

Estadística Aplicada a la Educación

Tema 8

Tutor.

UNED Madrid-Sur (A.U. Parla)

Miguel Ángel Daza

migdaza@madridsur.uned.es

1

- La Estadística en el proceso de investigación pedagógica empírica.

2

- Problema, hipótesis / objetivos, variables y datos. Niveles de medida

4

- Organización de los datos. análisis exploratorio de datos.

5

- Reducción de datos. Medidas descriptivas básicas y representaciones gráficas.

6

- Medidas individuales.

7

- Relación entre variables. Las correlaciones. La regresión.

8

- Aplicaciones de la correlación: fiabilidad y validez de las medida.

9

- Modelos estadísticos y probabilidad. La curva normal de probabilidades.

10

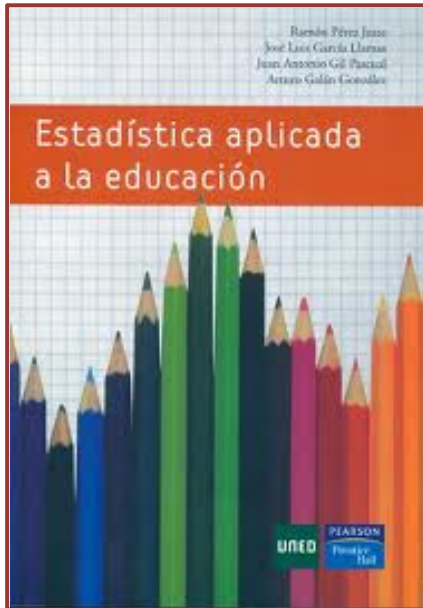
- Los baremos o normas. Muestreo. Aplicaciones.

11

- Estimación de parámetros. Errores de estimación.

12

- Introducción al contraste de hipótesis: la prueba t para el contraste de medias en los diseños de dos grupos.



8.1 Introducción.

8.2 Estudio de la fiabilidad.

8.2.1 Procedimientos para determinar la fiabilidad.

8.2.2 Algunas cuestiones relacionadas con la fiabilidad.

8.3 Estudio de la Validez.

8.3.1 La validez de contenido.

8.3.2 La validez predictiva.

8.3.3 La validez concurrente.

8.3.4 La validez de constructo.

8.3.5 La validez aparente.

8.3.6 Algunos coeficientes relacionados con la validez.

8.3.7 Validez y longitud de una prueba.

8.3.8 Predicción de puntuaciones.

8.4 Estudio de los elementos o ítems de una prueba.

8.4.1 El índice de dificultad (ID).

8.4.2 El índice de homogeneidad (IH).

8.4.3 El índice de validez (IV).

8. APLICACIONES DE LA CORRELACIÓN: FIABILIDAD Y VALIDEZ DE LAS MEDIDA.

8.1 Introducción.

En este capítulo veremos las **características técnicas de los instrumentos de medida**

Es fundamental que los instrumentos sean **precisos en sus apreciaciones (*fiabilidad*) y sirvan para lo que realmente se diseñaron (*validez*)**

En el tema se tratan 3 cuestiones:

- **FIABILIDAD** DE LOS INSTRUMENTOS
- **VALIDEZ** DE LOS INSTRUMENTOS (especialmente la *validez predictiva*)
- **ANÁLISIS DE ELEMENTOS (ÍTEMS)** DE UNA PRUEBA:
 - Índice de Dificultad*
 - Índice de Homogeneidad*
 - Índice de Validez*

8.2 Estudio de la fiabilidad.

La **fiabilidad** de las medidas se identifica con la **precisión**. **Un instrumento es fiable cuando mide algo con precisión**, independientemente de lo que se esté midiendo.

Puntuación observada = Puntuación verdadera \pm error de medida

Los errores pueden ser sistemáticos o aleatorios

Cuanto menor sea el error, más fiable es el instrumento.

También podemos expresar la **fiabilidad** como:

- la **constancia** en las puntuaciones de los sujetos
- o bien la **concordancia** entre varias mediciones sucesivas de una misma realidad.

Para el cálculo de la fiabilidad se procede siempre desde 2 conjuntos de datos a partir de los cuales se puede determinar el coeficiente de correlación que estime la fiabilidad del instrumentos

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como ESTABILIDAD

- O procedimiento de la repetición o del "test-retest"

Fiabilidad como EQUIVALENCIA

- O de "formas paralelas"

Fiabilidad como CONSISTENCIA INTERNA

- Por ejemplo, con las "dos mitades"

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como ESTABILIDAD

También conocida como **procedimiento de la repetición** o del **test-retest**.

Se calcula la **correlación** que existe entre las puntuaciones obtenidas por **un mismo grupo** de sujetos (debidamente identificados) en 2 aplicaciones sucesivas de una misma prueba, **en 2 momentos diferentes**.

Entre ambas aplicaciones debe transcurrir un tiempo:

ni muy corto (por el entrenamiento)

ni muy largo (por la maduración).

Algunos autores lo fijan en 20 ó 25 días.

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como ESTABILIDAD

Ejemplo 8.1

“Un investigador está interesado en construir una buena prueba objetiva para evaluar a los estudiantes de Estadística en la Facultad de Educación, para ello construye una prueba que consta de 20 ítems. Realiza una primera aplicación a una muestra de 40 sujetos, transcurridos 22 días, vuelve a aplicar la misma prueba a los mismos 40 sujetos, obteniendo los resultados en la siguiente tabla”.

8.2.1 Procedimientos para determinar la fiabilidad.

Ejemplo 8.1

Fiabilidad como ESTABILIDAD

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

8.2.1 Procedimientos para determinar la fiabilidad.

Ejemplo 8.1

Fiabilidad como ESTABILIDAD

Sujetos	1ª	2ª	X ²	Y ²	XY
	Aplicación	Aplicación			
	X	Y			
1	12	13	144	169	156
2	18	17	324	289	306
3	15	15	225	225	225
4	11	10	121	100	110
5	9	10	81	100	90
6	17	16	289	256	272
7	13	15	169	225	195
8	8	9	64	81	72
9	18	17	324	289	306
10	12	12	144	144	144
11	7	6	49	36	42
12	11	10	121	100	110
13	15	16	225	256	240
14	20	14	400	196	280
15	14	14	196	196	196
16	8	8	64	64	64
17	10	11	100	121	110
18	17	16	289	256	272
19	14	13	196	169	182
20	11	12	121	144	132
21	7	8	49	64	56
22	17	15	289	225	255
23	12	12	144	144	144
24	12	13	144	169	156
25	18	19	324	361	342
26	15	16	225	256	240
27	13	13	169	169	169
28	8	7	64	49	56
29	11	13	121	169	143
30	18	17	324	289	306
31	10	10	100	100	100
32	17	16	289	256	272
33	8	9	64	81	72
34	12	13	144	169	156
35	14	15	196	225	210
36	16	17	256	289	272
37	8	9	64	81	72
38	11	10	121	100	110
39	18	19	324	361	342
40	12	12	144	144	144
	517	517	7201	7117	7121

Sujetos	1ª Aplicación	2ª Aplicación	X ²	Y ²	XY
	X	Y			
	517	517	7201	7117	7121

$$r_{xy} = \frac{n \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] \cdot [n \sum Y^2 - (\sum Y)^2]}}$$

Puntuaciones Directas

$$r_{xy} = \frac{40(7121) - (517)(517)}{\sqrt{[40(7201) - (517)^2] \cdot [40(7117) - (517)^2]}} = 0,924$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como EQUIVALENCIA

También conocida como de **formas paralelas**.

Consiste en aplicar **2 pruebas diferentes** pero que miden el **mismo rasgo** o característica, de tal forma que los resultados de la aplicación de la primera prueba se correlacionan con los de la segunda.

Es un **procedimiento difícil y complicado**: es muy difícil encontrar pruebas equivalentes (mismos objetivos, mismos contenidos, longitud y condiciones de aplicación) También se utiliza el coeficiente de correlación de Pearson.

Si transcurren más de 20 días entre la aplicación de la prueba y la forma paralela, el coeficiente encontrado puede ser considerado como de equivalencia y de estabilidad.

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como EQUIVALENCIA

Ejemplo 8.2

“Queremos determinar la fiabilidad de una prueba de Física, para ello recurrimos a una muestra de 30 sujetos de Bachillerato a los que se aplica una primera prueba que cuenta con 25 ítems, transcurridos 3 días se aplica una prueba equivalente que también consta de 25 ítems cuya dificultad es similar. Los resultados globales son:”.

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como EQUIVALENCIA

Ejemplo 8.2

$$r_{xy} = \frac{n \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] \cdot [n \sum Y^2 - (\sum Y)^2]}}$$

n	X	Y	X ²	Y ²	XY
30					
Sumas	550	570	10270	11070	10645

$$r = \frac{5850}{\text{Raiz}(5600 \cdot 7200)} = \frac{5850}{6349,80} = 0,921$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Los instrumentos de medida están representados por una serie de elementos o ítems...

...cabe esperar que cada uno de ellos medirá una parte del rasgo o característica que mide la prueba en su conjunto...

...por tanto, debe existir una **coherencia** o **consistencia** en las respuestas que ofrecen un conjunto de sujetos a los diferentes elementos que componen el instrumento.

La **fiabilidad como consistencia interna**, también conocida como **de las 2 mitades** consiste en dividir la puntuación total del sujeto en una prueba en 2 partes (lo mejor, "ítems-pares" / "ítems-impares")... ...así obtenemos 2 puntuaciones por cada sujeto, cuya correlación será el coeficiente de fiabilidad como consistencia interna

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.3

“Se ha aplicado una prueba de 30 ítems a una muestra de 12 sujetos de Enseñanza Secundaria. Sobre esos datos queremos determinar el valor de la fiabilidad mediante la consistencia interna de esa misma prueba; para ello, la puntuación de cada sujeto se ha dividido en dos puntuaciones, una se corresponde con los resultados de los ítems impares (primera mitad: variable X) y la otra con los ítems pares (segunda mitad: variable Y). Con esos valores se calcula el coeficiente de consistencia interna”.

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.3

Sujetos	X	Y	X ²	Y ²	XY
1	12	13	144	169	156
2	8	7	64	49	56
3	11	10	121	100	110
4	14	15	196	225	210
5	7	6	49	36	42
6	9	11	81	121	99
7	13	11	169	121	143
8	9	9	81	81	81
9	5	6	25	36	30
10	13	12	169	144	156
11	6	8	36	64	48
12	11	9	121	81	99
sumas	118	117	1256	1227	1230

$$r_{xy} = \frac{n \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] \cdot [n \sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{954}{\text{Raiz}(1148 \cdot 1035)} = \frac{954}{1090,04} = 0,875$$

$$R = \frac{2 \cdot 0,875}{1 + 0,875} = \frac{1,75}{1,875} = 0,933$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.3

A) Procedimiento de Spearman-Brown

Sujetos	X	Y	X ²	Y ²	XY
1	12	13	144	169	156
2	8	7	64	49	56
3	11	10	121	100	110
4	14	15	196	225	210
5	7	6	49	36	42
6	9	11	81	121	99
7	13	11	169	121	143
8	9	9	81	81	81
9	5	6	25	36	30
10	13	12	169	144	156
11	6	8	36	64	48
12	11	9	121	81	99
sumas	118	117	1256	1227	1230

$$r_{xy} = \frac{n \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] \cdot [n \sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{954}{\text{Raiz}(1148 \cdot 1035)} = \frac{954}{1090,04} = 0,875$$

$$R = \frac{2 \cdot 0,875}{1 + 0,875} = \frac{1,75}{1,875} = 0,933$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.3

B) Procedimiento de Rulon

Sujetos	X	Y	X ²	Y ²	XY	d=X-Y	d ²	T=X+Y	T ²
1	12	13	144	169	156	-1	1	25	625
2	8	7	64	49	56	1	1	15	225
3	11	10	121	100	110	1	1	21	441
4	14	15	196	225	210	-1	1	29	841
5	7	6	49	36	42	1	1	13	169
6	9	11	81	121	99	-2	4	20	400
7	13	11	169	121	143	2	4	24	576
8	9	9	81	81	81	0	0	18	324
9	5	6	25	36	30	-1	1	11	121
10	13	12	169	144	156	1	1	25	625
11	6	8	36	64	48	-2	4	14	196
12	11	9	121	81	99	2	4	20	400
sumas	118	117	1256	1227	1230	1	23	235	4943

$$s_d^2 = \frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1} = \frac{23 - \frac{1^2}{12}}{12-1} = 2,08$$

$$s_d^2 = \frac{\sum T^2 - \frac{(\sum T)^2}{n}}{n-1} = \frac{4943 - \frac{(235)^2}{12}}{12-1} = 30,99$$

$$r_{xx} = 1 - \frac{s_d^2}{s_T^2}$$

$$r_{xx} = 1 - \frac{2,08}{30,99} = 1 - 0,067 = 0,93$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.3

C) Procedimiento de Guttman

Sujetos	X	Y	X ²	Y ²	XY	d=X-Y	d ²	T=X+Y	T ²
1	12	13	144	169	156	-1	1	25	625
2	8	7	64	49	56	1	1	15	225
3	11	10	121	100	110	1	1	21	441
4	14	15	196	225	210	-1	1	29	841
5	7	6	49	36	42	1	1	13	169
6	9	11	81	121	99	-2	4	20	400
7	13	11	169	121	143	2	4	24	576
8	9	9	81	81	81	0	0	18	324
9	5	6	25	36	30	-1	1	11	121
10	13	12	169	144	156	1	1	25	625
11	6	8	36	64	48	-2	4	14	196
12	11	9	121	81	99	2	4	20	400
sumas	118	117	1256	1227	1230	1	23	235	4943

$$s_{1a}^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{1256 - \frac{(118)^2}{12}}{12-1} = 8,70$$

$$s_{2a}^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n-1} = \frac{1227 - \frac{(117)^2}{12}}{12-1} = 7,84$$

$$s_d^2 = \frac{\sum T^2 - \frac{(\sum T)^2}{n}}{n-1} = \frac{4943 - \frac{(235)^2}{12}}{12-1} = 30,99$$

$$r_{xx} = 2 \left(1 - \frac{s_{1a}^2 + s_{2a}^2}{s_d^2} \right) \quad r_{xx} = 2 \left(1 - \frac{8,70 + 7,84}{30,99} \right) = 0,93$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.4

D) Procedimiento de Kuder-Richardson

"Tomando como referencia los datos anteriores, es decir, una prueba de 30 ítems, pero suponiendo que son dicotómicos, y después de calcular los valores de p y q de cada ítem, hemos obtenido el valor de $\Sigma pq = 3,70$ "

$$r_{xx} = \left(\frac{n_e}{n_e - 1} \right) \left(\frac{s_t^2 - \sum p \cdot q}{s_t^2} \right)$$

$$s_d^2 = \frac{\sum T^2 - \frac{(\sum T)^2}{n}}{n-1} = \frac{4943 - \frac{(235)^2}{12}}{12-1} = 30,99$$

$$r_{xx} = \left(\frac{30}{30-1} \right) \left(\frac{30,99 - 3,70}{30,99} \right) = 1,034 \cdot 0,881 = 0,91$$

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.1

E) Procedimiento alfa (α) de Crombach

Este procedimiento se emplea con frecuencia cuando recurrimos como instrumento de recogida de datos al CUESTIONARIO. Dada su dificultad de cálculo se suele utilizar algún paquete estadístico (SPSS, ...) para su cálculo.

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

n = número de elementos o ítems de la prueba

s_i^2 = Varianza de cada uno de los ítems

s_t^2 = Varianza de las puntuaciones totales de la prueba

8.2.1 Procedimientos para determinar la fiabilidad.

Fiabilidad como CONSISTENCIA INTERNA

Ejemplo 8.3

E) Procedimiento alfa (α) de Crombach

Escala: TODAS LAS VARIABLES

Resumen del procesamiento de los casos

		N	%
Casos	Válidos	40	100,0
	Excluidos ^a	0	,0
	Total	40	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
,958	2

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

8.2.2 Algunas cuestiones relacionadas con la fiabilidad.

c) *Algunas cuestiones relacionadas con la fiabilidad*

c.1. *Fiabilidad y longitud*

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

R_{xx} será la fiabilidad alcanzada y n es el número de veces que la prueba se alarga o se acorta. Este es, el cociente entre el número de elementos de va a tener la prueba y los que tenía la prueba original:

$$N = \frac{\text{Número de elementos finales}}{\text{Número de elemento iniciales}}$$

De la fórmula anterior despejamos n y nos queda:

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

c.2. *Intervalo para la puntuación verdadera*

El intervalo se expresa: $X_v = X_A \pm z_{\alpha/2} \cdot \sigma_{med}$

Donde el error típico de medida es: $\sigma_{med} = s_x \cdot \sqrt{1 - r_{xx}}$

c.3. *Límites para la puntuación verdadera de un sujeto*

El intervalo se expresa: $X_v = X_A \pm z_{\alpha/2} \cdot \sigma_{med}$

c.4. *Comparación de las puntuaciones de dos sujetos en la misma prueba*

$$R.C. = \frac{|X_B - X_C|}{\sigma_{dif.med}}$$

Donde el error típico de la diferencia de medida se calcula de la siguiente forma:

$$\sigma_{dif.med.} = s_x \cdot \sqrt{1 - r_{xx}} \cdot \sqrt{2}$$

NO

8.3 Estudio de la Validez.

Aunque la **fiabilidad** es un **requisito imprescindible-necesario** de toda medición, **no es una condición suficiente** para defender la **idoneidad** de los **datos** recogidos

Podemos afirmar que:

**un instrumento es válido cuando
mide lo que dice medir
y no otra cosa distinta.**

Ningún instrumento es absolutamente válido, sino que **tendremos que matizar el grado y la naturaleza de la validez**

8.3 Estudio de la Validez.

VALIDEZ DE CONTENIDO

- Los ítems o elementos de una prueba deben ser una muestra suficiente y representativa del universo de los ítems relacionados con el objeto de estudio. Validación a través de “tablas de especificaciones” y “juicios de expertos-

VALIDEZ PREDICTIVA

- Capacidad de la prueba para predecir futuros hechos o fenómenos relacionados. Se calcula haciendo la correlación entre las puntuaciones en la prueba a validar y las puntuaciones en otra prueba posterior denominada criterio.

VALIDEZ CONCURRENTES

- Es igual que la validez predictiva sólo que tanto la prueba a validar como el criterio se aplican al mismo tiempo

VALIDEZ DE CONSTRUCTO

- Trata de analizar la conexión que se manifiesta entre la teoría en la que se basa la prueba y los ítems que la componen

VALIDEZ APARENTE

- Validez didáctica o validez superficial. Aspectos externos cuya influencia puede ser relevante para alcanzar participación y respuestas válidas

8.3.1 La validez de contenido.

Una muestra ha de ser **SUFICIENTE** y **REPRESENTATIVA** en función del universo de los ítems relacionados con el objeto del estudio.

La **SUFICIENCIA** se relaciona con el número mínimo de elementos que debe incluir el elemento.

La **REPRESENTATIVIDAD** exige un conocimiento profundo sobre el tema. Seleccionar convenientemente los ítems.

Elaborar una tabla de especificaciones realizada por varios expertos, que será sometida a estudio y crítica, a este proceso se le denomina **VALIDACIÓN**.

8.3.2 La validez predictiva.

Ejemplo 8.7

“Se ha elaborado una prueba de comprensión lectora compuesta por 30 ítems y se ha aplicado a 20 sujetos de Primaria. Para determinar su validez se han recogido las calificaciones de ese mismo grupo en Lengua el pasado curso que nos van a servir como criterio de comparación, los resultados de ambas pruebas así como las operaciones previas de cálculo que se recogen en la tabla siguiente:”.

8.3.2 La validez predictiva.

Ejemplo 8.7

$$r_{xy} = \frac{n \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] \cdot [n \sum Y^2 - (\sum Y)^2]}}$$

$$r_{xy} = \frac{20(2244) - (355)(110)}{\sqrt{[20(7057) - (355)^2] \cdot [20(738) - (110)^2]}} = 0,919$$

Sujetos	X	Y	X ²	Y ²	XY
1	20	8	400	64	160
2	12	3	144	9	36
3	17	5	289	25	85
4	25	10	625	100	250
5	8	2	64	4	16
6	15	5	225	25	75
7	21	7	441	49	147
8	13	4	169	16	52
9	16	6	256	36	96
10	30	9	900	81	270
11	9	2	81	4	18
12	18	5	324	25	90
13	14	3	196	9	42
14	28	10	784	100	280
15	22	6	484	36	132
16	11	2	121	4	22
17	15	6	225	36	90
18	22	5	484	25	110
19	26	9	676	81	234
20	13	3	169	9	39
sumas	355	110	7057	738	2244

8.3.3 La validez concurrente.

Este tipo de validez es una modalidad de la validez predictiva, pues se calcula mediante una correlación entre las puntuaciones de los sujetos en la prueba de validez y el criterio externo.

La diferencia estriba en que ambas mediciones se llevan a cabo en el mismo tiempo, además los resultados permiten realizar pronósticos a corto plazo, es decir, de utilización inmediata. Mientras que en la validez predictiva no coinciden la aplicación de la prueba y la recogida de datos del criterio.

8.3.4 La validez de constructo.

En esencia trata de analizar la conexión que se manifiesta entre la teoría en la que se basa la prueba y los ítems que la componen.

Fox (1987) nos dice que la validez de constructo es la capacidad que tiene el instrumento para distinguir entre aquellos grupos cuyo comportamiento difiere en relación a la variable estudiada. Esta forma de comprobación del ajuste entre el instrumento de medida y la realidad es posible cuando se recoja información de la conducta, habrá que encontrar aquellos indicadores de la variable que es objeto de estudio.

8.3.5 La validez aparente.

También recibe la denominación de validez didáctica y se refiere a la serie de aspectos externos de la prueba.

Quizá no ofrezca información relevante sobre el ajuste del instrumento al objeto de medida, pero contribuye a mejorar aquellos aspectos externos, cuya influencia puede ser relevante para alcanzar una mayor participación y que las respuestas sean válidas.

8.3.6 Algunos coeficientes relacionados con la validez.

Coeficiente de determinación

$$d = r_{xy}^2$$

Coeficiente de alienación

$$k = \sqrt{1 - r_{xy}^2}$$

Coeficiente de valor predictivo

$$E = 1 - k = 1 - \sqrt{1 - r_{xy}^2}$$

NO

8.3.7 Validez y longitud de una prueba.

Validez

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

R_{xy} = nueva Validez

r_{xy} = validez original

r_{xx} = valor inicial del coef. de fiabilidad

n = cociente entre num elem. finales e iniciales

Longitud de una prueba

$$n = \frac{1-r_{xx}}{\frac{r_{xy}^2}{R_{xy}^2} - r_{xx}}$$

NO

8.3.8 Predicción de puntuaciones.

NO

Se trata de **predecir** las puntuaciones en el criterio, conociendo las puntuaciones alcanzadas en la prueba, una vez determinado el coeficiente de validez.

Esta predicción es **más segura y precisa a medida que aumenta el coeficiente de correlación (validez predictiva) entre las variables**

Cuando estimamos las puntuaciones en el criterio (Y) a partir de las puntuaciones en la prueba (X), no tenemos la seguridad total de que la puntuación predicha sea única y siempre la misma. Es decir estamos haciendo una estimación (Y') que conlleva un error:

$$\text{Error de estimación} = Y' - Y$$

Así, cada predicción lleva asociado un error de estimación.

La desviación típica de los errores de estimación es lo que recibe el nombre de

$$\text{error típico de estimación } (\sigma_{est})$$

8.3.8 Predicción de puntuaciones.

NO

A) En puntuaciones directas :

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

B) En puntuaciones diferenciales :

$$Y' = r_{xy} \frac{S_y}{S_x} x$$

C) En puntuaciones típicas :

$$z'_y = r_{xy} z_x$$

8.4 Estudio de los elementos o ítems de una prueba.

Para tener las suficientes garantías científicas de que la prueba que vamos a aplicar a un grupo de sujetos es la más adecuada, es preciso **conocer el comportamiento de cada uno de los ítems que la componen.**

Así se podrán:

**eliminar los ítems que no pasen el control
o añadir otros que sí cumplan esos requisitos.**

En el campo socioeducativo se recurre a 3 índices:

- ***ÍNDICE DE DIFICULTAD (ID)***
- ***ÍNDICE DE HOMOGENEIDAD (IH)***
- ***ÍNDICE DE VALIDEZ (IV)***

Estos índices **no son aplicables a las pruebas de medida de actitudes**

8.4.1 El índice de dificultad (ID).

El índice de dificultad de un elemento depende del grupo de sujetos que lo conteste. Se expresa numéricamente por el número de ellos **que lo resuelve satisfactoriamente**, de tal forma que cuando es contestado correctamente por un número pequeño de sujetos el ítem es muy difícil; y a la inversa, si casi todos los sujetos responden bien al ítem, éste será muy fácil.

“El índice de dificultad (ID) de un elemento o ítem nos indica la proporción de los sujetos que lo resuelven de forma correcta en relación con el total de los que lo contestan”

Una buena prueba debe presentar aproximadamente la siguiente **distribución de ítems**:

- Muy fáciles ($ID > 0,75$): 10 %
- Fáciles ($0,55 < ID < 0,75$): 20%
- Normales ($0,45 < ID < 0,55$): 40%
- Difíciles ($0,25 < ID < 0,45$): 20%
- Muy Difíciles ($ID < 0,25$): 10%

Lo más pedagógico es que los ítems fáciles se sitúen al principio de la prueba, y los difíciles al final.

8.4.1 El índice de dificultad (ID).

El índice de dificultad de un elemento depende del grupo de sujetos que lo conteste. Se expresa numéricamente por el número de ellos **que lo resuelve satisfactoriamente**, de tal forma que cuando es contestado correctamente por un número pequeño de sujetos el ítem es muy difícil; y a la inversa, si casi todos los sujetos responden bien al ítem, éste será muy fácil.

“El índice de dificultad (ID) de un elemento o ítem nos indica la proporción de los sujetos que lo resuelven de forma correcta en relación con el total de los que lo contestan”

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

$$\text{Puntuación de un sujeto} = A - \frac{E}{n_a - 1}$$

$$ID = \frac{A - \frac{E}{n_a - 1}}{n}$$

8.4.1 El índice de dificultad (ID).

Ejemplo 8.11

Calcule e interprete el índice de dificultad de los siguientes ítems:
Elemento de evocación dicotómico contestado de forma acertada por 27 sujetos y de forma errónea por 15 y otro contestado por 43 sujetos con 16 que han acertado.

$$ID = \frac{A}{n} = \frac{27}{42} = 0,64$$

$$ID = \frac{A}{n} = \frac{16}{43} = 0,37$$

Este valor nos indica un ítem fácil por estar entre 0,55 y 0,74

En este caso el ítem se considera difícil por estar entre 0,25 y 0,44

En los casos de varias alternativas de respuesta, calcular el índice de dificultad de un ítem con tres alternativas de respuesta al que han contestado 75 sujetos y en el que el número de aciertos es de 35.

$$ID = \frac{A - \frac{E}{n_a - 1}}{n} = \frac{35 - \frac{40}{3-1}}{75} = 0,20$$

Dificultad muy difícil $\rightarrow 0,20 < 0,25$

Calcular el índice de dificultad de un elemento con cuatro alternativas de respuesta, que ha sido contestado de forma acertada por 51 sujetos y de forma errónea por 29

$$ID = \frac{A - \frac{E}{n_a - 1}}{n} = \frac{51 - \frac{29}{4-1}}{80} = 0,517$$

Dificultad media $\rightarrow 0,517$

8.4.2 El índice de homogeneidad (IH).

La homogeneidad de los elementos pone de manifiesto **la coherencia de cada uno de ellos con el total de la prueba**: si ésta mide un rasgo o características es lógico que cada uno de los ítems mida hasta cierto punto lo mismo.

Se calcula a **partir del valor de la correlación entre cada elemento y el conjunto de los demás**, así pues, cuanto mayor es el coeficiente entre ambas puntuaciones, mayor será la homogeneidad.

$$IH = \frac{r_{AB} s_A - s_B}{\sqrt{x_A^2 + x_B^2 - 2r_{AB} \cdot s_A \cdot s_B}}$$
$$s_B = \sqrt{p \cdot q}$$

8.4.2 El índice de homogeneidad (IH).

Ejemplo 8.12

En el caso de puntuaciones continuas en la prueba y respuestas dicotómicas en el ítem, se han obtenido los siguientes valores:

$$r_{bp} = 0,52 = r_{AB} \quad s_A = 3,25 \quad s_B = 0,495$$

$$IH = \frac{r_{AB}s_A - s_B}{\sqrt{x_A^2 + x_B^2 - 2r_{AB} \cdot s_A \cdot s_B}}$$

$$IH = \frac{0,52 \cdot 3,25 - 0,495}{\sqrt{(3,25)^2 + (0,495)^2 - 2 \cdot 0,52 \cdot 3,25 \cdot 0,495}} = 0,395$$

Se puede concluir que este elemento es homogéneo con el resto de la prueba y puede ser seleccionado dado que el valor es superior a 0,20

8.4.3 El índice de validez (IV).

Los elementos de una prueba se consideran válidos cuando realmente miden lo que dicen medir y no otra cosa distinta. Ello exige la existencia de una clara **relación entre las puntuaciones obtenidas por los sujetos de la muestra en cada uno de los ítems y las alcanzadas en el criterio de validez: cuanto mayor sea esa relación mayor será el índice de validez (IV).**

Pueden usarse los coeficientes:

Biserial-puntual

Biserial

Tetracórico

Phi

Se consideran **satisfactorios** aquellos ítems cuyos **valores IV sean iguales o superiores a 0,20**

8.4.3 El índice de validez (IV).

Ejemplo 8.13

Queremos calcular el índice de validez de un ítem dicotómico cuyos datos globales del criterio referidos al ítem que se analiza son los siguientes: la media de los aciertos es 6,32; la media del total es 5,79; la desviación típica es 1,86; $p=0,61$ y $q=0,39$. Por lo tanto

$$r_{bp} = \frac{|\bar{X}_p - \bar{X}_t|}{s_t} \sqrt{\frac{p}{q}}$$
$$r_{bp} = \frac{|6,32 - 5,79|}{1,86} \sqrt{\frac{0,61}{0,39}} = 0,36$$

Este elemento es válido, dado que el valor 0,36 es mayor que 0,20 por lo tanto puede ser seleccionado para configurar la prueba.

(Los elementos sombreados corresponden a contenidos no obligatorios)

FIABILIDAD Y VALIDEZ

a) *Fiabilidad como estabilidad y equivalencia*

$$r_{xy} = \frac{n \cdot \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Puntuaciones directas

Correlación de Pearson entre las dos aplicaciones sucesivas (estabilidad) o bien entre la aplicación de la prueba y su equivalente.

b) *Fiabilidad como consistencia interna*

b.1. *Procedimiento de Spearman-Brown*

$$R_{xx} = \frac{2 \cdot r_{xx}}{1 + r_{xx}}$$

r_{xx} se calcula mediante el coeficiente de correlación de Pearson entre las mitades, así llamamos X_1 a las puntuaciones de los ítems impares (1ª mitad) y X_2 a la suma de los ítems pares (2ª mitad).

Siendo:
$$r_{xx} = \frac{n \sum X_1 X_2 - \sum X_1 \sum X_2}{\sqrt{[n \sum X_1^2 - (\sum X_1)^2][n \sum X_2^2 - (\sum X_2)^2]}}$$

b.2. Procedimiento de Rulon

$$r_{xx} = 1 - \frac{s_d^2}{s_t^2}$$

Para llegar a determinar el valor del coeficiente de fiabilidad debemos calcular previamente tanto la varianza de las diferencias como la total.

b.3. Procedimiento de Guttman

$$r_{xx} = 2 \left(1 - \frac{s_{1a}^2 + s_{2a}^2}{s_t^2} \right)$$

Debemos calcular la varianza total y los valores de las varianzas de las mitades (impares/pares).

$$s_{1a}^2 = \frac{\sum X_1^2 - \frac{(\sum X)^2}{n}}{n-1}$$

b.4. Procedimiento de Kuder-Richardson

$$r_{xx} = \left(\frac{n_e}{n_e - 1} \right) \left(\frac{s_t^2 - \sum p \cdot q}{s_t^2} \right)$$

Para ítems dicotómicos, dónde n_e se refiere al número de elementos de que consta la prueba. Además p es la proporción de sujetos que aciertan y $q = 1 - p$; esta operación se debe realizar con cada uno de los ítems, pues el valor que necesitamos es la suma de $p \cdot q$ de todos los elementos.

b.5. Procedimiento alfa de Cronbach

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

Donde: n : número de elementos o ítems de la prueba; s_i^2 : varianza de cada uno de los ítems y s_t^2 : varianza de las puntuaciones totales de la prueba

c) Algunas cuestiones relacionadas con la fiabilidad

c.1. Fiabilidad y longitud

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

R_{xx} será la fiabilidad alcanzada y n es el número de veces que la prueba se alarga o se acorta, esto es, el cociente entre el número de elementos de va a tener la prueba y los que tenía la prueba original:

$$N = \frac{\text{Número de elementos finales}}{\text{Número de elemento iniciales}}$$

De la fórmula anterior despejamos n y nos queda:

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

c.2. Intervalo para la puntuación verdadera

El intervalo se expresa: $X_v = X_A \pm z_{\alpha/2} \cdot \sigma_{med}$

Donde el error típico de medida es: $\sigma_{med} = s_x \cdot \sqrt{1 - r_{xx}}$

c.3. Límites para la puntuación verdadera de un sujeto

El intervalo se expresa: $X_v = X_A \pm z_{\alpha/2} \cdot \sigma_{med}$

c.4. Comparación de las puntuaciones de dos sujetos en la misma prueba

$$R.C. = \frac{|X_B - X_C|}{\sigma_{df.med}}$$

Donde el error típico de la diferencia de medida se calcula de la siguiente forma:

$$\sigma_{df.med.} = s_x \cdot \sqrt{1 - r_{xx}} \cdot \sqrt{2}$$

d) Validez predictiva de las pruebas:

$$r_{xy} = \frac{n \cdot \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \cdot \sum X^2 - (\sum X)^2][n \cdot \sum Y^2 - (\sum Y)^2]}}$$

Correlación de Pearson entre las puntuaciones en la prueba (X) y en el criterio (Y).

d.1. Algunos coeficientes relacionados con la validez

d.1.1. Coeficiente de determinación

Se representa por "d" y su valor consiste en elevar al cuadrado el coeficiente de validez:

$$d = r_{xy}^2 = 0,88^2$$

d.1.2. Coeficiente de alienación

Se representa por "k" y se obtiene mediante: $k = \sqrt{1 - r_{xy}^2}$

d.1.3. Coeficiente de valor predictivo

Se representa mediante "E" y se obtiene: $E = 1 - k = 1 - \sqrt{1 - r_{xy}^2}$

d.2. Validez y longitud de una prueba

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

Donde R_{xy} se trata de la nueva validez, r_{xy} es la validez original, r_{xx} es el valor inicial del coeficiente de fiabilidad y n es el cociente entre el número de elementos finales y los elementos iniciales.

De la ecuación anterior despejamos n: $n = \frac{1 - r_{xx}}{\frac{r_{xy}^2}{R_{xy}^2} - r_{xx}}$

d.3. Predicción de puntuaciones

a) En puntuaciones directas: $Y' = r_{xy} \frac{s_y}{s_x} (X_i - \bar{X}) + \bar{Y}$

b) En puntuaciones diferenciales: $y' = r_{xy} \frac{s_y}{s_x} x$

c) En puntuaciones típicas: $z'_y = r_{xy} z_x$

Error típico de estimación:

$$\sigma_{est} = s_y \sqrt{1 - (r_{xy})^2}$$

Intervalos de confianza:

a) Puntuaciones directas: $Y' \pm z_{\alpha/2} \cdot \sigma_{est}$

ESTUDIO DE LOS ELEMENTOS O ÍTEMS DE UNA PRUEBA

a. Índice de dificultad (I.D.)

$$I.D. = \frac{A}{n}$$

Donde A nos indica el número de sujetos que aciertan el ítem y n el número de sujetos que lo intentan.

Para elementos de varias alternativas de respuesta la fórmula que debemos aplicar es la siguiente:

$$I.D. = \frac{A - \frac{E}{n_a - 1}}{n}$$

Donde A es el número de aciertos, E el número de errores y n_a el número de alternativas de respuesta del ítem.

b. Índice de homogeneidad (I.H.)

$$I.H. = \frac{r_{AB} \cdot s_A - s_B}{\sqrt{s_A^2 + s_B^2 - 2 \cdot r_{AB} \cdot s_A \cdot s_B}}$$

Donde r_{AB} es la correlación entre el ítem y el total, s_A la desviación típica de las puntuaciones en la prueba, s_B la desviación típica en el ítem que se calcula $s_B = \sqrt{p \cdot q}$.

c. Índice de validez (I.V.)

Para seleccionar el coeficiente de correlación más adecuado habrá que tener en cuenta las condiciones de los datos, tanto las referidas a los elementos como al criterio. Pueden usarse los coeficientes biserial-puntual (dicotómicos), el biserial, el tetracórico o el Phi.

$$r_{bp} = \frac{|\bar{X}_p - \bar{X}_t|}{s_t} \sqrt{\frac{p}{q}}$$

Resumen

Fiabilidad

Fiabilidad como estabilidad

Fiabilidad como equivalencia

Fiabilidad como consistencia interna

Fiabilidad y longitud de las pruebas

Validez

Validez de contenido

Validez predictiva

Validez concurrente

Validez de constructo

Validez aparente

Coefficiente de determinación

Coefficiente de alienación

Coefficiente de valor predictivo

Validez y longitud de una prueba

Predicción de puntuaciones

Índice de Dificultad (ID)

Índice de Homogeneidad (IH)

Índice de Validez (IV)

Fe de erratas

TEMA 8:

En el ejemplo 8.7 de la página 164 hay varios errores en la Tabla 8.3., que se arrastran al calcular el coeficiente de correlación r_{xy} de validez predictiva.

La tabla y el coeficiente correctos son:

	A	B	C	D	E	F
1		X	Y	X ²	Y ²	X*Y
2		20	8	400	64	160
3		12	3	144	9	36
4		17	5	289	25	85
5		25	10	625	100	250
6		8	2	64	4	16
7		15	5	225	25	75
8		21	7	441	49	147
9		13	4	169	16	52
10		16	6	256	36	96
11		30	9	900	81	270
12		9	2	81	4	18
13		18	5	324	25	90
14		14	3	196	9	42
15		28	10	784	100	280
16		22	6	484	36	132
17		11	2	121	4	22
18		15	6	225	36	90
19		22	5	484	25	110
20		26	9	676	81	234
21		13	3	169	9	39
22	Suma	355	110	7057	738	2244
23						
24	Correlación r_{xy}	0,91944057				

PREGUNTAS

Exámenes
anteriores



1

El procedimiento conocido como de repetición, del test-retest, idéntica la fiabilidad como...

Seleccione una:

- a. Equivalencia
- b. Variabilidad
- c. Estabilidad

2

Por su propia naturaleza, el cálculo de la validez predictiva implica calcular la Correlación entre las puntuaciones de la prueba y:

Seleccione una:

- a. Un Criterio externo
- b. Una segunda aplicación de la misma
- c. Un Criterio medido tiempo después de la aplicación de la prueba

El procedimiento conocido como de repetición, del test-retest, idéntica la fiabilidad como...

1

Seleccione una:

- a. Equivalencia
- b. Variabilidad
- c. Estabilidad ✓

La respuesta correcta es: Estabilidad

2

Por su propia naturaleza, el cálculo de la validez predictiva implica calcular la Correlación entre las puntuaciones de la prueba y:

Seleccione una:

- a. Un Criterio externo
- b. Una segunda aplicación de la misma
- c. Un Criterio medido tiempo después de la aplicación de la prueba ✓

La respuesta correcta es: Un Criterio medido tiempo después de la aplicación de la prueba

¿Cuál de estos términos define mejor la fiabilidad?

3

Seleccione una:

- a. La variabilidad
- b. La precisión
- c. La constante

4

El procedimiento de Rulon se basa en la

Seleccione una:

- a. Correlación entre las mitades
- b. Intercorrelaciones de los ítems
- c. Varianza de las diferencias

3

¿Cuál de estos términos define mejor la fiabilidad?

Seleccione una:

- a. La variabilidad
- b. La precisión ✓
- c. La constante

La respuesta correcta es: La precisión

4

El procedimiento de Rulon se basa en la

Seleccione una:

- a. Correlación entre las mitades
- b. Intercorrelaciones de los ítems
- c. Varianza de las diferencias ✓

La respuesta correcta es: Varianza de las diferencias