TERCERA ACTIVITAT D'AVALUACIÓ CONTINUADA: "FIABILITAT"

Gemma Pérez Moliner

1.1.

Encara que els dos tests tindran característiques similars (ja que els ítems afegits al test A provenen de formes paral·leles d'aquest), no podem considerar els tests A i B com a formes paral·leles; si bé ambdós tests avaluarien el mateix contingut i podrien tenir mitjanes iguals, no es compleix que tinguin el mateix nombre d'ítems (el test B tindria el triple d'ítems que l'A).

1.2.

60 ítems.

1.3.

40 ítems. 20 de cada una de les formes paral·leles.

1.4.

Aplicarem la Profecia de Spearman-Brown que ens permet de relacionar la fiabilitat de dos tests compostos per un nombre diferent d'ítems, i en què la fiabilitat dels ítems (pel fet de pertànyer a formes paral·leles) té el mateix valor:

Profecia de Spearman-Brown: $\rho_h = \frac{k\rho_c}{1+(k-1)\rho_c}$; on la fiabilitat del test hipotètic és: $\rho_h = 0{,}80$ (perquè és la fiabilitat del nou test construït amb el triple d'ítems que el test original); i on la relació entre el nombre d'ítems d'un i altre test és: $k = \frac{n_h}{n_c} = \frac{60}{20} = 3$; llavors:

$$\rho_h = \frac{k\rho_c}{1+(k-1)\rho_c}; \quad 0{,}80 = \frac{3\rho_c}{1+(3-1)\rho_c}; \quad \rho_c = 0{,}571$$

la fiabilitat del test A (abans d'afegir 40 ítems) era molt pitjor que la fiabilitat del nou test, valia 0,571.

2.1.

La comparació de les puntuacions entre la primera passació i la segona passació mitjançant l'obtenció del coeficient de correlació (Pearson si es tracta de dades quantitatives, Spearman si són dades ordinals i kappa de Cohen o phi, si són dades categòriques).

Es tracta d'un disseny test-retest, ja que el psicòleg ha passat la prova als alumnes una primera vegada a començament de curs i, ha passat la mateixa prova una segona vegada al cap de dues setmanes.

L'objectiu de la comparació de puntuacions és esbrinar si existeix una relació entre les puntuacions obtingudes en la primera passació i les obtingudes en la segona. Donat que es tracta del mateix test, les puntuacions obtingudes en els dos moments haurien de resultar coherents.

2.2.

Es podria utilitzar l'estratègia de les dues meitats, especialment si l'examen és suficientment llarg (no té un nombre d'ítems massa petit) i si el nombre de preguntes és parell.

Suposem que aquest examen de psicofarmacologia pretén avaluar els coneixements de tota l'assignatura, per tant, que les preguntes cobreixen tots els conceptes de la mateixa. I també suposem que totes les preguntes són equivalents en quant a dificultat i puntuació.

Si l'examen és una mesura fiable pel que fa als coneixements de psicofarmacologia, al dividir aquest examen en dues meitats, les puntuacions obtingudes en cada una de les meitats haurien de ser similars.

La comparació (mitjançant un coeficient de correlació) de les dues puntuacions ens hauria de dir si existeix relació entre les preguntes de les dues parts i, seria un indicador de la fiabilitat de la meitat de l'examen. Hauríem d'aplicar la fórmula general de Spearman-Brown (o bé la de Guttman-Flanagan) per a conèixer la fiabilitat de la totalitat del test.

Val a dir que en aquest cas el disseny test-retest no seria adequat ja que els alumnes de farmacologia podrien recordar les respostes donades en la primera passació, produint-se així una millora de les puntuacions entre una passació i l'altra.

Depenent de la llargada de l'examen, tampoc no seria adequada la passació d'una forma paral·lela d'aquest examen, ja que podria produir un biaix de les puntuacions del segon test per cansament dels alumnes.

2.3.

Per recolzar un informe mèdic que s'ha fet, volem esbrinar si un grup de treballadors de la mateixa empresa presenta un símptoma concret d'estrès no detectable a través d'una exploració mèdica (símptoma psicològic) i que podria ser origen d'una malaltia física. Per a això, demanem els serveis d'un psicòleg que emprèn una sèrie d'entrevistes amb els treballadors afectats.

Durant les entrevistes, el psicòleg anota si detecta o no aquest símptoma en els treballadors, a través de les respostes directes o indirectes dels mateixos.

Una setmana més tard, un altre psicòleg (independent del primer) repeteix la mateixa operació. A través de les entrevistes, anota si creu que cada un dels treballadors presenta o no el símptoma elegit.

Volem "avaluar" la fiabilitat entre avaluadors (que actuen com a instrument de mesura del símptoma d'estrès).

La ocurrència o no del símptoma psicològic constitueix una variable categòrica. Per a calcular la correlació entre les observacions dels dos psicòlegs, i per tant la fiabilitat de les mesures, faríem servir el coeficient fi o el kappa de Cohen.

Es tracta d'una variant de la comparació entre instruments equivalents. Els dos psicòlegs representarien aquests instruments amb característiques similars (en formació, orientació teòrica i objectius); i per saber si hi ha coherència entre les dues observacions hem d'aplicar el corresponent coeficient de correlació, en aquest cas el que sigui adequat per a dades qualitatives.

3.1.

En separar els ítems senars dels parells per obtenir dues meitats de l'escala N, obtenim dues sumes de puntuacions totals: la suma de les puntuacions dels ítems 12, 18 i 22 i la suma de puntuacions dels ítems 13, 15 i 19; ambdues columnes amb 497 dades, igual que el nombre d'individus avaluats.

Calculem el coeficient de correlació (Pearson per ser dades quantitatives) entre aquestes dues noves matrius de dades. El resultat és: 0,665.

Aquest resultat és indicador de la fiabilitat de la meitat de l'escala N.

Per a calcular la fiabilitat de la puntuació total de l'escala, apliquem una correcció; en aquest cas la fórmula de Guttman-Flanagan ja que les dues matrius tenen variàncies diferents:

$$r_{total} = 2\left(1 - \frac{\sigma_A^2 + \sigma_B^2}{\sigma_T^2}\right) = 2\left(1 - \frac{1{,}050 + 1{,}265}{3{,}848}\right) = 0{,}797$$

Sabem que en analitzar la fiabilitat d'un test a través del disseny de mesures equivalents, se solen obtenir valors baixos de fiabilitat. En aquest cas, tenint en compte que el valor obtingut (0,797) ja és bastant alt (molt proper a 0,8), podríem dir que la fiabilitat d'aquest test és molt bona.

3.2.

Pretenem avaluar la fiabilitat de l'escala N a partir de la consistència interna entre cada un dels ítems que la componen, així que, per començar, veurem quines són les correlacions entre aquests ítems:

matriu de correlacions entre els ítems

	12	**13**	**15**	**18**	**19**	**22**
12						
13	0,228					
15	0,425	0,445				
18	0,210	0,323	0,353			
19	0,525	0,274	0,400	0,262		
22	0,244	0,412	0,398	0,298	0,303	

la mitjana de les correlacions: 0,340; és indicador de la fiabilitat de cada un dels sis ítems.

Per a calcular la fiabilitat de la puntuació total de l'escala, apliquem una correcció; la fórmula de Spearman-Brown generalitzada al cas en què el test es divideix en un nombre qualsevol de parts:

$r_{total} = \dfrac{k\bar{r}_{parts}}{1+(k-1)\bar{r}_{parts}}$; on k és igual al nombre d'ítems, i $\bar{r}_{parts}$ és el coeficient de fiabilitat de cada un d'ells; llavors:

$$r_{total} = \frac{k\bar{r}_{parts}}{1+(k-1)\bar{r}_{parts}} = \frac{6\cdot 0{,}340}{1+(6-1)\cdot 0{,}340} = 0{,}756$$

Podem considerar que el valor de la fiabilitat obtingut (0,756), és acceptable al estar per sobre de 0,7.

Ara bé, en els casos en que els ítems no tinguin la mateixa variància (com és el cas), haurem d'aplicar el coeficient alfa de Cronbach; i com que els ítems que componen el test són dicotòmics (en aquest cas prenen valors si/no) utilitzarem la variant proposada per Kuder i Richardson:

$$\alpha = KR20 = \frac{k}{k-1}\left(1-\frac{\sum p_j q_j}{\sigma_t^2}\right)$$

per a fer els diferents càlculs, construirem la matriu de variàncies i covariàncies:

	12	13	15	18	19	22
12	0,244					
13	0,056	0,244				
15	0,104	0,109	0,248			
18	0,050	0,077	0,085	0,235		
19	0,125	0,065	0,096	0,061	0,231	
22	0,057	0,095	0,093	0,068	0,068	0,221

D'on la variància de la puntuació total és: $\sigma_t^2 = \sum \sigma_j^2 + 2\cdot \sum \sigma_{jj'} = 1{,}424 + (2\cdot 1{,}210) = 3{,}844$; per l'obtenció de la qual hem fet servir el sumatori de la meitat de les covariàncies: $\sum \sigma_{jj'} = 1{,}210$; i el sumatori de les variàncies: $\sum \sigma_j^2 = 1{,}424$;

També, les proporcions de respostes codificades com a 1 (p_j) i com a 0 (q_j) són:

	12	13	15	18	19	22
Respostes totals	497	497	497	497	497	497
Respostes 1	210	209	225	187	180	163
p_i	0,423	0,421	0,453	0,376	0,362	0,328
Respostes 0	287	288	272	310	317	334
q_i	0,577	0,579	0,547	0,624	0,638	0,672
$p_i \times q_i$	0,244	0,244	0,248	0,235	0,231	0,220

i així:

$$\alpha = KR20 = \frac{k}{k-1}\left(1-\frac{\sum p_j q_j}{\sigma_t^2}\right) = \frac{6}{6-1}\left(1-\frac{1{,}422}{3{,}844}\right) = 0{,}756$$

de tota manera, com podem observar en les dues taules, el producte de les proporcions de respostes 1 i respostes 0 és igual a la variància de cada ítem, per tant, la fórmula general de l'alfa de Cronbach ens hauria de donar el mateix resultat que la fórmula de Kuder i Richardson:

$$\alpha = \frac{k}{k-1}\left(1-\frac{\sum \sigma_j^2}{\sigma_t^2}\right) = \frac{6}{6-1}\left(1-\frac{1{,}424}{3{,}844}\right) = 0{,}755$$

o bé:

$$\alpha = \frac{k}{k-1}\left(\frac{\sum \sigma_{jj'}}{\sigma_t^2}\right) = \frac{6}{6-1}\left(\frac{2{,}420}{3{,}844}\right) = 0{,}755$$

o bé:

$\alpha = \frac{k^2 \cdot \overline{\sigma}_{jj'}}{\sigma_t^2} = \frac{6^2 \cdot 0{,}081}{3{,}844} = 0{,}755$; pel qual hem fet servir la mitjana de les covariàncies: $\overline{\sigma}_{jj'} = 0{,}081$

Totes quatre maneres de calcular l'alfa de Cronbach ens donen el valor $\approx 0{,}756$, el qual, per ser un valor més proper a 1 que a 0 ja ens indica que els ítems del test estan relacionats entre sí, és adir, que existeix consistència interna. També, el fet que el valor no sigui massa proper a 1, ens indicaria que els ítems no són redundants i que representen bé el constructe.

Seguint Nunnally (1978), podríem dir que el valor trobat (0,756) garanteix una consistència interna acceptable.

4.

Farem servir la profecia de Spearman-Brown que, com hem dit abans, relaciona la fiabilitat de dos tests que tenen diferent nombre d'ítems (en aquest cas un primer test amb 6 ítems i un segon test amb 6+3 ítems).

Profecia de Spearman-Brown: $\rho_h = \frac{k\rho_c}{1+(k-1)\rho_c}$; on la fiabilitat del test conegut és: $\rho_c = 0{,}756$; i on la relació entre el nombre d'ítems d'un i altre test és: $k = \frac{n_h}{n_c} = \frac{9}{6} = 1{,}5$; llavors:

$$\rho_h = \frac{k\rho_c}{1+(k-1)\rho_c} = \frac{1{,}5 \cdot 0{,}756}{1+(1{,}5-1)\cdot 0{,}756} = 0{,}823$$

la fiabilitat de l'escala N després d'afegir 3 ítems similars als que ja teníem, és igual a 0,823; és a dir, la fiabilitat de l'escala ha augmentat en afegir-hi 3 ítems.

5.

Tornem a aplicar la profecia de Spearman-Brown; aquesta vegada amb la dada coneguda de la fiabilitat que volem obtenir ($\rho_h = 0{,}90$).

$\rho_h = \frac{k\rho_c}{1+(k-1)\rho_c}; \quad 0{,}90 = \frac{k \cdot 0{,}756}{1+(k-1)\cdot 0{,}756}; \quad k = 2{,}902$; és a dir, el nou test hauria de ser 2,902 vegades més llarg que l'actual; d'on

$$k = \frac{n_h}{n_c}; \quad 2{,}902 = \frac{n_h}{6}; \quad n_h = 17{,}412 \approx 17$$

Si volguéssim que l'escala N tingués una fiabilitat de 0,90 hauríem d'afegir-li 11 ítems similars als 6 que ja teníem. El nou test tindria 17 ítems.

6.

El segon subjecte del fitxer de dades de l'escala N ha obtingut una puntuació observada de 3.

Tenint en compte el valor de l'alfa de Cronbach i la variància de les puntuacions totals observades, obtinguts en l'activitat 3 ($\alpha = 0{,}756 \; i \; \sigma_t^2 = 3{,}844$) calcularem l'interval en que, amb un nivell de confiança del 95%, es troben les puntuacions vertaderes relacionades amb les puntuacions observades en l'escala N.

$$IC_v = X \pm 1{,}96\sigma_e = 3 \pm 1{,}96 \cdot \left(1{,}96\sqrt{1-0{,}756}\right) \Rightarrow$$

$$\Rightarrow IC_v = 3 + 1{,}897 = 4{,}897 \approx 5$$

$$\Rightarrow IC_v = 3 - 1{,}897 = 1{,}103 \approx 1$$

Podem dir que la puntuació vertadera obtinguda pel segon subjecte en l'escala N de l'EPQ-RA, amb un 95% de probabilitat, es troba entre 1 i 5 punts.

******