Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72-101.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18*, 161-169.

Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology, 5*, 417-426.

Student (1908). The probable error of a mean. *Biometrika, 6,* 1-25.

*Instituto de Investigaciones*
*de la Facultad de Psicología*
*Universidad de Buenos Aires (UBA)*
*Buenos Aires - República Argentina*

Fecha de recepción: 26 de diciembre de 2002
Fecha de aceptación: 5 de agosto de 2003

$$= \rho_{XX'}(1-\rho_{XX'})\sigma^2_X(1-\rho_{XX'}+\rho_{XX'}) =$$

$$= \sigma^2_X \rho_{XX'}(1-\rho_{XX'})$$

## Referencias bibliográficas


Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory.* California: Brooks/Cole Publishing Company.

Attorresi, H.F., Galibert, M.S. & Aguerri, M.E. (2002). Modelo de medición de componente no observable. Una presentación formal de la axiomatización de la teoría clásica de tests [Model of non observable component measurement. A formal presentation of an axiomatic formulation of the classical test theory]. *Psicothema, 14*(3), 665-668.

Cámara Sánchez, A. (2000). Aportaciones de la matemática a la metodología económica [Contributions of mathematics to economical methodology]. *Psicothema, 12* (Supl. 2), 103-107.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Florida: Harcourt Brace Jovanovich College Publishers.

Galibert, M.S., Aguerri, M.E., Lozzia, G.S. & Attorresi, H.F. (2003). El modelo lineal de puntuaciones. Una propuesta de formulación [The punctuations linear model. A formulation is proposed]. *Investigaciones en Psicología,* Año 8 (2), 79-91.

Gulliksen, H. (1950). *Theory of mental tests*. NY: John Wiley.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental tests scores.* Reading, MA: Addison-Wesley.

Meyer, P. (1970). *Introductory probability and statistical applications.* Reading, MA: Addison-Wesley.

Myers, R. (1990). *Classical and modern regression with application.* California: Duxbury Press.

Muñiz, J. (1998). La medición de lo psicológico [Psychological measurement]. *Psicothema, 10*(1), 1-21.

Demostración

$$\xi = \hat{V} - V = \mu_X + \rho_{XX'}(X - \mu_X) - V =$$
$$= \mu_X + \rho_{XX'}(V + \varepsilon - \mu_X) - V =$$
$$= (1 - \rho_{XX'})(\mu_X - V) + \rho_{XX}\varepsilon \qquad (19)$$

Por lo que $\xi$ es combinación lineal de $\mu_X$ - $V$ y $\varepsilon$ que, por (12), son variables aleatorias independientes normalmente distribuidas. Por tanto, $\xi$ tiene distribución normal.

$$\mu_\xi = E(\xi) = E(\hat{V} - V) =$$
$$= E(\hat{V}) - E(V)$$

Por (15) $$= \mu_V - \mu_V = 0$$

$$\sigma_\xi^2 = Var(\hat{V} - V) =$$

Por (19) $$= Var((1 - \rho_{XX'})(\mu_X - V) + \rho_{XX'}\varepsilon) =$$
$$= (1 - \rho_{XX'})^2 Var(V) + \rho_{XX'}^2 Var(\varepsilon) =$$
$$= (1 - \rho_{XX'})^2 \sigma_V^2 + \rho_{XX'}^2 \sigma_\varepsilon^2 =$$

Por (6) $$= (1 - \rho_{XX'})^2 \rho_{XX'} \sigma_X^2 + \rho_{XX'}^2 (1 - \rho_{XX'}) \sigma_X^2 =$$

$$E(\sigma^2_{\varepsilon_V}) = \sigma^2_{\varepsilon_v}$$

Por transitividad

$$Var(\varepsilon) = \sigma^2_{\varepsilon_v}$$

Luego

$$\varepsilon_v \sim N(0, \sigma^2_{\varepsilon})$$

Corolario

$$\sigma^2_{X_v} = \sigma^2_{\varepsilon}$$

Por (1)

$$Var(X / V = v) = Var(V + \varepsilon / V = v) = Var(v + \varepsilon_v) = \sigma^2_{\varepsilon_v} = \sigma^2_{\varepsilon}$$

Por tanto

$$\sigma^2_{X_v} = \sigma^2_{\varepsilon}$$

*Proposición 2*

$\xi$ se distribuye normalmente con

$\mu_{\xi} = 0$ y $\sigma^2_{\xi} = \sigma^2_X \rho_{XX'}(1 - \rho_{XX'})$

## Proposiciones referenciadas en las deducciones de las distribuciones de los estadísticos

*Proposición 1*

Demostración

$$\sigma^2_{\varepsilon_v} = Var(\varepsilon_v) = \sigma^2_{\varepsilon}$$

De (7) se deduce que

$$\sigma^2_{\varepsilon_v} = Var(\varepsilon_v) = \sigma^2_{\varepsilon}$$

En efecto,

$$Var(\varepsilon) = E[(\varepsilon - \mu_{\varepsilon})^2]$$

por la propiedad de esperanza condicional

$$= E\{E[(\varepsilon - \mu_{\varepsilon})^2/V]\}$$

por (2) y (3)

$$= E\{E[(\varepsilon - \mu_{\varepsilon/V})^2/V]\}$$

$$= E(\sigma^2_{\varepsilon_V})$$

De (7) resulta $\sigma^2_{\varepsilon_V} = \sigma^2_{\varepsilon_v}$ (donde $v$ es un valor realizado de la variable aleatoria $V$) y, siendo $\sigma^2_{\varepsilon_v}$ una constante, resulta

$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)}$$

Observaciones:

- $E(X/y)$ es la esperanza de $X$ condicionada al suceso $\{Y = y\}$. Es decir, puede interpretarse como la media de la variable $X$ calculada sobre la subpoblación de individuos con valor $y$ en la variable $Y$.
- $E(X/y)$ es una función de *y* y por tanto es una variable aleatoria. En rigor, $E(X/y)$ es el *valor* de la variable aleatoria $E(X/Y)$.
- Puesto que $E(X/Y)$ es una variable aleatoria, puede calcularse su esperanza. Así podemos considerar $E[E(X/Y)]$, por ejemplo. Obsérvese que la esperanza interna se toma respecto a la distribución condicional de $X$ dado que $Y$ es igual a $y$, mientras que la esperanza exterior se toma respecto a la distribución de probabilidades de $Y$.

Propiedad de la esperanza condicional

$$E[E(X/Y)] = E(X)$$

*Demostración* (caso continuo solamente): por definición,

$$E(X / y) = \int_{-\infty}^{+\infty} xf(x / y)dx = \int_{-\infty}^{+\infty} x \frac{f(x, y)}{f(y)} dx$$

Por tanto

$$E[E(X / Y)] = \int_{-\infty}^{+\infty} E(X/y) f(y)dy = \int_{-\infty}^{+\infty} [\int_{-\infty}^{+\infty} x \frac{f(x, y)}{f(y)} dx] f(y)dy.$$

Si todas las esperanzas existen, es posible escribir la integral iterada anterior con el orden de integración invertido. Así:

$$E[E(X / Y)] = \int_{-\infty}^{+\infty} x[\int_{-\infty}^{+\infty} f(x, y)dy]dx = \int_{-\infty}^{+\infty} xf(x)dx = E(X)$$

**Anexo**
Esperanza condicional

Definiciones

*Caso continuo*

Si $(X,Y)$ es una variable aleatoria continua bidimensional, se define la *esperanza condicional* de $X$ dado un valor $y$ de $Y$ como

$$E(X/y) = \int_{-\infty}^{+\infty} xf(x/y)dx$$

donde $f(x/y)$ es el cociente entre la función de densidad conjunta $f(x,y)$ y la función de densidad marginal de $Y$, $f(y)$. Esto es, la función de densidad condicional de $X$ dado $Y$: .

$$f(x/y) = \frac{f(x,y)}{f(y)}$$

*Caso discreto*

Si $(X,Y)$ es una variable aleatoria discreta bidimensional, se define la *esperanza condicional* de $X$ dado un valor $y_j$ de $Y$ como

$$E(X/y_j) = \sum_{i=1}^{\infty} x_i p(x_i/y_j)$$

donde $p(x_i/y_j)$ es el cociente entre la función de probabilidad conjunta $p(x_i,y_j) = P(X = x_i, Y = y_j)$ y la función de probabilidad marginal de $Y$, $p(y_j) = P(Y = y_j)$. Esto es, la función de probabilidad condicional de $X$ dado $Y$ calculada en $x_i$, $y_j$:

ción, habría que tomar en cuenta el error de muestreo de los estadísticos $S^2$ y $\hat{\rho}_{XX'}$ ; en ese caso la distribución y la variancia de los estadísticos $X_V$ y $\xi$ que se utilizan para la deducción de dichos intervalos se modificarían, con lo que el verdadero coeficiente de confianza 1-$\alpha$ de los intervalos (11) y (17) sería menor que el especificado. Lo mismo ocurría a principios del Siglo XX cuando se utilizaba la distribución normal en la obtención del intervalo de confianza para una media con desvío estándar desconocido hasta que Student (1908) dedujo su distribución exacta en el muestreo. Lord y Novick (1968) no mencionaron el problema y se limitaron a ejemplificar el cálculo de los intervalos, suponiendo como conocidos los parámetros. Otros autores admiten que se trata de estimadores pero aparentemente no se percatan de que no se está considerando el error de muestreo de los mismos en la construcción de los intervalos. Hallar la distribución exacta de $X_V$ y $\xi$ cuando se reemplazan la variancia y la confiabilidad por sus estimadores involucraría desarrollos matemáticos muy complejos. En tanto esto no se logre, usar los intervalos de confianza que se han calculado en el presente desarrollo supone, en rigor, que la población se restringe al grupo normativo y no que éste es una muestra de aquélla. Por tanto, claramente la estimación del puntaje verdadero de un sujeto mediante un intervalo de confianza depende del grupo normativo.

## La verificación de supuestos

Por ser $V$ una variable inobservable, los procedimientos para inferir los puntajes verdaderos presentados en los Casos 1 y 2 tienen la limitación de que no es posible verificar, a partir de una sola medición de los sujetos, si se cumplen los supuestos de los modelos estadísticos subyacentes. De este carácter inobservable de la variable $V$ se sigue la necesidad de disponer de dos o más medidas paralelas, lo cual no está exento de dificultades en la práctica.

$$\frac{(1 - \text{ñ}_{XX'})\text{ì}_{X} + \text{ñ}_{XX'}X - (1 - \text{ñ}_{XX'})\text{ì}_{X}}{\text{ñ}_{XX'}} = X$$

Por tanto, corregir el estimador de la regresión con el fin de que sea insesgado al condicionarlo a $v$ es equivalente a ignorar la regresión y volver al Caso 1. Es por ello que el intervalo de confianza para $v$ mediante la regresión que plantean Lord y Novick (1968) usando la desigualdad de Chebychev es correcto sólo si se comprende en el mismo sentido que el obtenido en (16): un intervalo para una variable aleatoria $V$ de extremos aleatorios $\hat{V}$ cuyos valores quedan fijados *simultáneamente* al elegir al individuo. Lord y Novick (1968) no explicitan la deducción de dicho intervalo, sólo muestran su cálculo en un ejemplo en el que suponen homogeneidad de variancias. Como se ha señalado, este supuesto es necesario para la deducción del intervalo de confianza (11) del Caso 1 pero no para el (17) del Caso 2.

## Dependencia de los intervalos de confianza con respecto a los parámetros poblacionales

La deducción y el cálculo de los intervalos de confianza de las expresiones (11) y (17) suponen el conocimiento de la confiabilidad del test y de la variancia de los puntajes observados en la población. Estos parámetros se estiman sobre la base de una muestra de sujetos que constituyen el grupo de referencia o normativo. Establecido el grupo normativo, los valores estimados para la media, la variancia y la confiabilidad son *valores fijos* que suplen a los valores verdaderos de los parámetros en todos los desarrollos que anteceden. Por esta razón es un error conceptual designarlos con $\bar{X}$, $S^2$ o supraindicarlos con ^, como $\hat{\text{n}}_{XX'}$. En efecto, estas notaciones designan estimadores cuya distribución en el muestreo no está siendo considerada para el cálculo de los intervalos de confianza. En (11) y (17) se infiere el puntaje verdadero de un sujeto supuesto, conocidas la variancia y la confiabilidad del test (los parámetros en el grupo normativo). Si se pretendiera considerar el grupo normativo como una muestra representativa de la pobla-

Si se considerara, como en el Caso 1, que una vez elegido el individuo su puntaje verdadero $v$ es un valor fijo y se pretendiera utilizar como su estimador a $\hat{V}$, éste sería sesgado.

En efecto, por (14)

$$E(\hat{V} / V = v) = E[(1 - \tilde{n}_{XX'})\grave{\imath}_X + \tilde{n}_{XX'}X / V = v] =$$

$$= (1 - \tilde{n}_{XX'})\grave{\imath}_X + \tilde{n}_{XX'}E(X / V = v) =$$

Por (5) $$= (1 - \tilde{n}_{XX'})\grave{\imath}_X + \tilde{n}_{XX'}v \qquad (18)$$

Por lo que el sesgo de estimación es:

$$v - E(\hat{V} / V = v) = v - (1 - \tilde{n}_{XX'})\grave{\imath}_X + \tilde{n}_{XX'}v =$$

$$= (1 - \tilde{n}_{XX'})(v - \grave{\imath}_X)$$

Luego, para obtener mediante la regresión un estimador que al condicionarlo a $v$ sea insesgado, basta con despejar $v$ de (18), con lo que resulta la siguiente corrección para $\hat{V}$:

$$\frac{\hat{V} - (1 - \tilde{n}_{XX'})\grave{\imath}_X}{\tilde{n}_{XX'}}$$

Siendo $\hat{V} = (1 - \tilde{n}_{XX'})\grave{\imath}_X + \tilde{n}_{XX'}X$ al reemplazar en la expresión anterior se obtiene:

Despejando $V$

$$P(\hat{V} - z_{1-\frac{\alpha}{2}}\sigma_X\sqrt{\rho_{XX'}(1-\rho_{XX'})} \leq V \leq \hat{V} + z_{1-\frac{\alpha}{2}}\sigma_X\sqrt{\rho_{XX'}(1-\rho_{XX'})} = 1-\alpha$$

Se obtiene así el siguiente intervalo para la variable aleatoria $V$:

$$\hat{V} \pm z_{1-\frac{\alpha}{2}}\sigma_X\sqrt{\rho_{XX'}(1-\rho_{XX'})} \qquad (16)$$

Hecha la elección aleatoria de un individuo y habiéndosele administrado el test, quedan fijados tanto su puntaje verdadero como su puntaje predicho por la regresión; esto es, se obtienen dos valores fijos $v$ y $\hat{v}$ de las correspondientes variables aleatorias $V$ y $\hat{V}$. Mientras $v$ permanece desconocido, siendo el parámetro por estimar, $\hat{v}$ se obtiene a partir del puntaje observado $x$ mediante (14). Reemplazando en (16) la variable aleatoria $\hat{V}$ por su valor observado $\hat{v}$ el siguiente intervalo de confianza de nivel 1-$\alpha$ para $v$:

$$\hat{v} \pm z_{1-\frac{\alpha}{2}}\sigma_X\sqrt{\rho_{XX'}(1-\rho_{XX'})} \qquad (17)$$

## Discusión

### El supuesto de homoscedasticidad

La deducción del intervalo (17) no requiere del *supuesto de homoscedasticidad* ya que no se construye condicionando a cada valor fijo de $V$ sino considerando la variabilidad conjunta de $V$ y $\hat{V}$ en la población.

$$= \mu_X + \rho_{XX'}(E(X) - \mu_X) = \mu_X$$

Por (4) resulta

$$E(\hat{V}) = \mu_V$$

En la práctica, el objetivo es estimar el puntaje verdadero *v* de un sujeto a partir de su puntaje observado *x* aprovechando la información dada por la ecuación de regresión (14) que vincula las variables *V* y *X.* Pero al reemplazar el valor *x* de *X* en dicha ecuación se obtiene *E(V / X = x)*; lo cual no es el puntaje verdadero del sujeto sino el *puntaje verdadero medio* de la población de sujetos con puntaje observado *x.*

### Definición 5

Se designará con la letra ξ a la diferencia entre el puntaje predicho por la regresión $\hat{V}$ y el puntaje verdadero *V*, lo que se conoce en la literatura psicométrica como *error de estimación* (Lord & Novick, 1968). Es decir:

$$\xi = \hat{V} - V.$$

## Cálculo del intervalo de confianza

Siendo

$$\xi \sim N(0, \sigma_X^2 \rho_{XX'}(1 - \rho_{XX'})) \quad \text{(Ver Proposición 2 del Anexo),}$$

es posible hallar $z_{1-\frac{\alpha}{2}}$ de modo tal que:

$$P\left(-z_{1-\frac{\alpha}{2}} \le \frac{\hat{V} - V}{\sigma_X \sqrt{\rho_{XX'}(1 - \rho_{XX'})}} \le z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

### Caso 2. Intervalo de confianza para un valor aleatorio del puntaje verdadero

### Estimación de los puntajes verdaderos mediante la ecuación de regresión lineal sobre los puntajes observados.

*Supuestos*

1.- $V$ y $\varepsilon$ son variables aleatorias independientes normalmente distribuidas. (12)

2.-La regresión de $V$ sobre $X$ es lineal, en consecuencia

$$E(V/X) = \mu_V + \rho_{XV} \frac{\sigma_V}{\sigma_X}(X - \mu_X) \tag{13}$$

La $E(V/X)$ se denomina *predicho de V* por la regresión y se la anota $\hat{V}$, la cual es una variable aleatoria por ser función de la variable aleatoria $X$.

*Propiedades*

a.- De (4) y (6) se deduce que:

$$\hat{V} = E(X/V) = \mu_X + \rho_{XX'}(X - \mu_X) \tag{14}$$

b.- $\hat{V}$ es un estimador insesgado de $\mu_V$ (15)

En efecto,

$$E(\hat{V}) = E(\mu_X + \rho_{XX'}(X - \mu_X)) =$$

$$X_v = v + \text{å}$$

Por la normalidad de los errores se sigue que $X_v$ es normal,

luego $$X_v \sim N(v, \sigma^2_{\text{å}})$$

$X_v$ es, por tanto, un estimador insesgado de *v* normalmente distribuido; luego, la construcción de un intervalo de confianza para *v* sobre la base de una observación de $X_v$ se reduce a la de un intervalo de confianza de nivel $\alpha$ para una media poblacional con una muestra de tamaño 1. Esto es:

$$x_v \pm z_{1-\frac{\text{á}}{2}} \text{ó}_{X_v} \qquad (10)$$

donde $x_v$ es una observación realizada de la variable $X_v$ y $z_{1-\frac{\text{á}}{2}}$ es el percentil de la distribución normal estándar que acumula una poobabilidad de $1-\frac{\text{á}}{2}$. Despejando en (6) se sigue que:

$$\text{ó}^2_{\text{å}} = \text{ó}^2_X (1 - \text{ñ}_{XX'})$$

Luego, por (9) $$\text{ó}_{X_v} = \text{ó}_X \sqrt{1 - \text{ñ}_{XX'}}$$

Sustituyendo en (10) se obtiene el intervalo de confianza:

$$x_v \pm z_{1-\frac{\text{á}}{2}} \text{ó}_X \sqrt{1 - \text{ñ}_{XX'}} \qquad (11)$$

## *Supuestos*

1.- *Homoscedasticidad* de $\varepsilon_v$
Los errores tienen la misma variancia para cada clase de sujetos.

$$\sigma^2_{\varepsilon_v} = \sigma^2_{\varepsilon_{v'}}$$

para todo par $v$ y $v'$. (7)

2.- *Normalidad*
Los errores se distribuyen normalmente en cada clase de sujetos.
En símbolos,

$$\varepsilon_v \sim N(\mu_{\varepsilon_v}, \sigma^2_{\varepsilon_v})$$

Por (2) $\mu_{\varepsilon_v} = 0$

En el Anexo se demuestra que la variancia de los errores en cada clase de sujetos coincide con la variancia de los errores en toda la población (Proposición 1). Esto es:

$$\sigma^2_{\varepsilon_v} = \sigma^2_{\varepsilon} \qquad (8)$$

Luego, $\varepsilon_v \sim N(0, \sigma^2_{\varepsilon})$

Por tanto $\sigma^2_{X_v} = \sigma^2_{\varepsilon}$ (9)

(Ver corolario de la Proposición 1 del Anexo). De (5) resulta

$\mu_{X_v} = v$ y, por (1),

b.- La esperanza de las puntuaciones observadas y verdaderas es la misma; o sea que tienen la misma media en la población de sujetos:

$$E(X) = E(V)$$
$$\textit{Notación}: \mu_X = \mu_V \qquad (4)$$

c.- El *puntaje verdadero* de cada sujeto es el promedio de las puntuaciones empíricas de todos los sujetos de su misma clase:

$$E(X/V) = V$$
$$\textit{Notación}: \mu_{X/V} = V \qquad (5)$$

d.- El *coeficiente de confiabilidad* es el cuadrado del índice de confiabilidad que, a su vez, coincide con el cociente de las variancias de los puntajes verdaderos y observados y es complementario a la unidad con el cociente de las variancias de los errores y de los observados. En símbolos:

$$\rho_{XX'} = \rho^2_{XV} = \frac{\sigma^2_V}{\sigma^2_X} = 1 - \frac{\sigma^2_\varepsilon}{\sigma^2_X} \qquad (6)$$

donde $\rho$ denota el coeficiente de correlación lineal y $\sigma^2$ la variancia.

## Caso 1. Intervalo de confianza para un valor fijo del puntaje verdadero

Considérense el puntaje observado y el error de medición de un individuo elegido al azar de una población de individuos con puntaje verdadero $v$. Es decir, se consideran los valores que toman las variables $X$ y $\varepsilon$ cuando la población se restringe a los sujetos cuyo puntaje verdadero es $v$. Serán denotadas respectivamente $X_v$ y $\varepsilon_v$.

### Definición 4

Se denomina *índice de confiabilidad* al coeficiente de correlación lineal entre las puntuaciones empíricas y las verdaderas: $\rho_{X,V}$.

### Supuestos

1.-Los errores tienen media *cero* para cada clase de sujetos. En símbolos,

$$E(\varepsilon/V) = 0$$

donde $E(\varepsilon/V)$ es la esperanza condicional del error dado el puntaje verdadero y "0" designa a la variable aleatoria nula. Una revisión del concepto y propiedades de la esperanza condicional pueden consultarse en Meyer (1970) y se sintetiza en el Anexo.
*Notación*: En adelante $E(\varepsilon/V)$ se denotará con $\mu_{\varepsilon/V}$, por lo cual esta propiedad quedará expresada como:

$$\mu_{\varepsilon/V} = 0 \qquad (2)$$

2.-Los errores asociados a medidas paralelas no tienen correlación lineal; es decir $Cov(\varepsilon,\varepsilon') = 0$ para toda medida paralela $X'$, donde *Cov* designa la covariancia.

De esta formulación del Modelo Lineal de Puntuaciones se deducen las propiedades sobre las que se fundamenta la Teoría Clásica de Tests. A continuación se mencionarán sólo las que se necesitan para la deducción de los intervalos de confianza.

### Propiedades

a.- La esperanza matemática de los errores es cero. En otras palabras, los errores tienen media cero sobre toda la población de sujetos. En símbolos:

$$E(\varepsilon) = 0$$
$$\textit{Notación}: \mu_{\varepsilon} = 0 \qquad (3)$$

damentar correctamente su deducción y para comprender su significado y limitaciones. Tal es el objetivo de este artículo. Para lograrlo se apela a la formalización matemática ya que ella permite la formulación de supuestos, concisión, claridad y precisión (Cámara Sánchez, 2000). En este trabajo se explicitan los supuestos sobre los cuales se basan los métodos para inferir los puntajes verdaderos. Se trata de dos modelos que dan lugar a la deducción de dos tipos de intervalos, y se expondrán como dos casos diferentes.

Para deducir con rigor formal los intervalos de confianza en el modelo lineal de puntuaciones es preciso hacer referencia a sus axiomas y resultados. Para ello se seguirá el ordenamiento axiomático propuesto por Galibert, Aguerri, Lozzia y Attorresi (2003).

## Formulación del Modelo Lineal de Puntuaciones

Se considera el experimento aleatorio de elegir un individuo al azar de una población y administrarle un determinado test. Se definen las siguientes variables aleatorias:

$X$ = Puntaje observado o empírico del individuo en el test.

$V$ = Puntaje verdadero de dicho individuo.

$\varepsilon$ = Error de medición.

El Modelo Lineal de Puntuaciones formula que:

$$X = V + \varepsilon \qquad (1)$$

### Definición 1

Se dice que las puntuaciones $X$ y $X'$ son medidas paralelas si se cumple que $X = V + \varepsilon$ y $X' = V' + \varepsilon'$ con $V = V'$ y $Var(\varepsilon) = Var(\varepsilon')$.

### Definición 2

Dos sujetos son de la misma clase si tienen el mismo puntaje verdadero.

### Definición 3

Se denomina *coeficiente de confiabilidad* al coeficiente de correlación lineal entre medidas paralelas: $\rho_{X,X'}$.

propios supuestos, que es preciso diferenciar. Los autores, en general, o no explicitan claramente los supuestos o no muestran cómo se vinculan éstos con la deducción de los intervalos de confianza. Por ejemplo, Gulliksen (1950) arriba a un intervalo de confianza para el puntaje verdadero a través de una explicación coloquial donde supone los puntajes observados normalmente distribuidos para cada puntaje verdadero pero sin deducirlo matemáticamente. Lord y Novick (1968) mencionaron la necesidad de la homogeneidad de variancias para construir un intervalo de confianza basado en el error de estimación pero no explicitaron cómo interviene este supuesto que, como se verá, no es necesario para ese tipo de intervalos. Las mayores dificultades se hallan en el uso de la ecuación de regresión de los puntajes verdaderos sobre los observados. Suelen aplicarse los supuestos propios del modelo de regresión con variable independiente de valores fijos cuando, en el modelo lineal clásico, ambos puntajes varían conjuntamente. Los autores se refieren sólo tangencialmente a los supuestos de la regresión lineal, por ejemplo a la homoscedasticidad, como si el cálculo de un intervalo de confianza para el puntaje verdadero fuera una mera aplicación particular de la Teoría de Regresión de la Estadística General (Myers, 1990). Sin embargo el modelo de regresión en este contexto es algo diferente. En el modelo de regresión general todas las variables son observables y suelen aplicarse para predecir el valor de una variable a través de otra u otras de diferente naturaleza.

Se distinguen dos modelos de regresión según el carácter fijo o aleatorio de la variable considerada como independiente donde la homoscedasticidad es un supuesto necesario sólo para el primer modelo. Si la variable independiente es aleatoria el modelo atañe a la distribución de probabilidades de un vector aleatorio bidimensional; por lo que no habría, en rigor, una variable dependiente y otra independiente. La regresión de los puntajes verdaderos sobre los observados se inscribe en este caso puesto que son dos variables que varían conjuntamente. Surgen, por tanto, dos diferencias con respecto al modelo de regresión de la teoría estadística general que dan cierta peculiaridad al uso de la regresión en el contexto de la teoría clásica de tests, a saber:

- Una de las dos variables, el puntaje verdadero, es inobservable.
- Para cada sujeto, el puntaje verdadero y el observado son variables de la misma naturaleza; en realidad se trata de la misma característica expresada en términos de sus valores verdaderos y éstos estimados por los observados.

Todas estas consideraciones deben tenerse en cuenta al aplicar el modelo de regresión en el cálculo de los intervalos de confianza para fun-

of this article. To accomplish this, it is necessary to make explicit the assumptions on which the methods to infer the true scores are based. It deals with two models, which allow the deduction of two types of intervals and they will be treated as two different cases.

*Key words:* Classical test theory – confidence intervals – true scores – assumptions.

## Introducción

La medición de variables psicológicas ha sido no sólo un tema más de los que aborda la psicología sino que ha estado presente en el origen mismo de la psicología como ciencia en áreas como la psicofísica y la psicometría. Muñiz (1998) analiza la problemática implicada en la medición de variables psicológicas y comenta las soluciones aportadas por diferentes enfoques psicométricos. Entre ellos, presenta el modelo lineal clásico propuesto por Spearman (1904, 1907, 1913) para hacer posible la estimación de los errores de medición. Este expresa la puntuación empírica $X$ de un sujeto en función de dos componentes: su puntuación verdadera $V$ más un error de medición $\varepsilon$; esto es $X = V + \varepsilon$. La teoría clásica de tests se desarrolla a partir de esta formulación del modelo al cual se añaden ciertos supuestos que permiten la deducción de sus propiedades fundamentales (Allen & Yen, 1979; Crocker & Algina, 1986; Lord & Novick, 1968). Attorresi, Galibert y Aguerri (2002) presentaron una formulación abstracta de la axiomatización del modelo que utiliza el concepto de esperanza condicional (Meyer, 1970) y que permite salvar algunos inconvenientes formales de la presentación clásica.

Dado que un objetivo de la medición es estimar los puntajes verdaderos de los sujetos, resultan de interés las distintas técnicas de inferencia de dichos puntajes a partir de los observados. Según cúal sea la técnica elegida se necesita agregar supuestos de distribución sobre las componentes del modelo lineal clásico. En la bibliografía suelen presentarse dos tipos de intervalos: uno basado en el *error de medición* y otro basado en el *error de estimación* cometido cuando los puntajes verdaderos se predicen con la ecuación de regresión sobre los observados. La construcción de ambos tipos de intervalos descansa en dos modelos, cada uno con sus the confidence interval construction. For example, Lord and Novick (1968) mentioned the need of variance homogeneity, without explaining how that assumption is used in the process. The assumption of variance homogeneity when the confidence intervals are built based on the error of estimation is not needed as it is shown later on. The greatest difficulties arise when the regression equation of true scores on observed scores is used. In that case, the assumptions for the usual regression model with fixed independent variables are used, while in the linear classical model, both scores vary jointly. The authors refer to the linear regression assumptions only in a slightly way, for example, about the homocedasticity, as if they took the true score confidence interval computation as a simple particular application of the Regression Theory on General Statistics. Nevertheless, in this context the regression model is something different. In the general regression model, all the variables are observable and it is applied to predict the value of one variable through the value of other or others of different nature. Two regression models can be distinguished depending whether the independent variable is considered as fixed or as random; the homocedasticity is a necessary assumption only for the first case. When the independent variable is random, the model is linked to the probability distribution of a bidimensional random vector; and therefore, there would not exist a dependent and another independent variable. The regression of the true scores on the observed ones is included in this case, since both variables vary jointly. Consequently, two different features regarding the regression model in the general statistic theory arise, and they confer certain peculiarity to the use of regression in the context of the classical test theory. The two differences are:

- One of the two variables, the true score, is not observable.
- For each individual, the true score and the observed score are variables of the same nature; actually, both are the same characteristic expressed in terms of their true values, and these are estimated by the observed ones.

All these considerations must be taken into account when calculating the confidence intervals by application of the regression model, in order to base their deduction and to understand its meaning and limitations. Such is the objective

(Caso 2). La construcción de ambos tipos de intervalos descansa en dos modelos, cada uno con sus propios supuestos, que es preciso diferenciar. El Caso 1 corresponde a la estimación de un valor fijo del puntaje verdadero y el Caso 2 a la de un valor aleatorio del mismo. Se señalan diferencias entre el modelo de regresión de la teoría estadística general y el uso de la regresión en el contexto de la teoría clásica. Se presentan las consideraciones que, por tanto, deben tenerse en cuenta al aplicarlo en el cálculo de los intervalos de confianza para fundamentar correctamente su deducción y para comprender su significado y limitaciones. Se concluye que el supuesto de homoscedasticidad es sólo necesario para el Caso 1 y que la construcción de ambos intervalos supone que la población se restringe al grupo normativo pues no considera el error de muestreo de los estimadores de la confiabilidad del test y de la variancia de los puntajes observados.

*Palabras clave:* Teoría clásica de tests – intervalos de confianza – puntajes verdaderos – supuestos.

## Abstract

The *classical test theory* was developed by starting with the linear model to which certain *assumptions* are added to allow the deduction of its fundamental properties. Since one of the measurement objectives is to estimate the individuals' *true scores*, the different inferential techniques for the true scores based on the observed ones, result of interest. Depending on the technique chosen, it will be necessary to add assumptions about the distribution of the classical linear model components. In the psychometric literature, two types of intervals are usually seen: one type is based on the measurement error ε; and the other, on the error of estimation obtained from the prediction of true scores are predicted from a regression equation on the observed ones. Both types of intervals are set down on two models, each one with its own assumptions, and the models should be differentiated. In general, the authors do not state the assumptions clearly or, at times, they do not show how the assumptions are involved in

# INTERVALOS DE CONFIANZA PARA LAS PUNTUACIONES VERDADERAS. EXPLICITACION DE SUS SUPUESTOS *

Horacio Félix **Attorresi** **, María Ester **Aguerri** ***,
Gabriela Susana **Lozzia** **** y María Silvia **Galibert** *****

## Resumen

En este artículo se explicitan los *supuestos* sobre los que se basan los métodos para inferir los puntajes verdaderos de la teoría clásica de tests. El objetivo del presente trabajo fue fundamentar la deducción de los dos tipos de *intervalos de confianza* que se mencionan corrientemente en la bibliografía precisando los modelos en los cuales se deducen. Uno de ellos se basa en el error de medición (Caso 1) y el otro en el error de estimación cometido cuando los *puntajes verdaderos* se predicen con la ecuación de regresión sobre los observados

---

* Esta investigación fue realizada en el marco de los siguientes proyectos: P054 y P605 de la Universidad de Buenos Aires (UBACyT), PIP Nº 2426 del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y PICT 4704 de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT).

** Licenciado en Ciencias Matemáticas. Profesor Regular Titular de Estadística y Director de Proyectos de Investigación en el Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires (UBA). Rivera Indarte 132-1er Piso Dpto. A(1406) Buenos Aires, Rep. Argentina. E-Mail: hattorre@psi.uba.ar

*** Magister Scientiae en Biometría y Licenciada en Ciencias Matemáticas. Profesora Regular Adjunta de Estadística e Investigadora en el Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires (UBA).
E-Mail: maguerri@psi.uba.ar

**** Licenciada en Psicología. Ayudante de Primera en las cátedras I y II de Estadística y Becaria del Proyecto de Investigación PICT 4704, de la Agencia Nacional para la Promoción Científica y Tecnológica (ANPCyT) en el Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires (UBA).
E-Mail: glozzia@psi.uba.ar

***** Magister Scientiae en Biometría y Profesora de Enseñanza Especial en Ciencias Matemáticas. Profesora Regular Adjunta de Estadística e Investigadora en el Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires (UBA). E-Mail: galibert@psi.uba.ar