PSICOMETRÍA

Tema 1

Introducción a la Psicometría

Los tests, como cualquier otro instrumento de medición, son imperfectos, y que las puntuaciones que obtienen los sujetos cuando se les aplican no representan con exactitud su verdadera puntuación en aquello que se está midiendo ya que están afectadas de errores de medida cuya cuantía es necesario estimar.

INTRODUCCIÓN

Definición

La Psicometría es una disciplina metodológica, dentro del área de la Psicología, cuya tarea fundamental es la **medición o cuantificación** de las variables psicológicas con todas las implicaciones que ello conlleva, tanto teóricas (posibilidades y criterios de medición) como prácticas (cómo y con qué se mide).

Según esta definición, la Psicometría deberá ocuparse:

- De la justificación y legitimación de la medición psicológica
- De las implicaciones prácticas y aplicadas que dicha medición conlleva

La psicología en el marco de las CC

La Psicología, como ciencia del comportamiento, tiene su propio objeto de estudio y sus propios problemas; por eso, deberá adaptar el método general de la ciencia, el método científico, al marco concreto de cada uno de los problemas.

La Psicometría, comparte con el resto de las disciplinas psicológicas tanto el objeto de estudio: la conducta humana, como el método: el método científico; entonces, lo que realmente la caracteriza es la peculiar adaptación que hace de éste al objeto de estudio de la Psicología. Al incluir todo la referente a la medición, la Psicometría proporciona las reglas que van a permitir llevar a cabo el proceso de operativización de las variables que se quieren medir.

La medición en psicología (I)

Para Nunnally (1970), la medición se reduce a algo muy sencillo, consiste en un conjunto de normas para asignar números a los objetos de modo tal que estos números representen cantidades de atributos, entendiendo por atributos las características de los objetos y no los objetos mismos.

La mayoría de los atributos psicológicos como por ejemplo la inteligencia, el autoritarismo y la introversión, son conceptos abstractos, denominados también **constructos teóricos (variables latentes),** cuya medición no puede llevarse a cabo de forma directa, por lo que se hace mediante **indicadores**. En este sentido, Zeller y Carmines (1980) plantean una nueva concepción de la medición; consideran que se trata de *un proceso mediante el cual se enlazan conceptos abstractos (los constructos inobservables directamente), con indicadores empíricos observables directamente (las conductas).*

La medición en psicología (II)

En Psicología, hay dos formas fundamentales de abordar el problema de las unidades de medida, una **referida a normas** y otra **referida al criterio**. La forma más habitual de proceder es la primera, la referida a normas, que consiste en comparar los resultados obtenidos con los obtenidos por un grupo de referencia llamado grupo normativo.

En otras ocasiones la interpretación se hace en relación a un criterio previamente establecido ; los resultados obtenidos se comparan con ese criterio (un punto crítico) y la superación o no del mismo es lo que va a dar significado a las puntuaciones obtenidas.

Orígenes y vías de desarrollo (I)

De acuerdo con Yela (1968), podemos decir que existieron dos motivos fundamentales que posibiltaron la introducción de la medición en Psicología:

- Uno de ellos, la tendencia a formular los problemas científicos en términos matemáticos;
- El otro motivo, el enfrentamiento de la Psicología, hacia la mitad del siglo XIX con dos problemas fundamentales:
  - el estudio cuantitativo de las relaciones entre las características físicas de los estímulos y las sensaciones que suscitan en os sujetos en función de las cuales se asignan valores numéricos a los estímulos: **el problema psicofísico**.
  - el problema de **la cuantificación de las diferencias individuales** mediante la asignación de números a los sujetos en función del grado en que manifiesten un atributo o conducta.

Orígenes y vías de desarrollo

Se puede situar el origen de la Psicometría hacia la mitad del siglo XIX y, a partir de ese momento, se va a desarrollar, fundamentalmente, a través de estas dos vías:

- los estudios de Psicofísica que dieron lugar al desarrollo de modelos que permitieron asignar valores numéricos a los estímulos y, por lo tanto, que permitieron el **escalamiento de estímulos**.
- los estudios acerca de las diferencias individuales que dieron lugar al desarrollo de los Tests y de las distintas Teorías de los Tests (modelos), que posibilitaron la asignación de valores numéricos a los sujetos y, por lo tanto , el **escalamiento de los sujetos**.

LOS MÉTODOS DE ESCALAMIENTO

Introducción

Dentro del marco de la Psicometría, hay dos supuestos básicos en todos los métodos de escalamiento:

- La existencia de un continuo latente o subyacente, a lo largo del cual varían los objetos psicológicos que se van a escalar (estímulos, sujetos, o respuestas), que no puede ser observado de forma directa.
- Que los objetos psicológicos (estímulos, sujetos o respuestas) pueden situarse de forma ordenada a lo largo de ese continuo.

En el **escalamiento de estímulos** los sujetos actúan como instrumentos de medida, mientras que en el **escalamiento de sujetos** son los estímulos los que actúan como instrumentos de medida.

Hay veces que lo que interesa es situar sobre el continuo, a lo largo del cual varía el atributo que se está estudiando, tanto a los estímulos como a los sujetos. Esta tercera aproximación a los métodos de escalamiento fue denominada por Torgerson (1958) **aproximación centrada en la respuesta**.

Escalamiento psicofísico (I)

Umbral mínimo (Herbart). Es la mínima intensidad que tiene que tener un estímulo para que se pueda percibir.

Ley de Weber. El incremento de magnitud que debe experimentar un estímulo (ΔE) para que el sujeto reciba que se ha producido un cambio, es una proporción constante de su magnitud inicial (E).

$$K = \frac{\Delta E}{E}$$

K → Constante de Weber

ΔE → Incremento mínimo que ha de experimentar la magnitud del estímulo (magnitud física) respecto a la magnitud inicial (E) para el sujeto perciba un cambio mínimo en la sensación (magnitud psicológica). A este cambio mínimo de sensación le denominó Weber **diferencia apenas perceptible (dap).**

Escalamiento psicofísico (II)

MÉTODOS DE ESCALAMIENTO PSICOFÍSICO (II)

Ley de Fechner (I)

En el escalamiento psicofísico se tienen dos continuos, un continuo físico lo largo del cual varían los estímulos y uno psicológico a lo largo del cual varían las sensaciones que dichos estímulos producen. Según Fechner, la función que mejor representa la relación entre los continuos es una función logarítmica:

$$S = C \,.\, \ln E + A$$

Además Fechner introdujo los conceptos de umbral absoluto y umbral diferencial.

El valor de un estímulo se puede calcular a partir del estímulo anterior:

$$E_n = E_{n-1} + K \cdot E_{n-1} = (1 + K)E_{n-1}$$

Supuestos de la ley de Fechner:

1. Asume la ley de Weber.
2. Asume que todas las diferencias apenas perceptibles son psicológicamente iguales.
3. Establece que el origen de la escala de sensación corresponde al umbral absoluto.

Ejemplo

Junio 2018, tipo A

14.- Supongamos que el umbral absoluto de un sujeto para escuchar un sonido es de 40Hz y que para que el sujeto detecte que ha habido un cambio en la intensidad del mismo hay que aumentar la intensidad 16Hz. Asumiendo que se cumpliera la Ley de Fechner, los valores para el estímulo que correspondería a la siguiente escala de sensación serían: a) 40, 50.25, 75.34, 109.16; b) 40, 56, 78.4, 109.76; c) 40, 50.25, 78.4, 109.16

E	**40**			
S (dap)	0	1	2	3

escalamiento psicofísico (III)

MÉTODOS DE ESCALAMIENTO PSICOFÍSICO (III)

Ley de Fechner (II)

El problema fundamental está en la obtención experimental de los umbrales absoluto y diferencial punto para ello, Fechner desarrollo una serie de procedimientos llamados **métodos indirectos**:

- Método de los límites o de los cambios mínimos.
- Método de ajuste o de error promedio.
- Método de los estímulos constantes punto es el más utilizado para el cálculo de los umbrales absolutos diferencial .

La ley de Fechner recibió varias críticas debidas a que los estímulos de muy alta o muy baja intensidad no se ajustan bien a ella y tampoco lo hacen todos los sistemas sensoriales.

Escalamiento psicofísico (IV)

Ley de Stevens

Stevens reformula la psicofísica defender con una nueva función conocida con el nombre de función potencial de Stevens:

$$R = q \cdot E^n$$

donde

R → respuesta de los sujetos

E → valor del estímulo

q → Constante que depende de las unidades de medida

n → exponente de la función que depende del atributo sensorial

Escalamiento psicofísico (V)

Métodos directos

- Métodos de emparejamiento de magnitudes
  - ✓ Estimación de magnitudes
  - ✓ Producción de magnitudes
  - ✓ Modalidad cruzada
- Método de emparejamiento de razones
  - ✓ Estimación de razones
  - ✓ Producción de razones
  - ✓ Modalidad cruzada
- Métodos de emparejamiento de intervalos
  - ✓ Estimación de intervalos
  - ✓ Producción de intervalos
  - ✓ Modalidad cruzada

Escalamiento psicológico (I)

Thurstone trató de elaborar un modelo a partir del cual pudiera elaborar una escala sobre un continuo psicológico y situar en ella los estímulos (también psicológicos) sin necesidad de recurrir a ninguna operación en un continuo físico. El modelo que desarrolló está basado, por una parte, en la variabilidad perceptual de los sujetos, e incluso en la de un mismo sujeto cuando se le presentan los mismos estímulos en distintas ocasiones y, por otra, en la limitación que tienen los sujetos para percibir las diferencias de magnitud entre dos estímulos cuando estas son muy pequeñas.

Escalamiento psicológico (II)

- **Escalas de Thurstone**
  - ✓ Ley del juicio comparativo
  - ✓ Ley del juicio categórico
- **Escalograma de Guttman**. Permite el escalamiento conjunto de sujetos y estímulos (escalamiento de respuestas). La escala resultante se denomina **escala de entrelazamiento.**
- **Teoría del despliegue de Coombs**. También permite escalar sujetos y estímulos conjuntamente.

Todos estos modelos escalamiento se desarrollaron para la construcción de escalas **unidimensionales**.

En los años 60 se desarrollaron los métodos de **escalamiento multidimensional***, que* permiten la ordenación de los objetos a escalar atendiendo simultáneamente a más de un atributo o característica y, por lo tanto, asumen la existencia de más de una dimensión subyacente al conjunto de observaciones. En este caso, en lugar de asignar un único valor en la escala a cada uno de los objetos, se les asignará un valor en cada una de las dimensiones analizadas.

ORÍGENES Y DESARROLLO DE LOS TEST

Introducción

El desarrollo del Método de los Tests y de las distintas Teorías de los Tests, se produjo por el intento de apreciar de forma sistemática las diferencias individuales; es decir, el intento de escalar a los sujetos. Hubo que esperar hasta la mitad del siglo XIX para que se desarrollaran en Europa y en Estados Unidos procedimientos formalizados para la evaluación de tales diferencias: **los tests**.

> *Un test es un instrumento de medición diseñado especialmente para estudiar de un modo objetivo y sistemático el nivel de los sujetos respecto a algún atributo, característica* o *dominio de conocimientos y, a partir de las puntuaciones que obtengan los sujetos en el test poder analizar las diferencias existentes entre ellos.*

Hay tres factores que se pueden considerar decisivos en el desarrollo de los tests: la apertura del laboratorio antropométrico de Galton en Londres, el desarrollo de la correlación de Pearson y la interpretación que Spearman hace de ella

Los trabajos de Galton

Galton todavía creía que la medición de las características mentales estaba estrechamente relacionada con la de las características físicas, *y* pensaba que para obtener una descripción completa del hombre, las medidas antropométricas de estatura, peso, color de la piel, etc., deberían completarse con medidas psicométricas de los sentidos, la memoria, etc.

En 1884 montó su laboratorio antropométrico en la Exposición Internacional de la Salud. Consideraba que la inteligencia podía medirse a través de la actividad sensomotora y, una vez recogidos los datos, comenzó elaborando distribuciones de frecuencias y comprobando que su forma era aproximadamente la misma para las variables psicológicas que para las anatómicas, la distribución normal. Pronto se plantea el problema de cómo encontrar relaciones entre medidas diferentes. Galton recurrió a la correlación como medida de asociación. Así, Galton fue el primero en aplicar el concepto estadístico de distribución normal, de media, de mediana, varianza y correlación a datos psicológicos. Esta tarea fue continuada por Karl Pearson.

Para poder llevar a cabo sus mediciones, Galton construyó y utilizó una serie de instrumentos Que pueden ser considerados los primeros *Tests psicométricos,* pero éstos se caracterizaban por un fuerte carácter antropométrico, sensorial y motor.

Los primeros test mentales

James McKeen Cattell en 1888, estuvo en la Universidad de Cambridge y allí se puso en contacto con Galton. Cattell utilizó por primera vez el término *Test Mental,* pero los tests a los que hace referencia Cattell, al igual que los de Galton, tenían un carácter sensorial y motor fundamentalmente. El análisis de los datos puso de manifiesto que la correlación entre este tipo de pruebas y el nivel intelectual de los sujetos era nula.

Los primeros test de inteligencia

Alfred Binet (1857-1911), fue el primero en darse cuenta de que las sensaciones no jugaban un papel demasiado importante en la Psicología diferencial y que era necesario centrarse en el estudio de los procesos mentales superiores.

Binet trabajó con Simon, y publicaron, en 1905, lo que puede ser considerada la primera escala de inteligencia. A diferencia de los tests de Galton, este test ya no se ocupaba de las funciones motoras o sensoriales sino de la capacidad de comprensión y razonamiento de los niños. La adaptación más conocida del test de Binet fue la que realizó en 1916 el psicólogo americano Terman en la Universidad de Stanford, de ahí el nombre de **Test Stanford-Binet de Terman**.

Aunque la finalidad inicial de la escala era detectar a los sujetos que presentaran algún retraso intelectual, posteriormente se utilizó para el estudio de las diferencias en otros niveles. Para poder interpretar las puntuaciones obtenidas, se desarrolló el concepto de **Edad Mental.**

Terman consideraba que la inteligencia se podía definir como la habilidad para llevar a cabo razonamientos abstractos y, a pesar de que sigue la táctica de Binet para la elaboración de sus tests, utiliza como medida de la inteligencia el concepto de **Cociente Intelectual (CI**) acuñado por Stern en 1912.

Los tests colectivos (I)

A raíz de la entrada de EE.UU. En la Primera Guerra mundial en 1917, se nombró una comisión especial de la Asociación Norteamericana de Psicología y se diseñaron los famosos tests conocidos como *Tests Alpha y Beta* que fueron publicados en 1918. El Test Alpha fue diseñado para la población general y el Test Beta, para los reclutas analfabetos o aquellos que no dominaban el inglés.

Test de inteligencia

Nombre	Publicación
Alpha	1918
Beta	1918
Wechler-Belleuve	1938
WISC	1949
Wechler Adult Intelligence Scale (WAIS)	1955, 1981
Wechler Preschool and Primary Scale of Intelligence	1967

El análisis factorial

En un intento de dar respuesta a qué miden los tests y si existen los rasgos que dicen medir, se desarrollaron el coeficiente de correlación de Pearson y una serie de técnicas correlacionales conocidas bajo el nombre genérico de **Análisis Factorial.**

El objetivo común de las técnicas reunidas bajo el término Análisis Factorial es representar un conjunto de variables observables (escalas, ítems, etc.) mediante un número más pequeño de variables no observables sino latentes llamadas factores de las que las primeras son indicadores.

La influencia del Análisis Factorial respecto a los tests de inteligencia fue doble.

- Por una parte dio una fundamentación teórica a su utilización, mostrando que en la mayor parte de las funciones cognoscitivas interviene un factor general «g» común a todas ellas.
- Por otra, los resultados del análisis factorial subrayan la importancia de estudiar otras aptitudes más específicas en el campo de la inteligencia.

Baterías de aptitud múltiple

A partir de los años treinta, y sobre todo a partir de los cuarenta del siglo XX, cuando ya la técnica del análisis factorial empezaba a dar sus frutos, comenzaron a desarrollarse las baterías de tests *(Baterías de Aptitud Múltiple),* cuya finalidad era procurarnos una medida de la posición de cada sujeto en un cierto número de rasgos.

Entre las primeras baterías merece destacar la *Batería de Aptitudes Mentales Primarias* de Thurstone

Los tests de personalidad

Se puede considerar a Kraepelin como un precursor de los *Tests de Personalidad,* pues ya en 1894 utilizó el *Test de Asociación Libre* con pacientes normales para estudiar los efectos psicológicos de la fatiga, el hambre y las drogas.

Nombre	Publicación
Test de Asociación Libre	1894
Hoja de Datos Personales de Woodworth	1917
Test de Psicodiagnóstico de Rorschach	1921
Test de Dibujo de la Figura Hiumana	1926
Inventario Multifásico de Personalidad de Minnesota (MMPI)	1942

Medición de intereses y actitudes

Entre los instrumentos clásicos para la medida de los intereses merecen citarse el *Cuestionario de Intereses Vocacionales de Strong (SVIB)* desarrollado por E.K. Strong (1927) y la *Escala de Preferencias de Kuder* introducida por Kuder en 1934.

En cuanto a la medida de las actitudes, creencias y opiniones, su desarrollo efectivo tuvo lugar a partir de 1928, fecha en la que Thurstone desarrolló sus dos modelos, la **Ley del juicio Comparativo** *y,* sobre todo, la **Ley del Juicio Categórico.**

Algunos años más tarde Likert (1932) desarrolló una nueva técnica que vino a paliar alguno de los inconvenientes de la de Thurstone, entre otros la prueba de jueces.

Con el fin de medir el significado connotativo, también llamado significado afectivo o subjetivo, que determinados estímulos tienen para los sujetos, Osgood (1952) revisó todas las teorías del significado *y* así pudo encontrar el marco teórico que le permitió desarrollar una escala de clasificación para medirlo: *El Diferencial Semántico.*

Institucionalización del uso de los tests

El uso masivo de los tests fue extendiendo su campo de aplicación a distintos sectores tanto públicos como privados, en la selección de puestos en el gobierno y en las escuelas, para ayudar a niños con problemas escolares, en orientación, clasificación, etc., lo que ocasionó la aparición de instituciones tanto públicas como privadas dedicadas a ello.

En la publicación llevada a cabo en 1999, la *American Psychological Association (APA),* la *American Educational Research Association (AERA),* y el *National Council on Measurements in Education (NCME),* publicaron los *Standards far Educational and Psychological Testing,* con el fin de proporcionar a los profesionales de la Psicometría criterios para la evaluación de los tests. La última edición es la de 2014.

Tests referidos a criterio y a normas (I)

Las diferencias entre los TRN y los TRC no siempre son aparentes a simple vista ya que ambos tipos de tests están formados por ítems de formatos similares, requieren el mismo tipo de operaciones cognitivas de los sujetos, etc. Sin embargo, como recoge Martínez-Arias (1995, pág. 657) hay grandes diferencias tanto en su construcción como en la interpretación de las puntuaciones obtenidas por los sujetos:

La finalidad de la evaluación:

Mientras que en los TRN el objetivo es poner de manifiesto las diferencias individuales en la conducta o rasgo que mide el test, en los TRC el objetivo es estimar el rendimiento o conducta del sujeto en los objetivos que mide el test.

La construcción del test y la especificación de los contenidos:

En los TRN a la hora de construir los elementos que formarán el test se suele recurrir a las teorías existentes respecto al rasgo o constructo que se quiere medir, pero no suelen tener una delimitación clara del dominio de contenidos que se van a evaluar. Por el contrario, en los TRC lo primero que hay que hacer es especificar de una manera clara el dominio de contenidos o conductas que se quieren evaluar y el uso que se pretende hacer del test.

Tests referidos a criterio y a normas (II)

La forma de seleccionar los ítems:

En los TRN los ítems deben poner de relieve las diferencias individuales maximizando la varianza del test, y seleccionando ítems de dificultad media y alto poder díscrímínatívo. En los TRC los ítems se seleccionan en función de los objetivos y del uso que se vaya a hacer del test.

El significado de las puntuaciones:

En los TRN la puntuación obtenida por los sujetos se considera un indicador de su puntuación verdadera en un rasgo latente. En los TRC la puntuación es un estimador de la conducta o rendimiento del sujeto en el dominio.

La interpretación de las puntuaciones:

Mientras que en los TRN la puntuación tiene significado únicamente con relación a los resultados del grupo normativo, en los TRC la puntuación tiene significado en términos absolutos.

Tests adaptativos informatizados

Se han modificado los procedimientos de administración de tests; cada vez son más populares las aplicaciones a través del ordenador. Un ejemplo de este tipo de aplicaciones lo constituyen los **Tests Adaptativos Informatizados (TAI's)** denominados así porque la selección de los ítems se va haciendo a lo largo del proceso de administración del test, de manera que los ítems se puedan *adaptar* al nivel de cada sujeto y no le resulten ni demasiado fáciles ni demasiado difíciles.

Desarrollo de la teoría de los tests

En todo proceso de medición, aunque sea en el campo de la física o de la biología, se cometen errores de medida, y muchos más se cometerán cuando se trata de medir variables psicológicas.

Ante estos problemas es lógico pensar en la necesidad de algún procedimiento que permita analizar, por una parte, la precisión de las medidas obtenidas **(fiabilidad de las puntuaciones)** y, por otra parte, la validez de las inferencias o conclusiones que se puedan sacar a partir de las mismas **(validez).**

Para dar respuesta a estos problemas se desarrolló un marco teórico general, la *Teoría de los Tests,* que va a permitir establecer una relación funcional entre las variables observables (a partir de las puntuaciones empíricas obtenidas por los sujetos en los tests o en los ítems que los componen), y las variables inobservables (las puntuaciones verdaderas o el nivel de habilidad de los sujetos en el rasgo)

Cada modelo de relación entre el nivel de rasgo y las puntuaciones empíricas del test podría dar origen a una Teoría de los Tests, pero las que han tenido una mayor incidencia en este campo han sido: la *Teoría Clásica de los Tests* (TCT), y la *Teoría de Respuesta al Ítem* (TRI).

Teoría clásica de los tests (I)

El modelo lineal de Spearman, es un modelo aditivo en el que la puntuación observada (variable dependiente) de un sujeto en un test (X) es el resultado de la suma de dos componentes: su puntuación verdadera (variable independiente) (V) y el error (E) que inevitablemente lleva asociado todo proceso de medición.

A partir de los supuestos y las deducciones que se extraigan a partir de los mismos, se podrá hacer una estimación de esos errores. La expresión formal del modelo es:

$$X = V + E$$

Puntuación empírica u observada (VD)

Puntuación verdadera (VI)

Error de medida. Engloba todas las fuentes de error

Teoría clásica de los tests (II)

Se han realizado algunos intentos para sistematizar y clasificar el error en función de las posibles fuentes que lo originan, siendo la **Teoría de la Ceneralizabilidad (TG)** propuesta por Cronbach y sus colaboradores el más ambicioso y global de los intentos realizados para estimar la fiabilidad de un instrumento de medida, analizando de forma sistemática las posibles fuentes de error.

A pesar del avance que en cierto modo supuso la TG, suele ser considerada más como una extensiónnde la TCT que como un modelo alternativo.

Teoría de respuesta al item

Lord (1953) observó que cuando a una muestra de sujetos se les aplicaba un test, o una serie de tests, para evaluar su nivel en un determinado rasgo, la puntuación obtenida dependía del conjunto de ítems o tests utilizados. También los estadísticos de los ítems, su índice de dificultad y de discriminación, dependían de la muestra de sujetos utilizada para su cálculo.

Estos dos problemas, junto con el de la indiferenciación del error (sólo había un componente error que englobaba a todos), fueron los que centraron las críticas hechas a la TCT.

La solución más adecuada se encontró en el marco de la **Teoría de Respuesta al Ítem (TRI),** que proporciona una serie de modelos que asumen una relación funcional entre los valores de la variable que miden los ítems (nivel de habilidad de los sujetos en el rasgo medido) y la probabilidad de que los sujetos, en función de su nivel de habilidad, acierten cada ítem. A esta función se la conoce con el nombre de *Curva Característica del Ítem*. La probabilidad de que un sujeto acierte a cada uno de los ítems no depende ya del propio ítem depende, exclusivamente, del nivel de los sujetos en la variable medida

A pesar del gran desarrollo de la TRI hacia 1980, la TCT sigue en auge ya que hay problemas que se pueden solucionar más eficaz *y* rápidamente dentro de este marco.

FIN DEL TEMA 1

Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 2

Principios básicos para la construcción de instrumentos de medición psicológica

TESTS, ESCALAS. CUESTIONARIOS E INVENTARIOS

DEFINICIONES

Test. En general, se ha utilizado el término test como un término general para referirnos a cualquier instrumento de medida. Pero sobre todo se utiliza para hacer referencia a los instrumentos de medición de **variables de tipo cognitivo**: aptitudes, conocimientos, rendimiento, habilidades … en los que las respuestas de los sujetos a cada uno de los elementos son correctas o incorrectas.

Escala. El término escala se suele utilizar para hacer referencia a los instrumentos elaborados para medir **variables no cognitivas**: actitudes, intereses, preferencias, opiniones… y se caracterizan porque los sujetos han de responder eligiendo, sobre una escala de categorías graduada y ordenada, aquella categoría que mejor represente su posición.

Cuestionarios. Suelen estar formados a menudo por una serie de ítems o elementos que no están necesariamente relacionados unos con otros, cuyas opciones de respuesta no están ordenadas ni graduadas y en los que tampoco hay respuestas correctas o incorrectas.

Inventarios. Son los instrumentos elaborados para medir **variables de personalidad**. Las respuestas de los sujetos a los ítems no son correctas o incorrectas, lo único que demuestran es conformidad o no de los sujetos con los enunciados de los ítems.

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
Características del test	Cuál va a ser el contenido Qué tipo de ítems se van a incluir Cuántos ítems Características psicométricas
Redacción de los ítems	Ítems de elección Ítems de construcción
Revisión por expertos	Qué ítems van a seleccionarse
Confección de la prueba piloto	Instrucciones de administración Formato de presentación Formato de registro de respuestas
Aplicación de la prueba piloto	Individual – colectiva Papel y lápiz – informatizada Correo, entrevista, por teléfono…
Corrección de la prueba piloto y asignación de puntuaciones	En test formados por ítems de elección En test formados por ítems de construcción

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
C	Cuál va a ser el conte Qué tipo de íte Cuántos ítems Características
Redacción de los ítems	Ítems de elecci Ítems de const
Revisión por expertos	Qué ítems van a seleccionarse
Confección de la prueba piloto	Instrucciones de administración Formato de presentación Formato de registro de respuestas
Aplicación de la prueba piloto	Individual – colectiva Papel y lápiz – informatizada Correo, entrevista, por teléfono…
Corrección de la prueba piloto y asignación de puntuaciones	En test formados por ítems de elección En test formados por ítems de construcción

Los constructos (variables teóricas inobservables), se manifiestan a través de una serie de conductas que ya sí son observables de forma directa

Selección
Clasificación/colocación
Diagnóstico
Certificación
Orientación/consejo
Descripción/orientación

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
C	Cuál va a ser el contenido Qué tipo de ítems se van a incluir Cuántos ítems Características psicométricas
R	Ítems de elección Ítems de construcción
R	Qué ítems van a seleccionarse
C	Instrucciones de administración Formato de presentación Formato de registro de respuestas
Aplicación de la prueba piloto	Individual – colectiva Papel y lápiz – informatizada Correo, entrevista, por teléfono…
Corrección de la prueba piloto y asignación de puntuaciones	En test formados por ítems de elección En test formados por ítems de construcción

- Ítems de elección
  - Dos alternativas
  - Elección múltiple
  - Emparejamiento
  - Escalas de clasificación
  - Listados
- Ítems de construcción
  - Ítems de respuesta corta
  - Ítems de respuesta extensa

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
	Cuál va a ser el contenido Qué tipo de ítems se van a incluir Cuántos ítems Características psicométricas
	Ítems de elección Ítems de construcción
	Qué ítems van a seleccionarse
	Instrucciones de administración Formato de presentación Formato de registro de respuestas
Aplicación de la prueba piloto	Individual – colectiva Papel y lápiz – informatizada Correo, entrevista, por teléfono…
Corrección de la prueba piloto y asignación de puntuaciones	En test formados por ítems de elección En test formados por ítems de construcción

- Dificultad
  - Test de velocidad
  - Test de ejecución máxima (test de potencia)
  - Test de ejecución típica
- Homogeneidad
  - Constructo unidimensional
  - Constructo multidimensional
- Discriminación

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
Características del test	Cu Qu Cu Ca
Redacción de los ítems	Íte
Revisión por expertos	
Confección de la prueba piloto	Ins For Formato de registro de respuestas
Aplicación de la prueba piloto	Individual – colectiva Papel y lápiz – informatizada Correo, entrevista, por teléfono…
Corrección de la prueba piloto y asignación de puntuaciones	En test formados por ítems de elección En test formados por ítems de construcción

- Evitar la ambigüedad de los enunciados
- Evitar enunciados que provocan respuestas sesgadas
- Expresar una única idea en el enunciado
- Evitar las dobles negaciones en los enunciados

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
Características del test	Cu O
Redacción de los ítems	Íte Íte
Revisión por expertos	Qu
Confección de la prueba piloto	Ins For For
Aplicación de la prueba piloto	Individual – colectiva Papel y lápiz – informatizada Correo, entrevista, por teléfono…
Corrección de la prueba piloto y asignación de puntuaciones	En test formados por ítems de elección En test formados por ítems de construcción

Sesgos de respuesta

- Aquiescencia o tendencia a responder sistemáticamente (de acuerdo o desacuerdo)
- Deseabilidad social o tendencia a responder de forma socialmente aceptable
- Indecisión o tendencia a seleccionar la alternativa central o neutra
- Respuesta extrema o tendencia a elegir una categoría extrema

PROCESO CONSTRUCCIÓN DEL TEST

Etapas	Decisiones a tomar
Finalidad del test	Qué se va a medir A quién se va a medir Para qué se va a medir
Características del test	Cuál va a ser el contenido Qué tipo de ítems se van a incluir Cuántos ítems Características psicométricas
Redacción de los ítems	Ít Ít
Revisión por expertos	C
Confección de la prueba piloto	Ir F F
Aplicación de la prueba piloto	
Corrección de la prueba piloto y asignación de puntuaciones	E En test formados por ítems de construcción

ASIGNACIÓN DE PUNTUACIONES

- Pruebas cognitivas
  - Puntuación = $\sum X_i$
  - Puntuaciones corregidas
    - Penalización de errores

$$X_c = A - A_a = A - \frac{E}{k-1}$$

    - Bonificación de omisiones

$$X_c = A + A_a = A + \frac{O}{k}$$

- Pruebas no cognitivas
  - Cada ítem tiene su valor

FIN DEL TEMA 2

Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 3

Técnicas para la construcción de escalas de actitudes

INTERESES, VALORES Y ACTITUDES

DEFINICIONES

Aunque el tema se centra en la construcción de escalas para la medición de las actitudes, las técnicas expuestas se pueden adaptar para su utilización en la medición de intereses y valores .

Nunally (1978) hace la siguiente distinción :

- Los **intereses** son preferencias por actividades particulares.
- Los **valores** hacen referencia a preferencias sobre objetivos de vida y formas de vida más que sobre actividades concretas
- Las **actitudes** hacen referencia a sentimientos acerca de un determinado objeto social.

Tanto en la Ley del Juicio Comparativo como en la Ley del juicio Categórico los sujetos que evalúan los ítems han de emitir un **juicio objetivo** acerca del grado de atributo que consideran que contiene cada uno de ellos y no mostrar su actitud o postura personal ante cada uno de los enunciados. Se les pide que emitan **juicios de hecho no juicios de valor**. A esta prueba se la denomina **prueba de jueces**.

EL MODELO ESCALAR DE THURSTONE

SUPUESTOS BÁSICOS DEL MODELO

1. Existe un continuo psicológico o subjetivo a lo largo del cual varía el atributo o característica que se está estudiando.
2. Cada uno de los estímulos que se va a estudiar origina en el sujeto un proceso subjetivo, proceso discriminante, a través del cual les asignará un valor subjetivo en el continuo psicológico.
3. Cuando un estímulo es presentado en repetidas ocasiones al mismo sujeto, no siempre origina en él el mismo proceso discriminante y, por tanto, el valor subjetivo asignado varía.
4. Si el número de veces que se presenta cada estímulo es muy grande se asume que la distribución de los valores subjetivos es normal.
5. La media de esta distribución, denominada **distribución discriminante**, recibe el nombre de **valor escalar** del estímulo. La desviación típica de la distribución es la **dispersión discriminante** y nos da idea de la ambigüedad suscitada por el estímulo.
6. Si se presentan varios estímulos en repetidas ocasiones a un mismo sujeto, cada uno de ellos dará lugar a una distribución discriminante distinta.
7. El modelo se cumple tanto si es un solo sujeto el que emite numerosos juicios, como si es una muestra grande de sujetos cada uno de los cual es emite un único juicio acerca del estímulo.

LA LEY DEL JUICIO COMPARATIVO

Este modelo utiliza el método de las comparaciones binarias para la obtención de los datos. La tarea de los sujetos consiste en comparar cada uno de los estímulos con todos los demás y decir cuál es el preferido.

A partir de estos datos obtendremos una serie de matrices:

- Matriz de frecuencias
- Matriz de frecuencias ordenadas
- Matriz de proporciones
- Matriz de puntuaciones típicas

A partir de esta última matriz podemos calcular la media de la puntuación típica de cada estímulo, corrigiendo en la magnitud necesaria para que no haya valores negativos.

EJEMPLO (I)

4.- Se desea conocer las preferencias de los españoles respecto a cuatro marcas de detergente. Para ello se pide a una muestra de 100 sujetos que indiquen sus preferencias respecto a los mismos. La matriz de frecuencias obtenida es la siguiente:

	1	**2**	**3**	**4**
1		60	75	20
2	40		70	35
3	25	30		10
4	80	65	90	

Construir la escala

EJEMPLO (II)

A partir de la matriz de frecuencias, reordenamos de mayor a menor. Para ello sumamos las columnas

Matriz de frecuencias

	1	**2**	**3**	**4**
1		60	75	20
2	40		70	35
3	25	30		10
4	80	65	90	
	145	155	235	65

Matriz de frecuencias ordenada

	3	**2**	**1**	**4**
3		30	25	10
2	70		40	35
1	75	60		20
4	90	65	80	
	235	155	145	65

Se observa que en ambas matrices de frecuencias, la suma de las celdillas simétricas da 100 (número de jueces)

EJEMPLO (III)

A partir de la matriz de frecuencias ordenadas obtenemos la matriz de proporciones, dividiendo cada frecuencia por el número de jueces. La matriz de típicas se obtiene buscando cada proporción en la tabla de la curva normal.

Matriz de proporciones

	3	**2**	**1**	**4**
3	0.5	0.30	0.25	0.10
2	0.70	0.5	0.40	0.35
1	0.75	0.60	0.5	0.20
4	0.90	0.65	0.80	0.5

Matriz de típicas

	3	**2**	**1**	**4**
3	0	-0.52	-0,67	-1,28
2	0.52	0	-0.25	-0.39
1	0.67	0.25	0	-0.84
4	1.28	0.39	0.84	0

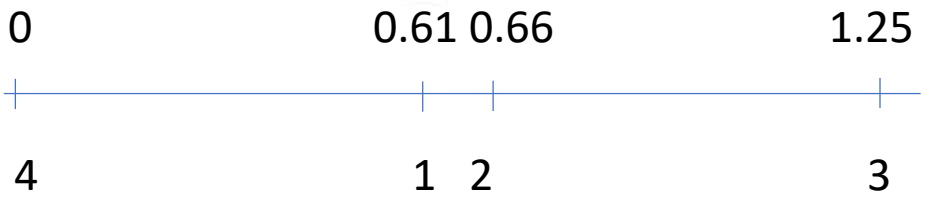
Se observa que en la matriz de proporciones, la suma de las celdillas simétricas da 1 y en la matriz de típicas da 0. En la diagonal principal de la matriz de proporciones hemos añadido 0,5

EJEMPLO (III)

A partir de la matriz de típicas hacemos la media para cada estímulo y corregimos si es necesario, para evitar los valores negativos

Matriz de típicas

	3	**2**	**1**	**4**
3	0	-0.52	-0,67	-1,28
2	0.52	0	-0.25	-0.39
1	0.67	0.25	0	-0.84
4	1.28	0.39	0.84	0
∑Z	2.47	0.12	-0.08	-2.51
$\bar{Z}$	0.62	0.03	-0.02	-0.63
+0.63	1.25	0.66	0.61	0



EJEMPLO 1

Junio 2017, segunda semana

4.- Se desea conocer las preferencias de los españoles respecto a cuatro líderes políticos. Para ello se pide a una muestra de 3000 sujetos que indiquen sus preferencias respecto a los mismos. La matriz de puntuaciones típicas obtenida es la siguiente:

	1	**2**	**3**	**4**
1	0			
2	0.65	0		
3	0.70	0.45	0	
4	0.85	0.55	0,34	0

El valor escalar transformado de los cuatro políticos valorados es: a) VE1 =0.99; VE2=0.53; VE3 =0.24; VE4 =0; b) VE1 =0.80; VE2=0.45; VE3 =0.18; VE4 =0; c) VE1 =0.78; VE2=0.40; VE3 =0.20; VE4 =0.18.

EJEMPLO 2

Junio 2019, segunda semana

Dada la siguiente matriz de frecuencias observadas obtenida tras aplicar el método de las comparaciones binarias de Thurstone, responder a las preguntas 12 y 13.

Estímulos	1	2	3
1	---	60	30
2	40	---	70
3	70	30	---

12. La ordenación escalar de los estímulos sería de menos preferido a más preferido: a) 1, 2, 3; b) 1, 3, 2; c) 2, 1, 3.

13. El valor escalar del estímulo 2, después de la transformación lineal correspondiente, es: a) 0,18; b) 0; c) 0,27.

LA LEY DEL JUICIO CATEGÓRICO (I)

En este modelo, además de los supuestos generales, es necesario asumir que el continuo psicológico de cada sujeto-juez puede ser dividido en una serie de categorías ordenadas.

Para la obtención empírica de los datos la Ley del Juicio Categórico utiliza fundamentalmente tres procedimientos: el de ordenación por rangos, el de los intervalos sucesivos y el de los intervalos aparentemente iguales.

LA LEY DEL JUICIO CATEGÓRICO (II)

MÉTODO DE LOS INTERVALOS APARENTEMENTE IGUALES

A cada uno de los jueces se les instruye para que imaginen un espacio dividido en una serie de categorías ordenadas (normalmente 11, pero siempre impar) desde la categoría con un grado de actitud más negativa hasta la categoría con un grado actitud más positiva.

La tarea de los jueces consistirá en clasificar cada uno de los estímulos en una de las categorías en función del grado y la dirección del atributo que manifiesten.

Para hallar el **valor escalar** de los estímulos, a través de este método, basta calcular la mediana de la distribución de frecuencias originada por cada uno de ellos.

Como medida del grado de acuerdo entre los jueces se puede utilizar el **coeficiente de ambigüedad** que es la distancia entre el tercer y el primer cuartil. ($CA=Q_3-Q_1$)

Si el coeficiente de ambigüedad es mayor que 2, el elemento será considerado ambiguo y deberá eliminarse de la escala definitiva. Elementos neutrales, es decir, en aquellos cuyo valor escalar este en la categoría central, el cociente de ambigüedad puede llegar a 3.

EJEMPLO 2

Septiembre de 2018, tipo A

Se quiere elaborar una escala que mida xenofobia. Para calcular el grado de actitud implícita en cada uno de los ítems se va a utilizar el método de los intervalos aparentemente iguales y 7 categorías de respuesta. El resultado de la evaluación de los jueces a uno de los ítems de la escala es el siguiente:

Escala del ítem	1	2	3	4	5	6	7
Jueces	2	5	8	10	25	40	10

12. El valor escalar del ítem es: a) 4,5; b) 4,75; c) 5,5.

13. ¿Se debería aceptar el elemento para formar parte de la escala definitiva?: a) SI porque el coeficiente de ambigüedad es menor que 2; b) NO, porque el coeficiente de ambigüedad es menor que 2; c) NO, porque el coeficiente de ambigüedad es mayor que 2.

LA TÉCNICA DE LIKERT

LA TÉCNICA DE LIKERT

Se trata de una escala sumativa en la que se asume que los ítems están monotónicamente relacionados con el rasgo subyacente que se quiere medir (a más rasgo, más puntuación). La puntuación total se obtiene sumando las puntuaciones de los sujetos a cada uno de los ítems.

En las escalas Likert lo que se pide a los sujetos son juicios de valor. La forma de responder a los ítems puede variar aunque normalmente se responden en función de cinco categorías:

a) Completamente de acuerdo

b) De acuerdo

e) Indiferente

d) En desacuerdo

e) Completamente en desacuerdo

EL DIFERENCIAL SEMÁNTICO DE OSGOOD

EL D. S. DE OSGOOD

Se trata de una escala de clasificación elaborada por Osgood y sus colaboradores (1957), con el fin de medir el significado connotativo, también llamado significado afectivo o subjetivo, que determinados estímulos tienen para los sujetos.

Osgood estaba interesado en las reacciones emocionales que las palabras o conceptos producen en las personas.

Los conceptos. Los estímulos pueden ser de lo más variado y aunque, en general, se refieren a conceptos verbales (Dios, madre, educación, acciones políticas, etc.), se pueden referir a conceptos no verbales (cuadros, esculturas, estímulos físicos, etc.)

Las escalas bipolares. Se encontraron tres factores:

- Factor evaluativo o valorativo. Es el que explica mayor porcentaje de varianza.
- Factor de potencia.
- Factor de actividad

EJEMPLO 3

Septiembre 2016, reserva

21. En la tabla adjunta se presenta el número de sujetos que respondieron a cada una de las categorías de las tres escalas bipolares utilizadas para medir la dimensión *potencia* del concepto *Politico.* El valor de la puntuación factorial en dicha dimensión es igual a: a) 2.33; b) 4.33; c) 1.86

	1	2	3	4	5	6	7
Agradable-Desagradable	8	15	12	40	25	35	15
Conciliador -Agresivo	15	10	20	30	25	40	10
Débil -Fuerte	12	20	20	30	23	35	10

EL ESCALOGRAMA DE GUTTMAN

EL ESCALOGRAMA DE GUTTMAN

Está diseñado de tal forma que es posible conseguir la ordenación tanto de los sujetos como de los estímulos (los ítems) respecto a una dimensión determinada, asegurando, si los resultados demuestran que los datos se ajustan al modelo propuesto por Guttman, que el conjunto de ítems que conforman la escala miden una única dimensión.

Guttman propuso que como criterio de bondad de ajuste se utilizará el Coeficiente de Reproductividad cuya fórmula es:

$$CR = 1 - \frac{\text{Nº de errores}}{\text{Total de respuestas}} = 1 - \frac{\text{Nº de errores}}{\text{Nº de ítems x Nº de sujetos}}$$

Se admite que los datos se ajustan al modelo de Guttman si CR ≥ 0,9

EJEMPLO 4

Junio de 2015, 2ª semana

16. En la tabla siguiente se recogen las respuestas de 5 sujetos ante la presentación de 4 estímulos en un experimento d detección de señales (1= acierto; 0= error). El coeficiente de reproductividad es: a) 0.80; b) 0.90; c) 0.85.

		EST.		
Suj.	X1	X2	X3	X4
1	1	0	0	1
2	1	1	1	0
3	1	1	0	1
4	1	0	0	0
5	1	0	0	0

EJEMPLO 5

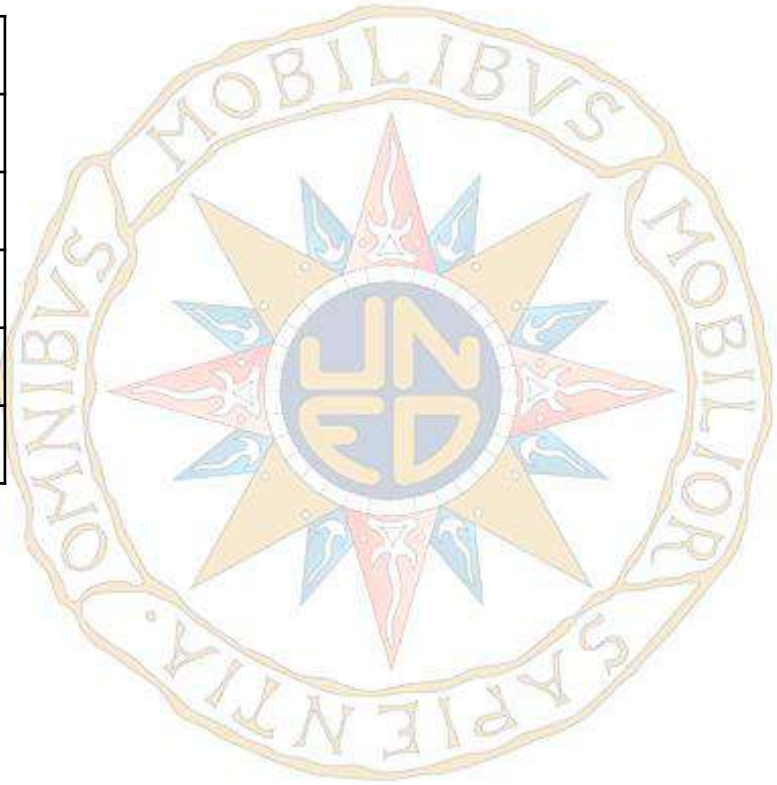
Junio de 2014, 2ª semana

14.- A continuación se presenta una escala de entrelazamiento obtenida al aplicar un cuestionario de 6 elementos a una muestra de 3 sujetos: 1_3_4_A_5_C_6_B_2. El valor escalar de los sujetos C y B son respectivamente: a) 4 y 5; b) 5 y 6; c) 6 y 2.

EJEMPLO 6

En la tabla se muestran las respuestas de 5 sujetos a 4 ítems. Comprobar si los datos se ajustan al modelo de Guttman y, si es así, hallar la escala de entrelazamiento.

	A	B	C	D
1	0	1	0	1
2	0	1	0	0
3	0	1	1	1
4	0	1	0	1
5	1	1	1	1



FIN DEL TEMA 3

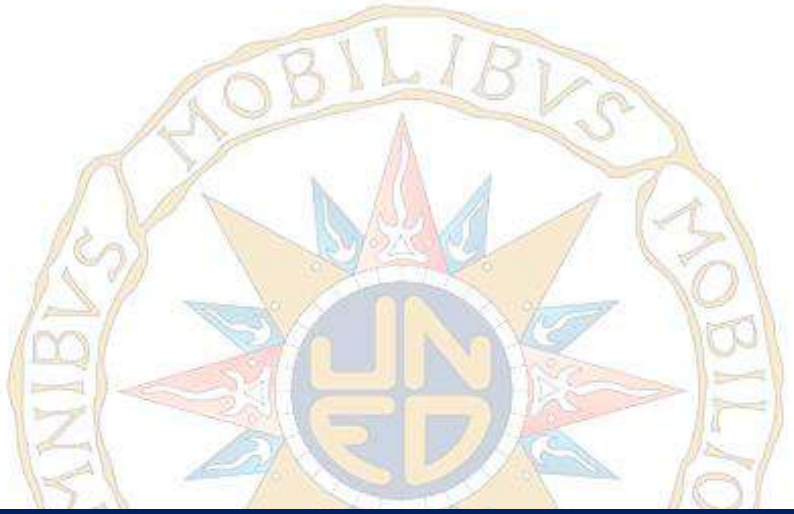
Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 4

La fiabilidad de las puntuaciones

Profesor D. Joaquín Antón Duce



EL MODELO LINEAL DE SPEARMAN

DEFINICIÓN Y SUPUESTOS

DEFINICIÓN

La puntuación empírica obtenida por un sujeto en un test (X) es suma de la puntuación verdadera (V) de ese sujeto y el error de medida (E) que la afecta.

$$X = V + E$$

SUPUESTOS

1. La puntuación verdadera (V) es la esperanza matemática de la empírica (X).

$$V = E(X)$$

2. La correlación existente entre las puntuaciones verdaderas de «n» sujetos en un test y los errores de medida es igual a cero.

$$r_{ve} = 0$$

3. La correlación entre los errores de medida $r_{e_1 e_2}$ que afectan a las puntuaciones de los sujetos en dos tests diferentes (X1 y X2) es igual a cero.

$$r_{e_1 e_2} = 0$$

DEDUCCIONES DEL MODELO (I)

1. El error de medida es la diferencia entre la puntuación empírica y la verdadera.

$$E = X - V$$

2. La esperanza matemática de los errores de medida es cero.

$$E(e) = 0$$

3. La media de las puntuaciones empíricas es igual a la media de las puntuaciones verdaderas.

$$\bar{X} = \bar{V}$$

4. La covarianza entre las puntuaciones verdaderas y los errores es igual a cero.

$$Cov(V,E)=0$$

DEDUCCIONES DEL MODELO (II)

5. La varianza de las puntuaciones empíricas es igual a la suma de la varianza de las puntuaciones verdaderas más la varianza de los errores.

$$S_x^2 = S_v^2 + S_e^2$$

6. La covarianza entre las puntuaciones empíricas y las verdaderas es igual a la varianza de las puntuaciones verdaderas.

$$Cov(X, V) = S_v^2$$

7. La correlación entre las puntuaciones empíricas y los errores es igual al cociente entre la desviación típica de los errores y la desviación típica de las puntuaciones empíricas.

$$r_{xe} = \frac{S_e}{S_x}$$

8. La covarianza entre las puntuaciones empíricas de dos tests es igual a la covarianza entre las puntuaciones verdaderas.

$$Cov(X_1, X_2) = Cov(V_1, V_2)$$

TEST PARALELOS. DEFINICIÓN

Si a una misma muestra de sujetos se le aplican dos tests, *X* y *X′*, podemos considerar que son paralelos si, además de cumplirse los supuestos anteriores, se cumplen las dos condiciones siguientes:

1. Las puntuaciones verdaderas de los sujetos son iguales en ambos tests.

$$V = V'$$

2. La varianza de los errores de medida es la misma en ambos tests.

$$S_e^2 = S_{e'}^2$$

TEST PARALELOS. DEDUCCIONES

1. La media de las puntuaciones empíricas obtenidas en dos tests supuestamente paralelos es la misma.

$$\overline{X} = \overline{X}'$$

2. Las varianzas de las puntuaciones empíricas obtenidas en dos tests paralelos son iguales.

$$S_x^2 = S_{x'}^2$$

3. La correlación entre las puntuaciones empíricas obtenidas en dos tests paralelos ($r_{xx'}$) es igual al cuadrado de la correlación entre las puntuaciones empíricas y las puntuaciones verdaderas (r_{xv})o bien, al cociente entre la varianza de las puntuaciones verdaderas y la varianza de las puntuaciones empíricas.

$$r_{xx'} = r_{xv}^2 = \frac{S_v^2}{S_x^2}$$

4. Dados dos o más tests paralelos, las intercorrelaciones entre cada dos de ellos son iguales.

$$r_{x_1x_2} = r_{x_1x_3} = r_{x_2x_3} = \cdots r_{x_jx_k}$$

EL COEFICIENTE DE FIABILIDAD

INTERPRETACIÓN TEÓRICA

Definimos el **coeficiente de fiabilidad** de un test, como:

... la correlación entre las puntuaciones empíricas obtenidas por una muestra de sujetos en dos formas paralelas del test.

Se puede expresar también como el cociente entre la varianza de las puntuaciones verdaderas y la varianza de las puntuaciones empíricas.

$$r_{xx'} = \frac{S_v^2}{S_x^2}$$

Índice de fiabilidad (r_{xv}). Es la raíz cuadrada del coeficiente de fiabilidad

$$r_{xv} = \sqrt{r_{xx'}}$$

Otras relaciones importantes:

$$r_{xx'} = 1 - \frac{S_e^2}{S_x^2} = 1 - r_{xe}^2 \qquad 0 \leq r_{xx'} \leq 1$$

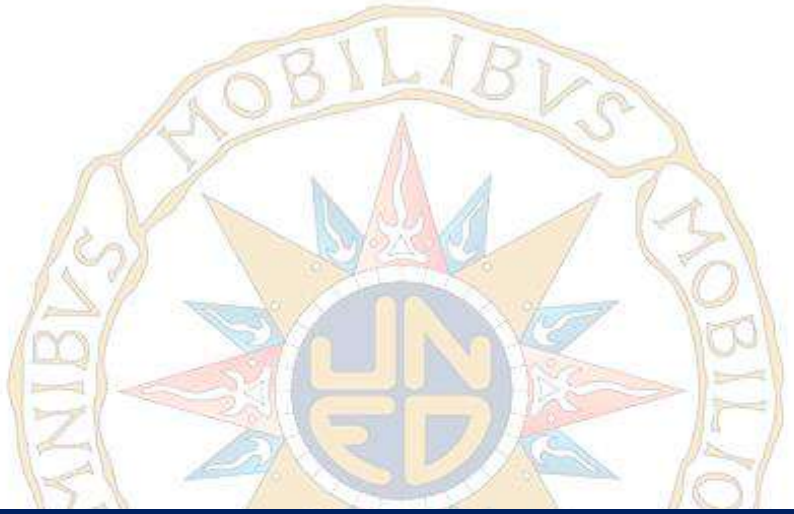
$$r_{xe} = \frac{S_e}{S_x} \Rightarrow r_{xe} = \sqrt{1 - r_{xx'}}$$

EJEMPLO 1

Junio 2016, 2ª semana

12. Sabiendo que en una muestra de sujetos la varianza de los errores de medida es la mitad de la varianza de sus puntuaciones verdaderas, el coeficiente de fiabilidad del test es: a) 0.50; b) 0.70; c) 0.67

13. La desviación típica de los errores de medida de un test es igual a 4, lo que implica el 20% de la varianza de las puntuaciones verdaderas. El coeficiente de fiabilidad del test es igual a: a) 0.56; b) 0.62; c) 0.74



TIPOS DE ERRORES

ERROR DE MEDIDA

DEFINICIÓN

El error de medida es la diferencia entre la puntuación empírica de un sujeto y su puntuación verdadera.

$$E = X - V$$

ERROR TÍPICO DE MEDIDA

Es la desviación típica de los errores de medida.

$S_e = S_x\sqrt{1 - r_{xx'}}$ en puntuaciones directas y diferenciales

$S_{Z_e} = \sqrt{1 - r_{xx'}}$ en puntuaciones típicas

ERROR DE ESTIMACIÓN DE LA PUNTUACIÓN VERDADERA

DEFINICIÓN

El error de estimación de la puntuación verdadera es la diferencia entre la puntuación verdadera de un sujeto y la puntuación verdadera pronosticada mediante el modelo de regresión.

$$E = V - V'$$

ERROR TÍPICO DE ESTIMACIÓN DE LA PUNTUACIÓN VERDADERA

Es la desviación típica de los errores de estimación.

$S_{v.x} = S_e\sqrt{r_{xx'}} = S_x\sqrt{1 - r_{xx'}}\sqrt{r_{xx'}}$ en puntuaciones directas y diferenciales

$S_{Z_v.Z_x} = S_{Z_e}\sqrt{r_{xx'}} = \sqrt{1 - r_{xx'}}\sqrt{r_{xx'}}$ en puntuaciones típicas

ERROR DE SUSTITUCIÓN

DEFINICIÓN

El error de sustitución es la diferencia entre las puntuaciones obtenidas por un sujeto en un test y las obtenidas en otro test paralelo. Es decir, el error que se cometería al sustituir las puntuaciones obtenidas en un test X_1 por las obtenidas en un test paralelo X_2.

$$e = X_1 - X_2$$

ERROR TÍPICO DE SUSTITUCIÓN

Es la desviación típica de los errores de sustitución.

$S_{x_1-x_2} = S_x\sqrt{1-r_{xx'}}\sqrt{2} = S_e\sqrt{2}$ en puntuaciones directas y diferenciales

$S_{z_1-z_2} = \sqrt{1-r_{xx'}}\sqrt{2} = S_{z_e}\sqrt{2}$ en puntuaciones típicas

ERROR DE PREDICCIÓN

DEFINICIÓN

El error de predicción es la diferencia entre las puntuaciones obtenidas por un sujeto en un test (X_1) y las puntuaciones pronosticadas en ese mismo test (X'_1) a partir de una forma paralela X_2.

$$e = X_1 - X'_1$$

ERROR TÍPICO DE PREDICCIÓN

Es la desviación típica de los errores de predicción.

$S_{e_p} = S_x\sqrt{1 - r_{xx'}}\sqrt{1 + r_{xx'}} = S_e\sqrt{1 + r_{xx'}}$ en punt. directas y diferenciales

$S_{Z_{ep}} = \sqrt{1 - r_{xx'}}\sqrt{1 + r_{xx'}} = S_{Z_e}\sqrt{1 + r_{xx'}}$ en puntuaciones típicas

FACTORES QUE AFECTAN A LA FIABILIDAD

LONGITUD DEL TEST

A mayor longitud, mayor fiabilidad. La relación existente entre la fiabilidad de un test y su longitud, siempre y cuando los ítems a añadir sean paralelos a los que ya tenía el test original, se puede evaluar mediante la ecuación de Spearman-Brown.

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

donde:

$R_{xx'}$ =coeficiente de fiabilidad del test alargado o acortado.

$r_{xx'}$ =coeficiente de fiabilidad del test inicial.

n = número de veces que se ha alargado o acortado el test.

$n = \frac{EF}{EI}$, siendo EF el número de elementos finales del test y EI el número de elementos iniciales del test.

En el caso de longitud doble (n=2), la fórmula quedaría así:

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

EJEMPLO 2

Septiembre 2016, reserva

Con los datos que se presentan a continuación responder a las preguntas 6 y 7

Un test de 100 ítems se ha aplicado a una muestra de sujetos obteniéndose un coeficiente de fiabilidad de 0.95.

6. ¿Cuál sería el coeficiente de fiabilidad del test si se duplicara su longitud?: a) 0.97; b) 0.86; c) 0.91.

7. Si fuera suficiente un test con un coeficiente de fiabilidad de 0.90, ¿Cuántos elementos debería tener?: a) 60; b) 54; c) 47

EJEMPLO 3

Septiembre2018, tipo B

17. Un test formado por 10 ítems presenta un coeficiente de fiabilidad de 0,50. ¿Cuántos ítems paralelos a los que tiene hay que añadir para alcanzar un coeficiente de fiabilidad de 0,80? a) 30; b) 40; c) 50.

VARIABILIDAD DE LA MUESTRA

A mayor variabilidad de la muestra, mayor fiabilidad. Partimos del supuesto de que el error típico de medida de un test se mantiene constante, independientemente de la variabilidad del grupo en que se aplique.

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

donde:

S_1^2 = varianza empírica de las puntuaciones en el grupo 1.

S_2^2 = varianza empírica de las puntuaciones en el grupo 2.

r_{11} = coeficiente de fiabilidad en el grupo 1.

r_{22} = coeficiente de fiabilidad en el grupo 2.

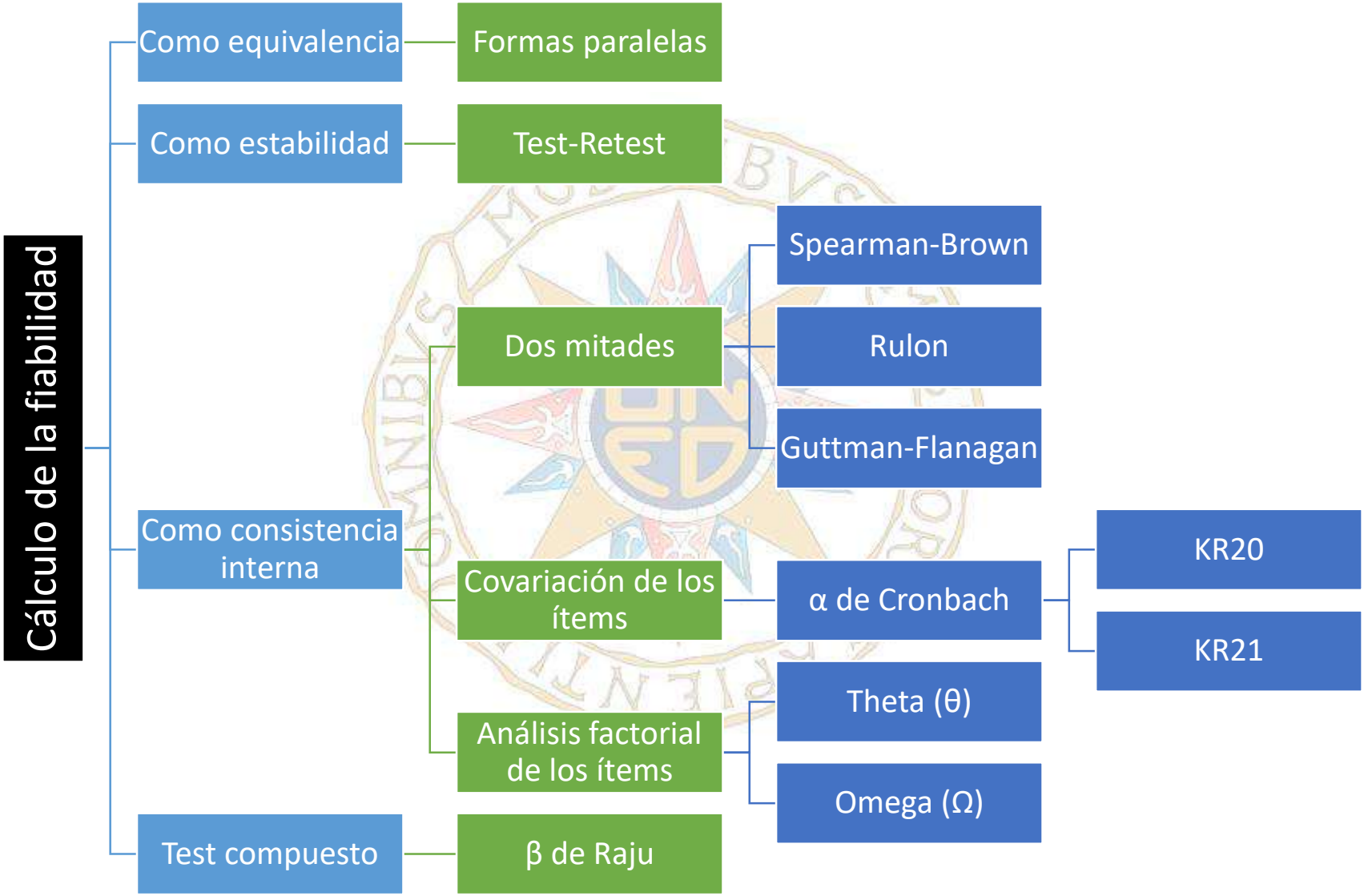
EJEMPLO 4

Septiembre 2019, tipo A

15. Tras aplicar un test a una muestra de sujetos se ha obtenido un coeficiente de fiabilidad de 0,76. ¿Cuál sería el índice de fiabilidad del test si se hubiera aplicado en una muestra con el triple de varianza?: a) 0,76; b) 0,92; c) 0,96.

CÁLCULO DEL COEFICIENTE DE FIABILIDAD

ESQUEMA


Cálculo de la fiabilidad
Como equivalencia
Formas paralelas
Como estabilidad
Test-Retest
Como consistencia interna
Dos mitades
Spearman-Brown
Rulon
Guttman-Flanagan
Covariación de los ítems
α de Cronbach
KR20
KR21
Análisis factorial de los ítems
Theta (θ)
Omega (Ω)
Test compuesto
β de Raju

FORMAS PARALELAS

Si disponemos de dos formas paralelas de un test, X_1 y X_2, se aplican los dos tests **una vez cada uno** y la fiabilidad de ambos se calcula mediante el coeficiente de correlación de Pearson entre las puntuaciones de los sujetos en ambas formas..

$$r_{xx'} = r_{x_1x_2} = \frac{N\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{[N\sum X_1^2 - (\sum X_1)^2][N\sum X_2^2 - (\sum X_2)^2]}}$$

El coeficiente de fiabilidad así obtenido recibe también el nombre de **coeficiente de equivalencia,** haciendo referencia al grado en que ambas formas son equivalentes.

TEST-RETEST

Con este método se aplica el mismo test en dos ocasiones diferentes a una misma muestra de sujetos. Calculamos el coeficiente de fiabilidad mediante la correlación entre las puntuaciones obtenidas por los sujetos en ambas aplicaciones.

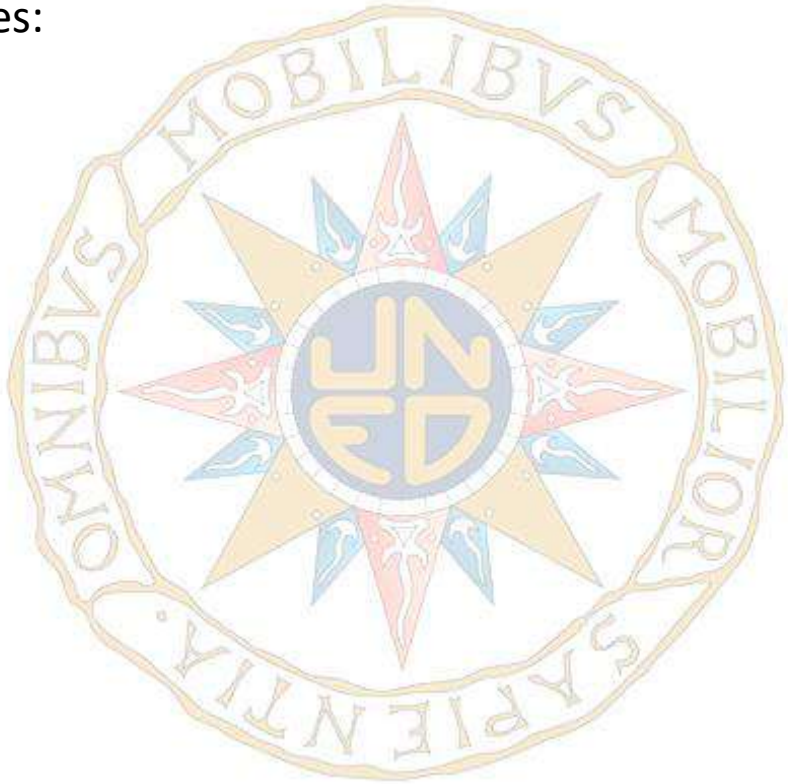
$$r_{xx'} = r_{x_1x_2} = \frac{N\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{[N\sum X_1^2 - (\sum X_1)^2][N\sum X_2^2 - (\sum X_2)^2]}}$$

El coeficiente de fiabilidad así obtenido recibe también el nombre de **coeficiente de estabilidad**.

DOS MITADES

Se aplica un test **una sóla vez** dividiendo el resultado en dos mitades, normalmente ítems pares e impares. Después, se calcula la fiabilidad por cualquiera de los tres procedimientos siguientes:

- Spearman-Brown
- Rulon
- Guttman-Flanagan



SPEARMAN-BROWN

Se calcula la correlación entre la mitad par y la mitad impar del test (r_{pi}), que han de ser paralelas. Después se aplica la formula de Spearman Brown para el caso de longitud doble:

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

Para aplicar este método de forma estricta habría que comprobar el paralelismo de las dos mitades (igualdad de medias e igualdad de los errores típicos de medida).

EJEMPLO 5

Junio de 2016, 1ª semana

19. En la tabla que se presenta a continuación aparecen recogidas las puntuaciones obtenidas en los elementos pares (P) e impares (I) de un test de razonamiento matemático por 4 sujetos de una muestra (se trata de un ejemplo). Utilizando el método que consideren más adecuado, el coeficiente de fiabilidad del test será: a) 0.86; b) 0.76; c) 0.84

Sujetos	P	I
A	4	3
B	8	9
C	3	6
D	6	7

RULON

Se utiliza cuando, aún no siendo las dos mitades definidas estrictamente paralelas, podemos considerarlas τ-equivalentes (tau-equivalentes) o esencialmente τ-equivalentes.

$$r_{xx'} = 1 - \frac{S_d^2}{S_x^2} = 1 - \frac{S_{p-i}^2}{S_x^2}$$

Lord y Novick (1968) definen los tests τ-equivalentes como aquellos en los que las puntuaciones verdaderas de los sujetos de una muestra son iguales en ambas formas, pero las varianzas de error no tienen porqué ser iguales, y definen los tests esencialmente τ-equivalentes como aquellos en los que la puntuación verdadera de cada sujeto en uno de los tests es igual a la del otro más una constante. Tanto en una situación como en otra se asume el cumplimiento del supuesto de igualdad de las varianzas verdaderas de ambas mitades.

Equivalentes	τ-equivalentes	esencialmente τ-equivalentes
$V_1=V_2$	$V_1=V_2$	$V_1=V_2+k$
$S_{e_1}^2 = S_{e_2}^2$	$S_{e_1}^2 \neq S_{e_2}^2$	$S_{e_1}^2 \neq S_{e_2}^2$

EJEMPLO 6

Septiembre de 2016, reserva

15. La tabla adjunta recoge las respuestas dadas por 5 sujetos a 6 ítems (1 indica acierto y 0 error). El coeficiente de fiabilidad según Rulon es: a) 0,53; b) 0,61; c) 0,79.

	1	2	3	4	5	6
A	1	0	0	0	1	0
B	1	1	1	0	1	1
C	1	1	1	1	1	0
D	0	1	1	0	1	1
E	1	0	1	1	1	0

GUTTMAN-FLANAGAN

Flanagan (1937) y Guttman (1945), de forma independiente llegaron a una fórmula equivalente a la de Rulon y que presenta una mayor sencillez de aplicación

$$r_{xx'} = 2\left(1 - \frac{S_p^2 + S_i^2}{S_x^2}\right)$$

COEFICIENTE α DE CRONBACH (I)

El coeficiente de Cronbach (1951) constituye un indicador de la consistencia interna del test. Este coeficiente expresa la fiabilidad del test en función del número de ítems y de la proporción de la varianza total del test debida a la covariación entre los ítems. Cuanto más covaríen los ítems entre sí mayor será la fiabilidad del test.

$$\alpha = \frac{n}{n-1}\left(\frac{\sum\sum Cov(j,k)}{S_x^2}\right) = \frac{n}{n-1}\left(1 - \frac{\sum S_j^2}{S_x^2}\right) = \frac{n}{n-1}\left(\frac{S_x^2 - \sum S_j^2}{S_x^2}\right) = \frac{nr_1}{1+(n-1)r_1}$$

Donde

n → número de elementos del test

$\sum S_j^2$ → suma de las varianzas de los elementos del test

$\sum\sum Cov(j,k)$ → suma de las covarianzas de los ítems

S_x^2 → varianza de las puntuaciones en el test

r_1 → Cociente entre la covarianza media de los ítems y su varianza media

COEFICIENTE α DE CRONBACH (II)

ESTIMADOR INSESGADO DE α

El estimador insesgado de a propuesto por Feldt, Woodruff y Salih (1987) se expresa como:

$$\overline{\alpha} = \frac{(N-3)\widehat{\alpha}+2}{N-1}$$

Donde

$\overline{\alpha}$ → estimador insesgado

$\widehat{\alpha}$ → valor de α de Cronbach

N → Número de sujetos de la muestra

COEFICIENTE α DE CRONBACH (III)

EL COEFICIENTE α COMO LÍMITE INFERIOR DEL COEFICIENTE DE FIABILIDAD

El coeficiente a puede ser considerado como una estimación del límite inferior del coeficiente de fiabilidad de un test, siendo su valor menor o igual que el coeficiente de correlación $r_{xx'}$ (Guttman, 1945):

$$\alpha \leq r_{xx'}$$

Otro estimador del límite inferior del coeficiente de fiabilidad es el coeficiente δ_3 (delta) propuesto por Guttman (1945):

$$\delta_3 = 1 - \left(\sum \frac{S_j^2}{S_x^2} \right) + \frac{\sqrt{\frac{n}{n-1} \sum \sum Cov(j,k)}}{S_x^2}$$

INFERENCIAS SOBRE α (I)

INFERENCIAS PARA UN SOLO VALOR DE α

Hipótesis

$H_0: \alpha = \alpha_0$

$H_1: \alpha \neq \alpha_0$

Estadístico

$$F = \frac{1 - \alpha_0}{1 - \hat{\alpha}}$$

Donde F se distribuye según F con (N-1) y (n-1)(N-1) grados de libertad

Siendo N el tamaño de la muestra y n el número de ítems

Intervalo de confianza

$$1\text{-}F_{0,975;(N-1),(n-1)(N-1)} \cdot (1 - \hat{\alpha}) \leq \alpha \leq 1\text{-}F_{0,025;(N-1),(n-1)(N-1)} \cdot (1 - \hat{\alpha})$$

EJEMPLO 7

Junio de 2016, primera semana

9. Se ha aplicado un test de inteligencia general formado por 26 ítems a una muestra de 61 adolescentes y se ha obtenido un coeficiente alfa de Cronbach de 0.78. Utilizando el NC del 95% comprobar si el coeficiente obtenido es estadísticamente significativo y establecer el intervalo confidencial en el que se encontrará el coeficiente alfa en la población: a) No es significativo y el intervalo es 0.60 y 0.85; b) Si es significativo y el intervalo es 0.60 y 0.79; c) Si es significativo y el intervalo es 0.69 y 0.85.

EJEMPLO 8

Septiembre de 2015, reserva

15.- Hemos aplicado un test de 16 ítems a una muestra de 61 sujetos, obteniendo un coeficiente de fiabilidad igual a 0,55. Decir si son compatibles estos datos con la hipótesis de que el valor de alpha en la población sea igual a 0.65 y establecer el intervalo confidencial (N.C. 95%): a) Sí, $0.67<0.78<1.39$; b) Sí, $0.69<0.78<1.28$; c) No, $0.65<1.28<1.23$.

INFERENCIAS SOBRE α (II)

INFERENCIAS SOBRE α PARA MUESTRAS INDEPENDIENTES

Hipótesis

$$H_0: \alpha_1 = \alpha_2$$

$$H_1: \alpha_1 \neq \alpha_2$$

Estadístico

$$W = \frac{1 - \hat{\alpha}_1}{1 - \hat{\alpha}_2}$$

Donde F se distribuye según F con $(N_1\text{-}1)$ y $(N_2\text{-}1)$ grados de libertad

INFERENCIAS SOBRE α (III)

INFERENCIAS SOBRE α PARA MUESTRAS DEPENDIENTES

Hipótesis

$$H_0: \alpha_1 = \alpha_2$$

$$H_1: \alpha_1 \neq \alpha_2$$

Estadístico

$$t = \frac{|\hat{\alpha}_1 - \hat{\alpha}_2|\sqrt{N-2}}{\sqrt{\left[4(1-\hat{\alpha}_1)(1-\hat{\alpha}_2)\left(1-r^2_{x_1x_2}\right)\right]}}$$

Donde t se distribuye según t de Student con N-2 grados de libertad

EJEMPLO 9

Septiembre 2015

20. Sabiendo que la correlación entre las puntuaciones obtenidas por los sujetos en el test original (con coeficiente de fiabilidad de 0,23) y en el duplicado (con coeficiente de fiabilidad de 0,37) es 0,60. Serían significativas las diferencias encontradas entre ambos coeficientes? (NC 95%): a) Si porque el valor de chi cuadrado obtenido es mayor que el de tablas; b) no porque la puntuación típica obtenida es mayor que la de las tablas; c) no porque el valor de t obtenido es menor que el de las tablas.

CASOS PARTICULARES DE α. KR_{20}

Las ecuaciones de Kuder-Richardson (1937) representan un caso particular del coeficiente «alpha» de Cronbach, en el supuesto de que los ítems sean dicotómicos.

$$KR_{20} = \frac{n}{n-1}\left(1 - \frac{\sum p_j q_j}{S_x^2}\right)$$

n = número de elementos del test.

p_j =proporción de aciertos en el elemento j.

q_j = proporción de errores en el elemento j.

$P_j q_j$ =varianza del elemento j.

S_x^2 = varianza total del test.

EJEMPLO 10

Junio 2015, segunda semana

Con estos datos contestar a las preguntas 23, 24 y 25. En la tabla adjunta se muestran las puntuaciones obtenidas por 5 estudiantes en un test de 5 ítems dicotómicos (se trata de un ejemplo).

Sujetos	**1**	**2**	**3**	**4**	**5**
A	0	1	1	1	1
B	1	1	0	1	0
C	1	1	0	0	0
D	1	1	1	1	0
E	1	0	0	0	0

23. La media del ítem 1 es : a) 0.80; b) 0.60; c) 0,50

24. La varianza del ítem 5 es: a) 0.21; b) 0.16; c) 0.25

25. El coeficiente α de Cronbach es: a) 0.49; b) 0.36; c) 0.69

CASOS PARTICULARES DE α. KR_{21}

Si los ítems que componen el test, además de ser dicotómicos, presentan la misma dificultad, podemos aplicar la ecuación de Kuder-Richardson 21 (KR_{21})

$$KR_{21} = \frac{n}{n-1}\left(1 - \frac{npq}{S_x^2}\right)$$

n = número de elementos del test.

npq = suma de las varianzas de los elementos. Al ser iguales las varianzas se sustituye el signo sumatorio por *«n»* veces la misma varianza.

S_x^2 = varianza total del test.

COEFICIENTES THETA (θ) Y OMEGA (Ω)

Los coeficientes Theta (θ) de Carmines (Carmines y Zeller, 1979) y Omega (Ω) de Heise y Bohrnstedt (1970) constituyen dos indicadores de la consistencia interna de los ítems de un test y una aproximación al coeficiente alpha. Se trata de dos coeficientes basados en el análisis factorial de los ítems.

$$\theta = \frac{n}{n-1}\left(1 - \frac{1}{\lambda_1}\right)$$

n = número de ítems del test.

λ_1 =primer autovalor de la matriz factorial; es decir, la varianza explicada por el primer factor antes de la rotación.

El coeficiente θ es además un indicador de la unidimensionalidad de los ítems.

En general, y para los mismos datos, se verifica que $\alpha \leq \theta \leq \Omega$. La igualdad entre los coeficientes se verifica cuando los ítems son paralelos.

EJEMPLO 11

Junio 2017, segunda semana

13.- Tras someter a un análisis factorial las puntuaciones obtenidas por una muestra de sujetos en un test de 25 ítems se obtuvieron 3 factores. La varianza explicada por el primer factor antes de la rotación es 8.50. El coeficiente Theta (θ) de Carmines y Zeller será: a) 0.92; b) 0.89; c) 0.97.

EL COEFICIENTE β DE RAJU

En el caso de que un test se divida en varios subtests, con desigual número de ítems, y se quiera estimar la consistencia interna del test total a partir de las puntuaciones totales de los sujetos en los subtests, el coeficiente alfa presenta el problema de que proporciona un valor infraestimado de la fiabilidad. El coeficiente β propuesto por Raju (Raju, 1977) permite superar este problema. Se aplica este coeficiente cuando se desconocen las puntuaciones de los sujetos en los ítems de los distintos subtests. En el caso de conocer los valores de estas puntuaciones es mejor emplear el coeficiente α.

$$\beta = \frac{S_x^2 - \sum S_j^2}{S_x^2\left(1 - \sum\left(\frac{n_j}{N}\right)^2\right)}$$

k = número de subtests.

S_x^2 = varianza del test.

S_j = varianza de cada subtest.

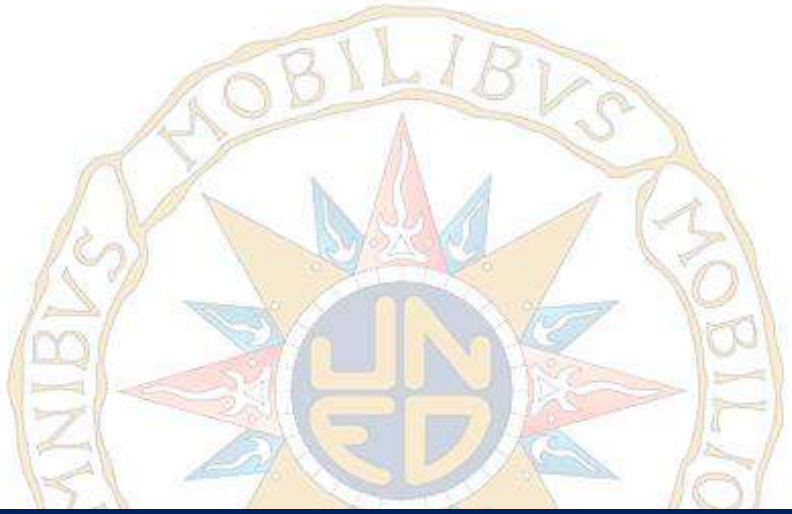
n_j = número de ítems en cada subtest.

N= número de ítems total de la batería.

EJEMPLO 12

Septiembre 2017, tipo A

15.- Se ha aplicado un test de razonamiento verbal compuesto de tres subtests a una muestra de 150 niños de Bachillerato. El subtest A está compuesto de 30 ítems y la varianza de las puntuaciones obtenidas por los sujetos fue de 6 puntos; el subtest B estaba formado por 25 ítems y su varianza fue de 5 y el subtest C tenía 20 ítems y su varianza fue de 4 puntos. La varianza total del test fue de 25 puntos. La consistencia interna del test total mediante el coeficiente de Raju será: a) 0.57; b) 0.65; c) 0.61.



ESTIMACIÓN DE LA PUNTUACIÓN VERDADERA

DESIGUALDAD DE CHEBYSHEV (Чебышёб)

Si no se hace ningún supuesto sobre la distribución de las puntuaciones empíricas o de los errores, se aplica la desigualdad de Chebyshev:

$$\forall K, P\{|X - V| \leq K(S_e)\} \geq 1 - \frac{1}{K^2}$$

donde:

K → constante que toma valores positivos

$1 - \frac{1}{K^2}$ → nivel de confianza utilizado.

S_e → error típico de medida.

EJEMPLO 13

Septiembre de 2017, reserva

15.- Se ha administrado un test de Razonamiento Numérico a una muestra de 300 alumnos de un Colegio. La media y la desviación típica de las puntuaciones obtenidas fueron 42 y 4 puntos respectivamente. Sabiendo que la proporción de varianza verdadera que hay en la varianza empírica del test es de 0.64, a un NC del 99% y utilizando la desigualdad de Chevyschef, el intervalo confidencial en el que se encontrará la puntuación verdadera de un sujeto que obtuvo una puntuación empírica de 50 puntos será:

a) $26 \leq V \leq 74$; b) $39.2 \leq V \leq 89.8$; c) $30.1 \leq V \leq 70.9$

DISTRIBUCIÓN NORMAL DE LOS ERRORES

Este método asume una distribución normal de los errores de medida (con media 0 *y* varianza S_e^2*)* y de las puntuaciones empíricas condicionadas a un determinado valor de V.

Directas $IC = X \pm E_{max}$

$Z_c . S_e$

$S_x\sqrt{1 - r_{xx'}}$

Diferenciales $IC = x \pm E_{max}$

$Z_c . S_e$

$S_x\sqrt{1 - r_{xx'}}$

Típicas $IC = Z_x \pm e_{max}$

$Z_c . S_{z_e}$

$\sqrt{1 - r_{xx'}}$

EJEMPLO 14

Septiembre de 2016, reserva

Con los datos que se muestran a continuación responder a las preguntas 19 y 20
Se ha administrado un test de percepción visual, cuya fiabilidad es cero, a una muestra de 200 sujetos y se obtiene una media de 15 y una varianza igual a 8.

19**.** Utilizando el modelo de la distribución normal de los errores, calcular el intervalo confidencial en el cual se encontraría la puntuación directa verdadera de un sujeto que en el test obtuvo una puntuación típica igual a 2.5. (NC, 95%): a) 14.53 -19.97; b) 16.53 - 27.63; c) 17.31- 32.92.

MÉTODO DE REGRESIÓN (I)

Las puntuaciones empíricas son siempre sesgadas y, por lo tanto, es más conveniente establecer el intervalo confidencial no a partir de las puntuaciones empíricas (que son sesgadas) sino a partir de la puntuación verdadera estimada, que podremos calcular mediante el modelo de regresión lineal

Directas

$$IC = V' \pm E_{max}$$

$$E_{max} = Z_c . S_{v.x}$$

$$S_{v.x} = S_x \sqrt{1 - r_{xx'}} \sqrt{r_{xx'}}$$

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

EJEMPLO 15

Junio de 2019, 2ª semana

14. Un test presenta una media de las puntuaciones verdaderas igual a 5, y una varianza de 2 puntos. Sabiendo que el coeficiente de fiabilidad es 0,70, estimar la puntuación verdadera, basada en el modelo de regresión, en puntuaciones directas de un sujeto que ha obtenido una puntuación empírica de 4: a) 3,9; b) 4,3; c) 5,6.

MÉTODO DE REGRESIÓN (II)

Diferenciales

$IC = v' \pm E_{max}$

$Z_c.S_{v.x}$

$S_x\sqrt{1 - r_{xx'}}\sqrt{r_{xx'}}$

$v' = r_{xx'}(X - \bar{X}) = r_{xx'} \cdot x$

Típicas

$IC = Z_{v'} \pm e_{max}$

$Z_c \cdot S_{Z_v Z_x}$

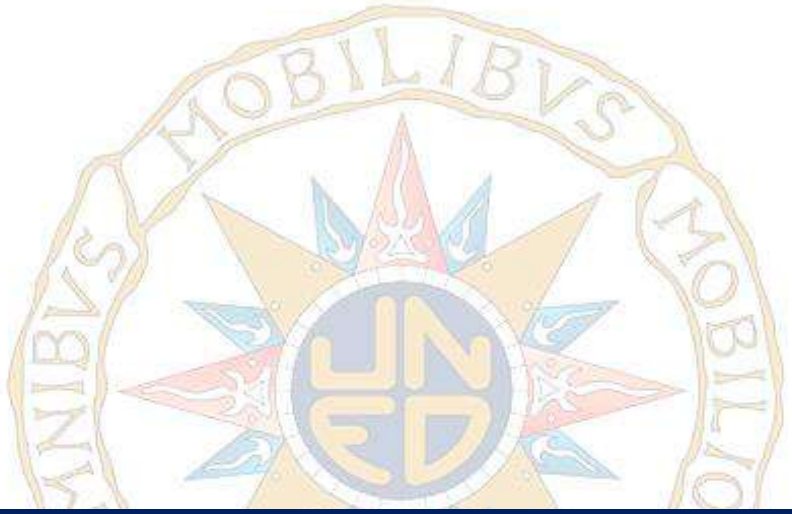
$\sqrt{1 - r_{xx'}}\sqrt{r_{xx'}}$

$Z_{v'} = r_{vx}.Z_x$

EJEMPLO 16

Septiembre de 2017, tipo A

18.- La correlación entre las puntuaciones empíricas obtenidas por una muestra de sujetos en un test y los errores de medida es 0,40. La desviación típica fue de 4 puntos. Si un sujeto ha obtenido en el test una puntuación diferencial igual a 4, el intervalo confidencia en el que se encontrará su puntuación diferencial verdadera será (NC 95%): a) 6,25 y 1,24; b) 6,50 y 0,22; c) 8,21 y 2,24.



FIABILIDAD DE UNA BATERÍS DE TESTS

FIABILIDAD DE UNA BATERÍA DE TESTS

Se trata de calcular la fiabilidad de la batería en función de los coeficientes de fiabilidad, varianzas y covarianzas de los subtests que la van a conformar.

La fórmula a utilizar en este caso será:

$$r_{tt} = 1 - \frac{\sum S_j^2 - \sum S_j^2 r_{jj}}{S_T^2}$$

Siendo:

S_j = varianza del subtest j.

r_{jj} =coeficiente de fiabilidad del subtest j.

S_T^2 = varianza de la batería total.

FIN DEL TEMA 4

Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 6

Validez de las inferencias

DEFINICIÓN DE VALIDEZ

Aunque es difícil dar una definición concreta acerca del concepto de validez, convencionalmente y en relación con los tests, se acepta que el término hace referencia *al grado en que el test mide aquello que pretende medir.*

OTRAS VARIANTES DE VALIDEZ

- **Validez referida a criterio**
  - **Validez predictiva.** Un test es válido en la medida en que exista correlación entre las puntuaciones obtenidas por los sujetos en el test y las obtenidas en el criterio externo seleccionado.
  - **Validez concurrente.** Se diferencia de la validez predictiva en que la recogida de la información, tanto del test como del criterio, se hace simultáneamente.
  - **Validez retrospectiva.** Ocurre cuando se obtiene la medida del criterio con anterioridad a la del test.
- **Validez de contenido**. Trata de estudiar hasta qué punto, a partir del contenido de los tests, se puede inferir el rendimiento en una determinada materia.
- **Validez de constructo.** Este tipo de validez implica recoger toda la información necesaria para poder tener garantía suficiente de que las conductas observables que se han elegido como indicadores del constructo que se quiere medir, lo son realmente.

VALIDACIÓN DE CONTENIDO

VALIDEZ DE CONTENIDO

El objetivo que se persigue al llevar a cabo un estudio de validación de contenido es analizar hasta qué punto los elementos o ítems que componen el test son una muestra **relevante** y **representativa** del constructo sobre el que se van a realizar las inferencias.

- Relevancia implica la necesidad de una clara y exhaustiva especificación de todas las posibles conductas observables que son representativas del contructo a medir (especificación del dominio de conductas).
- Representatividad hace referencia a la necesidad de que todas esas conductas estén representadas en el test (representatividad del dominio).

VALIDACIÓN DE CONSTRUCTO

VALIDEZ DE CONSTRUCTO

Partiendo de que los tests son instrumentos que permiten describir de forma indirecta, u operativizar, el grado en que los sujetos poseen alguna característica postulada a nivel teórico denominada *constructo,* la validación de constructo será el proceso que permitirá obtener evidencia acerca de la capacidad del test para medirlo.

Entre los métodos más utilizados para llevar a cabo la validación del constructo hemos de destacar el método de la matriz multimétodo-multirrasgo y el análisis factorial.

MATRIZ MULTIMÉTODO-MULTIRRASGO

MATRIZ MULTIMÉTODO-MULTIRRASGO

		V-F			E-M			F-I		
		RN	FE	RA	RN	FE	RA	RN	FE	RA
V-F	RN	.95								
	FE	**.20**	.90							
	RA	**.30**	**.28**	.92						
E-M	RN	.90	.31	.40	.93					
	FE	.26	.87	.33	**.37**	.94				
	RA	.43	.20	.84	**.26**	**.37**	.88			
F-I	RN	.79	.27	.31	.77	.15	.23	.89		
	FE	.11	.68	.22	.24	.67	.31	**.19**	.93	
	RA	.19	.18	.50	.19	.33	.72	**.41**	**.30**	.64

Fiabilidad

V. convergente

V. discriminante

MATRIZ MULTIMÉTODO-MULTIRRASGO

MATRIZ MULTIMÉTODO-MULTIRRASGO

		V-F			E-M			F-I		
		RN	FE	RA	RN	FE	RA	RN	FE	RA
V-F	RN	.95								
	FE	**.20**	.90							
	RA	**.30**	**.28**	.92						
E-M	RN	.90	.31	.40	.93					
	FE	.26	.87	.33	**.37**	.94				
	RA	.43	.20	.84	**.26**	**.37**	.88			
F-I	RN	.79	.27	.31	.77	.15	.23	.89		
	FE	.11	.68	.22	.24	.67	.31	**.19**	.93	
	RA	.19	.18	.50	.19	.33	.72	**.41**	**.30**	.64

Fiabilidad

V. convergente

V. discriminante

Hay validez convergente si las correlaciones entre mismo rasgo por diferentes métodos son altas

MATRIZ MULTIMÉTODO-MULTIRRASGO

MATRIZ MULTIMÉTODO-MULTIRRASGO

		V-F			E-M			F-I		
		RN	FE	RA	RN	FE	RA	RN	FE	RA
V-F	RN	.95								
	FE	**.20**	.90							
	RA	**.30**	**.28**	.92						
E-M	RN	.90	.31	.40	.93					
	FE	.26	.87	.33	**.37**	.94				
	RA	.43	.20	.84	**.26**	**.37**	.88			
F-I	RN	.79	.27	.31	.77	.15	.23	.89		
	FE	.11	.68	.22	.24	.67	.31	**.19**	.93	
	RA	.19	.18	.50	.19	.33	.72	**.41**	**.30**	.64

Fiabilidad

V. convergente

V. discriminante

Hay validez discriminante si las correlaciones entre diferentes rasgos por el mismo método son bajas

EL ANÁLISIS FACTORIAL

Es quizás la técnica más utilizada, tanto en su vertiente exploratoria como confirmatoria, para poner a prueba las hipótesis planteadas acerca de la estructura interna del constructo y de las relaciones del mismo con otras variables.

Las medidas que proporcionan los tests pueden hacer referencia a variables unidimensionales o multidimensionales y, precisamente, el análisis factorial nos va a permitir descubrir la estructura que subyace a las puntuaciones obtenidas por los sujetos en los distintos ítems del test o en un conjunto de tests.

EL ANÁLISIS FACTORIAL

Variables	Factor 1	Factor 2
RN (V-F)	.86	—
RN (E-M)	.75	—
RN (F-I)	.92	—
FE (V-F)	—	.82
FE (E-M)	—	.74
FE (F-I)	—	.63
RA (V-F)	.42	.33
RA (E-M)	.51	—
RA (F-I)	—	.54

Este tipo de estudios de validación permiten obtener evidencia acerca del grado en que las puntuaciones obtenidas en el test pueden utilizarse eficazmente para hacer inferencias acerca del comportamiento real de los sujetos en un criterio

VALIDACIÓN REFERIDA AL CRITERIO

PROCEDIMIENTOS ESTADÍSTICOS

- **Un único test predictor y un solo indicador de criterio**

Los procedimientos más utilizados son la correlación y el modelo de regresión lineal simple.

- **Varios predictores y un solo indicador del criterio**

En este caso los procedimientos qué se utilizan son la correlación y la regresión lineal múltiple. Si el criterio es cualitativo, se suele utilizar otra técnica multivariante denominada análisis discriminante y cuándo se utilizan criterios dicotómicos la regresión logística.

- **Varios predictores cuantitativos y varios indicadores del criterio cuantitativos**

En este caso las técnicas más adecuadas son las regresión lineal multivariante y la correlación canónica.

- **Procedimientos basados en la teoría de la decisión: validez y utilidad en las decisiones**

Los procedimientos propuestos se basan en diferentes métodos para optimizar las decisiones realizadas con el test: técnicas maximin y minimax y especialmente la Teoría de la Utilidad Multiatributo

VALIDACIÓN CON UN ÚNICO PREDICTOR Y UN SOLO INDICADOR DEL CRITERIO

COEFICIENTE DE VALIDEZ (I)

Se define como la correlación entre las puntuaciones obtenidas por los sujetos en el test predictor y las obtenidas en el criterio.

El tipo de correlación utilizada para el cálculo del coeficiente de validez dependerá de la naturaleza de las variables implicadas.

INDICADOR CRITERIO	TEST		
	Continua	**Dicotomizada**	**Dicotómica**
Continua	Pearson	Biserial	Biserial puntual
Dicotomizada	Biserial	Tetracórica	$\Phi_{biserial}$
Dicotómica	Biserial puntual	$\Phi_{biserial}$	Φ

COEFICIENTE DE VALIDEZ (II)

Se define como la correlación entre las puntuaciones obtenidas por los sujetos en el test pred

El de validez dependerá de la nat

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

INDICADOR CRITERIO			
	Cont ua	Dicotomizada	Dicotómica
Continua	Pearson	Biserial	Biserial puntual
Dicotomizada	Biserial	Tetracórica	$\Phi_{biserial}$
Dicotómica	Bi ial puntual	$\Phi_{biserial}$	Φ

$$r_b = \frac{\bar{X}_p - \bar{X}_q}{S_x} \cdot \frac{pq}{y} = \frac{\bar{X}_p - \bar{X}}{S_x} \cdot \frac{p}{y}$$

COEFICIENTE DE VALIDEZ (II)

Se define como la correlación entre las puntuaciones obtenidas por los sujetos en el test predictor y

El tipo de idez dependerá de la naturaleza

$$r_{bp} = \frac{\bar{X}_p - \bar{X}_q}{S_x} \cdot \sqrt{pq} = \frac{\bar{X}_p - \bar{X}}{S_x} \cdot \sqrt{\frac{p}{q}}$$

INDICADOR CRITERIO	TEST		
	Continua	**Dicotomizada**	**Dicotómica**
Continua	Pearson	Biserial	Biserial puntual
Dicotomizada	Biserial	Tetracórica	$\Phi_{biserial}$
Dicotómica	Biserial puntual	$\Phi_{biserial}$	Φ

COEFICIENTE DE VALIDEZ (II)

Se define como la correlación entre las puntuaciones obtenidas por los sujetos en el test predictor y las obtenidas en el criterio.

El tipo de correlación utilizada para el cálculo del coeficiente de validez dependerá de la naturaleza de las variables implicadas.

INDICADOR CRITERIO	TEST		
	Continua	**Dicotomizada**	**Dicotómica**
Continua	Pearson	Biserial	Biserial puntual
Dicotomizada	Biserial	Tetracórica	$\Phi_{biserial}$
Dicotómica	Biserial puntual	$\Phi_{biserial}$	Φ

		X	
		0	1
Y	1	a	b
	0	c	d

$$\Phi = \frac{bc - ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

COEFICIENTE DE VALIDEZ (III)

test p ...sujetos en el

de la ...z dependerá

Su cálculo es muy laborioso, pero se han ofrecido algunas aproximaciones muy sencillas, la más utilizada es calcular la razón *bc/ad* y consultar la tabla correspondiente. Si la razón es menor que la unidad se debe usar la recíproca *ad/bc* para consultar la tabla, en este caso la correlación será negativa.

INDICADOR CRITERIO	TEST		
	Continua	**Dicotor...**	**Dicotómica**
Continua	Pearson	Bis...	Biserial puntual
Dicotomizada	Biserial	Tetracórica	$\Phi_{biserial}$
Dicotómica	Biserial puntual	$\Phi_{biserial}$	

		X	
		0	1
Y	1	a	b
	0	c	d

$$\Phi_{biserial} = \frac{bc - ad}{\sqrt{(a+b)(c+d)}} \cdot \frac{\sqrt{pq}}{y}$$

EL MODELO DE REGRESIÓN LINEAL

ECUACIONES DE REGRESIÓN

Mediante el modelo de regresión se intenta buscar una ecuación lineal que haga mínimos los errores de pronóstico.

$$Y = a + bX$$

$$b = r_{xy}\frac{S_y}{S_x}$$

$$a = \overline{Y} - b\overline{X}$$

Ecuaciones de regresión

Puntuaciones directas $$Y' = r_{xy}\frac{S_y}{S_x}(X - \overline{X}) + \overline{Y} = r_{xy}\frac{S_y}{S_x}X + \left(\overline{Y} - r_{xy}\frac{S_y}{S_x}\overline{X}\right)$$

Puntuaciones diferenciales $$y' = r_{xy}\frac{S_y}{S_x}(X - \overline{X}) = r_{xy}\frac{S_y}{S_x}x$$

Puntuaciones típicas $$Z'_y = r_{xy}Z_x$$

VARIANZA RESIDUAL Y ERROR TÍPICO

El error de estimación es la diferencia entre la puntuación de un sujeto en el criterio y la que se le pronostica mediante la ecuación de regresión.

$$E = Y - Y'$$

La varianza de todos los errores de estimación se denomina varianza residual, varianza error o error cuadrático medio.

$$S^2_{Y.X} = \frac{\sum(Y - Y')^2}{N} = S^2_Y\left(1 - r^2_{XY}\right)$$

y su desviación típica es el error típico de estimación

$S_{Y.X} = \sqrt{\frac{\sum(Y-Y')^2}{N}} = S_Y\sqrt{(1 - r^2_{XY})}$ en puntuaciones directas o diferenciales

$S_{Z_x.Z_y=}\sqrt{(1 - r^2_{XY})}$ en puntuaciones típicas

Además se cumple

$\overline{Y} = \overline{Y}'$ $\sum E = 0$ $S^2_Y = S^2_{Y'}+S^2_{Y.X}$

INTERVALOS DE CONFIANZA

Directas

$IC = Y' \pm E_{max}$

$E_{max} = Z_c . S_{Y.X}$

$S_{Y.X} = S_Y \sqrt{1 - r_{xy}^2}$

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

Diferenciales

$IC = y' \pm E_{max}$

$E_{max} = Z_c . S_{Y.X}$

$S_{Y.X} = S_Y \sqrt{1 - r_{xy}^2}$

$$y' = r_{xy} \frac{S_y}{S_x} (X - \overline{X})$$

Típicas

$IC = Z_{y'} \pm e_{max}$

$e_{max} = Z_c S_{Z_x Z_y}$

$S_{Z_x Z_y} = \sqrt{1 - r_{xy}^2}$

$$Z_{y'} = r_{xy} Z_x$$

EJEMPLO 1

Junio de 2017, 2ª semana

20.- A una muestra de 500 sujetos se les ha aplicado un test de percepción compuesto por 10 ítems. El coeficiente de fiabilidad del test fue 0,78 y su correlación con un criterio fue 0,70. La media y varianza del test fueron respectivamente de 6 y 4 puntos y las del criterio de 100 y 25 puntos. ¿Entre qué límites estará la puntuación directa de un sujeto en el criterio si en el test obtuvo una puntuación de 7 puntos? (NC 99%): a) 80, 25 y 110, 72; b) 115,25 y 141,32; c) 92,54 y 110,96.

EJEMPLO 2

Septiembre de 2018

Con los datos que se muestran a continuación responder a las preguntas 16 a 19

Aplicado un test a una muestra normativa se obtuvieron los siguientes resultados: La desviación típica de las puntuaciones fue de 5 puntos, la correlación obtenida respecto a las puntuaciones de un criterio externo fue 0.80; la media y la desviación típica de las puntuaciones obtenidas en el criterio fueron 10 y 3 respectivamente y los coeficientes de fiabilidad del test y del criterio fueron 0.90 y 0.80. Si un sujeto obtuvo en el test una puntuación diferencial igual a 4.

16. A un NC del 95%, y utilizando el modelo de regresión, el intervalo confidencial en el que se encontrará la puntuación verdadera diferencial del sujeto será: a) 2,94 y 5,54; b) 0, 66 y 3,30; c) 0,66 y 6,54

17. ¿Qué puntuación diferencial le pronosticaríamos en el criterio al sujeto que en el test obtuvo una puntuación diferencial de 4 puntos?: a) 3,34; b) 1,92; c) 2,58.

COEFICIENTES

Coeficiente de determinación. Representa la proporción (o el porcentaje) de la varianza de las puntuaciones de los sujetos en el criterio (variable dependiente) que se puede pronosticar a partir del test (variable predictora o independiente). También se define como la varianza común o asociada entre el test y el criterio.

$$CD = r_{XY}^2$$

Coeficiente de alienación. El coeficiente de alienación representa la inseguridad, o el azar, que afecta a los pronósticos.

$$CA = K = \frac{S_{Y.X}}{S_Y} = \sqrt{(1 - r_{XY}^2)}$$

Coeficiente de valor predictivo. Es el complementario del coeficiente de alienación y es otra forma de expresar la capacidad del test para pronosticar el criterio ya que representa la proporción (o el porcentaje si se multiplica por cien) de seguridad en los pronósticos.

$$CVP = 1 - \sqrt{(1 - r_{XY}^2)}$$

EJEMPLO 3

Septiembre de 2021. Ávex

Un psicólogo está interesado en estudiar la validez predictiva de un test de aptitudes. Dicho test consta de seis elementos dicotómicos y pretende predecir el rendimiento académico tal y como es evaluado por medio de las calificaciones escolares. En la siguiente tabla se muestran las puntuaciones de seis alumnos en cada uno de los elementos del test y las calificaciones otorgadas al final de curso

Suj.	X1	X2	X3	X4	X5	X6	Y
1	1	1	0	1	1	0	7
2	1	0	0	1	0	0	4
3	1	1	1	1	1	0	6
4	1	0	1	0	1	0	5
5	1	0	1	1	1	1	8
6	0	1	0	0	0	0	3

El índice de fiabilidad del test es:
a) 0,73 b) 0,54 c) 0,84

¿Qué porcentaje de varianza error hay en las puntuaciones empíricas del test?
a) 46% b) 8% c) 54%

El coeficiente de validez del test vale:
a) 0,92 b) 0,73 c) 0,84

¿Qué porcentaje de varianza de las puntuaciones en el criterio no es explicada por las puntuaciones en el test?
a) 8% b) 46% c) 15%

FIN DEL TEMA 6

Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 7

Validez de las inferencias (II)

Cuando se desea conocer el influjo de varias variables predictoras cuantitativas en otra también cuantitativa (criterio), los procedimientos estadísticos que van a permitir obtener esta información y dar solución a estos problemas son, fundamentalmente, la correlación múltiple y el modelo de regresión lineal múltiple.

VALIDACIÓN CON VARIOS PREDICTORES Y UN SOLO INDICADOR DEL CRITERIO

CORRELACIÓN PARCIAL

Permite interpretar el grado de correlación entre la variable criterio (Y) y una de las variables predictoras, eliminando de antemano el efecto que sobre dicha correlación puedan estar ejerciendo el resto de las variables.

$$R_{YX_1 \cdot X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{(1 - r^2_{YX_2})(1 - r^2_{X_1X_2})}}$$

$$R_{YX_2 \cdot X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{(1 - r^2_{YX_1})(1 - r^2_{X_1X_2})}}$$

En la primera se calcula la correlación entre la variable criterio Y y la variable predictora X_1 eliminando la influencia que, en esa correlación, pueda estar ejerciendo la variable X_2 .

En la segunda, al contrario, se calcula la correlación entre la variable criterio y la variable predictora X_2 eliminando el influjo que, en esa correlación, pueda estar ejerciendo la variable predictora X_1 .

CORRELACIÓN SEMIPARCIAL

Permite conocer el grado de correlación entre la variable criterio (Y) y una de las variables predictoras, eliminando el efecto que sobre esta variable predictora puedan estar ejerciendo el resto de las variables.

$$R_{Y(X_1 \cdot X_2)} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1 - r^2_{X_1X_2}}}$$

$$R_{Y(X_2 \cdot X_1)} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{1 - r^2_{X_1X_2}}}$$

La primera fórmula expresa la correlación entre la variable criterio (Y) y la variable predictora X_1 cuando de esta variable se elimina la influencia que pueda estar ejerciendo la variable predictora X_2.

La segunda fórmula expresa la correlación entre la variable criterio (Y) y la variable predictora X_2 cuando de esta variable se elimina la influencia que pueda estar ejerciendo la variable predictora X_1.

COEFICIENTE DE VALIDEZ MÚLTIPLE

La correlación múltiple, va a permitir analizar el grado de asociación entre la variable dependiente (el criterio) y el conjunto de variables predictoras, en nuestro caso X_1 y X_2.

$$R_{Y \cdot X_1 X_2} = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2 r_{YX_1} r_{YX_2} r_{X_1 X_2}}{1 - r_{YX_2}^2}}$$

Otra forma de expresar la correlación múltiple es en función de los coeficientes de regresión múltiple en puntuaciones típicas, cuya fórmula expondremos más adelante, y de las correlaciones de cada variable predictora con el criterio:

$$R_{Y \cdot X_1 X_2} = \sqrt{b_1^* r_{YX_1} + b_2^* r_{YX_2}}$$

EJEMPLO 1

Junio 2016, primera semana

Con los datos que se presentan a continuación responder a las preguntas 11 a la 13

Se quiere comprobar la influencia que dos variables (X1 y X2) tienen sobre la Empatía de las personas (Y). Para ello se han aplicado las pruebas correspondientes y se han obtenido los siguientes resultados:

$$r_{Y.X_1} = 0{,}69; r_{Y.X_2} = 0{,}45; r_{X_1.X_2} = 0{,}52$$

11. Si se eliminara el efecto de la variable X_2, la correlación entre el criterio (Y) y la variable X_1 sería: a) 0.59; b) 0.65; c) 0.60

ECUACIONES DE REGRESIÓN (I)

EN PUNTUACIONES TÍPICAS

$$Z_{Y'} = b_1^* Z_{X_1} + b_2^* Z_{X_2}$$

donde

$$b_1^* = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{1 - r_{X_1X_2}^2}$$

$$b_2^* = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{1 - r_{X_1X_2}^2}$$

ECUACIONES DE REGRESIÓN (II)

EN PUNTUACIONES DIFERENCIALES

$$y' = b_1 x_1 + b_2 x_2$$

EN PUNTUACIONES DIRECTAS

$$Y' = a + b_1 X_1 + b_2 X_2$$

donde, en ambos casos:

$$b_1 = b_1^* \frac{S_Y}{S_{X_1}} \qquad b_2 = b_2^* \frac{S_Y}{S_{X_2}}$$

Cálculo de a

$a = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2$ en puntuaciones directas

$a = 0$ en puntuaciones diferenciales

VARIANZA RESIDUAL

A la varianza de todos los errores de estimación cometidos con los sujetos de la muestra seleccionada se denomina **Varianza residual, Varianza error** o **Error cuadrático medio** y su fórmula es:

$$S^2_{Y \cdot X_1 X_2} = \frac{\sum (Y - Y')^2}{N}$$

A la desviación típica de estos errores se la denomina: **Error típico de estimación múltiple** y su fórmula es:

- En puntuaciones directas y diferenciales

$$S_{Y \cdot X_1 X_2} = \sqrt{\frac{\sum (Y - Y')^2}{N}} = S_Y \sqrt{1 - R^2_{Y \cdot X_1 X_2}}$$

- En puntuaciones típicas

$$S_{Z_Y \cdot Z_{X_1} Z_{X_2}} = \sqrt{1 - R^2_{Y \cdot X_1 X_2}}$$

EJEMPLO 2

Junio 2016, primera semana

Con los datos que se presentan a continuación responder a las preguntas 11 a la 13

Se quiere comprobar la influencia que dos variables (Xl y X2) tienen sobre la Empatía de las personas (Y). Para ello se han aplicado las pruebas correspondientes y se han obtenido los siguientes resultados:

$$r_{Y.X_1} = 0{,}69; r_{Y.X_2} = 0{,}45; r_{X_1.X_2} = 0{,}52$$

12. La proporción de la varianza de las puntuaciones en el criterio (Y) que se puede pronosticar a partir de las puntuaciones obtenidas en las variables X_1 y X_2 es: a) 0.49; b) 0,55; c) 0.34

13. El error típico de estimación múltiple en puntuaciones típicas será: a) 0.51; b) 0.61; c) 0.71

COEFICIENTES

Coeficiente de determinación múltiple. Representa la proporción (o el porcentaje si se multiplica por cien) de la varianza de las puntuaciones de los sujetos en el criterio (variable dependiente) que se puede pronosticar a partir del conjunto de variables predictoras. También se define como la varianza común o asociada entre el criterio y las variables predictoras.

$$CD = R^2_{Y \cdot X_1 X_2}$$

Coeficiente de alienación múltiple. El coeficiente de alienación representa la inseguridad, o el azar, que afecta a los pronósticos.

$$CA = K = \frac{S_{Y \cdot X_1 X_2}}{S_Y} = \sqrt{1 - R^2_{Y \cdot X_1 X_2}}$$

Coeficiente de valor predictivo múltiple. Es el complementario del coeficiente de alienación y es otra forma de expresar la capacidad de las variables predictoras para pronosticar el criterio ya que representa la proporción (o el porcentaje si se multiplica por cien) de seguridad en los pronósticos.

$$CVP = 1 - \sqrt{1 - R^2_{Y \cdot X_1 X_2}}$$

EJEMPLO 3

Septiembre 2017

Con los datos que se dan a continuación responder a las preguntas 20 y 21

A una muestra de sujetos se les aplican dos tests A y B. La correlación del test A con el Rendimiento Académico (RA) es de 0,80 y la del test B 0,70. La correlación entre ambos tests es 0,15. Uno de los sujetos de la muestra ha obtenido en el test A una puntuación típica igual a 1 y en el test B una puntuación típica igual a 0.

20.- A partir de la puntuación obtenida en los dos tests, la puntuación típica pronosticada en RA será: a) 0,80; b) 0,59; c) 0,71.

21.- Los coeficientes de determinación y de valor predictivo múltiple serán respectivamente: a) 0,98 y 0,86; b) 0,71 y 0,98; c) 0,98 y 0,14

MÉTODOS PARA SELECCIONAR LAS VARIABLES PREDICTORAS MÁS ADECUADAS

MÉTODOS *FORWARD*

Dentro de estos métodos vamos a explicar el más utilizado que es el *stepwise* (paso a paso

- Se calculan las intercorrelaciones entre las distintas variables.
- Se selecciona en primer lugar la variable predictora (independiente) cuya correlación con el criterio sea más alta y se construye la ecuación de regresión.
- Se van añadiendo en la ecuación de regresión, una a una, las demás variables predictoras pero siguiendo la siguiente pauta: la segunda variable a incluir será aquella cuya correlación semiparcial con el criterio sea más alta; es decir, sea más alta después de haber eliminado de antemano el efecto que pueda estar ejerciendo dicha variable sobre la variable que se había seleccionado en primer lugar, y así sucesivamente.
- Cada vez que se incluye una variable predictora en la ecuación de regresión se calcula el aumento que se produce en el porcentaje de varianza del criterio que explican el conjunto de variables seleccionadas (aumento en el coeficiente de determinación múltiple) y se analiza si ese aumento es estadísticamente significativo o no. El proceso se detiene cuando el aumento no es significativo.)

MÉTODOS *BACKWARD*

Es un método inverso al anterior y menos utilizado. Al utilizar este método se procede de la siguiente manera:

- Se calcula la correlación múltiple al cuadrado (coeficiente de determinación) entre la variable criterio y todo el conjunto de predictores de que se dispone.
- Se van eliminando una a una las variables menos relevantes calculando en cada proceso de eliminación la reducción que se produce en el coeficiente de determinación.
- El proceso se detiene cuando la reducción observada sea significativa.

VALIDEZ Y UTILIDAD DE LAS DECISIONES

ÍNDICES DE VALIDEZ (I)

Clasificación de los sujetos en función del test y del criterio

		Criterio		
		A	R	
Test	A	N_{AA}	N_{AR}	N_{AT}
	R	N_{RA}	N_{RR}	N_{RT}
		N_{AC}	N_{RC}	N

Coeficiente Kappa

$$K = \frac{p_c - p_a}{1 - p_a} = \frac{F_c - F_a}{N - F_a}$$

Proporción de clasificaciones correctas

$$PCC = \frac{N_{AA} + N_{RR}}{N}$$

ÍNDICES DE VALIDEZ (II)

Sensibilidad. Proporción de aspirantes correctamente seleccionados respecto al total de los que tuvieron éxito en el criterio

$$S = \frac{N_{AA}}{N_{AC}}$$

Especifidad. Proporción de aspirantes que fueron correctamente rechazados respecto al total de los aspirantes que no alcanzaron un rendimiento adecuado en el criterio.

$$E = \frac{N_{RR}}{N_{RC}}$$

Razón de eficacia. Proporción de aspirantes seleccionados mediante la prueba de admisión que rindieron satisfactoriamente en el criterio.

$$RE = \frac{N_{AA}}{N_{AT}}$$

ÍNDICES DE SELECCIÓN

Razón de idoneidad. Cuando se lleva a cabo una selección, la razón de idoneidad equivale a la proporción de aspirantes que rindieron satisfactoriamente en el criterio.

$$RI = \frac{N_{AC}}{N}$$

Razón de selección. En un proceso de selección, es la proporción de aspirantes que han sido seleccionados mediante el test.

$$RS = \frac{N_{AT}}{N}$$

EJEMPLO 4

Junio 2017, segunda semana

Con los siguientes datos responder a las preguntas 16 y 17.

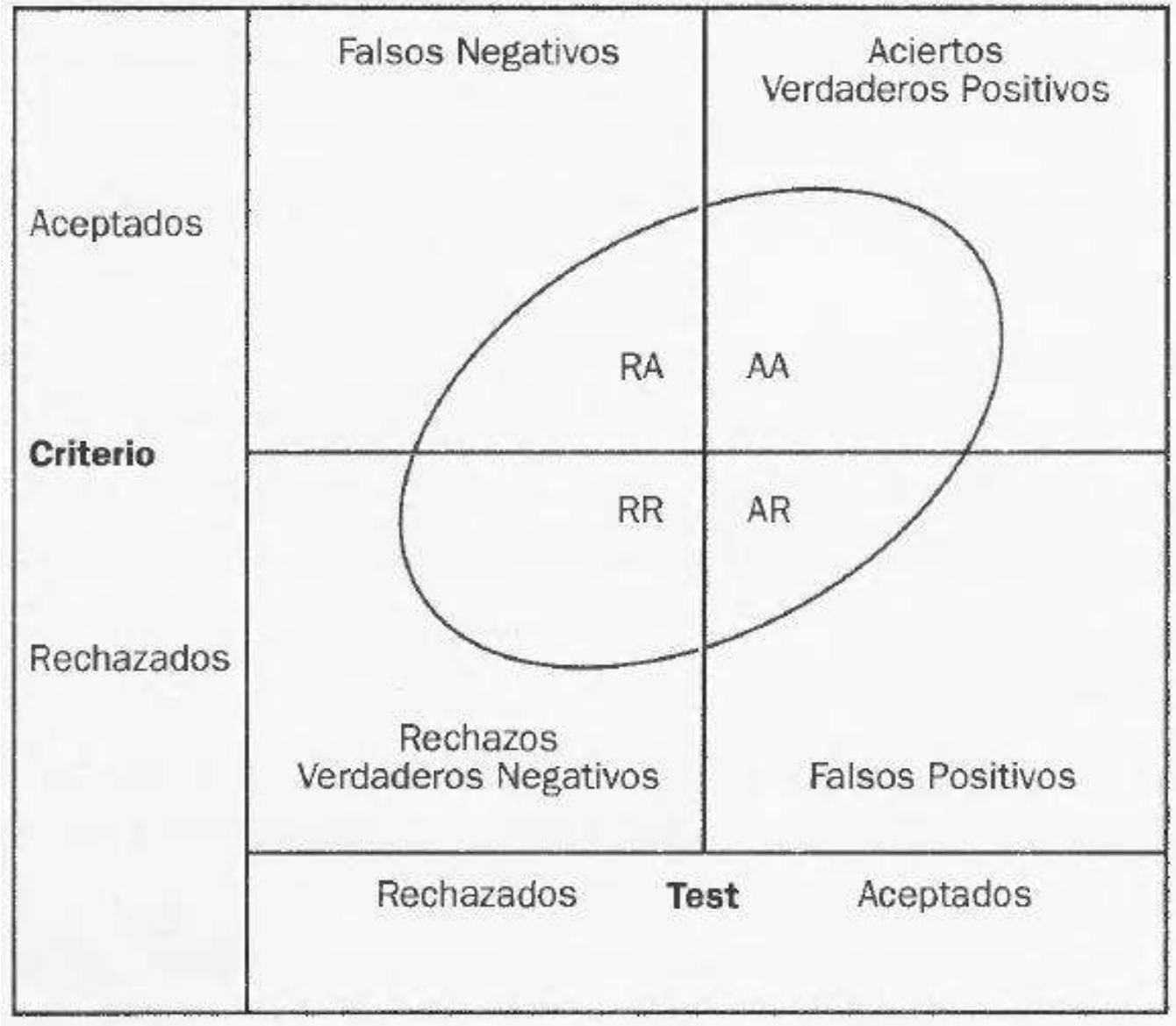
A una muestra de 100 niños se les ha aplicado un test en proceso de validación que trata de diagnosticar el Trastorno de Hiperactividad con Déficit de Atención (TDAH). A su vez, estos mismos niños han sido evaluados por un experto. La clasificación realizada por el test y el experto aparece a continuación:

		Experto		
		Diagnóstico TDAH	No-diagnóstico TDAH	Total
Test B	Diagnóstico TDAH	34	10	44
	No-diagnóstico TDAH	15	41	56
	Total	49	51	100

16. La sensibilidad es igual a: a) 0,34; b) 0,69; c) 0,77.

17. El número de falsos negativos es: a) 10; b) 15; c) 41.

UBICACIÓN DEL PUNTO DE CORTE (I)



UBICACIÓN DEL PUNTO DE CORTE (II)

El punto de corte habría que situarlo teniendo en cuenta la matriz de pagos, es decir, la matriz que refleje las *«pérdidas»* y *«ganancias»* derivadas de las decisiones adoptadas.

Cuando un decisor utiliza un criterio **maximin**, deberá elegir aquella alternativa que entre los resultados más desfavorables, le permita obtener la máxima ganancia (máximo de los mínimos). Cuando utiliza el criterio **mínimax**, el decisor estudiaría las alternativas que le van a proporcionar las máximas «pérdidas» y dentro de esas alternativas elegiría aquella que le proporcionara una «pérdida» menor (mínimo de los máximos).

MODELOS DE SELECCIÓN

Modelo compensatorio. Se trata de un modelo aditivo en el que a cada sujeto se le asigna una única puntuación global. El nombre alude a que los sujetos pueden compensar una baja puntuación en una de las pruebas con una puntuación alta en otras de manera que el resultado final sea una única puntuación.

Modelo conjuntivo. En este modelo se fijan de antemano unos mínimos en cada una de las pruebas utilizadas para la selección, de manera que sólo se seleccionarán aquellas personas que hayan superado esos mínimos en todas y cada una de las pruebas.

Modelo disyuntivo. En este modelo sólo se exige superar determinado nivel de competencia en al menos alguno de los predictores o bloque de predictores.

Modelo conjuntivo-compensatorio. Se aplica, en un primer momento, el modelo conjuntivo y se seleccionan aquellos sujetos que superan los mínimos establecidos en cada uno de los predictores. A continuación, a los sujetos seleccionados se les aplica el modelo compensatorio de manera que queden ordenados en función de la puntuación global obtenida.

Modelo disyuntivo-compensatorio. Se hace una primera selección aplicando el modelo disyuntivo y a los sujetos seleccionados se les aplica el modelo compensatorio.

ESTIMACIÓN DE LA EFICACIA DE UNA SELECCIÓN

Entre los índices que hemos expuesto anteriormente uno de ellos es la razón de eficacia que representa la proporción de personas seleccionadas que tienen éxito en el criterio.

Otra forma de estimar la eficacia de la selección es utilizando el modelo de regresión, siempre que se verifiquen los supuestos que implica, pues permite estimar la probabilidad de que los seleccionados tengan éxito en el criterio.

EJEMPLO 5

Septiembre 2016

Con los datos siguientes responder a las preguntas 15 y 16

Para la participación en un curso de "Coaching" se han presentado 200 alumnos y sólo hay 50 plazas por lo que es necesario llevar a cabo un proceso de selección mediante una determinada prueba **(X).** De los 50 seleccionados 40 terminaron el curso con buen rendimiento y 10 no alcanzaron la calificación mínima obtenida. La media y la desviación típica obtenida por los sujetos en la prueba (X) fueron de 6 y 2 puntos respectivamente. Asumiendo que las puntuaciones obtenidas por los aspirantes en la prueba X se distribuyen según la curva normal de probabilidad, que la ecuación de regresión de Y/X es Y= 0.40 + 3X, que la S_y = 4, y r_{xy}= 0.60

15. La razón de selección y la razón de eficacia fueron respectivamente: a) 0.25 y 0.80; b) 0.25 y 0.20; c) 0.20 y 0.80

16. Si la puntuación mínima para considerar que se ha tenido éxito en el curso de Coaching fuera de 20. La probabilidad de éxito en el curso de las personas seleccionadas hubiera sido: a) 0.60; b) 0.78; c) 0.84

FACTORES QUE INFLUYEN EN LA VALIDEZ

VARIABILIDAD DE LA MUESTRA

$$R_{XY} = \frac{S_X \cdot r_{xy}}{\sqrt{S_X^2 \cdot r_{xy}^2 + s_x^2 - s_x^2 \cdot r_{xy}^2}}$$

$$S_Y = s_y \sqrt{1 - r_{xy}^2 + r_{xy}^2 \frac{S_X^2}{s_x^2}}$$

Donde

R_{XY} → Coeficiente de validez en el grupo de aspirantes

r_{xy} → Coeficiente de validez en el grupo de seleccionados

S_X → Desviación típica en el test del grupo de aspirantes

S_Y → Desviación típica en el criterio del grupo de aspirantes (estimada)

S_x → Desviación típica en el test del grupo de seleccionados

S_y → Desviación típica en el criterio del grupo de seleccionados

EJEMPLO 6

Avex 2020

El coeficiente de validez de un test es 0,70 y su coeficiente de fiabilidad 0,60. ¿Cuál será el coeficiente de valor predictivo de dicho test si se calculara en un grupo con una varianza doble de la del grupo normativo en el que fue calculado nuestro coeficiente de validez? a) 0,42 b) 0,56 c) 0,55

LONGITUD DEL TEST

$$R_{XY} = \frac{r_{xy}\sqrt{n}}{\sqrt{1 + (n-1)r_{xx'}}}$$

$$n = \frac{R_{XY}^2(1 - r_{xx'})}{r_{xy}^2 - R_{XY}^2 r_{xx'}}$$

Donde

R_{XY} → Coeficiente de validez estimado al modificar la longitud del test

r_{xy} → Coeficiente de validez inicial del test

n → Número de veces que se aumenta o disminuye la longitud del test

$r_{xx'}$ → Coeficiente de fiabilidad del test

EJEMPLO 7

Junio de 2019, primera semana

16. Un test compuesto por 50 ítems presenta un coeficiente de fiabilidad de 0.70 y un coeficiente de validez de 0.80. ¿Cuántos ítems paralelos tendríamos que añadir si deseamos alcanzar un coeficiente de validez de 0.90?: a) 122; b) 172; c) 343.

FIABILIDAD DEL TEST Y DEL CRITERIO

		Fiabilidad del CRITERIO		
		No se modifica	Se modifica	Fiabilidad perfecta
Fiabilidad del TEST	No se modifica		$R_{XY} = \frac{r_{xy}}{\sqrt{\frac{r_{yy}}{R_{yy}}}}$	$R_{XV_y} = \frac{r_{xy}}{\sqrt{r_{yy}}}$
	Se modifica	$R_{XY} = \frac{r_{xy}}{\sqrt{\frac{r_{xx}}{R_{xx}}}}$	$R_{XY} = \frac{r_{xy}}{\sqrt{\frac{r_{xx}}{R_{xx}}}\sqrt{\frac{r_{yy}}{R_{yy}}}}$	
	Fiabilidad perfecta	$R_{V_xY} = \frac{r_{xy}}{\sqrt{r_{xx}}}$		$R_{V_xV_y} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}$

Valor máximo del coeficiente de validez → $r_{xy} \leq r_{vx}$

EJEMPLO 8

Junio de 2017, primera semana

Con los datos que se dan a continuación responder a las preguntas 21 a la 23

Para comprobar si un test de aptitud espacial (X) se puede utilizar en la selección de diseñadores gráficos se lleva a cabo un estudio de validación referida al criterio. Para ello se selecciona algún indicador del criterio (Y), una muestra representativa de la población de estudio a la que se aplica el test (dado que se trata de un ejemplo supongan que son 10 sujetos) y se obtienen los datos siguientes: r_{xx} = 0.90; ΣX= 70; ΣY= 52; ΣX^2=720; ΣY^2= 344; ΣXY= 482

21. El coeficiente de valor predictivo del test será: a) 0.83; b) 0.59; c) 0.91

22. La puntuación típica pronosticada en el criterio a un sujeto que en el test se encontraba a una desviación típica por encima de la media será: a) 0.91; b) 0.54; c) 0.83

23. Si se duplicara la fiabilidad del test, el coeficiente de validez sería: a) 0.93; b) 0.97; c) 0.90

FIN DEL TEMA 7

Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 8

Análisis de los ítems

DIFICULTAD DE LOS ÍTEMS

ÍNDICE DE DIFICULTAD

Formalmente el índice de dificultad viene expresado por:

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

donde

A → número de sujetos que aciertan el ítem

N → número de personas que intentan responder el ítem

El índice de dificultad oscila entre 0 y 1. Donde 0 indica que ningún sujeto ha acertado el ítem, mientras que 1 hace referencia a que todos los sujetos respondieron correctamente. Es por ello que debería llamarse índice de facilidad más que de dificultad. En general, se recomienda que los ítems con valores extremos para la población a la que van dirigidos sean eliminados del test final ya que no contribuyen a diferenciar entre sujetos con distinto nivel en el rasgo medido, puesto que o todos los aciertan o todos los fallan.

MEDIA Y VARIANZA

El ID está relacionado directamente con la media y varianza del test. Respecto a la media, en ítems dicotómicos encontramos la siguiente relación:

$$\overline{X} = \sum_{j=1}^{n} ID_j$$

La relación entre la dificultad y la varianza del test es aún más directa, sobre todo si consideramos que en ítems dicotómicos la varianza viene dada por:

$$S_j^2 = p_j q_j$$

CORRECCIÓN DE ACIERTOS POR AZAR

$$ID_c = \frac{A - \frac{E}{k-1}}{N} = \frac{A}{N} - \frac{\frac{E}{k-1}}{N} = p - \frac{q}{k-1}$$

donde

ID_c → índice de dificultad corregido

A → aciertos

E → errores

p → proporción de aciertos

q → proporción de errores

k → número de alternativas del ítem

N → número de personas que intentan resolver el ítem

EJEMPLO 1

Junio 2017, primera semana

24. En la tabla aparecen las respuestas de 400 sujetos a las 4 alternativas (A, B, C, D) (siendo la opción A la correcta) de un ítem. Se presenta el número de sujetos que han seleccionado cada alternativa y que han obtenido puntuaciones superiores e inferiores al 50% de su muestra. El índice de dificultad del ítem corrigiendo el azar es: a) 0.08; b) 0.12; c) 0.17.

	A*	**B**	**C**	**D**
50% superior	100	39	32	14
50% inferior	24	50	70	71

El concepto de discriminación referido a un ítem nos indica si el ítem discrimina, distingue entre sujetos con alto y bajo nivel en el rasgo medido

DISCRIMINACIÓN

BASADO EN GRUPOS EXTREMOS

Kelly (1939) aconseja tomar el 27% (o el 25%) superior y el 27% (o el 25%) inferior de la muestra total para obtener un índice *D* sensible y estable.

$$D = p_s - p_i$$

El índice *D* oscila entre -1 y 1. Tomará el valor ideal de 1 cuando todas las personas del grupo superior hayan acertado el ítem y las del inferior lo hayan fallado.

Tabla de Ebel

Valores	Interpretación
D≥0.40	El ítem discrimina muy bien
0.30≤D≤0.39	El ítem discrimina bien
0.20≤D≤0.29	El ítem discrimina poco
0.10≤D≤0.19	El ítem necesita revisión
D<0.10	El ítem carece de utilidad

EJEMPLO 2

Septiembre 2017, reserva

Con los datos que se dan a continuación responder a las preguntas 23, 24 y 25

Una muestra de 300 sujetos contesta a un ítem de un test y lo aciertan 200 sujetos. Separados los sujetos en el 27% superior e inferior se observa que de los 200 sujetos que acertaron el ítem hay 70 que pertenecen al 27% de sujetos que mejores puntuaciones obtuvieron en el test y 30 al 27% de los que obtuvieron peores puntuaciones.

23.- El índice de dificultad sin corregir del elemento fue: a) 0.67; b) 0.77; c) 0.81

24.-. El índice de discriminación fue: a) 0.35; b) 0.49; c) 0.51

25.-El número de discriminaciones máximo que realiza el elemento en esa muestra fue: a) 30000; b) 40000; c) 20000.

BASADO EN CORRELACIÓN

Si un ítem discrimina adecuadamente entonces la correlación entre las puntuaciones obtenidas por los sujetos en el ítem y las obtenidas en el test total será positiva. Es decir, los sujetos que puntúan alto en el test tendrán más probabilidad de acertar el ítem. La puntuación total de los sujetos en el test ha de calcularse descontando la puntuación del ítem. En caso contrario, estaríamos incrementando artificialmente el índice de discriminación ya que estaríamos correlacionando una variable (ítem) con otra variable (test) que contiene a la primera.

Ahora bien, el índice de correlación utilizado ha de ser coherente con el tipo de puntuaciones del ítem y del test.

Coeficiente de correlación Φ. Se utiliza cuando las puntuaciones del ítem y del test son estrictamente dicotómicas.

Correlación biserial puntual (r_{bp}). Cuando el ítem es una variable dicotómica *y* la puntuación en el test es continua, el índice de correlación más apropiado es el biserial - puntual.

Correlación biserial (r_b). En la correlación biserial se entiende que ambas variables son inherentemente continuas, aunque una de ellas se haya dicotomizado (ítem).

EJEMPLO 3

Septiembre 2018, reserva

En la tabla adjunta se presentan las puntuaciones de 4 sujetos en un test de 4 ítems de tres alternativas. Responder a las preguntas 19 y 20

	1	2	3	4	X_{total}
A	0	0	1	1	2
B	1	1	1	0	3
C	1	0	0	0	1
D	1	1	1	0	3

19. El índice de discriminación del ítem 3 está entre el rango: a) 0,40_0,45; b) 0,55_0,60; c) 0,65_0,70.

20. La puntuación directa corregida del sujeto D es igual a: a) 1; b) 2,5; c) 3.

EJEMPLO 4

Septiembre 2019, reserva

A partir de los datos siguientes responder a las preguntas 22 y 23

Una muestra de 6 alumnos universitarios obtuvo las siguientes puntuaciones en un test de comprensión verbal, con ítems de 5 alternativas de respuesta, siendo una la correcta. Únicamente los sujetos B, D y E acertaron el último ítem del test, aunque todos lo intentaron resolver.

22. ¿Cuál es el índice de dificultad corregido del último ítem? a) 0.38; b) 0.50; c) 0.12.

23. ¿Cuál es el índice de discriminación del último ítem según la correlación biserial-puntual? a) -0.10; b) 0.10; c) 0.5.

Sujetos	Test
A	5
B	7
C	7
D	8
E	9
F	12

DISCRIMINACIÓN EN LOS ÍTEMS DE ACTITUDES (I)

Los ítems de actitudes se caracterizan porque no existen respuestas correctas o incorrectas. Un procedimiento para estimar la discriminación de los ítems de actitudes pasa por calcular la correlación entre ítem y test. En este caso, al tratarse de ítems que no son dicotómicos el coeficiente de correlación adecuado sería el de Pearson, también llamado **Índice de Homogeneidad (IH)**.

$$IH = R_{jX} = \frac{N\sum jX - \sum j \sum X}{\sqrt{[N\sum j^2 - (\sum j)^2][N\sum X^2 - (\sum X)^2]}}$$

Si las puntuaciones del ítem están contando a la hora de calcular la puntuación total del test, habría que aplicar una corrección. Como ya vimos, dicha corrección puede implicar, simplemente, descontar de la puntuación total la del ítem o aplicar la siguiente fórmula:

$$R_{j(X-j)} = \frac{R_{jX}S_X - S_j}{\sqrt{S_X^2 + S_j^2 - 2R_{jX}S_X S_j}}$$

EJEMPLO 5

Septiembre 2019

Con los siguientes datos, responder a las preguntas 21, 22 y 23.

En la tabla adjunta se presentan las puntuaciones de 4 sujetos en una prueba cognitiva de 4 ítems.

	Ítems				X_{total}
	1	2	3	4	
A	0	0	1	1	2
B	1	1	1	0	3
C	1	0	0	0	1
D	1	1	1	0	3

21. El índice de discriminación corregido (descontando de la puntuación total la correspondiente al ítem) utilizando la correlación biserial puntual del ítem 2 es: a) 0,57; b) 0,76; c) 0,90.

22. El índice de fiabilidad del ítem 3 es: a) 0,26; b) 0,45; c) 0,55.

23. Si los ítems constaran de 5 opciones de respuesta, ¿cuál sería la puntuación empírica obtenida por el sujeto A, una vez corregido el efecto del azar?: a) 1; b) 1,5; c) 2.

DISCRIMINACIÓN EN LOS ÍTEMS DE ACTITUDES (II)

Otro procedimiento extremadamente útil (aunque menos eficiente que el anterior porque no utiliza toda la muestra) para averiguar si un ítem diferencia entre grupos extremos de actitud consiste en calcular si la media en el ítem de los sujetos con puntuaciones más altas en el test total es estadísticamente superior a la media de los sujetos con puntuaciones más bajas, mediante la prueba de *T de Student.*

$$T = \frac{\overline{X}_{sj} - \overline{X}_{ij}}{\sqrt{\frac{(n_s - 1)S_{sj}^2 + (n_i - 1)S_{ij}^2}{n_s + n_i - 2}\left[\frac{1}{n_s} + \frac{1}{n_i}\right]}}$$

La T de Student obtenida se distribuye con (n_s + n_i - 2) grados de libertad. La hipótesis nula que se pone a prueba es que las medias de ambos grupos son iguales.

EJEMPLO 6

Junio 2014, segunda semana

25.-A continuación se presentan las puntuaciones obtenidas en un ítem que corresponde a una escala tipo Likert, por el 25% de los sujetos que obtuvieron una puntuación más alta en la escala total y por el 25 % que obtuvieron una puntuación más baja.

25% superior	4	3	5	4	1	2	4	5	4	3	4	3
25% inferior	1	2	3	2	3	4	1	2	3	3	1	2

Utilizando el estadístico de contraste "T", calcular si el ítem discrimina adecuadamente entre ambos grupos. (NC 95%): a) el ítem no discrimina; b) el ítem discrimina; c) el ítem no discrimina porque el valor de $T=2.12<t_{_0\,95'24}$ •

FACTORES QUE AFECTAN A LA DISCRIMINACIÓN

VARIABILIDAD

Cuando la varianza de un ítem es cero, implica que todos los sujetos han respondido lo mismo. Y cuando esto ocurre el ítem no presenta ningún poder discriminativo, dado que si su varianza es igual a cero, entonces su correlación con las puntuaciones del test también es cero.

La relación entre la variabilidad del test y la discriminación de los ítems se puede formular algebraicamente:

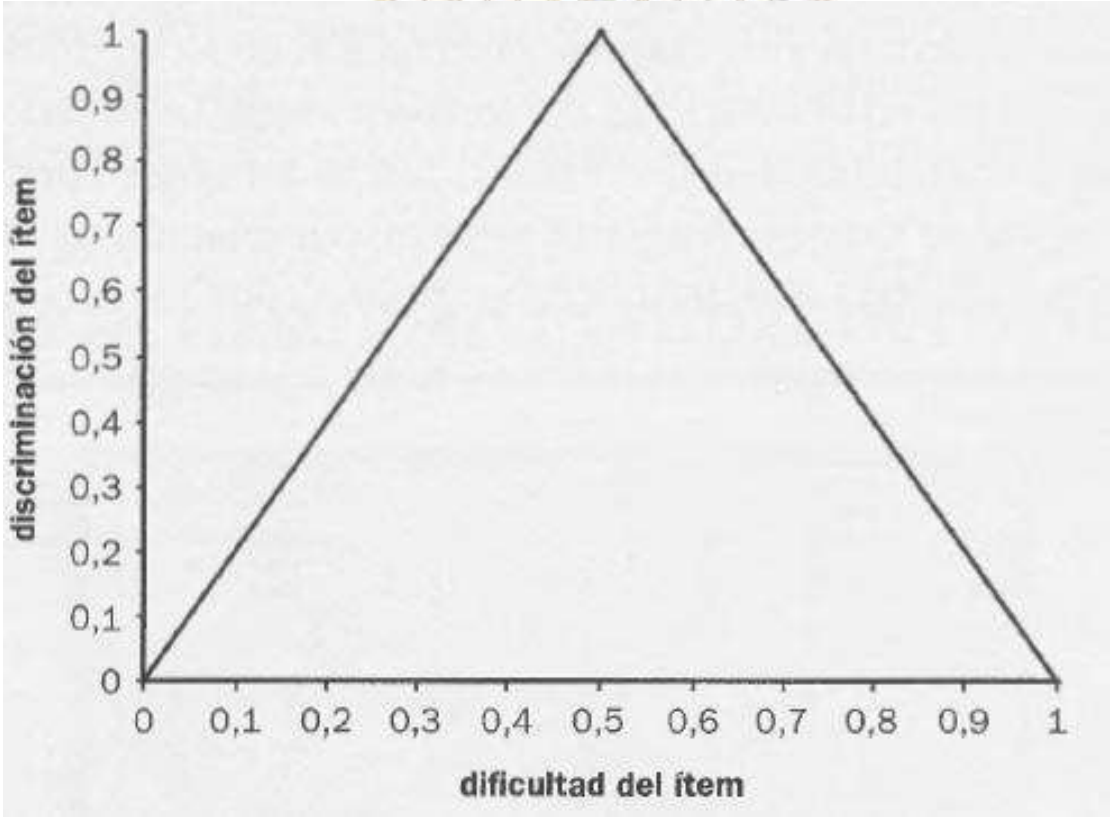
$$S_X = \sum_{j=1}^{n} S_j r_{jX}$$

Si el test estuviera compuesto por ítems dicotómicos, dado que la varianza de una variable dicotómica es igual a la proporción de aciertos por la proporción de fallos:

$$S_X = \sqrt{\sum_{j=1}^{n} p_j q_j r_{jX}^2}$$

DIFICULTAD DEL ÍTEM

Un ítem alcanza su máximo poder discriminativo cuando su dificultad es media (p = 0,5 en ítems dicotómicos)



DIMENSIONALIDAD DEL TEST

Cuando se construye un test, se trata de que sólo mida un único concepto, es decir, que sea unidimensional. Si tras someter el test a un Análisis Factorial encontráramos varias dimensiones subyacentes, implicaría la existencia de distintas escalas, lo que sería similar a una batería de test que mide tantos aspectos como escalas o dimensiones distintas hubiera. Si fuera este el caso, la correlación entre las puntuaciones en el test y las del ítem se verá afectada a la baja, y tanto más cuanto más dimensiones contenga el test.

FIABILIDAD DEL TEST

Si la discriminación se define como la correlación entre las puntuaciones obtenidas por los sujetos en el ítem y las del test, entonces fiabilidad y discriminación han de estar íntimamente relacionados. Tan es así que es posible expresar el coeficiente *alpha de Cronbach* a partir de la discriminación de los ítems (también de su dificultad). Para ello, basta con sustituir S_X por $\sum S_j r_{jX}$.

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum S_j^2}{S_x^2}\right) = \frac{n}{n-1}\left(\frac{S_x^2 - \sum S_j^2}{\left[\sum S_j r_{jX}\right]^2}\right)$$

A medida que aumenta la discriminación media del test, el coeficiente de fiabilidad se incrementa.

ÍNDICES DE FIABILIDAD Y VALIDEZ DE LOS ITEMS

ÍNDICE DE FIABILIDAD

Se utiliza para cuantificar el grado en que el ítem en cuestión está midiendo con precisión el atributo de interés.

$$IF = S_j ID_j$$

donde

S_j → desviación típica de las puntuaciones en el ítem.

ID_j → índice de discriminación del ítem.

Cuando se utiliza algún coeficiente de correlación para calcular la discriminación de los ítems entonces:

$$IF = S_j r_{jX}$$

En la medida que seleccionemos los ítems con mayor IF, mayor será su sumatorio ($\Sigma S_j r_{jX} = \Sigma IF$), y por ende mejor la fiabilidad del test.

ÍNDICE DE VALIDEZ

En el caso de un ítem concreto, la validez implicará correlacionar las puntuaciones obtenidas por una muestra de sujetos en el ítem con las puntuaciones obtenidas por los mismos sujetos en algún criterio externo de interés.

$$IV = S_j r_{jY}$$

En el caso de que el criterio sea una variable continua y el ítem una variable dicotómica, la correlación a utilizar sería la biserial puntual; pero ahora no es necesario descontar de la puntuación total del criterio externo la del ítem ya que ésta no está incluida y, por lo tanto, no contribuye de ninguna manera en su cómputo.

La validez del test también puede expresarse en función de los IV de los ítems, de manera que cuanto mayores sean los IV de los ítems, más optimizarán la validez del test.

$$r_{xy} = \frac{\sum S_j r_{jY}}{\sum S_j r_{jX}} = \frac{\sum IV_j}{\sum IF_j}$$

ANÁLISIS DE DISTRACTORES

REQUISITOS

En este tipo de análisis, hay que seguir los siguientes pasos:

1. Controlar que todas las opciones incorrectas sean elegidas por un mínimo de sujetos.
2. A ser posible, que sean equiprobables, es decir, que sean igualmente atractivas para las personas que no conocen la respuesta correcta.
3. Que el rendimiento en el test de los sujetos que han seleccionado cada alternativa incorrecta sea menor al de los sujetos que han seleccionado la correcta.
4. En relación con el punto anterior, es de esperar que a medida que aumente el nivel de aptitud de los sujetos, el porcentaje de ellos que seleccionen las alternativas incorrectas disminuya, y viceversa.

EQUIPROBABILIDAD DISTRACTORES

Una forma de comprobar la equiprobabilidad de los distractores es mediante la prueba de independencia χ^2.

$$\chi^2 = \sum \frac{(f_t - f_o)^2}{f_t}$$

Que se distribuye según k-1 grados de libertad, donde k es el número de alternativas incorrectas. La hipótesis nula a poner a prueba es que $f_t = f_o$, que significa que para los sujetos que no conocen la respuesta correcta la elección de cualquiera de los distractores es igualmente atractiva.

EJEMPLO 6

Junio 2015, segunda semana

19. En la tabla que se muestra a continuación se muestran las respuestas que dieron una muestra de sujetos a un ítem de un test de aptitudes. La respuesta correcta es la A. A un NC del 95%, ¿ podemos decir que las alternativas incorrectas son equiprobables?: a) sí porque el valor teórico del estadístico es menor que el valor obtenido; b) sí porque el valor teórico del estadístico es mayor que el valor obtenido; c) no porque el valor teórico del estadístico es mayor que el obtenido.

A	B	C
389	52	36

PODER DISCRIMINATIVO

Si son buenos distractores, lo lógico es que discriminen en sentido contrario a como lo hace la opción correcta. Es decir, si se espera que la correlación entre las puntuaciones del test y la opción correcta sea positiva, y cuanto más mejor, lo esperable de un buen distractor es que su correlación sea negativa. Lo que implica que a medida que aumenta el nivel de aptitud de los sujetos la proporción de sujetos que elige el distractor disminuya.

Figura 8.7 Ejemplo de un buen ítem

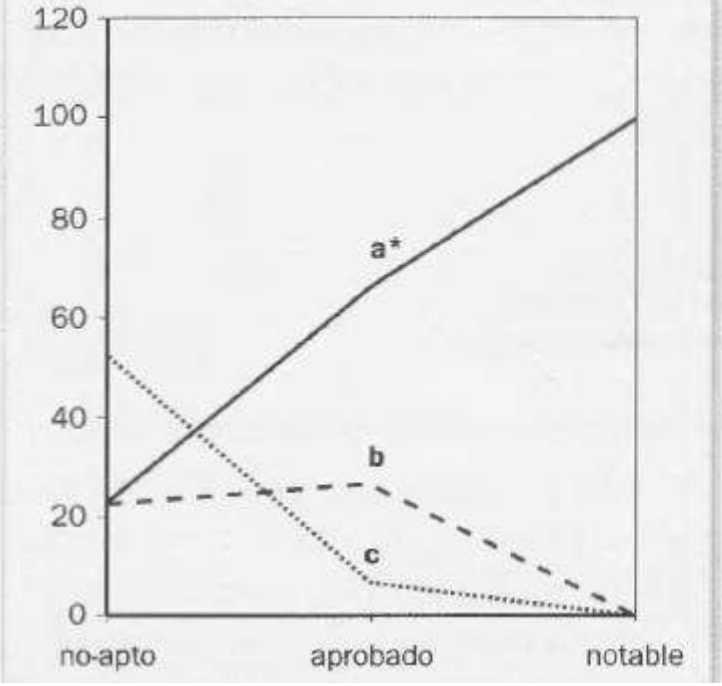
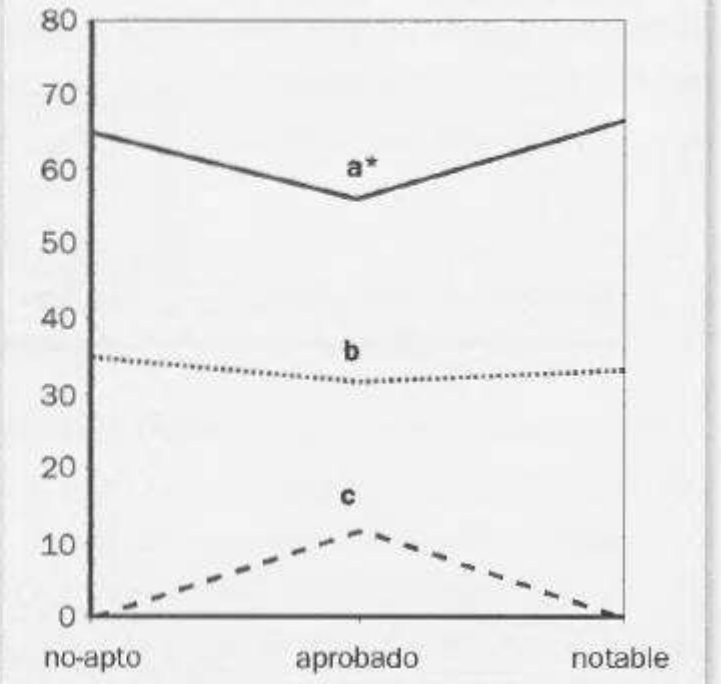


Figura 8.8 Ejemplo de un mal ítem



EJEMPLO 7

Septiembre de 2017

23.-En la siguiente tabla se presenta el número de sujetos que ha seleccionado cada una de las alternativas de un ítem, siendo la opción B* la correcta. La muestra se ha separado entre el 27% superior y 27% inferior. La muestra total estaba formada por 300 sujetos. Si calculamos el poder discriminativo de las alternativas incorrectas mediante el índice basado en las proporciones de acierto *D,* por orden de mejor a peor alternativas encontramos la siguiente ordenación: a) C, D y A; b) A, D y C; c) D, A y C.

	A	B*	C	D
27% superior	10	40	16	15
27% inferior	20	9	22	30

FUNCIONAMIENTO DIFERENCIAL DE LOS ÍTEMS

INTRODUCCIÓN

Otro aspecto a evaluar dentro del análisis de ítems, es si de manera sistemática sujetos de distintos grupos de pertenencia pero con el mismo nivel en el rasgo medido tienen distintas probabilidades de éxito en el ítem en cuestión (Shultz y Whitney, 2005). A esta circunstancia se la conoce como **funcionamiento diferencial de los ítems (FDI)**, reservando la palabra **sesgo** para la interpretación de las causas de dicho funcionamiento diferencial. Por el contrario, si dichas diferencias son debidas a una diferencia real en la variable medida y no a fuentes sistemáticas de variación entonces hablamos de **impacto** (Ackerman, 1992).

MANTEL-HAENSZEL (I)

Para detectar el FDI existen una amplia variedad de procedimientos estadísticos. Por su parsimonia y buenos resultados el método de Mantel-Haenszel (1959) es uno de los más utilizados y además se encuentra implementado en gran parte de las aplicaciones informáticas sobre FDl.

Para aplicar Mantel-Haenszel, en primer lugar habrá que identificar una variable que sea la posible causante del FDl. Una vez seleccionada, hemos de conformar dos grupos, uno de Referencia (GR), y otro Focal (GF). El GR suele coincidir con el grupo favorecido. Por el contrario, el GF suele ser el conformado por los sujetos perjudicados. Luego se establecen distintos niveles de aptitud tomando la puntuación empírica obtenida en el test y, finalmente, se cuenta el número de respuestas correctas e incorrectas por cada grupo (GR y GF) y nivel de habilidad *i.*

MANTEL-HAENSZEL (II)

La hipótesis nula a poner a prueba es:

$$H_0: \frac{A_i}{B_i} = \frac{C_i}{D_i}$$

	Correctas	Incorrectas	
GR	A_i	B_i	n_{Ri}
GF	C_i	D_i	n_{Fi}
	n_{1j}	n_{0j}	N_i

Una vez confeccionadas las tablas anteriores (una para cada nivel de aptitud i) aplicamos el estadístico de Mantel-Haenszel.

$$\alpha_{MH} = \frac{\sum \frac{A_i D_i}{N_i}}{\sum \frac{B_i C_i}{N_i}}$$

Los valores obtenidos oscilan entre cero e infinito. Valores mayores que 1 indican que el ítem favorece al GR y menores al GF. Valores iguales a 1 o próximos indican que el ítem no presenta FDI.

FIN DEL TEMA 8

Muchas gracias por vuestra atención

PSICOMETRÍA

Tema 9

Asignación, transformación y equiparación de puntuaciones

TRANSFORMACIÓN DE PUNTUACIONES

Para la interpretación de los resultados surge la necesidad de poder contar con procedimientos que nos permitan dar un significado a las puntuaciones obtenidas.

Los dos procedimientos de interpretación propuestos son: la interpretación normativa y la interpretación criterial .

En la interpretación referida a la norma, o normativa, se compara la puntuación obtenida por un sujeto en un test con las obtenidas, en el mismo test, por un grupo de referencia o **grupo normativo**. A las puntuaciones obtenidas por los sujetos que constituyen el grupo normativo, así como a las transformaciones que se hagan de dichas puntuaciones, se las denomina **normas**. El conjunto de todas las normas constituye el baremo del test.

La interpretación referida al criterio se basa en determinar el grado de dominio que un sujeto tiene sobre un criterio preestablecido. Para ello, se suele tomar una puntuación de corte, que permita clasificar a los sujetos en dos grupos: los que dominan el criterio definido y los que no lo dominan.

PUNTUACIÓN CORREGIDA POR AZAR

PENALIZACIÓN DE ERRORES

$$X_c = A - A_a = A - \frac{E}{k-1}$$

BONIFICACIÓN DE OMISIONES

$$X_c = A + A_o = A + \frac{O}{k}$$

EJEMPLO 1

Junio 2016, primera semana

22. Hemos aplicado un test formado por 60 ítems a una muestra de sujetos. Los 15 primeros son de dos alternativas, los 20 siguientes de tres alternativas y los 25 restantes son de cuatro alternativas. La puntuación que le corresponde en el test a un sujeto que contestó correctamente a 10 de los 15 primeros ítems, a 16 de los 20 siguientes y a 16 de los veinticinco restantes, sabiendo que no dejó ningún ítem sin contestar, es igual a: a) 30; b) 32; c) 42.

EJEMPLO 2

Junio 2016, segunda semana

20. Hemos administrado un test compuesto por 50 ítems de tres alternativas cada uno, a una muestra de 300 sujetos. Cuál sería la puntuación, corrigiendo el azar, en una escala de 0 a 30, de un sujeto que acierta 30 ítems y falla el resto: a) 10; b) 12; c) 20.

TRANSFORMACIÓN DE LAS PUNTUACIONES EN LOS TESTS REFERIDOS A NORMAS

TRANSFORMACIONES LINEALES (I)

ESCALAS TÍPICAS

La puntuación típica nos indica el número de desviaciones típicas a las que se encuentra la puntuación de un sujeto respecto de la media del grupo normativo o de referencia.

$$Z_x = \frac{X - \overline{X}}{S_x}$$

La escala de puntuaciones típicas tiene de media 0 y desviación típica 1. Asimismo, la distribución de puntuaciones típicas de una variable normal suele oscilar de -3 a +3, lo que implica la existencia de valores negativos y decimales. Una forma de evitar este inconveniente es el empleo de las *escalas típicas derivadas.*

TRANSFORMACIONES LINEALES (II)

ESCALAS TÍPICAS DERIVADAS

Las escalas típicas derivadas, son transformaciones lineales de las escalas típicas. Esta transformación consiste, esencialmente, en multiplicar la puntuación típica por una constante *b,* desviación típica de la nueva escala, y sumarle otra constante *a,* la media en la escala resultante.

$$Y = a + bZ_x$$

Si bien existen diversas posibles transformaciones, las más utilizadas suelen ser la escala *D* y la escala *T:*

- **Escala D:** $D = 50 + 20Z_x$

  Se trata de una escala en la que la media es 50 y la desviación típica es 20.

- **Escala T (McCall):** $T = 50 + 10Z_x$

  En esta escala la media es igual a 50 y la desviación típica es igual a 10.

TRANSFORMACIONES NO LINEALES (I)

RANGO DE PERCENTILES (I)

Se define el percentil como aquella puntuación del test que deja por debajo de sí un determinado porcentaje de casos del grupo normativo. El percentil nos proporciona una idea de la posición de un determinado sujeto dentro del grupo normativo. Los percentiles constituyen una escala ordinal.

$$P_x = C_x = \frac{100}{N}\left[f_b + \frac{f_d}{I}(X_c - L_i)\right] = f_{ac}\frac{100}{N}$$

P_x ó C_x = porcentaje de sujetos que obtienen una puntuación inferior a la puntuación directa X. (En Análisis de Datos, a P_x se le llamaba k)

N = número de sujetos de la muestra.

f_b = frecuencia absoluta acumulada bajo del intervalo crítico.

f_d = frecuencia absoluta dentro del intervalo crítico.

I = amplitud de los intervalos.

X_c = puntuación del test correspondiente al centil C_x.

L_i = límite inferior del intervalo crítico.

f_{ac} = frecuencia acumulada al punto medio del intervalo donde se encuentra X_c.

TRANSFORMACIONES NO LINEALES (II)

RANGO DE PERCENTILES (II)

Si queremos saber la puntuación que le corresponde a un sujeto que supera al X% de los sujetos de la muestra, simplemente despejamos de la expresión anterior el término X_c.

$$X_c = L_i + \left(\frac{N \cdot C_x}{100} - f_b\right)\frac{I}{f_d}$$

En Análisis de Datos, a X_c se le llamaba P_k.

TRANSFORMACIONES NO LINEALES (III)

UNED CALATAYUD

ESCALAS TÍPICAS NORMALIZADAS

Las puntuaciones típicas normalizadas se obtienen a partir de los percentiles, y se definen, como la puntuación típica que le corresponde a una puntuación empírica obtenida por un sujeto en un test en una distribución normal.

X	F_a	Z_x	Frecuencias Acumuladas al Punto Medio	Percentil	Típica Normalizada
10	1	1,95	24,5	98	2,05
9	2	1,53	23	92	1,39
8	2	1,11	21	84	0,99
7	3	0,69	18,5	74	0,64
6	4	0,34	15	60	0,25
5	4	–0,15	11	44	–0,15
4	3	–0,57	7,5	30	–0,52
3	2	–0,99	5	20	–1,28
2	3	–1,41	2,5	10	–1,28
1	1	–1,83	0,5	2	–2,00

TRANSFORMACIONES NO LINEALES (IV)

UNED CALATAYUD

ESCALAS NORMALIZADAS DERIVADAS

La escala normalizada derivada más utilizada es la escala de **estaninos** o **eneatipos**. Se utilizó por primera vez durante la Segunda Guerra Mundial por el ejército de los Estados Unidos. La escala de estaninos consiste en una escala de valores enteros y positivos de 9 unidades, del 1 al 9 (se redondea por exceso). Esta escala derivada tiene de media 5 y desviación típica 2.

$$E = 5 + 2(Z_n)$$

EJEMPLO 3

Septiembre de 2019

25. En un test en el que las puntuaciones de los sujetos se ajustan a la curva normal, la media y desviación típica son 30 y 5, respectivamente. El percentil y el eneatipo de los sujetos que obtuvieron una puntuación de 25 son, respectivamente: a) 20,56 y -1; b) 25,56 y 1; c) 15,87 y 3.

EJEMPLO 4

Junio de 2015, primera semana

21.- Si la distribución de las puntuaciones de los sujetos en un determinado test se ajusta a la curva normal, el estanino, el percentil y la puntuación típica derivada de media 8 y desviación típica 2 que le corresponde a un sujeto que en el test se situó a dos desviaciones típicas por encima de la media son: a)9, 98 y 12 b) 7, 98 y 10 c) 6, 96 y 4.

NORMAS CRONOLÓGICAS

Con las normas cronológicas, la interpretación de la puntuación obtenida por un sujeto en un test, se lleva a cabo con relación a su edad y con la puntuación media obtenida por los sujetos de su edad.

Las normas cronológicas más utilizadas son la *edad mental* y el *cociente intelectual.* Las escalas de *edad mental* fueron propuestas, en principio, por Alfred Binet y, posteriormente, por las investigaciones de Binet-Simon.

Para la obtención del *cociente intelectual,* se calcula la edad mental del sujeto y se divide por su edad cronológica, multiplicando por 100 el valor obtenido.

$$\mathrm{CI} = \frac{\mathrm{EM}}{\mathrm{EC}} \cdot 100$$

El cociente intelectual es poco discriminativo para los adultos debido a que la edad mental medida por los tests se estabiliza a partir de una determinada edad cronológica con lo que se produce el efecto de techo.

EJEMPLO 5

Septiembre de 2018, reserva

18. En una prueba de inteligencia, un sujeto ha obtenido una puntuación equivalente a una edad mental de 9 años. Si el sujeto tiene 10 años, cuál es su cociente intelectual: a) 90; b) 100; c) 111.

EQUIPARACIÓN DE PUNTUACIONES

DISEÑOS DE EQUIPARACIÓN

DE UN SOLO GRUPO

En los diseños de un solo grupo se administran las dos formas del test, cuyas puntuaciones se desean equiparar, al mismo grupo de sujetos. Las dos formas del test deben medir la misma característica objeto de estudio y presentar el mismo grado de dificultad. Se suelen contrabalancear las aplicaciones por mitades.

DE GRUPOS EQUIVALENTES

En este diseño, se extraen de la población *y* de forma aleatoria dos muestras de sujetos, *y* a cada muestra se le aplica una forma del test. Por lo tanto, cada sujeto responde solamente a una de las formas.

DE GRUPOS NO EQUIVALENTES CON ÍTEMS COMUNES (DISEÑO DE ANCLAJE)

A cada una de las muestras de sujetos se le administra solamente una forma del test, la Forma X o la Forma Y. Ambas muestras no tienen porqué ser equivalentes entre sí y, además, a ambas muestras se les aplica un test común (Z) que permite establecer las equivalencias entre los tests a equiparar. Consiguientemente, cada sujeto contesta un test diferente y un test común. A este test común que contestan ambos grupos se le conoce como **test de anclaje**, y puede ser interno (preguntas intercaladas con el resto del test) o externo (test independiente).

MÉTODOS DE EQUIPARACIÓN (I)

MÉTODO DE LA MEDIA

En el método de la media se asume que las puntuaciones de uno de los test difieren en una cuantía constante de las puntuaciones del otro test. En esencia, lo que se pretende con este método es hacer corresponder las medias de los tests a equiparar.

$$X^* = Y = X - \overline{X} + \overline{Y}$$

donde

X^* = puntuación del test Y equivalente a una del test X.

X = puntuación del test X.

$\overline{X}$ = media del test X.

$\overline{Y}$ = media del test Y.

MÉTODOS DE EQUIPARACIÓN (II)

MÉTODO LINEAL, DISEÑO DE GRUPOS EQUIVALENTES

Este método se basa en la equiparación de aquellas puntuaciones directas que tienen la misma puntuación típica. Es decir, una determinada puntuación perteneciente a un test Y, es equivalente a una puntuación perteneciente a un test X si ambas puntuaciones tienen idéntica puntuación Z, con lo que $Z_x = Z_y$

$$\frac{X - \overline{X}}{S_x} = \frac{Y - \overline{Y}}{S_y} \quad \text{y despejando, } X^* = Y = \frac{S_y}{S_x}(X - \overline{X}) + \overline{Y}$$

donde

X^* = puntuación del test Y equivalente a una del test X.

S_y = desviación típica de las puntuaciones del test Y.

S_x = desviación típica de las puntuaciones del test X.

X = puntuación del test X.

$\overline{X}$ = media del test X.

$\overline{Y}$ = media del test Y.

EJEMPLO 6

Junio 2015, primera semana

24.- Se ha aplicado un test a una muestra de 150 escolares. La media y varianza de las puntuaciones empíricas son 20 y 6 respectivamente. Si un sujeto ha obtenido en el test una puntuación empírica igual a 22 puntos, ¿utilizando el método lineal, cuál sería la puntuación equivalente que le correspondería a dicho sujeto en un segundo test cuya media fuera igual a 15 y la desviación típica igual a 4? a) 18.26; b) 16.87; c) 14.45.

MÉTODOS DE EQUIPARACIÓN (III)

MÉTODO LINEAL, DISEÑO DE UN SOLO GRUPO

Si se hubiera utilizada un *diseño de un solo grupo,* en el que se deben administrar los dos tests, cuyas puntuaciones se desean equiparar, al mismo grupo de sujetos pero en orden inverso, la transformación lineal se expresaría de la siguiente manera:

$$X^* = Y = \sqrt{\left(\frac{S^2_{y1} + S^2_{y2}}{S^2_{x1} + S^2_{x2}}\right)}\left(X - \frac{\bar{X}_1 + \bar{X}_2}{2}\right) + \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

MÉTODOS DE EQUIPARACIÓN (IV)

MÉTODO LINEAL, DISEÑO DE ANCLAJE

Si se hubiera utilizado un *diseño de anclaje* en el que se cuenta con dos grupos de sujetos y a cada grupo se les administra una forma diferente del test, y un test de anclaje (Z) que es común a ambos grupos, la transformación lineal quedaría expresada en los siguientes términos:

$$X^* = Y = \left(\frac{\sqrt{S_{y2}^2 + b_{yz2}^2 \left(S_z^2 - S_{z2}^2 \right)}}{\sqrt{S_{x1}^2 + b_{xz1}^2 \left(S_z^2 - S_{z1}^2 \right)}} \right) \left[X - \left(\overline{X}_1 + b_{xz1} \left(\overline{Z} - \overline{Z}_1 \right) \right) \right] + \left[\overline{Y}_2 + b_{yz2} \left(\overline{Z} - \overline{Z}_2 \right) \right]$$

MÉTODOS DE EQUIPARACIÓN (V)

MÉTODO EQUIPERCENTIL (I)

El método equipercentil es el método de equiparación más habitual, consiste en equiparar aquellas puntuaciones cuyos percentiles son iguales.

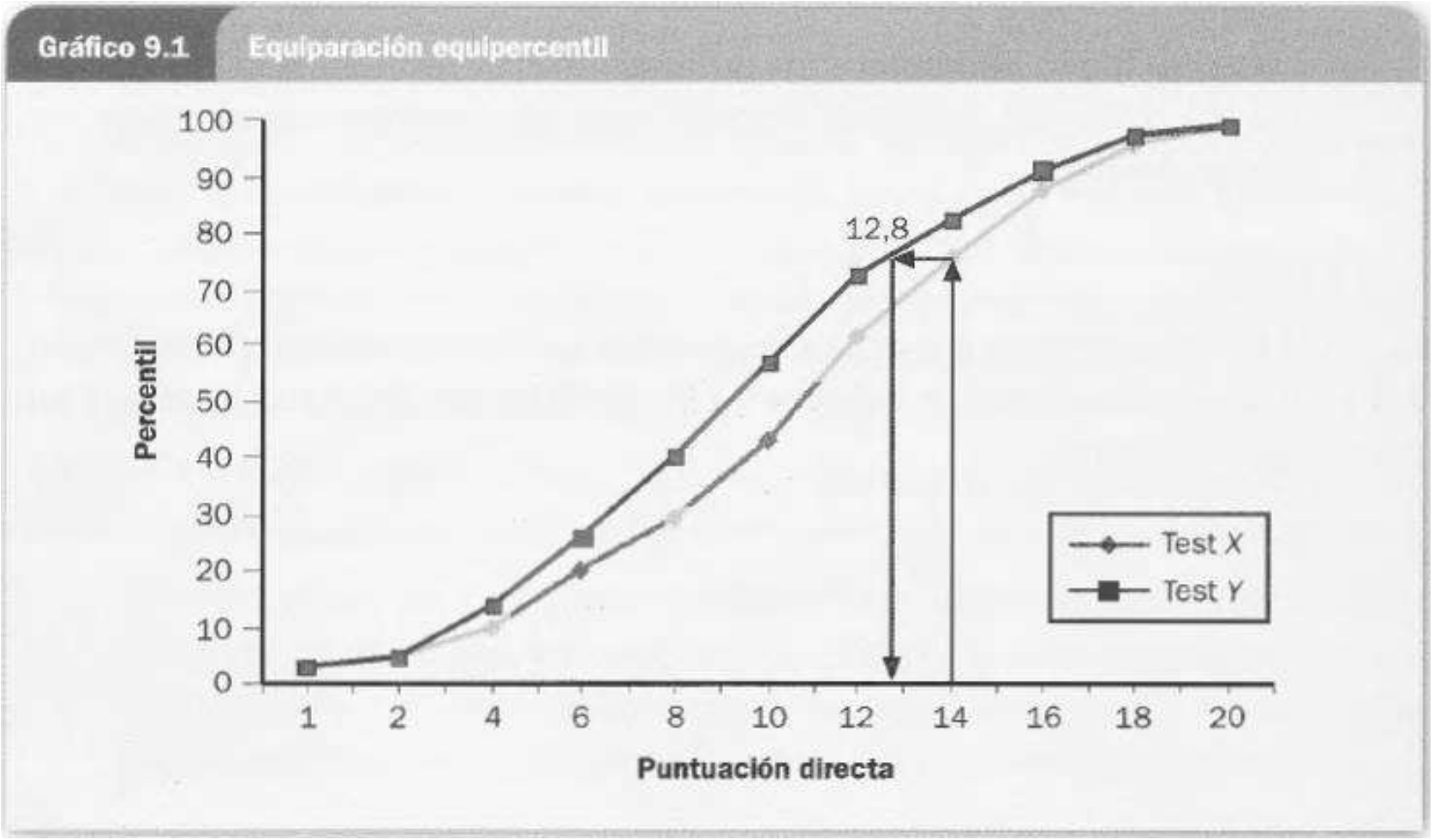
Según Crocker y Algina (1986), los pasos a seguir para llevar a cabo el proceso de equiparación percentil, se pueden resumir en los siguientes apartados:

Tenemos dos tests *X* e *Y,* cuyas puntuaciones queremos equiparar.

1. Calculamos en cada test las puntuaciones percentiles que corresponden a cada una de las puntuaciones de ambos tests. Para calcular dichas puntuaciones aplicamos la ecuación de las puntuaciones percentiles (k)
2. Representamos gráficamente las dos distribuciones de percentiles.
3. Obtenemos las puntuaciones equivalentes en los dos tests *X* e *Y* a partir del gráfico anterior.

MÉTODOS DE EQUIPARACIÓN (VI)

MÉTODO EQUIPERCENTIL (II)



ERROR TÍPICO DE EQUIPARACIÓN

DISEÑO DE GRUPOS EQUIVALENTES

$$S_e = \sqrt{\frac{2S_y^2}{N_1 + N_2}(Z_x^2 + 2)}$$

DISEÑO DE UN SOLO GRUPO

$$S_e = \sqrt{\frac{(S_y^2)(1 - r_{xy})\left(Z_x^2(1 + r_{xy}) + 2\right)}{N}}$$

DISEÑO DE ANCLAJE

$$S_e = \sqrt{\frac{2S_y^2(1 - r^2)\left[Z_x^2(1 + r^2) + 2\right]}{N}}$$

FIN DEL TEMA 9

Muchas gracias por vuestra atención