



Nombre del Alumno: Ernesto López Sánchez.

Nombre del Profesor: Lic. Liliana Villegas López.

Nombre del Trabajo: Resumen del tema “Confiabilidad en las Pruebas de Medición”

Materia: Psicometría.

Grado: 3er Cuatrimestre.

Grupo: "A"

Pichucalco, Chiapas a 18 De Julio De 2020.

CONFIABILIDAD EN LAS PRUEBAS DE MEDICIÓN.

Para que un test pueda ser utilizado responsablemente es necesario que cumpla con determinados estándares técnicos.

Anteriormente los instrumentos de medición en psicología fueron aplicados con escaso control de su calidad, se administraron e interpretaron sin una clara demostración de su utilidad para los fines propuestos, ni de sus límites o alcances.

Esto ocasionó innumerables críticas y, en muchos casos, un abierto rechazo social a la utilización de test. Como consecuencia de estos cuestionamientos surgió una corriente de revisión y análisis de la fundamentación científica de las pruebas.

En los Estados Unidos se publicaron documentos generados por organizaciones especializadas (American Psychological Association, American Educational Research Association), cuya meta esencial fue establecer los requisitos técnicos mínimos que debían reunir los test utilizados en el ámbito de la psicología y la educación.

En 1966 se publicaron las “Normas técnicas para test psicológicos y educativos” y en 1999 apareció la última revisión de las mismas, que incluye modificaciones importantes, particularmente en la concepción de validez de los test.

Los profesionales usuarios de test deben conocer estas normas y ajustarse estrictamente a ellas en lo concerniente a la administración, validación e interpretación de los resultados de estos instrumentos.

En la actualidad coexisten dos teorías generales de los test, la teoría clásica de los test y la de respuesta al ítem.

La hipótesis fundamental de la teoría clásica de los test (TCT) es que la puntuación observada de una persona en un test es una función de dos componentes: su puntaje verdadero (que es inobservable) y el error de medición implícito en toda medición.

El postulado esencial de la TCT se expresa como:

$$O_i = V_i + E_i,$$

Esto es, la puntuación observada de un individuo es igual a la puntuación verdadera más el error. En el plano teórico, la puntuación verdadera puede entenderse como la media de

las puntuaciones obtenidas por una persona en infinitas aplicaciones de un test (en diferentes momentos y condiciones), asumiendo que la forma de distribución de esas infinitas puntuaciones se aproxima a la normal.

La puntuación de error, por otra parte, es la suma de todos aquellos factores aleatorios que influyen y afectan el registro de los datos, introduciendo inconsistencia en el proceso y alejando la puntuación observada de la puntuación verdadera.

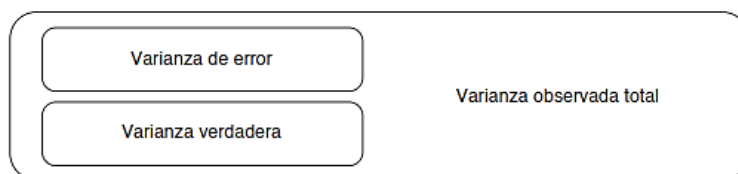
La confiabilidad puede entenderse como la exactitud o precisión de una medición, o el grado en el cual las puntuaciones de un test están libres de esos errores de medición. Esta exactitud o precisión de las puntuaciones permite que éstas se mantengan constantes en diferentes circunstancias.

La confiabilidad significa la consistencia entre los puntajes de un test obtenido por los mismos individuos en distintas ocasiones o entre diferentes conjuntos de ítems equivalentes (APA, 1999).

Debemos distinguir entre la falta de consistencia debido a un cambio real en el rasgo medido y las fluctuaciones azarosas de las puntuaciones ocasionadas por cambios en circunstancias personales u otras que afecten la situación de evaluación.

En general, la confiabilidad se verifica mediante un coeficiente de correlación entre las medidas repetidas de un fenómeno. Para entender la lógica de la utilización de este coeficiente cabe realizar algunas observaciones previas. La puntuación verdadera también puede expresarse en términos de varianza de las puntuaciones de los test.

En la teoría clásica de los tests, la confiabilidad puede expresarse como la proporción de varianza observada de las puntuaciones de tests que se debe a la varianza verdadera (la variabilidad de la puntuación verdadera), es decir, a la varianza del rasgo que se está evaluando y no a otros factores. Puede representarse lo afirmado anteriormente de la siguiente manera:



Los diferentes métodos utilizados para evaluar la confiabilidad utilizan el coeficiente de correlación como estadístico fundamental. De este modo, un coeficiente de 0,80 sugiere que un 64% de la varianza observada es asimilable a la variabilidad de la puntuación verdadera, con un margen de error del 36%. Los diferentes métodos que se expondrán más adelante están diseñados para estimar la cuantía de error presente en las puntuaciones de un test determinado.

Existen múltiples factores que afectan la confiabilidad de las puntuaciones obtenidas mediante la aplicación de un test. La APA (1999) los clasifica en factores internos (fluctuaciones en el nivel de atención del examinado, por ejemplo) o externos (diferentes criterios de corrección de un test, según evaluadores distintos, por ejemplo). A continuación, revisamos estos factores de error, agrupándolos en cuatro categorías de acuerdo al criterio propuesto por Hogan (2004). Esta clasificación no pretende ser exhaustiva, sino solamente destacar aquellos errores fundamentales que afectan la confiabilidad de las puntuaciones.

-Contenido del test

Al construir un test debe tenerse en cuenta que la elección de los ítems, los materiales a través de los cuales estimulamos la respuesta del individuo, y la formulación de las consignas pueden ser una fuente de error aleatorio en la medición. Si se desea estimar el puntaje de un individuo en un cierto dominio, los errores en el muestreo de contenido pueden afectar la confiabilidad.

Si consideramos a un test como un conjunto de indicadores operacionales de un rasgo o dominio de comportamiento, es comprensible que la elección de los mismos (muestra de ítems) pueda constituir una fuente de error. Tanto si el muestreo de contenido es sesgado, como si no es suficientemente extenso, los puntajes resultantes serán poco confiables. Un test de inteligencia con solo 10 ítems representa un ejemplo de muestreo de contenido muy limitado, pues en este caso los puntajes dependerían de la capacidad del individuo con respecto al pequeño conjunto de reactivos utilizados.

-Administración

En el momento de administrar un test también pueden introducirse errores que afecten la confiabilidad de los resultados. Por ese motivo es esencial examinar a todos los participantes en condiciones uniformes, estandarizadas. Así, por ejemplo, las condiciones generales del ambiente en que se administran los tests deben ser lo más semejantes que sea posible (iluminación, nivel de ruido o confort del lugar).

Por consiguiente, siempre es deseable que las instrucciones del test sean lo suficientemente claras y unívocas, para que todos los evaluadores las impartan de la misma manera y presenten los materiales en idéntico orden y forma a todos los examinados. La falta de consistencia en la administración de un test influirá en la estabilidad de las puntuaciones obtenidas por las personas medidas por ese test.

-Calificación

En el momento de calificar (puntuar) un test existen otros factores que pueden influir negativamente en la confiabilidad de las puntuaciones obtenidas. En los tests de opción múltiple pueden cometerse errores cuando la corrección es manual, tales como calificar accidentalmente algunas respuestas correctas como erróneas y viceversa, o realizar mal la sumatoria de las respuestas clave o correctas. La posibilidad de cometer este tipo de errores prácticamente se elimina en los tests computarizados.

-Factores internos del examinado.

En los métodos utilizados para corroborar la confiabilidad de un instrumento de medición se pueden distinguir dos instancias. Por una parte, es necesario administrar el instrumento a una muestra según un diseño de investigación específico y, por otro lado, los datos que resulten de tal aplicación deben ser analizados mediante procedimientos apropiados para obtener un estadístico que represente la confiabilidad de las puntuaciones del test (Cortada de Kohan, 1999)

-Test-retest:

Este método consiste en administrar un test en dos oportunidades a la misma muestra de sujetos, con un determinado intervalo entre las dos administraciones, y calcular la correlación entre los puntajes obtenidos en la primera y segunda vez.

-Formas equivalentes:

A través de este método, también denominado “formas paralelas”, se puede evaluar la consistencia interna pero también la estabilidad temporal de un conjunto de puntuaciones. El procedimiento básico consiste en administrar dos formas equivalentes de un test a un mismo grupo de individuos.

-Partición en mitades:

A través de este método se verifica la consistencia interna de las puntuaciones de un test, es decir, el grado en que las diferentes partes del test miden la misma variable.

Se administra el test en una ocasión a una muestra de individuos y posteriormente se divide la prueba en dos mitades comparables, obteniendo de esta manera dos puntuaciones para cada individuo de la muestra. Finalmente, se correlacionan las puntuaciones correspondientes a ambas mitades del test por medio de un coeficiente de correlación. Este método fue popular antes de que se dispusiera de computadoras personales, debido a que los estadísticos requeridos son más fáciles de calcular manualmente que el coeficiente alfa

-Métodos de covarianza de los ítems:

Estos métodos comparten con el anterior dos aspectos importantes: por un lado, permiten verificar la consistencia interna de los puntajes del test y, por otra parte, requieren una sola administración de la prueba (Thorndike, 1989).

A partir de una única aplicación del test a una muestra se obtiene una estimación del grado de covarianza de los ítems, utilizando como estadístico el coeficiente alfa de Cronbach o la fórmula alternativa de Kuder Richardson (KR-20), cuando se trabaja con ítems dicotómicos (verdadero-falso, por ejemplo).

-Acuerdo entre examinadores.

La dimensión evaluada por este método es la confiabilidad entre examinadores. El método consiste en administrar un test a una muestra, entregar los resultados (protocolos de respuesta) del test a un conjunto de jueces que los puntuarán independientemente. A continuación, se verifica el grado de acuerdo que alcanzan los jueces luego de leer, registrar y codificar los mismos datos (Murat, 1985).

Naturalmente, este procedimiento no se aplica en tests que se puntúan de manera objetiva (de opción múltiple, por ejemplo) y sólo adquiere importancia cuando interviene el criterio del examinador en el proceso de calificación.

BIBLIOGRAFÍA:

PDF/ Recursos UDS / Introducción a la psicometría / Silvia Tornimbeni, Buenos Aires: Paidós, 2008