TEMA 11. LA FIABILIDAD DE LOS TESTS

11.1. Métodos empíricos para estimar el coeficiente de Fiabilidad

11.2. Factores que afectan al Coeficiente de Fiabilidad

11.3. Interpretación del Coeficiente de Fiabilidad

11. La Fiabilidad de los tests

La fiabilidad es un tópico constante en todos los instrumentos de medida. Su estudio trata de establecer la precisión con la que mide cualquier instrumento de medida en general y los tests en particular. Es un concepto muy asociado al error de medida. Cuanto más fiable es un test, con mayor precisión mide y, por lo tanto, menos error de medida se comete.

Se asume que toda medida empírica (X) tiene dos componentes: La medida verdadera (V) y el error de medida (E).

En la práctica es imposible conocer V y E pero, bajo determinados supuestos, podemos estimar la Varianza de V y la Varianza de E.

La varianza de E puede derivar en lo que llamamos el “error típico” de E (desviación típica de los errores de medida). Este error típico oscila entre 0 y la desviación típica de X (Sx). Cuanto más tiende a 0 más fiable es el test; cuanto más tiende a Sx menos fiable es. Basándonos en ese error típico podemos llevar a cabo una estimación probabilística de la Puntuación Verdadera (V) de un sujeto a partir del conocimiento de su Puntuación empírica (X)

El error típico de medida (bajo los supuestos de la Teoría Clásica de los Tests) se puede establecer formalmente como:

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

El error típico constituye la llamada “fiabilidad absoluta” del test, frente a rxx que constituye la “fiabilidad relativa” y que es un coeficiente de correlación que se puede estimar (bajo determinados supuestos) empíricamente (es el coeficiente de fiabilidad)

11. La Fiabilidad de los tests (cont.)

11.1. Métodos empíricos para estimar el coeficiente de Fiabilidad (r_{xx})

- **Método de las formas paralelas**. Identifica la Fiabilidad como Equivalencia de las medidas. Se construye el test y una forma paralela. Se aplican a una muestra de sujetos y se correlacionan ambas medidas. Se supone que si las formas son paralelas, ambas deberían medir lo mismo y con igual precisión. En la medida en que ello sea así, nos podremos "fiar" de la medida obtenida con cualquiera de esas formas.
- **Método del Test-Retest**. Identifica la fiabilidad como Estabilidad de la medida. Se aplica el test a una muestra de sujetos en dos momentos temporales distintos y se correlacionan ambas medidas. Se supone que, si el test es preciso, las medidas deberán ser muy parecidas y el coeficiente de fiabilidad tenderá a 1. Si ello es así se asume que el test es fiable porque independientemente de cuándo se aplique se tiende a obtener con él siempre la misma medida.

(En el escalamiento de estímulos visto en el tema 9 sólo se pueden aplicar estos dos primeros métodos para el cálculo de la fiabilidad).

- **Método "Alfa" de Cronbach**. Identifica la fiabilidad como Consistencia Interna. Se denomina así porque analiza hasta qué punto medidas parciales obtenidas con los diferentes ítems son "consistentes" entre sí y por tanto representativas del universo posible de ítems que podrían medir ese constructo.
- **Método del seccionamiento en dos mitades**. Es un método de consistencia pero también podría considerarse como un método de formas paralelas. Se divide el test en dos mitades y se analiza la consistencia que existe entre las mismas.

11. La Fiabilidad de los tests (cont.)

11.2. Factores que afectan al Coeficiente de Fiabilidad:

11.2.1. Longitud del Test. Se asume que, en ausencia de otros condicionantes, cuantos más ítems tenga un test más fiable será. Para analizar la conveniencia de añadir ítems a un test con el fin de mejorar su fiabilidad hasta cotas aceptables se emplea, generalmente, la ecuación de Spearman-Brown que pone en relación, bajo determinadas condiciones, el coeficiente de fiabilidad con el número de items del test:

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

Donde
R_{xx} es el nuevo coeficiente de fiabilidad
r_{xx} es el coeficiente de fiabilidad inicial.
n es el número de veces que se ve incrementada la longitud inicial (LI) de un test con respecto a su longitud final (LF)

$$\boxed{n = \frac{LF}{LI}}$$

Normalmente, cuando se realiza este tipo de análisis, lo que interesa es conocer el número de ítems que hay que añadir (o quitar) para operar con un determinado coeficiente de fiabilidad deseado, con lo que la incógnita suele ser "n".

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

11. La Fiabilidad de los tests (cont.)

11.2. Factores que afectan al Coeficiente de Fiabilidad (cont.)

11.2.1. Longitud del Test (cont.)

Un caso particular en la aplicación de esta corrección es el cálculo del coeficiente de fiabilidad por el método del seccionamiento en dos mitades.

En este caso, el valor de la correlación entre las puntuaciones de las dos mitades no se corresponde con el coeficiente de fiabilidad del test total ya que estamos correlacionando la mitad del test con la otra mitad. Para conocer el coeficiente de fiabilidad del test total es necesario aplicar la ecuación de Spearman-Brown para n = 2.

$$\boxed{r_{xx} = \frac{2r_{pi}}{1 + r_{pi}}}$$

Donde

r_{xx} es el coeficiente de fiabilidad del test total

r_{pi} es la correlación entre las dos mitades del test.

Convencionalmente se denota como r_{pi} porque el método tradicional de seccionamiento en dos mitades es el de evaluar los ítems pares por un lado y los impares por otro; si bien conviene resaltar en este punto que es admisible cualquier procedimiento de seccionamiento en dos mitades. De lo que se trata es de establecer dos mitades lo más “paralelas” posibles.

11. La Fiabilidad de los tests (cont.)

11.2. Factores que afectan al Coeficiente de Fiabilidad (cont.)

11.2.2. Restricción del rango (variabilidad de la muestra).

Otro factor que afecta a la cuantía del coeficiente de fiabilidad es la desviación típica (variabilidad) de la muestra a la que se aplica el test. Como hemos visto, la cuantía del error típico de medida depende de la variabilidad de los datos (S_x) y del coeficiente de fiabilidad del test (r_{xx}). Es lógico que, para mantener el mismo valor de ese error típico, si la variabilidad de la muestra se ve afectada, se altere también el coeficiente de fiabilidad para mantener esa cuantía de error constante.

$$\boxed{r_{22} = 1 - \frac{S_1^2 (1 - r_{11})}{S_2^2}}$$

Donde:

r_{11} es el coeficiente de fiabilidad que corresponde a la muestra de variabilidad S_1

r_{22} es el coeficiente de fiabilidad que corresponde a la muestra de variabilidad S_2

Ejemplo

Un test experimental formado por 50 items tiene un coeficiente de fiabilidad de 0,90 obtenido con una muestra de población general con media 30 y D.Típica 12. Al resultar demasiado extenso decidimos reducir el nº de items hasta dejarlo con 20. ¿Sigue siendo aceptable su fiabilidad con esa reducción?

Por otra parte, otro investigador quiere utilizar ese mismo test con fines diagnósticos pero necesita que su coeficiente de fiabilidad sea, al menos, de 0,95. ¿Cuántos items tendría que añadir para conseguir sus propósitos?

Un tercer investigador utiliza el test original en una subpoblación de estudiantes con media 38 y D. Típica 10. ¿Cuál sería en este caso la cuantía del coeficiente de fiabilidad?

Solución:

Cuestión 1. Modificamos la longitud del test y queremos conocer la nueva fiabilidad.

Datos: $r_{xx} = 0,90$; LI = 50 ; LF = 20; n = 20/50 = 0,4

$$R_{xx} = \frac{0,4x0,90}{1+(0,4-1)x0,90} = 0,783$$

El nuevo coeficiente pasa a ser 0,783 que, para determinados fines, puede ser aceptable.

Ejemplo

Solución:

Cuestión 2. Averiguar nueva longitud para obtener un coeficiente de fiabilidad deseado.

Datos: $r_{xx} = 0{,}90$; LI = 50 ; $R_{xx} = 0{,}95$

$$n = \frac{0{,}95(1-0{,}90)}{0{,}90(1-0{,}95)} = 2{,}11$$

2,11 = LF/50 → LF = 50x2,11 = 105,5 → 106

Tendría que añadir 56 items más ya que 106 – 50 = 56.

Cuestión 3. Averiguar nuevo coeficiente de fiabilidad por distinta variabilidad de la muestra.

Datos: $r_{11} = 0{,}90$; $S_1 = 12$; $S_2 = 10$

$$r_{22} = 1 - \frac{12^2(1-0{,}90)}{10^2} = 0{,}856$$

El coeficiente de fiabilidad de 0,90 en la muestra de variabilidad 12 equivale a un coeficiente de 0,856 en la muestra de variabilidad 10

11. La Fiabilidad de los tests (cont.)

11.3. Interpretación del Coeficiente de Fiabilidad

11.3.1. Estimación de la puntuación verdadera cuando se asume la distribución normal de los errores de medida.

$$\boxed{V = X \pm S_e\, Z_{(\alpha)}}$$

Donde,

V es la puntuación verdadera estimada
X es la puntuación empírica de un sujeto obtenida tras la aplicación del test
$\mathbf{S_e}$ es el error típico de medida
$\mathbf{Z_{(\alpha)}}$ es el valor Z de la curva normal que corresponde a un nivel "alfa" preestablecido del margen de error en la estimación

11. La Fiabilidad de los tests (cont.)

11.3. Interpretación del Coeficiente de Fiabilidad (cont.)

11.3.2. Estimación de la puntuación verdadera cuando se ignora el tipo de distribución que generan los errores de medida.

$$\boxed{V = V' \pm S_{v.x}\, Z_{(\alpha)}}$$

Donde,

V es la puntuación verdadera estimada

V' es la puntuación verdadera estimada por regresión a partir de la puntuación empírica de un sujeto obtenida tras la aplicación del test

$$\boxed{V' = r_{xx}\left(X - \bar{X}\right) + \bar{X}}$$

$S_{v.x}$ es el error típico de de estimación de V a partir de X

$$\boxed{S_{v.x} = S_e \sqrt{r_{xx}}}$$

$Z_{(\alpha)}$ es el valor Z de la curva normal que corresponde a un nivel "alfa" preestablecido del margen de error en la estimación

<u>Ejemplo</u>

Un test experimental formado por 50 items tiene un coeficiente de fiabilidad de 0,90 obtenido con una muestra de población general con media 30 y D.Típica 12. ¿Qué Puntuación Verdadera, con un margen de error del 5% (α = 0,05) cabría pronosticarle a un sujeto que ha obtenido 45 puntos si suponemos que los errores de medida se distribuyen normalmente?
¿Cuál sería ese pronóstico si ignoramos el tipo de distribución que generan esos errores de medida?

Solución:

Cuestión 1.

Datos: $r_{xx} = 0,90$ $X = 45$ Media = 30 $S_x = 12$ $\alpha = 0,05$ $Z_{1-\alpha/2} = 1,96$

$$S_e = S_x\sqrt{1-r_{xx}} = 12\sqrt{1-0,90} = 3,79$$

V = 45 ± 3,79 x 1,96

Límite superior = 52,44
Límite inferior = 37,56

<u>Ejemplo</u> (cont.)

Cuestión 2.

Datos:

$r_{xx} = 0,90$ $X = 45$ Media $= 30$ $S_x = 12$ $S_e = 3,79$ $\alpha = 0,05$ $Z_{1-\alpha/2} = 1,96$

$$V' = r_{xx}\left(X - \overline{X}\right) + \overline{X} \rightarrow V' = 0,90(45 - 30) + 30 = 43,5$$

$$S_{v.x} = S_e\sqrt{r_{xx}} = 3,79\sqrt{0,90} = 3,60$$

V = 43,5 ± 3,60 x 1,96

Límite superior = 50,55
Límite inferior = 36,45