

Activitats de Psicometria

Primer quadrimestre

Laura Galiana

Curs 2025-2026

Índex de continguts

Índex de continguts	2
Tema 1. Introducció a la psicometria	4
1.1. Un segle usant tests	4
Un segle usant tests. Solucions.....	5
Tema 2. Fonaments de l'escalament per a construir escales i tests.....	8
2.1. Preguntes d'autoavaluació.....	8
Preguntes d'autoavaluació. Solucions	10
Tema 3. Mètodes d'escalament psicològic	11
3.1. L'escalament de Thurstone	11
L'escalament de Thurstone. Solucions	12
3.2. Les escales de Likert. Part 1	16
Les escales de Likert. Part 1. Solucions.....	18
3.3. Les escales de Likert. Part 2	24
Les escales de Likert. Part 2. Solucions.....	26
3.4. L'escalograma de Guttman	33
L'escalograma de Guttman. Solucions.....	34
Tema 4. Construcció de tests	39
4.1. Construcció d'ítems.....	39
4.2. Anàlisi d'ítems.....	40
Anàlisi d'ítems. Tutorial amb exemple de l'ítem 1	42

Anàlisi d'ítems. Solucions	58
Tema 5. Teoria clàssica de tests: fonaments del model clàssic lineal	111
5.1. Concepte de fiabilitat.....	111
Concepte de fiabilitat. Solucions	112
Tema 6. Estimació de la fiabilitat	115
6.1. Procediments d'estimació de la fiabilitat i factors que l'afecten	115
Procediments d'estimació de la fiabilitat i factors que l'afecten. Solucions	119
6.2. Variants d'alfa i estimació de puntuacions vertaderes i contrastos de puntuacions observades.....	126
Variants d'alfa i estimació de puntuacions vertaderes i contrastos de puntuacions observades. Solucions.....	128
6.3. Estimació de la fiabilitat en Jamovi	133
Estimació de la fiabilitat en Jamovi. Tutorial amb exemple de la dimensió <i>Reconeixement del patiment</i>	135
Estimació de la fiabilitat en Jamovi. Solucions	139
Tema 7. Teoria de la resposta a l'ítem.....	144
7.1. Corba característica de l'ítem (CCI) i models de la TRI	144
7.1. Corba característica de l'ítem (CCI) i models. Solucions.....	146
7.2. Qüestions generals sobre la TRI.....	150
Qüestions generals sobre la TRI. Solucions	153

Tema 1. Introducció a la psicometria

1.1. Un segle usant tests

A continuació tens unes quantes preguntes sobre l'article "Un siglo utilizando tests"

(<http://www.unioviedo.es/reunido/index.php/rema/article/view/9743/9489>).

Llegeix l'article i respon les preguntes següents.

1. Quan va ser l'origen dels tests?
2. En quin any es van aplicar tests de manera col·lectiva i amb quina finalitat? Quins tests eren i quines característiques tenia cadascun?
3. Quan es va produir la generalització dels tests? En quins sectors? Quina utilitat tenien en cada sector?
4. Quan comencen les crítiques als tests? Diques cinc crítiques que apareixen al text.
5. Quines són les mesures que s'han pres per a regular l'ús de tests?
6. Quina és la situació actual pel que fa a l'ús de tests?

Un segle usant tests. Solucions

1. Quan va ser l'origen dels tests?

Cattell, en 1890, va ser el primer que va emprar el terme *test mental*.

El primer test pròpiament d'intel·ligència el van crear a principi del segle XX els psicòlegs francesos Binet i Simon per a diagnosticar infants amb discapacitat intel·lectual a les escoles públiques de París.

2. En quin any es van aplicar tests de manera col·lectiva i amb quina finalitat? Quins tests eren i quines característiques tenia cadascun?

Durant la Primera Guerra Mundial (1914-1918) es van aplicar tests de manera col·lectiva per seleccionar i classificar reclutes.

Es tractava de dos tests molt senzills, d'aplicació col·lectiva i automàtica. El *Test alfa* era una prova verbal, adequada a subjectes amb capacitat lectora, i el *Test beta* era una prova no verbal, apropiada a subjectes amb problemes d'alfabetització.

3. Quan es va produir la generalització dels tests? En quins sectors? Quina utilitat tenien en cada sector?

La generalització de l'ús de tests es produeix després de la Primera Guerra Mundial.

Hi hagué una ràpida implantació de tests a l'escola, a la indústria i en l'àmbit clínic.

En el sector educatiu es van emprar tests no sols per al diagnòstic d'infants amb necessitats educatives especials, sinó també per a avaluar infants sense necessitats educatives especials. En 1920 apareix la primera bateria estandarditzada de rendiment per a avaluar diverses àrees acadèmiques. A més, els tests van cobrir les necessitats de selecció i d'orientació individual que van aparèixer a conseqüència de la ràpida expansió de l'oferta educativa. També s'empraren per a monitoritzar el progrés dels estudiants mitjançant l'avaluació i, així, proporcionaven un control de l'accés a programes amb places limitades i eines per a certificar el nivell d'aprofitament i d'adquisició de coneixements de cada alumne.

En la indústria i en el món empresarial en general, l'avaluació dels subjectes i de les diferències individuals va proporcionar un mitjà adequat per a seleccionar i classificar les persones més adequades per al lloc de treball. En aquest àmbit, els tests es van usar fonamentalment en la classificació, selecció i promoció laboral dels treballadors.

En l'àmbit clínic es van començar a usar amb finalitats diagnòstiques, però també per a predir la conducta del subjecte i la resposta al tractament o la teràpia, en el marc de la psicologia de la salut. Posteriorment, l'avaluació clínica s'ha emprat a l'hora de decidir sobre la veracitat d'un testimoni, per a determinar la custòdia legal dels fills en processos de separació i divorci, per a dictar resolucions temporals de llibertat condicional o sentències fermes i, també, per a dilucidar sobre la pertinència d'una al·legació de pertorbació mental transitòria en el moment de cometre el delicte jutjat.

4. Quan comencen les crítiques als tests? Digues cinc crítiques que apareixen al text.

Les crítiques als tests comencen a final de la dècada dels vint, especialment en l'àmbit de la selecció i provenen de les persones que queden fora del procés, és a dir, de les no seleccionades.

La crítica que els fan és que no són prou vàlids ni fiables per a justificar l'ús que se'n fa.

També s'afirma que no són útils per a fer prediccions significatives i a llarg termini.

A més, es qüestiona el contingut i el format dels ítems (especialment els d'opció múltiple), com també les normes per a la interpretació de les puntuacions.

També s'assenyalen els efectes secundaris que els tests han tingut en el procés d'ensenyament i aprenentatge dels escolars.

Finalment, es critica que siguin eines al servei del poder, instruments de control o repressió social al servei de la classe econòmica i políticament dominant i que al·len barreres contra la igualtat social i d'oportunitats econòmiques. Per exemple, han servit per a discriminar grups socials que tendeixen a puntuar com a grup per davall de la mitjana d'altres grups socialment més afavorits.

5. Quines són les mesures que s'han pres per a regular l'ús dels tests?

Per a fer front a les diverses crítiques s'han proposat mesures des de la indústria dels tests, des de les organitzacions, des dels col·legis professionals i des de les administracions públiques.

Per exemple, el Consell General de la Psicologia (abans Col·legi Oficial de Psicòlegs) ha emprès diverses activitats per a millorar l'ús i la pràctica de tests a Espanya. A escala nacional, es va crear en el seu moment una comissió de tests constituïda per universitaris, professionals i cases editores amb la finalitat de discutir i proposar mesures tendents a millorar l'ús dels tests a Espanya. A escala internacional, el Consell General de la Psicologia participa en la International Test Commission, organisme encarregat de potenciar un ús adequat dels tests a escala global. Aquest organisme ha elaborat diverses normes internacionals per a l'ús de tests. L'última edició és del 2017.

També hi ha una comissió europea per a l'estudi i la millora de la construcció i l'ús de tests a Europa.

6. Quina és la situació actual pel que fa a l'ús de tests?

Els tests constitueixen la tecnologia amb més repercussió social que ha produït la psicologia i es continuen emprant en multitud d'àmbits.

Actualment, el desafiament dels tests és respondre a les demandes creixents, com més va més exigents i diversificades, que es generen en la societat i satisfer la necessitat de més claredat i transparència al llarg de totes les fases implicades en l'ús dels tests, des de la construcció fins al moment en què es comunica a les persones la puntuació obtinguda.

A més, també cal investigar a fi de crear ítems menys restrictius i noves formes de tests, tant pel que fa al contingut com al format i mode d'administració.

Tema 2. Fonaments de l'escalament per a construir escales i tests

2.1. Preguntes d'autoavaluació

Respon a les preguntes següents. Només una alternativa és correcta.

- 1. La part de la psicometria que s'encarrega de desenvolupar regles sistemàtiques que permeten obtenir unitats de mesura amb significat per a quantificar les observacions empíriques es coneix com a:**
 - a) Teoria del mesurament.
 - b) Escalament.
 - c) Teoria de tests.
- 2. Un dels autors que va iniciar els mètodes d'escalament psicològic es considera que fou:**
 - a) Erns Heinrich Weber.
 - b) Gustav Theodor Fechner.
 - c) Louis Leon Thurstone.
- 3. Els mètodes d'escalament que tenen com a objectiu assignar valors escalars a les persones, a partir de les respostes o manifestacions sobre el grau en què representen la seua pròpia posició respecte de l'atribut psicològic que es mesura davant de certs estímuls o ítems, es coneixen com a:**

- a) Mètodes d'escalament centrats en els estímuls.
 - b) Mètodes d'escalament centrats en els subjectes.
 - c) Mètodes d'escalament centrats en les respostes.
- 4. Els mètodes d'escalament en què les diferències observades en les respostes dels subjectes se suposa que es deuen tant a les diferències que hi ha entre els estímuls respecte al grau amb què presenten determinat atribut psicològic, com també a les diferències individuals que les mateixes persones presenten en la magnitud en l'atribut es coneixen com a:**
- a) Mètodes d'escalament centrats en els estímuls.
 - b) Mètodes d'escalament centrats en els subjectes.
 - c) Mètodes d'escalament centrats en les respostes.

Preguntes d'autoavaluació. Solucions

- 1. La part de la psicometria que s'encarrega de desenvolupar regles sistemàtiques que permeten obtenir unitats de mesura amb significat per a quantificar les observacions empíriques es coneix com a:**
 - a) Teoria del mesurament.
 - b) *Escalament.*
 - c) Teoria de tests.

- 2. Un dels autors que va iniciar els mètodes d'escalament psicològic es considera que fou:**
 - a) Erns Heinrich Weber.
 - b) Gustav Theodor Fechner.
 - c) *Louis Leon Thurstone.*

- 3. Els mètodes d'escalament que tenen com a objectiu assignar valors escalars a les persones, a partir de les respostes o manifestacions sobre el grau en què representen la seua pròpia posició respecte de l'atribut psicològic que es mesura davant de certs estímuls o ítems, es coneixen com a:**
 - a) Mètodes d'escalament centrats en els estímuls.
 - b) *Mètodes d'escalament centrats en els subjectes.*
 - c) Mètodes d'escalament centrats en les respostes.

- 4. Els mètodes d'escalament en què les diferències observades en les respostes dels subjectes se suposa que es deuen tant a les diferències que hi ha entre els estímuls respecte al grau amb què presenten un determinat atribut psicològic, com també a les diferències individuals que les mateixes persones presenten en la magnitud en l'atribut es coneixen com a:**
 - a) Mètodes d'escalament centrats en els estímuls.
 - b) Mètodes d'escalament centrats en els subjectes.
 - c) *Mètodes d'escalament centrats en les respostes.*

Tema 3. Mètodes d'escalament psicològic

3.1. L'escalament de Thurstone

En una classe s'obtenen les dades següents sobre quins són els presentadors de televisió preferits:

Taula 1

Quadre resum dels presentadors preferits per una mostra de 33 subjectes.

	Prefereix el primer presentador	Prefereix el segon presentador
<i>Presentador 1 vs. presentador 2</i>	32	1
<i>Presentador 1 vs. presentador 3</i>	27	6
<i>Presentador 1 vs. presentador 4</i>	29	4
<i>Presentador 2 vs. presentador 3</i>	6	27
<i>Presentador 2 vs. presentador 4</i>	20	13
<i>Presentador 3 vs. presentador 4</i>	26	7

A partir de les dades obtingudes, respon a les preguntes següents.

1. Quin és el valor escalar dels quatre estímuls presentats?
2. Quin és el presentador de televisió preferit?

3. Quina és la distància entre el presentador de televisió més preferit i el menys preferit?
4. Representa gràficament els estímuls en un contínuum unidimensional.

L'escalament de Thurstone. Solucions

Demaneu als subjectes que comparen els parells d'estímuls. Les respostes sobre la comparació d'estímuls es mostren a la taula 1.

Taula 1

Quadre resum dels presentadors preferits per una mostra de 33 subjectes.

	Prefereix el primer presentador	Prefereix el segon presentador
<i>Presentador 1 vs. presentador 2</i>	32	1
<i>Presentador 1 vs. presentador 3</i>	27	6
<i>Presentador 1 vs. presentador 4</i>	29	4
<i>Presentador 2 vs. presentador 3</i>	6	27
<i>Presentador 2 vs. presentador 4</i>	20	13
<i>Presentador 3 vs. presentador 4</i>	26	7

Crea la taula o matriu de freqüències (*F*) (taula 2).

Taula 2

Matriu *F* de freqüències.

	Presentador 1	Presentador 2	Presentador 3	Presentador 4
<i>Presentador 1</i>	16,5	1	6	4
<i>Presentador 2</i>	32	16,5	27	13
<i>Presentador 3</i>	27	6	16,5	7
<i>Presentador 4</i>	29	20	26	16,5

Passen les freqüències a proporcions i creem la taula o matriu de proporcions (*P*) (taula 3).

Taula 3*Matriu P de proporcions.*

	Presentador 1	Presentador 2	Presentador 3	Presentador 4
<i>Presentador 1</i>	0,50	0,03	0,18	0,12
<i>Presentador 2</i>	0,97	0,50	0,82	0,39
<i>Presentador 3</i>	0,82	0,18	0,50	0,21
<i>Presentador 4</i>	0,88	0,61	0,79	0,50

Passem les proporcions a puntuacions típiques i creem la taula o matriu de puntuacions típiques (*Z*) (taula 4).

Taula 4*Matriu Z de puntuacions típiques.*

	Presentador 1	Presentador 2	Presentador 3	Presentador 4
<i>Presentador 1</i>	0,0000	-1,88	-0,91	-1,17
<i>Presentador 2</i>	1,88	0,0000	0,91	-0,28
<i>Presentador 3</i>	0,91	-0,91	0,0000	-0,81
<i>Presentador 4</i>	1,17	0,28	0,81	0,0000

A partir de les puntuacions típiques calculem la mitjana d'aquests valors per a l'obtenció del valor escalar. Busquem el valor escalar més menut i el sumem en positiu a tots els valors, a fi d'evitar els negatius i tenir l'origen de l'escala en 0 (taula 4).

Taula 4

Valors escalars.

	Presentador 1	Presentador 2	Presentador 3	Presentador 4
<i>Presentador 1</i>	0,0000	-1,88	-0,91	-1,17
<i>Presentador 2</i>	1,88	0,0000	0,91	-0,28
<i>Presentador 3</i>	0,91	-0,91	0,0000	-0,81
<i>Presentador 4</i>	1,17	0,28	0,81	0,0000
<i>Suma Z</i>	3,96	-2,51	0,81	-2,26
<i>Mitjana Z</i>	0,99	-0,63	0,20	-0,57
<i>Suma del valor escalar més menut</i>	+0,63	+0,63	+0,63	+0,63
<i>Valors escalars positius</i>	1,62	0,00	0,83	0,06

Amb aquestes dades, ja podem respondre a les preguntes.

1. Quin és el valor escalar dels quatre estímuls presentats?

Els valors escalars són 0,00 per a l'estímul 2; 0,06 per a l'estímul 4; 0,83 per a l'estímul 3 i 1,62 per a l'estímul 1.

2. Quin és el presentador de televisió preferit?

El presentador 1, perquè té el valor escalar més alt (1,62).

3. Quina és la distància entre el presentador de televisió més preferit i el menys preferit?

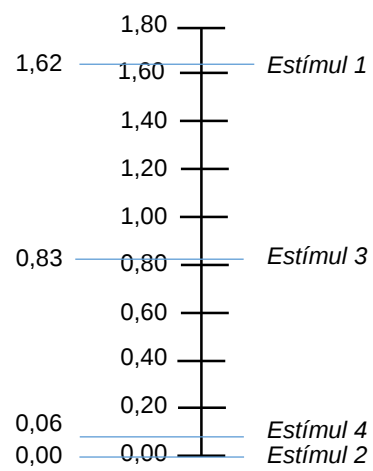
La distància és 1,62 (el presentador menys preferit té un valor escalar de 0,00; i el més preferit d'1,62).

4. Representa gràficament els estímuls en un contínuum unidimensional.

La figura 1 mostra la representació gràfica dels presentadors.

Figura 1

Representació gràfica dels valors escalars.



3.2. Les escales de Likert. Part 1

L'activitat s'ha de fer amb la base de dades *Base de dades per a activitat Likert*.

En aquesta primera part començarem amb els preliminars de l'anàlisi de les propietats d'una escala de tipus Likert. Necessitem fer aquestes primeres anàlisis quan construïm una escala de Likert, però també quan apliquem aquesta escala a una nova mostra de subjectes.

En aquest exemple emprem l'escala Sussex-Oxford Compassion for Others Scale (SOCS-O; Gu et al., 2020), que mesura la compassió pels altres.

Els ítems que formen l'escala són:

1. M'adone que algú està angoixat sense que m'ho diga.
2. Sé que tothom pateix en algun moment de la vida.
3. Quan una persona ho passa malament, la compadisc.
4. Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.
5. Quan algú ho passa malament, procure ajudar-lo d'alguna manera.
6. Percep si algú està angoixat.
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupe pel seu benestar.
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.
10. Quan una persona passa un mal moment, tracte de cuidar-la.
11. Detecte de seguida els primers signes d'angoixa en els altres.
12. Igual que jo, sé que altres persones també ho passen malament en la vida.
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.
14. Connecte amb el patiment dels altres sense jutjar-los.
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.
16. Detecte els signes de patiment en els altres.
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.
18. Soc sensible a l'angoixa dels altres.
19. Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angoixa.
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.

L'Escaleta de compassió pels altres d'Oxford (SOCS-O) avalua cinc dimensions de la compassió: 1. Reconeixement del patiment; 2. Comprensió del patiment com un sentiment universal. 3 Sentiment per la persona que pateix. 4. Tolerància als sentiments incòmodes; 5. Actuar o estar motivat a actuar per a alleujar el patiment.

Aquestes cinc dimensions es consideren aspectes relacionats amb la compassió però distingibles. Cada dimensió es mesura amb quatre ítems, dels quals cal calcular la mitjana, que a la base de dades reben el nom de SOCS:

- Reconeixement del patiment: ítems 1, 6, 11 i 16.
- Comprensió del patiment com un sentiment universal: ítems 2, 7, 12 i 17.
- Sentiment per la persona que pateix: ítems 3, 8, 13 i 18.
- Tolerància als sentiments incòmodes: ítems 4, 9, 14 i 19.
- Actuar o estar motivat a actuar per a alleujar el patiment d'altri: ítems 5, 10, 15 i 20.

En l'enllaç següent hi ha més informació sobre aquesta escaleta:

<https://journals.sagepub.com/doi/10.1177/1073191119860911>

A partir de la base de dades, i amb la informació de l'escaleta SOCS-O, fes les activitats següents:

- 1. DESCRIPCIÓ DE LA MOSTRA.** Quines característiques té la mostra a la qual s'ha aplicat l'escaleta? Calcula el *N* de la mostra i els estadístics descriptius per a gènere, edat i curs.
Els resultats s'han de presentar en format APA.
- 2. CREACIÓ DE VARIABLES.** Crea les variables de les cinc dimensions de compassió.
- 3. ESTADÍSTIQUES DESCRIPTIVES DELS ÍTEMS I LES DIMENSIONS.** Amb les dimensions calculades, ja podem presentar les estadístiques descriptives dels ítems juntament amb les dimensions a les quals pertanyen. És important incloure-hi la mitjana i la desviació típica, com també les puntuacions mínimes i màximes en cadascun dels ítems i de les dimensions.
Els resultats s'han de presentar en format APA.

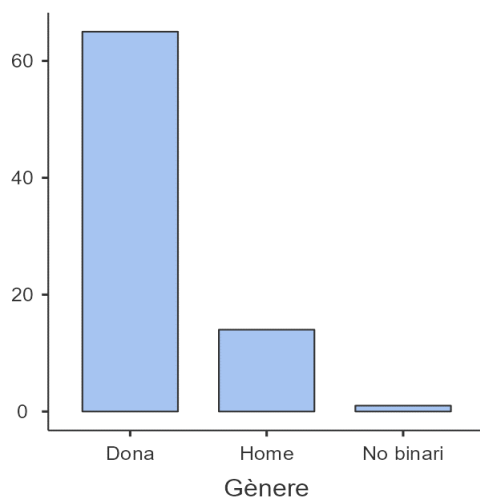
Les escales de Likert. Part 1. Solucions

Descripció de la mostra

La mostra objecte d'estudi està formada per 80 participants. La mitjana d'edat és de 19,8 anys ($DT = 2,64$), amb una edat mínima de 18 anys i màxima de 38. Pel que fa a la distribució per gènere, el 81,3% són dones ($n = 65$), el 17,5% homes ($n = 14$) i l'1,3% de gènere no binari ($n = 1$). La distribució es mostra a la figura 1.

Figura 1

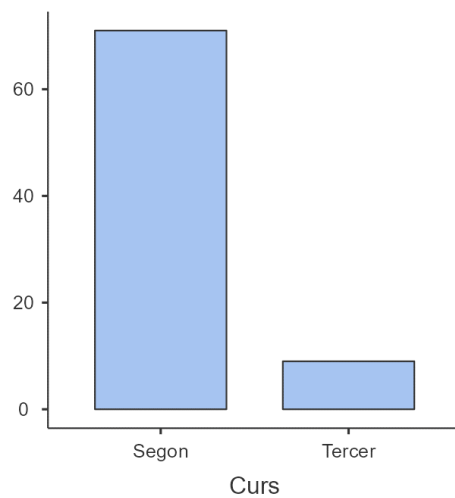
Distribució dels participants per gènere.



Finalment, pel que fa al curs, la majoria són estudiants de segon del Grau de Psicologia (88,8%, $n = 71$) i hi ha nou estudiants de tercer curs de la mateixa titulació (11,3%). La distribució per curs es mostra a la figura 2.

Figura 2

Distribució dels participants per curs.



Descripció de les variables objecte d'estudi

La descripció dels ítems i les dimensions de l'escala SOCS-O s'ha efectuat calculant la mitjana, la desviació típica i les puntuacions mínimes i màximes de cada variable.

Per a la dimensió *Reconeixement del patiment*, tal com s'observa a la taula 1, la mitjana és d'aproximadament quatre punts, cosa que implica puntuacions altes en l'escala, que oscil·la entre un i cinc punts. L'ítem amb la mitjana més alta és el 6, "Percep si algú està angoixat", mentre que el de puntuació més baixa és l'ítem 11, "Detecte de seguida els primers signes d'angoixa en els altres".

Taula 1

Estadístiques descriptives dels ítems i puntuació total de la dimensió Reconeixement del patiment.

Variables	M	DT	Mín.	Màx.
1. M'adone que algú està angoixat sense que m'ho diga.	4,01	0,58	3,00	5,00
6. Percep si algú està angoixat.	4,14	0,62	2,00	5,00
11. Detecte de seguida els primers signes d'angoixa en els altres.	3,89	0,71	3,00	5,00
16. Detecte els signes de patiment en els altres.	4,03	0,70	2,00	5,00
Reconeixement del patiment.	4,01	0,56	2,25	5,00

Notes: *M* = mitjana; *DT* = desviació típica; *Mín.* = puntuació mínima; *Màx.* = puntuació màxima.

En la dimensió *Comprensió del patiment com un sentiment universal*, la mitjana és de 4,62 punts, molt pròxima a la puntuació màxima de l'escala de cinc punts. Pel que fa als ítems, el de puntuació més alta és l'ítem 12, "Igual que jo, sé que altres persones també ho passen malament en la vida", mentre que el de mitjana més baixa és l'ítem 7, "Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana", tal com es mostra a la taula 2. Val a dir que en tots els casos les puntuacions mitjanes dels ítems són molt elevades.

Taula 2

Estadístiques descriptives dels ítems i puntuació total de la dimensió Comprensió del patiment com un sentiment universal.

Variables	M	DT	Mín.	Màx.
2. Sé que tothom pateix en algun moment de la vida.	4,64	0,56	3,00	5,00
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.	4,55	0,63	2,00	5,00
12. Igual que jo, sé que altres persones també ho passen malament en la vida.	4,70	0,56	3,00	5,00
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.	4,58	0,55	3,00	5,00
Comprensió del patiment com un sentiment universal.	4,62	0,44	3,25	5,00

Notes: *M* = mitjana; *DT* = desviació típica; *Mín.* = puntuació mínima; *Màx.* = puntuació màxima.

Pel que fa a la dimensió *Sentiment per la persona que pateix*, la mitjana és de 4,38 punts, elevada en l'escala. Si ens fixem en els ítems, la puntuació mitjana més elevada és per a l'ítem 13, "Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments", mentre que la mitjana més baixa la té l'ítem 18, "Soc sensible a l'angoixa dels altres". La taula 3 mostra aquesta informació.

Taula 3

Estadístiques descriptives dels ítems i puntuació total de la dimensió Sentiment per la persona que pateix.

Variables	M	DT	Mín.	Màx.
3. Quan una persona ho passa malament, la compadisc.	4,49	0,57	3,00	5,00
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupe pel seu benestar.	4,44	0,59	3,00	5,00
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.	4,56	0,59	2,00	5,00
18. Soc sensible a l'angoixa dels altres.	4,05	0,87	2,00	5,00
Sentiment per la persona que pateix.	4,38	0,54	3,00	5,00

Notes: *M* = mitjana; *DT* = desviació típica; *Mín.* = puntuació mínima; *Màx.* = puntuació màxima.

En la dimensió *Tolerància als sentiments incòmodes*, la mitjana en la mostra és de 4,13 punts. L'ítem amb una mitjana més baixa és l'ítem 19, "Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angoixa", l'únic ítem que arriba a la puntuació mitjana de quatre de tota l'escala. L'ítem amb la mitjana més alta és l'ítem 4, "Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los". Les estadístiques descriptives poden consultar-se a la taula 4.

Taula 4

Estadístiques descriptives dels ítems i puntuació total de la dimensió Tolerància als sentiments incòmodes.

Variables	<i>M</i>	<i>DT</i>	<i>Mín.</i>	<i>Màx.</i>
4. Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.	4,38	0,68	2,00	5,00
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.	4,40	0,59	3,00	5,00
14. Connecte amb el patiment dels altres sense jutjar-los.	4,13	0,74	2,00	5,00
19. Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angoixa.	3,61	0,91	2,00	5,00
Tolerància als sentiments incòmodes.	4,13	0,49	2,75	5,00

Notes: *M* = mitjana; *DT* = desviació típica; *Mín.* = puntuació mínima; *Màx.* = puntuació màxima.

Finalment, pel que fa a la dimensió *Actuar o estar motivat a actuar per a alleujar el patiment*, i tal com es mostra a la taula 5, l'ítem amb la mitjana més alta és el número 5 "Quan algú ho passa malament, procure ajudar-lo d'alguna manera", mentre que l'ítem amb la mitjana més baixa és el 20 "Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la", encara que tots superen la puntuació de quatre.

Taula 5

Estadístiques descriptives dels ítems i puntuació total de la dimensió Actuar o estar motivat a actuar per a alleujar el patiment.

Variables	<i>M</i>	<i>DT</i>	<i>Mín.</i>	<i>Màx.</i>
5. Quan algú ho passa malament, procure ajudar-lo d'alguna manera.	4,46	0,62	3,00	5,00
10. Quan una persona passa un mal moment, tracte de cuidar-la.	4,44	0,67	2,00	5,00
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.	4,39	0,67	2,00	5,00
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.	4,22	0,76	2,00	5,00
Actuar o estar motivat a actuar per a alleujar el patiment.	4,38	0,56	2,50	5,00

Notes: *M* = mitjana; *DT* = desviació típica; *Mín.* = puntuació mínima; *Màx.* = puntuació màxima.

3.3. Les escales de Likert. Part 2

L'activitat s'ha de fer amb la base de dades *Base de dades per a activitat Likert*.

En aquesta segona part, una vegada completats els passos previs per a l'anàlisi de les propietats d'una escala tipus Likert, passarem a analitzar els ítems de l'escala.

Continuarem usant el mateix exemple, l'escala Sussex-Oxford Compassion for Others Scale (SOCS-O; Gu et al., 2020), que mesura la compassió pels altres.

Els ítems que formen l'escala són:

1. M'adone que algú està angoixat sense que m'ho diga.
2. Sé que tothom pateix en algun moment de la vida.
3. Quan una persona ho passa malament, la compadisc.
4. Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.
5. Quan algú ho passa malament, procure ajudar-lo d'alguna manera.
6. Percep si algú està angoixat.
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupe pel seu benestar.
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.
10. Quan una persona passa un mal moment, tracte de cuidar-la.
11. Detecte de seguida els primers signes d'angoixa en els altres.
12. Igual que jo, sé que altres persones també ho passen malament en la vida.
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.
14. Connecte amb el patiment dels altres sense jutjar-los.
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.
16. Detecte els signes de patiment en els altres.
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.
18. Soc sensible a l'angoixa dels altres.
19. Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angoixa.
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.

L'*Escala de compassió pels altres d'Oxford (SOCS-O)* avalua cinc dimensions de la compassió: 1. Reconeixement del patiment; 2. Comprensió del patiment com un

sentiment universal; 3. Sentiment per la persona que pateix; 4. Tolerància als sentiments incòmodes. 5. Actuar o estar motivat a actuar per a alleujar el patiment.

Aquestes cinc dimensions es consideren aspectes relacionats amb la compassió però distingibles. Cada dimensió es mesura amb quatre ítems, dels quals cal calcular la mitjana, que a la base de dades reben el nom de SOCS:

- Reconeixement del patiment: ítems 1, 6, 11 i 16.
- Comprensió del patiment com un sentiment universal: ítems 2, 7, 12, 17 i 27.
- Sentiment per la persona que pateix: ítems 3, 8, 13 i 18.
- Tolerància als sentiments incòmodes: ítems 4, 9, 14 i 19.
- Actuar o estar motivat a actuar per a alleujar el patiment: ítems 5, 10, 15 i 20.

En aquest enllaç hi ha més informació sobre l'escala:

<https://journals.sagepub.com/doi/10.1177/1073191119860911>

A partir de la base de dades, i amb la informació de l'escala SOCS-O, fes les activitats següents:

- 1. ANÀLISI DE L'HOMOGENEÏTAT DELS ÍTEMS.** Els ítems de l'escala són homogenis? Calcula l'índex d'homogeneïtat per a cada ítem (respecte de la dimensió corresponent).
Els resultats s'han de presentar en format APA.
- 2. ANÀLISI DE LA DISCRIMINACIÓ DELS ÍTEMS.** Els ítems de l'escala són discriminants? Calcula les proves *t* per a cada ítem (respecte de la dimensió corresponent).
Els resultats s'han de presentar en format APA.
- 3. AVALUACIÓ DE L'ESCALA.** Tenint en compte els resultats de les anàlisis d'homogeneïtat i discriminació, rebutjaries o modificaries algun ítem? Interpreta aquests resultats i trau-ne conclusions.

Les escales de Likert. Part 2. Solucions

Anàlisi de l'homogeneïtat dels ítems

Per a estudiar l'homogeneïtat dels ítems es calculen els índexs d'homogeneïtat respectius per als ítems, a partir del càlcul de les correlacions ítem-total. La taula 1 mostra els resultats de la dimensió *Reconeixement del patiment*. Tots els ítems mostren correlacions positives, grans i estadísticament significatives amb la dimensió corresponent.

Taula 1

Resultats dels IH dels ítems de la dimensió Reconeixement del patiment.

	<i>IH</i>
1. M'adone que algú està angoixat sense que m'ho diga.	0,794 ^{***}
6. Percep si algú està angoixat.	0,872 ^{***}
11. Detecte de seguida els primers signes d'angoixa en els altres.	0,884 ^{***}
16. Detecte els signes de patiment en els altres.	0,855 ^{***}

Notes: ^{***} $p < 0,001$

Pel que fa a la dimensió *Comprensió del patiment com un sentiment universal*, els resultats van en la mateixa línia. Tal com es veu a la taula 2, tots els ítems mostren correlacions positives, grans i estadísticament significatives amb la dimensió corresponent que oscil·len entre 0,715 (ítem 17) i 0,868 (ítem 12).

Taula 2

Resultats dels IH dels ítems de la dimensió Comprensió del patiment com un sentiment universal.

	<i>IH</i>
2. Sé que tothom pateix en algun moment de la vida.	0,729***
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.	0,785***
12. Igual que jo, sé que altres persones també ho passen malament en la vida.	0,868***
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.	0,715***

Notes: *** $p < 0,001$.

També els ítems de la dimensió *Sentiment per la persona que pateix* mostren correlacions positives, grans i estadísticament significatives amb la puntuació en la dimensió (taula 3).

Taula 3

Resultats dels IH dels ítems de la dimensió Sentiment per la persona que pateix.

	<i>IH</i>
3. Quan una persona ho passa malament, la compadisc.	0,774***
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupa pel seu benestar.	0,801***
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.	0,787***
18. Soc sensible a l'angoixa dels altres.	0,869***

Notes: *** $p < 0,001$

Pel que fa a la dimensió *Tolerància als sentiments incòmodes*, les correlacions entre els ítems que la formen i la puntuació final en la dimensió són, de nou, positives, grans i estadísticament significatives, encara que, en general, el valor és inferior a l'observat fins ara. Excedeixen el punt de tall de 0,30, tal com es mostra a la taula 4 i es consideren, de nou, evidències adequades d'homogeneïtat dels ítems.

Taula 4

Resultats dels IH dels ítems de la dimensió Tolerància als sentiments incòmodes.

	<i>IH</i>
4. Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.	0,715***
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.	0,666***
14. Connecte amb el patiment dels altres sense jutjar-los.	0,736***
19. Quan algú està disgustat puc acompanyar-lo sense sentir-me aclaparat per la seua angouxa.	0,597***

Notes: *** $p < 0,001$

Finalment, es calculen els índexs d'homogeneïtat de la dimensió *Actuar o estar motivat a actuar per a alleujar el patiment*. De nou, les correlacions observades entre els ítems i la dimensió són positives, altes, estadísticament significatives i evidencien homogeneïtat. Els valors es poden consultar a la taula 5.

Taula 5

Resultats dels IH dels ítems de la dimensió Actuar o estar motivat a actuar per a alleujar el patiment.

	<i>IH</i>
5. Quan algú ho passa malament, procure ajudar-lo d'alguna manera.	0,793***
10. Quan una persona passa un mal moment, tracte de cuidar-la.	0,802***
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.	0,802***
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.	0,883***

Notes: *** $p < 0,001$

Anàlisi de la discriminació dels ítems

Per a dur a terme les anàlisis de discriminació dels ítems, en primer lloc es calculen els valors dels percentils 25 i 75 de cadascuna de les dimensions de l'escala SOCS-O. Els resultats es mostren a la taula 6.

Taula 6

Percentils 25 i 75 per a les dimensions de l'escala SOCS-O emprats com a punt de tall per a identificar els grups extrems.

	Pc25	Pc75
Reconeixement del patiment.	3,75	4,25
Comprensió del patiment com un sentiment universal.	4,50	5,00
Sentiment per la persona que pateix.	4,00	5,00
Tolerància als sentiments incòmodes.	3,75	4,50
Actuar o estar motivat a actuar per a alleujar el patiment.	4,00	5,00

Notes: *** $p < 0,001$

Aquests valors dels Pc25 i Pc75 s'han usat com a punt de tall per a crear els grups extrems en les cinc dimensions avaluades pel test. Una vegada creats els grups, es comparen les mitjanes dels dos grups en cadascun dels ítems de cada dimensió.

Tal com es mostra a la taula 7, tots els ítems de la dimensió *Reconeixement del patiment* són discriminants, ja que les mitjanes en tots els ítems d'alts i baixos en la dimensió han resultat estadísticament diferents.

Taula 7

Resultats de l'anàlisi de discriminació dels ítems de la dimensió Reconeixement del patiment.

	Prova t	gll	p	Diferència de mitjanes	d de Cohen
1. M'adone que algú està angoixat sense que m'ho diga.	-6,27 ^s	52	< 0,001	-0,929	-1,71
6. Percep si algú està angoixat.	-7,64 ^s	51	< 0,001	-1,064	-2,10
11. Detecte de seguida els primers signes d'angoixa en els altres.	-9,00 ^s	52	< 0,001	-1,324	-2,45
16. Detecte signes de patiment en els altres.	-8,38 ^w	47,9	< 0,001	-1,285	-2,30

Notes: ^s = prova t de Student; ^w = prova t de Welch; gll = graus de llibertat.

En el cas de la dimensió *Comprensió del patiment com un sentiment universal*, també tots els ítems tenen capacitat de discriminar entre grups extrems d'actitud, ja que en tots s'han trobat diferències en l'ítem entre els grups alts i baixos en el constructe. La informació relacionada amb les proves estadístiques es mostra a la taula 8.

Taula 8

Resultats de l'anàlisi de discriminació dels ítems de la dimensió Comprensió del patiment com un sentiment universal.

	Prova <i>t</i>	<i>gll</i>	<i>p</i>	Diferència de mitjanes	<i>d</i> de Cohen
2. Sé que tothom pateix en algun moment de la vida.	-6,55 ^w	30,0	< 0,001	-0,742	-1,66
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.	-8,804 ^w	30,0	< 0,001	-0,774	-1,64
12. Igual que jo, sé que altres persones també ho passen malament en la vida.	-6,45 ^w	30,0	< 0,001	-0,774	-1,64
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.	-9,71 ^w	30,0	< 0,001	-0,871	-2,47

Notes: ^w = prova *t* de Welch; *gll* = graus de llibertat.

Continuant amb la dimensió *Sentiment per la persona que pateix*, de nou tots els ítems tenen capacitat de discriminar entre grups extrems d'actitud, ja que en tots s'han trobat diferències en l'ítem entre els grups d'alts i baixos en el constructe, tal com es mostra a la taula 9.

Taula 9

Resultats de l'anàlisi de discriminació dels ítems de la dimensió Sentiment per la persona que pateix.

	Prova t	gll	p	Diferència de mitjanes	d de Cohen
3. Quan una persona ho passa malament, la compadisc.	-14,50 ^w	26,0	< 0,001	-1,07	-3,95
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupa pel seu benestar.	-13,63 ^w	26,0	< 0,001	-1,11	-3,71
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.	-9,37 ^w	26,0	< 0,001	-1,00	-2,55
18. Soc sensible a l'angoixa dels altres.	-15,15 ^w	26,0	< 0,001	-1,81	-4,12

Notes: ^w = prova t de Welch; gll = graus de llibertat.

Pel que fa a la dimensió *Tolerància als sentiments incòmodes*, de nou tots els ítems poden discriminar entre grups extrems d'actitud (taula 10).

Taula 10

Resultats de l'anàlisi de discriminació dels ítems de la dimensió Tolerància als sentiments incòmodes.

	Prova t	gll	p	Diferència de mitjanes	d de Cohen
4. Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.	-8,42 ^w	36,2	< 0,001	-1,165	-2,37
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.	-9,11 ^s	49,0	< 0,001	-0,962	-2,55
14. Connecte amb el patiment dels altres sense jutjar-los.	-7,04 ^s	49,0	< 0,001	-1,171	-1,97
19. Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angoixa.	-5,81 ^s	49,0	< 0,001	-1,194	-1,63

Notes: ^s = prova t de Student; ^w = prova t de Welch; gll = graus de llibertat.

Finalment, pel que fa als ítems de la dimensió *Actuar o estar motivat a actuar per a alleujar el patiment*, els resultats de les anàlisis estadístiques mostren que, també en aquesta dimensió, els ítems són discriminants. Els detalls es mostren a la taula 11.

Taula 11

Resultats de l'anàlisi de discriminació dels ítems de la dimensió Actuar o estar motivat a actuar per a alleujar el patiment.

	Prova <i>t</i>	<i>gll</i>	<i>p</i>	Diferència de mitjanes	<i>d</i> de Cohen
5. Quan algú ho passa malament, procure ajudar-lo d'alguna manera.	-15,9 ^w	31,0	< 0,001	-1,09	-3,96
10. Quan una persona passa un mal moment, tracte de cuidar-la.	-11,4 ^w	31,0	< 0,001	-1,16	-2,85
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.	-11,5 ^w	31,0	< 0,001	-1,13	-2,87
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.	-12,9 ^w	31,0	< 0,001	-1,41	-3,23

Notes: ^w = prova *t* de Welch; *gll* = graus de llibertat.

Avaluació de l'escala

Tal com s'ha pogut observar, tots els ítems de les cinc dimensions de l'escala SOCS-O han obtingut evidències adequades d'homogeneïtat i de discriminació. Per tant, no caldria eliminar-ne ni reformular-ne cap. En conclusió, podem dir que els ítems de les cinc dimensions tenen propietats psicomètriques adequades. El pas següent seria estudiar el funcionament de les dimensions de l'escala, en termes de fiabilitat i validesa.

3.4. L'escalograma de Guttman

Una psicòloga interessada a saber quin és el grau d'empatia dels joves ha creat una escala de sis ítems que mesuren comportaments empàtics. En un estudi pilot administra l'escala a una mostra de 10 joves. Les respostes es mostren a la taula 1 (1 = duu a terme el comportament empàtic; 0 = no duu a terme el comportament empàtic).

Taula 1

Matriu de respostes.

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

A partir de les dades de la matriu, respon a les preguntes següents:

1. Quin subjecte presenta el grau més alt de conducta empàtica?
2. Quants errors segons Guttman hi ha a l'escala?
3. Analitza les dades i determina si constitueixen un escalograma. Es pot dir que aquesta escala és un escalograma? Justifica la resposta.
4. Si aquesta escala fora un escalograma, estaria esbiaixat?
5. Si aquesta escala no fora un escalograma, què es podria fer perquè ho fora?
6. Identifica els ítems amb millor i pitjor funcionament.

L'escalograma de Guttman. Solucions

En primer lloc, cal executar els diversos passos de l'escalograma de Guttman. El primer consisteix a sumar el nombre d'encerts dels subjectes i en cadascun dels ítems.

Taula 1

Suma del nombre d'encerts en la mostra de 10 subjectes de l'escalograma de Guttman per a avaluar la conducta empàtica.

Subjectes	Ítems						Puntuació total en l'escalograma
	1	2	3	4	5	6	
1	1	1	0	1	0	0	3
2	0	0	0	0	0	0	0
3	1	1	1	1	1	1	6
4	1	0	1	0	0	0	2
5	1	1	1	0	0	0	3
6	1	1	1	0	1	0	4
7	1	0	0	0	0	0	1
8	1	0	1	0	0	0	2
9	1	1	1	1	0	0	4
10	1	1	1	1	1	0	5
Total d'encerts	9	6	7	4	3	1	

Amb aquestes dades ja podem respondre a la pregunta 1.

1. Quin subjecte té el grau més alt de conducta empàtica?

El subjecte amb més conducta empàtica és el 3, que assoleix una puntuació de 6 ($X = 6$).

El segon pas de la matriu consisteix a reordenar els ítems segons la dificultat. Si observem la taula 1, l'ítem 3 és més fàcil (té més nombre de respostes 1) que l'ítem 2. Per tant, caldria invertir l'ordre d'aquests dos ítems. Creem, doncs, la taula 2, en què canviem l'ordre dels ítems 2 i 3.

Taula 2

Matriu de respostes de l'escalograma de Guttman després de la reordenació dels ítems segons la dificultat.

<i>Subjectes</i>	Ítems						Puntuació total en l'escalograma
	1	3	2	4	5	6	
<i>1</i>	1	0	1	1	0	0	3
<i>2</i>	0	0	0	0	0	0	0
<i>3</i>	1	1	1	1	1	1	6
<i>4</i>	1	1	0	0	0	0	2
<i>5</i>	1	1	1	0	0	0	3
<i>6</i>	1	1	1	0	1	0	4
<i>7</i>	1	0	0	0	0	0	1
<i>8</i>	1	1	0	0	0	0	2
<i>9</i>	1	1	1	1	0	0	4
<i>10</i>	1	1	1	1	1	0	5
<i>Total d'encerts</i>	9	7	6	4	3	1	

Sobre aquesta matriu ja podem calcular els errors en els patrons de resposta, tal com es mostra a la taula 3.

Taula 3

Comptabilització dels errors en la matriu de respostes reordenada.

<i>Subjectes</i>	Ítems						Puntuació total en l'escalograma
	1	3	2	4	5	6	
1	1	0*	1	1*	0	0	3
2	0	0	0	0	0	0	0
3	1	1	1	1	1	1	6
4	1	1	0	0	0	0	2
5	1	1	1	0	0	0	3
6	1	1	1	0*	1*	0	4
7	1	0	0	0	0	0	1
8	1	1	0	0	0	0	2
9	1	1	1	1	0	0	4
10	1	1	1	1	1	0	5
<i>Total d'encerts</i>	9	7	6	4	3	1	
<i>Errors</i>	0	1	0	2	1	0	

2. Quants errors segons Guttman hi ha a l'escala?

Com es pot observar a la taula 3, hi ha quatre errors en el patró de resposta. A partir d'aquesta informació es pot calcular el coeficient de reproductibilitat de l'escala.

3. Analitza les dades obtingudes i determina si constitueixen un escalograma.
Es pot dir que aquesta escala és un escalograma? Justifica la resposta.

Per a saber si es tracta d'un escalograma hem de calcular el coeficient de reproductibilitat de l'escala:

$$\begin{aligned}
 \text{Coeficient de reproductibilitat (CR)} &= 1 - \frac{\text{Nombre total d'errors}}{(\text{Nombre d'ítems}) \times (\text{Nombre de subjectes})} \\
 &= 1 - \frac{4}{6 \times 10} = 1 - \frac{4}{60} = 1 - 0,067 = 0,933
 \end{aligned}$$

En aquest cas, com que el $CR \geq 0,90$, sembla que és una escala perfecta. No obstant això, caldria saber si aquest valor es deu al fet que l'escala està esbiaixada o

no. Comprovem si l'escala està esbiaixada mitjançant el càlcul del coeficient d'escalabilitat (CE).

Per a fer-ho, cal calcular les categories modals de cada ítem, a partir de les proporcions d'uns i zeros que s'observen en les respostes. Aquest pas es pot veure a la taula 4.

Taula 4

Comptabilització dels errors en la matriu de respostes reordenada.

Subjectes	Ítems						Puntuació total en l'escalograma
	1	3	2	4	5	6	
1	1	0*	1	1*	0	0	3
2	0	0	0	0	0	0	0
3	1	1	1	1	1	1	6
4	1	1	0	0	0	0	2
5	1	1	1	0	0	0	3
6	1	1	1	0*	1*	0	4
7	1	0	0	0	0	0	1
8	1	1	0	0	0	0	2
9	1	1	1	1	0	0	4
10	1	1	1	1	1	0	5
Total d'encerts	9	7	6	4	3	1	
<i>p</i>	0,9	0,7	0,6	0,4	0,3	0,1	
<i>q (1 - p)</i>	0,1	0,3	0,4	0,6	0,7	0,9	
Errors	0	1	0	2	1	0	

Nota: les categories modals van en negreta.

Una vegada tenim les categories modals, ja podem calcular la reproductibilitat marginal mínima (RMM), que és la mitjana aritmètica de les categories modals de tots els ítems:

$$\begin{aligned} \text{Reproductibilitat marginal mínima (RMM)} &= \frac{\text{Suma de les categories modals dels ítems}}{\text{Nombre d'ítems}} \\ &= \frac{0,90 + 0,70 + 0,60 + 0,60 + 0,70 + 0,90}{6} = \frac{4,4}{6} = 0,733 \end{aligned}$$

El resultat de l'RMM ens permet calcular el CE de l'escala:

$$\text{Coeficient d'escalabilitat(CE)} = \frac{CR - RMM}{1 - RMM} = \frac{0,933 - 0,733}{1 - 0,733} = \frac{0,200}{0,267} = 0,749$$

4. Si aquesta escala fora un escalograma, estaria esbiaixat?

Com que el CE és més gran que 0,60, podríem dir que l'escala no està esbiaixada.

5. Si aquesta escala no fora un escalograma, què es podria fer perquè ho fora?

Si l'escala no haguera sigut un escalograma, caldria examinar els coeficients de reproductibilitat dels ítems (CRi) per a determinar quins presenten un funcionament inadequat i eliminar l'ítem amb el pitjor funcionament. Una vegada eliminat l'ítem, caldria calcular de nou el CR de l'escala.

6. Identifica els ítems amb millor i pitjor funcionament.

Per a saber com han funcionat els ítems, calculem el CRi de cadascun:

$$CRi_1 = 1 - \left(\frac{0}{10}\right) = 1 - 0 = 1$$

$$CRi_3 = 1 - \left(\frac{1}{10}\right) = 1 - 0,1 = 0,90$$

$$CRi_2 = 1 - \left(\frac{0}{10}\right) = 1 - 0 = 1$$

$$CRi_4 = 1 - \left(\frac{2}{10}\right) = 1 - 0,2 = 0,80$$

$$CRi_5 = 1 - \left(\frac{1}{10}\right) = 1 - 0,1 = 0,90$$

$$CRi_6 = 1 - \left(\frac{0}{10}\right) = 1 - 0 = 1$$

Els ítems que funcionen millor són els ítems 1, 2 i 6, amb un CRi = 1. Per la seua banda, el que funciona pitjor és l'ítem 4, amb un CRi = 0,80.

Tema 4. Construcció de tests

4.1. Construcció d'ítems

Tenint en compte les recomanacions per a redactar ítems d'execució màxima vistes en el tema 4, crea i resol tres exemples de cadascun dels tipus d'ítems següents:

1. D'elecció múltiple (de tres alternatives)
2. Vertader-fals
3. D'assaig, desenvolupament o construcció (breu)

El contingut dels ítems ha de ser dels temes 1, 2 i 3 de l'assignatura de Psicometria. Indica les alternatives correctes o, si s'escau, desenvolupa les respostes corresponents.

4.2. Anàlisi d'ítems

En aquesta activitat treballarem sobre la base de dades del qüestionari DAT-5, amb la subescala que avalua la dimensió de raonament verbal. Aquesta prova està formada per 10 ítems d'opció múltiple, amb cinc alternatives de resposta i només una és la correcta.

Analitzarem els 10 ítems que conté la prova a fi de veure si compleixen les propietats psicomètriques fonamentals.

Segueix els passos del *Tutorial amb exemple de l'ítem 1* i respon a les preguntes.

1. Calcula i interpreta l'índex de dificultat (ID) i l'índex de dificultat corregit (ID') dels ítems. Respon a les preguntes següents:
 - A partir dels resultats obtinguts en cadascun dels ítems, quina dificultat té cada ítem?
 - I si la corregim considerant la possibilitat d'encert per atzar?
 - Quin és l'ítem més difícil?
 - I el més fàcil?
 - Tenint en compte les recomanacions dels autors sobre la dificultat que han de tenir els ítems que tinguen possibilitat d'encert per atzar, rebutjaríem algun ítem perquè és massa difícil?
2. Calcula i interpreta l'índex de discriminació (D) dels ítems.
 - A partir dels resultats obtinguts en cadascun dels ítems, quin és l'ítem més discriminant?
 - I el menys discriminant?
 - N'hi ha cap que no arriba al punt de tall i que hauríem de rebutjar?
3. Calcula i interpreta l'índex d'homogeneïtat (IH) dels ítems.
 - Segons els resultats obtinguts en cadascun dels ítems, quin és més homogeni?
 - I l'ítem menys homogeni?
 - Si prenem com a punt de tall 0,30, rebutjaríem algun ítem?

4. Calcula i interpreta l'índex de validesa (IV) dels ítems.

- A partir dels resultats en cadascun dels ítems, quin és l'ítem amb més IV?
- I l'ítem amb menys IV?
- Rebutjaríem algun ítem per falta d'evidències de validesa, és a dir, perquè el valor de la correlació no siga estadísticament significatiu?

5. Analitza com funcionen els distractors dels ítems. Per a cada ítem, respon a les preguntes següents:

- L'alternativa correcta és la més seleccionada?
- Els distractors són equiprobables?
- El rendiment mitjà dels participants que han triat cada distractor (opció incorrecta) és inferior al dels participants que han seleccionat la correcta, i no mostra diferències entre els qui han triat les alternatives incorrectes?
- Tenint en compte aquesta informació, modificaries o eliminaries algun distractor o algun ítem?

Anàlisi d'ítems. Tutorial amb exemple de l'ítem 1

Tasques prèvies

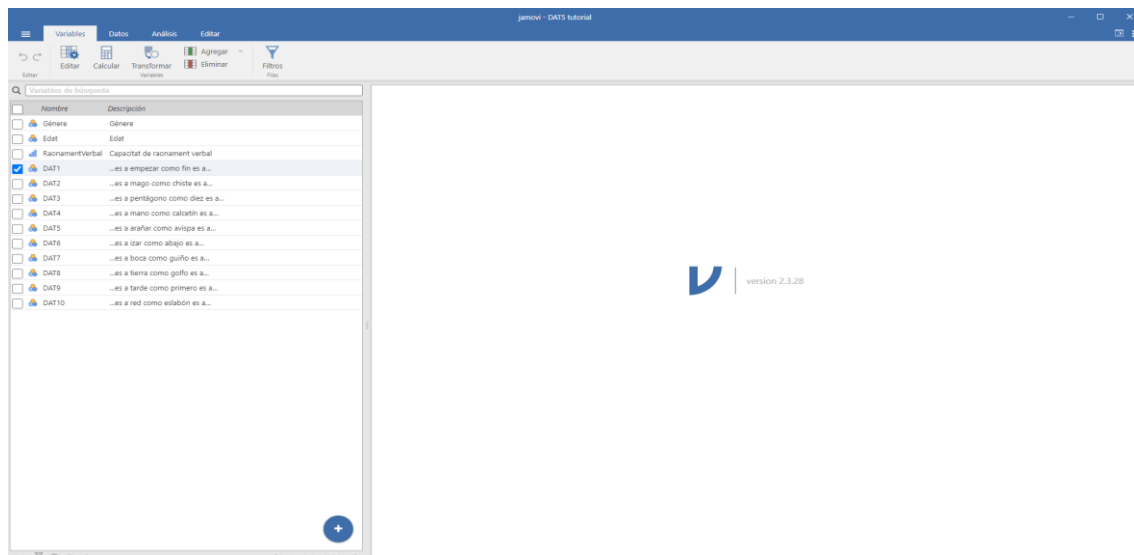
Per al càlcul de molts dels índexs amb què treballarem és imprescindible saber **quines persones han encertat els ítems i quines no**. És a dir, el primer pas és puntuar els subjectes segons que hagen triat l'alternativa correcta de resposta o no.

Per a puntuar els subjectes hem de saber quines són les alternatives de resposta correcta en cada pregunta del DAT-5:

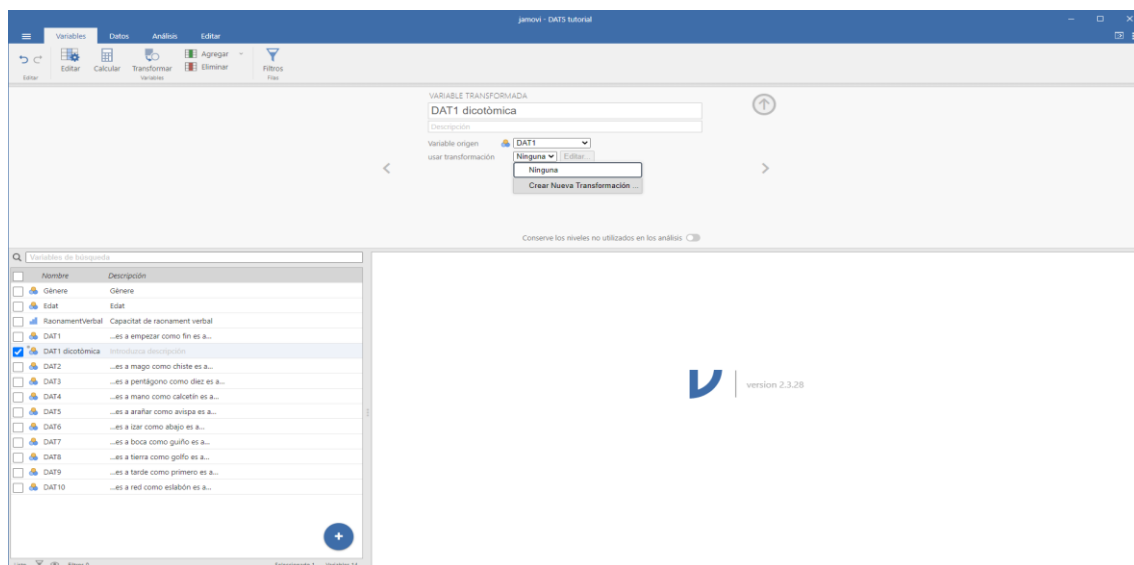
- Ítem 1 – opció C
- Ítem 2 – opció E
- Ítem 3 – opció B
- Ítem 4 – opció B
- Ítem 5 – opció B
- Ítem 6 – opció C
- Ítem 7 – opció A
- Ítem 8 – opció A
- Ítem 9 – opció D
- Ítem 10 – opció B

Si volguérem, per exemple, recodificar l'ítem 1 en una variable dicotòmica d'encert-error, hauríem de transformar la variable en una de nova, en què l'opció C (la correcta) valguera un punt i les altres opcions (les incorrectes) valgueren 0 punts.

En Jamovi, anem a *Variables*, seleccionem la que volem transformar i fem clic en el boto *Transformar*.



Una vegada ací, donem nom a la nova variable, per exemple, DAT1 dicotòmica, ens assegurem que la variable d'origen és el DAT1 i fem clic en l'opció *Crear nova transformació*.



La transformació ha de donar dues instruccions:

- Si en la variable d'origen, la puntuació és un 3 (opció C), en la nova variable ha de ser un 1 (encert).
- Si en la variable d'origen, la puntuació no és un 3 (opció C), en la nova variable ha de ser un 0 (error).

En Jamovi, les ordres han de prendre aquest format:

The screenshot shows the 'TRANSFORMAR' (Transform) window in Jamovi. The title is 'Opció correcta la C'. Under 'Descripción', it says 'Variable de sufijo'. There are three recoding rules listed:

- Rule 1: f_k si $\$source == 3$ usar 1
- Rule 2: f_k si $\$source != 3$ usar 0
- Rule 3: f_k si no usar e.g. $\$source$

The 'Tipo de medida' (Measure type) is set to 'Nominal'.

Procedim d'aquesta manera per als altres ítems, tenint en compte quina és l'opció correcta en cada cas. Per als altres ítems que tinguem com a opció correcta l'opció C, com l'ítem 1, podem aplicar aquesta mateixa transformació. En el cas dels ítems que tinguem una altra opció de resposta correcta cal crear una nova transformació.

Igualment, també és necessari disposar d'una puntuació del constructe, que, com hem vist, sol ser la **puntuació total en el test**.

Una vegada tenim creades les variables dicotòmiques dels 10 ítems, podem procedir a calcular la puntuació total per al DAT-5. Per a fer-ho, anem a *Variables – Calcular*, donem nom a la nova variable i seleccionem l'opció SUM:

The screenshot shows the 'Calcular' (Calculate) window in Jamovi. The title is 'Puntuació total DAT'. The formula entered is: $\text{DAT2} + \text{DAT3} + \text{DAT4} + \text{DAT5} + \text{DAT6} + \text{DAT7} + \text{DAT8} + \text{DAT9} + \text{DAT10} + \text{DAT11}$. The window shows a list of variables on the left and a list of functions on the right. The 'Variables' list includes DAT2, DAT3, DAT4, DAT5, DAT6, DAT7, DAT8, DAT9, DAT10, and DAT11. The 'Functions' list includes SUM, which is selected.

Hem d'incloure dins de la fórmula de suma els 10 ítems del DAT recodificats en encerts-error, separats per comes:

VARIABLE CALCULADA

Puntuació total DAT

Descripción

Fórmula

f_x

= SUM('DAT1 dicotòmica', 'DAT2 dicotòmica', 'DAT3 dicotòmica', 'DAT4 dicotòmica', 'DAT5 dicotòmica', 'DAT6 dicotòmica', 'DAT7 dicotòmica', 'DAT8 dicotòmica', 'DAT9 dicotòmica', 'DAT10 dicotòmica')

Una vegada tenim aquesta variable, passem a la fase d'anàlisi d'ítems.

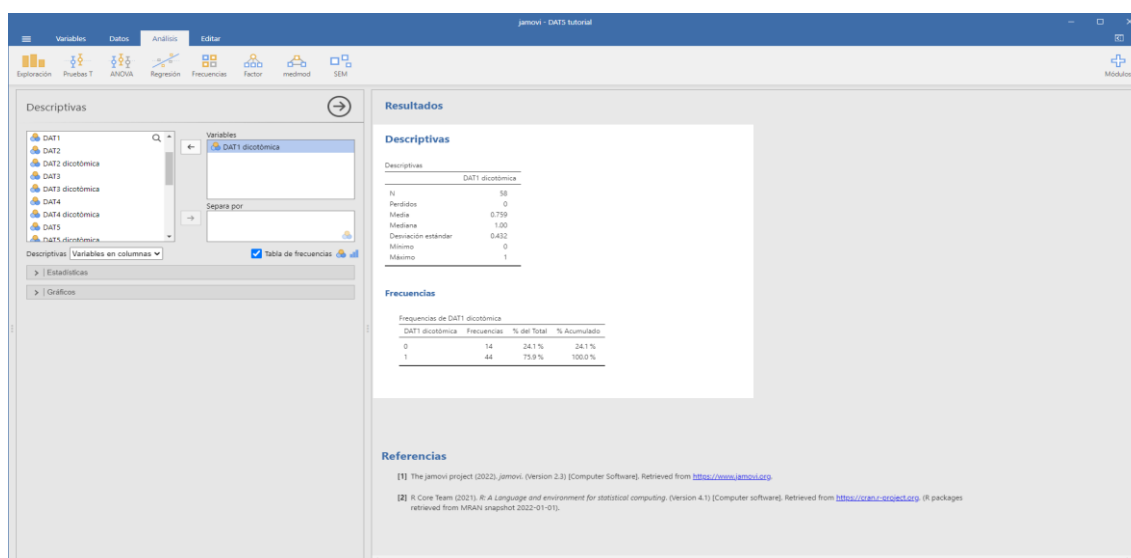
Exemple de càlcul de l'índex de dificultat per a l'ítem 1.

Per a calcular l'índex de dificultat (ID) és necessari saber la proporció d'encerts en cadascun dels ítems. Recordem que tenim aquesta informació en les noves variables recodificades.

Així, només cal anar al menú *Anàlisi – Exploració – Descriptives*.

Ítem	Puntuació total DAT
1	10
2	10
3	10
4	10
5	10
6	10
7	10
8	10
9	10
10	10
11	10
12	10
13	10
14	10
15	10
16	10
17	10
18	10
19	10
20	10
21	10
22	10
23	10
24	10
25	10
26	10
27	10
28	10
29	10
30	10
31	10
32	10
33	10
34	10
35	10
36	10
37	10
38	10
39	10
40	10

Una vegada en aquest apartat, passem l'ítem a la columna de la dreta, *Variables*, i seleccionem l'opció *Taula de freqüències*, per a saber, del total dels subjectes, quants han encertat l'ítem 1:



Amb aquesta informació ja podem calcular l'índex de dificultat per a l'ítem 1, que recordem que és el quocient entre el nombre d'avaluats que l'han encertat ($E_j = 44$) i el total d'avaluats que han respost a l'ítem ($N_j = 58$). En aquest cas seria:

$$ID_j = \frac{E_j}{N_j} = \frac{44}{58} = 0,76$$

L'índex de dificultat corregit (ID') seria el següent:

$$ID_j^c = ID_j - \frac{q_j}{K - 1} = 0,76 - \frac{0,24}{5 - 1} = 0,76 - \frac{0,24}{4} = 0,76 - 0,06 = 0,70$$

En què ID_j és l'índex de dificultat sense corregir ($ID_j = 0,76$), K és el nombre d'alternatives o opcions de resposta ($K = 5$) i q_j és la proporció d'errors comesos ($q_j = 1 - ID_j = 1 - 0,76 = 0,24$).

Tenint en compte aquests resultats, podem dir que l'ítem 1 té una dificultat adequada, ja que és superior a 0,50 (Crocker i Algina, 1986).

Exemple de càlcul de l'índex de discriminació per a l'ítem 1.

Per a avaluar el poder discriminant d'un ítem és necessari tenir una mesura de la competència en el constructe que el test pretén mesurar, un criteri. Aquest criteri serà la puntuació total en el test, la variable que prèviament hem creat.

L'índex de discriminació (D) requereix establir dos grups d'avaluats a partir de les seues puntuacions en el test: el grup amb puntuacions altes (25% superior) i el grup amb les puntuacions més baixes (25% inferior). Per a saber qui són en la mostra, caldrà

demanar els percentils 25 i 75 de la variable *Puntuació total DAT*. Anem a *Anàlisi – Exploració – Descriptives*:

Resultados

Descriptivas

Descriptivas

DAT1 dicotómica	
N	58
Perdidos	0
Media	0.759
Mediana	1.00
Desviación estándar	0.432
Mínimo	0
Máximo	1

Frecuencias

Frecuencias de DAT1 dicotómica

DAT1 dicotómica	Frecuencias	% del Total	% Acumulado
0	14	24.1 %	24.1 %
1	44	75.9 %	100.0 %

Referencias

[1] The Jamovi project (2022). *Jamovi*. (Version 2.3) [Computer Software]. Retrieved from <https://www.jamovi.org>

[2] R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from <https://cran.r-project.org> (R packages retrieved from MIRROR snapshot 2022-01-01).

I demanem els percentils 25 i 75 per a la variable *Puntuació total DAT*:

Frecuencias

Frecuencias de DAT1 dicotómica

DAT1 dicotómica	Frecuencias	% del Total	% Acumulado
0	14	24.1 %	24.1 %
1	44	75.9 %	100.0 %

Descriptivas

Descriptivas

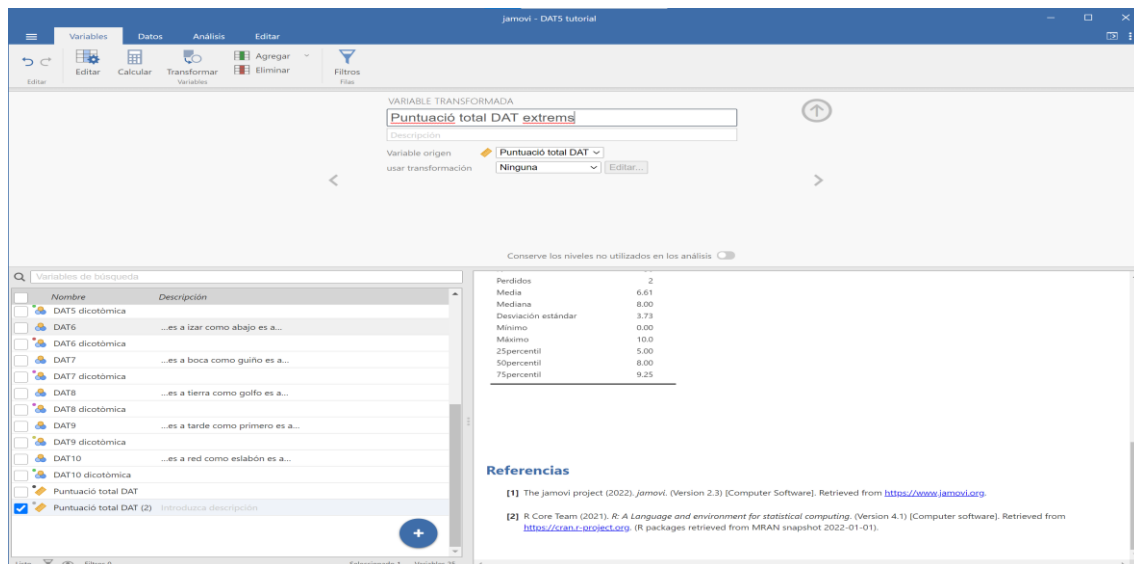
Puntuació total DAT	
N	56
Perdidos	2
Media	6.61
Mediana	8.00
Desviación estándar	3.73
Mínimo	0.00
Máximo	10.00
25percentil	5.00
50percentil	8.00
75percentil	9.25

Referencias

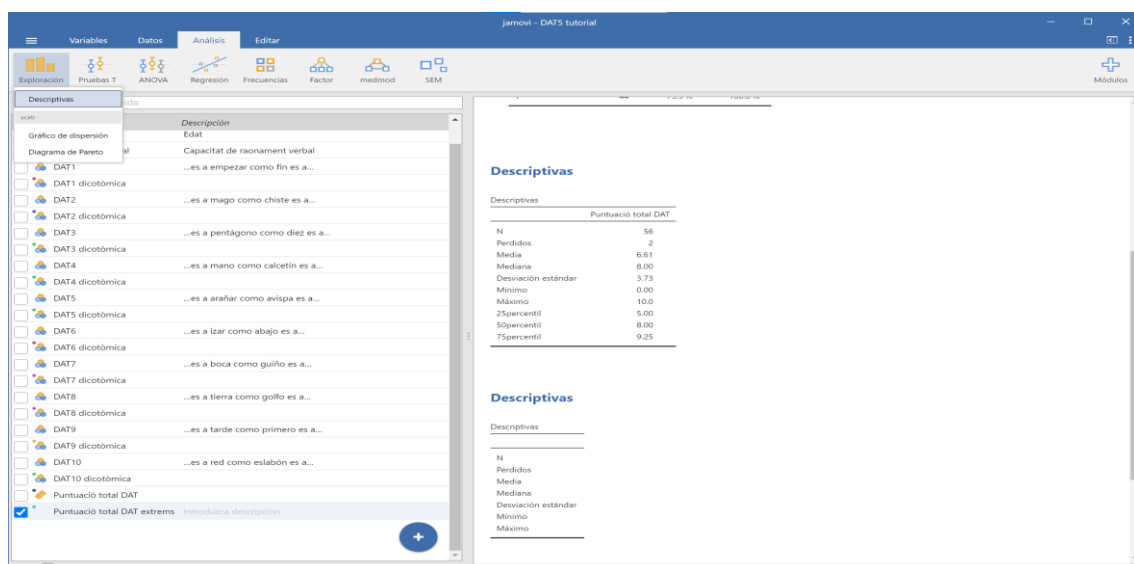
[1] The Jamovi project (2022). *Jamovi*. (Version 2.3) [Computer Software]. Retrieved from <https://www.jamovi.org>

[2] R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from <https://cran.r-project.org>

Coneixent aquests percentils, transformem la variable *Puntuació total* en una nova variable, en què recodificarem els qui hagen tingut 5 punts o menys com a 0 (baixos) i els qui hagen tingut 9,25 punts o més com a 1 (alts):



Quan ja tenim aquesta nova variable, només caldrà veure la proporció d'encerts de l'ítem 1 que té cadascun dels dos grups, alts i baixos en el DAT total. Anem a *Anàlisi – Exploració – Descriptives*:



I demanem la taula de freqüències per a l'ítem 1 (la variable dicotòmica d'encerts i errors que hem creat), separada en funció de la variable *DAT total extrems*, és a dir, que ens diga la proporció d'encerts i errors en funció dels grups d'alts i baixos:

Ens fixem primer en el grup de baixos o grup inferior. Tal com es mostra a la taula de freqüències, dels baixos en el DAT total, tenim 13 persones que han fallat l'ítem 1 (tenen un 0 en DAT1 dicotòmica) i dues persones que l'han encertat (tenen un 1 en DAT1 dicotòmica):

Frecuencias de DAT1 dicotòmica

DAT1 dicotòmica	Puntuació total DAT extremes	Frecuencias	% del Total	% Acumulado
0	0	13	44.8 %	44.8 %
	1	0	0.0 %	44.8 %
1	0	2	6.9 %	51.7 %
	1	14	48.3 %	100.0 %

Per tant, en total són 15 subjectes. La proporció d'encerts del grup inferior (p_i) per a l'ítem 1 seria:

$$p_i = \frac{2}{15} = 0,13$$

Si ens fixem ara en el grup superior, dels alts en el DAT total, no tenim cap persona que haja fallat l'ítem 1 (0 en DAT1 dicotòmica) i 14 persones que l'han encertat (tenen un 1 en DAT1 dicotòmica):

Frecuencias de DAT1 dicotòmica

DAT1 dicotòmica	Puntuació total DAT extrems	Frecuencias	% del Total	% Acumulado
0	0	13	44.8 %	44.8 %
	1	0	0.0 %	44.8 %
1	0	2	6.9 %	51.7 %
	1	14	48.3 %	100.0 %

Per tant, en total són 14 subjectes. La proporció d'encerts del grup superior (p_s) per a l'ítem 1 seria:

$$p_s = \frac{14}{14} = 1$$

Amb aquestes dades ja podem calcular l'índex de discriminació (D) per a l'ítem 1:

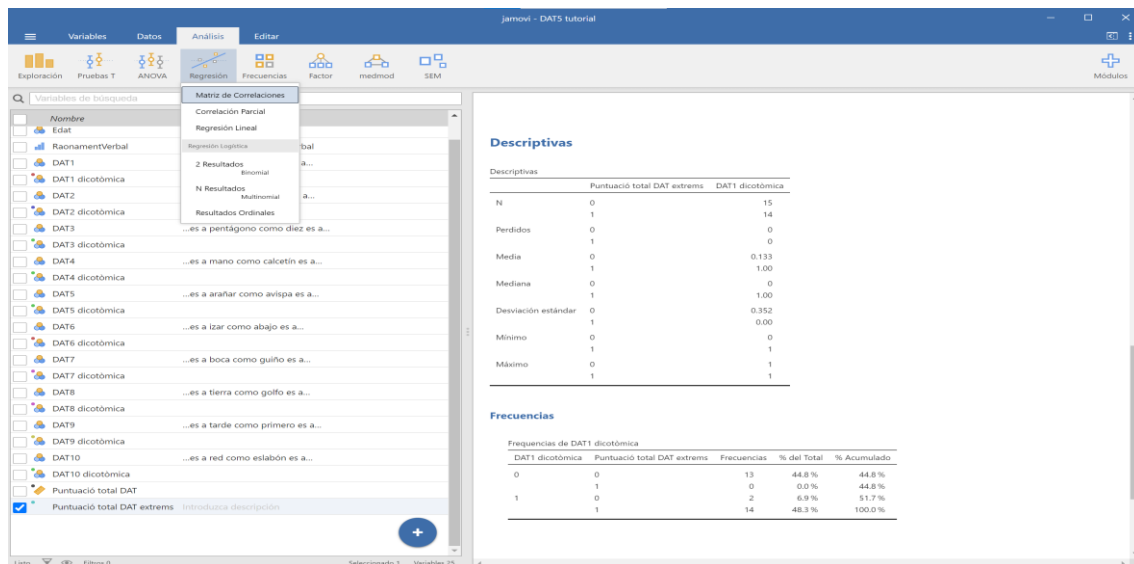
$$D_j = p_s - p_i = 1 - 0,13 = 0,87$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 1 discrimina molt bé, ja que supera el punt de tall de 0,40.

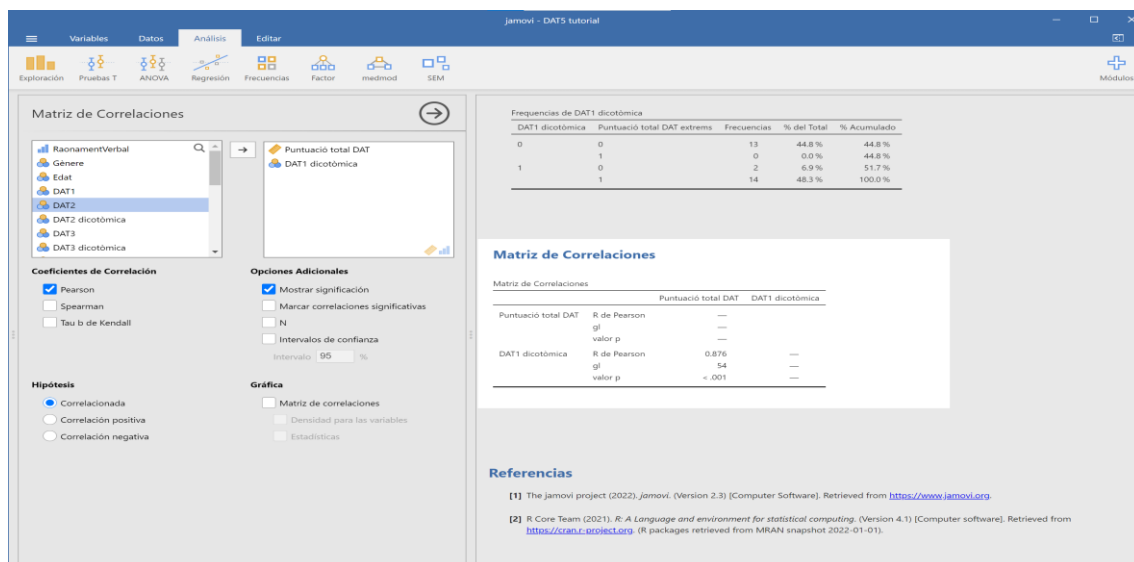
Exemple de càlcul de l'índex d'homogeneïtat per a l'ítem 1.

Per a calcular l'índex d'homogeneïtat (IH) és necessari correlacionar els ítems (en la versió dicotòmica, on tenim recodificats els subjectes que l'han encertat en 1 i els qui l'han fallat en 0) amb la *Puntuació total en el DAT* (l'original, no la que classifica alts i baixos).

Per a calcular l'IH de l'ítem 1, anem a *Anàlisi – Regressió – Matriu de correlacions*:



I correlacionem l'ítem 1 dicotòmic amb la puntuació total en el DAT:



Com que en aquest cas, la correlació ítem-test pren un valor de 0,876 i és positiva i estadísticament significativa ($p < 0,050$) podem dir que l'ítem 1 és homogeni o que té un IH adequat.

Exemple de càlcul de l'índex de validesa per a l'ítem 1.

Per a calcular l'índex de validesa (IV) és necessari correlacionar els ítems (en la versió dicotòmica, on tenim recodificats els subjectes que l'han encertat en 1 i els qui l'han fallat en 0) amb un altre criteri extern, distint del test. En aquest cas, a la base de dades tenim una mesura de *Raonament verbal percebut*: si els ítems mesuren raonament verbal, considerem que els avaluats que tinguen més puntuació en l'ítem han de ser també els que es perceben més competents en raonament verbal.

Per a calcular l'IV de l'ítem 1 anem a *Anàlisi – Regressió – Matriu de correlacions*:

Descriptivas

Descriptivas	Puntuació total DAT extremes	DAT1 dicotòmic
N	0	15
	1	14
Pèrdues	0	0
	1	0
Media	0	0.133
	1	1.00
Mediana	0	0
	1	1.00
Desviació estàndard	0	0.352
	1	0.00
Mínimo	0	0
	1	1
Máximo	0	1
	1	1

Frecuencias

Frecuencias de DAT1 dicotòmic

DAT1 dicotòmic	Puntuació total DAT extremes	Frecuencias	% del Total	% Acumulado
0	0	13	44.8 %	44.8 %
	1	0	0.0 %	44.8 %
1	0	2	6.9 %	51.7 %
	1	14	48.3 %	100.0 %

I correlacionem l'ítem 1 dicotòmic amb la variable *Raonament verbal*:

Matriu de Correlacions

Frecuencias de DAT1 dicotòmic

DAT1 dicotòmic	Puntuació total DAT extremes	Frecuencias	% del Total	% Acumulado
0	0	13	44.8 %	44.8 %
	1	0	0.0 %	44.8 %
1	0	2	6.9 %	51.7 %
	1	14	48.3 %	100.0 %

Matriu de Correlacions

	RaonamentVerbal	DAT1 dicotòmic
RaonamentVerbal	R de Pearson	—
	gl	—
	valor p	—
DAT1 dicotòmic	R de Pearson	0.537
	gl	56
	valor p	< .001

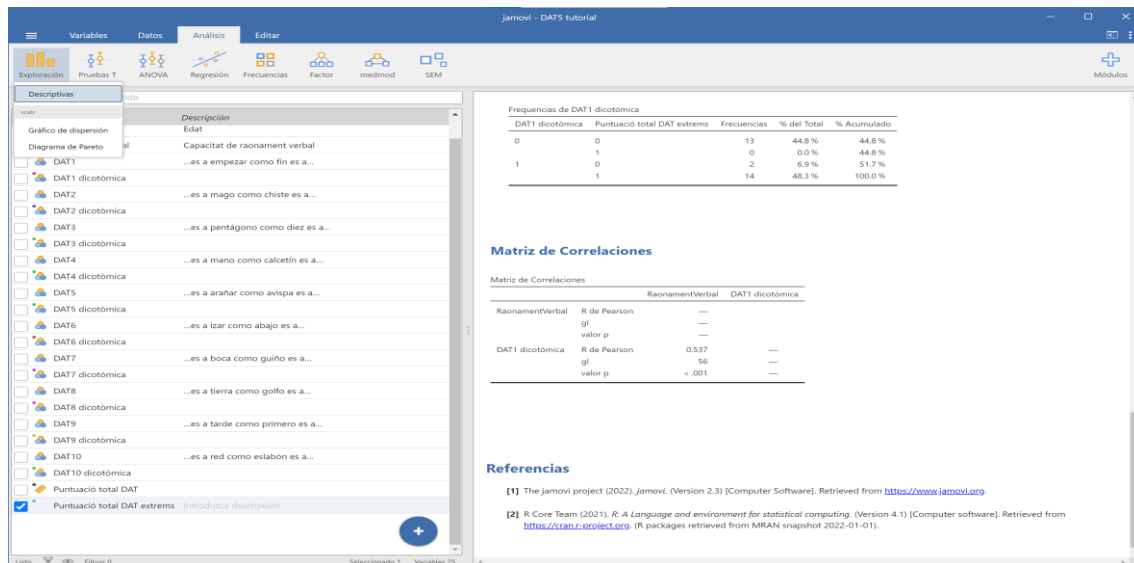
Com que en aquest cas, la correlació ítem-criteri extern pren un valor de 0,537 (mida mitjana) i és positiva i estadísticament significativa ($p < 0,050$), podem dir que l'ítem 1 té evidències de validesa.

Exemple de l'anàlisi de distractors per a l'ítem 1.

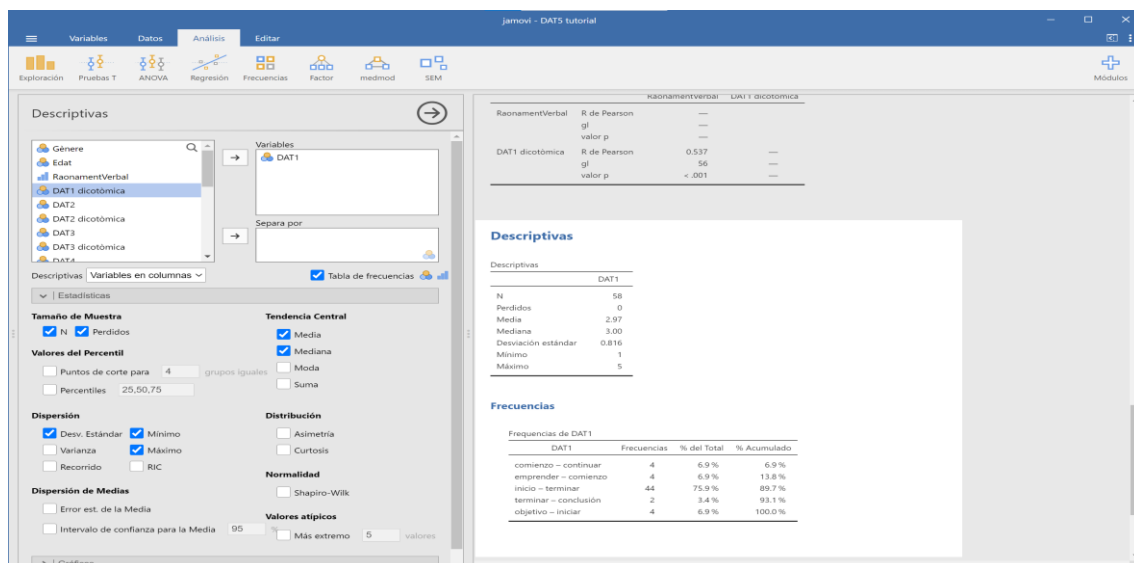
Per a analitzar els distractors, treballarem amb les variables originals (amb els ítems originals). Volem saber qui/quants han triat l'opció A (1), l'opció B (2), la C (3), la D (4) i la E (5).

Aquesta anàlisi de distractors inclou tres passos:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives, en aquest cas, de l'ítem 1. Cal anar a *Anàlisi – Exploració – Descriptives*:



I demanar la taula de freqüències per a l'ítem 1 (l'original, no el dicotòmic):



Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. De primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{4 + 4 + 44 + 2 + 4}{5} = \frac{58}{5} = 11,6$$

És a dir, si la distribució fora equiprobable entre totes les alternatives, trobaríem 11,6 subjectes que haurien triat la A, 11,6 haurien triat la B, 11,6 l'opció C, etc.

Amb les freqüències teòriques ja podem calcular la χ^2 i veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}
\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\
&= \frac{(11,6 - 4)^2 + (11,6 - 4)^2 + (11,6 - 44)^2 + (11,6 - 2)^2 + (11,6 - 4)^2}{11,6} \\
&= \frac{(7,6)^2 + (7,6)^2 + (-32,4)^2 + (9,6)^2 + (7,6)^2}{11,6} \\
&= \frac{57,76 + 57,76 + 1049,76 + 92,16 + 57,76}{11,6} = \frac{1315,2}{11,6} = 113,38
\end{aligned}$$

Com que el valor de χ^2 en la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 113,38$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa C, la correcta en l'ítem 1, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

De primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{4 + 4 + 2 + 4}{4} = \frac{14}{4} = 3,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 i veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}
\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(3,5 - 4)^2 + (3,5 - 4)^2 + (3,5 - 2)^2 + (3,5 - 4)^2}{3,5} \\
&= \frac{(0,5)^2 + (0,5)^2 + (1,5)^2 + (0,5)^2}{3,5} = \frac{0,25 + 0,25 + 2,25 + 0,25}{3,5} = \frac{3}{3,5} = 0,86
\end{aligned}$$

Com que el valor de χ^2 en la taula per a alfa = 0,05 i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,86$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa

correcta, i que no hi ha diferències entre els qui han triat les incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Caldria anar a *Anàlisi – ANOVA – ANOVA d'un factor*.

The screenshot shows the JAMOVI software interface. On the left, the 'Variables' list includes 'Gènere', 'Edat', 'RaonamentVerbal', 'DAT1 dicotòmica', 'DAT2', 'DAT2 dicotòmica', 'DAT3', and 'DAT3 dicotòmica'. The 'Anàlisi' menu is open, showing 'ANOVA de Un Factor' as the selected option. On the right, the 'Descriptivas' panel displays summary statistics for 'DAT1 dicotòmica'.

DAT1 dicotòmica	
N	58
Pèrdues	0
Media	0.759
Mediana	1.00
Desviación estándar	0.432
Mínimo	0
Máximo	1

Frecuencias de DAT1 dicotòmica			
DAT1 dicotòmica	Frecuencias	% del Total	% Acumulado
0	14	24.1 %	24.1 %
1	44	75.9 %	100.0 %

I seleccionem com a variable agrupadora l'ítem 1 original i com a variable dependent la *Puntuació total DAT*:

The screenshot shows the JAMOVI software interface with the 'ANOVA de Un Factor' configuration panel on the left and the results panel on the right. In the configuration panel, 'Puntuació total DAT' is set as the dependent variable and 'DAT1' as the grouping variable. The 'Varianzas' section has 'No asumir iguales (Welch)' and 'Asumir iguales (Fisher)' checked. The 'Estadísticas Adicionales' section has 'Tabla de descriptivas' and 'Gráficas descriptivas' checked. The 'Comprobaciones de Supuestos' section has 'Prueba de homogeneidad' and 'Prueba de Normalidad' checked. The 'Pruebas Post-Hoc' section has 'Games-Howell (varianzas diferentes)' selected. The results panel shows the ANOVA table and descriptive statistics for the groups.

	F	gl1	gl2	p
Puntuació total DAT	90.5	4	51	< .001

	DAT1	N	Media	DE	EE
Puntuació total DAT	comienzo – continuar	4	0.00	0.00	0.000
	emprender – comienzo	4	0.00	0.00	0.000
	inicio – terminar	42	8.48	1.45	0.224
	terminar – conclusión	2	7.00	2.83	2.000
	objetivo – iniciar	4	0.00	0.00	0.000

Prueba de Levene para homogeneidad de varianzas				
	F	gl1	gl2	p
Puntuació total DAT	8.31	4	51	< .001

Prueba Post-Hoc de Games-Howell – Puntuació total DAT							
	comienzo – continuar	emprender – comienzo	inicio – terminar	terminar – conclusión	objetivo		
comienzo – continuar							
	Diferencia de medias	—	0.00	-8.48	-7.00		
	valor p	—	NaN	< .001	NaN		

És important seleccionar les opcions següents:

- La *F* de Fisher i la *t* de Welch, i usar l'una o l'altra depenent de si es pressuposen o no variàncies iguals (prova de Levene).
- Estadístiques descriptives.

- Prova d'homogeneïtat (la prova de Levene).
- Proves *post hoc* de Games-Howell o de Tuckey depenent de si les variàncies són iguals o diferents.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

ANOVA de Un Factor

		F	gl1	gl2	p
Puntuació total DAT	t de Welch	NaN	4	NaN	NaN
	Fisher	90.5	4	51	< .001

Tal com es pot veure a la taula, la prova *t* de Welch no s'ha pogut calcular (NaN: *not a number*), però sí la *F*: $F(4, 51) = 90,5$ ($p < 0,001$), i és estadísticament significativa. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons siga l'opció de resposta que han triat.

En segon lloc, hauríem de mirar les proves *post hoc* per veure entre quins grups hi ha aquestes diferències:

Prueba Post-Hoc de Games-Howell – Puntuació total DAT

		comienzo – continuar	emprender – comienzo	inicio – terminar	terminar – conclusión	objetivo – iniciar
comienzo – continuar	Diferencia de medias	—	0.00	-8.48	-7.00	0.00
	valor p	—	NaN	< .001	NaN	NaN
emprender – comienzo	Diferencia de medias		—	-8.48	-7.00	0.00
	valor p		—	< .001	NaN	NaN
inicio – terminar	Diferencia de medias			—	1.48	8.48
	valor p			—	NaN	< .001
terminar – conclusión	Diferencia de medias				—	7.00
	valor p				—	NaN
objetivo – iniciar	Diferencia de medias					—
	valor p					—

Recordem, en primer lloc, que la resposta correcta és la C: “acabar – conclusió”. Busquem les proves *post hoc* d'aquesta opció, marcades a la taula.

Atesa la poca grandària dels grups, no s'han pogut fer totes les proves *post hoc*. Si s'hagueren pogut calcular, esperaríem diferències estadísticament significatives de les alternatives incorrectes respecte de la correcta. En aquest cas s'han observat aquestes diferències en els casos dels distractors A, B i E.

En el cas del distractor D, que només havien seleccionat dues persones, haurem de mirar la taula d'estadístiques descriptives a fi de consultar-ne les mitjanes:

Descriptivas de Grupo

	DAT1	N	Media	DE	EE
Puntuació total DAT	comienzo – continuar	4	0.00	0.00	0.000
	emprender – comienzo	4	0.00	0.00	0.000
	inicio – terminar	42	8.48	1.45	0.224
	terminar – conclusión	2	7.00	2.83	2.000
	objetivo – iniciar	4	0.00	0.00	0.000

Si ens fixem en les mitjanes dels distractors A, B i E, totes són més baixes que la mitjana de l'opció correcta, la C. La mitjana de l'opció D, distractor, també és més baixa que la C, però els subjectes tenen una puntuació elevada en el DAT (7,00 punts).

Podríem dir que, en general, els distractors de l'ítem 1 funcionen bé, excepte per l'alternativa D: s'assembla massa a l'opció correcta i confon els participants amb alt nivell de tret, de raonament verbal. Caldria revisar i modificar aquest distractor.

Anàlisi d'ítems. Solucions*Dificultat dels ítems*

Resultats de l'ítem 1:

Taula 1*Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 1.*

Resposta	Freqüència
0 - error	14
1 - encert	44

$$ID_1 = \frac{E_1}{N_1} = \frac{44}{58} = 0,76$$

$$ID_1^c = ID_1 - \frac{q_1}{K-1} = 0,76 - \frac{0,24}{5-1} = 0,76 - \frac{0,24}{4} = 0,76 - 0,06 = 0,70$$

Tenint en compte aquests resultats, podem dir que l'ítem 1 té una dificultat adequada, ja que és superior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986). **IMPORTANT: els punts de tall s'apliquen a l'ID, no a l'ID corregit*.*

Resultats de l'ítem 2:

Taula 2*Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 2.*

Resposta	Freqüència
0 - error	14
1 - encert	44

$$ID_2 = \frac{E_2}{N_2} = \frac{44}{58} = 0,76$$

$$ID_2^c = ID_2 - \frac{q_2}{K-1} = 0,76 - \frac{0,24}{5-1} = 0,76 - \frac{0,24}{4} = 0,76 - 0,06 = 0,70$$

Tenint en compte aquests resultats, podem dir que l'ítem 2 té una dificultat adequada, ja que és superior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 3:

Taula 3

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 3.

Resposta	Freqüència
0 - error	14
1 - encert	44

$$ID_3 = \frac{E_j}{N_j} = \frac{44}{58} = 0,76$$

$$ID_3^c = ID_3 - \frac{q_3}{K-1} = 0,76 - \frac{0,24}{5-1} = 0,76 - \frac{0,24}{4} = 0,76 - 0,06 = 0,70$$

En vista dels resultats, podem dir que l'ítem 3 té una dificultat adequada perquè és superior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 4:

Taula 4

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 4.

Resposta	Freqüència
0 - error	18
1 - encert	40

$$ID_4 = \frac{E_4}{N_4} = \frac{40}{58} = 0,69$$

$$ID_4^c = ID_4 - \frac{q_4}{K-1} = 0,69 - \frac{0,31}{5-1} = 0,69 - \frac{0,31}{4} = 0,69 - 0,08 = 0,61$$

En vista dels resultats, podem dir que l'ítem 4 té una dificultat adequada perquè és superior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 5:

Taula 5

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 5.

Resposta	Freqüència
0 - error	24
1 - encert	34

$$ID_5 = \frac{E_5}{N_5} = \frac{34}{58} = 0,59$$

$$ID_5^c = ID_5 - \frac{q_5}{K-1} = 0,59 - \frac{0,41}{5-1} = 0,59 - \frac{0,41}{4} = 0,59 - 0,10 = 0,49$$

En vista dels resultats, podem dir que l'ítem 5 no té una dificultat adequada perquè és inferior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 6:

Taula 6

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 6.

Resposta	Freqüència
0 - error	22
1 - encert	36

$$ID_6 = \frac{E_6}{N_6} = \frac{36}{58} = 0,62$$

$$ID_6^c = ID_6 - \frac{q_6}{K-1} = 0,62 - \frac{0,38}{5-1} = 0,62 - \frac{0,38}{4} = 0,62 - 0,10 = 0,52$$

En vista dels resultats, podem dir que l'ítem 6 té una dificultat adequada perquè és igual a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 8:

Taula 7

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 7.

Resposta	Freqüència
0 - error	19
1 - encert	38

$$ID_7 = \frac{E_7}{N_7} = \frac{38}{58} = 0,66$$

$$ID_7^c = ID_7 - \frac{q_7}{K-1} = 0,66 - \frac{0,34}{5-1} = 0,66 - \frac{0,34}{4} = 0,66 - 0,09 = 0,57$$

En vista dels resultats, podem dir que l'ítem 7 té una dificultat adequada perquè és superior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 8:

Taula 8

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 8.

Resposta	Freqüència
0 - error	24
1 - encert	33

$$ID_8 = \frac{E_8}{N_8} = \frac{33}{58} = 0,57$$

$$ID_8^c = ID_8 - \frac{q_8}{K-1} = 0,57 - \frac{0,43}{5-1} = 0,57 - \frac{0,43}{4} = 0,57 - 0,11 = 0,46$$

En vista dels resultats, podem dir que l'ítem 8 no té una dificultat adequada perquè és inferior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 9:

Taula 9

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 9.

Resposta	Freqüència
0 - error	22
1 - encert	34

$$ID_9 = \frac{E_9}{N_9} = \frac{34}{58} = 0,59$$

$$ID_9^c = ID_9 - \frac{q_9}{K-1} = 0,59 - \frac{0,41}{5-1} = 0,59 - \frac{0,41}{4} = 0,59 - 0,10 = 0,49$$

En vista dels resultats, podem dir que l'ítem 9 no té una dificultat adequada perquè és inferior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Resultats de l'ítem 10:

Taula 10

Taula de freqüències per a les alternatives de resposta correcta/incorrecta de l'ítem 10.

Resposta	Freqüència
0 - error	22
1 - encert	34

$$ID_{10} = \frac{E_{10}}{N_{10}} = \frac{34}{58} = 0,59$$

$$ID_{10}^c = ID_{10} - \frac{q_{10}}{K-1} = 0,59 - \frac{0,41}{5-1} = 0,59 - \frac{0,41}{4} = 0,59 - 0,10 = 0,49$$

En vista dels resultats, podem dir que l'ítem 10 no té una dificultat adequada perquè és inferior a 0,62, el punt de tall recomanat per a quatre o més alternatives de resposta (Crocker i Algina, 1986).

Discriminació dels ítems

Resultats de l'ítem 1:

Taula 11

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 1.

Resposta a l'ítem 1	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	13
1 - encert	1 - grup superior	14
	0 - grup inferior	2

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{2}{15} = 0,13$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 1:

$$D_1 = p_s - p_i = 1 - 0,13 = 0,87$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 1 discrimina molt bé perquè supera el punt de tall de 0,40.

Resultats de l'ítem 2:

Taula 12

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 2.

Resposta a l'ítem 2	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	13
1 - encert	1 - grup superior	14
	0 - grup inferior	2

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{2}{15} = 0,13$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 2:

$$D_2 = p_s - p_i = 1 - 0,13 = 0,87$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 2 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 3:

Taula 13

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 3.

Resposta a l'ítem 3	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	13
1 - encert	1 - grup superior	14
	0 - grup inferior	2

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{2}{15} = 0,13$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 3:

$$D_3 = p_s - p_i = 1 - 0,13 = 0,87$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 3 discrimina molt bé perquè supera el punt de tall de 0,40.

Resultats de l'ítem 4:

Taula 14

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 4.

Resposta a l'ítem 4	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	14
1 - encert	1 - grup superior	14
	0 - grup inferior	1

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{1}{15} = 0,07$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 4:

$$D_4 = p_s - p_i = 1 - 0,07 = 0,93$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 4 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 5:

Taula 15

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 5.

Resposta a l'ítem 5	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	14
1 - encert	1 - grup superior	14
	0 - grup inferior	1

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{1}{15} = 0,07$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 5:

$$D_5 = p_s - p_i = 1 - 0,07 = 0,93$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 5 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 6:

Taula 16

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 6.

Resposta a l'ítem 6	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	13
1 - encert	1 - grup superior	14
	0 - grup inferior	2

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{2}{15} = 0,13$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 6:

$$D_6 = p_s - p_i = 1 - 0,13 = 0,87$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 6 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 7:

Taula 17

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 7.

Resposta a l'ítem 7	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	13
1 - encert	1 - grup superior	14
	0 - grup inferior	2

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{2}{15} = 0,13$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 7:

$$D_7 = p_s - p_i = 1 - 0,13 = 0,87$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 7 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 8:

Taula 18

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 8.

Resposta a l'ítem 8	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	14
1 - encert	1 - grup superior	14
	0 - grup inferior	1

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{1}{15} = 0,07$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 8:

$$D_8 = p_s - p_i = 1 - 0,07 = 0,93$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 8 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 9:

Taula 19

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 9.

Resposta a l'ítem 5	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	14
1 - encert	1 - grup superior	14
	0 - grup inferior	1

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{1}{15} = 0,07$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 9:

$$D_9 = p_s - p_i = 1 - 0,07 = 0,93$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 9 discrimina molt bé, ja que supera el punt de tall de 0,40.

Resultats de l'ítem 10:

Taula 20

Freqüències d'encerts i errors per als grups extrems (inferior i superior) en l'ítem 10.

Resposta a l'ítem 10	Puntuació als extrems	Freqüències
0 - error	1 - grup superior	0
	0 - grup inferior	14
1 - encert	1 - grup superior	14
	0 - grup inferior	1

Proporció d'encerts en el grup inferior (p_i):

$$p_i = \frac{1}{15} = 0,07$$

Proporció d'encerts del grup superior (p_s):

$$p_s = \frac{14}{14} = 1$$

Índex de discriminació (D) per a l'ítem 10:

$$D_{10} = p_s - p_i = 1 - 0,07 = 0,93$$

Prenent com a referència els punts de tall establits per Ebel (1965), l'ítem 10 discrimina molt bé, ja que supera el punt de tall de 0,40.

Homogeneïtat dels ítems

Per a calcular l'índex d'homogeneïtat (IH) dels ítems és necessari correlacionar-los (en la versió dicotòmica, on tenim recodificats els subjectes que l'han encertat en 1 i els qui l'han fallat en 0) amb la puntuació total en el DAT (l'original, no la que classifica alts i baixos).

Tots els ítems tenen evidències d'homogeneïtat, ja que la correlació ítem-total supera el punt de tall de 0,30. L'ítem més homogeni és l'ítem 2, amb el valor de la correlació més alt, i el menys homogeni és l'ítem 9, amb el valor més baix per a la correlació ítem-total o índex d'homogeneïtat.

Taula 21

Matriu de correlacions entre la puntuació total en el DAT i els ítems recodificats.

Ítem	Correlació amb el DAT total
1	0,876***
2	0,898***
3	0,887***
4	0,854***
5	0,738***
6	0,756***
7	0,785***
8	0,708***
9	0,706***
10	0,735***

Notes: *** $p < 0,001$

Validesa dels ítems

Per a calcular l'índex de validesa (IV) és necessari correlacionar els ítems (en la versió dicotòmica, on tenim recodificats els subjectes que l'han encertat en 1 i els qui l'han fallat en 0) amb un altre criteri extern i distint del test. En aquest cas, a la base de dades tenim una mesura de raonament verbal percebut: si els ítems mesuren raonament verbal, considerem que els subjectes que traguin més puntuació en l'ítem han de ser també els qui es perceben com a més competents en raonament verbal.

Tots els ítems tenen evidències de validesa (taula 22), ja que la correlació entre els ítems i el criteri ha sigut positiva i estadísticament significativa, tal com s'esperava. L'ítem amb més evidències de validesa és l'ítem 7, amb el valor de la correlació més alt, i el menys homogeni l'ítem 8, amb el valor més baix per a l'índex de validesa.

Taula 22

Matriu de correlacions entre la puntuació en raonament verbal i els ítems recodificats.

Ítem	Correlació amb el raonament verbal
1	0,537***
2	0,537***
3	0,537***
4	0,345**
5	0,431***
6	0,324*
7	0,615***
8	0,290*
9	0,339**
10	0,291*

Notes: * $p < 0,050$; ** $p < 0,010$; *** $p < 0,001$.

Anàlisi de distractors

Resultats de l'ítem 1:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 1.

Taula 23

Freqüències de resposta per a l'ítem 1.

Alternativa de resposta	<i>n</i>
Opció A	4
Opció B	4
Opció C	44
Opció D	2
Opció E	4

Nota: l'alternativa de resposta correcta és l'opció C.

Amb aquestes dades ja podem calcular el χ^2 amb totes les alternatives. De primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{4 + 4 + 44 + 2 + 4}{5} = \frac{58}{5} = 11,6$$

Amb les freqüències teòriques ja podem calcular el χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\ &= \frac{(11,6 - 4)^2 + (11,6 - 4)^2 + (11,6 - 44)^2 + (11,6 - 2)^2 + (11,6 - 4)^2}{11,6} \\ &= \frac{(7,6)^2 + (7,6)^2 + (-32,4)^2 + (9,6)^2 + (7,6)^2}{11,6} \\ &= \frac{57,76 + 57,76 + 1049,76 + 92,16 + 57,76}{11,6} = \frac{1315,2}{11,6} = 113,38\end{aligned}$$

Com que el valor de χ^2 en la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 113,38$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa C, la correcta en l'ítem 1, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat per comprovar si els distractors són equiprobables: si els selecciona un nombre similar de persones. Així, efectuem una prova khi quadrat només amb els distractors.

De primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{4 + 4 + 2 + 4}{4} = \frac{14}{4} = 3,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(3,5 - 4)^2 + (3,5 - 4)^2 + (3,5 - 2)^2 + (3,5 - 4)^2}{3,5} \\ &= \frac{(-0,5)^2 + (-0,5)^2 + (1,5)^2 + (-0,5)^2}{3,5} = \frac{0,25 + 0,25 + 2,25 + 0,25}{3,5} = \frac{3}{3,5} \\ &= 0,86\end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,3)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,86$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) en què emprem com a variable agrupadora l'ítem original, i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 24

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 1.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	NaN	4	NaN	NaN
	Fisher	90,5	4	51	< 0,001

Tal com es pot veure a la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 90,5$ ($p < 0,001$), i és estadísticament significativa. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han triat.

En segon lloc, hauríem d'examinar les proves *post hoc* per veure entre quins grups hi ha aquestes diferències. Com que el resultat de la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 8,31$, $p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 25

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 1.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	0,00	−8,48	−7,00	0,00
	Valor p	—	NaN	<0,001	NaN	NaN
Opció B	Diferència de mitjanes		—	−8,48	−7,00	0,00
	Valor p		—	<0,001	NaN	NaN
Opció C	Diferència de mitjanes			—	1,48	8,48
	Valor p			—	NaN	<0,001
Opció D	Diferència de mitjanes				—	7,00
	Valor p				—	NaN
Opció E	Diferència de mitjanes					—
	Valor p					—

Com que l'opció correcta és la C, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Atesa la poca grandària dels grups, no s'han pogut fer totes les proves *post hoc*. Si s'hagueren pogut calcular, esperaríem diferències estadísticament significatives en les alternatives incorrectes respecte de la correcta. En aquest cas s'han observat aquestes diferències en els casos dels distractors A, B i E.

En el cas del distractor D, que només havien seleccionat dues persones, hem de mirar la taula d'estadístiques descriptives a fi de consultar la informació de les mitjanes:

Taula 26

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 1.

	N	Mitjana	DE	EE
Opció A	4	0,00	0,00	0,000
Opció B	4	0,00	0,00	0,000
Opció C	42	8,48	1,45	0,224
Opció D	2	7,00	2,83	2,000
Opció E	4	0,00	0,00	0,000

Notes: *DE* = desviació estàndard; *EE* = error típic d'estimació.

Si ens fixem en les mitjanes dels distractors A, B i E, totes són més baixes que la mitjana de l'opció correcta, la C. La mitjana de l'opció D, distractor, també és més baixa que la C, però els subjectes tenen una puntuació elevada en el DAT (7.00 punts).

Podríem dir que, en general, els distractors de l'ítem 1 funcionen bé, excepte per l'alternativa D: s'assembla massa a l'opció correcta i confon els participants amb alt nivell de tret de raonament verbal. Caldria revisar i modificar aquest distractor.

Resultats de l'ítem 2:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 2.

Taula 27

Freqüències de resposta per a l'ítem 2.

Alternativa de resposta	n
Opció A	4
Opció B	4
Opció C	3
Opció D	3
Opció E	44

Nota: l'alternativa de resposta correcta és l'opció E.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. De primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{4 + 4 + 3 + 3 + 44}{5} = \frac{58}{5} = 11,6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\ &= \frac{(11,6 - 4)^2 + (11,6 - 4)^2 + (11,6 - 3)^2 + (11,6 - 3)^2 + (11,6 - 44)^2}{11,6} \\ &= \frac{(7,6)^2 + (7,6)^2 + (8,6)^2 + (8,6)^2 + (-32,4)^2}{11,6} \\ &= \frac{57,76 + 57,76 + 73,96 + 73,96 + 1049,76}{11,6} = \frac{1313,2}{11,6} = 113,21\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 113,21$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa E, la correcta en l'ítem 2, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat per comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

De primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{4 + 4 + 3 + 3}{4} = \frac{14}{4} = 3,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(3,5 - 4)^2 + (3,5 - 4)^2 + (3,5 - 3)^2 + (3,5 - 3)^2}{3,5} \\ &= \frac{(-0,5)^2 + (-0,5)^2 + (+0,5)^2 + (+0,5)^2}{3,5} = \frac{0,25 + 0,25 + 0,25 + 0,25}{3,5} = \frac{1}{3,5} \\ &= 0,29\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,29$) és inferior al de la taula, acceptem la hipòtesi

nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 28

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 2.

		F	gl1	gl2	p
Puntuació DAT total	t de Welch	NaN	4	NaN	NaN
	Fisher	57,0	4	51	<0,001

Tal com es pot veure a la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 57,0$ ($p < 0,001$), i el resultat és estadísticament significatiu. Això indicaria que hi ha diferències entre les mitjanes dels subjectes depenent de l'opció de resposta que han triat.

En segon lloc, hauríem de mirar les proves *post hoc* per veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies ha donat un resultat estadísticament significatiu ($F(4, 51) = 6,07$, $p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 29

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 2.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	−0,500	1,25	1,25	−7,27
	Valor <i>p</i>	—	0,999	0,843	0,843	0,036
Opció B	Diferència de mitjanes		—	1,75	1,75	−6,77
	Valor <i>p</i>		—	0,843	0,843	0,110
Opció C	Diferència de mitjanes			—	0,00	−8,52
	Valor <i>p</i>			—	NaN	<0,001
Opció D	Diferència de mitjanes				—	−8,52
	Valor <i>p</i>				—	<0,001
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la E, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb tots els distractors, excepte amb l'opció incorrecta B.

En el cas del distractor B, si ens fixem en la taula d'estadístiques descriptives, observem que es tracta d'un distractor amb una mitjana de 0; sembla, per tant, que funciona correctament:

Taula 30

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 2.

	N	Mitjana	DE	EE
Opció A	4	1,25	2,50	1,250
Opció B	4	1,75	3,50	1,750
Opció C	3	0,00	0,00	0,000
Opció D	3	0,00	0,00	0,000
Opció E	42	8,52	1,44	0,221

Notes: DE = desviació estàndard; EE = error típic d'estimació.

Finalment, com que no es detecten diferències estadísticament significatives entre els distractors, podem dir que tant l'alternativa correcta com els distractors de l'ítem 2 funcionen correctament.

Resultats de l'ítem 3:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 3.

Taula 31

Freqüències de resposta per a l'ítem 3.

Alternativa de resposta	<i>n</i>
Opció A	4
Opció B	44
Opció C	3
Opció D	4
Opció E	3

Nota: l'alternativa de resposta correcta és l'opció B.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. De primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{4 + 44 + 3 + 4 + 3}{5} = \frac{58}{5} = 11,6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}
 \chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\
 &= \frac{(11,6 - 4)^2 + (11,6 - 44)^2 + (11,6 - 3)^2 + (11,6 - 4)^2 + (11,6 - 3)^2}{11,6} \\
 &= \frac{(7,6)^2 + (-32,4)^2 + (8,6)^2 + (7,6)^2 + (8,6)^2}{11,6} \\
 &= \frac{57,76 + 1049,76 + 73,96 + 57,76 + 73,96}{11,6} = \frac{1313,2}{11,6} = 113,21
 \end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2

observat ($\chi^2 = 113,21$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa B, la correcta en l'ítem 3, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat per comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

De primer, calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{4 + 3 + 4 + 3}{4} = \frac{14}{4} = 3,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(3,5 - 4)^2 + (3,5 - 3)^2 + (3,5 - 4)^2 + (3,5 - 3)^2}{3,5} \\ &= \frac{(-0,5)^2 + (0,5)^2 + (-0,5)^2 + (0,5)^2}{3,5} = \frac{0,25 + 0,25 + 0,25 + 0,25}{3,5} = \frac{1}{3,5} \\ &= 0,86\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,29$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 32

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 3.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	NaN	4	NaN	NaN
	Fisher	50,9	4	51	<0,001

Tal com es mostra a la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 50,9$ ($p < 0,001$), i el resultat és estadísticament significatiu. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han triat.

En segon lloc, hauríem de mirar les proves *post hoc* per veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 7,18$, $p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 33

Proves *post hoc* de comparació de mitjanes en els grups de respostes de l'ítem 3.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	-6,50	2,00	0,750	2,00
	Valor <i>p</i>	—	0,168	0,843	0,997	0,843
Opció B	Diferència de mitjanes		—	8,50	7,250	8,50
	Valor <i>p</i>		—	<0,001	0,036	<0,001
Opció C	Diferència de mitjanes			—	-1,250	0,00
	Valor <i>p</i>			—	0,843	NaN
Opció D	Diferència de mitjanes				—	1,25
	Valor <i>p</i>				—	0,843
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la B, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb tots els distractors, excepte amb l'opció incorrecta A.

En el cas del distractor A, si ens fixem en la taula d'estadístiques descriptives, veiem que es tracta d'un distractor amb una mitjana de 2,00, i només l'han escollit quatre subjectes. Pe tant, sembla que funciona correctament, i l'absència de diferències estadísticament significatives es deu a la mida mostral insuficient:

Taula 34

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 3.

	<i>N</i>	Mitjana	<i>DE</i>	<i>EE</i>
Opció A	4	2,00	4,00	2,000
Opció B	42	8,50	1,45	0,224
Opció C	3	0,00	0,00	0,000
Opció D	4	1,25	2,50	1,250
Opció E	3	0,00	0,00	0,000

Notes: *DE* = desviació estàndard; *EE* = error típic d'estimació.

Finalment, com que no s'observen diferències estadísticament significatives entre els distractors, podem dir que tant l'alternativa correcta com els distractors de l'ítem 3 funcionen correctament.

Resultats de l'ítem 4:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 4.

Taula 35

Freqüències de resposta per a l'ítem 4.

Alternativa de resposta	<i>n</i>
Opció A	5
Opció B	40
Opció C	4
Opció D	5
Opció E	4

Nota: l'alternativa de resposta correcta és l'opció B.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. De primer, calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{5 + 40 + 4 + 5 + 4}{5} = \frac{58}{5} = 11,6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}
 \chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\
 &= \frac{(11,6 - 5)^2 + (11,6 - 40)^2 + (11,6 - 4)^2 + (11,6 - 5)^2 + (11,6 - 4)^2}{11,6} \\
 &= \frac{(6,6)^2 + (-28,4)^2 + (7,6)^2 + (6,6)^2 + (7,6)^2}{11,6} \\
 &= \frac{43,56 + 806,56 + 57,76 + 43,56 + 57,76}{11,6} = \frac{1009,2}{11,6} = 87
 \end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 87$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa B, la correcta en l'ítem 4, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

De primer, calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{5 + 4 + 5 + 4}{4} = \frac{18}{4} = 4,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 per veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(4,5 - 5)^2 + (4,5 - 4)^2 + (4,5 - 5)^2 + (4,5 - 4)^2}{4,5} \\ &= \frac{(-0,5)^2 + (0,5)^2 + (-0,5)^2 + (0,5)^2}{4,5} = \frac{0,25 + 0,25 + 0,25 + 0,25}{4,5} = \frac{1}{4,5} \\ &= 0,22\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,3-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,22$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 36

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 4.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	NaN	4	NaN	NaN
	Fisher	49,6	4	51	<0,001

Tal com mostra la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 49,6$ ($p < 0,001$), i el resultat és estadísticament significatiu. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem d'examinar les proves *post hoc* per veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 22,8, p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 37

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 4.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	−5,29	3,40	−0,100	3,40
	Valor <i>p</i>	—	0,088	0,286	1,000	0,286
Opció B	Diferència de mitjanes		—	8,69	5,192	8,69
	Valor <i>p</i>		—	<0,001	0,288	<0,001
Opció C	Diferència de mitjanes			—	−3,500	0,00
	Valor <i>p</i>			—	0,540	NaN
Opció D	Diferència de mitjanes				—	3,50
	Valor <i>p</i>				—	0,540
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la B, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb tots els distractors, excepte amb l'opció incorrecta D.

En el cas del distractor D, si ens fixem en la taula d'estadístiques descriptives, veiem que es tracta d'un distractor amb una mitjana de 3,50, i només l'han escollit quatre subjectes, per la qual cosa sembla que, com abans, funciona correctament i que l'absència de diferències estadísticament significatives es deu a la mida mostral insuficient:

Taula 38

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 4.

	N	Mitjana	DE	EE
Opció A	5	3,40	3,21	1,435
Opció B	39	8,69	1,30	0,208
Opció C	4	0,00	0,00	0,000
Opció D	4	3,50	4,12	2,062
Opció E	4	0,00	0,00	0,000

Notes: *DE* = desviació estàndard; *EE* = error típic d'estimació.

Finalment, com que no s'observen diferències estadísticament significatives entre els distractors, podem dir que tant l'alternativa correcta com els distractors de l'ítem 4 funcionen correctament.

Resultats de l'ítem 5:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 5.

Taula 39

Freqüències de resposta per a l'ítem 5.

Alternativa de resposta	n
Opció A	9
Opció B	34
Opció C	5
Opció D	6
Opció E	4

Nota: l'alternativa de resposta correcta és l'opció B.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. Primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{9 + 34 + 5 + 6 + 4}{5} = \frac{58}{5} = 11,6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\ &= \frac{(11,6 - 9)^2 + (11,6 - 34)^2 + (11,6 - 5)^2 + (11,6 - 6)^2 + (11,6 - 4)^2}{11,6} \\ &= \frac{(2,6)^2 + (-22,4)^2 + (6,6)^2 + (5,6)^2 + (7,6)^2}{11,6} \\ &= \frac{6,76 + 501,76 + 43,56 + 31,36 + 57,76}{11,6} = \frac{641,2}{11,6} = 55,28\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 55,28$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa B, la correcta en l'ítem 5, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

De primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{9 + 5 + 6 + 4}{4} = \frac{24}{4} = 6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(6 - 9)^2 + (6 - 5)^2 + (6 - 6)^2 + (6 - 4)^2}{6} \\ &= \frac{(-3)^2 + (1)^2 + (0)^2 + (2)^2}{6} = \frac{9 + 1 + 0 + 4}{6} = \frac{14}{6} = 2,33\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 2,33$) és inferior al de la taula, acceptem la hipòtesi

nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 40

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 5.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	NaN	4	NaN	NaN
	Fisher	48,8	4	51	<0,001

Tal com es pot veure a la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 48,8$ ($p < 0,001$), i ha resultat estadísticament significativa. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem de mirar les proves *post hoc* a fi de saber entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 8,65$, $p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 41

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 5.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	−1,97	5,20	5,1667	7,00
	Valor <i>p</i>	—	0,001	0,178	0,031	<0,001
Opció B	Diferència de mitjanes		—	7,17	7,1354	8,97
	Valor <i>p</i>		—	0,070	0,008	<0,001
Opció C	Diferència de mitjanes			—	−0,0333	1,80
	Valor <i>p</i>			—	1,000	0,844
Opció D	Diferència de mitjanes				—	1,83
	Valor <i>p</i>				—	0,567
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la B, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb tots els distractors, excepte amb l'opció incorrecta C, amb la qual les diferències són marginalment significatives ($p = 0,070$).

En el cas del distractor C, si ens fixem en la taula d'estadístiques descriptives, veiem que es tracta d'un distractor amb una mitjana d'1,80, i només l'han escollit cinc subjectes. Per tant, sembla que, com abans, funciona correctament i l'absència de diferències estadísticament significatives es deu a la mida mostral insuficient:

Taula 42*Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 5.*

	N	Mitjana	DE	EE
Opció A	9	7,00	1,00	0,333
Opció B	32	8,97	1,20	0,213
Opció C	5	1,80	4,02	1,800
Opció D	6	1,83	2,86	1,167
Opció E	4	0,00	0,00	0,000

Notes: DE = desviació estàndard; EE = error típic d'estimació.

El problema, però, és que sí que es detecten diferències estadísticament significatives entre els distractors: l'opció incorrecta A mostra diferències estadísticament significatives amb les opcions D i E. De fet, si observem la taula de les estadístiques descriptives, la mitjana en el distractor A és molt superior a les obtingudes en els distractors C, D i E, i molt similar a la de l'opció correcta B. Per tant, podem dir que aquest distractor no funciona correctament.

Resultats de l'ítem 6:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 6.

Taula 43*Freqüències de resposta per a l'ítem 6.*

Alternativa de resposta	n
Opció A	7
Opció B	4
Opció C	36
Opció D	5
Opció E	6

Nota: l'alternativa de resposta correcta és l'opció C.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. De primer, calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{7 + 4 + 36 + 5 + 6}{5} = \frac{58}{5} = 11,6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\ &= \frac{(11,6 - 7)^2 + (11,6 - 4)^2 + (11,6 - 36)^2 + (11,6 - 5)^2 + (11,6 - 6)^2}{11,6} \\ &= \frac{(4,6)^2 + (7,6)^2 + (-24,4)^2 + (6,6)^2 + (5,6)^2}{11,6} \\ &= \frac{21,16 + 57,76 + 595,36 + 43,56 + 31,36}{11,6} = \frac{749,2}{11,6} = 64,59\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 64,59$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa C, la correcta en l'ítem 6, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

Primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{7 + 4 + 5 + 6}{4} = \frac{22}{4} = 5,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de saber si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(5,5 - 7)^2 + (5,5 - 4)^2 + (5,5 - 5)^2 + (5,5 - 6)^2}{5,5} \\ &= \frac{(-1,5)^2 + (1,5)^2 + (0,5)^2 + (-0,5)^2}{5,5} = \frac{2,25 + 2,25 + 0,25 + 0,25}{5,5} = \frac{5}{5,5} \\ &= 0,91\end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,3-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,91$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 44

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 6.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	NaN	4	NaN	NaN
	Fisher	40,8	4	51	<0,001

Tal com es pot veure a la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 40,8$ ($p < 0,001$), i resulta estadísticament significativa. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem d'examinar les proves *post hoc* a fi de veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 7,83$, $p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 45

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 6.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	2,17	-6,60	0,567	-4,83
	Valor <i>p</i>	—	0,566	0,023	0,999	0,080
Opció B	Diferència de mitjanes		—	-8,77	-1,600	-7,00
	Valor <i>p</i>		—	<0,001	0,844	<0,001
Opció C	Diferència de mitjanes			—	7,171	1,77
	Valor <i>p</i>			—	0,047	0,129
Opció D	Diferència de mitjanes				—	-5,40
	Valor <i>p</i>				—	0,112
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la C, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb tots els distractors, excepte amb l'opció incorrecta E.

En el cas del distractor E, si ens fixem en la taula d'estadístiques descriptives, veiem que es tracta d'un distractor amb una mitjana de 7,00 punts, molt similar a la de l'opció correcta C, que és de 8,77 punts.

Taula 46

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 6.

	N	Mitjana	DE	EE
Opció A	6	2,17	3,37	1,376
Opció B	4	0,00	0,00	0,000
Opció C	35	8,77	1,37	0,232
Opció D	5	1,60	3,58	1,600
Opció E	6	7,00	1,41	0,577

Notes: DE = desviació estàndard; EE = error típic d'estimació.

Quan s'estudien les diferències entre els distractors torna a presentar-se el problema amb el distractor E: no mostra diferències estadísticament significatives entre els distractors. Per tant, podem dir que aquest distractor no funciona correctament.

Resultats de l'ítem 7:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 7.

Taula 47

Freqüències de resposta per a l'ítem 7.

Alternativa de resposta	<i>n</i>
Opció A	38
Opció B	5
Opció C	4
Opció D	6
Opció E	4

Nota: l'alternativa de resposta correcta és l'opció A.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. Primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{38 + 5 + 4 + 6 + 4}{5} = \frac{57}{5} = 11,4$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}
 \chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\
 &= \frac{(11,4 - 38)^2 + (11,4 - 5)^2 + (11,4 - 4)^2 + (11,4 - 6)^2 + (11,4 - 4)^2}{11,4} \\
 &= \frac{(-26,6)^2 + (6,4)^2 + (7,4)^2 + (5,4)^2 + (7,4)^2}{11,4} \\
 &= \frac{707,56 + 40,96 + 54,76 + 29,16 + 54,76}{11,4} = \frac{887,2}{11,4} = 77,82
 \end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 77,82$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa A, la correcta en l'ítem 7, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

Primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{5 + 4 + 6 + 4}{4} = \frac{19}{4} = 4,75$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(4,75 - 5)^2 + (4,75 - 4)^2 + (4,75 - 6)^2 + (4,75 - 4)^2}{4,75} \\ &= \frac{(-0,25)^2 + (0,75)^2 + (-1,25)^2 + (0,75)^2}{4,75} = \frac{0,06 + 0,56 + 1,56 + 0,56}{4,75} \\ &= \frac{2,74}{4,75} = 0,58\end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,58$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 48

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 7.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	NaN	4	NaN	NaN
	Fisher	33,7	4	51	<0,001

Tal com es veu a la taula, la prova *t* de Welch no s'ha pogut calcular, però sí la *F*: $F(4, 51) = 33,7$ ($p < 0,001$), i resulta estadísticament significativa. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem de mirar les proves *post hoc* a fi de veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 10,4$, $p < 0,001$), calculem les proves *post hoc* de Games-Howell:

Taula 49

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 7.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	7,21	8,61	5,94	1,94
	Valor <i>p</i>	—	0,028	<0,001	0,079	0,424
Opció B	Diferència de mitjanes		—	1,40	-1,27	-5,27
	Valor <i>p</i>		—	0,844	0,975	0,096
Opció C	Diferència de mitjanes			—	-2,67	-6,67
	Valor <i>p</i>			—	0,563	0,052
Opció D	Diferència de mitjanes				—	-4,00
	Valor <i>p</i>				—	0,317
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la A, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb els distractors B i C, marginalment significatives amb el D ($p = 0,079$) i estadísticament no significatives amb el distractor E ($p = 0,424$).

En el cas del distractor D, la mitjana és de 2,67, i en el cas del distractor E, la mitjana és de 6,67, enfront de la de l'opció correcta A, que és de 8,61 punts.

Taula 50

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 7.

	N	Mitjana	DE	EE
Opció A	38	8,61	1,48	0,240
Opció B	5	1,40	3,13	1,400
Opció C	4	0,00	0,00	0,000
Opció D	6	2,67	4,13	1,687
Opció E	3	6,67	1,53	0,882

Notes: DE = desviació estàndard; EE = error típic d'estimació.

Si estudiem les diferències entre els distractors, torna a presentar-se el problema amb el distractor E: mostra diferències marginalment significatives amb el distractor C ($p = 0,052$). Per tant, podem dir que aquest distractor no funciona correctament.

Resultats de l'ítem 8:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 8.

Taula 51

Freqüències de resposta per a l'ítem 8.

Alternativa de resposta	<i>n</i>
Opció A	33
Opció B	2
Opció C	8
Opció D	9
Opció E	5

Nota: l'alternativa de resposta correcta és l'opció A.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. Primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{33 + 2 + 8 + 9 + 5}{5} = \frac{57}{5} = 11,4$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}
 \chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\
 &= \frac{(11,4 - 33)^2 + (11,4 - 2)^2 + (11,4 - 8)^2 + (11,4 - 9)^2 + (11,4 - 5)^2}{11,4} \\
 &= \frac{(-21,6)^2 + (9,4)^2 + (3,4)^2 + (2,4)^2 + (6,4)^2}{11,4} \\
 &= \frac{466,56 + 88,36 + 11,56 + 5,76 + 40,96}{11,4} = \frac{613,2}{11,4} = 53,79
 \end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 53,79$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa A, la correcta en l'ítem 8, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

Primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{2 + 8 + 9 + 5}{4} = \frac{24}{4} = 6$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(6-2)^2 + (6-8)^2 + (6-9)^2 + (6-5)^2}{6} \\ &= \frac{(4)^2 + (-2)^2 + (-3)^2 + (1)^2}{6} = \frac{16 + 4 + 9 + 1}{6} = \frac{30}{6} = 5\end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 5$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 52

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 8.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	7,98	4	5,39	0,018
	Fisher	14,9	4	51	<0,001

Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 16,0, p < 0,001$), haurem de fixar-nos en la prova *t* de Welch, que aquesta vegada sí que s'ha pogut calcular. Tal com es pot veure a la taula, la prova *t* de Welch ha resultat estadísticament significativa: $F(4, 5,29) = 7,98$ ($p = 0,018$). Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem d'examinar les proves *post hoc* a fi de veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies hem dit que ha resultat estadísticament significativa, calculem les proves *post hoc* de Games-Howell:

Taula 53

Proves *post hoc* de comparació de mitjanes en els grups de respostes de l'ítem 8.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	4,88	5,250	6,43	3,47
	Valor <i>p</i>	—	NaN	0,042	0,005	0,275
Opció B	Diferència de mitjanes		—	0,375	1,56	-1,40
	Valor <i>p</i>		—	NaN	NaN	NaN
Opció C	Diferència de mitjanes			—	1,18	-1,78
	Valor <i>p</i>			—	0,968	0,898
Opció D	Diferència de mitjanes				—	-2,96
	Valor <i>p</i>				—	0,555
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la A, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb els distractors C i D, però no amb l'opció incorrecta E. Per al distractor B no s'han pogut calcular.

Si ens fixem en la taula de les estadístiques descriptives, observem que la mitjana del distractor E és de 5,40 punts, mentre que la mitjana de l'opció correcta és de 8,88 punts. En el cas del distractor B, en què no s'ha pogut calcular la prova *post hoc*, observem que la mitjana és de 4,00 punts.

Taula 54

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 8.

	<i>N</i>	Mitjana	<i>DE</i>	<i>EE</i>
Opció A	32	8,88	1,34	0,237
Opció B	2	4,00	5,66	4,000
Opció C	8	3,63	4,00	1,413
Opció D	9	2,44	3,75	1,248
Opció E	5	5,40	3,21	1,435

Notes: *DE* = desviació estàndard; *EE* = error típic d'estimació.

Si estudiem les diferències entre els distractors, veiem que no mostren diferències estadísticament significatives entre si. Per tant, podem dir que, malgrat que les mitjanes entre els distractors són similars (no hi ha diferències estadísticament significatives entre si), en el cas del distractor E no es diferencia suficientment de l'opció correcta, la A.

Resultats de l'ítem 9:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 9.

Taula 55

Freqüències de resposta per a l'ítem 9.

Alternativa de resposta	<i>n</i>
Opció A	6
Opció B	6
Opció C	5
Opció D	34
Opció E	5

Nota: l'alternativa de resposta correcta és l'opció D.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. Primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{6 + 6 + 5 + 34 + 5}{5} = \frac{56}{5} = 11,2$$

Amb les freqüències teòriques ja podem calcular la χ^2 i veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}
 \chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\
 &= \frac{(11,2 - 6)^2 + (11,2 - 6)^2 + (11,2 - 5)^2 + (11,2 - 34)^2 + (11,2 - 5)^2}{11,2} \\
 &= \frac{(5,2)^2 + (5,2)^2 + (6,2)^2 + (-22,8)^2 + (6,2)^2}{11,2} \\
 &= \frac{27,04 + 27,04 + 38,44 + 519,84 + 38,44}{11,2} = \frac{650,8}{11,2} = 58,11
 \end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 58,11$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa D, la correcta en l'ítem 9, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

Primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{6 + 6 + 5 + 5}{4} = \frac{22}{4} = 5,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de saber si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(5,5 - 6)^2 + (5,5 - 6)^2 + (5,5 - 5)^2 + (5,5 - 5)^2}{5,5} \\ &= \frac{(-0,5)^2 + (-0,5)^2 + (0,5)^2 + (0,5)^2}{5,5} = \frac{0,25 + 0,25 + 0,25 + 0,25}{5,5} = \frac{1}{5,5} \\ &= 0,18\end{aligned}$$

Com que el valor de χ^2 a la taula per a $\alpha = 0,05$ i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,3-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,18$) és inferior al de la taula, acceptem la hipòtesi nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 56

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 9.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	24,8	4	9,84	<0,001
	Fisher	33,4	4	51	<0,001

Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 11,2$, $p < 0,001$), haurem de fixar-nos en la prova *t* de Welch, que aquesta vegada sí que s'ha pogut calcular. Tal com es mostra a la taula, la prova *t*

de Welch ha resultat estadísticament significativa: $F(4, 9,84) = 24,8$ ($p < 0,001$). Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem d'examinar les proves *post hoc* a fi de veure entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies hem dit que ha sigut estadísticament significativa, calculem les proves *post hoc* de Games-Howell:

Taula 57

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 9.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	−6,83	−0,367	−7,87	−2,57
	Valor <i>p</i>	—	<0,001	0,999	<0,001	0,782
Opció B	Diferència de mitjanes		—	6,467	−1,04	4,27
	Valor <i>p</i>		—	0,017	0,554	0,400
Opció C	Diferència de mitjanes			—	−7,51	−2,20
	Valor <i>p</i>			—	0,013	0,883
Opció D	Diferència de mitjanes				—	5,31
	Valor <i>p</i>				—	0,245
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la D, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives amb els distractors A i C, però no amb l'opció incorrecta B ni amb el distractor E.

Si ens fixem en la taula de les estadístiques descriptives, veiem que la mitjana del distractor B és de 7,67 punts, mentre que la mitjana de l'opció correcta és de 8,71. En el cas del distractor E, la mitjana és de 3,40 punts.

Taula 58

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 9.

	N	Mitjana	DE	EE
Opció A	6	0,833	2,04	0,833
Opció B	6	7,667	1,51	0,615
Opció C	5	1,200	2,68	1,200
Opció D	34	8,706	1,38	0,237
Opció E	5	3,400	4,67	2,088

Notes: *DE* = desviació estàndard; *EE* = error típic d'estimació.

Si estudiem les diferències entre els distractors, veiem que no hi ha diferències estadísticament significatives, excepte en el cas del distractor B, que sí que mostra diferències amb el distractor A i el C. Per tant, podem dir que el distractor B no funciona correctament.

Resultats de l'ítem 10:

1. Comprovem que l'alternativa correcta siga la més seleccionada. Calculem les freqüències de resposta de cadascuna de les alternatives de l'ítem 10.

Taula 59

Freqüències de resposta per a l'ítem 10.

Alternativa de resposta	n
Opció A	6
Opció B	34
Opció C	6
Opció D	4
Opció E	6

Nota: l'alternativa de resposta correcta és l'opció B.

Amb aquestes dades ja podem calcular la χ^2 amb totes les alternatives. Primer calculem quina és la freqüència teòrica o esperada:

$$FT_i = \frac{6 + 34 + 6 + 4 + 6}{5} = \frac{56}{5} = 11,2$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} \\ &= \frac{(11,2 - 6)^2 + (11,2 - 34)^2 + (11,2 - 6)^2 + (11,2 - 4)^2 + (11,2 - 6)^2}{11,2} \\ &= \frac{(5,2)^2 + (-22,8)^2 + (5,2)^2 + (7,2)^2 + (5,2)^2}{11,2} \\ &= \frac{27,04 + 519,84 + 27,04 + 51,84 + 27,04}{11,2} = \frac{652,8}{11,2} = 58,29\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i quatre graus de llibertat (ja que els ítems tenen cinc alternatives) és 9,49 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 58,29$) és superior al de la taula, rebutgem la hipòtesi nul·la: les alternatives de resposta no són equiprobables. De fet, l'alternativa B, la correcta en l'ítem 10, és la més seleccionada.

2. Comprovem que els distractors o alternatives incorrectes són equiprobables.

Fem un contrast d'hipòtesi khi quadrat a fi de comprovar si els distractors són equiprobables: si són seleccionats per un nombre similar de persones. Així, fem una prova khi quadrat només amb els distractors.

Primer calculem quina és la freqüència teòrica o esperada, aquesta vegada només amb les alternatives incorrectes o distractors:

$$FT_i = \frac{6 + 6 + 4 + 6}{4} = \frac{22}{4} = 5,5$$

Amb les freqüències teòriques ja podem calcular la χ^2 a fi de veure si, efectivament, els subjectes s'han distribuït equitativament entre les respostes incorrectes:

$$\begin{aligned}\chi^2 &= \sum_{j=1}^k \frac{(FT_i - FO_i)^2}{FT_i} = \frac{(5,5 - 6)^2 + (5,5 - 6)^2 + (5,5 - 4)^2 + (5,5 - 6)^2}{5,5} \\ &= \frac{(-0,5)^2 + (-0,5)^2 + (1,5)^2 + (-0,5)^2}{5,5} = \frac{0,25 + 0,25 + 2,25 + 0,25}{5,5} = \frac{3}{5,5} \\ &= 0,55\end{aligned}$$

Com que el valor de χ^2 a la taula per a alfa = 0,05 i tres graus de llibertat (ja que treballem amb els quatre distractors, $K - 1 = 4 - 1 = 3$) és 7,81 ($\chi^2_{(0,05,5-1)}$) i, per tant, el valor de χ^2 observat ($\chi^2 = 0,55$) és inferior al de la taula, acceptem la hipòtesi

nul·la: les alternatives de resposta són equiprobables. És a dir, aparentment, els distractors funcionen bé.

3. Finalment, comprovem que el rendiment mitjà en el test dels participants que han triat cada distractor siga inferior al dels participants que han seleccionat l'alternativa correcta, i que no hi ha diferències entre els qui han triat les alternatives incorrectes: la mitjana en el test dels participants que trien la correcta ha de ser superior a la mitjana dels participants que seleccionen la incorrecta, i entre aquests últims no ha d'haver-hi diferències. Fem una anàlisi de variància (ANOVA) usant com a variable agrupadora l'ítem original i com a variable dependent la puntuació total en el test.

Per a la interpretació, examinem en primer lloc els resultats de l'ANOVA:

Taula 60

ANOVA per a la comparació de mitjanes en DAT total en els grups de respostes de l'ítem 10.

		<i>F</i>	<i>gl1</i>	<i>gl2</i>	<i>p</i>
Puntuació DAT total	<i>t</i> de Welch	10,1	4	8,89	0,002
	Fisher	16,5	4	51	<0,001

Com que la prova de Levene d'igualtat de variàncies ha sigut estadísticament significativa ($F(4, 51) = 31,6, p < 0,001$), haurem de fixar-nos en la prova *t* de Welch, que aquesta vegada sí que s'ha pogut calcular. Tal com es mostra a la taula, la prova *t* de Welch ha resultat estadísticament significativa: $F(4, 8,89) = 10,1 (p = 0,002)$. Això indicaria que hi ha diferències entre les mitjanes dels subjectes segons l'opció de resposta que han escollit.

En segon lloc, hauríem d'examinar les proves *post hoc* a fi de saber entre quins grups hi ha aquestes diferències. Com que la prova de Levene d'igualtat de variàncies hem dit que ha resultat estadísticament significativa, calculem les proves *post hoc* de Games-Howell:

Taula 61

Proves post hoc de comparació de mitjanes en els grups de respostes de l'ítem 10.

		Opció A	Opció B	Opció C	Opció D	Opció E
Opció A	Diferència de mitjanes	—	−5,96	−1,50	1,08	−0,667
	Valor <i>p</i>	—	0,028	0,964	0,985	0,997
Opció B	Diferència de mitjanes		—	4,46	7,04	5,294
	Valor <i>p</i>		—	0,283	0,100	0,095
Opció C	Diferència de mitjanes			—	2,58	0,833
	Valor <i>p</i>			—	0,854	0,997
Opció D	Diferència de mitjanes				—	−1,750
	Valor <i>p</i>				—	0,940
Opció E	Diferència de mitjanes					—
	Valor <i>p</i>					—

Com que l'opció correcta és la B, busquem les proves *post hoc* d'aquesta opció, marcades en cursiva a la taula. Les proves *post hoc* indiquen diferències estadísticament significatives només amb el distractor A, però no amb el C, D ni E.

Si ens fixem en la taula de les estadístiques descriptives, observem que la mitjana de l'opció correcta B és superior a les altres, de manera que podem pensar que la falta d'aquestes diferències estadísticament significatives pot ser a causa de la petita mida mostral dels grups que escullen les opcions incorrectes.

Taula 62

Estadístiques descriptives per a DAT total en els grups de respostes de l'ítem 10.

	N	Mitjana	DE	EE
Opció A	6	2,83	3,19	1,302
Opció B	34	8,79	1,32	0,226
Opció C	6	4,33	4,76	1,944
Opció D	4	1,75	3,50	1,750
Opció E	6	3,50	3,89	1,586

Notes: *DE* = desviació estàndard; *EE* = error típic d'estimació.

Si estudiem les diferències entre els distractors, veiem que no hi ha diferències estadísticament significatives. Per tant, podem dir que sembla que els distractors i l'alternativa correcta funcionen correctament, i que l'absència d'aquestes diferències entre l'opció correcta i els altres distractors siga per la petita mida dels grups.

Tema 5. Teoria clàssica de tests: fonaments del model clàssic lineal

5.1. Concepte de fiabilitat

Respon a les preguntes següents:

- 1. Administrem tres formes paral·leles d'un test a 40 subjectes i obtenim una variància vertadera de 55 i una variància observada de 70,50. Quin és el coeficient de fiabilitat del test? I l'índex de fiabilitat?**
- 2. Administrem dues formes paral·leles d'un test i obtenim un coeficient de fiabilitat de 0,80, una mitjana de 25 i una variància de 4. Quin és l'índex de fiabilitat del test? I l'error típic de mesura? I la variància vertadera?**
- 3. Si el test A té un coeficient de fiabilitat de 0,70 i el test B un índex de fiabilitat de 0,70, quin té millors evidències de fiabilitat (pressuposant les mateixes condicions de variabilitat)?**
- 4. Si les puntuacions del test estan estandarditzades i el coeficient de fiabilitat és 0,90, quin és l'error típic de mesura?**
- 5. Interpreta què significa un coeficient de fiabilitat igual a 0,80 en un test que mesura masclisme.**
- 6. Es pot obtenir empíricament un coeficient de fiabilitat negatiu? Raona la resposta.**

Concepte de fiabilitat. Solucions

- 1. Administrem tres formes paral·leles d'un test a 40 subjectes i obtenim una variància vertadera de 55 i una variància observada de 70,50. Quin és el coeficient de fiabilitat del test? I l'índex de fiabilitat?**

Dividim la variància de la puntuació vertadera entre la variància de la puntuació observada per a obtenir el coeficient de fiabilitat del test:

$$r_{XX} = \frac{s_V^2}{s_X^2} = \frac{55}{70,50} = 0,78$$

L'índex de fiabilitat es calcula amb l'arrel quadrada del resultat anterior:

$$r_{XV} = \sqrt{r_{XX}} = \sqrt{0,78} = 0,88$$

Així, el coeficient de fiabilitat prendria un valor de 0,78, mentre que l'índex de fiabilitat seria de 0,88.

- 2. Administrem dues formes paral·leles d'un test i obtenim un coeficient de fiabilitat de 0,80, una mitjana de 25 i una variància de 4. Quin índex de fiabilitat té el test? I l'error típic de mesura? I la variància vertadera?**

L'índex de fiabilitat es calcula a partir del coeficient de fiabilitat:

$$r_{XV} = \sqrt{r_{XX}} = \sqrt{0,80} = 0,89$$

L'índex de fiabilitat és 0,89.

Calculem l'error típic de mesura:

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX}} = 2 \sqrt{1 - 0,80} = 2 \sqrt{0,20} = 2 \cdot 0,45 = 0,90$$

L'error típic de mesura és 0,90.

La variància vertadera es pot calcular de diverses maneres:

- Mitjançant el coeficient de fiabilitat:

$$r_{XX} = \frac{s_V^2}{s_X^2}$$

$$0,8 = \frac{s_V^2}{4} \rightarrow s_V^2 = 0,80 \cdot 4 = 3,2$$

- Mitjançant l'índex de fiabilitat, calculem primer la desviació típica vertadera:

$$r_{XV} = \frac{s_V}{s_X}$$

$$0,89 = \frac{s_V}{2} \rightarrow s_V = 0,89 \cdot 2 = 1,78$$

I després elevem al quadrat la desviació típica de les puntuacions vertaderes:

$$s_V^2 = 1,78^2 = 3,2$$

- O mitjançant la variància observada i la variància d'error:

$$\sigma_{X_f}^2 = \sigma_V^2 + \sigma_{E_f}^2$$

Calculem la variància o proporció d'error a partir de l'error típic de mesura (desviació típica de l'error):

$$\sigma_E^2 = 0,90^2 = 0,80$$

I després restem a la variància observada la variància d'error:

$$\sigma_V^2 = \sigma_{X_f}^2 - \sigma_{E_f}^2 = 4 - 0,8 = 3,2$$

En tots els casos, la variància vertadera és 3,2.

*Resultats aproximats per decimals.

3. Si el test A té un coeficient de fiabilitat de 0,70 i el test B un índex de fiabilitat de 0,70, quin té millors evidències de fiabilitat (pressuposant les mateixes condicions de variabilitat)?

Podem calcular l'índex de fiabilitat per al test A:

$$r_{XV} = \sqrt{r_{XX}} = \sqrt{0,70} = 0,84$$

I el comparem amb l'índex de fiabilitat del test B que era $r_{XV} = 0,70$. L'índex de fiabilitat del test A és més alt que el del test B.

O podem calcular el coeficient de fiabilitat del test B:

$$r_{XX} = r_{XV}^2 = 0,70^2 = 0,49$$

I el comparem amb el coeficient de fiabilitat del test A, que era $r_{XV} = 0,70$. El coeficient de fiabilitat del test A és més alt que el del test B.

Per tant, el test A té millors evidències de fiabilitat.

4. Si les puntuacions del test estan estandarditzades i el coeficient de fiabilitat és 0,90, quin és l'error típic de mesura?

Les puntuacions estandarditzades impliquen que la desviació típica és 1.

Per tant:

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX}} = 1 \sqrt{1 - 0,90} = 1 \sqrt{0,10} = 1 \cdot 0,32 = 0,32$$

L'error típic de mesura és 0,32.

5. Interpreta què significa un coeficient de fiabilitat igual a 0,80 en un test que mesura masclisme.

Que el coeficient de fiabilitat prengui un valor de 0,80 indica que el 80% de les diferències que observem en les puntuacions que obtenen les persones a qui s'administra el test es deu al fet que els subjectes tenen diferent nivell de masclisme (diferents puntuacions vertaderes).

6. És possible obtenir empíricament un coeficient de fiabilitat negatiu? Raona la resposta.

Sí, ja que en la pràctica es pot obtenir com la correlació entre dues formes paral·leles d'un test i, com a coeficient de correlació, pot tenir un valor entre -1 i 1. No obstant això, aquesta circumstància és molt poc probable que es produïska amb dades empíriques: un valor negatiu és conceptualment inadmissible perquè el coeficient de fiabilitat és una proporció de variàncies, el valor va de 0 a 1, de manera que els valors negatius no tenen sentit.

Tema 6. Estimació de la fiabilitat

6.1. Procediments d'estimació de la fiabilitat i factors que l'afecten

Respon a les preguntes següents:

- 1. Els procediments d'estimació de la fiabilitat segons la TCT es veuen afectats per tres tipus d'error: errors per efectes transitoris, errors per l'especificitat i errors per factors aleatoris. A què es refereix cada tipus d'error i com afecta cada tipus de procediment d'estimació?**
- 2. Apliquem un test a una mostra de 10 subjectes i les dades obtingudes es mostren a la taula 1. El coeficient de fiabilitat pel mètode de les dues meitats ha sigut $r_{XX} = 0,817$. Aplica la fórmula adequada per a corregir-lo i interpreta'l.**
- 3. Apliquem un test a una mostra de 10 subjectes i les dades resultants es mostren a la taula 2. Fes servir la fórmula de Rulon i la fórmula de Guttman-Flanagan per a obtenir el coeficient de fiabilitat del test i interpreta els resultats obtinguts.**
- 4. Calcula el coeficient de fiabilitat d'alfa de Cronbach a partir de les dades de la taula 2. Interpreta els resultats d'aquest càlcul. Per què és diferent del que s'obté amb les fórmules de Rulon i de Guttman-Flanagan?**
- 5. Passem un test de 30 ítems en què obtenim les dades següents: longitud inicial = 30 ítems; mitjana = 35; variància = 26; coeficient de fiabilitat = 0,75. Quin seria el coeficient de fiabilitat del test si n'augmentàrem la longitud a 45 ítems?**

6. Passem un test compost per 10 ítems i calculem que té un coeficient de fiabilitat de 0,80. S'afegeixen 20 ítems paral·lels més a la longitud inicial. Quin coeficient de fiabilitat tindria després d'afegir-hi els 20 ítems?
7. Sabent que el coeficient de fiabilitat d'un test d'aptitud compost per 100 ítems és igual a 0,90, quin coeficient de fiabilitat tindria si eliminem 25 ítems?
8. Passem un test de 50 ítems en què obtenim les dades següents: longitud inicial = 50 ítems; variància = 46; coeficient de fiabilitat = 0,75. Quants ítems hauríem d'afegir-hi perquè el coeficient de fiabilitat augmentara a 0,90?
9. Passem un test de 80 ítems en què obtenim les dades següents: longitud inicial = 80 ítems; mitjana = 55; variància = 64; coeficient de fiabilitat = 0,85. Quants ítems hauríem d'afegir-hi perquè el coeficient de fiabilitat fora de 0,95?
10. Passem dos tests en què obtenim les dades següents. El test A té una mitjana de 45, una variància de 10 i un coeficient de fiabilitat de 0,73. El test B té una mitjana de 30, una variància de 25 i una fiabilitat de 0,86. Si igualem les condicions de variabilitat, quin test és més fiable?
11. Passem dos tests en què obtenim les dades següents: el test A té una mitjana de 45, una variància de 25 i un coeficient de fiabilitat de 0,72. El test B té una mitjana de 65, una variància de 47 i una fiabilitat de 0,83. Si les condicions de variabilitat s'igualen, quin test és més fiable?

Taula 1.*Puntuació de 10 subjectes en els quatre ítems d'un test.*

	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Total	Meitat A (ítems 1 i 2)	Meitat B (ítems 3 i 4)	$d (A - B)$
Subjecte 1	3	4	4	4	15	7	8	-1
Subjecte 2	4	4	4	3	15	8	7	1
Subjecte 3	4	4	4	4	16	8	8	0
Subjecte 4	4	2	3	3	12	6	6	0
Subjecte 5	4	4	4	4	16	8	8	0
Subjecte 6	3	5	4	4	16	8	8	0
Subjecte 7	5	4	4	5	18	9	9	0
Subjecte 8	5	5	5	5	20	10	10	0
Subjecte 9	3	4	4	4	15	7	8	-1
Subjecte 10	4	4	3	4	15	8	7	1
	$\bar{X}_1 = 3,90$	$\bar{X}_2 = 4,00$	$\bar{X}_3 = 3,90$	$\bar{X}_4 = 4,00$	$\bar{X} = 15,80$	$\bar{X}_A = 7,90$	$\bar{X}_B = 7,90$	$\bar{X}_d = 0,00$
	$s_{X1}^2 = 0,544$	$s_{X2}^2 = 0,816$	$s_{X3}^2 = 0,568$	$s_{X4}^2 = 0,667$	$s_X^2 = 4,40$	$s_A^2 = 1,211$	$s_B^2 = 1,211$	$s_d^2 = 0,444$

Taula 2*Puntuació de 10 subjectes en els quatre ítems d'un test.*

	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Total	Meitat A (ítems 1 i 2)	Meitat B (ítems 3 i 4)	$d (A - B)$
Subjecte 1	3	4	4	4	15	7	8	-1
Subjecte 2	4	2	4	3	13	6	7	-1
Subjecte 3	4	4	4	4	16	8	8	0
Subjecte 4	4	2	3	3	12	6	6	0
Subjecte 5	4	3	4	4	15	7	8	-1
Subjecte 6	3	5	4	4	16	8	8	0
Subjecte 7	5	4	4	5	18	9	9	0
Subjecte 8	5	4	5	5	19	9	10	-1
Subjecte 9	3	4	4	4	15	7	8	-1
Subjecte 10	4	1	3	4	12	5	7	-2
	$\bar{X}_1 = 3,90$	$\bar{X}_2 = 3,30$	$\bar{X}_3 = 3,90$	$\bar{X}_4 = 4,00$	$\bar{X} = 15,10$	$\bar{X}_A = 7,20$	$\bar{X}_B = 7,90$	$\bar{X}_d = -0,70$
	$s_{X1}^2 = 0,544$	$s_{X2}^2 = 1,567$	$s_{X3}^2 = 0,322$	$s_{X4}^2 = 0,444$	$s_X^2 = 5,433$	$s_A^2 = 1,733$	$s_B^2 = 1,211$	$s_d^2 = 0,456$

Procediments d'estimació de la fiabilitat i factors que l'afecten. Solucions

- 1. Els procediments d'estimació de la fiabilitat segons la TCT es veuen afectats per tres tipus d'error: errors per efectes transitoris, errors per l'especificitat i errors per factors aleatoris. A què es refereix cada tipus d'error i com afecta cada tipus de procediment d'estimació?**

Els errors per efectes transitoris comporten modificacions en les respostes d'una persona a causa de factors que canvien d'una sessió d'aplicació a l'altra. Per tant, afecten el coeficient d'estabilitat temporal, que implica l'aplicació repetida d'aquest instrument de mesura sobre el mateix subjecte o subjectes en dos moments temporals distints.

Els errors per l'especificitat comporten canvis en les respostes a causa del contingut concret dels ítems que es presenten. Així doncs, afecten el coeficient d'equivalència, ja que les formes paral·leles d'un temps estan formades per ítems diferents, i el coeficient de consistència interna, tant quan es calcula amb el mètode de les dues meitats, ja que les meitats d'un test estan formades per diversos ítems, com quan es calcula mitjançant l'alfa de Cronbach, ja que considera cada ítem com una part del test, i cada ítem és diferent de l'altre.

Els errors per factors aleatoris es refereixen a inconsistències en les respostes que no poden atribuir-se directament al contingut dels ítems, ni a factors de la persona ni a la situació, i que actuen de manera sistemàtica en el temps. Poden deure's a variacions en l'atenció, a distraccions momentànies, a labilitat del sistema nerviós, etc. Per tant, afecten tots els procediments de càlcul de la fiabilitat.

- 2. Apliquem un test a una mostra de 10 subjectes i les dades obtingudes es mostren a la taula 1. El coeficient de fiabilitat pel mètode de les dues meitats ha sigut $r_{XX} = 0,817$. Aplica la fórmula adequada per a corregir-lo i interpreta'l.**

Taula 1*Puntuació de 10 subjectes en els quatre ítems d'un test.*

	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Total	Meitat A (ítems 1 i 2)	Meitat B (ítems 3 i 4)	$d (A - B)$
Subjecte 1	3	4	4	4	15	7	8	-1
Subjecte 2	4	4	4	3	15	8	7	1
Subjecte 3	4	4	4	4	16	8	8	0
Subjecte 4	4	2	3	3	12	6	6	0
Subjecte 5	4	4	4	4	16	8	8	0
Subjecte 6	3	5	4	4	16	8	8	0
Subjecte 7	5	4	4	5	18	9	9	0
Subjecte 8	5	5	5	5	20	10	10	0
Subjecte 9	3	4	4	4	15	7	8	-1
Subjecte 10	4	4	3	4	15	8	7	1
	$\bar{X}_1 = 3,90$	$\bar{X}_2 = 4,00$	$\bar{X}_3 = 3,90$	$\bar{X}_4 = 4,00$	$\bar{X} = 15,80$	$\bar{X}_A = 7,90$	$\bar{X}_B = 7,90$	$\bar{X}_d = 0,00$
	$s_{X1}^2 = 0,544$	$s_{X2}^2 = 0,816$	$s_{X3}^2 = 0,568$	$s_{X4}^2 = 0,667$	$s_X^2 = 4,40$	$s_A^2 = 1,211$	$s_B^2 = 1,211$	$s_d^2 = 0,444$

Com que les mitjanes i les variàncies de les formes són iguals, es compleixen els supòsits del paral·lelisme i podem dir que les dues meitats són paral·leles. Quan aquest és el cas, es pot aplicar la fórmula de Spearman-Brown, que informa sobre quin hauria sigut el valor del coeficient de correlació de Pearson estimat si, en lloc de calcular-se amb el nombre d'ítems de la meitat del test, s'haguera calculat amb un test de longitud doble.

$$r_{XX} = \frac{2r_{X_A X_B}}{1 + r_{X_A X_B}} = \frac{2 \cdot 0,817}{1 + 0,817} = \frac{1,634}{1,817} = 0,899$$

Per tant, si el test s'haguera aplicat sobre un test de quatre ítems de longitud en lloc de dos ítems (els que té cada meitat), el coeficient de fiabilitat hauria sigut de 0,899. Segons els punts de tall del CET-R, es tracta d'un valor excel·lent.

3. Apliquem un test a una mostra de 10 subjectes i les dades resultants es mostren a la taula 2. Fes servir la fórmula de Rulon i la fórmula de Guttman-

Flanagan per a obtenir el coeficient de fiabilitat del test i interpreta els resultats obtinguts.

Taula 2.

Puntuació de 10 subjectes en els quatre ítems d'un test.

	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Total	Meitat A (ítems 1 i 2)	Meitat B (ítems 3 i 4)	$d (A - B)$
Subjecte 1	3	4	4	4	15	7	8	-1
Subjecte 2	4	2	4	3	13	6	7	-1
Subjecte 3	4	4	4	4	16	8	8	0
Subjecte 4	4	2	3	3	12	6	6	0
Subjecte 5	4	3	4	4	15	7	8	-1
Subjecte 6	3	5	4	4	16	8	8	0
Subjecte 7	5	4	4	5	18	9	9	0
Subjecte 8	5	4	5	5	19	9	10	-1
Subjecte 9	3	4	4	4	15	7	8	-1
Subjecte 10	4	1	3	4	12	5	7	-2
	$\bar{X}_1 = 3,90$	$\bar{X}_2 = 3,30$	$\bar{X}_3 = 3,90$	$\bar{X}_4 = 4,00$	$\bar{X} = 15,10$	$\bar{X}_A = 7,20$	$\bar{X}_B = 7,90$	$\bar{X}_d = -0,70$
	$s_{X1}^2 = 0,544$	$s_{X2}^2 = 1,567$	$s_{X3}^2 = 0,322$	$s_{X4}^2 = 0,444$	$s_X^2 = 5,433$	$s_A^2 = 1,733$	$s_B^2 = 1,211$	$s_d^2 = 0,456$

Apliquem la fórmula de Rulon:

$$r_{XX} = 1 - \frac{s_d^2}{s_X^2} = 1 - \frac{0,456}{5,433} = 1 - 0,08 = 0,92$$

Apliquem la fórmula de Guttman-Flanagan:

$$r_{XX} = 2 \left(1 - \frac{s_A^2 + s_B^2}{s_X^2} \right) = 2 \left(1 - \frac{1,733 + 1,211}{5,433} \right) = 2 \left(1 - \frac{2,944}{5,433} \right) = 2(1 - 0,54) = 2 \cdot 0,46 = 0,92$$

Els resultats de les dues fórmules indiquen que el coeficient de fiabilitat del test és 0,92. Ambdues fórmules donen el mateix resultat perquè són equivalents. Segons els punts de tall del CET-R, es tracta d'un valor excel·lent.

- 4. Calcula el coeficient de fiabilitat d'alfa de Cronbach a partir de les dades de la taula 2. Interpreta els resultats d'aquest càlcul. Per què és diferent del que s'obté amb les fórmules de Rulon i de Guttman-Flanagan?**

Calculem el coeficient de fiabilitat alfa de Cronbach:

$$\alpha = \frac{J}{J-1} \left(1 - \frac{\sum_j s_{x_j}^2}{s_x^2} \right) = \frac{4}{4-1} \left(1 - \frac{0,544 + 1,567 + 0,322 + 0,444}{5,433} \right) = \frac{4}{3} \left(1 - \frac{2,877}{5,433} \right) \\ = 1,33(1 - 0,53) = 1,33 \cdot 0,47 = 0,63$$

El coeficient de fiabilitat alfa de Cronbach té un resultat de 0,63. Segons els punts de tall del CET-R, es tracta d'un valor adequat, però amb mancances.

Es tracta d'un valor inferior al calculat amb els mètodes de les dues meitats perquè el coeficient alfa és el límit inferior del coeficient de fiabilitat, atès que sempre pren valors iguals o per davall d'aquest. El coeficient pressuposa que els ítems són paral·lels o, almenys, tau-equivalents, per la qual cosa, quan això no ocorre, com en aquest cas (no tenen les mateixes mitjanes ni variàncies), el coeficient alfa subestima el coeficient de fiabilitat.

- 5. Passem un test de 30 ítems en què obtenim les dades següents: longitud inicial = 30 ítems; mitjana = 35; variància = 26; coeficient de fiabilitat = 0,75. Quin seria el coeficient de fiabilitat del test si n'augmentàrem la longitud a 45 ítems?**

Calculem primer quin és el valor de k :

$$k = \frac{\text{longitud final}}{\text{longitud original}} = \frac{45}{30} = 1,5$$

I després apliquem la fórmula:

$$r_{xx'} = \frac{k \cdot r_{xx}}{1 + (k - 1) \cdot r_{xx}} = \frac{1,5 \cdot 0,75}{1 + (1,5 - 1) \cdot 0,75} = \frac{1,125}{1 + (0,5) \cdot 0,75} = \frac{1,125}{1 + 0,375} = \frac{1,125}{1,375} \\ = 0,82$$

El coeficient de fiabilitat del test si n'augmentem la longitud a 40 ítems augmenta fins a 0,82. Per tant, l'evidència de fiabilitat passaria de ser adequada a ser bona.

- 6. Passem un test compost per 10 ítems i calculem que té un coeficient de fiabilitat de 0,80. S'afegeixen 20 ítems paral·lels més a la longitud inicial. Quin coeficient de fiabilitat tindria després d'afegir-hi els 20 ítems?**

Calculem primer quin és el valor de k :

$$k = \frac{\text{longitud final}}{\text{longitud original}} = \frac{30}{10} = 3$$

I després apliquem la fórmula:

$$r_{XX'} = \frac{k \cdot r_{XX}}{1 + (k - 1) \cdot r_{XX}} = \frac{3 \cdot 0,80}{1 + (3 - 1) \cdot 0,80} = \frac{2,4}{1 + (2) \cdot 0,80} = \frac{2,4}{1 + 1,60} = \frac{2,4}{2,60} = 0,923$$

El coeficient de fiabilitat si la longitud augmenta a 30 ítems creix fins a 0,923. Passa de ser bona a excel·lent, però implica triplicar la longitud del test.

- 7. Sabent que el coeficient de fiabilitat d'un test d'aptitud compost per 100 ítems és igual a 0,90, quin coeficient de fiabilitat tindria si eliminem 25 ítems?**

La longitud final del test seran 75 ítems (100 – 25).

Calculem ara quin és el valor de k :

$$k = \frac{\text{longitud final}}{\text{longitud original}} = \frac{75}{100} = 0,75$$

I apliquem la fórmula:

$$\begin{aligned} r_{XX'} &= \frac{k \cdot r_{XX}}{1 + (k - 1) \cdot r_{XX}} = \frac{0,75 \cdot 0,90}{1 + (0,75 - 1) \cdot 0,90} = \frac{0,675}{1 + (-0,25) \cdot 0,90} = \frac{0,675}{1 + (-0,225)} \\ &= \frac{0,675}{0,775} = 0,87 \end{aligned}$$

El coeficient de fiabilitat del test si eliminem 25 ítems disminueix de 0,90 a 0,87, de manera que continua sent excel·lent.

- 8. Passem un test de 50 ítems en què obtenim les dades següents: longitud inicial = 50 ítems; variància = 46; coeficient de fiabilitat = 0,75. Quants ítems hauríem d'afegir-hi perquè el coeficient de fiabilitat augmentara a 0,90?**

Per a saber quant hem d'augmentar la longitud del test, abans hem de calcular el valor de k :

$$k = \frac{r_{XX'}(1 - r_{XX})}{r_{XX}(1 - r_{XX'})} = \frac{0,90(1 - 0,75)}{0,75(1 - 0,90)} = \frac{0,90 \cdot 0,25}{0,75 \cdot 0,10} = \frac{0,225}{0,075} = 3$$

Per tant, hauríem de triplicar la longitud del test. És a dir, el nou test hauria de tenir 150 ítems:

$$longitud\ final = k \cdot longitud\ inicial = 3 \cdot 50 = 150$$

I, per tant, hauríem d'afegir 100 ítems al test original.

- 9. Passem un test de 80 ítems en què obtenim les dades següents: longitud inicial = 80 ítems; mitjana = 55; variància = 64; coeficient de fiabilitat = 0,85. Quants ítems hauríem d'afegir-hi perquè el coeficient de fiabilitat fora de 0,95?**

Per a saber quant hem d'augmentar la longitud del test, abans hem de calcular el valor de k :

$$k = \frac{r_{XX'}(1 - r_{XX})}{r_{XX}(1 - r_{XX'})} = \frac{0,95(1 - 0,85)}{0,85(1 - 0,95)} = \frac{0,95 \cdot 0,15}{0,85 \cdot 0,05} = \frac{0,1425}{0,0425} = 3,35$$

Per tant, hauríem de multiplicar per 3,35 la longitud inicial del test. És a dir, el nou test hauria de tenir 268 ítems:

$$longitud\ final = k \cdot longitud\ inicial = 3,35 \cdot 80 = 268$$

I, per tant, hauríem d'afegir 188 ítems al test original.

- 10. Passem dos tests en què obtenim les dades següents. El test A té una mitjana de 45, una variància de 10 i un coeficient de fiabilitat de 0,73. El test B té una mitjana de 30, una variància de 25 i una fiabilitat de 0,86. Si igualem les condicions de variabilitat, quin test és més fiable?**

Igualem les condicions de variabilitat del test A (amb menys variabilitat) amb les condicions de variabilitat del test B. És a dir, calculem quin hauria sigut el coeficient de fiabilitat per al test A si s'haguera calculat en una mostra amb la mateixa variabilitat que B.

Per a fer-ho, apliquem la fórmula de correcció del coeficient de correlació per restricció de rang, tenint en compte que la variància del test A en la mostra original ha sigut de 10 ($s_{X(1)}^2 = 10$), que la variància del test A igualada a la del test B seria 25 ($s_{X(2)}^2 = 25$) i que el coeficient de fiabilitat del test A en la mostra original és 0,73 ($r_{XX(1)} = 0,73$):

$$r_{XX(2)} = 1 - \frac{s_{X(1)}^2(1 - r_{XX(1)})}{s_{X(2)}^2} = 1 - \frac{10(1 - 0,73)}{25} = 1 - \frac{10(0,27)}{25} = 1 - \frac{2,7}{25} = 1 - 0,108 = 0,892$$

El test A és més fiable que el test B si igualem les condicions de variabilitat (0,892 respecte a 0,86).

11. Passem dos tests en què obtenim les dades següents: el test A té una mitjana de 45, una variància de 25 i un coeficient de fiabilitat de 0,72. El test B té una mitjana de 65, una variància de 47 i una fiabilitat de 0,83. Si les condicions de variabilitat s'igualen, quin test és més fiable?

Igualem les condicions de variabilitat del test A (amb menys variabilitat) a les condicions de variabilitat del test B. És a dir, calculem quin hauria sigut el coeficient de fiabilitat per al test A si s'haguera calculat en una mostra amb la mateixa variabilitat que B.

Per a fer-ho, apliquem la fórmula de correcció del coeficient de correlació per restricció de rang, tenint en compte que la variància del test A en la mostra original ha sigut de 25 ($s_{X(1)}^2 = 25$), que la variància del test A igualada a la del test B seria de 47 ($s_{X(2)}^2 = 47$) i que el coeficient de fiabilitat del test A en la mostra original és de 0,72 ($r_{XX(1)} = 0,72$):

$$r_{XX(2)} = 1 - \frac{s_{X(1)}^2(1 - r_{XX(1)})}{s_{X(2)}^2} = 1 - \frac{25(1 - 0,72)}{47} = 1 - \frac{25(0,28)}{47} = 1 - \frac{7}{47} = 1 - 0,149 = 0,851$$

El test A és més fiable que el test B si les condicions de variabilitat s'igualen (0,851 respecte a 0,83).

6.2. Variants d'alfa i estimació de puntuacions vertaderes i contrastos de puntuacions observades

Respon a les preguntes següents:

1. Hem passat a 10 subjectes un test de comprensió lectora compost per tres ítems amb escala de resposta de tipus Likert. El coeficient de fiabilitat, calculat amb el coeficient alfa de Cronbach, ha sigut de 0,76 ($\alpha = 0,76$). Quin valor tindria el coeficient no esbiaixat? Interpreta el resultat obtingut.
2. Hem aplicat un test de matemàtiques compost per sis ítems dicotòmics a 10 subjectes i volem calcular-ne el coeficient de fiabilitat mitjançant el mètode de consistència interna més adient. Calcula el coeficient de consistència interna del test per a les dades que apareixen a la taula 1, tenint en compte que la variància observada és 1,656 ($s_X^2 = 1,656$), i interpreta'l.
3. Hem aplicat un test d'anglès compost per tres ítems dicotòmics a 10 subjectes i volem calcular-ne el coeficient de fiabilitat mitjançant el mètode de consistència interna més adient. Calcula el coeficient de consistència interna d'aquest test segons les dades que apareixen a la taula 2 i tenint en compte, així mateix, que la mitjana és 2,10 ($\bar{X} = 2,10$) i que la variància observada és 1,656 ($s_X^2 = 1,656$). Interpreta els resultats obtinguts.
4. Sabem que una persona ha obtingut una puntuació directa de 20 en un test de masclisme. Pressuposant que la distribució de les puntuacions observades d'una persona en les diverses formes paral·leles d'un test és normal, i sabent que l'error típic de mesura és de 0,34, entre quins valors se situa la puntuació vertadera de la persona, pressuposant un 5% d'error?
5. Una persona ha obtingut una puntuació directa de 35 en un test de raonament matemàtic. Una altra ha obtingut en el mateix test una puntuació de 32. Pressuposant que la diferència entre aquestes puntuacions observades es distribueix normalment i que l'error de mesura és 1,2, podem afirmar que les puntuacions observades difereixen en la puntuació vertadera? Respon a la pregunta tenint en compte un error del 5%.

Taula 1.*Puntuació de 10 subjectes en els sis ítems d'un test.*

	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Total
<i>Subjecte 1</i>	1	1	1	1	1	1	6
<i>Subjecte 2</i>	1	1	0	1	0	1	4
<i>Subjecte 3</i>	0	1	1	0	1	0	3
<i>Subjecte 4</i>	1	0	1	1	0	1	4
<i>Subjecte 5</i>	0	1	1	0	1	1	4
<i>Subjecte 6</i>	1	0	1	1	1	0	4
<i>Subjecte 7</i>	1	0	1	0	0	0	2
<i>Subjecte 8</i>	1	0	0	1	1	0	3
<i>Subjecte 9</i>	1	1	0	1	1	1	5
<i>Subjecte 10</i>	1	1	1	1	1	1	6

Taula 2.*Puntuació de 10 subjectes en els tres ítems d'un test.*

	Ítem 1	Ítem 2	Ítem 3	Total
<i>Subjecte 1</i>	1	1	1	3
<i>Subjecte 2</i>	1	1	1	3
<i>Subjecte 3</i>	1	0	1	2
<i>Subjecte 4</i>	1	1	1	3
<i>Subjecte 5</i>	1	1	1	3
<i>Subjecte 6</i>	0	0	0	0
<i>Subjecte 7</i>	0	1	0	1
<i>Subjecte 8</i>	0	0	0	0
<i>Subjecte 9</i>	1	1	1	3
<i>Subjecte 10</i>	1	1	1	3

Variants d'alfa i estimació de puntuacions vertaderes i contrastos de puntuacions observades. Solucions

1. Hem passat a 10 subjectes un test de comprensió lectora compost per tres ítems amb escala de resposta de tipus Likert. El coeficient de fiabilitat, calculat amb el coeficient alfa de Cronbach, ha sigut de 0,76 ($\alpha = 7,96$). Quin valor tindria el coeficient no esbiaixat? Interpreta el resultat obtingut.

Apliquem la fórmula sabent que la mida mostral és 10 ($N = 10$) i el coeficient de fiabilitat 0,76 ($\alpha = 7,6$):

$$\bar{\alpha} = \frac{(N-3)\alpha + 2}{N-1} = \frac{(10-3)0,76 + 2}{10-1} = \frac{7 \cdot 0,76 + 2}{9} = \frac{5,32 + 2}{9} = \frac{7,32}{9} = 0,81$$

L'estimador d'alfa no esbiaixat ha sigut 0,81 ($\bar{\alpha} = 0,81$). Per tant, parlaríem d'una evidència de fiabilitat bona.

2. Hem aplicat un test de matemàtiques compost per sis ítems dicotòmics a 10 subjectes i volem calcular-ne el coeficient de fiabilitat mitjançant el mètode de consistència interna més adient. Calcula el coeficient de consistència interna del test per a les dades que apareixen a la taula 1, tenint en compte que la variància observada és 1,656 ($s_X^2 = 1,656$), i interpreta'l.

En primer lloc, i com que són ítems dicotòmics, cal aplicar alguna de les fórmules de Kuder-Richardson. Per a saber si és més adequat aplicar el KR-20 (ítems amb distinta dificultat) o el KR-21 (ítems amb la mateixa dificultat), cal calcular la proporció d'encerts (p) per a cada ítem.

Taula 1

Puntuació de 10 subjectes en els sis ítems d'un test.

	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Total
Subjecte 1	1	1	1	1	1	1	6
Subjecte 2	1	1	0	1	0	1	4
Subjecte 3	0	1	1	0	1	0	3
Subjecte 4	1	0	1	1	0	1	4
Subjecte 5	0	1	1	0	1	1	4
Subjecte 6	1	0	1	1	1	0	4
Subjecte 7	1	0	1	0	0	0	2
Subjecte 8	1	0	0	1	1	0	3
Subjecte 9	1	1	0	1	1	1	5
Subjecte 10	1	1	1	1	1	1	6
p	0,80	0,60	0,70	0,70	0,70	0,60	
q	0,20	0,40	0,30	0,30	0,30	0,40	
$p \cdot q$	0,16	0,24	0,21	0,21	0,21	0,24	

Com que la proporció d'encerts varia entre ítems, tal com es mostra a la taula, són ítems amb distinta dificultat i cal aplicar el KR-20.

Per a fer-ho, cal calcular també la proporció d'errors (q), informació que ens permet calcular les variàncies dels ítems ($p \cdot q$), dades necessàries per a aplicar la fórmula. Sabent que el nombre d'ítems és sis ($J = 6$), ja podem calcular la fórmula del KR-20:

$$\begin{aligned}
 KR_{20} &= \frac{J}{J-1} \left(1 - \frac{\sum_j p_j q_j}{s_x^2} \right) = \frac{6}{6-1} \left(1 - \frac{0,16 + 0,24 + 0,21 + 0,21 + 0,21 + 0,24}{1,656} \right) \\
 &= \frac{6}{5} \left(1 - \frac{1,27}{1,656} \right) = 1,2(1 - 0,767) = 1,2 \cdot 0,233 = 0,28
 \end{aligned}$$

Com que el coeficient de fiabilitat del test pren un valor de 0,28 ($KR - 20 = 0,28$), concloem que el test té una fiabilitat inadequada.

3. Hem aplicat un test d'anglès compost per tres ítems dicotòmics a 10 subjectes i volem calcular-ne el coeficient de fiabilitat mitjançant el mètode

de consistència interna més adient. Calcula el coeficient de consistència interna d'aquest test segons les dades que apareixen a la taula 2 i tenint en compte, així mateix, que la mitjana és 2,10 ($\bar{X} = 2,10$) i que la variància observada és 1,656 ($s_X^2 = 1,656$). Interpreta els resultats obtinguts.

En primer lloc, i com que són ítems dicotòmics, caldrà aplicar alguna de les fórmules de Kuder-Richardson. Per a saber si és més adequat aplicar el KR-20 (ítems amb distinta dificultat) o el KR-21 (ítems amb la mateixa dificultat), s'ha de calcular la proporció d'encerts (p) per a cada ítem.

Taula 2

Puntuacions de 10 subjectes en els tres ítems d'un test.

	Ítem 1	Ítem 2	Ítem 3	Total
<i>Subjecte 1</i>	1	1	1	3
<i>Subjecte 2</i>	1	1	1	3
<i>Subjecte 3</i>	1	0	1	2
<i>Subjecte 4</i>	1	1	1	3
<i>Subjecte 5</i>	1	1	1	3
<i>Subjecte 6</i>	0	0	0	0
<i>Subjecte 7</i>	0	1	0	1
<i>Subjecte 8</i>	0	0	0	0
<i>Subjecte 9</i>	1	1	1	3
<i>Subjecte 10</i>	1	1	1	3
<i>p</i>	0,70	0,70	0,70	

Com que la proporció d'encerts és la mateixa, són ítems amb la mateixa dificultat i cal aplicar el KR-21.

Per a fer-ho, només cal resoldre la fórmula corresponent, sabent que el nombre d'ítems és tres ($J = 3$), la mitjana és 2,10 ($\bar{X} = 2,10$) i la variància és 1,656 ($s_X^2 = 1,656$), i ja podem calcular la fórmula del KR-21:

$$\begin{aligned}
 KR_{21} &= \frac{J}{J-1} \left(1 - \frac{\bar{X} - \frac{\bar{X}^2}{J}}{s_X^2} \right) = \frac{3}{3-1} \left(1 - \frac{2,1 - \frac{2,1^2}{3}}{1,656} \right) = \frac{3}{2} \left(1 - \frac{2,1 - \frac{4,41}{3}}{1,656} \right) \\
 &= 1,5 \left(1 - \frac{2,1 - 1,47}{1,656} \right) = 1,5 \left(1 - \frac{0,63}{1,656} \right) = 1,5(1 - 0,38) = 1,5 \cdot 0,62 = 0,93
 \end{aligned}$$

El coeficient de fiabilitat del test pren un valor de 0,93 ($KR - 21 = 0,93$), xifra que indica una evidència de fiabilitat excel·lent.

- 4. Sabem que una persona ha obtingut una puntuació directa de 20 en un test de masclisme. Pressuposant que la distribució de les puntuacions observades d'una persona en les diverses formes paral·leles d'un test és normal, i sabent que l'error típic de mesura és de 0,34, entre quins valors se situa la puntuació vertadera d'aquesta persona, pressuposant un 5% d'error?**

Si tenim una puntuació observada d'un subjecte de 20 ($X_i = 20$), pressuposem un $\alpha = 0,05$ ($z_{1-\alpha/2} = z_{1-0,025/2} = z_{0,025} = 1,96$) i sabem que l'error típic de mesura és 0,34 ($s_E = 0,34$), aleshores a aquesta persona li estimarem els límits següents per a la seua puntuació vertadera:

$$V_{Li} = X_i - z_{1-\alpha/2} s_E = 20 - 1,96 \cdot 0,34 = 20 - 0,6664 = 19,3336$$

$$V_{Ls} = X_i + z_{1-\alpha/2} s_E = 20 + 1,96 \cdot 0,34 = 20 + 0,664 = 20,6664$$

Podem dir que la puntuació vertadera (V_i) del subjecte se situa entre 19,3336 i 20,6664, amb un nivell de confiança del 95%.

- 5. Una persona ha obtingut una puntuació directa de 35 en un test de raonament matemàtic. Una altra ha obtingut en el mateix test una puntuació de 32. Pressuposant que la diferència entre aquestes puntuacions observades es distribueix normalment i que l'error de mesura és 1,2, podem afirmar que les puntuacions observades difereixen en la puntuació vertadera? Respon a la pregunta tenint en compte un error del 5%.**

Per a contrastar les puntuacions de dues persones en un test de raonament matemàtic, en què han obtingut puntuacions observades de 35 ($X_1 = 35$) i 32 ($X_2 = 32$), i sabent que l'error típic de mesura és 1,2, el valor de Z és:

$$Z = \frac{(X_1 - X_2)}{S_E \sqrt{2}} = \frac{(35 - 32)}{1,2 \sqrt{2}} = \frac{(3)}{1,2 \cdot 1,41} = \frac{3}{1,692} = 1,773$$

Si pressuposem un nivell de confiança del 95% ($\alpha = 0,05$), la zona crítica de Z seria $Z \leq -1,96$ i $Z \geq +1,96$. Com que el valor de Z resta dins de l'interval ($-1,96 < +1,773 < +1,96$), mantenim la hipòtesi nul·la i no podem afirmar que les puntuacions observades diferisquen en la puntuació vertadera.

6.3. Estimació de la fiabilitat en Jamovi

L'activitat es realitzarà amb la base de dades *Base de dades per a activitat Likert*. Usarem l'escala *Sussex-Oxford Compassion for Others Scale* (SOCS-O; Gu et al., 2020), que mesura la compassió pels altres, per a estudiar les evidències de fiabilitat.

Els ítems que formen l'escala són:

1. M'adone que algú està angoixat sense que m'ho diga.
2. Sé que tothom pateix en algun moment de la vida.
3. Quan una persona ho passa malament, la compadisc.
4. Quan algú està disgustat, procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.
5. Quan algú ho passa malament, procure ajudar-lo d'alguna manera.
6. Percep si algú està angoixat.
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupe pel seu benestar.
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.
10. Quan una persona passa un mal moment, tracte de cuidar-la.
11. Detecte de seguida els primers signes d'angoixa en els altres.
12. Igual que jo, sé que altres persones també ho passen malament en la vida.
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.
14. Connecte amb el patiment dels altres sense jutjar-los.
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.
16. Detecte els signes del patiment en els altres.
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.
18. Soc sensible a l'angoixa dels altres.
19. Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angoixa.
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.

L'*Escala de compassió pels altres d'Oxford* (SOCS-O) avalua cinc dimensions de la compassió: 1. Reconeixement del patiment; 2. Comprensió del patiment com un sentiment universal; 3. Sentiment per la persona que pateix; 4. Tolerància als sentiments incòmodes; 5. Actuar o estar motivat a actuar per a alleujar el patiment.

Aquestes cinc dimensions es consideren diversos aspectes relacionats de la compassió però distingibles. Cada dimensió es mesura amb quatre ítems, que a la base de dades reben el nom de SOCS:

- Reconeixement del patiment: ítems 1, 6, 11 i 16.
- Comprensió del patiment com un sentiment universal: ítems 2, 7, 12 i 17.
- Sentiment per la persona que pateix: ítems 3, 8, 13 i 18.
- Tolerància als sentiments incòmodes: ítems 4, 9, 14 i 19.
- Actuar o estar motivat a actuar per a alleujar el patiment: ítems 5, 10, 15 i 20.

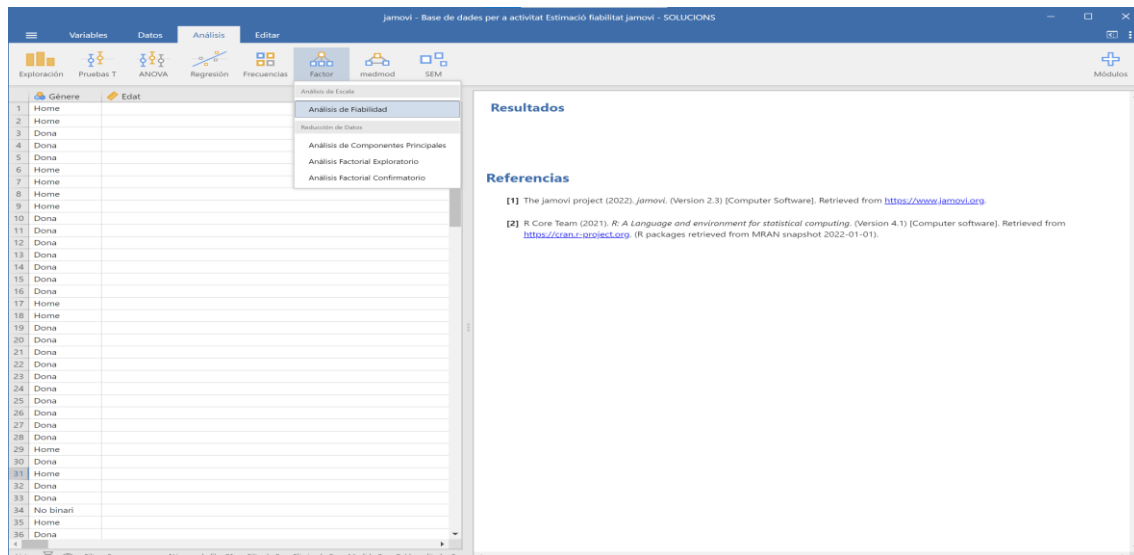
En l'enllaç següent hi ha més informació sobre aquesta escala:

<https://journals.sagepub.com/doi/10.1177/1073191119860911>

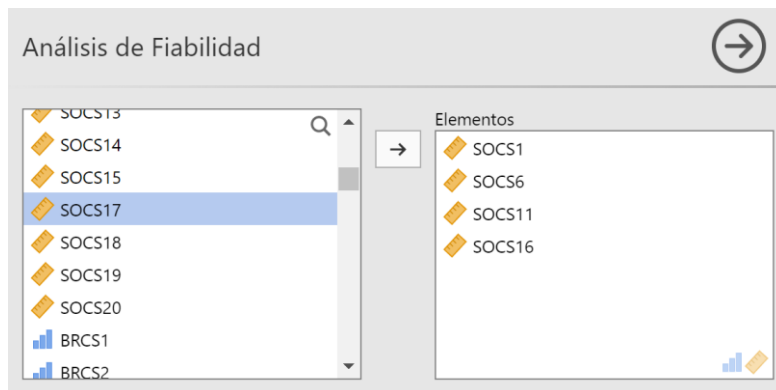
A partir de la base de dades, i seguint els passos del tutorial, calcula l'alfa de Cronbach per a cadascuna de les seues dimensions, el valor d'alfa si s'elimina cadascun dels ítems i l'homogeneïtat corregida dels ítems. Interpreta els resultats obtinguts.

Estimació de la fiabilitat en Jamovi. Tutorial amb exemple de la dimensió *Reconeixement del patiment*

Si volem calcular l'alfa de Cronbach per a la dimensió *Reconeixement del patiment*, que està formada pels ítems 1, 6, 11 i 16, només cal que anem a la pestanya *Anàlisi – Factor – Anàlisi de fiabilitat*.



Ací hem de seleccionar els ítems que formen l'escala i passar-los a la columna *Elements*.



En *Estadístiques de l'escala*, hem d'assegurar-nos que l'alfa de Cronbach està seleccionat:

Estadísticas de Escala

- ☒ Alfa de Cronbach
- ☐ ω de McDonald
- ☐ Media
- ☐ Desviación estándar

I en *Estadísticas dels elements*, és a dir, dels ítems, hem de seleccionar l'alfa de Cronbach si s'elimina un element i la correlació de l'element amb els altres (homogeneïtat).

Estadísticas del Elemento

- ☒ α de Cronbach (si se elimina un elemento)
- ☐ ω de McDonald (si se elimina el elemento)
- ☐ Media
- ☐ Desviación estándar
- ☒ Correlación del elemento con otros

Una vegada tenim els resultats de l'operació, només cal interpretar-los.

Com podem veure, l'alfa de Cronbach pren un valor de 0,868, valor que indica una evidència excel·lent de fiabilitat segons el CET-R.

Estadísticas de Fiabilidad de Escala

Alfa de Cronbach	
escala	0.868

Pel que fa a les estadístiques dels ítems, podem observar que, si eliminem l'ítem 1, el valor de l'alfa de Cronbach disminueix fins a 0,857. D'aquesta manera, l'ítem 1 contribueix a millorar les evidències de fiabilitat de l'escala (mirem en la columna de la dreta, "Si se descarta l'element – Alfa de Cronbach"). El mateix passa amb els ítems 6, 11 i 16. No recomanem en cap cas eliminar cap ítem, ja que en tots els casos, si eliminem un ítem, l'alfa de Cronbach disminueix.

Pel que fa a l'homogeneïtat corregida dels ítems (observem la columna de l'esquerra, "Correlació de l'element amb d'altres"), en aquest cas la informació que ens presenta la taula és la correlació de l'ítem amb el test si haguérem eliminat el mateix ítem del total del test. En l'exemple de l'ítem 1, l'homogeneïtat corregida (el valor de la correlació entre l'ítem 1 i els altres de la dimensió, ítems 6, 11 i 16) pren un valor de 0,655, valor que supera els punts de tall de 0,30 i 0,50 que hem vist en el tema 3. Per tant, té evidències d'homogeneïtat adequades. Això també ocorre amb els altres ítems: l'ítem més homogeni és el 6, i el menys homogeni és l'1.

Estadísticas de Fiabilidad de Elemento

	Correlación del elemento con otros	Si se descarta el elemento
		Alfa de Cronbach
SOCS1	0.655	0.857
SOCS6	0.763	0.816
SOCS11	0.761	0.815
SOCS16	0.716	0.835

Ara només caldria presentar aquesta informació en format APA.

Anàlisi de les evidències de fiabilitat de la dimensió Reconeixement del patiment i característiques dels seus ítems.

Per a obtenir evidències de fiabilitat de la dimensió *Reconeixement del patiment* es calcula l'alfa de Cronbach de l'escala —0,868—, xifra que indica unes evidències de fiabilitat excel·lents segons els punts de tall del CET-R (Hernández et al., 2016).

Específicament, i tal com es mostra a la taula 1, tots els ítems ofereixen evidències de fiabilitat adequades que contribueixen a la fiabilitat del test, ja que en cap cas eliminar un ítem no fa disminuir l'alfa de Cronbach. En tot cas, l'ítem 1 és el que menys contribueix a les evidències de fiabilitat de l'escala (si l'eliminàrem, l'alfa disminuiria, però menys que si eliminàrem els altres).

Finalment, pel que fa a les evidències d'homogeneïtat corregida, també incloses a la taula 1, es pot observar que tots els ítems presenten valors adequats d'homogeneïtat, ja que les correlacions corregides amb el test superen el valor mínim de 0,30 i el més desitjable de 0,50. L'ítem amb més evidències d'homogeneïtat és el 6, i el que té menys evidències d'homogeneïtat és l'1.

Taula 1

Evidències de fiabilitat i homogeneïtat dels ítems de la dimensió Reconeixement del patiment.

	α^c	r_{jX}^c
1. M'adone que algú està angoixat sense que m'ho diga.	0,857	0,655
6. Percep si algú està angoixat.	0,816	0,763
11. Detecte de seguida els primers signes d'angoixa en els altres.	0,815	0,761
16. Detecte els signes de patiment en els altres.	0,835	0,716

Notes: α^c = alfa corregit per eliminació d'ítem; r_{jX}^c = índex d'homogeneïtat corregit per eliminació d'ítem.

Estimació de la fiabilitat en Jamovi. Solucions

Anàlisi de les evidències de fiabilitat de la dimensió Reconeixement del patiment i característiques dels seus ítems.

Per a obtenir evidències de fiabilitat de la dimensió *Reconeixement del patiment* es calcula l'alfa de Cronbach de l'escala, que és 0,868, xifra que indica unes evidències de fiabilitat excel·lents segons els punts de tall del CET-R (Hernández et al., 2016).

Específicament, i tal com es mostra a la taula 1, tots els ítems ofereixen evidències de fiabilitat adequades que contribueixen a la fiabilitat del test, ja que en cap cas eliminar un ítem no fa disminuir l'alfa de Cronbach. En tot cas, l'ítem 1 és el que menys contribueix a les evidències de fiabilitat de l'escala (si l'eliminem, l'alfa disminueix, però menys que si eliminàrem els altres).

Finalment, pel que fa a les evidències d'homogeneïtat corregida, també incloses a la taula 1, es pot observar que tots els ítems presenten valors adequats d'homogeneïtat, ja que les correlacions corregides amb el test superen el valor mínim de 0,30 i el més desitjable de 0,50. L'ítem amb més evidències d'homogeneïtat és el 6, i el que en té menys és l'1.

Taula 1

Evidències de fiabilitat i homogeneïtat dels ítems de la dimensió Reconeixement del patiment.

	α^c	r_{jX}^c
1. M'adone que algú està angoixat sense que m'ho diga.	0,857	0,655
6. Percep si algú està angoixat.	0,816	0,763
11. Detecte de seguida els primers signes d'angoixa en els altres.	0,815	0,761
16. Detecte els signes de patiment en els altres.	0,835	0,716

Notes: α^c = alfa corregit amb l'ítem eliminat; r_{jX}^c = índex d'homogeneïtat corregit amb l'ítem eliminat.

Anàlisi de les evidències de fiabilitat de la dimensió Comprensió del patiment com un sentiment universal i característiques dels seus ítems

Per a obtenir evidències de fiabilitat de la dimensió *Comprensió del patiment* es calcula l'alfa de Cronbach de l'escala, i el resultat és 0,780, xifra que indica unes evidències de fiabilitat adequades segons els punts de tall del CET-R (Hernández et al., 2016).

Específicament, i tal com es mostra a la taula 2, tots els ítems ofereixen evidències de fiabilitat adequades que contribueixen a la fiabilitat del test, ja que en cap cas eliminar un ítem no fa disminuir l'alfa de Cronbach. En tot cas, l'ítem 17 és el que menys contribueix a les evidències de fiabilitat de l'escala (si l'eliminem, l'alfa disminueix, però menys que si eliminàrem els altres).

Finalment, pel que fa a les evidències d'homogeneïtat corregida, també incloses a la taula 2, es pot observar que tots els ítems presenten valors adequats d'homogeneïtat, ja que les correlacions corregides amb el test superen el valor mínim de 0,30 i el més desitjable de 0,50. L'ítem amb més evidències d'homogeneïtat és el 12, i el que menys el 17.

Taula 2

Evidències de fiabilitat i homogeneïtat dels ítems de la dimensió Comprensió del patiment.

	α^c	r_{jX}^c
2. Sé que tothom pateix en algun moment de la vida.	0,752	0,533
7. Sé que sentir-se disgustat de tant en tant forma part de la naturalesa humana.	0,737	0,572
12. Igual que jo, sé que altres persones també ho passen malament en la vida.	0,642	0,746
17. Sé que tots ens disgustem alguna vegada quan patim una injustícia.	0,765	0,506

Notes: α^c = alfa corregit amb l'eliminació de l'ítem; r_{jX}^c = índex d'homogeneïtat corregit amb l'eliminació de l'ítem.

Anàlisi de les evidències de fiabilitat de la dimensió Sentiment per la persona que pateix i característiques dels seus ítems.

Per a obtenir evidències de fiabilitat de la dimensió *Sentiment per la persona que pateix* es calcula l'alfa de Cronbach de l'escala, que té com a resultat 0,813, xifra que indica unes evidències de fiabilitat bones, segons els punts de tall del CET-R (Hernández et al., 2016).

Específicament, i tal com es mostra a la taula 3, tots els ítems ofereixen evidències de fiabilitat adequades que contribueixen a la fiabilitat del test, ja que en cap cas eliminar un ítem no fa disminuir l'alfa de Cronbach. L'ítem 3 és el que menys contribueix a les evidències de fiabilitat de l'escala (si l'eliminàrem, l'alfa disminuiria, però menys que si eliminàrem els altres ítems).

Pel que fa a les evidències d'homogeneïtat corregida, que també es mostren en la taula 3, es pot observar que tots els ítems presenten valors adequats d'homogeneïtat, ja que les correlacions corregides amb el test superen el valor mínim de 0,30 i el més desitjable de 0,50. L'ítem amb més evidències d'homogeneïtat és l'ítem 18, i el de menys el 3.

Taula 3

Evidències de fiabilitat i homogeneïtat dels ítems de la dimensió Sentiment per la persona que pateix.

	α^c	r_{jX}^c
3. Quan una persona ho passa malament, la compadisc.	0,773	0,624
8. Quan sé que a algú li ha passat una cosa dolenta, em preocupa pel seu benestar.	0,757	0,659
13. Quan algú està disgustat, tracte d'empatitzar amb els seus sentiments.	0,766	0,638
18. Soc sensible a l'angoixa dels altres.	0,765	0,683

Notes: α^c = alfa corregit amb l'eliminació de l'ítem; r_{jX}^c = índex d'homogeneïtat corregit amb l'eliminació de l'ítem.

Anàlisi de les evidències de fiabilitat de la dimensió Tolerància als sentiments incòmodes i característiques dels seus ítems

Per a obtenir evidències de fiabilitat de la dimensió *Tolerància als sentiments incòmodes* es calcula l'alfa de Cronbach de l'escala, i el resultat és 0,580, xifra que indica unes evidències de fiabilitat inadequades, segons els punts de tall del CET-R (Hernández et al., 2016).

Tal com es mostra a la taula 4, l'ítem que més perjudica la fiabilitat del test és el 19, que si s'elimina fa que augmente l'alfa de Cronbach de l'escala fins a 0,705. Igualment, l'ítem 19 no presenta tampoc una homogeneïtat corregida adequada, i no arriba fins al punt de tall de 0,30. Per tant, seria recomanable modificar-lo o eliminar-lo.

Taula 4

Evidències de fiabilitat i homogeneïtat dels ítems de la dimensió Tolerància als sentiments incòmodes.

	α^c	r_{jX}^c
4. Quan algú està disgustat procure mostrar-me receptiu/iva als seus sentiments en lloc d'evitar-los.	0,432	0,466
9. Acompanye i escolte els altres quan estan disgustats, encara que no siga fàcil de suportar.	0,468	0,442
14. Connecte amb el patiment dels altres sense jutjar-los.	0,419	0,470
19. Quan algú està disgustat, puc acompanyar-lo sense sentir-me aclaparat per la seua angonya.	0,705	0,165

Notes: α^c = alfa corregit amb l'eliminació de l'ítem; r_{jX}^c = índex d'homogeneïtat corregit amb l'eliminació de l'ítem.

Anàlisi de les evidències de fiabilitat de la dimensió Actuar o estar motivat a actuar per a alleujar el patiment i característiques dels seus ítems

Finalment, per a obtenir evidències de fiabilitat de la dimensió *Actuar o estar motivat a actuar per a alleujar el patiment* es calcula l'alfa de Cronbach de l'escala, i el resultat és 0,808, xifra que indica unes evidències de fiabilitat bones, segons els punts de tall del CET-R (Hernández et al., 2016).

Tal com es mostra a la taula 5, tots els ítems ofereixen evidències de fiabilitat adequades que contribueixen a la fiabilitat del test, ja que en cap cas eliminar un ítem no fa disminuir l'alfa de Cronbach. En tot cas, l'ítem 20 és el que més contribueix a les evidències de fiabilitat de l'escala (si s'eliminara, l'alfa disminuiria en més grau).

Pel que fa a les evidències d'homogeneïtat corregida, també incloses a la taula 5, es pot observar que tots els ítems presenten valors adequats d'homogeneïtat, ja que les correlacions corregides amb el test superen el valor mínim de 0,30 i el més desitjable de 0,50. L'ítem amb més evidències d'homogeneïtat és el 20.

Taula 5

Evidències de fiabilitat i homogeneïtat dels ítems de la dimensió Actuar o estar motivat a actuar per a alleujar el patiment.

	α^c	r_{jX}^c
5. Quan algú ho està passant malament, procure ajudar-lo d'alguna manera.	0,805	0,647
10. Quan una persona passa un mal moment, tracte de cuidar-la.	0,806	0,642
15. Quan veig que una persona necessita ajuda, tracte de fer el millor per a ella.	0,805	0,644
20. Quan veig que una persona no es troba bé, faig tot el possible per cuidar-la.	0,755	0,755

Notes: α^c = alfa corregit amb l'eliminació de l'ítem; r_{jX}^c = índex d'homogeneïtat corregit amb l'eliminació de l'ítem.

Tema 7. Teoria de la resposta a l'ítem

7.1. Corba característica de l'ítem (CCI) i models de la TRI

Respon a les preguntes següents.

- 1. En la figura 1 es presenten les CCI dels quatre ítems que formen un test. A partir de la figura 1, respon a les preguntes següents:**
Quin és l'ítem més difícil? I el més fàcil?
Quin nivell de tret es necessita per a encertar l'ítem 3?
Quin ítem és més discriminant?
Quin ítem té més probabilitat d'encert per atzar?
Tenint en compte les respostes donades a les preguntes anteriors, quin model logístic seria adequat per a estudiar el comportament dels ítems?
- 2. Un subjecte amb un determinat nivell d'habilitat té la probabilitat de 0,70 d'encertar l'ítem 1 i la probabilitat de 0,60 d'encertar l'ítem 2. Tenint en compte que els ítems són localment independents, respon a les preguntes següents:**
Quina és la probabilitat que encerte els dos ítems?
Quina és la probabilitat que encerte el primer ítem i que falle el segon?
Quina és la probabilitat que falle el primer ítem i que encerte el segon?
Quina és la probabilitat que falle els dos ítems?
- 3. Mitjançant el model logístic de dos paràmetres (ML2P) s'han estimat els valors següents dels paràmetres per a un dels ítems d'un test: $a = 1,50$, $b = 1$. Per al nivell $\theta = +2$, quina és la probabilitat d'encertar aquest ítem?**

4. Apliquem un test de cinc ítems a una mostra i obtenim els paràmetres que es mostren a la taula 1. Tenint en compte la informació que es presenta a la taula 1, respon a les preguntes següents:

Quin model has d'usar per a poder representar les dades de la taula 1?

Si ens centrem en l'ítem 5, quina és la probabilitat d'encertar-lo per atzar?

Quin ítem és el més fàcil i per què?

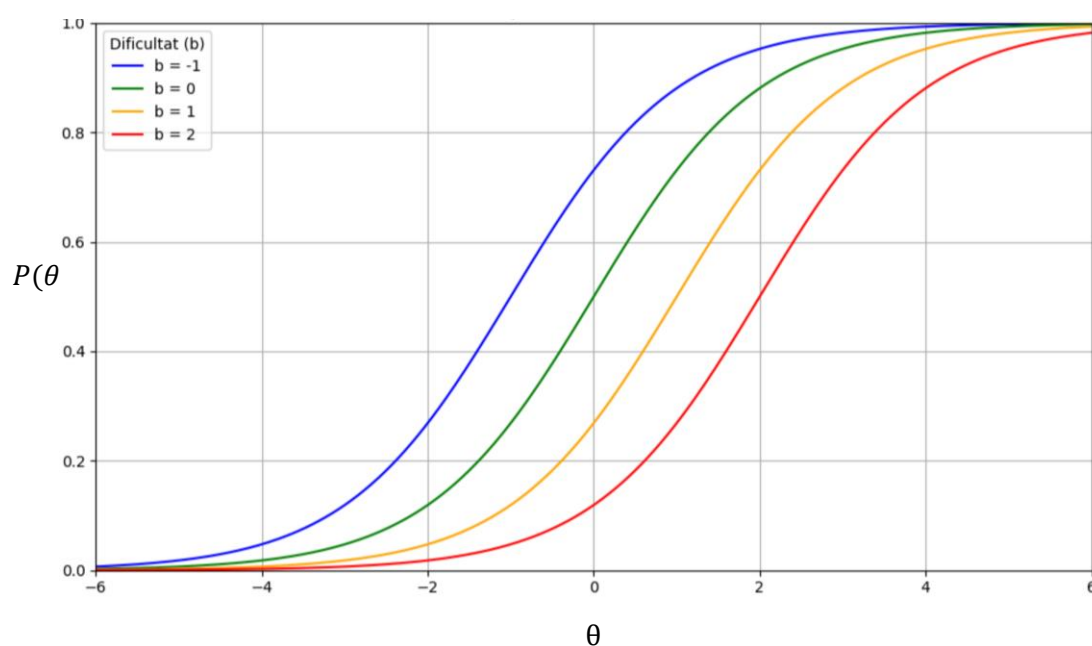
Quin ítem és el més difícil i per què?

Quin ítem és el que té més poder discriminant i per què?

Quin ítem és el que té menys poder discriminant i per què?

Figura 1

CCI dels quatre ítems d'un test.



Notes: línia blava = ítem 1; línia verda = ítem 2; línia groga = ítem 3; línia roja = ítem 4.

Taula 1

Paràmetres estimats per als cinc ítems d'un test.

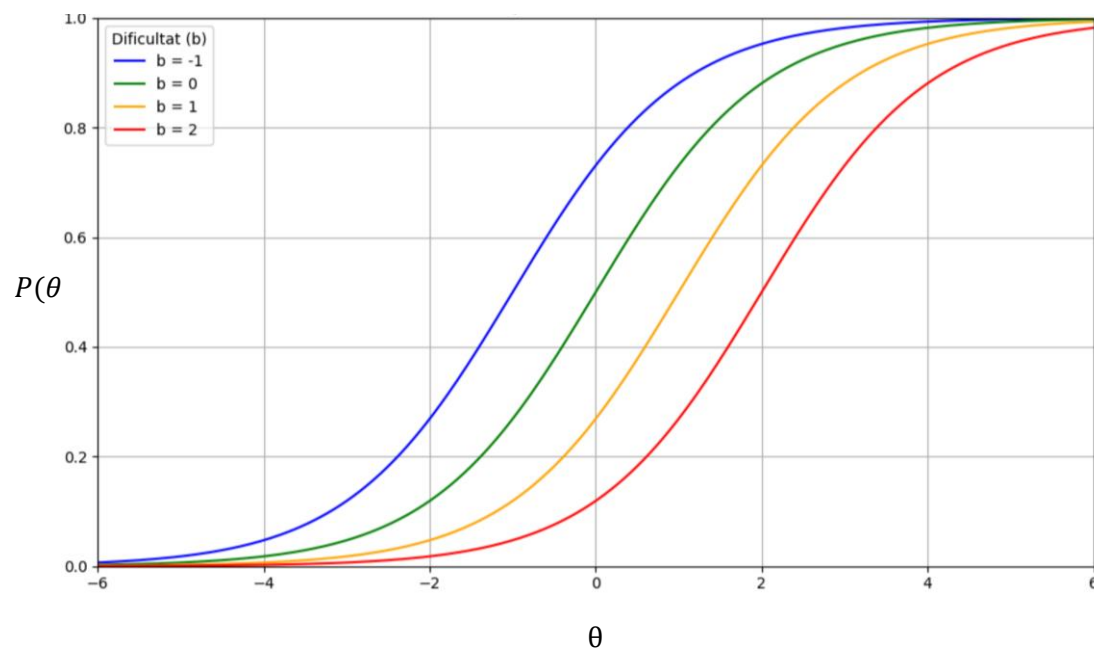
Ítem	<i>a</i>	<i>b</i>	<i>c</i>
1	0,50	2,69	0,34
2	0,88	-1,30	0,15
3	1,63	0,32	0,13
4	0,50	0,45	0,02
5	2,00	-2,43	0,14

7.1. Corba característica de l'ítem (CCI) i models. Solucions

1. En la figura 1 es presenten les CCI dels quatre ítems que formen un test. A partir de la figura 1, respon a les preguntes següents:

Figura 1

CCI dels quatre ítems d'un test.



Notes: línia blava = ítem 1; línia verda = ítem 2; línia groga = ítem 3; línia roja = ítem 4.

Quin és l'ítem més difícil? I el més fàcil?

L'ítem més difícil és el 4 perquè té l'índex de dificultat (paràmetre *b*) més alt.
L'ítem més fàcil és l'1 perquè té l'índex de dificultat més baix.

Quin nivell de tret es necessita per a encertar l'ítem 3?

Es necessita un nivell de tret de +1 per a encertar l'ítem 3, ja que és el valor que es pren per a la probabilitat: $b = 0,50$ ($P(+1) = 0,50$).

Quin ítem és més discriminant?

Aparentment, tots els ítems són igualment discriminants (paràmetre a), perquè el seu pendent és igual quan el tret pren el valor de probabilitat d'encertar l'ítem de 0,50 ($\theta = b$).

Quin ítem té més probabilitat d'encert per atzar?

Aparentment, cap ítem no es pot encertar per atzar. En tots els ítems, el valor de la probabilitat d'encertar, $P(\theta)$, quan θ tendeix a $-\infty$ és 0.

Tenint en compte les respostes donades a les preguntes anteriors, quin model logístic seria adequat per a estudiar el comportament dels ítems?

Tenint en compte que els ítems no varien en els paràmetres a i c , només en el paràmetre b , la resposta a un ítem només depèn de la competència del subjecte (θ) i de la dificultat de l'ítem (b). Per tant, el model més adequat seria el model logístic d'un paràmetre (ML1P).

- 5. Un subjecte amb un determinat nivell d'habilitat té la probabilitat de 0,70 d'encertar l'ítem 1 i la probabilitat de 0,60 d'encertar l'ítem 2. Tenint en compte que els ítems són localment independents, respon a les preguntes següents:**

Quina és la probabilitat que encerte els dos ítems?

Com que es compleix el supòsit d'independència local, la probabilitat d'encertar els ítems que componen un test és el producte de les probabilitats d'encertar cadascun dels ítems.

$$P1(\theta) \cdot P2(\theta) = (0,70) \cdot (0,60) = 0,42$$

Quina és la probabilitat que encerte el primer ítem i que falle el segon?

Com que es compleix el supòsit d'independència local, la probabilitat d'encertar un ítem i fallar l'altre és el producte de les probabilitats d'encertar o fallar cadascun dels ítems per separat.

$$Q(\theta) = 1 - P(\theta)$$

Per tant, la probabilitat de fallar l'ítem 2 és:

$$Q2(\theta) = 1 - P2(\theta) = 1 - 0,60 = 0,40.$$

Així, la probabilitat d'encertar el primer ítem i fallar el segon és:

$$P1(\theta) \cdot Q2(\theta) = (0,70) \cdot (0,40) = 0,28$$

Quina és la probabilitat que falle el primer ítem i que encerte el segon?

Com que es compleix el supòsit d'independència local, la probabilitat de fallar el primer ítem i encertar el segon és el producte de les probabilitats d'encertar o fallar cadascun dels ítems per separat.

La probabilitat de fallar l'ítem 1 és $Q1(\theta) = 1 - P1(\theta) = 1 - 0,70 = 0,30$.

Així, la probabilitat de fallar el primer ítem i encertar el segon és:

$$Q1(\theta) \cdot P2(\theta) = (0,30) \cdot (0,60) = 0,18$$

Quina és la probabilitat que falle els dos ítems?

Com que es compleix el supòsit d'independència local, la probabilitat de fallar els dos ítems és el producte de les probabilitats d'encertar o fallar cadascun dels ítems per separat.

$$Q1(\theta) \cdot Q2(\theta) = (0,30) \cdot (0,40) = 0,12$$

Així, la probabilitat de fallar els dos ítems és 0,12.

- 2. Mitjançant el model logístic de dos paràmetres (ML2P) s'han estimat els valors següents dels paràmetres per a un dels ítems d'un test: $a = 1,50$, $b = 1$. Per al nivell $\theta = +2$, quina és la probabilitat d'encertar aquest ítem?**

$$\begin{aligned} P_1(\theta = +2) &= \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}} = \frac{2,72^{(1,7) \cdot (1,50) \cdot (+2 - 1)}}{1 + 2,72^{(1,7) \cdot (1,50) \cdot (+2 - 1)}} = \frac{2,72^{2,55}}{1 + 2,72^{2,55}} = \frac{12,81}{1 + 12,81} \\ &= \frac{12,81}{13,81} = 0,93 \end{aligned}$$

Per al nivell de tret $\theta = +2$, la probabilitat d'encertar l'ítem és aproximadament 0,93 ($P_1(\theta = +2) = 0,93$).

- 3. Apliquem un test de cinc ítems a una mostra i obtenim els paràmetres que es mostren a la taula 1. Tenint en compte la informació que es presenta a la taula 1, respon a les preguntes següents:**

Taula 1

Paràmetres estimats per als cinc ítems d'un test.

Ítem	<i>a</i>	<i>b</i>	<i>c</i>
1	0,50	2,69	0,34
2	0,88	-1,30	0,15
3	1,63	0,32	0,13
4	0,50	0,45	0,02
5	2,00	-2,43	0,14

Quin model has d'usar per a poder representar les dades de la taula 1?

El model logístic de tres paràmetres (ML3P), atès que hi ha probabilitat d'encertar els ítems per atzar.

Si ens centrem en l'ítem 5, quina és la probabilitat d'encertar-lo per atzar?

La probabilitat d'encertar per atzar l'ítem 5 és de 0,14.

Quin ítem és el més fàcil i per què?

D'acord amb els valors dels paràmetres, l'ítem 5 és l'ítem més fàcil perquè el seu paràmetre de dificultat és -2,43 ($b = -2,43$).

Quin ítem és el més difícil i per què?

L'ítem 1 és l'ítem més difícil perquè el seu paràmetre de dificultat és +2,69 ($b = +2,69$).

Quin ítem té més poder discriminant i per què?

D'acord amb els valors dels paràmetres, l'ítem amb més capacitat discriminant és l'ítem 5 ($a = 2$).

Quin ítem té menys poder discriminant i per què?

D'acord amb els valors dels paràmetres, els ítems amb menys capacitat discriminant són l'ítem 1 i el 4 ($a = 0,50$).

7.2. Qüestions generals sobre la TRI

Respon a les preguntes següents.

1. Quins són els supòsits de la TRI?
2. Per què es diu que si es compleix la unidimensionalitat es compleix també la independència local?
3. Quin model de la TRI preveu valors distints per al paràmetre d'endevinament?
4. Quina és la tècnica que se sol emprar per a comprovar si es compleix el supòsit d'unidimensionalitat?
5. Quines són les fases de l'estimació de paràmetres?
6. Una vegada aplicat un model de TRI i estimats els paràmetres pertinents, què hem de fer amb el model estimat?
7. Quina informació ens proporciona una funció d'informació d'un test? Quin és el seu anàleg (equivalent) en la TCT?
8. Calculem la funció d'informació dels ítems i del test d'un test format per set ítems i obtenim les dades que es mostren a la figura 1. A partir de la figura 1, respon a les preguntes següents:

Quin ítem és el que proporciona més informació?

I l'ítem que proporciona menys informació?

Quin ítem seleccionaries per a aplicar-lo a un subjecte amb un nivell de tret mitjà ($\theta = 0$)?

Si ens fixem en la funció d'informació del test, en quins nivells de tret podem dir que és més precís aquest test?

Creus que és un bon test per a avaluar subjectes que tinguen nivells de tret molt baixos (per exemple, -2)?

9. La taula 1 mostra les funcions d'informació de dos tests per a distints nivells de tret. Tenint en compte la informació inclosa en la taula 1, respon a les preguntes següents:

Quin test és més precís per a un nivell de tret de -1 ? Quina és l'eficiència relativa del test B per a aquest nivell de tret?

Quin test és més precís per a un nivell de tret de +1? Quina és l'eficiència relativa del test A per a aquest nivell de tret?

- 10. La figura 2 mostra la corba característica del test (CCT) que s'ha calculat per a un test de vuit ítems. Tenint en compte els resultats de la taula 2, respon a les preguntes següents:**

Quina puntuació aproximada prediu el test a un subjecte amb un nivell de tret de -1?

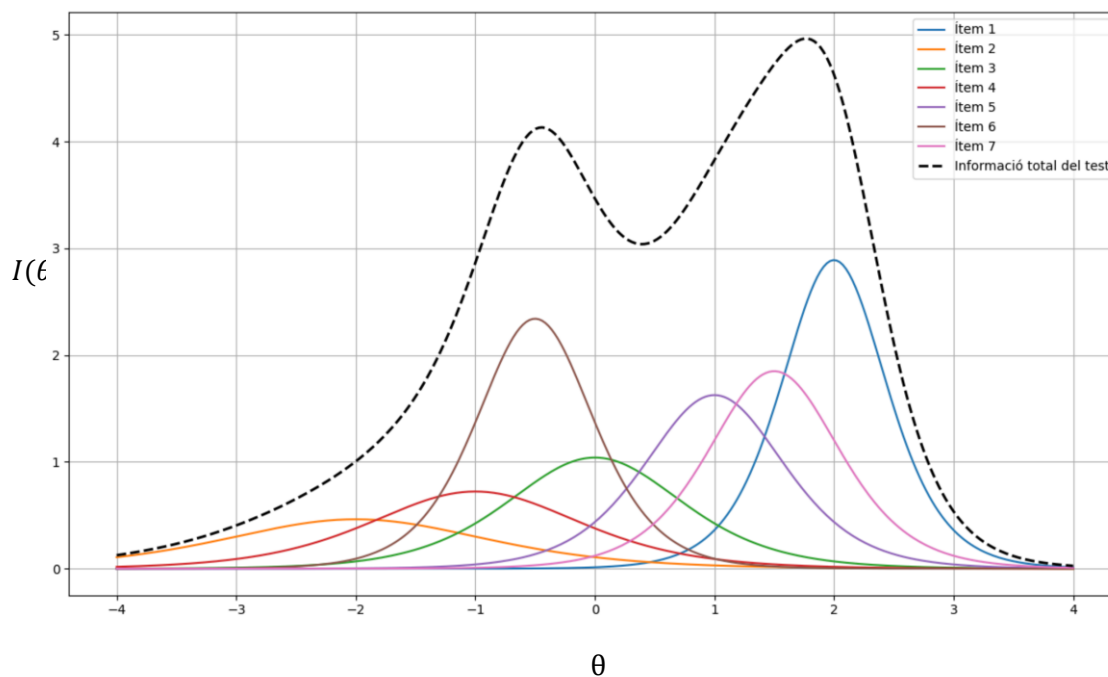
I a un subjecte amb un nivell de tret de +2?

Quin nivell de tret és necessari, aproximadament, per a encertar els vuit ítems que formen el test?

- 11. Digues dues diferències entre la TRI i la TCT.**
- 12. Quines són les tres principals limitacions de la TRI?**
- 13. Quines són les dues aplicacions principals derivades de la TRI?**

Figura 1

Funcions d'informació dels ítems i del test per a un test de set ítems.



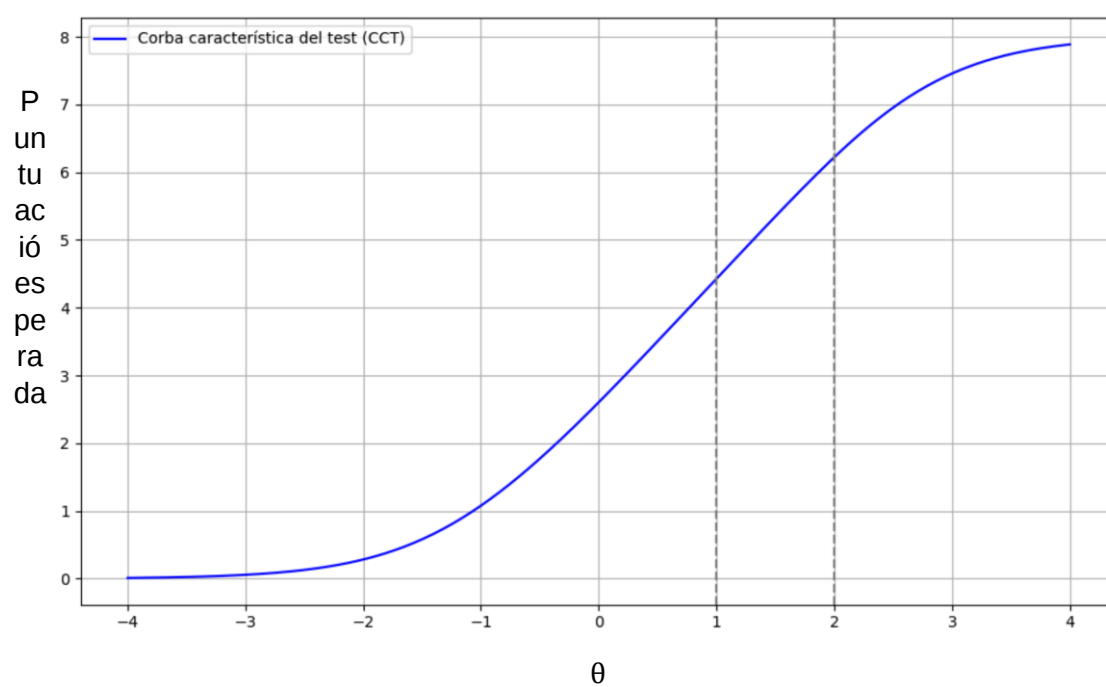
Taula 1

Funcions d'informació de dos tests per a cinc nivells de tret.

θ	Test A	Test B
-2	0,60	7,36
-1	2,31	3,88
0	4,52	1,29
+1	8,30	0,83
+2	2,54	0,23

Figura 2

Exemple de CCT d'un test de vuit ítems.



Qüestions generals sobre la TRI. Solucions

1. Quins són els supòsits de la TRI?

Els supòsits de la TRI són:

1. La unidimensionalitat.
2. La independència local.
3. La invariància dels paràmetres dels ítems.
4. La invariància del paràmetre tret.
5. La continuïtat de l'espai latent.
6. La monotonicitat simple.

2. Per què es diu que si es compleix la unidimensionalitat es compleix també la independència local?

Perquè el supòsit d'unidimensionalitat pressuposa que tots els ítems que componen el test mesuren un mateix i únic tret, és a dir, que els ítems mesuren una única dimensió. Quan això ocorre, les respostes d'una persona a l'ítem només depenen dels paràmetres d'aquest ítem (els paràmetres que hem vist que defineixen la seua CCI) i del nivell de tret que tinga el subjecte (θ). Per tant, no l'afecten les respostes que haja donat a altres ítems, que és el que implica el supòsit d'independència local: la independència local significa que la resposta d'una persona a l'ítem d'un test no depèn de les seues respostes als altres tests.

3. Quin model de la TRI preveu valors diferents per al paràmetre d'endevinament?

El model logístic de tres paràmetres (ML3P), en el qual es pressuposa que la resposta dels subjectes als ítems depèn, a més del seu nivell de tret, de la dificultat de l'ítem i de la discriminació de l'ítem, de la probabilitat d'encert per atzar (c).

4. Quina és la tècnica que se sol emprar per a comprovar si es compleix el supòsit d'unidimensionalitat?

La tècnica de l'anàlisi factorial és la que se sol emprar per a comprovar si es compleix el supòsit d'unidimensionalitat, que serveix per a identificar els factors latents que expliquen els ítems. Direm que es compleix el supòsit d'unidimensionalitat, i per tant el d'independència local, si a la base dels ítems del test hi ha un únic factor latent que explique el comportament dels ítems.

5. Quines són les fases de l'estimació de paràmetres?

S'han de dur a terme dues fases. En la primera, *fase de calibratge dels ítems*, es calculen conjuntament els paràmetres dels ítems (a , b , c) i del tret (θ). És la més complicada perquè requereix estimar tots els paràmetres i, per tant, requereix també grans mides mostrals. Hem de dur-la a terme si desconexem tant els paràmetres dels ítems (a , b i c) com els de les persones (θ).

En cas que els ítems ja estiguen calibrats, aleshores la fase que durem a terme quan apliquem el test és la *fase d'estimació del nivell de tret*. Aquesta fase s'implementa quan ja sabem quins són els paràmetres dels ítems, podem usar-los en les noves aplicacions del test per a estimar el nivell de tret de les persones, i només cal estimar el nivell de tret dels subjectes quan apliquem el test.

6. Una vegada aplicat un model de TRI i estimats els paràmetres pertinents, què hem de fer amb el model estimat?

Quan ja hem aplicat el model de TRI i hem estimat els paràmetres pertinents, cal fer l'*avaluació de l'ajust del model*, que consisteix a avaluar en quina mesura els resultats pronosticats pel model coincideixen amb els obtinguts de fet.

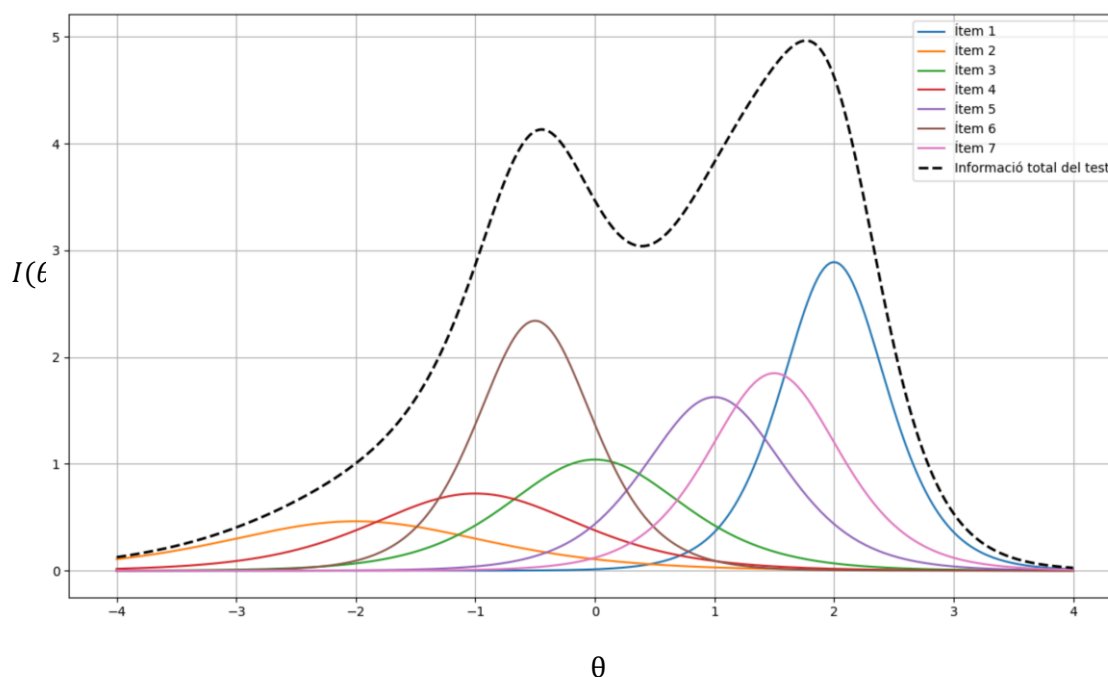
7. Quina informació ens proporciona una funció d'informació d'un test? Quin és el seu anàleg (equivalent) en la TCT?

La funció d'informació indica per a quins nivells d'habilitat (θ) l'ítem o el test proporcionen mesures més precises. És a dir, és un indicador de la precisió de l'ítem o del test. En el cas del test, l'anàleg (equivalent) en la TCT seria el coeficient de fiabilitat, ja que també ens dona informació sobre com de precís és el test quan mesura.

8. Calculem la funció d'informació dels ítems i del test d'un test format per set ítems i obtenim les dades que es mostren a la figura 1. A partir de la figura 1, respon a les preguntes següents:

Figura 1

Funcions d'informació dels ítems i del test per a un test de set ítems.



Quin ítem proporciona més informació?

L'ítem 1, perquè té una $I_1(\theta = 2)$ d'aproximadament 3.

I l'ítem que proporciona menys informació?

L'ítem 2, perquè té una $I_2(\theta = -2)$ d'aproximadament 0,50.

Quin ítem seleccionaries per a aplicar-lo a un subjecte amb un nivell de tret mitjà ($\theta = 0$)?

L'ítem 3 o l'ítem 6, perquè són els que més informació donen per a $\theta = 0$.

Si ens fixem en la funció d'informació del test, en quins nivells de tret podem dir que aquest test és més precís?

El test és més precís quan avalua nivells de tret d'aproximadament +2, ja que és on la funció del test arriba al seu màxim i, per tant, on dona més informació sobre el nivell de tret dels subjectes. No obstant això, també és un bon test per a avaluar persones mitjanes-baixes en el tret (−0,50).

Creus que és un bon test per a avaluar subjectes que tinguen nivells de tret molt baixos (per exemple, -2)?

No és un bon test per a avaluar subjectes que tinguen nivells de tret molt baixos, ja que és un test que dona menys informació en les parts baixes del tret, per comparació amb la informació que dona quan s'aplica a nivells de tret alts.

9. A la taula 1 es poden consultar les funcions d'informació de dos tests per a distints nivells de tret. Tenint en compte la informació inclosa en la taula 1, respon a les preguntes següents:

Taula 1

Funcions d'informació de dos tests per a cinc nivells de tret.

θ	Test A	Test B
-2	0,60	7,36
-1	2,31	3,88
0	4,52	1,29
+1	8,30	0,83
+2	2,54	0,23

Quin test és més precís per a un nivell de tret de -1? Quina és l'eficiència relativa del test B per a aquest nivell de tret?

El test més precís per a un nivell de tret de -1 és el test B, ja que ofereix més informació per a aquest nivell de tret.

$$ER = \frac{I(\theta_B)}{I(\theta_A)} = \frac{3,88}{2,31} = 1,68$$

El test B dona 1,68 vegades més informació que el test A per a un nivell de tret de -1.

Quin test és més precís per a un nivell de tret de +1? Quina és l'eficiència relativa del test A per a aquest nivell de tret?

El test més precís per a un nivell de tret de +1 és el test A, ja que ofereix més informació per a aquest nivell de tret.

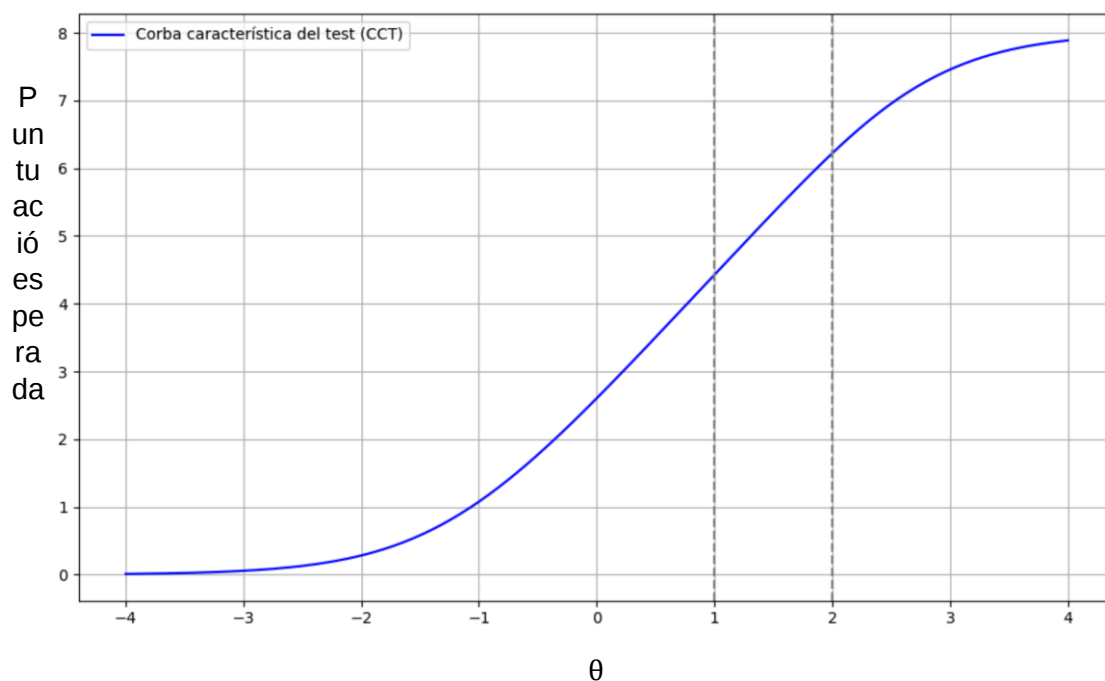
$$ER = \frac{I(\theta_A)}{I(\theta_B)} = \frac{8,30}{0,83} = 10,4$$

El test A dona 10 vegades més informació que el test B per a un nivell de tret de -1.

10. La figura 2 mostra la corba característica del test (CCT) que s'ha calculat per a un test de vuit ítems. Tenint en compte els resultats de la taula 2, respon a les preguntes següents:

Figura 2

Exemple de CCT d'un test de vuit ítems.



Quina puntuació aproximada prediu el test a un subjecte amb un nivell de tret de -1?

El test prediu una puntuació observada d'aproximadament 1 a un subjecte amb un nivell de tret de -1.

I a un subjecte amb un nivell de tret de +2?

El test prediu una puntuació observada d'aproximadament 6 a un subjecte amb un nivell de tret de +2.

Quin nivell de tret és necessari, aproximadament, per a encertar els vuit ítems que formen el test?

És necessari un nivell de tret de +4 o superior per a encertar els vuit ítems del test.

11. Digues dues diferències entre la TRI i la TCT.

Algunes de les principals diferències entre la TRI i la TCT són:

1. En la TRI la unitat d'anàlisi és l'ítem, mentre que en la TCT la unitat d'anàlisi és el test.
2. La TRI estableix una funció no lineal entre la puntuació observada i la puntuació en el tret (la corba característica del test, CCT), mentre que la TCT estableix una relació lineal entre la puntuació observada i la puntuació vertadera.
3. En la TRI, els subjectes i els ítems se situen en la mateixa escala de mesura, mentre que en la TCT no (l'escala dels subjectes és el tret, però l'escala dels ítems és la mètrica del test).
4. La TRI té supòsits verificables, com ara la unidimensionalitat i la independència local, mentre que en la TCT els supòsits no es poden verificar.
5. En la TRI els paràmetres dels subjectes avaluats són invariants, tenen la mateixa mètrica, la mètrica del tret, a través de tests, mentre que en la TCT els paràmetres dels subjectes varien segons el test emprat per a avaluar-los, ja que estan en la mètrica del test.
6. En la TRI els paràmetres dels ítems són invariants, ja que, si es compleixen els supòsits del model, els paràmetres dels ítems no depenen de la mostra on s'obtenen. En la TCT els paràmetres del test varien segons la mostra on es calculen.
7. En la TRI no es pressuposa l'homoscedasticitat de l'error: la precisió del test en la TRI està condicionada al nivell de tret i als ítems aplicats, i pot ser més gran, en un mateix ítem, per a un cert nivell de tret i més petita per a un altre. En la TCT es pressuposa un error incondicionat, igual per a tots els nivells de tret de les persones avaluades.
8. En la TRI es disposa d'indicadors de bondat d'ajust (estadístic khi quadrat i anàlisi de residus), mentre que en la TCT no es pot avaluar l'ajust del model.

12. Quines són les tres principals limitacions de la TRI?

Requereix mostres grans, requereix crear un nombre elevat d'ítems i té supòsits molt restrictius.

13. Quines són les dues aplicacions principals derivades de la TRI?

El desenvolupament dels tests adaptatius informatitzats (TAI) i l'estudi del funcionament diferencial dels ítems (DIF).