UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

PLANTEAMIENTO TEORICO ACERCA DEL RENDIMIENTO ESCOLAR APOYADO EN TECNICAS ESTADISTICAS

TESIS
Que para obtener el Título de:
ACTUARIO
Presenta:
Patricio Alberto Rosen Robles

México, D. F. 1988

INDICE

INTRODUCCION

El rendimiento escolar es y ha sido estudiado desde diversos enfoques: sociológicos, psicológicos, filosóficos, por mencionar algunos; sin embargo, muchos de los planteamientos que se establecen acerca del rendimiento escolar necesitan fundamentarse en aspectos cuantificables del mismo, es decir, el rendimiento escolar debe ser medido. La forma más generalizada y quizá la mejor de medir el rendimiento escolar es por medio de la aplicación de exámenes o pruebas. Sin embargo, estos exámenes o pruebas no siempre miden verazmente el rendimiento escolar, pues existen características que deben satisfacer para hacerlo, como son principalmente la confiabilidad y la validez que son definidas ampliamente en este trabajo y que pueden medirse empleando procedimientos estadísticos.

Entre las intenciones de este trabajo, no se encuentra la de indicar a los docentes la forma de elaborar pruebas adecuadas o perfectas, sino que más bien permite la posibilidad de que los mismos se percaten por medio de la aplicación de procedimientos estadísticos, si los instrumentos de evaluación que diseñan son confiables y válidos.

Para lograr el propósito mencionado, se definen previamente los elementos conceptuales que integran el marco teórico de este trabajo, entre los que se cuentan fundamentalmente los siguientes:

Rendimiento escolar, naturaleza de la medición, medición y evaluación, pruebas y su clasificación, confiabilidad y validez.

Una vez que se han definido estos elementos se desarrolla el planteamiento teórico sustancial, que consiste en explicitar los modelos matemáticos y los procedimientos estadísticos utilizados para medir la confiabilidad y la validez de las pruebas que se emplean en la evaluación del rendimiento escolar.

Asimismo se plantean y resuelven algunos ejemplos que ilustran la aplicación de los modelos y procedimientos que se señalan.

I. CONCEPTOS GENERALES SOBRE RENDIMIENTO ESCOLAR, MEDICION Y EVALUACION.

¿ QUE ES EL RENDIMIENTO ESCOLAR?

Para responder esta interrogante, es conveniente citar previamente algunas de las definiciones que sobre el concepto proporcionan algunos estudiosos en la materia.

Thorndike y Hagen definen al rendimiento escolar como los resultados diferenciales de aprendizaje que cada estudiante muestra en comparación con los que muestran sus condiscípulos. (Thorndike, 1973, 9 - 17).

Se refieren por supuesto a aprendizajes que son medidos por medio de la aplicación de pruebas.

Galán y Marín lo definen en la forma siguiente:

"Es la expresión de la calidad del proceso de vida académica de los alumnos en una escuela y que está proyectado a la formación del alumno en una práctica profesional determinada" (Galán, 1983 27).

A pesar de que la definición de Galán está claramente referida al nivel profesional, podría de manera análoga aplicarse al campo de estudios propedeúticos, considerando que los esfuerzos del proceso de vida académica de los alumnos está proyectado a la preparación del alumno para realizar estudios mas complejos e incluso para la vida, como lo señalan las nuevas Pedagogías.

Una de las maneras en que puede estudiarse al rendimiento escolar es en base al análisis de los alcances académicos, considerando que su estudio se centra en la determinación de las variables que mayor influencia tienen en el éxito o fracaso escolar, determinado en términos del logro de las metas -- académicas; ya sea a partir de calificaciones obtenidas en un - curso o del logro o alcance de los objetivos de aprendizaje prefijados.

En este sentido, puede definirse al rendimiento escolar de la manera siguiente: ES EL NIVEL ALCANZADO POR UN ESTUDIANTE EN EL LOGRO DE LOS OBJETIVOS DE UN PROGRAMA DE ESTUDIOS DETERMINADO.

Por extensión, puede hablarse también del rendimiento escolar de un grupo, instituciónо sistema.

El problema estriba ahora en medir ese nivel alcanzado de la mejor manera posible, es decir, en medir eficientemente el rendimiento escolar, para lo cual es preciso que se defina - previamente el concepto de medición y que se establezca su naturaleza así como su diferenciación con el concepto de evaluación con el que frecuentemente es confundido.

IDEAS GENERALES SOBRE LA MEDICION

ASPECTOS GENERALES SOBRE TEORIA DE LA MEDICION.

El principal objetivo de la ciencia más que la simple descripción de los fenómenos empíricos es establecer mediante leyes y teorías los principios generales con los cuales se puedan explicar y pronosticar dichos fenómenos.

Para poder alcanzar este propósito, las ciencias recogen y comparan datos para obtener las correlaciones y ecuaciones matemáticas que son el fin de su búsqueda. Lo que hace posible este proceso es la MEDICION.

Se puede decir que la MEDICION permite que el instrumental matemático se aplique a la ciencia con buenos resultados.

Las distintas ciencias que se postulan como tales, difieren unas de otras de diversas formas. Estas formas distintas hace que se clasifiquen de manera distinta. Torgerson (1925, --40)señala la clasificación de Morgenau que divide a las ciencias de acuerdo al tipo de explicación que proporcionan, proponiendo lo siguiente:

TIPO DE CIENCIA	DEFINICION	EJEMPLOS
CORRELACIONALES	Estas ciencias están constituídas principalmente por afirmaciones que describen el grado de relación entre variables más o menos observables directamente.	Biología Psicología Pedagogía

DE EXPLICACION TEORETICA.	Estas ciencias tratan de explicar o derivar las relaciones entre las variables observables por principios no dados en forma inmediata, sino que subyacen al conocimiento estrictamente empírico.	Física Química

Es conveniente aclarar que ninguna ciencia es totalmente correlacional o teorética y, precisamente el grado en que se apoyan en uno u otro nivel de explicación es lo que las diferencia. Del mismo modo si clasificáramos a las ciencias apoyándonos en el grado de obtención de una MEDICION satisfactoria, - las clasificaríamos del mismo modo anterior ya que, el desarrollo de una ciencia teorética en el sentido en que nos la explica Morgenau, no sería posible si no se pudieran MEDIR sus variables en forma adecuada.

Una de las diferencias más importantes entre las --- ciencias sociales y de la conducta (p.ej. Psicología y Pedagogía por un lado, y la Física y Química por el otro) está precisamente en los procedimientos para MEDIR sus propiedades más importantes.

SISTEMAS, PROPIEDADES Y MEDICION.

Las construcciones científicas se pueden dividir en dos tipos principales: Sistemas y Propiedades. Entre estos dos tipos siempre existirá una relación muy estrecha. Los sistemas pueden ser p. ej. edificios, plantas, papel, o también los átomos, genes, rayos de luz, etc. (las cosas de la vida cotidiana)

Las propiedades son aspectos observables del mundo empírico, -- p. ej. el peso, la longitud, el calor, masa y carga de un electrón, coeficiente intelectual, índice de aprovechamiento, etc.

Es importante asentar aquí que lo que se MIDE son -- siempre las propiedades y no los sistemas.

Los sistemas no se pueden medir, aunque sí se puedan clasificar o contar. Las propiedades, también se pueden clasificar, pero la mayoría de los conceptos de clasificación que -- trabajan con propiedades, resulta más adecuado expresarlos cuantitativamente. Las ventajas de MEDIR las propiedades, son señalados por Hempel (1986, 70), y se mencionan a continuación:

1). Con conceptos métricos es posible diferenciar - situaciones que en una clasificación simple se tendrían que colocar juntas.

2). El concepto cuantitativo establece una posición relativa y, por ende, una idea de ordinalidad.

3). Permite que la formulación de las leyes tengan una mayor flexibilidad.

4). La introducción de términos métricos hace posible la aplicación de conceptos y teorías matemáticas y expresa leyes generales presentadas como relaciones funcionales entre cantidades distintas.

Por otro lado, la forma en que se manifiestan las -- ventajas de definir en términos de MEDICION las propiedades de los sistemas en las funciones de la ciencia son:

A). DESCRIPCION: En esta función de la ciencia, al considerar el número de clases como prácticamente ilimitado, aumenta la flexibilidad descriptiva.

B). INFORMACION; En cuanto a esta función, nos dice bastante sobre las relaciones que existen entre las diversas cantidades de la propiedad.

C). EXPLICACION: Al permitir una formulación más precisa de las leyes generales, relaciona distintos contenidos y posibilita que el sistema matemático se extienda y se aplique a la ciencia. Con todo lo anterior, la ciencia puede utilizar las relaciones funcionales entre las distintas construcciones que resulten.

EL METODO CIENTIFICO, SUS PASOS Y LA IMPORTANCIA DE LA MEDICION

Para poder afirmar que se ha llegado a la solución científica de un problema, se suele proceder utilizando dos aspectos de la lógica: el inductivo y el deductivo. Si bien es cierto que se acostumbra comenzar por un proceso inductivo la observación de cualquier fenómeno, se suele terminar demostrando la evidencia que justifica una conclusión. El método científico de investigación experimental debe su fuerza al método lógico inductivo; en cambio, la comprobación de la validez de los resultados de la experimentación recibe su valor de la lógica deductiva. El vínculo de ambas metodologías nos permite presentar los cuatro pasos fundamentales de toda investigación científica:

1) OBSERVACION. Este primer paso lo utilizan los científicos para la detección de las relaciones

aparentes entre los factores implicados y la conexión con la situación total para destacar el hecho aislado.

2). CLASIFICACION. En esta fase se implica una búsqueda de un marco de referencia desde el cual se estudien y analicen los datos. También aquí mismo se deben hacer conjeturas sobre cual es la causa de la relación observada y, por supuesto, se deben de elaborar las hipótesis que, al fín y al cabo, son una forma de clasificación.

3). VERIFICACION. Después de observar un problema y de plantear hipótesis que lo expliquen, se diseña un experimento para comprobar la validez de las respuestas sugeridas. Para esto se controlan las condiciones del fenómeno haciendo que se cambien únicamente las variables cuya influencia se desea medir.

Los datos que así se obtienen nos servirán de base para derivar las proposiciones a las que se llegue por el proceso de inducción.

4). GENERALIZACION. Una vez que se establece la influencia de ciertos factores o variables en un fenómeno determinado, se enunciarán ciertas inferencias generales, principios, teorías o leyes de las cuales se deducen ciertos juicios sobre la presentación del fenómeno.

El paso más importante es la VERIFICACION y es allí en donde el problema de la MEDICION de las variables dependientes y de los controles cobra importancia al definir el carácter de éstos y que pueden ser experimentales o estadísticos.

LA MEDICION Y SU NATURALEZA.

El concepto de MEDICION que nos interesa desarrollar aquí es el que se refiere al proceso y a la lógica que va implicada en la construcción de un instrumento de medición o escala - y a las características que debe de reunir este instrumento. --- Cuando usamos la palabra MEDICION, nos referimos al proceso para construir un instrumento de medición y no a su uso una vez que - éste ha sido establecido. Esta idea es análoga a la construcción de una balanza y no el uso de la misma para pesar cereales, por ejemplo.

Algunas definiciones sobre MEDICION:

Según B. Rusell "La medición de magnitudes es en el sentido más general cualquier método por el cual se establece -- una correspondencia única y recíproca de toda o algunas magnitudes de una clase y todos o algunos números, sean éstos enteros, racionales o reales, según el caso".

Para Campbell "MEDICION" es la asignación de números para representar propiedades de los sistemas materiales que no - son números, en virtud de las leyes que rigen estas propiedades"

Para Stevens, "MEDIR" es asignar números a los objetos o hechos, de acuerdo con reglas.*

Se observa de las anteriores definiciones que tanto Rusell como Campbell coinciden en aplicar los números a propiedades (la magnitud es, en última instancia una propiedad) mientras que para Stevens se deben aplicar números a los objetos.

---

* Las 3 definiciones son citadas por Torgerson, op cit.

Para poder medir una propiedad debe existir una correspondencia recíproca entre las características del sistema - numérico y las relaciones entre las cantidades de la propiedad por medir.

Las características principales de la serie de los números reales son: orden, distancia y origen.

1). ORDEN: Los números están ordenados.

2). DISTANCIA: Las diferencias entre los números, - es decir, las diferencias entre un par de números, pueden ser "mayor que", "igual que" ó -- "menor que". (Ley de Tricotomía).

3). ORIGEN: La serie tiene un origen único llamado "cero".

Las tres características anteriores, constituyen la fase de dos métodos para distinguir entre las distintas clases - de medición:

A) Según que los números reflejen dos o tres de las características, esto lleva a tipos de escala.

B) Según el significado de la característica misma,- es decir, clases de medición.

TIPOS DE ESCALAS.

Se refieren a toda la información que sobre la propiedad está representada por los números. Según Torgerson todas -- las escalas poseen la característica de orden, pero no acepta la escala nominal por considerarla meramente clasificatoria y el número es totalmente ajeno a la función de clasificar, la cual pue de hacerse con cualquier otro símbolo, señal o nombre.

En realidad el número, tiene potencial para darnos mucha información. Entonces de acuerdo con Torgerson las escalas de medición según tengan o no la distancia igual y el origen natural, se pueden clasificar de acuerdo al siguiente cuadro:

ORIGEN NATURAL.

		NO	SI
DISTANCIA IGUAL	NO	ESCALA ORDINAL.	ESCALA ORDINAL CON ORIGEN NATURAL.
	SI	ESCALA INTERVALOS IGUALES	ESCALA DE COCIENTES.

EJEMPLOS:

De Escala Ordinal:

- Grados escolares
- Orden de llegada de una competencia.

De Escala Ordinal con origen natural:

- Escalas psicofísicas
- Escalas de inteligencia

En las pruebas psicofísicas, el umbral absoluto.

En las escalas de inteligencia, el puntaje cero en las pruebas.

De la Escala de intervalos iguales:

El origen es arbitrario, la distancia es igual, corresponde escalas de tiempo.

- Termómetro
- Pruebas de inteligencia con puntaje estándar.

De Escala de cocientes:

- Número
- Longitudes
- Pesos
- Volúmenes

Si consideramos un origen natural, los números dados a las instancias corresponden a las distancias de los puntos desde el origen natural de la propiedad.

CLASES DE MEDICION

Se refiere a la clase de información que representan los números.

1). MEDICION DERIVADA. La característica de una escala (orden, distancia y origen) obtiene significado mediante leyes que relacionan la propiedad con otras propiedades, por ejemplo, la densidad: Hay una ley que dice que la relación entre masa y volumen para determinado material es una constante. Esta relación difiere con los distintos materiales y por lo tanto, el valor de esta relación se puede tomar como la densidad del material.

2). MEDICION POR FIAT. El significado afín se debe a una definición arbitraria, que depende de las relaciones supuestas entre las observaciones y el concepto en cuestión, p. ejm. en Ciencias de la Educación NO PODEMOS MEDIR DIRECTAMENTE, SINO QUE LO HACEMOS POR MEDIO DE INDICES QUE SOLAMENTE TIENEN UN SIGNIFICADO OPERACIONAL.

3). MEDICION DE MAGNITUDES. En este caso los números pueden asignarse según leyes naturales que representan la propiedad y no presuponen la medición de ninguna otra variable. Una construcción medida por magnitudes posee significado constitutivo y operacional por sí mismo, ejemplo: la longitud el peso, resistencia. Una escala determinada puede ser una mezcla de distintas clases de medición, por ejemplo: en Ciencias de la Educación es algo muy usual que el orden sea determinado por el método de magnitudes y los intervalos por fiat.

DISTINCION ENTRE MEDICION Y EVALUACION.

En el campo educativo se usa muy frecuentemente el concepto de EVALUACION, y, se hace necesario analizarlo un poco para diferenciarlo del de MEDICION.

La EVALUACION en el contexto educativo es un proceso sistemático para determinar en qué medida los alumnos han logrado los objetivos educacionales. Por lo anterior diremos que el concepto de EVALUACION es mucho más amplio que el de MEDICION -- La EVALUACION implica tanto descripciones cuantitativas como cualitativas de la conducta del educando, pero más importante aún, es el rasgo siguiente: supone siempre un juicio de valor en relación con esta conducta, por ejemplo, si decimos que una persona obtuvo en un examen una calificación de 6, únicamente hemos medido algo. Para que lo anterior sea una EVALUACION se deben de emitir algunos juicios como los siguientes: Esta calificación es inferior al del promedio de la población escolar, es probable que esta persona tenga dificultad en el aprendizaje o cualquier otro juicio que se pueda emitir.

La EVALUACION puede suponer descripciones cuantitativas y cualitativas de la conducta o puede suponer solamente uno de los dos tipos de descripciones, pero lo que nunca debe de dejar de exponerse es un juicio de valor. El énfasis en la EVALUACION educativa está en determinar en qué medida se han logrado - los fines educacionales.

II. MEDICION DEL RENDIMIENTO ESCOLAR Y CARACTERISTICAS DE LAS PRUEBAS.

MEDICION DEL RENDIMIENTO ESCOLAR

El rendimiento de los estudiantes, tradicionalmente y a todos los niveles escolares es y ha sido medido por medio de la aplicación de instrumentos específicos conocidos comúnmente como exámenes o pruebas.

Una prueba o examen es simplemente un conjunto de tareas que debe realizar el estudiante y que se emplea para medir una muestra de la conducta de ese estudiante en un momento determinado.

En adelante nos referiremos a cada una de estas tareas, con la palabra reactivo o ítem.

Se habla aquí de una muestra, por la imposibilidad práctica de asignar todo el universo de tareas posibles al estudiante, así se trate de un programa de estudios reducido.

Estas pruebas pueden ser escritas u orales; las pruebas escritas suelen clasificarse en pruebas de ensayo y pruebas objetivas.

Una prueba de ensayo es calificada en forma subjetiva, es decir, la opinión del maestro calificador influye sobre los resultados, en cambio las otras pruebas pueden calificarse "objetivamente", lo que significa que cualquier persona que dé la calificación llegará a los mismos resultados.

Será necesario primeramente aclarar las diferencias entre las pruebas objetivas y las de ensayo, para que pueda entenderse mejor el planteamiento anterior.

Una prueba de ensayo es un conjunto de reactivos planteados en diversos grados de amplitud, en las que el alumno tiene toda libertad para responder (Lafourcade, 1973, 75).

Las principales características de este tipo de instrumentos son las siguientes:

a) El estudiante organiza sus respuestas con total libertad.
b) El estudiante emplea su propio vocabulario.
c) El estudiante generalmente debe de contestar pocos reactivos.
d) Las respuestas que da el estudiante pueden tener diversos grados de exactitud.

Ejemplos de reactivos para este tipo de pruebas, son los siguientes:

- Mencione las características de la Distribución Normal Estándar.
- Demuestre el teorema de Pitágoras.
- Demuestre el teorema de Pitágoras empleando Semejanza de Triángulos.
- Explique la interpretación geométrica de la derivada de una función para un punto de su gráfica

Las pruebas objetivas, de las que existen diversos tipos, tienen en común la característica de que la respuesta o respuestas correctas a una pregunta o planteamiento (generalmente sólo una), queda (n) determinada (s) cuando se elabora la prueba.

En este caso, el examinado:

a). Trabaja en una tarea estructurada y no libre.

b). Selecciona una respuesta de una cantidad limitada de posibles opciones.

c) Generalmente responde una muestra grande de preguntas.

d). Recibe una puntuación o calificación para cada respuesta (según sea acertada o no) de acuerdo a una clave predeterminada.

Existen diversas clasificaciones de los tipos de pruebas objetivas, sin embargo no es propósito de este trabajo analizarlas en detalle. A pesar de ello se considera de utilidad mencionar alguna de éstas, habiendo seleccionado la que propone Pedro D. Lafourcade.

"Tipos de Pruebas Objetivas. Habitualmente Empleadas por el Docente en el Registro y Valoración de los Resultados del Aprendizaje" (Lafourcade, 1973, 73-74).

1. Pruebas que requieren la selección de algún tipo de respuesta.

   1.1. Pruebas de Alternativas Constantes
   1.2. Pruebas de Falso y Verdadero
   1.3. Pruebas de Falso y Verdadero con corrección.
   1.4. Pruebas de Falso y Verdadero con requisitos.

2. Pruebas de Tres Opciones.

   2.1. Falso - Verdadero - Dudoso.
   2.2. Siempre - Nunca - Algunas veces.

3. Pruebas de Opciones Múltiples.

   3.1. Pregunta Directa.
   3.2. Enunciado Incompleto

   - 3.3. La Respuesta más Aceptable.
   - 3.4. La Respuesta Correcta.

4. Pruebas por Pares.

   - 4.1. Pares Simples
   - 4.2. Pares Compuestos

5. Pruebas que requieren el ordenamiento de un contexto.

   - 5.1. Ordenamiento Cronológico
   - 5.2. Ordenamiento Lógico

6. Pruebas Múlti-ítem de Base Común.

Ejemplos de reactivos para algunos de los tipos de pruebas objetivas señaladas son los siguientes:

Tipo Verdadero - Falso

- Dados 3 puntos no alineados puede trazarse una circunferencia que los contenga.

Tipo Opción Múltiple de Pregunta Directa.

Si se sabe que $x + 8x = 54$ ¿Cuánto vale x?

a) 6
b) 20
c) 15
d) 4

Tipo Opción Múltiple de Enunciado Incompleto.

Si el producto de las pendientes de dos rectas dadas es - 1, dichas rectas son:

a) paralelas.
b) coincidentes.
c) perpendiculares.

VENTAJAS DE LAS PRUEBAS OBJETIVAS SOBRE LAS DE ENSAYO.

Resulta casi evidente, por lo expuesto anteriormente, que la principal ventaja de las pruebas objetivas sobre las de ensayo se presenta a la hora de calificar las respuestas a cada uno de los reactivos, pues mientras en las de ensayo pueden presentarse diferencias en la asignación de puntajes a las respuestas, ya sea por el diferente criterio de los calificadores o bien debido a que un mismo calificador modifique criterios de una prueba a otra; en las pruebas objetivas, en cambio, no importa quien o quienes califiquen las respuestas, pues lo harán de acuerdo a un criterio predeterminado en el que una respuesta a un reactivo será correcta o incorrecta sin exponerse a interpretaciones. Asimismo los pesos que se den a cada uno de los reactivos serán predeterminados, situación que no ocurre generalmente en las pruebas de ensayo.

Hay quien piensa que las pruebas objetivas son más superficiales que las de ensayo, porque al presentar al alumno las posibles respuestas se le facilita la tarea. En efecto, ésto ocurre con pruebas objetivas mal construidas, sin embargo los reactivos objetivos bien diseñados exigen al examinado desarrollar formas de pensar originales para establecer la base de elección entre las alternativas propuestas.

Pensemos por ejemplo en el juego de ajedrez en el que cada jugada resulta ser un reactivo de opción múltiple en el que debe seleccionarse la respuesta más adecuada de una serie de alternativas limitadas por las reglas del juego.

Las consideraciones que hacen que un movimiento de las piezas sea el adecuado son semejantes a las que realiza quien responde un reactivo de opción múltiple, en el que dichas opciones han sido elegidas, de tal forma que todas sean verosímiles pero entre las que solamente hay una correcta.

CARACTERISTICAS DE LAS PRUEBAS.

Existen algunas características de suma importancia que deben cumplir las pruebas para que midan eficientemente el rendimiento escolar. (Lemus, 1974, 40 - 63).

Las principales son: la CONFIABILIDAD Y LA VALIDEZ; aunque también pueden considerarse la OBJETIVIDAD, la AMPLITUD y la PRACTICABILIDAD.

La CONFIABILIDAD es la exactitud o precisión con que una prueba mide el rendimiento escolar y se refiere a su seguridad o confianza; es decir, que proporcione el mismo o casi el mismo resultado en diferentes ocasiones, ya sea aplicando el mismo instrumento o alguno equivalente a un mismo grupo de estudiantes. Cuando se habla de casi el mismo resultado, se hace referencia a fluctuaciones de carácter aleatorio, las que deben ser mínimas.

La VALIDEZ es el grado en que una prueba mide aquello que se propone medir cuando se le compara con un criterio aceptado. Es el hecho de que una prueba sea de tal manera planeada y aplicada que mida realmente si los estudiantes han alcanzado los objetivos del curso de que se trate.

Es importante manifestar algunas relaciones entre la confiabilidad y la validez de una prueba; pués una prueba puede -

ser confiable pero no válida; es decir, si se repite puede producir resultados muy similares, pero sin embargo puede no medir lo que se propone medir. Dicho de otra manera, mide confiablemente algo, pero no lo deseado.

Sin embargo, para que una prueba sea válida se requiere que sea confiable.

III. MODELOS ESTADISTICOS PARA MEDIR LA CONFIABILIDAD.

COMO SE MIDE LA CONFIABILIDAD DE UNA PRUEBA.

La medición de la confiabilidad de una prueba se puede contemplar desde dos enfoques: el enfoque experimental y el enfoque estadístico. Por un lado se cree conveniente aplicar la prueba a un grupo definido de casos de acuerdo a un plan prediseñado y manteniendo constantes las condiciones particulares de tipo experimental, y por el otro es necesario que los resultados de esa aplicación se analicen mediante técnicas estadísticas apropiadas para generar un estimador de la confiabilidad que caracteriza a la prueba.

Desde el punto de vista estadístico, después de aplicar a un grupo de individuos un instrumento de evaluación y obtener un puntaje para cada individuo del grupo, ex istirá un rango de variación en las mediciones obtenidas.

Estas diferencias se deben principalmente a dos fuentes de variación en los puntajes:

a). Las atribuibles a las verdaderas diferencias entre los examinados.

b). Las atribuibles a inexactitudes o errores en la medición.

La evaluación de la confiabilidad de una prueba sería la determinación de cuanto en la variación de un conjunto de puntajes es atribuible a las verdaderas diferencias entre los examinados y cuánto a las inexactitudes de medición.

Para medir esta variación, se requiere emplear una medida de dispersión. La que mejor evalua la variabilidad de una serie de datos es la varianza que se define de la siguiente manera:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

cuando se tienen, N datos: $x_1, x_2, x_3, \ldots\ldots\ x_N$ provenientes de una población, y como $S^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n - 1}$ cuando se tienen n datos: $x_1, x_2, x_3, \ldots\ldots, x_n$ provenientes de una muestra; donde μ es la media poblaciónal de los N datos. Y $\overline{X}$ es la media muestral de los n datos.

Las ventajas de la varianza sobre otras medidas de dispersión, para el caso que nos ocupa, es la de que la varianza puede descomponerse como una suma de varianzas que la integren, Asi, por ejemplo, si la varianza de los puntajes de un examen en un grupo es de 15 puntos (en una escala de 0 a 100), ésta podría descomponerse quizá en una varianza de 12 puntos asociada con -- las verdaderas diferencias en el rendimiento escolar de los alumnos y una varianza de 3 puntos asociada con la inexactitud de la medición.

Estas partes sumadas constituyen la varianza total de 15 puntos para todo el conjunto de valores.

Para generalizar el planteamiento anterior, definamos lo siguiente.

Sean: O_i el puntaje observado en el i-ésimo examinado que se obtiene luego de calificar la prueba.

V_i el puntaje verdadero del i-ésimo examinado, que se desconoce.

e_i el error de medición en el i-ésimo examinado.

Así pués, se tiene que $O_i = v_i + e_i$ para $i = 1 \ldots, m$

Es decir, el puntaje observado (resultado del examen) será igual al puntaje verdadero, más el error de medición. Esto se verifica para cada uno de los examinados.

Como $O_i = v_i + e_i$, para i = 1,...., m se puede demostrar fácilmente que : $\overline{O} = \overline{v} + \overline{e}$, es decir, la media aritmética de los puntajes observados en igual a la media aritmética - de los verdaderos puntajes más la media aritmética de los errores de medición.

Podemos ahora establecer lo siguiente:

σ_0^2 es la varianza de los puntajes observados.

σ_v^2 es la varianza atribuible a las verdaderas diferencias en los examinados.

σ_e^2 es la varianza debida a los errores de medición.

Y debemos probar que: $\sigma_0^2 = \sigma_v^2 + \sigma_e^2$

Sabemos que: $\sigma_0^2 = \frac{1}{m} \sum_{i=1}^{m} (o_i - \overline{o})^2$

Pero podemos expresar $o_i - \overline{o}$ de la manera siguiente: $o_i - \overline{o} = (v_i + e_i) - (\overline{v} + \overline{e})$

Por lo que:

$$\sigma_0^2 = \frac{1}{m} \sum_{i=1}^{m} [(v_i + e_i) - (\overline{v} + \overline{e})]^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} [(v_i - \overline{v}) + (e_i - \overline{e})]^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} [(v_i - \overline{v})^2 + 2(v_i - \overline{v})(e_i-\overline{e})+(e_i-\overline{e})^2]$$

$$= \frac{1}{m}\sum_{i=1}^{m}(v_i-\overline{v})^2 + \frac{1}{m}\sum_{i=1}^{m}(e_i-\overline{e})^2 + \frac{2}{m}\sum_{i=1}^{m}(v_i-\overline{v})(e_i-\overline{e})$$

Es decir:

$$\sigma_0^2 = \sigma^2_v + \sigma_e^2 + \frac{2}{m}\sum_{i=1}^{m}(v_i - \overline{v})\ (e_i - \overline{e})$$

La expresión $\sum_{i=1}^{m} (v_i - \overline{v})(e_i-\overline{e})$ es llamada la covarianza de las variables aleatorias v y e

Cuando 2 variables aleatorias son independientes su covarianza es cero.

Y si hacemos la suposición razonable de que las variables v y e son independientes, es decir, que los puntajes verdaderos son independientes de los errores de medición se tiene que:

$$\sigma_o^2 = \sigma_v^2 + \sigma_e^2$$

El coeficiente de confiabilidad que denotaremos como r, puede ser definido como la razón de la varianza atribuible a las verdaderas diferencias en los examinados entre la varianza de los puntajes observados.

Es decir: $r = \frac{\sigma_v^2}{\sigma_0^2}$

Que también puede expresarse: $r = \frac{\sigma_o^2 - \sigma_e^2}{\sigma_o^2}$

Es decir: $r = 1 - \frac{\sigma_e^2}{\sigma_o^2}$

El problema para calcular el coeficiente de confiabilidad se reduce ahora a estimar la varianza atribuida al ---- error de medición.

Para poder estimarla es conveniente que analicemos - lo que representa un puntaje de un examinado cualquiera en una - prueba.

Llamemos x_0 al puntaje observado a partir de una -- prueba de n reactivos que han sido seleccionados de un universo de reactivos sobre aquello que deseamos evaluar con la prueba - Los n reactivos, son, lo tanto, una muestra de esta población de reactivos.

N unca sabremos los resultados que hubiera logrado - el examinado si le hubiésemos aplicado toda la población de reactivos, es decir, si le hubiéramos preguntado todo lo posible sobre el tema.

Por lo que el puntaje x_0 debe considerarse como un - estimador del puntaje verdadero, pués es obtenido de una muestra y no de la población.

Así pués, consideremos el puntaje verdadero Xv, que es el puntaje que realmente habría logrado el examinado si hubiera sido posible preguntarle toda la población de reactivos y si además se hubiera podido eliminar factores de error aleatorios.

Puede plantearse entonces que:

$$X_0 = X_v + e$$

Es decir, el puntaje observado estimado es igual al puntaje verdadero más el error de medición.

La dificultad consiste en que no es posible determinar el valor de e. Sin embargo, aunque no podemos calcular e, si es posible estimar el error estandar de e.

Llamaremos el error estandar de medición a la desviación estandar de los errores de medición, es decir a Se.

Para calcular Se, Frederick Davis propone 2 posibles fórmulas. (Davis, 1980, 340 - 341).

La primera de ellas debe emplearse en caso de que se administren dos formas equivalentes de prueba al mismo grupo de alumnos, y se plantea como sigue:

$$Se = \sqrt{\frac{\sum_{i=1}^{m} d_i^{\,2}}{2\,m - 2} - \frac{(\sum_{i=1}^{m} d_i)^2}{2\,m\,(m - 1)}}$$

En la que:

m es el número de examinados.

di es la diferencia de los puntajes obtenidos por el i-ésimo alumno en las 2 pruebas, es decir, el puntaje de la primera prueba, menos el puntaje de la segunda.

Consideremos un ejemplo para ilustrar la aplicación de la fórmula señalada.

Examinados	Puntaje con la 1a. Prueba.	Puntaje con la 2a. Prueba.	di	di^2
A	97	92	5	25
B	81	79	2	4
C	60	68	-8	64
D	72	53	19	361
E	58	53	5	25
F	64	73	-9	81
G	47	46	1	1
H	51	64	-13	169
I	43	43	0	0
J	35	38	-3	9
TOTALES			-1	739

En este caso, cada prueba tenía un puntaje máximo de 100 y como puede observarse en la tabla: m = 10, es decir eran - 10 examinados, además $\sum_{i=1}^{10} d_i = -1$ y $\sum_{i=1}^{m} d_i^2 = 739$

Por lo que:

$$Se = \sqrt{\frac{\sum_{i=1}^{m} d_i^2}{2m - 2} - \frac{(\sum_{i=1}^{m} d_i)^2}{2\ m\ (m - 1)}}$$

y sustituyendo los valores obtenidos:

$$Se = \sqrt{\frac{739}{2\ (10) - 2} - \frac{(-\ 1)^2}{2\ (10)(10 - 1)}}$$

$$Se = \sqrt{\frac{739}{18} - \frac{1}{180}}$$

$$Se = \sqrt{41.06 \quad - .01}$$

$$Se = \sqrt{41,05}$$

$$Se = 6.41$$

Se es pués un estimador de σ_e y por consiguiente debe ser interpretado como todos los errores estándares de acuerdo con un intervalo de confianza y niveles fiduciales.

Considerado lo anterior, podemos estimar con este criterio a partir de la información del ejemplo, la confiabilidad de la primera prueba y la de la segunda.

Definimos anteriormente al coeficiente de confiabilidad como:

$$r = \frac{\sigma_v^2}{\sigma_o^2}$$

o como: $$r = 1 - \frac{\sigma_e^2}{\sigma_o^2}$$

σ_e^2 puede estimarse por medio de S_e^2 como ya se explicó y σ_o^2 puede estimarse a partir de la información obtenida preferentemente por S_o^2 por tratarse de una muestra.

Si por conveniencia, llamamos A a la primera prueba y B a la segunda podríamos escribir:

$$r_A = 1 - \frac{S_e^2}{S_{oA}^2} \qquad y \qquad r_B = 1 - \frac{S_e^2}{S_{oB}^2}$$

Los puntajes de los examinados en la prueba A son: 97, 81, 60, 72, 58, 64, 47, 51, 43 y 35 por lo que su varianza es 347.96, es decir $S_{oA}^2 = 347.96$.

Y la confiabilidad de la prueba A puede estimarse como:

$$r_A = 1 - \frac{41.05}{347.96}$$

$$r_A = .88$$

De manera análoga, los puntajes de los examinados en la prueba B son: 92, 79, 68, 53, 73, 46, 64, 43 y 38, por lo que su varianza es 299.21, es decir $S^2_{oB} = 299.21$.

Y la confiabilidad de la prueba B puede estimarse --como:

$$r_B = 1 - \frac{41.05}{299.21}$$

$$r_B = .86$$

¿Qué nos indica la confiabilidad de una prueba?

ya se demostró que $\sigma^2_o = \sigma^2_v + \sigma^2_e$

σ^2_e puede entonces tomar valores desde 0 (caso ideal, en el que $\sigma^2_o = \sigma^2_v$) hasta σ^2_o (en cuyo caso $\sigma^2_v = 0$). cuando $\sigma^2_e = 0$, $r = 1 - \frac{0}{\sigma^2_o} = 1$, en cuyo caso no existiría diferencia en los puntajes obtenidos por los examinados en las dos - pruebas aplicadas, lo que indicaría una confiabilidad máxima de la prueba.

Cuando $\sigma_e^2 = \sigma_0^2$, r = 1 - 1 = 0, en cuyo caso existirían diferencias muy significativas en los puntajes obtenidos por los examinados en las dos pruebas aplicadas, lo que indicaría una confiabilidad nula de la prueba.

En general, el coeficiente de confiabilidad de una prueba, teóricamente puede tomar valores entre - 1 y 1, como todos los coeficientes de correlación, aunque en la práctica es muy probable que no tome valores negativos.

También de manera práctica, podríamos afirmar que cuando el coeficiente de confiabilidad es mayor a 0.80, como en el ejemplo que fué descrito anteriormente se trata de pruebas que miden realmente con bastante seguridad, exactitud o confianza.

Anteriormente mencionamos que Davis planteó 2 fórmulas para estimar el error estandar de medición, de la cual sólo hemos explicado la primera, en la que se aplican dos pruebas equivalentes a los examinados.

Ahora explicaremos la segunda fórmula que se emplea cuando se aplica solamente una prueba.

Los reactivos de la prueba se separan en dos mitades y se clasifican en forma independiente y los puntajes de la primera mitad se correlacionan con los de la segunda mitad.

Calcular la confiabilidad por éste método proporciona esencialmente una medida de la consistencia interna de la prueba y presupone un nivel de ejecución homogéneo similar en la prueba por cada examinado.

Para aplicarlo suelen considerarse por un lado todos los reactivos pares y por otro todos los impares y obtener los puntajes de cada examinado en cada una de las mitades.

La fórmula que propone Davis es la siguiente:

$$S_e = \sqrt{\frac{\sum_{i=1}^{m} d_i^2}{m - 1} - \frac{(\sum_{i=1}^{m} d_i)^2}{m(m-1)}}$$

en donde:

m es el número de examinados.

di es la diferencia de los puntajes del i-ésimo examinado en la primera y segunda mitad de la prueba, es decir, corresponde al puntaje de la primera mitad menos el puntaje de la segunda.

Consideremos un ejemplo para ilustrar la aplicación de la fórmula señalada:

Examinados	Puntaje en Reactivos Impares.	Puntaje en Reactivos Pares	Puntaje Total	d_i	d_i^2
A	46	42	88	4	16
B	40	41	81	-1	1
C	32	32	64	0	0
D	30	40	70	-10	100
E	30	28	58	2	4
F	34	34	68	0	0
G	25	23	48	2	4
H	26	30	56	-4	16
I	22	22	44	0	0
J	18	19	37	-1	1
TOTALES				-8	142

La prueba tiene en este caso un puntaje máximo de -- 100 y como se observa en la tabla: m = 10, $\sum_{i=1}^{10} d_i = - 8$ y

$\sum_{i=1}^{m} d_i^2 = 142$ por lo que:

$$S_e = \sqrt{\frac{\sum_{i=1}^{m} d_i^{\,2}}{m - 1} - \frac{(\sum_{i=1}^{m} d_i)^2}{m\,(m - 1)}}$$

y sustituyendo valores:

$$S_e = \sqrt{\frac{142}{9} - \frac{(-8)}{90}}$$

$$S_e = \sqrt{15.78 - 0.71}$$

$$S_e = \sqrt{15.07}$$

$$S_e = 3.88$$

Así pués, la confiabilidad de la prueba sería entonces estimada por:

$$r = 1 - \frac{S_e^2}{S_0^2}$$

Para calcular S_0 se consideran los puntajes totales obtenidos por cada examinado que son los siguientes: 88, 81, 64, 70, 58, 68, 48, 56, 44 y 37, por lo que su varianza es 259.37, - es decir $S_0 = 259.37$

De donde: $r = 1 - \frac{15.07}{259.37}$

$$r = 0.94$$

Lo que indica que el examen muestra gran confiabilidad en su consistencia interna.

OTRAS TECNICAS PARA MEDIR LA CONFIABILIDAD

William Angoff desarrolló dos fórmulas estadísticas para estimar el coeficiente de confiabilidad (Angoff, 1983, 10---14); una de ellas cuando se aplican a los examinados dos pruebas equivalentes y la otra cuando se separan en dos partes los reactivos de una prueba:

Estas fórmulas son las siguientes:

$$r_A = r_B = 1 - \frac{S^2_{(A-B)}}{S^2_A + S^2_B}$$ (Para dos pruebas equivalentes). donde:

S^2_A es la varianza de los puntajes obtenidos con la prueba A.

S^2_B es la varianza de los puntajes obtenidos con la prueba B.

$S^2_{(A-B)}$ es la varianza de las diferencias de los puntajes obtenidos en las pruebas A y B.

y $$r = \frac{2\ S^2(S^2 - S_1^2 - S_2^2)}{(S^2+S_1^2-S_2^2)(S^2+S_2^2-S_1^2)}$$ (Para una prueba dividida).

donde:

S^2 es la varianza de los puntajes de los examinados considerando la prueba completa.

S_1^2 es la varianza de los puntajes de los examinados considerando los reactivos impares.

S_2^2 es la varianza de los puntajes de los examinados considerando los reactivos pares.

Vale la pena ejemplificar el uso de estas fórmulas, rescatando los casos que se consideraron previamente para comparar los resultados que se obtienen.

En el caso de pruebas equivalentes.

Examinados	Puntaje con la Prueba A	Puntaje con la Prueba B	Puntaje A - B
A	97	92	5
B	81	79	-2
C	60	68	-8
D	72	53	19
E	58	53	5
F	64	73	-9
G	47	46	1
H	51	64	13
I	43	43	0
J	35	38	-3

De donde podemos calcular las varianzas de los puntajes de cada columna obteniendo los siguientes resultados:

$S_A^2 = 347.96$ $\quad$ $S_B^2 = 299.21$ $\quad$ $S^2_{(A-B)} = 82.10$

y sustituyendo se tiene:

$$r_A = r_B = 1 - \frac{S^2_{(A-B)}}{S^2_A + S^2_B}$$

$$r_A = r_B = 1 - \frac{82.10}{347.96 + 299.21}$$

$$r_A = r_B = 1 - 0.13 = 0.87$$

que como puede observarse, es bastante aproximado al valor de -- .88 que se había obtenido al calcular el error estandar de medición.

En el caso de la prueba dividida en mitades:

Examinados	Puntaje en Reactivos Impares	Puntaje en Reactivos Pares	Puntaje Total
A	42	42	88
B	40	41	81
C	32	32	64
D	30	40	70
E	30	28	58
F	34	34	68
G	25	23	48
H	26	30	56
I	22	22	44
J	18	19	37

De donde podemos calcular las varianzas de los puntajes de cada columna obteniendo los siguientes resultados:

$S^2 = 259.37$ $S_1^2 = 69.34$ $S_2^2 = 67.87$

y sustituyendo se tiene:

$$r = \frac{2(259.37)(259.37 - 69.34 - 67.87)}{(259.37 + 69.34 - 67.87)(259.37 + 67.87 - 69.34)}$$

$$r = \frac{518.74\ (122.16)}{260.84\ (257.9)}$$

$$r = \frac{63369.278}{67270.636} = 0.94$$

que es el mismo valor que habíamos obtenido calculando el error estandar de medición.

CONFIABILIDAD EN TERMINOS DEL NUMERO DE REACTIVOS DE UNA PRUEBA.

En la teoría del muestreo, cuando se habla de la representatividad de una muestra probabilística se menciona que si se aumenta el tamaño de la muestra, se reduce la probabilidad de cometer error al estimar a los parámetros.

Aplicando este principio, se puede considerar razonablemente que se puede aumentar la confiabilidad de una prueba planteando al examinado un mayor número de reactivos.

Kuder y Richardson idearon algunas fórmulas empíricas para estimar la confiabilidad de una prueba en función del número de reactivos. Una de ellas puede ser aplicada fácilmente ya que solamente requiere conocer la media y la desviación estandar de los puntajes de una prueba.

La fórmula a que se hace referencia se expresa de la siguiente forma:

$$r = \frac{k}{k - 1} \left(1 - \frac{\overline{x}\,(k - \overline{x})}{k\ S^2}\right)$$

donde:

r es el coeficiente de confiabilidad
k es el número de reactivos de la prueba
$\overline{x}$ es la media de los puntajes de la prueba
S es la desviación estandar de los puntajes de la prueba.

IV. MODELOS ESTADISTICOS PARA MEDIR LA VALIDEZ.

COMO PUEDE ESTIMARSE LA VALIDEZ DE UNA PRUEBA.

Un índice de validez debe indicar en que grado una prueba mide el rendimiento escolar, es decir, que tanto mide lo que se propone medir, para lo cual debe comparársele con una medida externa, que llamaremos criterio.

La técnica mas simple para calcular este índice sería la de calcular el coeficiente de correlación entre los puntajes de la prueba que se aplica y los puntajes del criterio externo que se seleccione.

De los diversos coeficientes de correlación, el que se emplea con mayor frecuencia es el coeficiente de correlación producto-momento de Pearson, que para su aplicación requiere que se cumplan dos condiciones:

La primera que los datos que se correlacionen se puedan ajustar a un modelo lineal, es decir que la recta de regresión sea un buen modelo predictivo y la segunda es la homocedasticidad (*) en las dos variables.

Sean: X_1, X_2, X_3 ,......., X_n los puntajes de los examinados en la prueba cuya validez desea conocerse, con media

---

(*) Homocedasticidad: "Al estudiar determinado fenómeno y considerar 2 o más poblaciones con el mismo grado de generalidad se pueden esperar cambios en las medias, pero no en las varianzas, lo cual nos lleva a suponer una homogeneidad en las mismas, llamada Homocedasticidad". (Méndez, 1980, 38)

$M_x = \frac{\sum_{i=1}^{m} X_i}{n}$ y desviación estandar σ_X

y: y_1 , y_2 , y_3 ,........,y_n los puntajes del criterio externo seleccionado, con media $M_y = \frac{\sum_{i=1}^{m} y_i}{n}$ y desviación estandar σ_y

Por medio del método de mínimos cuadrados puede obtenerse la ecuación de la línea recta de regresión como un modelo predictivo entre estas dos variables.

La ecuación de esta recta (como la de cualquier otra) puede expresarse en la forma siguiente:

$\hat{y} = a + b\ x$ ($\hat{y}$ es un estimador, que predice los puntajes y conocido un puntaje X)

Un caso de especial interés para nuestra discusión se presenta cuando la recta de regresión pasa por el origen, en cuyo caso su ecuación puede expresarse:

$$\hat{y} = b\ x$$

Si transformamos tanto los puntajes X, como los Y a puntajes estándar, mediante las transformaciones

$z_{xi} = \frac{x_i - M_x}{\sigma_x}$ y $z_{yi} = \frac{Y_i - M_y}{\sigma_y}$

cada una de las distribuciones de los puntajes tendrá media cero y la ecuación de la recta de regresión para los puntajes estandar se puede expresar asi:

$$\hat{z}_y = b\ z_x$$

En la práctica, conocemos los puntajes estándar de cada examinado y los puntajes estándar del criterio con el que se correlacionan.

Los puntos (z_x, z_y) pueden colocarse en un plano de coordenadas y cada punto da información de un examinado. Luego debemos determinar la constante b que se ajuste mejor a esos puntos.

Mediante el método de los mínimos cuadrados se fija la posición de la recta de tal forma que, cuando las desviaciones de los puntajes estándar conocidos (z_y) respecto de los puntajes estandar predichos ($\hat{z}_y$) son elevadas al cuadrado, la suma de estos cuadrados sea mínima.

Es decir, la expresión $\sum_{i=1}^{n} (z_{y_i} - \hat{z}_{y_i})^2$ debe ser mínima.

Pero como $\hat{z}_y = b\ z_x$, la expresión que debe ser mínima puede expresarse como sigue:

$$\sum_{i=1}^{n} (z_{y_i} - b\ z_{x_i})^2$$

Los puntajes estandar z_y y z_x son conocidos, así que debemos encontrar el valor de b que haga mínima la expresión:

$$\sum_{i=1}^{n} (z_{y_i} - b\ z_{x_i})^2$$

Entonces:

$$\sum_{i=1}^{n} (Z_{y_i} - b\, Z_{x_i})^2 = (Z_{y_i} - b\, Z_{x_1})^2 + (Z_{y_2} - b\, Z_{x_2})^2 + \dots + (Z_{y_n} - bZ_{x_n})^2$$

Derivando respecto a b, se tiene:

$$\frac{d\,[\sum_{i=1}^{n} (Z_{y_i} - b\, Z_{x_i})^2\,]}{db}$$

$$= 2\,(Z_{y_1} - bZ_{x_1})(-Z_{x_1}) + 2(Z_{y_2} - bZ_{x_2})(-Z_{x_2}) + \dots + 2(Z_{y_n} - bZ_{x_n})(-Z_{x_n})$$

$$= 2[b(Z_{x_1}^2 + Z_{x_2}^2 + \dots\dots + Z_{x_n}^2) - (Z_{x_1} Z_{y_1} + Z_{x_2} Z_{y_2} + \dots + Z_{x_n} Z_{y_n})]$$

$$= 2\,[b(\sum_{i=1}^{n} Z_{x_i}^2) - \sum_{i=1}^{n} Z_{x_i} Z_{y_i}] \qquad (*)$$

Esta expresión admite una simplificación, pues como

$$\sum_{i=1}^{n} Z_{x_i}^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu_x}{\sigma_x}\right)^2 = \frac{(x_1 - \mu_x)^2 + (x_2 - \mu_x)^2 + \dots\dots + (x_n - \mu_x)^2}{\sigma_x^2}$$

$$= \frac{(x_1 - \mu_x)^2 + (x_2 - \mu_x)^2 + \dots\dots + (x_n - \mu_n)^2}{\dfrac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n}} = n\ .$$

la expresión señalada con asterisco puede expresarse:

$$2\,[b\,n - \sum_{i=1}^{n} Z_{x_i} Z_{y_i}]$$

Igualando a cero la expresión para encontrar el valor crítico de b, se tiene:

$$2 \left[b\,n - \sum_{i=1}^{n} z_{x_i}\, z_{y_i} \right] = 0$$

de donde:

$$b = \frac{\sum_{i=1}^{n} z_{x_i}\, z_{y_i}}{n}$$

Este valor de b se conoce como el coeficiente de correlación producto-momento de Pearson y podemos usarlo para estimar la validez de una prueba correlacionándola con otra que se considere como criterio externo.

Sin embargo, los cálculos con la expresión obtenida resultan engorrosos, por lo que es conveniente obtener una expresión para b en función de puntajes X e Y no de puntajes estándar.

Así que, como $z_{x_i} = \frac{x_i - \mu_x}{\sigma_x}$ y $z_{y_i} = \frac{y_i - \mu_y}{\sigma_y}$

$$b = \frac{\sum_{i=1}^{n} z_{x_i}\, z_{y_i}}{n} = \frac{\sum_{i=1}^{n} \frac{(x_i - \mu_x)}{\sigma x} \frac{(y_i - \mu y)}{\sigma y}}{n}$$

$$b = \frac{\left(\frac{x_1 - \mu_x}{\sigma x}\right)\left(\frac{y_1 - \mu_y}{\sigma y}\right) + \left(\frac{x_2 - \mu_x}{\sigma x}\right)\left(\frac{y_2 - \mu y}{\sigma y}\right) + \ldots\ldots + \left(\frac{x_n - \mu x}{\sigma X}\right)\left(\frac{y_n - \mu y}{\sigma y}\right)}{n}$$

$$b = \frac{\frac{\sum_{i=1}^{n} (X_i - \mu_x)(y_i - \mu y)}{\sigma X \quad \sigma y}}{n}$$

$$b = \frac{\sum_{i=1}^{n} (x_i - \mu x)(y_i - \mu y)}{n \sigma x \sigma y}$$

que podemos escribir en la forma:

$$b = \frac{\sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{(\sum_{i=1}^{n} (x_i - \mu_x)^2)(\sum_{i=1}^{n} (y_i - \mu_y)^2)}} \qquad (*)$$

si se tiene en cuenta que:

$$\sigma x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n}} \qquad y \qquad \sigma y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \mu_y)^2}{n}}$$

Además como:

$$\sum_{i=1}^{n} (x_i - \mu_x) = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

$$\sum_{i=1}^{n} (y_i - \mu_y)^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$\sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y) = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

y sustituyendo en la expresión con asterisco, se tiene:

$$b = \frac{\sum_{i=1}^{n} x_i\, y_i - [(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)/n]}{\sqrt{[\sum_{i=1}^{n} x_i^2 - [(\sum_{i=1}^{n} x_i)^2 / n][\sum_{i=1}^{n} y_i^2 - [(\sum_{i=1}^{n} y_i)^2 /n]]}}$$

y multiplicando por n, numerador y denominador, se obtiene:

$$b = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[\, n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2\,][n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2\,]}}$$

que es la expresión mas conocida para estimar el coeficiente de correlación producto momento de Pearson y con el que podemos estimar la validez de una prueba correlacionando los puntajes de los examinados con puntajes de un criterio externo, que puede ser obtenido por los puntajes obtenidos por los examinados en un examen con reactivos equivalentes.

Resulta conveniente ahora calcular el coeficiente de correlación b, para un ejemplo práctico que permita estimar la validez de una prueba al compararla con un criterio externo.

Consideremos el ejemplo planteado en el capítulo de Confiabilidad, en el que se tienen los puntajes de dos pruebas equivalentes. La segunda de ellas la consideraremos el criterio externo, aunque si se procediera considerando como criterio externo a la primera, se obtendría el mismo resultado; pues el coeficiente de correlación mide el grado de asociación entre las variables, en este caso de los puntajes.

Examinados	x_i Puntaje en la Prueba A	y_i Puntaje en la Prueba B (criterio Externo)	Z_{xi} Puntaje estandar en la prueba A	Z_{yi} Puntaje estandar en la prueba B	$Z_{xi}Z_{yi}$	X_i^2	Y_i^2	$X_i Y_i$
A	97	92	2.04	1.89	3.86	9409	8464	8924
B	81	79	1.14	1.10	1.25	6561	6241	6399
C	60	68	-0.05	0.43	-0.02	3600	4624	4080
D	72	53	0.63	-0.48	-0.30	5184	2809	3816
E	58	53	-0.16	-0.48	0.08	3364	2809	3074
F	64	73	0.18	0.74	0.13	4096	5329	4672
G	47	46	-0.78	-0.90	0.70	2209	2116	2162
H	51	64	-0.55	0.18	-0.10	2601	4096	3264
I	43	43	-1.01	-1.09	1.10	1849	1849	1849
J	35	38	-1.46	-1.39	2.03	1225	1444	1330

TOTALES $\sum_{i=1}^{10} X_i = 608$ $\sum_{i=1}^{10} Y_i = 609$ $\sum_{i=1}^{10} Z_{xi}Z_{yi} = 8.73$

$\mu_x = 60.8$ $\mu_y = 60.9$ $\sum_{i=1}^{10} X_i^2 = 40098$

$\sigma_x = 17.7$ $\sigma_y = 16.4$ $\sum_{i=1}^{10} y_i^2 = 39781$

$\sum_{i=1}^{10} X_i Y_i = 39570$

El cálculo de el coeficiente de correlación, que es el de validez puede obtenerse con los puntajes estandar o con los puntajes naturales, ya sea que se use la fórmula.

$$b = \frac{\sum_{i=1}^{n} z_{x_i} \; z_{y_i}}{n}$$

ó la fórmula:

$$b = \frac{n \sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{\sqrt{[n \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2][n \sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$

Sustituyendo los valores calculados en la fórmula, se tiene:

$$b = \frac{8.73}{10} = 0.873$$

ó

$$b = \frac{10(39570 - (608)\ (609)}{\sqrt{[10\ (40098) - (608)^2]\ [10(39781) - (609)^2]}}$$

es decir:

$$b = \frac{395700 - 370272}{\sqrt{(31316)\ (26929)}} = 0.873$$

Como se esperaba, se comprueba la equivalencia de las dos expresiones para medir el coeficiente de correlación de Pearson, por lo que se estima la validez de la prueba en 0.87, que es una validez satisfactoria.

OTRAS FORMAS ALTERNATIVAS DE ESTIMAR LA VALIDEZ.

Como se señala previamente, el coeficiente de correlación Producto-momento es el mejor para estimar la validez de una prueba. Sin embargo, no siempre es posible aplicarlo, por ejemplo cuando la relación entre los puntajes no es lineal; cuando - una o las dos variables no son continuas o cuando el número de - parejas es pequeño, en cuyo caso se puede recurrir a otro tipo - de coeficientes de correlación, como son el biserial-puntual el biserial y el tetracórico (Guilford, 1965,537-540)

EL COEFICIENTE DE CORRELACION BISERIAL PUNTUAL r_{bp}

Este coeficiente es conveniente aplicarlo cuando una de las variables es continua y la otra dicotómica. La variable es dicotómica, cuando se califican las respuestas a los reactivos en correctas o incorrectas, lo que ocurre principalmente en las pruebas objetivas. Puede considerarse un caso particular -- del coeficiente de Pearson. Si llamamos X_i puntaje obtenido en - la prueba X por el i-ésimo examinado y Y_i a las respuestas (X eY son variables aleatorias) en la prueba Y, se tiene que Y_i es una respuesta correcta o incorrecta y se califica con 1 o con 0, respectivamente según el caso. Llamemos N_t al número de examinados

Por lo que $\sum_{i=1}^{N_t} Y_i$ corresponde en realidad al número de examinados que respondieron correctamente a un reactivo.

Si llamamos N_e al número de examinados que responden correctamente, se tiene que $\sum_{i=1}^{N_t} Y_i = N_e$

Además $\sum_{i=1}^{N_t} Y_i^2 = N_e$, pues como las calificaciones son 0 ó 1, - la suma de cuadrados de éstas será una suma de "unos" igual a la suma de "unos" de las calificaciones.

Calculemos ahora la suma de cuadrados de las desviaciones de las calificaciones respecto a la media.

Así pues:

$$\sum_{i=1}^{N_t} (Y_i - \mu_y)^2 = \sum_{i=1}^{N_t} Y_i^2 - \frac{(\sum_{i=1}^{N_t} Y_i)^2}{N_t}$$

$$= N_e - \frac{(N_e)^2}{N_t}$$

$$= \frac{N_e N_t - N_e^2}{N_t}$$

$$= \frac{N_e (N_t - N_e)}{N_t}$$

Si llamamos N_w al número de examinados cuya respuesta es incorrecta, se tiene que: $N_w = N_t - N_e$

Así que sustituyendo se tiene:

$$\sum_{i=1}^{N_t} (Y_i - \mu_y)^2 = \frac{N_e N_w}{N_t}$$

Al calcular $\sum_{i=1}^{N_t} X_i Y_i$ solo entrarán en los cálculos aquellos valores en los que $Y_i = 1$, por lo que $\sum_{i=1}^{N_t} X_i Y_i$ -- puede escribirse en la forma $\sum_{i=1}^{N_t} f_e X_i$, en la que cada valor - de Y_i se multiplica por la frecuencia de las respuestas correc-- tas y se suman luego esos productos.

Si llamamos r_{bp} al coeficiente de correlación bise--

rial puntual y empleamos una expresión conveniente del coeficiente de correlación producto momento e igualamos, se tiene que:

$$r_{bp} = \frac{\sum_{i=1}^{N_t} X_iY_i - \frac{(\sum_{i=1}^{N_t} X_i)(\sum_{i=1}^{N_t} Y_i)}{Nt}}{\sqrt{(\sum_{i=1}^{N_t} X_i^2 - \frac{(\sum_{i=1}^{N_t} X_i)^2}{N_t})(\sum_{i=1}^{N_t} Y_i^2 - \frac{(\sum_{i=1}^{N_t} y_i)^2}{N_t})}}$$

y sustituyendo los valores obtenidos:

$$r_{bp} = \frac{\sum_{i=1}^{N_t} f_eX_i - \frac{(N_e)(\sum_{i=1}^{N_t} f X_i)}{N_t}}{\sqrt{\frac{N_e N_w}{N_t}[\sum_{i=1}^{N_t} f X_i - \frac{(\sum_{i=1}^{N_t} fy_i)^2}{N_t}]}}$$

y simplificando la expresión anterior obtenemos:

$$r_{bp} = \frac{N_t (\sum_{i=1}^{N_t} f_eX_i) - N_e (\sum_{i=1}^{N_t} f X_i)}{\sqrt{N_eN_w[N_t(\sum_{i=1}^{N_t} fX_i^2)-(\sum_{i=1}^{N_t} fX_i)^2]}}$$

Veamos ahora un ejemplo para ilustrar la aplicación de la fórmula:

En la primera columna de la tabla siguiente aparecen las puntuaciones obtenidas por 100 examinados (en escala de 0 a 10) en una prueba de rendimiento escolar (Puntuaciones X_i); en - la segunda columna aparecen las frecuencias de los que respondieron correctamente la prueba Y, de acuerdo con las puntuaciones - obtenidas en la prueba X; en la tercera columna aparecen las fre cuencias de los que contestaron de manera incorrecta la prueba Y_y en la cuarta columna se muestran las frecuencias totales -- para cada puntuación en X.

Debe aclararse que las respuestas de la prueba Y se calificaron como correctas o incorrectas con calificaciones 1 ó 0 respectivamente.

X	f_e	f_w	f	f_x	f_x^2	f_eX
10	3	0	3	30	300	30
9	4	1	5	45	405	36
8	7	1	8	64	512	56
7	7	2	9	63	441	49
6	9	2	11	66	396	54
5	6	5	11	55	275	30
4	6	6	12	48	192	24
3	3	9	12	36	108	9
2	3	7	10	20	40	6
1	1	9	10	10	10	1
0	0	9	9	0	0	0
TOTALES	$N_e=49$	$N_w=51$	$N_t=100$	437	2679	295

Debe observarse que $N_e = \Sigma f_e$, $N_w = \Sigma f_w$ y $N_t = \Sigma f$, así

como $f \doteq f_e + f_w$

Sustituyendo ahora los datos en la expresión obtenida para cálculo del coeficiente de correlación biserial puntual - para estimar la validez, se tiene:

$$r_{bp} = \frac{N_t \ (\Sigma \ f_eX) - N_e \ (\Sigma \ fX)}{\sqrt{N_e \ N_w \ [N_t(\Sigma \ fx^2) - (\Sigma fx)^2]}}$$

$$r_{bp} = \frac{100\ (295) - 49\ (437)}{\sqrt{49\ (51)[100(2679)-(437)^2]}}$$

$$r_{bp} = \frac{29500 - 21413}{\sqrt{2499\ [267900 - 190969]}} = \frac{8087}{13865.445} = 0.58$$

CORRELACION BISERIAL

El coeficiente de correlación biserial, que denotaremos r_b se emplea al igual que el de correlación biserial puntual para situaciones en las que alguna de las variables (en este caso, puntajes) que se correlacionan se ha reducido a dos categorías.

Sin embargo, este coeficiente de correlación exige - que las variables que se estudian se distribuyan normalmente, -- pues no existe una tansformación Z para este coeficiente que evite las distribuciones de muestras sesgadas asociadas a valores - altos de r.

El cálculo de este coeficiente es más sencillo que - el del biserial-puntual, pero en la actualidad con el uso de la computadora, no hay razón justificable para preferirlo, pues es un estimador de menor confianza que el coeficiente producto momento de Pearson o que el biserial puntual, pues se ha demostrado que su cuantía varia más de una muestra a otra.

La fórmula que se emplea para el cálculo de la correlación biserial, es la que sigue:

$$r_b = \frac{\mu p - \mu t}{\sigma t} \left(\frac{P}{y} \right)$$

en la que:

μp es la media de los puntajes de los que contestan correctamente el reactivo.

μt es la media de los puntajes de todos los examinados.

σt es la desviación estandar del puntaje total

p es la proporción de examinados que contestan correctamente el reactivo.

y es la ordenada de la distribución normal estandar tal que el area a la derecha de la misma es igual a p.

Para ejemplificar su uso consideremos el ejemplo presentado cuando se explicó el coeficiente de correlación biserial puntual.

X	F_e	F_w	F	F_eX	F_x
10	3	0	3	30	30
9	4	1	5	36	45
8	7	1	8	56	64
7	7	2	9	49	63
6	9	2	11	54	66
5	6	5	11	30	55
4	6	6	12	24	48
3	3	9	12	9	36
2	3	7	10	6	20
1	1	9	10	1	10
0	0	9	9	0	0
TOTALES	49	51	100	295	437

Obsérvese que: $\mu_p = \frac{\Sigma f_e x}{\Sigma f_e} = \frac{295}{49} = 6.02$

$$\mu_t = \frac{\Sigma fx}{\Sigma f} = \frac{437}{100} = 4.37, \qquad \sigma_t = 2.77$$

$$P = \frac{49}{100} = 0.49$$

El valor de Y debe localizarse en tablas de ordenadas de la curva normal estándar y observando en las mismas se encuentra que su valor es 0.3989.

Por lo que sustituyendo en la expresión para el cálculo del coeficiente de correlación biserial, se tiene:

$$r_b = \frac{\mu p - \mu t}{\sigma_t} \left(\frac{P}{y} \right)$$

$$r_b = \frac{6.02 - 4.37}{2.77} \left(\frac{0.49}{0.3989}\right) = \frac{0.8085}{1.10} = 0.73$$

coeficiente por medio del que se estima la validez del reactivo.

CORRELACION F$_i$ (Φ) y CORRELACION TETRACORICA

Ambos coeficientes de correlación se emplean cuando las variables que se estudian han sido dicotomizadas y requiere que los datos se dispongan en una tabla de contingencia de 2 renglones y 2 columnas.

Antes de formarlos consideremos la matriz de puntajes asignados a M examinados con una prueba de n reactivos, como sigue:

Reactivos

$$
\begin{array}{lcccccc}
\underline{\text{Examinados}} & 1 & 2 & 3 & 4 & \cdots\cdots & n \\
1 & a_{11} & a_{12} & a_{13} & a_{14} & \cdots\cdots & a_{1_n} \\
2 & a_{21} & a_{22} & a_{23} & a_{24} & \cdots\cdots & a_{2_n} \\
3 & a_{31} & a_{32} & a_{33} & a_{34} & \cdots\cdots & a_{3_n} \\
4 & a_{41} & a_{42} & a_{43} & a_{44} & \cdots\cdots & a_{4_n} \\
\cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\
m & a_{m_1} & a_{m_2} & a_{m_3} & a_{m_4} & \cdots\cdots & a_{m_n}
\end{array}
$$

En donde a_{ij} denota el puntaje obtenido por el i-ésimo examinado en el j-ésimo reactivo. Como cada reactivo es contestado correcta o incorrectamente (solo dos posibilidades), cada a_i^J puede tomar el valor 1 ó 0 respectivamente.

Logicamente, el puntaje total del i-ésimo examinado, está dado por la siguiente suma: $\sum_{j=1}^{n} a_{ij}$, cuyo valor máximo sería n.

Analogamente, el número de examinados que contestan correctamente elj-ésimo reactivo está dado por la suma: $\sum_{i=1}^{m} a_{ij}$, cuyo valor máximo sería m (en caso de que todos los examinados contesten el reactivo correctamente).

Si deseamos estimar la validez de un reactivo, podemos correlacionarlo con otros reactivos de la misma prueba.

Por ejemplo, supongamos que se desea correlacionar el reactivo K con el reactivo 1.

Puede construirse la siguiente tabla de contingencia:

	Reactivo 1		
Reactivo k	Examinados que responden correctamente	Examinados que responden incorrectamente	totales
Examinados que responden correctamente	A	B	A + B
Examinados que responden incorrectamente	C	D	C + D
TOTALES	A + B	B + D	A + B + C + D

En donde A es el número de examinados que responden correctamente ambos reactivos; B el número de examinados que responden correctamente el reactivo K, pero incorrectamente el reactivo 1; C el número de examinados que responden correctamente el reactivo 1, pero incorrectamente al reactivo K y D el número de examinados que responden incorrectamente ambos reactivos.

El grado de realción expresado por el coeficiente de correlación nos indica el grado de exactitud con el que es posible predecir de una de las dos distribuciones a la otra. Aquí se podría establecer una predicción exacta si todos los que resolvieron el reactivo K también hubiesen resuelto el reactivo 1 y todos los que fallaron en uno también hubiesen fallado en el otro.

Por lo que si conocieramos la ejecución del examinado en el reactivo K (por ejemplo) podríamos predecir su ejecución en el reactivo 1 con seguridad. Bajo estas condiciones, el coeficiente de correlación sería 1.

El coeficiente de correlación Φ puede ser calculado con la información de la tabla de contingencia por medio de la expresión siguiente:

$$\Phi = \frac{A\,D - B\,C}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

El coeficiente de correlación tetracórico, que denotaremos r_π , también puede calcularse a partir de la información de la tabla de contingencia, obteniendose una estimación trigonométrica del mismo por la expresión:

$$r_\pi = \cos\left(\frac{180°}{1+\sqrt{\frac{AD}{BC}}}\right) \qquad *$$

Valdría la pena en este caso ilustrar con un ejemplo el cálculo de estos coeficientes de correlación para estimar la -

---

* Esta expresión trigonométrica fué propuesta por Davidov y Go-Heen.

validez entre los reactivos de una prueba iniciando el procedimiento desde la matriz de puntajes.

En el ejemplo que se desarrolla se considera una --- prueba de 5 reactivos para 10 examinados. Lo anterior se hace - solo por fines explicativos, pues en la práctica las pruebas tienen por lo general muchos mas reactivos y se aplican a un mayor número de examinados.

Matriz de Puntajes del examen A

	Reactivos					
Examinados	1	2	3	4	5	TOTALES
1	1	1	0	1	0	3
2	1	1	1	1	1	5
3	1	0	1	0	0	2
4	0	1	0	1	0	2
5	1	1	0	1	1	4
6	1	0	1	0	0	2
7	0	1	1	1	1	4
8	1	1	0	0	0	2
9	1	0	1	0	0	2
10	1	1	1	0	1	4
TOTALES	8	7	6	5	4	

La tabla de contingencias para la correlación de los reactivos 1 y 2 es la siguiente:

	Reactivo 2				
Reactivo 1	Correctos		Incorrectos		Totales
Correctos	5	A	3	B	8 A + B
Incorrectos	2	C	0	D	2 C + D
TOTALES	7	A + C	3	B + D	10 A + B + C + D

Usando Φ y sustituyendo:

$$\Phi_{12} = \frac{0 - 6}{\sqrt{(8)(2)(7)(3)}} = \frac{-6}{18.33} = - 0,33$$

En este caso el número de intercorrelaciones que se pueden estimar es de 10, es decir $\binom{5}{2}$

Procediendo de manera análoga al ejemplo se pueden calcular:

Φ_{13} = 0.10, Φ_{14} = - 0.50 Φ_{15} = - 0.10

Φ_{23} = - 0.53 Φ_{24} = 0.66 Φ_{25} = 0.53

Φ_{34} = - 0.41 Φ_{35} = 0.25 Φ_{45} = 0.41

y construir la matriz de intercorrelaciones siguientes:

		Reactivos				
		1	2	3	4	5
	1	--	-0.33	0.10	-0.50	-0.10
	2	--	-----	0.53	0.66	0.53
Reactivos	3	--	-----	----	-0.41	0.25
	4	--	-----	----	-----	0.41
	5	--	-----	----	-----	----

LA TEORIA DEL ANALISIS FACTORIAL.

Hemos discutido ya que el puntaje obtenido en una prueba por un examinado (o_i) es igual al puntaje verdadero (v_i) mas el error de medición (e_i), es decir: $o_i = v_i + e_i$

Asimismo hemos demostrado que la varianza de los puntajes obtenidos es igual a la varianza de los puntajes verdaderos mas la varianza de los errores en la medición, es decir:

$$\sigma_o^2 = \sigma_v^2 + \sigma_e^2$$

Esta teória acepta que la varianza de los puntajes verdaderos puede descomponerse en términos de partes de varianza atribuibles a un número finito de factores que son independientes entre sí.

Esta consideración supone que el puntaje verdadero de cada examinado (v_i) puede separarse en los componentes: $v_{iA} + v_{iB} + v_{iC} + \ldots\ldots + v_{iM}$ donde la magnitud de cada componente está determinada por un solo factor (A, B, C,....., M), que no está correlacionado con ningún otro.

Como estos m factores son independientes, la varianza verdadera estará integrada por la suma de las varianzas determinadas por los M factores diferentes, pudiendo expresarse con la ecuación:

$$\sigma_v^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \ldots\ldots\ldots + \sigma_M^2$$

donde:

σ_A^2 es la varianza atribuible al factor A,

σ_B^2 es la varianza atribuible al factor B, hasta

σ_M^2 es la varianza atribuible al factor M

y como $\sigma_o^2 = \sigma_v^2 + \sigma_e^2$

podemos expresar σ^2_0 como sigue:

$$\sigma^2_0 = \sigma^2_A + \sigma^2_B + \sigma^2_C + \ldots\ldots + \sigma^2_M + \sigma^2_e$$

Si dividimos ambos miembros entre σ^2_0 podemos expresar la contribución de cada factor a la varianza observada como una proporción:

$$\frac{\sigma^2_0}{\sigma^2_0} = \frac{\sigma^2_A + \sigma^2_B + \sigma^2_e + \ldots + \sigma^2_M + \sigma^2_e}{\sigma^2_0}$$

es decir:

$$1 = A^2_0 + B^2_0 + C^2_0 + \ldots\ldots + M^2_0 + e^2_0$$

donde A^2_0 es la proporción de la varianza de la prueba compuesta de varianza verdadera determinada por el factor A, B^2_0 es la proporción de la varianza total de la prueba compuesta de varianza verdadera determinada por el factor B y así hasta el factor M y e^2_0 es la proporción de la varianza total de la prueba formada por varianza de error.

En la teoría del Analísis Factorial, se definen algunos términos especiales como la comunalidad y la especificidad. (Guilford, 1965)

La Comunalidad es la proporción de la varianza total que una prueba tiene en común con otras pruebas en una matriz de correlación dada, es una varianza que ocasiona la correlación entre la prueba dada y las otras pruebas basada en los componentes verdaderos de la varianza.

Guilford la representa como h_0^2 y como ecuación puede expresarse como sigue:

$$h_0^2 = A_0^2 + B_0^2 + \ldots\ldots + M_0^2$$

La especificidad de una prueba es la parte de la varianza verdadera que no aparece en otras pruebas en una matriz de correlación dada y por lo tanto no contribuye a la correlación entre la prueba cuya validez se estudia y las pruebas o criterios seleccionados.

Guilford la representa como v_0^2

Recordemos ahora que la confiabilidad de una prueba se definió como la razón entre la varianza de los puntajes verdaderos y la varianza de los puntajes obtenidos, es decir:

$$r = \frac{\sigma_v^2}{\sigma_0^2}$$

Como tanto la comunalidad como la especificidad componen a la fracción de varianza verdadera en la varianza obtenida, podemos redefinir a la confiabilidad como sigue:

$$r = h_0^2 + v_0^2$$

Es decir la confiabilidad de una prueba es igual a la comunalidad mas la especificidad.

Debido a que el objeto de este análisis es ejemplificar la dependencia de la validez de la estructura factorial de la prueba, distinguiremos entre la comunalidad y lo que llamaremos la varianza única de la prueba que esta formada por la especificidad y la proporción de varianza de error.

Si llamamos U_0^2 a la varianza única tendremos que:

$$U_0^2 = v_0^2 + e_0^2$$

Como $$r = 1 - \frac{\sigma_e^2}{\sigma_0^2}$$

es decir:

$$r = 1 - e_0^2$$

pero: $$r = h_0^2 + v_0^2$$

por lo que: $$h_0^2 + v_0^2 = 1 - e_0^2$$

es decir: $$h_0^2 + v_0^2 + e_0^2 = 1 \text{ y como } U_0^2 = v_0^2 + e_0^2$$

se tiene: $$h_0^2 + U_0^2 = 1$$

Es importante observar que la comunalidad es la única fracción de la varianza total de la prueba que puede usarse para predicciones y pone un límite a la validez del mismo.

La especificidad con respecto a un criterio dado, aunque es parte de la varianza verdadera y contribuye a la confiabilidad no contribuye a la validez de la prueba con ese criterio.

Se presenta a continuación un ejemplo numérico en -- forma de matriz, en el que se analizan 4 pruebas con 5 factores:

	Factores						
Prueba	A	B	C	D	E	h_0^2	U_0^2
1	0.04	0.36	0.10	0.35	0.00	0.85	0.15
2	0.16	0.00	0.12	0.00	0.64	0.92	0.08
3	0.00	0.49	0.25	0.00	0.16	0.90	0.10
4	0.20	0.30	0.17	0.08	0.04	0.79	0.21

Así, por ejemplo se tiene que el 4% de la varianza de la prueba 1 está determinada por el factor A, el 36% por el factor B, el 10 % por el factor C y el 35% por el factor D, mientras que el factor E no influye en la varianza de la prueba.

En total, la varianza de los factores comunes da su comunalidad de 85%, por lo que la varianza única de la prueba constituye el 15%, de la varianza total.

Pueden realizarse las interpretaciones de la distribución de la varianza de las otras 3 pruebas en forma análoga.

La raíz cuadrada de la proporción de la varianza total a la que contribuye cada componente es considerada como la correlación entre la variable y el factor.

Esas raíces cuadradas se conocen como saturaciones de los factores.

Así que, para nuestro ejemplo, podemos construir la matriz de saturación de los factores, como sigue:

	Factores				
Prueba	A	B	C	D	E
1	0.20	0.60	0.31	0.59	0
2	0.40	0	0.35	0	0.80
3	0	0.70	0.50	0	0.40
4	0.45	0.55	0.41	0.28	0.20

Guilford menciona que Thurstone propone la siguiente expresión para estimar la correlación entre una prueba X y otra y, una de las cuales se emplea como criterio externo:

$$r_{xy} = A_x A_y + B_x B_y + C_x C_y + \ldots\ldots + M_x\ M_y$$

cuando se tienen los factores A, B, C,......, M y donde: A_x, B_x, C_x,......, M_x son las saturaciones de los factores A, B, C.,....,M respectivamente de la prueba X.

A_y, B_y, C_y,....., M_y son las saturaciones de los factores A, B, C,.....,M respectivamente de la prueba Y

Por lo que si deseamos conocer la validez de la prueba 1, considerando como criterio externo la prueba 4, para nuestro ejemplo, se tiene:

$$r_{14} = A_1 \ A_4 + B_1 \ B_4 + C_1 \ C_4 + D_1 \ D_4 + E_1 \ E_4$$

$$r_{14} = (0.20)(0.45)+(0.60)(0.55)+(0.31)(0.41)+(0.59)(0.28) + (0)(0.20)$$

$$r_{14} = 0.71$$

Analogamente, pueden correlacionarse el resto de las pruebas para estimar su validez

COMO INTERPRETAR LOS INDICES DE VALIDEZ

Como los índices de validez son estimados por coeficientes de correlación, debemos recordar que aun una correlación perfecta entre dos variables no indica una relación causal necesaria o específica entre ellas; simplemente nos indica que las variables aumentan en perfecta concordancia cuando r = 1 y en oposición perfecta si r = - 1

En segundo término, al interpretar los coeficientes debemos comenzar por restringir su significación.

Si los coeficientes no son significativamente diferentes de cero puede pensarse en invalidez. Sin embargo, si difieren significativamente de cero, su interpretación depende del propósito para el que han sido calculados.

En el caso de la estimación de validez de una prueba o de un reactivo los coeficientes deben ser bastante altos para ser útiles.

Sin embargo la interpretación de la validez no es tan fácil de realizar como la de la confiabilidad, pues la dificultad estriba en encontrar un criterio de comparación apropiado de lo - que la prueba está tratando de medir, fuera de la prueba misma.

No hay regla general para la selección de criterios de validación, éstas variarán de prueba a prueba. Sin embargo, este problema puede dejarse al docente o aplicador individual de pruebas.

Es adecuado considerar válida una prueba desde el -- punto de vista estadístico si se encuentra una correlación alta - (mayor de .7) entre la prueba y un criterio independiente satisfactorio, siempre y cuando las confiabilidades tanto de la prueba como del criterio sean satisfactorias, por lo que se reitera que para que una prueba sea válida debe ser confiable.

V. CONCLUSIONES.

CONCLUSIONES

Debe precisarse que las técnicas estadísticas que se plantean en este trabajo para estimar la confiabilidad y la validez de las pruebas, tienen virtudes y limitaciones tanto desde el punto de vista estadístico como general.

La principal virtud y quizá la mas importante es que auxilian al investigador o docente que en una actitud profesional desea darse cuenta si los instrumentos de medición del rendimiento escolar que emplea pueden estimarse estadísticamente como confiables y válidos.

Contribuyen también estas técnicas a las teorías sobre el rendimiento escolar, aportando a los estudiosos elementos estadísticos elementales, que les permita verificar estadísticamente algunas de las hipótesis que planteen.

Como las técnicas que se plantean para estimar la confiabilidad y la validez son principalmente correlacionales, no es posible a partir de ellas establecer relaciones de causalidad entre las variables que se estudian.

No debe olvidarse que los puntajes de las pruebas, deben entenderse como valores muestrales, por que los resultados que se obtienen son estimadores y no parámetros, por lo que si se desea conocer la significancia de un estimador será necesario probar las hipótesis que se planteen, a determinado nivel de confianza. Además, debe recordarse que en los métodos paramétricos (por ejemplo, la correlación producto momento de Pearson), deben cumplirse algunas condiciones como: la distribución normal de las variables, el tamaño de la muestra y la homocedasticidad. En los métodos no paramétricos o de libre distribución se pueden realizar inferencias sin necesidad de que se cumplan condiciones como las mencionadas.

Por otra parte, no puede pensarse que con la estimación de la confiabilidad y la validez de las pruebas, pueda resolverse el problema del rendimiento escolar, pues éste tiene -- aparte determinantes histórico-sociales, políticos, institucionales y psicopedagógicos, está influido por el proyecto académico que lo rige, por las características socioeconómicas y culturales y por los antecedentes académicos de alumnos y maestros.

BIBLIOGRAFIA.

Bibliografía Consultada

1) "EVALUACION DE LOS APRENDIZAJES"
PEDRO D. LA FOURCADE
EDITORIAL KAPELUSZ
BUENOS AIRES, 1973

2) "EVALUACION DEL RENDIMIENTO ESCOLAR"
LUIS ARTURO LEMUS
EDITORIAL KAPELUSZ
BUENOS AIRES, 1974

3)"EDUCATIONAL MEASUREMENTS AND THEIR INTERPRETATION"
FREDERICK B. DAVIS
WADSWORTH PUBLISHING
BELMONT, CALIFORNIA, 1980

4) "PSYCHOMETRIC METHODS"
J.P. GUILFORD
M.C. GRAW HILL
NEW YORK WILEY, 1965

5) "ANALISIS DE LOS ITEMES EN LA CONSTRUCCION DE INSTRUMENTOS PSICOMETRICOS"
NICOLAS M. TAVELLA
TRILLAS, MEXICO, 1978

6) "STATISTICS: METHODS AND ANALYSES"
LINDOLN L. CHAD
MC. GRAW HILL-KOGAKISHA
TOKYO, 1969

7)"PROBABILITY AND MATHEMATICAL STATISTICS"
LESTER D TAYLOR
HARPER AND ROW
NEW YORK, 1974

8) "METODOS ESTADISTICOS APLICADOS"
N.M. DOWNIE
R.W. HEATH
HARLA, MEXICO,1973

9) "INTRODUCCION A LOS METODOS ESTADISTICOS" VOLUMEN 3
VARIOS AUTORES
UNIVERSIDAD PEDAGOGICA NACIONAL, 1983

10) "METODOS DE INVESTIGACION EDUCATIVA"
VARIOS AUTORES
SISTEMA DE UNIVERSIDAD ABIERTA UNAM, 1980

11)"MEDICION Y EVALUACION EN PSICOLOGIA Y EDUCACION "
ROBERT . THORNDIKE L.
JOHN WILEY AND SONS INC. LONDON, 1973

12) "THEORY AND METHODS OF SCALING"
W. TORGERSON
JOHN WILEY AND SONS NUEVA YORK, 1954

13) "MODELOS ESTADISTICOS LINEALES"
IGNACIO MENDEZ RAMIREZ
CONACYT, MEXICO, 1980

14) "ELABORACION DE TESTS"
DOROTHY ADKINS WOOD
TRILLAS, MEXICO, 1979

15)"FUNDAMENTALS OF CONCEPT FORMATION IN EMPIRICAL SCIENCE"
HEMPEL
CHICAGO UNIV PRESS 1986

Artículos Consultados

1) "MARCO TEORICO PARA EL ESTUDIO DEL RENDIMIENTO ESCOLAR" EVALUACION DEL CURRICULO
María Isabel Galán Giral y Dora Elena Marín Mendez
Revista Perfiles Educativos CISE UNAM Enero-Junio 1983

2) "TESIS PARA UNA TEORIA DE LA EVALUACION Y SUS DERIVACIONES EN LA DOCENCIA"
Angel Díaz Barriga
Revista Perfiles Educativos CISE UNAM Enero-marzo 82

3) "EVALUACION CURRICULAR: UNA PROPUESTA DE TRABAJO PARA EL ESTUDIO DEL RENDIMIENTO ESCOLAR"
María Isabel Galán Giral y Dora Elena Marín Méndez
Revista Perfiles Educativos CISE UNAM Abril-Junio 1986

4) "LA EVALUACION EN LA EDUCACION"
Ernesto García Cortés
Revista Perfiles Educativos CISE UNAM enero-Marzo 1979.