Pro Mathematica Vol. XVIII, Nos. 35-36, 2004

INTRODUCCIÓN AL MODELO PSICOMÉTRICO DE LA TEORÍA CLÁSICA DE LOS TESTS (PARTE 1)

Jorge Luis Bazán Guzmán

Artículo de difusión adaptado de conferencia en el XIV Coloquio Nacional de Matemática

Resumen

Se hace un esbozo de la justificación estadística y matemática propia de la Psicometría para el problema de medición psicológica, con una introducción a la axiomática y principales relaciones de la teoría clásica de los tests. Nuestro abordaje se basa en una revisión de la literatura pertinente, pero, adicionalmente, contiene resultados y derivaciones propias. Un valor adicional está constituido por el tratamiento como modelo y su conexión desde la teoría matemática de medición y la justificación de la aleatoriedad de los puntajes observados, así como incorporar la propuesta de Zimmerman sobre la definición de validez. El presente estudio pretende ofrecer una presentación del modelo psicométrico de la Teoría clásica de los tests, el cual contiene parte del modelo lineal clásico para una medición en una población de sujetos.

Sección Matemática, Departamento de Ciencias, PUCP

Introducción

El uso de los Tests en Psicología tienen larga data pero la construcción de pruebas empíricas de tipo psicométrico ha tenido y tiene un gran desarrollo, pero desigual, que está constituido por técnicas y metodologías con cierta racionalidad muchas veces no explícita.[1]

Otro aspecto tardío fue sustentar, de alguna manera, la Teoría de los Tests a través de un modelo matemático.[2] Entre estos esfuerzos pueden citarse a Gulliksen (1950), Guilford (1954) y Torgenson (1958). Una sistematización exhaustiva y más o menos sólida se alcanzó finalmente con la obra de Lord y Novick de 1968, reeditada en 1974.

El test o prueba de tipo psicométrico se construye teniendo como base modelos estadísticos tales como: Modelos lineales con error de medición. Análisis Factorial, Análisis de Variables Latentes y modelos de Item-Respuesta como los modelos de Rash. En su mayor parte estos modelos se generalizan en el Modelo de Ecuaciones Estructurales o de Estructuras de Covarianza (véase Lord y Novick, 1974; Weiss y Davison 1981).

Uno de los modelos estadísticos más utilizados para la construcción de un test psicométrico es la *Teoría Clásica de los Tests* (TCT), denominada también Teoría del Puntaje verdadero. Esta se apoya en un modelo lineal con error de medición formulado por Spearman en 1904.

Relacionada con la TCT está la Teoría de Medición, teoría matemática sistematizada en los tres volúmenes de Krantz *et al.* (1971).[3]

En nuestro medio la literatura es escasa en ofrecer los detalles matemáticos y técnicos de la fundamentación estadística de los tests psicométri-

[1]En 1954 la APA, recién, se reunió para uniformizar la terminología usada y para formular reglas básicas para la estandarización de los Tests y la Construcción de Pruebas. APA(1954)

[2]Esta necesidad de formalización está en la tradición de toda aplicación de la matemática y estadística. Matemáticos y estadísticos participaron activamente en este propósito, por ejemplo Pearson y Spearman entre los clásicos y Young, Rash y Jöreskog entre los más recientes.

[3]Luce y Narens (1986) han trazado la historia moderna de la teoría de la medición a través de las contribuciones de autores como Campbell, Stevens, von Neuman y Morgenstern, Pfanzagl, Savage, Suppes, Debreu, Luce y Tukey. En Narens y Luce (1986) aparece un sumario de trabajos recientes sobre esos tópicos.

cos, así como de relacionar ésta con la teoría matemática de la medición. (véase Jones y Appelbaum, 1989; Magnunson, 1990; Pfanzagl, 1973).

En este estudio se esbozará un modelo constructivo que partiendo de la Teoría de Escalas de Medición (TEM) nos introduzca a las principales propiedades de los instrumentos psicométricos dentro del marco de la TCT. Pasaremos así de un modelo determinístico, como es el caso de la TEM, a uno probabilístico, como es el caso de la TCT. Se esbozará una justificación estadística y matemática propia de la Psicometría para el problema de medición psicológica introduciendo la axiomática y principales relaciones de la TCT. También se presenta una discusión detallada acerca de los modelos presentados y de las principales relaciones identificadas.

En otro estudio, se puede abordar los aspectos metodológicos y técnicas, que forman parte de la construcción de pruebas empíricas (ver Bazán, 1997), las cuales se apoyan en este modelo teórico que se constituye en normativo.

El artículo está organizado de la siguiente manera. La sección 2 presenta los principios y supuestos subyacentes a la psicometria que generalmente no son explícitos en la literatura psicométrica. El modelo matemático de la medición psicométrica es presentado en la sección 3 y el modelo estadístico de dicha medición es presentado en la sección 4. En la sección 5, se presenta una propuesta para la definición de validez debida a Zimmerman (1983). En la sección 6, se da alcances finales con respecto a lo presentado en las secciones 2, 3, 4 y 5.

1. Principios y Supuestos de la Psicometría

La Psicometría podemos definirla como una temática de la Psicología que a través de la Estadística y la Matemática modela la medición psicológica y sus principales aspectos conexos, así como el escalamiento y la construcción de escalas, test o pruebas, que se constituyen en instrumentos psicométricos. Así la Psicometría tiene dos aspectos que la caracterizan: teórico, ligados a la fundamentación y justificación de la medición psicológica, y práctico, ligado a las técnicas para la elaboración de los instrumentos psicométricos.

Lo que la Psicometría recoge de la Teoría Psicológica muchas veces no es explícito, tanto en textos como en manuales y artículos. En la obra de Kantor, J. (1978) puede revisarse un "ejemplo de construcción sistemática", inusual en la diversas teorías psicológicas. De lo que considera Kantor, "intentos históricos para hacer científica a la Psicología desde la teoría lógica", presentamos, desde una perspectiva más bien histórica, algunos "principios" que consideramos subyacentes a la Teoría Psicométrica.

1. *La Psicología es una ciencia cuantitativa*, (viene de Herbart), aplicable a cosas y eventos no reales (constructos de las acciones) y son indicadores de entidades psíquicas.

2. *La Psicología es una ciencia experimental*, (viene de Weber-Fechner) que utiliza métodos psicofísicos analógicos al comportamiento orgánico siendo una experimentación indirecta.

3. *La Psicología es una ciencia paralela, isomorfa a la física,* (viene de Wundt), lo que se observa son movimientos y cambios fisiológicos que son "isomorfos" a movimientos y cambios psíquicos.

4. *La Psicología es una ciencia analógica,* (viene de Herbart-Wundt-James), las entidades mentales son factores similares a las entidades físicas o naturales, (los elementos psíquicos están relacionados con los factores físicos y fisiológicos).

5. *Los procesos o funciones psicológicas pueden establecerse mediante operaciones y, por tanto, definirse operacionalmente.* No interesa la descripción de las cosas sino los procesos, pues permiten establecer la existencia y estabilidad de algo mediante la ejecución de cierto tipo de operaciones.

6. *La Psicología es traducible a lo físico.* Lo psíquico es idéntico a lo físico, la conducta es observable y, por tanto, está atravesando por una experiencia psíquica mientras actúa de determinada manera.

7. *La Psicología es una ciencia deductiva,* pues es posible construir un sistema psicológico deductivo.

Por otro lado, Raven (1989) ha sistematizado los siguientes supuestos, comúnmente hechos por los psicómetras y usuarios de tests:

- Los puntajes de Test deben ser "normalmente"distribuidos.
- Los puntajes deben tener distribución amplia de manera que recoja las diferencias individuales.
- Los puntajes de test deben tener alta consistencia interna.
- Los psicómetras no deben confundirse con sus respuestas sociales o creencias políticas.
- Los individuos pueden ser descritos independientemente de la situación o contexto en el cual se encuentran.
- Debemos encontrar, del modo más convincente, una manera parsimoniosa de manejar las diferencias individuales para buscar un número pequeño de variables que pueden resumir la varianza total de intereses y habilidades humanas.
- Es importante encontrar maneras de hacer finas discriminaciones entre individuos a lo largo de un número pequeño de dimensiones.
- La práctica de hacer enunciados descriptivos acerca de los individuos no es científica.
- Hacer discriminaciones finas en un número pequeño de dimensiones es "objetivo", antes que un enunciado acerca de las propiedades distintivas, o acerca de los efectos de programas educacionales, sociales y políticos.
- Para la utilidad científica o para propósitos prácticos, cualquier rasgo que es indexado debe ser estable.

Muchos de ambos, principios y supuestos, subyacen también en este trabajo. No discutiremos aquí su validez sino se busca dar sus alcances a través del modelo presentado, para ser explícito aquello que aparece como poco riguroso en mucho de los libros clásicos de psicometría de que disponemos y, por tanto, no sujeto de demostración y sí sujeto de aceptación tácita. Nuestra posición debe ser crítica frente a este modelo, pero toda crítica debe basarse en un profundo y riguroso conocimiento.

2. Modelo Matemático de la Medición Psicométrica

2.1. Medición Psicométrica

Los psicólogos coinciden en reconocer la importancia de la medición de las propiedades psicológicas del individuo. Así Thorndike escribió en 1913: "todo lo que existe, existe en cierta cantidad; todo lo que existe en cierta cantidad puede ser medido".

Como es fácil conjeturar, las *propiedades psicológicas del individuo no son observables.* Por lo tanto, responde a la pregunta ¿cómo medir las propiedades psicológicas? constituye el problema fundamental de la medición psicológica. El área temática que aborda el problema es la Psicometría.

El supuesto fundamental en el que se apoya la Psicometría es el que a través de los instrumentos psicométricos podemos "medir" las propiedades psicológicas.

Esquemáticamente:

Figura 1: Supuesto fundamental de los instrumentos psicométricos

En el contexto de la Psicología, dada una propiedad psicológica no observable ("*propiedad medible*") "traducimos" dicha propiedad a través del instrumento psicométrico en una propiedad psicológica observable ("*propiedad medida*") que es una definición operacional de la propiedad expresada en números. De suerte que, si estamos midiendo dicha propiedad en una población de individuos en un instante específico, es decir, en una aplicación, pasamos de una población de individuos a una población de puntajes. En este contexto, *medir consiste en pasar del mundo de las propiedades de los individuos, respecto a una propiedad medible, al mundo de las propiedades entre los números* (en especial en los reales), respecto a esa misma propiedad medida.

Por otro lado, en el contexto de la Teoría matemática de medición, como ha explicado Valdivieso (1991 a) (1991 b), siguiendo a Pfanzagl (1973), *medir significa estimar una propiedad dentro de un sistema relacional dado*, es decir los sujetos de la medición son las propiedades del objeto "medible". Existen dos sistemas relacionales: numéricos y empíricos, donde el primero está conformado por los reales y por relaciones entre sus elementos, las que determinarán la escala a emplear, la cual denotamos por m, y el segundo por el conjunto empírico de propiedades a medir y por relaciones entre estas propiedades.

Con esto, el instrumento psicométrico es una *medición* de una propiedad medible definida como constructo, es decir la definición operacional de la propiedad medible especificando su proceso de medición empírica, y la escala de medición es una función que relaciona la propiedad medible directamente con la propiedad medida definida en forma teórica.

De manera que se plantea un paralelismo entre la escala de medición y el instrumento psicométrico, así la escala tiene una interpretación más sintáctica (formal) y el instrumento una interpretación más semántica (empírica). Es, por ello, que el instrumento psicométrico visto como una forma empírica se conoce como escala o prueba. Esquemáticamente podemos presentar lo siguiente:

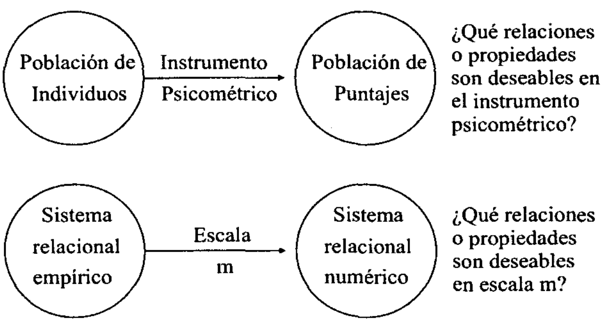

Figura 2: Paralelismo en la medición psicométrica de la escala de medición y el instrumento psicométrico.

Las preguntas que aparecen en la figura surgen de manera natural, con el modelo presentado en las secciones siguientes, pretendemos hacerlas explícitas.

Para estudiar las relaciones y propiedades tanto de la escala como del instrumento psicométrico es necesario formular un modelo matemático y un modelo estadístico. En las siguientes secciones desarrollamos estos modelos.

2.2. Modelo Matemático de las Escalas de Medición

Definiciones

Sea I un subconjunto contable que denota una población de individuos sometidos a un test, T un conjunto que denota una población de propiedades medibles de $card(T) = I$ y sea X un subconjunto de números reales que denota una población de propiedades medidas de $card(T) = I$. Considere $(R_j)_{j\in L}$, la familia de relaciones entre los $t \in T$ definida por las propiedades deseables del instrumento psicométrico con respecto a cada propiedad medible en su forma de función característica, y sea $(S_j)_{j\in L}$ que denota la familia de relaciones entre las $x \in X$, definida por las propiedades deseables del instrumento psicométrico con respecto a cada propiedad medida en su forma de función característica. En ambos casos L expresa un conjunto arbitrario pero usualmente finito de tales relaciones.

Un sistema relacional empírico (s.r.e) $\boldsymbol{T} = \langle T, (R_j)_{j\in L}\rangle$ es una estructura matemática formada por el constructo T y las relaciones entre sus elementos. Análogamente se define un *sistema relacional numérico* (s.r.n) correspondiente $\boldsymbol{X} = \langle X, (S_j)_{j\in L}\rangle$,

Una escala de medición entre un s.r.e irreducible y un s.r.n es una función $\boldsymbol{m} : \boldsymbol{T} \to \boldsymbol{X}$, $\boldsymbol{m}([t]) = [x]$ que preserva las relaciones empíricas a través de relaciones numéricas en los s.r dados, es decir lleva los objetos $[t]$ (la propiedad medible y sus relaciones) a objetos [x] (la propiedad medida y sus relaciones). En un sentido más general puede decirse que $\boldsymbol{m}(t_i) = x_i$ denota la medición de la propiedad t en el individuo i la cual podemos escribir como x_i^t.

En palabras simples, el término irreducible significa que $\boldsymbol{T}$ debe garantizar la inyectividad de la escala $\boldsymbol{m}$, es decir para cada valor numérico x existe una propiedad medible t correspondiente a un sólo individuo. Si $\boldsymbol{T}$ no es irreducible, no hay inconveniente pues es posible a partir de $\boldsymbol{T}$, construir un sistema relacional cociente asociado a una relación de congruencia (relación que permite sustituir internamente en $\boldsymbol{T}$ objetos equivalentes en relación a la propiedad medible). Este nuevo sistema resulta irreducible. Mayores detalles pueden consultarse en Pfanzagl (1973) y Valdivieso (1991a, 1991b).

Propiedades del instrumento y de la escala para el modelo psicométrico:

Considerando las definiciones anteriores, proponemos que las propiedades deseables del modelo psicométrico son:

Para el instrumento psicométrico

1. Debe medir cierta propiedad t unívocamente y no otra. (existencia y unicidad para la propiedad medible t), a esto se denominará la *validez* del instrumento psicométrico.

2. Debe proporcionar la misma medida x (valor numérico) de la propiedad medible t al volver a medir bajo "condiciones similares" al mismo individuo i (invarianza para la propiedad medida x), a esto se denominará la *confiabilidad* del instrumento psicométrico.

3. Debe poseer sensibilidad para diferenciar a los individuos con respecto a la propiedad medible t a través de la propiedad medida x (invarianza para las relaciones de orden, entre los individuos sometidos a un test con respecto a la posesión de la propiedad t), a esto se denominará la *diferenciabilidad* del instrumento psicométrico.

Para la escala psicométrica

1. Debe garantizar una "traducción" de la propiedad medible en una propiedad medida. Es decir la escala debe ser un *Homomorfismo*, esto es, $\boldsymbol{T}$ es homomorfo a $\boldsymbol{X}$ si $\forall\, l \in L$, $t_{(li)} \in T^{li}$, entonces se cumple que $R_i(t_1, t_2, ..., t_{(li)}) = S_i(x_i, x_2, ..., x_{(li)})$

2. Debe garantizar una traducción única de la propiedad medible en una propiedad medida. Es decir la escala debe ser *única*. Dada la clase de homomorfismos de $\boldsymbol{T}$ en $\boldsymbol{X}$ denotada por $\mathbf{M}(\boldsymbol{T}, \boldsymbol{X})$ podemos definir un conjunto de transformaciones admisibles de la escala $m_0 \in \mathbf{M}(\boldsymbol{T}, \boldsymbol{X})$ denotada por $\Gamma_{\boldsymbol{X}}(m_0(\boldsymbol{T}))$ de manera que sea única.

3. Debe garantizar la traducción de la propiedad medible en una propiedad medida que corresponda por lo menos a una medición de intervalo. Es decir la escala debe ser *Intervalar*. Dada $\boldsymbol{m}_0 \in \mathbf{M}(\boldsymbol{T}, \boldsymbol{X})$, se dice que es intervalar si $\Gamma_{\boldsymbol{X}}(\boldsymbol{m}_0(\boldsymbol{T})) = \{\gamma : \Re \rightarrow \Re / \gamma(x) = \alpha x + \beta, \forall x \in \Re, \alpha \in \Re^+ \wedge \beta \in \Re\}$ donde $\Re$ denota el conjunto de los reales. Esta escala corresponde a una escala donde son válidas las transformaciones que corresponde a dilataciones (parámetro α) y traslaciones (parámetro β) de los puntajes.

3. Modelo Estadístico de la Medición Psicométrica

En el modelo matemático de la medición psicométrica se pudo establecer relaciones y propiedades para la escala y el instrumento psicométrico. Sin embargo, el modelo no es un modelo que pueda ser verificado empíricamente. Adicionalmente, en situaciones comunes de medición psicométrica mediante test es apropiado establecer que las relaciones entre la propiedad medible T y la propiedad medida X están sujetas a un error aleatorio considerando que X es una medida indirecta de T y por tanto susceptible de error para una población I determinada. Para esto supondremos una sola propiedad medible para un test la cual puede ser medida a través de diversos estímulos o elementos j del test.

Inicialmente considere el siguiente experimento aleatorio dentro de esta sección Ξ: dado un individuo i al que se aplica un cierto estímulo j sucesivas veces para medir cierta propiedad t determinada. Estos estímulos generalmente son denominados ítems pero pueden ser tests que componen un test mayor.

$\boldsymbol{S} = \{s_1, s_2, ..., s_k, ...\}$ es el espacio muestral asociado, donde s_k el resultado de la aplicación del estímulo j al individuo i en la repetición k.

Una función X_{ij} llamado ***puntaje observado del individuo i en el estímulo j***, que asigna a cada uno de los elementos $s \in \boldsymbol{S}$ un número real $X_{ij}(s)$ llamado valor del puntaje observado correspondiente o simplemente puntuación del individuo i en el ítem j), se considera una variable aleatoria.

Esquemáticamente tenemos:

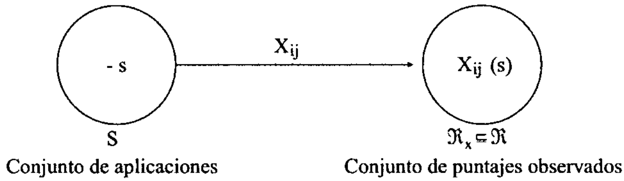

Figura 3: Definición de la variable aleatoria puntaje observado

Bajo las condiciones consideradas tenemos que X_{ij} es aleatoria puesto que la elección del individuo proviene de una muestra aleatoria de la población y el estímulo también es elegido de una población aleatoria de estímulos (test o ítems), más precisamente, doblemente aleatoria o experimento compuesto.

Al proceso de obtener el resultado de aplicar un cierto estímulo j elegido al azar, un conjunto amplio de veces, a un cierto individuo elegido también al azar para medir cierta propiedad determinada, lo denominaremos una medición mediante un instrumento psicométrico o simplemente *medición psicométrica* y constituye una definición semántica paralela a la que correpondía a la escala de medición en el modelo determinístico visto en la sección anterior.

Esta especificación de la variable aleatoria X_{ij} supone que la aleatoriedad corresponde a las repeticiones y una muestra aleatoria de ellas pude ser obtenida bajo una medición psicométrica.

Precisemos ahora otro experimento aleatorio Ξ^*, consistente apenas en seleccionar individuos donde el espacio asociado a Ξ^* es la población I de individuos, el cual corresponde a la medición psicométrica mediante un único estímulo .

La variable aleatoria X que asigna a cada elemento i en I (a cada individuo) uno y solamente un número real $x = X(i)$ correspondiente a un valor obtenido por un individuo en un instrumento psicométrico de una propiedad medible de estímulo único se llama solamente *puntaje observado*, y es tal que, Rango$(X) = R_X = \{x \in \Re / x = X(i), i \in I\} = X(I)$, $Dom.(X) = I$. En este caso la aleatoriedad está referida a los diferentes valores que toma X para diferentes individuos.

En Psicometría se tiene dos modelos estadísticos para la medición psicométrica, uno asociado a X_{ij} y otro asociado a X. En lo que sigue, presentamos el modelo estadístico de la medición psicométrica correspondiente a X denominado teoría del puntaje verdadero o Teoría clásica de los Tests.

3.1. Axiomas del Modelo Lineal Clásico

Sea T la variable aleatoria *puntaje verdadero* asociado a X que toma valores t correspondiente al verdadero valor, de la propiedad medible en tales individuos, esto es, Rango$(T) = R_T = \{t/t = T(i), i \in I\} = T(I)$, $Dom.(T) = I$.

Definición 1. *Definimos la variable aleatoria puntaje de error ε por la relación*

$$\varepsilon = X - T, \text{ en } I \tag{1.1}$$

Observación 1. *Pese a que la tríada (X, T, ε) constituye una medición, a veces nos referimos a X como tal, y conviene precisar que nos referimos a la aplicación de un estímulo a una subpoblación de n individuos, por lo que en verdad se tendría $\varepsilon_{ij} = X_{ij} - T_i$, donde $j = 1$, $i \in \{1, 2, ..., n\}$. En sentido práctico j puede ser considerado como un test o un ítem.*

Axiomas del Modelo Lineal Clásico

Para toda subpoblación no nula de I, es decir, $\forall N \subset I$, $N \neq \phi$, $N \in 2^I$, se asume lo siguiente:

$$E(\varepsilon) = 0 \tag{1.2}$$

$$r(T, \varepsilon) = 0 \tag{1.3}$$

$$r(\varepsilon_1, \varepsilon_2) = 0 \tag{1.4}$$

$$r(\varepsilon_1, T_2) = 0 \tag{1.5}$$

donde E denota el operador esperanza y r denota la correlación de las variables en el paréntesis y el subíndice dos mediciones diferentes cualesquiera (ver Bickel e Dobsum (1977), que pueden ser la medida de la propiedad t para dos individuos distintos en una misma ocasión a la medida de la propiedad t en dos ocasiones distintas.

También para evitar la trivialidad, asumamos que:

$$0 < V(X) < \infty, 0 < V(T) < \infty \text{ y } V(\varepsilon) < \infty \qquad (1.6)$$

Observación 2. *Hay que observar que tanto las definiciones como los resultados obtenidos en este estudio se cumplen para el modelo poblacional pues en todo momento se trabaja con parámetros poblacionales.*

Proposición 1. *Para toda subpoblación no nula de I, los siguientes grupos de axiomas son equivalentes:*

$$\begin{array}{lll} \text{(I)} & E(X) = E(T) \;\; ... & \text{(i)} \\ & E(X_1T_2) = E(T_1T_2) \;\; ... & \text{(ii)} \\ & E(X_1X_2) = E(X_1T_2) \;\; ... & \text{(iii)} \\ \text{(II)} & \text{(1.2) a (1.5)} & \end{array}$$

Corolario 1. *Para toda subpoblación no nula de I y para cualesquiera dos mediciones se tiene:*

$$COV\,(X_1, X_2) = COV\,(X_1, T_2) = COV\,(T_1, T_2) = COV\,(T_1, X_2).$$

En particular $COV\,(X, T) = V(T)$.

3.2. Propiedades Derivadas para toda Medición

Proposición 2. *Para toda medición se cumple lo siguiente*

$$E(T) = E(X) \qquad (2.1)$$

$$E(X/T = t) = t \qquad (2.2)$$

$$V(X) = V(T) + V(\varepsilon) \qquad (2.3)$$

$$r^2(X,T) = V(T)/V(X) = 1 - V(\varepsilon)/V(X) \qquad (2.4)$$

$$r^2(X,T) + r^2(X,\varepsilon) = 1 \qquad (2.5)$$

3.3. Relaciones Basadas en Mediciones Estrictamente Paralelas

Definición 2. *Sean dos mediciones X y $X^{'}$ de una misma propiedad medible que satisfacen (1.1) con distintro error pero de varianza homogénea y obviamente los supuestos (1.2) a (1.5), tales mediciones son llamadas* mediciones estrictamente paralelas. *Así satisfacen lo siguiente: Para toda subpoblación no nula de I, cualesquiera dos mediciones son estrictamente paralelas si se tiene:*

$$X = T + \varepsilon, \quad X^{'} = T + \varepsilon^{'} \quad y \quad V(\varepsilon) = V(\varepsilon^{'}) \tag{3.1}$$

Observación 3. *En verdad, para dos mediciones estrictamente paralelas se tiene $T = T^{'}$ y se puede adelantar aquí que los tests paralelos tendrán igual promedio de puntajes observados, varianzas y correlaciones con otros puntajes observados, pero no necesariamente perfectamente correlacionados entre ellos (a menos que no halla varianza de error).*

Proposición 3. *Para toda subpoblación no nula de I y para cualesquiera dos mediciones estrictamente paralelas de una misma propiedad medida, se cumple:*

$$E(X) = E(X^{'}) \text{ y } V(X) = V(X^{'}) \tag{3.2}$$

$$COV(X, X^{'}) = V(T) \tag{3.3}$$

$$r^2(X, Y) = r(X, X^{'}) \tag{3.4}$$

$$V(\varepsilon) = V(X)[1 - r(X, X^{'})] \tag{3.5}$$

$$r(X, \varepsilon) = \sqrt{1 - r(X, X')} \tag{3.6}$$

Observación 4. *El recíproco de (3.2) no necesariamente es cierto.*

3.4. La Confiabilidad de un Test

Definición 3. *La confiabilidad de un test es definida como la correlación al cuadrado $r^2(X, T)$ entre puntaje observado y puntaje verdadero.*

Proposición 4. *La confiabilidad es un número en el intervalo $< 0, 1]$.*

3.5. La Validez de un Test

Definición 4. *El coeficiente de validez de una medición X con respecto a un criterio Y (una segunda medición) es definida como el valor absoluto de su coeficiente de correlación, es decir, $|r(X,Y)|$.*

Proposición 5. *La confiabilidad de un test es justo su validez con respecto a un test paralelo.*

3.6. Fórmulas de Atenuación

Lema 1. *Dadas X e Y dos variables aleatorias psicológicas con errores de medición que satisfacen (1.1), se cumple lo siguiente:*

$$r(T_X,T_Y) = r(X,T_Y)/r(X,T_X) = r(Y,T_X)/r(Y,T_Y) \tag{6.1}$$

$$r(X,Y) = r(X,T_Y)r(Y,T_Y) = r(Y,T_X)r(X,T_x) \tag{6.2}$$

donde T_X, T_Y denotan los puntajes verdaderos de las mediciones distintas correspondientes a X e Y respectivamente.

Observación 5. *El cociente en (6.1) no se indetermina puesto que $r(X,T_X), r(Y,T_Y) > 0$, por la Proposición 1, dado que ambas expresan los índices de confiabilidad de las mediciones correspondientes.*

Proposición 6. *Dadas X e Y variables aleatorias psicológicas con errores de medición que satisfacen (1.1), se cumple lo siguiente:*

$$r(T_X,T_Y) = r(X,Y)/[r(X,T_X)r(Y,T_Y)] = r(X,Y)/[r(X,X')r(Y,Y')]^{1/2} \tag{6.3}$$

$$r(X,Y) = r(X,T_Y)[r(Y,Y')]^{1/2} - r(Y,T_X)[r(X,X')]^{1/2} \tag{6.4}$$

$$r(X,T_X) \geq r(X,T_Y) \text{ y } r(Y,T_Y) \geq r(Y,T_X) \tag{6.5}$$

$$|r(T_X,T_Y)| \geq |r(X,T_Y)| \geq |r(X,Y)| \text{ y}$$

$$|r(T_X,T_Y)| \geq |r(Y,T_X)| \geq |r(X,Y)| \tag{6.6}$$

Proposición 7. *La validez de un test con respecto a cualquier criterio no puede exceder del índice de confiabilidad. Si excede a la confiabilidad entonces el segundo puntaje observado correlaciona más altamente con el primer puntaje verdadero que con el primer puntaje observado.*

4. Propuesta de Zimmerman para la Definición de Validez

Zimmerman (1983) ha presentado una nueva propuesta para la definición de validez que presentamos a continuacíón.

Proposición 8. *Sean X e Y dos variables aleatorias de la misma propiedad, tal que $X = T_X + e_X$, y $Y = T_Y + e_Y$. Se tiene*

$$COV(X,Y) = COV(T_X,T_Y)) + COV(e_X,e_Y) \tag{7.1}$$

4.1. Definición de Validez según Zimmerman

Zimmerman (1983) propone definir la **validez** de una medición X con respecto a un criterio Y como $|V_{XY}|$, donde

$$V_{XY} = \frac{COV(T_X,T_Y)}{\sqrt{V(X)V(Y)}} \tag{7.2}$$

4.2. Relaciones entre la Definición de Validez de Zimmerman y la Usual

Si denotamos $r(X,Y)$ en (4.5) por ρ_{XY} tenemos las siguientes relaciones entre V_{XY} y ρ_{XY}

Proposición 9. *Se cumple*

$$\frac{V_{XY}}{\rho_{XY}} = \frac{COV(T_X,T_Y)}{COV(X,Y)} \tag{7.4}$$

$$\rho_{XY} - V_{XY} = \rho(e_X,e_Y)[1 - r(X,X')]^{1/2}[1 - r(Y,Y')]^{1/2} \tag{7.5}$$

5. Discusión y Alcances Finales

Como los tests psicométricos empleados en el país, en su mayoría, se basan en la Teoría clásica de los Tests, la cual a su vez se apoya en el modelo con error en medición formulado por Spearman en 1904 y extendido desde allí, nos propusimos abordar este tema para así contribuir a su esclarecimiento y difusión. Nuestro abordaje se basó en una revisión de la literatura pertinente, pero adicionalmente contiene resultados y derivaciones propias. Una justificación adicional lo constituye el tratamiento como modelo y su conexión desde la teoría matemática de medición y la justificación de la aleatoriedad de los puntajes observados.

Con respecto a los Principios y supuestos de la Psicometría

Un primer alcance a establecer corresponde a los principios y supuestos subyacentes a la teoría psicométrica. Los principios identificados en la sección 2 pueden reinterpretarse en: *Los constructos psicológicos existen, pero no son observables y sin embargo son susceptibles de experimentación indirecta. Los constructos responden, isomórficamente, a estímulos como la conducta medible, por tanto los constructos psicológicos pueden ser medidos. Los constructos psicológicos son operacionalizados para ser medidos empíricamente. El constructo operacionalizado se traduce por conductas obsevables. Por lo tanto es posible establecer modelos a través de un sistema psicológico deductivo.*

Por otro lado, los supuestos subyacentes a la Psicometría pueden reinterpretarse en: *Los puntajes observados en un test muestran regularidades, variabilidad y alta consistencia interna. Una condición requerida es la objetividad del evaluador y que la evaluación sea un experimento controlado. Lo que se busca es un Modelo Parsimonioso de la diferenciabilidad humana con alto poder de discriminación. La cuantificación mediante test es científica, objetiva y mide constructos estables.*

La validez de estos principios y supuestos no serán discutidos en este artículo.

Con respecto al modelo matemático de la medición psicométrica.

Con respecto a las Propiedades del instrumento y de la escala para el modelo psicométrico correpondientes a un modelo sintáctico presentado en la sección 3, podemos dar el alcance de que estas propiedades son difíciles de cumplir en la práctica, tanto por teoría de test que exige por lo menos una medición de intervalo, como en el escalamiento o construcción de pruebas que toman medición desde el nivel nominal como puede verse, por ejemplo, en Magnusson(1990).

El modelo determinístico de las escalas de medición, plantea un homomorfismo de la escala m, $m(t_i) = x_i \ \forall i \in I$, donde x_i es la propiedad medida (en verdad x_i^t). Sin embargo en la sección 4, x_i tomará la forma de puntaje observado, cuando se precisa el modelo probabilístico por el cual se obtiene x esto es, mediante el experimento aleatorio compuesto propuesto en el modelo TCT, de manera que el valor proporcionado es obtenido por un instrumento psicométrico.

Así el modelo matemático es un modelo normativo dentro del cual fue posible definir formalmente tres propiedades deseables para el instrumento psicométrico como son la validez, confiabilidad y diferenciabilidad, y también 3 propiedades que la escala debe satisfacer como son que sea un homomorfismo, sea única e intervalar.

Con respecto al modelo estadístico de la medición psicométrica

En la sección 4 fueron presentados un conjunto de resultados de la axiomática propuesta sin mayores interpretaciones. A continuación presentamos las interpretaciones de los principales enunciados.

De la definición (1.1), se puede interpretar que el puntaje observado, X, de un individuo es la suma de dos partes un puntaje verdadero, T, y un puntaje de error o error de medición, ε. T y ε son constructos teóricos que no son observables.

(1.2) es equivalente a asumir $E(X) = T$, que se interpreta como que el valor esperado (el promedio poblacional) de X es T. T es el promedio de la *distribución teórica* de los puntajes X que se hallarían en repetidas aplicaciones independientes del mismo test al mismo individuo. Hay que hacer notar que T se define en términos del valor esperado de los puntajes, no en términos de una propiedad real del examinado. El puntaje verdadero no necesariamente refleja "el verdadero" nivel de la propiedad a medir, esto corresponde a un problema de validez. (1.3) se interpreta como que puntajes de error y puntajes verdaderos obtenidos por una población de individuos en un test no están correlacionados. (1.4) se interpreta como que los puntajes de error, ε_1 y ε_2 en dos tests diferentes no están correlacionados. (1.5) se interpreta como que los puntajes de error en un test, ε_1 no están correlacionados con los puntajes verdaderos en otro test, T_2.

En Bohrntedt (1978) y Lord y Novick (1974) se presenta una construcción del modelo en base a los axiomas equivalentes. Estas caracterizaciones equivalentes son comunes en Matemática y prueban que los axiomas de cualquier sistema deductivo no son únicos.

Otros autores toman a (3.2) como definición de medidas paralelas reservando (3.4) como definición de estrictamente paralela.

La expresión $E(X) = E(T)$ es a menudo referida como *puntaje medio observado* o *puntaje medio verdadero.* La varianza de los puntajes de error $V(\varepsilon)$ se refiere a la varianza de errores de medición o simplemente a la *varianza de error*, y su raíz cuadrada σ_ε es referida como desviación estándar de los errores de medición o simplemente como el *error estándar de medición.* Cada una de las expresiones obtenidas en las secciones previas tienen un uso importante en la Teoría de los tests que no trataremos de interpretar aquí, pero puede revisarse a Magnusson (1990) y Lord y Novick (1974).

Por (2.4) la confiabilidad es la proporción de la varianza total del puntaje observado que es explicada por la influencia lineal de T y por (2.2) diremos que es una medida del ajuste de esta relación lineal, o de otra forma, podemos decir que la confiabilidad de un test es una medida del grado de variación relativa del puntaje verdadero sobre la variación del puntaje observado. De esta relación también podemos ver que la confiabilidad es una medida inversa de la varianza relativa de error sobre la varianza observada.

La definición 3 es sólo teórica puesto que T es una variable aleatoria no observable y por tanto $r^2(X,T)$ también, sin embargo $r(X,X')$ si lo es y puede estimarse; como por (3.4) son iguales, cualesquiera de esas cantidades puede tomarse como una definición de confiabilidad. En Magnusson (1990), Lord y Novick (1974), Bazán (1993), y Kristof (1963) puede revisarse diversas formas de estimar la confiabilidad. La expresión $r(X,T)$ es algunas veces referida como el *índice de confiabilidad,* y la Proposición 3 es igualmente válida para dicho índice.

El coeficiente de validez de una medición o predictor puede además establecerse únicamente en relación a una segunda medición llamada un criterio. Aunque lo anterior no debe de olvidarse, es usual referirse a dicho coeficiente como la validez de un test.

Un problema de importancia para los psicólogos teóricos consiste en determinar la relación entre dos variables psicológicas teóricas, *propiedades medibles o latentes*, con instrumentos psicométricos construidos para medirlas. Si la relación entre dichos instrumentos psicométricos es mucho menor que la correlación entre las propiedades medidas (constructos) como se ve en (6.3), y si asumimos como aproximación razonable que el puntaje verdadero en la medición puede ser tomada como el constructo en la cuestión, entonces podemos usar ciertas fórmulas llamadas *fórmulas de atenuación*, para calcular la verdadera correlación entre los constructos. La ecuación derecha de (6.3) da la correlación entre puntajes verdaderos en términos de la correlación entre puntajes observados y la confiabilidad de cada observación. La ecuación (6.4), al despejarse, da la correlación entre puntaje observado en una medición y el puntaje verdadero en una segunda medición en términos de la correlación entre los puntajes observados en las dos mediciones y la confiabilidad en la segunda medición. Ambas son ecuaciones de estimación de cantidades no observables cuyas cotas se obtienen de (6.5) y (6.6), y son las llamadas fórmulas de atenuación.

De la proposición 7 se deduce que una alta confiabilidad es una condición necesaria pero no suficiente para una alta validez. Que la validez de un test (con respecto a otro test) pueda exceder su confiabilidad puede en principio parecer contraintuitivo. La paradoja, sin embargo, resulta del carácter no paralelo de la definiciones de confiabilidad y validez; así la última fue definida como un coeficiente de correlación y el primero como una correlación al cuadrado. Como $|\rho| \leq 1$, se tiene que $\rho^2 \leq |\rho|$, lo cual explica el resultado dado. Caso en el que la validez excede a la confiabilidad pueden ocurrir cuando la medida del predictor es quizás no confiable y la medida del criterio es relativamente confiable. La propuesta de Zimmerman (1983) ofrece mayores perspectivas.

Con respecto a la propuesta de Zimmerman para la definición de validez

Una primera cosa a decir con respecto a la definición propuesta por Zimmerman es que la definición de validez es “simétrica” a la de confiabilidad. En la definición de confiabilidad, esta se definía como una cantidad no observable sobre otra observable (véase 2.4) para el caso de la validez en la definición propuesta $COV\ (T_X, T_Y)$ no es observable y $V(X)$, $V(Y)$ si lo son. A diferencia de la definición de validez anterior, en ésta el signo depende de $COV\ (T_X, T_Y)$.

Ventajas de $\mathbf{V_{XY}}$ *frente a* $\rho_{\mathbf{XY}}$

(i) Independencia del axioma (1.4)

- Si se cumple el axioma (1.4), esto es $E(e_X, e_Y) = 0$, en (7.3) $V_{XY} = \rho_{XY}$ la cual es análoga a la definición convencional (ver sección 5).
- Si no se cumple el axioma (1.4), esto es $E(e_X, e_Y) \neq 0$, en (7.3) V_{XY} ofrece otras ventajas.

(ii) "Correcta"desde el punto de vista de la axiomática de la TCT.

Si X e Y son paralelos, $\Rightarrow V_{XY} = r(XX^{'})$ y no como $\rho_{XY} = 1$

En efecto si el predictor X y el criterio Y cumplen (3.1) se tiene $T_X = T_Y = T$ por lo que en (7.1) $COV\ (e_X, e_Y) = COV\ (X, Y) - COV\ (T_X, T_Y) = 0$, en base a (3.3) pues de $COV\ (X, Y) = COV(X, X^{'})$. Luego en (7.3), $V_{XY} = r(X, Y) = r(X, X^{'})$. Es decir la validez es interna, y la validez coincide con la confiabilidad, por lo que ésta última sería un caso particular de validez para predictores paralelos (paralelismo entre el predictor y el criterio). Así, el valor de la confiabilidad se puede interpretar como validez interna, lo cual resulta correcto desde el punto de vista de la axiomática considerada.

(iii) Si el predictor X y el predictor Y son variables aleatorias con distribución normal de media cero y varianza 1 i.e. $(X, Y \sim N(0, 1))$, esto es $E(X) = E(Y) = 0$, $V(X) = V(Y) = 1$, se tiene de (3.2) y (1.2) que X e Y son paralelos y por lo anterior se tiene que $V_{xy} = COV\ (X, Y)$.

Estos resultados sugieren que la definición inducida (7.2) es la "correcta" desde el punto de vista de la axiomática de la TCT que se trabaja.

De (7.2) se obtiene

$$V_{XY} = \rho_{XY} - \frac{COV\,(e_X, e_Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \tag{7.3}$$

que es directo al usar (7.1), separar los términos en el numerador y usar la definición de correlación.

(7.5) se puede interpretar como la discrepancia entre $\rho_{xy} - V_{xy}$ está en función del grado correlación entre su puntajes de error.

Con respecto a la atenuación del error de medición en la definición de Zimmerman se tiene que de (7.2) la $cov\,(T_X, T_Y) = V_{xy}[V(x)V(y)]^{1/2}$ y en la siguiente expresión de la correlación se tiene

$$r(T_X, T_Y) = (V_{xy}[V(x)V(y)]^{1/2})/[V(T_X)V(T_Y)]^{1/2},$$

así usando (2,4) y (3,4) se tiene:

$$r(T_X, T_Y) = \frac{V_{XY}}{\sqrt{r(X,X')}\sqrt{r(Y,Y')}} \tag{7.6}$$

La cual es análoga a la confiabilidad en (7.2) pues si $E(e_x, e_y) = 0 \Rightarrow V_{xy} = \rho_{xy}$. Pero si $E(e_x, e_y) \neq 0$ por (7.2) se tiene

$$r(T_X, T_Y) = \frac{COV\,(T_X, T_Y)}{\sqrt{V(X)V(Y)}\sqrt{r(X,X')}\sqrt{r(Y,Y')}},$$

y usando (7.1) para el numerador y separando y dándole la forma de la definición de correlación se tiene

$$r(T_X, T_Y) = \frac{r(X,Y)}{\sqrt{r(X,X')}\sqrt{r(Y,Y')}} - \frac{r(e_X, e_Y)}{\sqrt{\frac{V(X)}{V(e_X)}}\sqrt{\frac{V(Y)}{V(e_Y)}}} \tag{7.7}$$

Finalmente, es fácil deducir, como $|r(T_X, T_Y)| \leq 1$, que $|V_{XY}| \leq [V(T_x)/V(X)]^{1/2}[V(T_Y)/V(Y)]^{1/2}$, y por (2.4) y (3.4) se tiene

$$|V_{XY}| \leq \sqrt{r(X,X')r(Y,Y')} \tag{7.8}$$

Tanto como (7,7) y (7.8) no tienen correspondientes en la TCT (ver Lord y Novik, 1974); por lo que la definición propuesta de validez resulta ser más general y no necesita del axioma (1.4).

En este artículo revisamos uno de los modelos estadísticos más utilizados para la construción de un test psicométrico que es la Teoría Clásica de los Tests denominada también Teoría del puntaje verdadero. Esta se apoya en un modelo lineal con error de medición formulado por Spearman en 1904. En este modelo se establece una axiomática relativamente simple para los componentes del modelo; y diversas definiciones de equivalencia entre *mediciones* para una aplicación de un test a cierto individuo. El término medición se refiere aquí a un sustantivo que denota un elemento de un test,(ítem) o el test mismo y se refiere a diversas medidas de un mismo constructo. El propósito del artículo es ofrecer una presentación del modelo psicométrico de la Teoría Clásica de los Tests, el cual contiene parte del modelo lineal clásico para una medición en una población de individuos. A partir de él pueden derivarse el Modelo Lineal Clásico para un individuo en una población de ítems (para mediciones de orden k) donde se deriven las conocidas fórmulas de predicción de Spearman-Brown, Alfa de Cronbach, Lamda de Guttman, y las fórmulas de Kuder-Richardson, las cuales son cotas inferiores de la confiabilidad de un test compuesto. Esto es materia de otra entrega para no recargar la extensión de este artículo.

Finalmente, esta presentación también tiene el propósito de sentar las bases para motivar en nuestro medio futuras contribuciones disciplinarias, interdiciplinarias y multidiciplinarias como se dan, en extenso, en otros contextos.

Referencias

[1] APA: *Standars for Educational and Psychological Test and Manuals.* Washigton DC, (1954).

[2] APA: *Methematical Statistic.* Helden-Day. Springer, (1977).

[3] J. BAZÁN: *Modelos estadísticos, clásico y reciente, de la confiabilidad de un instrumento psicométrico.* III Coloquio Nacional de Estadística. UNMSM, (1993).

[4] J. BAZÁN: *Metodología estadística de construcción de pruebas. Una aplicación al estudio de actitudes hacia la Matemática en la UNALM.* Tesis de Ingienería Estadística. UNALM, (1997).

[5] G. Bohrnstedt: *Evaluación de la confiabilidad y validez en la medición de actitudes.* En: Summers, Medición de actitudes. (1978) 103-127.

[6] J.P. Guilford: *Psychometrics Methods.* New York; McGraw-Hill, (1954).

[7] H. Guilleksen: *Teory of Mental Test.* Nueva York; Wiley, (1950).

[8] Jones y Appelbaum: *Psychometrics Methods.* Ann. Rev. Psichol. 40 (1989) 23-43.

[9] D. Krantz, R. Luce, P. Suppes y Tversky: *Foundations of Measurement.* New York. Academic. 3 vol. (1971).

[10] W. Kristoh: *The Statistical Theory of stepped-up reliability coefficient when a test has been divided into several equivalents parts.* Psychometrika. 28 (1963) 221-238.

[11] F. Lord y M. Novic: *Statistical Theories of Mental Test Scores.* Reading Mass. Addison- Wesley. 2da reimpresión, (1974).

[12] R. Luce y L. Narens: *The mathematics underlying measurement on the continuum.* Science 236 (1986) 1527-32

[13] Megnussson: *Teoría de los Test.* Edit. Trillas, México, (1990) 77-149.

[14] L. Narens y R. Luce: *Abstract Measurement Theory.* Cambridge, Mass: MITT Press, (1986).

[15] Pfanzangl. *Theory of measurement.* 2da ed. reimpresión. Physica Verlag, (1973).

[16] J. Raven: *Questionable assumptions in Test Construction.* Bulletin of International Test Commission. N° 29 (1989) 67-95.

[17] W. Torgerson: *Methods of Scaling.* Library of Congress. USA, (1958).

[18] L. Valdivieso: *Teoría de Escalas de medición.* Tesis de Bachillerato de Estadística PUCP, (1991a).

[19] L. Valdivieso: *Escalas de medición.* Pro Mathematica, Vol V, N°9-10 (1991b) 51-65.

[20] Weiss y Davison: *Test theory and methods.* Ann. Rev. Psichol. 32 (1981) 629-58.

[21] D. Zimmerman: *The Mathematical Definition of test Validity.* Educational and Psychological Measurement. 43 (1983). 791-796.

Jorge Luis Bazán Guzmán
Sección Matemática, Departamento de Ciencias
Pontificia Universidad Católica del Perú
bazan.jl@pucp.edu.pe.

ANEXO

PRUEBA DE LAS PROPOSICIONES Y LEMAS

Prueba de la Proposición 1

⌈ Veamos (I)$\Rightarrow$ (II)

(2) $E(\varepsilon) = E(X - T) = E(X) - E(T) = 0$, por (1) y (i).

(3) Bastará probar que $COV\,(T, \varepsilon) = 0$, en efecto $COV\,(T,\varepsilon) = E(T\varepsilon) - E(T)E(\varepsilon) = E(T(X - T)) = E(TX) - E(T^2) = 0$, por (1), (2) y (ii).

(4) Bastará probar que $COV\,V(\varepsilon_1, \varepsilon_2) = 0$, en efecto $COV\,(\varepsilon_1,\varepsilon_2) = E(\varepsilon_1,\varepsilon_2) - E(\varepsilon_1)E(\varepsilon_2) = E((X_1 - T_1)(X_2 - T_2)) - E(X_1 - T_1)E(X_2 - T_2) = E(X_1X_2) - E(T_1X_2) - E(X_1T_2) + E(T_1T_2) - E(X_1)E(X_2) + E(T_1)E(X_2) + E(X_1)E(T_2) - E(T_1)E(T_2) = 0$ por (1), (ii) y (iii).

(5) Bastará probar que $COV\,(\varepsilon_1, T_2) = 0$, en efecto $COV\,(\varepsilon_1, T_2) = E(\varepsilon_1 T_2) - E(\varepsilon_1)E(T_2) = E((X_1 - T_1)T_2) - E(X_1 - T_1)E(T_2) = E(X_1T_2) - E(T_1T_2) - E(X_1T_2) + E(T_1T_2) = 0$, por (1), (ii), (iii).

(II)$\Rightarrow$ (I)

(i) $E(X) = E(T + \varepsilon) = E(T) + E(\varepsilon) = E(T)$, por (1) y (2).

(ii) $E(X_1T_2) = E((T_1 + \varepsilon_1)T_2) = E(T_1T_2) + E(\varepsilon_1, T_2) = E(T_1T_2)$, por (1) y (5).

(iii) Bastará probar que $E(X_1X_2) = E(T_1T_2)$, pues por (ii) se obtiene la igualdad deseada, en efecto: $E(X_1X_2) = E((T_1 + \varepsilon_1)(T_2 + \varepsilon_2)) = E(T_1T_2) + E(\varepsilon_1T_2) + E(T_1\varepsilon_2) + E(\varepsilon_1, \varepsilon_2) = E(T_1T_2)$, por (1), (4) y (5).⌋

Prueba del Corolario 1

⌈ Sigue directamente de (I) (ii) y (I) (iii).⌋

Prueba de la Proposición 2

⌈(2.1) Sigue directamente de (1.1) y de (1.2).

(2.2) Sigue directamente de que $E(\varepsilon/T = t) = 0, \forall \tau$ es un caso particular de $E(\varepsilon) = 0, \forall N \subset I,\ N \neq \phi$ y de (1.1).

(2.3) Sigue directamente de (1.1) y (1.3).

(2.4) Sigue directamente de $COV(X, T) = V(T)$ del Corolario 1 y de (2.3).

(2.5) Sigue directamente de $r^2(X, \varepsilon) = V(\varepsilon)/V(T)$ y de (2.4).⌋

Prueba de la proposición 3

⌈(3.2) Son triviales a partir de (3.1).

(3.3) Usando (1.1),(2.1),(1.3) y (1.4) se tiene que:
$COV\,(X, X^{'}) = E(XX^{'}) - E(X)E(X^{'}) = E[(T + \varepsilon)(T + \varepsilon^{'})] - E(T)E(T^{'}) = E(T^2) - E(\varepsilon T) + E(\varepsilon^{'} T) + E(\varepsilon\varepsilon^{'}) - E^2(T) = E(T^2) - E^2(T) = V(T)$ o directamente del Corolario 1.

(3.4) Usando (3.2),(3.3) y (2.4) se tiene:
$r(X, X^{'}) = COV\,(X, X^{'})/[V(X)V(X^{'})]^{1/2} = V(T)/V(X) = r^2(X, T)$.

(3.5) Usando (2.3), (3.3) y (3.4) se tiene:
$V(\varepsilon) = V(X - T) = V(X) - V(T) = V(X) - COV\,(XX^{'}) = V(X) - V(X)r(X, X^{'}) = V(X)[1 - r(X, X')]$.

(3.6) Usando (2.5) y (3.4) se tiene:
$r(X, \varepsilon) = \sqrt{1 - r(X, T)} = \sqrt{1 - r(X, X)}$.⌋

Prueba de la proposición 4

⌈ En efecto, como por (2.4) $r^2(X, T) = 1 - V(\varepsilon)/V(X)$, podemos ver que si la varianza de errores de medición es cero, entonces la confiabilidad de un test será la unidad, de otro modo, como por (1.6) $V(X) > V(\varepsilon)$, si la varianza de errores de medición tiende hacer igual a la varianza del puntaje observado, entonces por (2.3) la varianza del puntaje verdadero tendería a cero y la confiabilidad del test debe también tender a cero.⌋

Prueba de la Proposición 5

⌈ Si X e Y son medidas estrictamente paralelas por (3.2), (3.3) se tiene que: $r(X,Y) = V(T)/V(X) = r^2(X,T)$.⌋

Prueba del Lema 1

⌈(6.1) Usando (1.2), el Corolario 1, sustituyendo y multiplicando por una expresión equivalente a la unidad se tiene que:
$[r(T_X,T_Y)/COV(T_X,T_Y)^{1/2}V(X)]^{1/2}V(X) =$
$[COV\,(X,T_Y)/V(X)V(T_Y)][V(T_X V(X))/COV\,(X,T_X)]$
$= r(X,T_Y)/r(X,T_X)$. De manera análoga se tiene $r(T_X,T_Y)$
$= r(Y,T_Y)/r(Y,T_Y)$.

(6.2) Usando (1.2), la demostración de (2.4), sustituyendo y multiplicando por una expresión equivalente a la unidad, se tiene que:
$r(X,Y) = COV\,(X,Y)/V(X)^{1/2}V(Y)^{1/2}$
$= [COV\,(X,T_Y)/V(X)^{1/2}V(Y)^{1/2}]\ [COV\,(Y,T_Y)/V(T_Y)^{1/2}]$
$[V(T_Y)^{1/2}/V(T_Y)^{1/2}] = [COV\,(X,T_Y)/V(X)^{1/2}V(T_Y)^{1/2}]$
$[COV\,(Y,T_Y)/V(T_Y)^{1/2}V(Y)^{1/2}] = r(X,T_Y), r(Y,T_Y)$.
De manera análoga se tiene $r(X,Y) = r(Y,T_X), r(X,T_X)$⌋

Prueba de la Proposición 6

⌈(6.3) La primera igualdad sigue directamente de (6.2) y reemplazando en (6.1). La segunda igualdad se obtiene por (3.4).

(6.4) Sigue directamente de (6.2) y de (3.4).

(6.5) Sigue directamente de (6.1) y del hecho que los índices de confiabilidad $r(X,T_X)$ y $r(Y,T_Y)$ son menores que uno como se observó en la Proposición 3.

(6.6) De la observación 4 $r(T_X,T_Y)$ es del mismo signo que $r(X,T_Y)$, por lo que de (6.1)
$|r(T_X,T_Y)| \geq |r(X,T_Y)|$, $|r(T_X,T_Y)| \geq |r(Y,T_X)|$.
Análogamente, de (6.2)
$|r(X,T_Y)| \geq |r(X,Y)|$ y $|r(Y,T_X)| \geq |r(X,Y)|$.
De ambas desigualdades se concluye.⌋

Prueba de la Proposición 7:

⌈ En efecto para X' e Y' medida estrictamente paralelas para X e Y, se cumple: $r(X,Y) \leq r(X,T_X) = r(X,X')^{1/2}$ y $r(X,Y) \leq r(Y,T_Y) = r(Y,Y')^{1/2}$, por (6.5) y (3.4). Por otro lado si suponemos que $r(X,Y) > r(X,X') = r^2(X,T_X)$ entonces $r(X,Y)/r(X,T_X) > r(X,TX)$ y por (6.2) $r(Y,T_X) > r(X,T_X)$. Análogamente se prueba que $r(X,T_Y) \geq r(Y,T_Y)$, es decir $r(X,Y) > r(X,X') \Rightarrow r(Y,T_X) > r(X,T_X)$ y $r(X,Y) > r(Y,Y') \Rightarrow r(X,T_Y) > r(Y,T_Y)$.⌋

Prueba de la proposición 8:

⌈ Por definición $COV\,(X,Y) = E(X,Y) - E(X)E(Y) = E(T_X + e_X)(T_Y + e_Y) - E(T_X + e_x)E(T_Y + e_y)]$ multiplicando y tomando esperanza, al usar la definición de covarianza se tiene $COV\,(X,Y) = COV\,(T_X,T_Y) + COV\,(e_X,e_Y)$, pues usando (1.2) $COV\,(T_X,e_Y)$ y $COV\,(T_Y,e_X)$ son cero.⌋

Prueba de la Proposición 9

⌈ De la definición de correlación y de (7.2) se tiene

(7.4) Se obtiene dividiendo en forma directa

(7.5) Restando, se obtiene porque en el numerador $COV\,(e_x,e_y) = COV\,(X,Y) - COV\,(T_X,T_Y)$ por (7.1), y en el denominador se reemplaza la raíz cuadrada del despeje de (3.5) finalmente de la definición de correlación para e_X y e_Y. ⌋