

CONFIABILIDAD Y VALIDEZ

SANTIAGO AGUDELO MEJIA Y HERNANDO RAMIREZ ARDILA

I—EL CONCEPTO DE CONFIABILIDAD DE MEDIDAS

Las pruebas que utilizamos en la medición psicológica nos permiten una descripción cuantitativa de los individuos en la medida en que éstos poseen varios rasgos y habilidades psicológicas. Ordinariamente estamos interesados en que estas descripciones cuantitativas o puntajes nos permitan hacer comparaciones interindividuales respecto de una característica como dentro de los individuos en diferentes rasgos; también para predecir otros tipos de conducta y estimar los efectos de varios factores en la conducta individual. Como consecuencia, cuando se evalúa a un individuo se espera obtener un puntaje que nos indique una caracterización precisa de él.

Si administramos la misma prueba varias veces al mismo individuo, encontramos que no obtiene exactamente el mismo puntaje en todas las aplicaciones. Algunas veces los puntajes cambian en forma sistemática, aumentando o disminuyendo o fluctuando regularmente de manera cíclica. En otras ocasiones los puntajes obtenidos por un individuo parecen fluctuar al azar y en forma no sistemática. Por ejemplo, cuando medimos la altura de una persona varias veces con la misma regla, la primera medida obte-

nida es de 1.76 metros y en sucesivas medidas podemos obtener alturas de 1.86; 1.79; 1.80. En forma similar, cuando una prueba psicológica se administra a un individuo en ocasiones diferentes, puede obtener puntajes de 73, 75, 79 y 81.

Si sucede esto último nos indica que no podemos confiar en un simple puntaje obtenido por un individuo, ya que en otra aplicación de la misma prueba él puede obtener un puntaje diferente. Es en esta observación de la variación no sistemática en los puntajes obtenidos por un individuo en aplicaciones repetidas de la misma prueba, en lo que se basa el término "confiabilidad de medidas". Nuestro objetivo en este caso es considerar los tipos de variación que se presentan en los puntajes obtenidos por un individuo en una prueba dada y formalizar el concepto de confiabilidad de medidas. Solamente si nosotros formalizamos este concepto podremos aplicarlo en forma teórica y práctica.

II—IMPORTANCIA DE LA VARIACION NO SISTEMATICA EN LOS PUNTAJES DE LAS PRUEBAS

Cuando un individuo obtiene diferentes puntajes en administraciones sucesi-

vas de la misma prueba y la variación de ellas no es sistemática, podemos decir que las descripciones cuantitativas dadas por la Prueba no son confiables. La confiabilidad de las medidas, por lo tanto, se refiere al grado de consistencia entre los puntajes obtenidos por un individuo. Si los puntajes suministrados por la prueba en un individuo difieren notablemente de una ocasión a otra, tienen muy poco valor. Tal prueba no nos da puntajes confiables o consistentes y por lo tanto podemos predecir solamente con un grado bajo de seguridad los puntajes que un individuo obtendrá en cualquier otra administración de la prueba a partir de una anterior. Los puntajes no confiables son de poco valor cuando deseamos comparar dos o más individuos con la misma prueba, para clasificar individuos en grupos o clases, para predecir otros tipos de conducta, para comparar diferentes rasgos o habilidades de un individuo, o para evaluar el efecto de varios factores sistemáticos que afectan una conducta individual. Vamos a considerar ejemplos de cada uno, en tal forma que podamos ver la importancia de la confiabilidad de las medidas.

1. Confiabilidad y comparación de individuos con la misma prueba.

Algunas veces deseamos conocer si una persona es superior a otra en los rasgos o habilidades medidos por una prueba específica. Si sabemos que las personas varían en sus puntajes de una aplicación a otra de la prueba por más de 10 puntos y la diferencia entre los puntajes es de 30, probablemente estaríamos en capacidad de concluir que una persona es realmente superior a otra y que la diferencia podría mantenerse si les aplicamos la prueba en otra ocasión. Podríamos conocer, al administrar la Prueba a las dos personas por segunda vez, que el individuo que fue superior en la primera ocasión indudablemente deberá ser superior en la segunda, si quien obtuvo el puntaje mayor en la pri-

mera aplicación obtiene el más bajo en la segunda, y quien obtuvo el más bajo en la primera, en la segunda obtiene el más alto; la diferencia será de 10 puntos como mínimo. Si por el contrario la diferencia entre las dos aplicaciones fuera de 20 puntos, no podríamos predecir seguramente quién será superior. El grado en que estamos en capacidad de verificar la diferencia entre los puntajes obtenidos por dos individuos en una prueba como reflejo de una diferencia real o estable entre ellos, en el rasgo que está siendo medido por la prueba, está en función directa de la confiabilidad de esa prueba.

2. Confiabilidad y clasificación de individuos en grupos o clases.

Supongamos que los alumnos de una escuela deben ser clasificados en grupos de matemáticas sobre la base de los puntajes que ellos han obtenido en una prueba de matemáticas. Con puntajes de 60 o más se asignarán a cursos acelerados, con 50 a 59 en cursos intermedios y aquellos con puntajes de 49 y menores a cursos inferiores. Suponiendo que la variación en puntajes de un individuo sea de 6 puntos en repetidas aplicaciones y que un alumno haya obtenido 55 puntos en la prueba, se le asignaría al grupo intermedio. Sin embargo, si se hubiera aplicado la prueba en otra ocasión pudiera haber obtenido un puntaje tan alto como 61 y se hubiera clasificado en el grupo superior, o pudiera haber obtenido un puntaje tan bajo como 49 y se le habría clasificado en el grupo inferior. Es por lo tanto obvio que el grado de confiabilidad de la capacidad de medición de esta prueba es suficiente para clasificar con bastante certeza a los alumnos en secciones.

Por otra parte, si la variación entre los puntajes en administraciones sucesivas de la prueba es solamente de un punto, una amplia proporción de alumnos puede clasificarse con un alto grado de seguridad. En consecuencia puede

verse que la confiabilidad de las medidas es importante en la consideración de los términos de precisión con los cuales los individuos pueden ser clasificados en grupos o clases.

3. *Confiabilidad y predicción.*

Los puntajes de pruebas se utilizan frecuentemente para predecir otros tipos de conducta.

Por ejemplo, los puntajes en una prueba de inteligencia se utilizan generalmente para predecir éxito en los estudios académicos. Si se emplea una prueba de inteligencia para hacer predicción de esta clase, y es muy poco confiable, para los puntajes obtenidos por un individuo en una ocasión, puede anticiparse un alto grado de éxito escolar, pero a partir de los puntajes que él haya obtenido en otra ocasión puede sacarse la conclusión justamente opuesta. Esto podría, por lo tanto, ser obstáculo en ciertas circunstancias para predecir con certeza satisfactoria. En consecuencia se puede apreciar que la seguridad de predicción de una variable a otra está limitada por el grado de confiabilidad con el cual estas variables hayan sido medidas.

4. *Confiabilidad y comparación de rasgos y habilidades de un individuo.*

En ciertas circunstancias es necesario conocer en cuál de los rasgos un individuo es superior. Por ejemplo, como ayuda para aconsejar a un estudiante podríamos desear conocer si es superior en aptitud mecánica o de oficina. Supongamos que la variación de esta prueba en puntajes estandarizados es de 20 puntos de una aplicación a otra y que las administramos a un estudiante en una sola ocasión encontrando que en aptitud mecánica su puntaje es 85 y en aptitud de oficina es de 70; en este caso no podemos decir con mucha certeza que su aptitud mecánica es superior a su aptitud de oficina. Sin embargo, si la varia-

ción de puntajes en repetidas aplicaciones de las pruebas fuera solamente 5 puntos, estaríamos más seguros de sacar esta conclusión. Por lo tanto la confianza que hemos depositado en las diferencias entre los puntajes obtenidos por un individuo en diferentes pruebas, está en función del grado de confiabilidad de estas pruebas.

5. *Confiabilidad y evaluación de los efectos de factores sistemáticos en la administración.*

Si deseamos averiguar si un cambio en el resultado de la administración de una prueba, respecto de una segunda administración, pueda deberse a algún factor sistemático que se había introducido en la administración o si es mejor considerarlo como resultado de una variación no-sistemática que ocurre entre repetidas aplicaciones de las pruebas. Si conocemos el grado de variación no-sistemática y encontramos que es mayor que los cambios en los puntajes, no estamos en capacidad de decir que estos cambios sean resultado de factores introducidos. Por otra parte, si la cantidad de variación no-sistemática es menor que la cantidad por la cual cambian los puntajes individuales, podemos concluir con considerable confianza que los cambios en los puntajes son el resultado de los factores introducidos.

Supongamos que cuando administramos una prueba un número de veces a un individuo bajo las mismas condiciones, sus puntajes varían alrededor de 20 puntos. Así, en administraciones posteriores de la misma prueba una persona puede obtener un puntaje tan bajo como 50 y tan alto como 70, y otra persona puede obtener un puntaje tan bajo como 80 y tan alto como 100. Si administramos la prueba a una persona, sometiéndola a un tratamiento, por ejemplo a una instrucción especial, y al aplicarle de nuevo la prueba observamos que su puntaje aumenta 15 puntos, podríamos dudar de si el aumento se debe al trata-

miento, pues cambios de tal magnitud ocurren frecuente y justamente en el curso normal de acontecimientos cuando no hay cambios especiales en las condiciones de aplicación. Sin embargo, si la variación en las diversas aplicaciones de la prueba, cuando no hay variaciones especiales, es solamente de 5 puntos, se podría suponer que el tratamiento fue benéfico.

6. *Variación sistemática y no-sistemática en los puntajes de las pruebas.*

Como hemos visto, hay dos tipos principales de variación en los puntajes obtenidos por un individuo en repetidas aplicaciones de una prueba. Uno lo hemos denominado variación sistemática y el otro no-sistemática. Debemos examinar ambos tipos de variación en tal forma que podamos diferenciarlos para desarrollar el concepto de confiabilidad de medidas.

Variación sistemática de puntajes por oposición a la variación no-sistemática.

La variación sistemática está caracterizada por un orden progresivo o por un patrón como cuando los puntajes obtenidos por un individuo cambian de una ocasión a otra siguiendo cierto orden determinado. Los cambios aparecen como decremento o incremento regular, o pueden seguir algún ciclo. La variación no-sistemática, por otra parte, está caracterizada por una completa ausencia de orden. Los puntajes de un individuo fluctúan de una ocasión a otra en completo azar.

Los dos tipos de variación por supuesto, pueden ocurrir simultáneamente. Así, encontramos que el promedio de puntajes de un individuo aumenta cuando se le hacen repetidas aplicaciones de la prueba y alrededor de esta tendencia su puntaje parece variar al azar, algunas veces por encima del aumento y otras por debajo.

7. *Procedimientos para estimar la confiabilidad.*

La evaluación de la confiabilidad de un instrumento de medida requiere una determinación de la consistencia absoluta o relativa de medidas repetidas del mismo objeto o grupos de objetos. En las ciencias físicas muchas repeticiones de la medida de un solo objeto o fenómeno nos dará el método razonable para estimar la precisión de las medidas al tener como objeto la conducta humana; sin embargo, el individuo probablemente será alterado como resultado de la operación de medida, y esto hará necesaria la limitación del número de veces que un individuo es sometido a medición. En la práctica, por lo tanto, todos los procedimientos de la determinación de la confiabilidad utilizados en la investigación personal se basan en adquirir un número pequeño de medidas, por lo general dos para cada individuo dentro de un grupo representativo. La estabilidad de los resultados se logra por aumento del número de individuos medidos más que por el número de mediciones hechas en cada uno de ellos. Esta medida nos da un conjunto de puntajes, generalmente dos para cada individuo con el objeto de efectuar el análisis. El análisis general consiste en el cálculo del coeficiente de correlación entre los conjuntos de puntajes.

Hemos definido el coeficiente de confiabilidad como la correlación entre dos conjuntos de medidas equivalentes de una característica para un grupo de individuos. A continuación vamos a considerar las operaciones que debemos ejecutar de modo que correspondan satisfactoriamente a los requisitos lógicos de las pruebas equivalentes. En seguida consideraremos los mejores procedimientos estadísticos y experimentales, describiendo y evaluando cada uno en términos del tratamiento de las diferentes categorías de variación. Dichos procedimientos son los siguientes:

1. Administración de dos pruebas equivalentes y correlación de los puntajes obtenidos.

2. Administración de la misma forma de la prueba o administración doble y correlación de los puntajes obtenidos.

3. Subdivisión de una sola prueba en dos grupos de preguntas, calificando cada parte por separado, y correlacionando los dos puntajes obtenidos.

4. Por medio del análisis de la variación entre las preguntas de una prueba, y determinando el error de la variación a partir del análisis.

1. *Confiabilidad definida por formas equivalentes de pruebas.*

Las formas paralelas de una Prueba son una serie de dos o más pruebas, que tienen el mismo contenido o naturaleza. Dos pruebas objetivas que tengan la misma clase y número de preguntas son formas paralelas. Una pregunta de forma paralela en una prueba de aritmética podría ser $27 + 84 = \dots$, y otra pregunta de otra forma paralela podría ser $48 + 72 = \dots$. En un inventario de estabilidad emocional dos preguntas paralelas podrían ser: “¿Duerme bien usted durante la noche?” y “¿Tiene usted pesadillas durante la noche?”.

Las formas paralelas de una prueba no deben confundirse con pruebas paralelas ya que las primeras son similares en contenido y naturaleza y diseñadas para medir los mismos rasgos; en cambio las pruebas paralelas son aquellas que cumplen con ciertos criterios estadísticos, tales como poseer los mismos promedios y desviaciones estándar o correlacionarse ambas pruebas en igual grado y el mismo nivel de correlación con cualquiera otra variable. Cuando se han utilizado dos formas paralelas tomamos como estimación de la confiabilidad la correlación entre los puntajes de ellas; si hay más de dos, la práctica común es tomar el promedio de las in-

tercorrelaciones como estimación del coeficiente de confiabilidad.

2. *Confiabilidad determinada a partir de la repetición de la misma prueba.*

Este método consiste en la administración de la prueba dos o más veces al mismo grupo de individuos; las intercorrelaciones entre los puntajes de las diversas aplicaciones se toman como coeficiente de confiabilidad. Hay dos ventajas principales con este método “test-retest”: a) no se necesita más que una forma de la prueba, y b) el instrumento de medida permanece constante. La más seria desventaja de este método está en la variedad de efectos que conlleva el aplicar una prueba en dos ocasiones.

3. *Confiabilidad determinada a partir de partes comparables de una prueba.*

En muchas situaciones los individuos no pueden ser sometidos a pruebas sino una sola vez y en consecuencia no se dispone sino de una sola aplicación. En tales circunstancias podemos obtener una estimación de la confiabilidad si consideramos la prueba no como una sola sino como la suma total de un número de formas paralelas de ella. Supongamos, por ejemplo, que tenemos una prueba objetiva de 100 preguntas, todas ellas pertenecientes al mismo rasgo. En vez de decir que tenemos una prueba de 100 preguntas, digamos que tenemos dos de 50 o cuatro de 25. Teniendo dos o más formas paralelas utilizables podemos proceder a estimar la confiabilidad por el método de correlación de puntajes en formas paralelas. Nótese que no podemos obtener la confiabilidad de una prueba de 100 preguntas, sino la de una más corta. Si la hemos dividido en dos partes de 50 preguntas cada una, hemos obtenido solo la confiabilidad de una prueba de 50 preguntas.

Si dividimos nuestra prueba en varias pruebas más cortas, encontramos que la confiabilidad de éstas será adecuada

para ellas; sin embargo puede desearse medir el rasgo con la prueba total para lograr una mayor precisión, para lo cual se empleará la fórmula de Spearman-Brown:

$$r_{nn} = \frac{n r_{11}}{1 + (n - 1) r_{11}}$$

En donde:

r_{11} es la correlación entre las partes, y n es el número de veces que es más grande la prueba total que las partes. Por medio de esta fórmula podremos estimar, a partir de la intercorrelación entre las partes, la confiabilidad de la prueba total.

El procedimiento más común es dividir la prueba en dos partes y a partir de la correlación entre las dos estimar la confiabilidad de la prueba total, por medio de la fórmula de Spearman-Brown mitad-mitad, que es la siguiente:

$$r_{11} = \frac{2 r_{\frac{1}{2} \frac{1}{2}}}{1 + r_{\frac{1}{2} \frac{1}{2}}}$$

4. Análisis de Varianza entre ítems.

Varios investigadores han desarrollado diversos procedimientos para obtener toda la información acerca de la consistencia en la solución de una pregunta a otra dentro de una prueba y lograr así una estimación de la consistencia interna. Los procedimientos y fórmulas básicas fueron presentados por Kuder y Richardson; posteriores modificaciones basadas en suposiciones menos restrictivas, fueron realizadas a la fórmula Kuder-Richardson número 20 por Jackson y Ferguson.

Este cálculo general da un tipo de estimación de la confiabilidad análogo al obtenido a partir de la subdivisión de un test y tiene muchas de las mismas características y limitaciones. En particular estos procedimientos no se aplican a pruebas que incluyan la velocidad y que tengan un tiempo limitado. Se supone implícitamente en el método,

que los individuos deben abordar cada una de las preguntas. Las características de las preguntas, tales como la dificultad, varianza e intercorrelación, llegan a ser de poco significado cuando un apreciable porcentaje de personas del grupo no tiene tiempo de leer y tratar de resolverlas. Por ejemplo, si las preguntas omitidas se consideran como erradas, la intercorrelación de ellas hacia el final de la prueba de velocidad llega a ser toscamente aumentada por el grupo de personas que no las abordaron. Si se calculan las intercorrelaciones, no hay una población uniforme sobre la cual pueda basarse el análisis estadístico. En otras palabras, las mismas dificultades que se tienen con las pruebas de velocidad, y con las pruebas que se dividen en partes, se presentan en el análisis de la consistencia de solución de preguntas. La consistencia de solución no puede ser evaluada a menos que las personas hayan tenido la oportunidad de contestarlas.

La fórmula más utilizada para la estimación de la confiabilidad a partir de la relación de la varianza de la prueba y la varianza de la pregunta, es la fórmula número 20 de Kuder-Richardson. Esta fórmula es la siguiente:

$$r_{tt} = \frac{n}{n-1} \left(\frac{\sigma_t^2 - \sum pq}{\sigma_t^2} \right)$$

Donde:

r_{tt} = confiabilidad de toda la prueba.

n = número de preguntas de la prueba.

σ_t^2 = varianza de toda la prueba.

p = proporción de los que aciertan la pregunta.

$q = 1 - p$.

Kuder y Richardson proponen una fórmula adicional (Fórmula 21) la cual reduce los cálculos a un mínimo. Esta fórmula supone que todas las preguntas tienen la misma dificultad, la fórmula es:

$$r_{tt} = \frac{n}{n-1} \bullet \frac{\sigma_t^2 - n \bar{p} \bar{q}}{\sigma_t^2}$$

donde $\bar{p} = \frac{M_t}{n}$ y $\bar{q} = 1 - \bar{p}$

Por lo tanto los valores que exige esta fórmula son el número de preguntas, el promedio y la desviación estándar de la prueba total. Cuando las dificultades de las preguntas no son completamente iguales, el valor obtenido por esta fórmula será menor y algunas veces bastante más bajo que el que podría obtenerse a partir de la fórmula Kuder-Richardson número 20. Sin embargo, en un número de casos bastante grande y con pruebas de poder la diferencia entre las dos fórmulas no es mayor que .05.

III—VALIDEZ Y VALIDACION

Como en el caso de la confiabilidad, el término validez significa diferentes hechos para diferentes personas.

Cuando se persigue una variedad de propósitos tanto teóricos como prácticos, para los cuales los individuos se someten a mediciones psicológicas, no es sorprendente encontrar diversas definiciones del término validez. Cada uno de los propósitos han conducido a diversas nociones de lo que es la validez, y en consecuencia, a desarrollar diferentes procedimientos para determinarla. Estos procedimientos a su vez han conducido a la revisión posterior del concepto de validez, con el resultado de que ha habido variaciones en su significado. Por lo tanto, no podemos definir la validez de una sola manera. Lo que podemos hacer, como lo hemos hecho con el problema de la confiabilidad, es describir más o menos los principales puntos de vista.

Tal vez la definición más común del término de validez, es la que se refiere al grado con el cual una prueba o con-

junto de operaciones miden lo que se supone que miden. Sin embargo, en algunos casos la definición del rasgo puede no ser totalmente clara y sería difícil asegurar la validez de un conjunto especial de operaciones. En otros casos, cuando parece obvio que las operaciones o pruebas miden completamente el rasgo incluido en la definición, puede suceder que también se midan otras propiedades que no estén incluidas en ella. Pero si estas circunstancias se presentan, no podemos concluir que las operaciones sean completamente inútiles. Más bien, esto significa simplemente que las propiedades medidas son diferentes de las que se habían definido. Las operaciones o pruebas pueden tener menos valor que el que se suponía, o pueden adquirir un gran valor práctico debido a la información adicional que ellas dan.

En consecuencia, con respecto a la validez prefieren hacer la pregunta "¿Qué rasgos miden las pruebas u operaciones?"

1. Validez Predictiva.

La validez predictiva se refiere a la seguridad con la cual nosotros podemos hacer suposiciones sobre cierta característica de una persona a partir de otra característica. Así, conociendo el grado en que una persona manifiesta o posee alguna propiedad, se puede estimar el grado en que manifiesta o posee otra. Con la validez predictiva tratamos de determinar el grado en el cual un conjunto de operaciones o pruebas pueden sustituirse por otras. El hecho más importante de la validez predictiva es que hay una medida de la propiedad que se considera como perfecta —el criterio—, y se buscan los puntajes estimados de ella a partir de los puntajes obtenidos en otra medida —el predictor—. El término predicción, como se utiliza aquí, no implica que estemos solamente interesados en la predicción de hechos futuros, sino que lo que se nos ha dado es anterior a lo que queremos conocer. La

predicción puede basarse en hechos futuros, presentes o pasados.

2. Descripción de la validez predictiva.

La validez predictiva de una prueba está descrita de manera objetiva y cuantitativa por el grado de relación entre los puntajes del predictor y los del criterio. Por lo tanto el coeficiente de correlación de Pearson o sus variantes, tales como el biserial, el punto biserial, el Phi o el tetracórico se usan ordinariamente para describir el grado de validez predictiva. Cuando estos coeficientes se utilizan para indicar el grado de relación entre los puntajes del predictor y el criterio, se denominan coeficientes de validez.

Es obvio que la relación entre dos variables puede indicarse por otras maneras diferentes del coeficiente de correlación. Así, por ejemplo, la diferencia entre los promedios de las variables X y Y puede indicarnos algo de la relación entre las dos variables, pero el coeficiente de correlación, como coeficiente de validez, se usa más generalmente porque da la descripción más exacta del grado de relación en la forma sencilla y conveniente.

3. Diversas situaciones predictivas.

La validez predictiva describe la precisión con la cual podemos estimar, a partir del grado en el cual un individuo manifiesta o posee una propiedad actualmente, el grado en que el mismo individuo manifestará en el futuro alguna otra propiedad o es manifestada o poseída por él en el presente, o fue manifestada o poseída en el pasado. Todos estos tres tipos de situaciones predictivas suponen la relación entre los puntajes del predictor y el criterio. Solamente difieren en los términos del tiempo de ocurrencia de los puntajes del criterio. En las tres situaciones, cuanto más alta sea la relación entre los puntajes del predictor y el criterio, más alta será la validez,

esto es, con más precisión pueden los puntajes del criterio estimarse a partir de los del predictor. Cuando las pruebas se utilizan para predecir un éxito futuro se denominan pruebas de aptitudes. Estas se emplean para estimar el nivel de eficiencia que un individuo probablemente posee en alguna ocupación, o en su eficacia en alguna situación educativa o de entrenamiento. Por ejemplo, decimos de un individuo que tiene una baja aptitud de oficina, lo cual significa que si él realiza un trabajo de oficina, sus oportunidades de éxito son pequeñas; y de otros individuos que su aptitud académica sea alta, significa que si va a la Universidad probablemente obtendrá notas altas.

Algunas veces estamos interesados en la precisión con la cual pueda estimarse el grado de un rasgo actual en un individuo, a partir del grado que él posee en algún otro rasgo. Utilizamos los puntajes de una variable para estimar los puntajes de la otra; ambas variables miden propiedades presentes del individuo. Puede ser muy difícil, costoso o dispendioso medir una propiedad directamente, y por lo tanto preferimos utilizar algún otro procedimiento para obtener descripciones cuantitativas de esa propiedad. Este tipo de validez es lo que algunos autores llaman "Validez Concurrente".

4. Métodos para calcular la validez de una sola Prueba.

Para calcular la validez de una prueba, se utilizan diferentes coeficientes de correlación según la naturaleza de las variables, así:

a) Coeficiente de correlación de Pearson.

Este coeficiente es un índice de la inter-dependencia entre dos variables y solo puede emplearse cuando sea conveniente hacerlo, pues hay casos en los cuales esta clase de correlación no puede ser aplicada, y otros en los cuales sí pero por razones prácticas, sea más conve-

niente emplear otros procedimientos. El coeficiente de Pearson no podría ser aplicado a menos que las dos variables X y Y estén medidas sobre la base de una escala métrica continua y su regresión sea lineal.

Para calcularlo se emplea la siguiente fórmula básica de Pearson:

$$r_{xy} = \frac{E_{xy}}{N \sigma_x \sigma_y}$$

En donde:

- x es igual a la desviación de cualquier puntaje X a partir del promedio en la prueba X.
- y es igual a la desviación de cualquier puntaje Y a partir del promedio en la prueba Y.
- E_{xy} es igual a la suma de todos los productos de las desviaciones.
- σ_x y σ_y son iguales a las desviaciones estándar de las distribuciones de los puntajes X y Y.

b) *Coeficiente de correlación biserial.*

El coeficiente de correlación biserial se utiliza cuando ambas variables correlacionadas son continuas, pero una de las dos se ha reducido artificialmente a dos categorías.

Una de las fórmulas empleadas para calcularlo es:

$$b = \frac{M_p - M_q}{\sigma_t} X \frac{pq}{y}$$

Donde:

- M_p es igual al promedio de los valores X del grupo superior en la variable dicotomizada.
- M_q igual al promedio de los valores X para el grupo inferior de la misma variable.
- p igual a proporción de los casos en el grupo alto.
- q proporción de casos en el grupo bajo.

— y igual al valor de la ordenada de la curva normal reducida, correspondiente a estas proporciones.

c) *Coeficiente de correlación punto biserial.*

Cuando una de las dos variables en un problema de correlación está genuinamente dicotomizada, el tipo apropiado de correlación que se utiliza es el punto biserial.

La siguiente fórmula es la que se utiliza para el cálculo de este tipo de correlación:

$$pb = \frac{M_p - M_q}{\sigma_t} X \sqrt{pq}$$

Cuyos símbolos ya son conocidos.

d) *Coeficiente de correlación tetracórico.*

Este coeficiente se calcula a partir de datos en los cuales ambas variables X y Y han sido reducidas artificialmente a dos categorías. En condiciones apropiadas permite obtener un coeficiente que es numéricamente equivalente al r de Pearson y puede considerársele como aproximación de él.

La correlación tetracórica exige que ambas variables sean continuas, que se distribuyan normalmente y que tengan relación lineal.

Una estimación del coeficiente de correlación tetracórico se obtiene por la fórmula:

$$r_{\cos} = \cos \left(\pi \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \right)$$

Siendo a , b , c y d las frecuencias.

e) *Coeficiente de correlación Phi.*

Cuando dos distribuciones correlacionadas están realmente dicotomizadas y las dos mitades están separadas por una división real entre ellas, podemos aplicar el coeficiente Phi.

$$\phi = \frac{ad - bc}{\sqrt{pq p' q'}}$$

En donde a , b , c y d son las proporciones en las cuatro celdas de la tabla de doble entrada, y p , q , p' y q' son los porcentajes en cada categoría de cada una de las dos dicotomías.

5. Método de Aitken de condensación Pivotal para calcular la validez total de dos o más pruebas.

Para la comprensión de este método tomaremos como ejemplo un ejercicio práctico del SENA, denominado el criterio como X_c y a los predictores como X_1 , X_2 , X_3 y X_4 . El criterio fue el promedio de notas obtenidas durante una etapa lectiva por un grupo de aprendizaje en Secretariado de la Sección del SENA de Cundinamarca y los predictores las partes I (Aptitud de Oficina), II (Rapidez de Cálculo), III (Factor Espacial) y IV (Factor Verbal) del GATB.

Los pasos para el cálculo de los coeficientes de regresión y la correlación múltiple son los siguientes:

1. Calcule las intercorrelaciones de los predictores y las correlaciones de cada uno de ellos con el criterio.

	X_c	X_1	X_2	X_3	X_4
X_c	1.000	.2571	.4604	.2586	.3846
X_1		1.000	.2793	.0101	.2921
X_2			1.000	.0036	.3205
X_3				1.000	.3082
X_4					1.000

2. Transcriba la anterior matriz de correlaciones, colocando la hilera de las correlaciones del criterio, en la parte inferior.

3. A la derecha de la matriz transcrita sitúe otra matriz con -1 en toda la diagonal. Todos los otros elementos de la matriz son ceros y están representados por puntos.

METODO DE AITKEN PARA COMPUTAR LOS COEFICIENTES DE REGRESION

	(1)								Control
	.2793	.0101	.2921	— 15815
	1.0000	.0036	.3205	.	— 16034
A	.0101	.0036	1.0000	.3082	.	.	— 1	.	.3219
	.2921	.3205	.3082	1.0000	.	.	.	— 1	.9208
	.2571	.4604	.2586	.3846	1.3607
	(.9220)	.0008	.2390		.2793	— 1	.	.	.4411
	1.0000	.0008	.2591		.3029	—1.0845	.	.	.4783
B	.0008	.9999	.3053		.0101	.	— 1	.	.3161
	.2390	.3053	.9147		.2921	.	.	— 1	.7511
	.3886	.2561	.3096		.2571	.	.	.	1.2114
	(.9999)	.3051			.0099	.0008	— 1	.	.3157
C		1.0000	.3051		.0099	.0008	—1.0001	.	.3157
		.3052	.8528		.2198	.2591	.	— 1	.6369
		.2558	.2090		.1394	.4214	.	.	1.0256
		(.7597)			.2168	.2589	.3052	— 1	.5406
D			1.0000		.2853	.3407	.4017	—1.3163	.7114
			.1310		.1369	.4212	.2558	.	.9449
E					.0996	.3766	.2032	.1724	.8518

Coefficiente de Regresión

4. Sume los valores de las hileras obtenidas y coloque el resultado en la columna de control.

5. Calcule la diferencia entre los productos de la primera matriz (matriz A), utilizando el número (1) de la primera casilla superior como elemento axial (o pivotal). Las siguientes diferencias de productos nos darán por consiguiente los resultados que deseamos encontrar:

$$\begin{aligned} 1 \times 1 & - .2793 \times .2793 = .9220 \\ 1 \times .0036 & - .2793 \times .0101 = .0008 \\ 1 \times .3205 & - .2793 \times .2921 = .2390 \\ 1 \times 0 & - .2793 \times (-1) = .2793 \\ 1 \times (-1) & - .2793 \times 0 = -1 \end{aligned}$$

Estos valores se colocan en la primera hilera de la matriz (B). El valor del control se obtiene al efectuar la diferencia de productos siguientes:

$$1 \times .6034 - .2793 \times .5815 = .4411.$$

Si los cálculos son correctos hasta este momento, la suma de los elementos de la primera hilera de la matriz B será igual a la diferencia de productos encontrada anteriormente o sea .4411.

6. La segunda hilera de la matriz B se obtendrá al dividir cada elemento de la primera hilera de esta matriz por el elemento superior izquierdo de esta misma hilera (.9220), o también multiplicando cada elemento por el recíproco de este número (1.0845). El resultado es una hilera con una unidad como elemento axial (pivotal).

7. Los elementos restantes de la matriz B se forman de las diferencias de

$$\begin{aligned} R &= \sqrt{(.2571 \times .0996) \times (.4604 \times .3766) \times (.2586 \times .2032) \times (.3846 \times .1724)} \\ R &= .5636. \end{aligned}$$

11. Calcule la significación del coeficiente de correlación múltiple. Esta significación se obtiene por medio de la razón F, con base en la siguiente fórmula:

los productos diagonales de la primera hilera de la matriz A con la 3ª, 4ª y 5ª hileras de la misma matriz, utilizando siempre el (1) superior izquierdo de la matriz A como elemento axial. Así:

$$\begin{aligned} 1 \times .0036 & - .0101 \times .2793 = .0008 \\ 1 \times .3205 & - .2921 \times .2793 = .2390 \\ 1 \times .4604 & - .2571 \times .2793 = .3886 \end{aligned}$$

Estos resultados forman la columna de la izquierda de la matriz B.

Para obtener los datos de la segunda columna se efectúa un procedimiento similar al anterior, que es el siguiente:

$$\begin{aligned} 1 \times 1 & - .0101 \times .0101 = .9999 \\ 1 \times .3082 & - .2921 \times .0101 = .3053 \\ 1 \times .2586 & - .2775 \times .0101 = .2561 \end{aligned}$$

Los datos de las columnas restantes se obtienen en forma similar a la anterior.

8. El procedimiento se repite para obtener las matrices C, D y E. En cada etapa, para obtener la segunda hilera de cada matriz a excepción de la última, la hilera superior se divide por el valor izquierdo de dicha hilera o se multiplica por su recíproco.

9. Al efectuar los cálculos, la matriz original se condensa en una matriz de una sola hilera (Matriz E). Los cuatro valores de la matriz E son los coeficientes de regresión múltiple.

10. Multiplicando los coeficientes de correlación entre el criterio y los predictores por los coeficientes de regresión respectivos; sumando estos productos y extrayendo la raíz cuadrada, obtenemos el coeficiente de correlación múltiple, el cual será en nuestro ejemplo:

$$F = \frac{R^2}{1 - R^2} \times \frac{N - (k - 1)}{k}$$

Donde R es la correlación múltiple; N número de observaciones; k número

de variables independientes o predictores.

Para consultar la tabla de F $df_1 = k$ y $df_2 = N - (k - 1)$.

En nuestro ejemplo los cálculos serán los siguientes:

$$F = \frac{.3177}{1 - .3177} \times \frac{28 - (4 - 1)}{4} = 2.9101$$

El cual es significativo al nivel del 5%.

BIBLIOGRAFIA

1. ANASTASI, ANNE: *Psychological Testing*. New York: Macmillan Co., 1961. 2ª Ed.
2. FAVERGE, J. M.: *Méthodes Statistiques en Psychologie Appliquée*. Paris: Press Universitaires de France, 1965. I, II y III.
3. FERGUSON, A. GEORGE: *Statistical Analysis in Psychology and Education*. New York: McGraw Hill: 1959.
4. GHISELLI, E. EDWIN: *Theory of Psychological Measurement*. New York: Mac-Graw Hill: 1964.
5. GUILFORD, J. P.: *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill: 1956, 3ª Ed.
6. GUILFORD, J. P.: *Psychometric Methods*. New York: McGraw Hill: 1954, 2ª Ed.
7. LINDLQUIST, E. F.: *Educational Measurement*. Washington, American Council on Education: 1961, 4ª Ed.
8. OLERON, PIERRE: *Les Composantes de l'intelligence*. Paris: Press Universitaires de France, 1957.
9. REUCHLIN, MAURICE: *Les Méthodes Quantitatives en Psychologie*. Paris: Press Universitaires de France, 1962.
10. THOMSON, H. GODFREY: *L'analyse Factorielle des Aptitudes Humaines*. Trad. Pierre Naville. Paris, 1950. Press Universitaires de France, 1ª Ed.
11. THORNDIKE, L. ROBERT: *Personnel Selection*. New York: John Wiley & Sons, Inc.: 1962.