

TEMA 8

Este tema va a tratar las características técnicas de los instrumentos de medida, con los cálculos de fiabilidad de instrumentos y algunas cuestiones sobre el uso de estas

8.2 ESTUDIO DE FIABILIDAD

Un instrumento de medida (una encuesta, un test...) debe ser fiable. En algunos casos esta fiabilidad depende del contenido del instrumento. Necesitamos que mida con precisión e independientemente de lo que está midiendo eso es lo que lo convierte en fiable o no. A la hora de medir podemos cometer un error eso sí por eso a la hora de dar la puntuación de un sujeto la descompondremos en dos partes, la puntuación de verdad, o sea la dada por el test, y el posible error que tendría ese instrumento.

Los errores son de dos tipos sistemáticos y aleatorios.

Los sistemáticos tienen que ver con las características del instrumento y los tendrían todo el estudio (todo lo que se mida con el instrumento). El error aleatorio, son variables cuyos efectos nos resultan desconocidos, que surgen por azar y lo grande que sea este error viene condicionado por el margen de error o nivel de confianza del investigador. Si menor el error mejor el instrumento. Vamos que si el instrumento de medida es bueno los posibles errores que surgen aleatoriamente por azar) será muy pocos. Estos errores pueden ser por ejemplo que el entrevistador rellene mal las casillas, que se borre una... Si está bien hecho el instrumento será difícil que ocurra algún error de estos. La fiabilidad también se puede llamar constancia en las puntuaciones de los sujetos, o concordancia entre varias mediciones sucesivas de una misma realidad.

8.2.1 Procedimiento para determinar la fiabilidad

Hay tres procedimientos para establecer la fiabilidad. El de fiabilidad como estabilidad, el de fiabilidad como equivalencia y el último el de fiabilidad como consistencia interna.

8.2.1.1 La fiabilidad como estabilidad

También llamada procedimiento de repetición o del test retest. Donde un mismo grupo hacen el mismo test varias veces, en un tiempo ni muy corto ni muy largo, dependiendo de las características del grupo y del instrumento. Normalmente del grupo y del instrumento. Normalmente se hará entre 20 -25 días entre uno y otro. Una vez obtenidos los datos de los dos test de cada sujeto, deberemos establecer una relación entre datos, para ello utilizaremos un coeficiente de correlación (los vistos en el tema 7) y escogeremos entre los siete tipos el más adecuado según el tipo, naturaleza (dicotómica, continua, discreta) de la variable que posean. Tomando los resultados de la primera aplicación como variable x y de la segunda aplicación como la variable y .

Y si la correlación entre valores supera 0,95 se considerará que la fiabilidad es muy buena. Y recordemos lo dicho en el tema 7 si la correlación es inferior a 0,95 para buscar la fiabilidad de un instrumento de medida no lo consideraremos suficiente. Aunque algunos autores lo bajan hasta 0,85.

8.2.1.2 FIABILIDAD COMO EQUIVALENCIA

También llamado de formas paralelas. Se hace aplicando dos pruebas diferentes que miden lo mismo (estudian lo mismo). Y buscaremos la correlación entre ambos. Es un procedimiento complicado porque no es fácil encontrar dos pruebas que sean exactamente lo mismo porque generalmente se consideran pruebas paralelas no solo si miden lo mismo (buscan lo mismo) sino que han de tener los mismos objetivos, los mismos contenidos y la misma forma. Vamos que no puedo mezclar dos pruebas de estadística una que estudie todos los temas como las del año pasado con otras que mida solo 11 temas como ocurre este año, también puedo comparar dos pruebas una que busque saber tus conocimientos de estadística con otra que mida conocimientos de matemáticas. O comparar dos test uno que tenga 11 preguntas y otro 12, o uno que sea a elegir entre sí o no y otra a elegir entre tres opciones. Si la aplicación de una prueba y otra supera los 20 días, el cálculo de fiabilidad será como equivalencia y estabilidad, y el índice que obtenemos se conoce como índice de estabilidad o equivalencia. Pero si transcurre poco tiempo es solo de equivalencia. Normalmente se recurre al coeficiente Pearson.

8.2.1.3 La fiabilidad como consistencia interna

Las pruebas están formadas por un grupo de ítems elementos, cada uno de los cuales mide un rasgo diferente, o característica diferente. A de haber una lógica, coherencia entre entre estos ítems (preguntas). Vamos este sistema de fiabilidad lo que busca es dividir en dos partes la prueba (instrumento) para tener de cada individuo dos resultados, con los cuales poder crear un coeficiente de correlación e intermanente saber si hay consistencia, o sea si hay correlación entre una y otra parte. Por medio de una prueba que ya veremos. Por ello podemos dividir las preguntas o ítems en dos partes o mitades, o bien eligiendo como criterios los ítems, que nos dará lugar a pares o impares o bien agrupándolos por primera y segunda mitad. Así para cada sujeto tendremos dos puntuaciones.

Lo más correcto sería el sistema de las dos mitades, primera parte y segunda parte. Porque en muchos test para que comiencen contentos y se embalancen ponen las preguntas más fáciles al principio por eso pueden encontrar más diferencia en las preguntas de principio y las del final (ojo, fíjate en esto, en los exámenes de estadística las preguntas más complejas están al final por eso no os enrolléis mucho con las primeras preguntas y reservar más tiempo para las últimas) Ahora veremos las pruebas que se aplicarán a estas mitades

A) La prueba o procedimiento de Spearman - Brown . Es la correlacion entre dos mitades para lo cual se aplica un coeficiente de correlacion de los aprendidos en el tema 7 , normalmente el Pearson , pero a cuyo resultado se le ha de aplicar otra prueba para corregirlo que seria esta formula.

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

r_{xx} es el coeficiente de correlacion . Los datos que surgen de este calculo se interpretaran igual que los del coeficiente de correlacion si es mas de 0,90 es que si es fiable si es de 0,95 o mas es que es una fiabilidad alta. Con esta prueba se buscara tener el valor que la correlacion hubiese tenido si las dos partes hubiesen tenido el mismo numero de items que el test completo

B) Procedimiento de Rulon

Este procedimiento se fija en las varianzas recordar σ^2 o s^2 (lo que se desvia de la media , la mayoria de los datos) Y se basa en hacer la varianza de las diferencias la resta de una parte y otra $d=x_1-x_2$ y la varianza total del test y compararlas por medio de esta formula.

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

Donde s_d^2 es la varianza de las diferencias entre puntuaciones S_t varianza total.

A la hora de analizar los resultados no sera igual que los coeficientes de correlacion, aunque los resultados tambien estaran entre 0 y 1 . En este caso si el resultado de r_{xx} es 0 significara la correlacion es perfecta, porque pensar, si yo divido dos varianzas iguales 5 de la varianza de la diferencia y otro 5 de la varianza total , $5:5=1$. Entonces si a 1 (la constante de la formula) le resto el 1 (de las divisiones de las varianzas) me dara 0 , eso nos dira que la relacion es muy fuerte. Pero cuanto mas se acerque a 0 este r_{xx} menos relacion entre ambas mitades habra porque mayor habra sido las diferencias entre las dos varianzas.

Vamos lo importante de este metodo es crear una columnita nueva que se volveria llamar d, y en el cual restaria la variable x a la variable sesgada y con esta columna

hallar una varianza (como si fuese una marca de clase) de esta resta. Ese sera lo que en la formula se llama s^2_d y a la que tendremos que dividirle la varianza total del test entero (sin cortar) y cuyo resultado restarle a 1.

C) Procedimiento de Guttman

Es parecido al anterior porque se basa en la varianza, pero ahora haremos la de la primera mitad y la de la segunda mitad. Con esta formula

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

. Esta formula es una re expresion vamos que es lo mismo de la de Rulon , por eso ambos daran siempre el mismo resultado (si los hemos hecho bien) y tambien tendria que tener el mismo resultado que la de Spearman Brown pero solo si la varianza de ambas partes de la parte 1 y 2 es = , sino fuese iguales entonces la formula de Rulon y de Guttman sera mas pequena que la de Spearman Brown

d) Procedimiento de Kuder Richardson

Este procedimiento es totalmente diferente por eso se divide la prueba en tantas partes como item hay . Asi al encontrar los coeficientes seran de consistencia interna , y de homogeneidad . Hay varios procedimientos de calculo (varias formulas segun el tipo de variable, segun la dificultad de las preguntas , aqui solo se ve el de dicotomicas cuya formula es

$$r_{xx} = \left(\frac{n}{n-1} \right) \cdot \left(\frac{s_t^2 - \sum pq}{s_t^2} \right)$$

Donde n es el numero de elementos que consta la prueba (11 preguntas, 12 preguntas)

P = proporcion de sujetos que aciertan

Q = proporcion de sujetos que fayan

Es muy larga de hacer porque hay que encontrar los porcentajes de cada una de las preguntas y para bien y para mal , multiplicarlos y sumarla de cada pregunta.

e) Procedimiento à de Grombach

Es muy usado para cuestionarlas y como es muy difícil de hacer la fórmula se utiliza los programas como el SPSS Y en principio es parecido al anterior de Kuder Richardson pero cuando no son dicotómicas las variables

La fórmula es

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

s_{xi} = es la varianza de cada ítem

N = es el número de elementos de la prueba (no creo que entre porque no se puede hacer a mano , solo caera en la parte del examen tipo test , si es que cae)

8.3 ESTUDIO DE VALIDEZ

Me debo poder fiar de los datos que tengo, pero no solo eso, sino que además deben proporcionarme la información sobre la que quiero saber, no me sirve de nada un estudio fiabilísimo que me da diferencia proporciona los gustos alimenticios de los alumnos si yo quiero saber su absentismo escolar. O sea un instrumento es válido si mide lo que dice medir y no otra cosa. Aunque ya os digo que por supuesto no va a ser tan sencillo porque nada mide realmente lo que quiere medir siempre hay una ligera diferencia . Y ahora veremos como comprobar esa validez

8.3.1 LA VALIDEZ DE CONTENIDOS

Para saber si un contenido es válido debemos saber si es suficiente , y si es representativo. Un instrumento está formada por muchas partes o ítems y para comprobar su validez no hace falta estudiarlos todos solo necesitaremos tomar una muestra de estos ítems aunque claro esta la muestra ha de ser suficiente y representativa (por lo cual debemos conocer bien el tema para seleccionar bien los ítems) sino sabemos nada de estadística difícilmente sabremos escoger las preguntas más representativas de un examen de estadística) Hay varios procedimientos para hallar la validez , como tablas de especificaciones pero nada propio (ninguna fórmula)

8.3.2 LA VALIDEZ PREDICTIVA

Nos dice si una prueba nos permite predecir o no un valor relacionado.

Se halla buscando la correlacion entre las puntuaciones alcanzadas en una prueba y los obtenidos en otra prueba que llaman criterios(porque ya estan validados) o sea que quiero saber si un examen que he creado esta bien tomo el examen officia, hago hacer los dos examenes a unos alumnos y si hay una correlacion entre las notas sera bueno validado

Par determinar la correlacion se usa el coeficiente de correlacion casi siempre es el Pearson

8.3.3 VALIDEZ CONCURRENTENTE

Es igual que la predictiva pero las dos pruebas se hacen a la vez y se hace igual que la predictiva y por eso solo permite pronosticar a corto plazo

8.3.4 VALIDEZ DE CONSTRUCTO

Tambien llamada de elaboracion de construcciòn, es la mas dificil y busca la relacion entre la teoria de la prueba y las preguntas de esta. Para esta validez debemos tener una hipotesis inicial que la prueba deba corroborar . Si la hipotesis no es confirmada es o porque las pruebas carecen de validez o es que las pruebas son insuficientes. Tambien puede ser que la hipotesis este mal planteada o que no se pueda validar con la investigacion . Fox en 1987 dice que la validez de contrecto es la capacidad que tiene el instrumento para distinguir entre quello grupos cuyo comportamiento es diferente a la variable estudiada.

8.3.5 LA VALIDEZ APARENTE

Es la validez didactica que mide las disquisiciones profundas (las preguntas profundas, sobre lo que mide , si es solo superficial . Tiene poco interes scientifico . Es util porque nos puede indicar los problemas que puede darse a la hora de rellenar la encuesta su aspecto externo.

Ej: - un instrumento largo provoca rechazo fatiga y no reflejar opiniones y valoraciones reales.

-Si el formato es atractivo puede motivar a la persona

-El lenguaje a de ser adecuado a la edad y la cultura de la persona

-Las instrucciones han de ser claras

8.3.6 Algunos coeficientes relacionados con la validez

Hay muchos tipos diferentes de coeficientes y cada uno con una funcion diferente y todos utilizan el coeficiente de validez

8.3.6.2 COEFICIENTE DE ALIENACION

Se representa por k y se obtiene mediante la raíz cuadrada de 1- el coeficiente de validez al cuadrado vamos esto

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

Nos indica si hacemos una predicción lo que puede quedar en azar ese valor y cuando se multiplica por 100 sabremos el porcentaje de azar cuando se produce Y para x

8.3.6.3 COEFICIENTE PREDICTIVO

Se obtiene restandole a q el coeficiente de alienación que acabamos de hacer y nos dice el grado de seguridad de una predicción y por 100 da el porcentaje

8.4.1 EL INDICE DE DIFICULTAD ID

Debemos poner mayor o menor dificultad según el grupo al que vaya dirigido. Y se sabe según el número de gente que lo ha sabido contestar si ha habido mucha gente que sabía contestarlas, pues será un índice de dificultad muy bajo y si muy poca gente sabía contestarlas es que la dificultad es alta

Pero no va a ser tan fácil según el % de la gente que contestó bien. Hay una tabla que indica el grado de dificultad de esta

Muy fáciles	10%	0,75 < ID
Fáciles	20%	0,55 < ID < 0,74
Normales (medios)	40%	0,45 < ID < 0,54
Difíciles	20%	0,25 < ID < 0,44
Muy difíciles	10%	ID < 0,24

Peor tampoco podía ser tan fácil para saber este porcentaje debemos hacer el ID que es una miniformula que nos indica la proporción de sujetos que lo resuelven bien. Pero no solo eso sino que este ID índice de dificultad varía dependiendo del tipo de respuesta que pida la prueba. Cuando la prueba pregunta pero no da posibles respuestas a la hora de sumar los aciertos tenidos no hay que restarle (no penalizan los fallos). O sea que la formula de ID sería

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

A → nº de sujetos que aciertan el ítem
n → nº de sujetos que lo contestan

Pero si la prueba te da a elegir en cada ítem entre varias soluciones posibles, los errores penaliza porque así evita que alguien conteste a boteo por eso la fórmula tendremos primero que hallar el nuevo valor A en el que afecta la penalización

corrección por respuesta al azar → $PD = A - \frac{E}{na-1}$; en la que A es el nº de ítems acertados; E es el nº de ítems fallados ó errores, y na es el nº de alternativas de respuesta que se ofrecen. El índice queda entonces:

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

Según este índice oscilará entre 0 y 1 y tendremos que ir a la tabla de arriba para saber el índice de dificultad si es fácil o difícil

Índice de Homogeneidad Ih

Nos indica la coherencia de cada elemento de la prueba. Y en cuanto a todos los ítems miden lo mismo. Si mayor es el coeficiente más homogénea es la prueba. Por supuesto este coeficiente dependerá del tipo de prueba y la forma de respuesta. Si los elementos son dicotómicos será el coeficiente biserial puntual. Si hay una variable continua y otra dicotomizada se usará la correlación biserial. Si las dos variables son dicotomizadas se usará el tetracórico. Y corregiremos el valor de índice con esta fórmula

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

r_{AB} → correlación entre el ítem y el total

s_A → S de las puntuaciones en la prueba

s_B → \sqrt{pq}