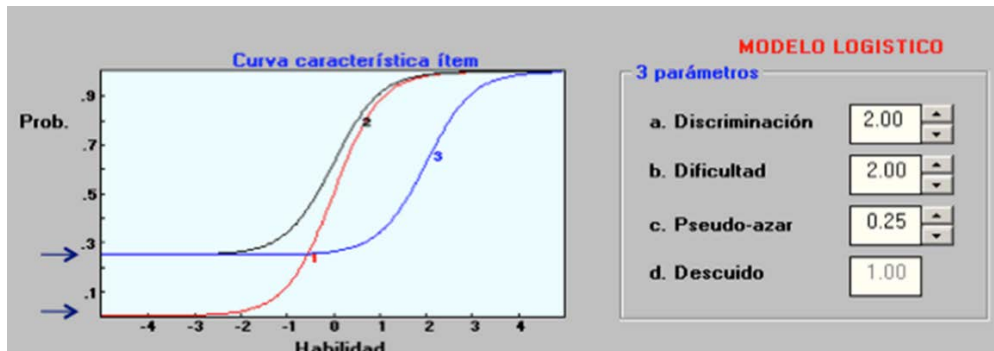
- Reflecteix la taxa de canvi en la probabilitat d'encert en funció de l'augment del nivell de tret.
- Assenyala quanta sensibilitat té l'ítem.
- El valor és proporcional al pendent de la recta tangent a la corba característica d'ítem en el màxim pendent.
- Normalment des de $-4 < a < 4$



En aquest cas, l'ítem 3 no discrimina res, té igual el nivell d'habilitat que presente la persona, la probabilitat d'encertar-lo no canvia. Per contra, l'ítem 2 és el més discriminador, encara únicament ho és per a nivells de tret entre -1 i 1.

Divinació (c)

- Únicament s'utilitza en ítem d'elecció múltiple.
- Indica la probabilitat d'encertar quan el nivell de tret és molt baix i suposadament es contesta per atzar.
- El valor no varia en funció del tret.
- El rang teòric està entre 0 i 1 (1 dividit les alternatives de resposta).



Com es pot observar, l'ítem 1 no presenta cap probabilitat d'encertar per atzar, ja que la corba comença directament en una probabilitat nul·la d'encert. Per contra, els ítems 2 i 3 presenten un valor de 0,25 de divinació i, per tant, una probabilitat entre 4 de ser encertat per atzar. És el cas dels ítems amb quatre alternatives de resposta.

EXERCICIS

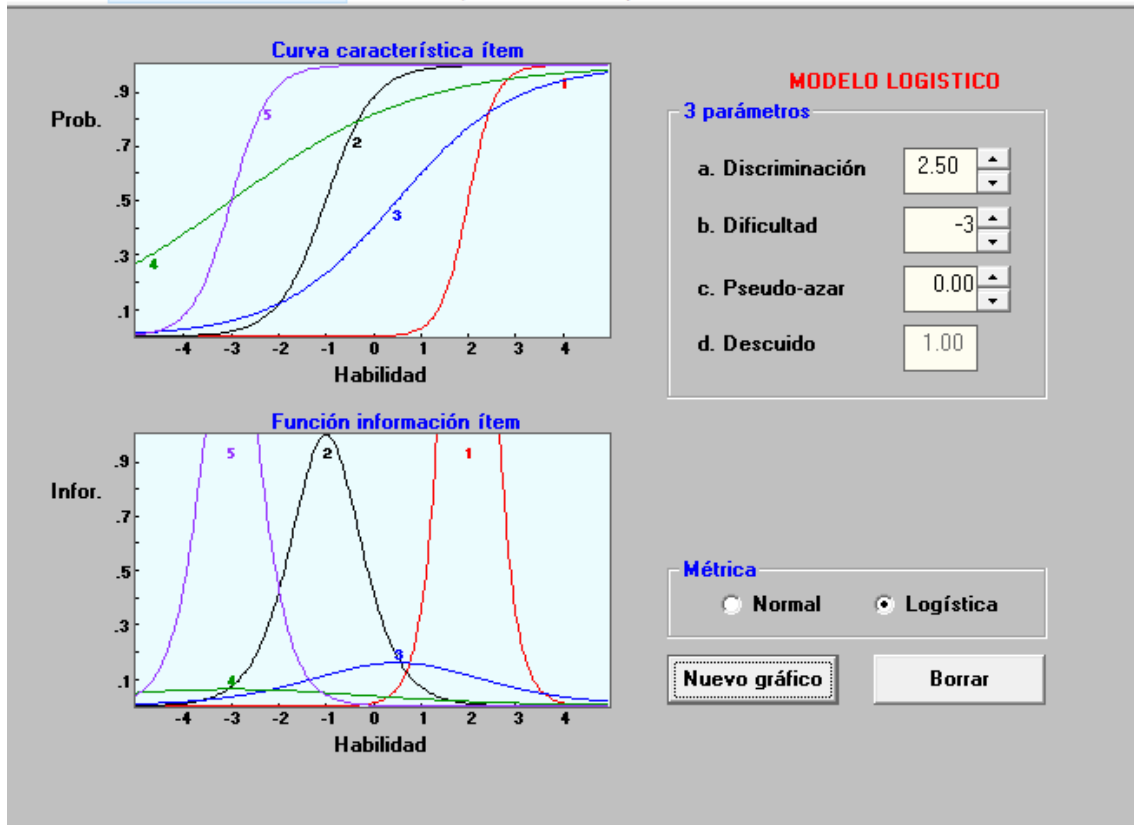
Per a fer el exercicis següents, treballarem amb el programa dedicat a la docència ANATRI. En la taula següent es presenten les dades de cinc ítems, representa les corbes i respon a les preguntes següents:

ITEM	a	b	c
1	0,500	2,451	0,250
2	0,875	2,834	0,146
3	1,250	0,318	0,118
4	1,625	-0,245	0,125
5	2,000	1,150	0,001
6	2,000	-2,725	0,154

1. Quin model ha d'utilitzar-se per a representar les dades anteriors?

2. L'ítem 5, quina probabilitat té d'encertar-se per atzar? Quina és la probabilitat que una persona amb un nivell de tret d'1,150 encerte aquest ítem?
3. Quina és la probabilitat d'encertar per atzar l'ítem 1? Quina és la probabilitat aproximada que una persona amb un nivell de tret de 2,651 encerte aquest ítem?
4. Si comparem els resultats dels ítems anteriors, per què creus que les probabilitats d'encert dels ítems 5 i 1 són diferents?
5. Per a l'ítem 6, quina és la probabilitat aproximada que una persona amb -3,5 de tret encerte l'ítem? És fàcil per a aquesta persona? I en el cas que la persona presente un nivell de tret de 2,5?
6. Quin són els ítems més discriminadors? Què observem en les corbes d'aquests ítems?
7. Fixant-nos en la funció d'informació dels ítems, quin interval de tret pareix que està mesurat amb menor precisió per aquests sis ítems?

Ara, fixa't en les corbes següents per a respondre a les preguntes:

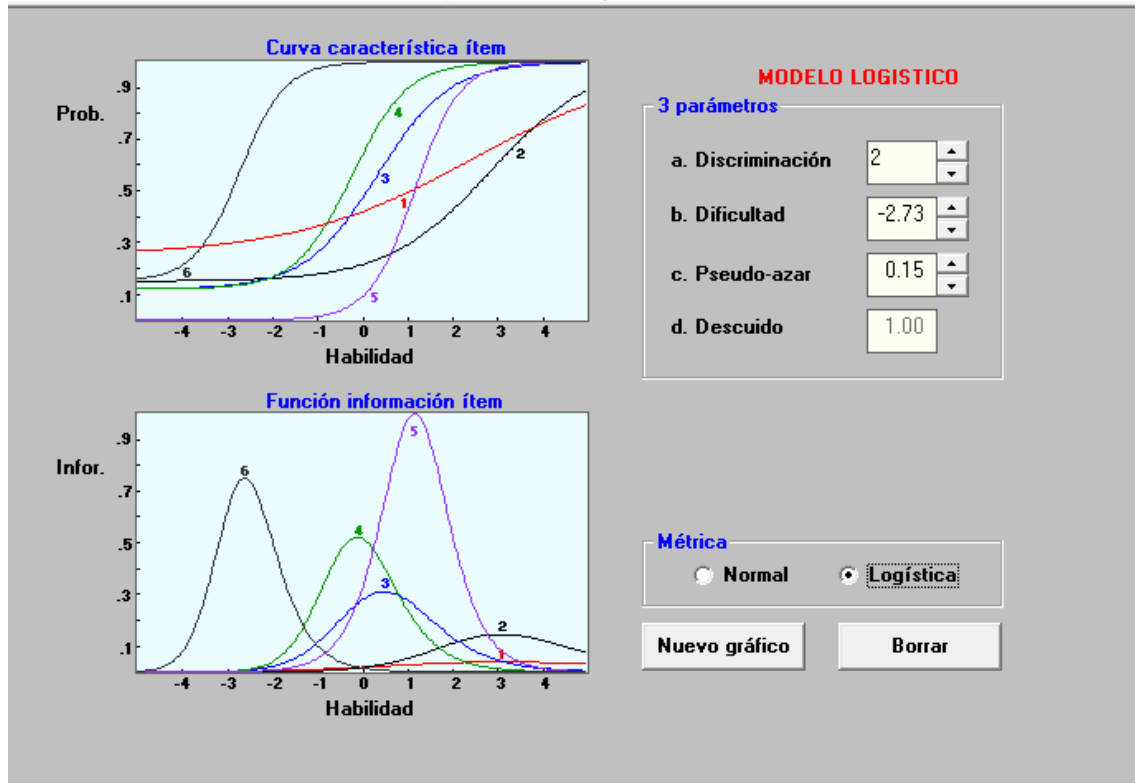


8. Quin és l'ítem menys discriminador? És vàlid aquest ítem per a una persona amb alta habilitat?
9. Quin és l'ítem més difícil? Per a quins nivells de tret es pot considerar que és discriminador aquest ítem?
10. A quin ítem creus que corresponen les dades situades a la dreta?

SOLUCIONS

teoria de la resposta al ítem - [Curva característica y función información de ítems]

Archivo Modelos dicotómicos Modelos politómicos Opciones Tablas datos



1. El model de tres paràmetres, perquè recull informació de dificultat, discriminació i diviniació.
2. La probabilitat d'encertar per atzar l'ítem 5 és quasi nul·la, 0,001. La probabilitat que una persona amb 1,150 encerte aquest ítem és de 0,50, ja que iguala la dificultat de l'ítem.
3. La probabilitat d'encertar per atzar l'ítem 1 és 0,250, és a dir, una probabilitat entre quatre. En aquest cas, si la persona té el mateix nivell de tret que la dificultat de l'ítem (2.451), la probabilitat d'encertar l'ítem serà major que 0,50, ja que és possible encertar-la per atzar.
4. La probabilitat d'encertar és diferent en els dos ítems perquè intervé el paràmetre de l'atzar.
5. La probabilitat que una persona amb nivell de tret -3,5 encerte l'ítem 6 és aproximadament de 30%. Aquest ítem és moderadament difícil per a aquesta persona. En el cas que la persona presente 2,5 de tret, l'ítem serà molt fàcil, i la probabilitat d'encertar-lo superarà el 90%.
6. Els ítems més discriminadors són el 5, el 4 i el 6. Si ens fixem en el gràfic de funció d'informació, es pot observar que són les corbes més apuntades.

7. Si observem el gràfic de funció d'informació, pareix que els nivells elevats de tret no tenen ítem prou discriminador, ja que totes les corbes es presenten en nivell baixos i intermedis de l'escala.
8. L'ítem 4 és el menys discriminador perquè és el més proper a l'horitzontalitat. En el gràfic de funció d'informació es veu que no presenta gens d'apuntament.
9. L'ítem més difícil és l'ítem 1. Si ens fixem en el primer gràfic és la corba més a la dreta i en el segon gràfic podem observar que aquest ítem és discriminador únicament per a nivells elevats de tret.
10. L'ítem 5 perquè presenta una discriminació alta (2,5), molt poca dificultat i nul·la diviniació.

B4. Pràctica 2. Models de la TRI i càlculs de probabilitats

La teoria de resposta a l'ítem presenta tres models diferents amb unes característiques compartides:

1. Model d'un paràmetre: model de Rasch. Considera únicament la dificultat de l'ítem.
2. Model de dos paràmetres: considera la dificultat de l'ítem i la discriminació.
3. Model de tres paràmetres: considera dificultat, discriminació i diviniació.

Les seues característiques en comú són:

- Unidimensionalitat de l'espai latent
- Independència local
- Per a ítems dicotòmics
- Expliquen la probabilitat d'encert en funció de la relació entre la quantitat de tret de la persona i la que demana l'ítem.
- La representació gràfica d'aquesta funció és la CCI: corba característica de l'ítem.

L'equació que descriu la funció logística és:

$$y = \frac{e^x}{1 + e^x}$$

- És la funció més simple que creix de 0 a 1 (igual que la probabilitat) a mesura que x creix de $-\infty$ a $+\infty$
- En tots els models la x conté la diferència entre la quantitat de tret de la persona (θ), i la quantitat que demana l'ítem (b), és a dir, ($\theta-b$).

$y = \frac{e^x}{1 + e^x}$	x ($\theta-b$)	-3	-2	2	3
	y (Pi)	0,047	0,119	0,881	0,953

Fixeu-vos, com més difícil és l'ítem per a la persona, menor és x i, per tant, menor és la probabilitat d'encertar-ho (y). Per contra, com més fàcil és l'ítem per a la persona, major és x i major la probabilitat d'encertar-ho.

Model de Rasch:

Únicament considera el paràmetre de dificultat

$$P(X=1|\theta) = \frac{e^{(\theta-b)}}{1+e^{(\theta-b)}} = \frac{1}{1+e^{-(\theta-b)}}$$

Sent:

e = sempre constant 2,718

b = dificultat de l'ítem

θ = tret

Exemple:

- Quina és la probabilitat que les persones amb nivells de tret 2 encerten un ítem a dificultat 1?

$$P(X=1|2) = \frac{1}{1+e^{-(\theta-b)}} = \frac{1}{1+e^{-(2-1)}} = 0,73$$

Model de dos paràmetres

Considera la dificultat i discriminació

$$P(X = 1 | \theta) = \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}} = \frac{1}{1 + e^{-Da(\theta-b)}}$$

Sent:

D = sempre constant 1,7

b = dificultat de l'ítem

θ = tret

a = discriminació

Exemple:

- Quina és la probabilitat que les persones amb un nivell de tret d'1,5 encerten un ítem amb dificultat 1,2 i discriminació 2,5?

$$P(X = 1 | 1,5) = \frac{1}{1 + e^{-Da(\theta-b)}} = \frac{1}{1 + e^{-1,7*2,5(1,5-1,2)}} = 0,7$$

Model de tres paràmetres

Considera la dificultat de l'ítem, la discriminació i la probabilitat d'encertar-ho per atzar.

$$P(X = 1 | \theta) = c + (1 - c) \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}} = c + (1 - c) \frac{1}{1 + e^{-Da(\theta-b)}}$$

Sent:

D = sempre constant 1,7

b = dificultat de l'ítem

θ = tret

a = discriminació

c = atzar

Exemple:

- Quina és la probabilitat que les persones amb un nivell de tret d'1,5 encerten un ítem amb dificultat 1,2, discriminació 2,5 i possibilitat d'encertar per atzar de 0,15?

$$P(X = 1 | 1,5) = c + (1 - c) \frac{1}{1 + e^{-Da(\theta-b)}}$$

$$= 0,15 + (1 - 0,15) * 0,7 = 0,745$$

EXERCICIS

1. Utilitzant el model d'un paràmetre de Rasch, calcula quina és la probabilitat d'encert d'un ítem amb dificultat 1,5, si la persona té un nivell de tret d'1 i de 2.
2. Si coneixem que aquest ítem presenta una discriminació d'1,3:
 - a) Quin model hauríem d'utilitzar?
 - b) Quina probabilitat d'encert tindrien ara les persones amb nivell de tret 1 i 2?
3. En un instrument d'elecció múltiple amb quatre alternatives de resposta, la dificultat d'un ítem és 1,4 i la discriminació d'1,7. Si una persona presenta nivell de tret de 2, quina probabilitat d'encert té?

SOLUCIONS

1. La probabilitat d'encert per a la persona que té un tret = 1 és de 0,38; d'altra banda, la persona que té 2 de tret té una probabilitat d'encert de 0,62.
2.
 - a) S'ha d'utilitzar el model de dos paràmetres perquè hi ha una variable més: la discriminació.
 - b) Per a una persona amb 1 de tret, la probabilitat d'encert és únicament de 0,25. Per a una persona amb 2 de tret la probabilitat d'encert és de 0,75.
3. La probabilitat d'encert per a una persona amb 2 de tret, serà de 0,87.