

Transformación Z de Fisher para la determinación de intervalos de confianza del coeficiente de correlación de Pearson

Alfonso Sánchez-Bruno (†) y África Borges del Rosal
Universidad de La Laguna

En el presente trabajo se pone a prueba la bondad de la transformación Z de Fisher para la estimación de intervalos de confianza para el coeficiente de correlación de Pearson con distintos valores de éste y tamaños muestrales y en condiciones de normalidad y no normalidad de los residuos. Se concluye que, asegurando la normalidad de los residuos y dado un número suficiente de sujetos, la determinación de los intervalos de confianza mediante dicha transformación resulta adecuada.

Fisher's transformation to determine confidence intervals for Pearson's product-moment correlation coefficient. In this study we test the accuracy of Fisher's Z transformation for estimating confidence intervals for Pearson's product-moment correlation coefficient with several correlation values and sample sizes and under several conditions of normality and non normality of residuals. In conclusion, under normality of residuals and given a sufficiently large number of subjects, the determination of confidence intervals by means of such transformation is adequate.

Las dificultades provocadas por la inadecuada interpretación del contraste de hipótesis concluyeron, finalizando el siglo, con las recomendaciones en torno a la investigación en general y al análisis de datos en particular, recogidos en el trabajo de Wilkinson y la Task Force (1999), donde se presentan directrices claras sobre la exposición de resultados que deben incluir los investigadores en sus informes. Entre ellos, la inclusión de tamaños de efecto parece la más enriquecedora, por diversas razones. Informar del tamaño de efecto permite entender la importancia teórica o clínica de los resultados (terminando también con la eterna dicotomía entre la significación estadística o clínica), a la vez que facilita la comparación entre los resultados de distintos estudios, acumulando de esta forma el conocimiento (Schmidt y Hunter, 2002). Esta práctica ya se va encontrando en la literatura más reciente (Cebollero y Guardia, 2002; Díaz y Sánchez-López, 2002; Martínez, Cifre, Llorens y Salanova, 2002; Prieto y Sorbal, 2003). Por todo esto, además, la medida del tamaño del efecto y su inclusión en los informes científicos se ha constituido en un requisito obligatorio para publicar en un número creciente de revistas de Psicología.

Por otra parte, la A.P.A., en el mismo artículo de Wilkinson y la Task Force (1999) ya citado, recomienda el uso de los intervalos de confianza para las medidas del tamaño de efecto, ya que tales intervalos resultan mucho más informativos.

No obstante, ni el tamaño del efecto ni sus intervalos de confianza constituyen todavía un dato del que se informe con asiduidad (Borges y otros, 2001), debido fundamentalmente a lo que podríamos definir como «*lo que no está en el paquete SPSS no*

existe». Por ello, son relativamente abundantes las investigaciones en las que se informa del valor del coeficiente de correlación de Pearson, pero muy infrecuentemente se acompaña este valor de un intervalo de confianza. Sin embargo, aunque dicho paquete no aporta intervalos de confianza para la correlación producto-momento, no obstante sí que facilita el cálculo del error típico, para ambas distribuciones, centrada y no centrada.

A la hora de determinar un intervalo de confianza para el coeficiente de correlación de Pearson, nos enfrentamos al problema de la no-centralidad de la distribución, lo que hace difícil la estimación directa de los límites. No obstante, Fisher (1915, 1921) propuso una transformación del coeficiente de correlación, hoy en día llamada Z de Fisher, cuya distribución se asume normal con desviación típica $1/\sqrt{n-3}$:

$$Z_f = \text{arco tanh}(r_{xy}) = 0.5 \ln \frac{1+r_{xy}}{1-r_{xy}}$$

Recientemente, Zimmerman y Zumbo (2003) mostraron la ausencia de sesgo de este estimador, si bien se muestra sensible a la violación de la normalidad bivariada.

En un estudio previo (Sánchez Bruno y Borges, en prensa) llegamos a la conclusión de que si la distribución de los residuos es normal, la distribución de la Z de Fisher tiende, efectivamente, a la normalidad al incrementarse el tamaño muestral, si bien para valores de n iguales o menores a 30 el apuntamiento es ligeramente superior a cero.

Concluimos también que en ausencia de normalidad el efecto del apuntamiento de los errores incide sobre la desviación típica real de la distribución muestral de la Z de Fisher, pero no sobre su asimetría y apuntamiento. Este efecto vimos que se produce también bajo la condición de asimetría, pero como no es posible, al menos con el procedimiento de generación de distribuciones no

Fecha recepción: 11-4-03 • Fecha aceptación: 1-3-04

Correspondencia: África Borges del Rosal

Facultad de Psicología

Universidad de La Laguna

38271 Islas Canarias (Spain)

E-mail: aborges@ull.es

normales que usamos (Fleishman, 1978), generar distribuciones asimétricas «puras», sino que éstas necesariamente deben tener algún grado de apuntamiento, es posible que este efecto se deba al apuntamiento y no a la asimetría de la distribución de los residuos.

En caso de que la violación del supuesto de normalidad de los residuos se deba tanto a la asimetría como al apuntamiento, la conclusión a la que llegamos es que la distribución muestral real de la Z de Fisher se aleja bastante de la teórica, de forma que tanto su asimetría como su apuntamiento se incrementan con el valor del coeficiente aunque disminuyen con el tamaño de la muestra. Además, la distribución muestral de Z se «retrasa» con respecto a sus valores teóricos, ya que la distribución se desplaza hacia la derecha a medida que las correlaciones se hacen positivas y más altas (Sánchez-Bruno y Borges, en prensa), con lo que la proporción de aciertos debe ser inferior a la prevista.

En el trabajo que aquí presentamos, en la línea del anterior, hemos puesto a prueba la bondad de la transformación Z de Fisher para la estimación de intervalos de confianza para el coeficiente de correlación de Pearson con distintos valores de éste y tamaños muestrales y en condiciones de normalidad y no normalidad de los residuos.

Método

Variables

Como variables independientes se consideraron las siguientes: valor del coeficiente de correlación, desde 0 hasta 0,9 con incrementos de 0,1; tamaño de la muestra, desde 10 hasta 50 con incrementos de 10; normalidad-no normalidad, con los siguientes valores de los coeficientes de asimetría y apuntamiento para las seis condiciones de no normalidad:

a ₃	a ₄
0	-1,15
0	5,2
-0,85	0
0,85	0
-2	5,2
2	5,2

Como variables dependientes se calcularon la proporción de «aciertos» al estimar el coeficiente de correlación poblacional a partir del muestral, entendiéndose como «acierto» que el valor del coeficiente de correlación poblacional esté entre los límites del intervalo de confianza, y la proporción de «errores» (valor del coeficiente de correlación poblacional fuera de los límites del intervalo de confianza) por encima y por debajo de los límites del intervalo de confianza.

Procedimiento

La investigación se llevó a cabo mediante simulación de Monte Carlo. Los programas fueron escritos en Gauss para Windows versión 3.2.38. Los números pseudo-aleatorios se generaron utilizando las funciones propias del Gauss, mediante el método multiplicativo congruencial de 32 bit (Kennedy y Gentle, 1980).

Para generar las variables (x e y) bajo las condiciones de normalidad se generaron en primer lugar dos vectores x y e distribuidos normalmente N(0,1). El vector y se generó mediante la aplicación de:

$$y = x + e\sqrt{1 - r_{xy}^2}$$

siendo r_{xy} el valor del coeficiente de correlación deseado.

Para las condiciones de no normalidad se transformó el vector e tras su generación, mediante el algoritmo de Fleishman (1978). Este algoritmo permite obtener un valor mínimo de -1,15 para a₄; teniendo en cuenta que la distribución uniforme tiene un coeficiente de apuntamiento de -1,20, podemos decir que la que generamos es extremadamente platocúrtica. En cuanto a la asimetría, este algoritmo no permite generar valores tan altos de la misma manteniendo nulo el apuntamiento, por lo que nos vemos obligados a seleccionar la combinación descrita, que por lo menos es comparable al efecto producido por el apuntamiento por sí solo.

Para llevar a cabo la simulación se generaron y analizaron 10.000 muestras con las características (correlación, tamaño muestral y normalidad o ausencia de ella) de cada una de las diferentes condiciones.

Resultados

Presentamos los resultados diferenciados en función de la normalidad o no-normalidad de los residuos.

Normalidad

En la figura 1A se puede observar gráficamente la proporción de «aciertos», entendiéndose como tales que el valor del coeficiente de correlación poblacional esté entre los límites del intervalo de confianza, para un nivel de confianza de 0,95. Como puede verse, la proporción de aciertos es prácticamente 0,95 para cualquier valor del coeficiente de correlación y cualquier tamaño muestral igual o superior a 10 sujetos.

En la figura 1B vemos la representación gráfica de los porcentajes de veces que el verdadero valor del parámetro estaba por encima del nivel superior del intervalo de confianza y por debajo del nivel inferior del mismo, sobre el total de errores. Puesto que pretendemos que el intervalo de confianza constituya el 95% central de la distribución, el porcentaje de errores por encima y por debajo debería ser aproximadamente el mismo, es decir, en la representación gráfica ambas líneas deberían superponerse en valores cercanos al 50%. En realidad lo que observamos es que el porcentaje de errores por la izquierda se incrementa con el valor del coeficiente, aunque este efecto tiende a disminuir con el tamaño muestral. Es decir, la transformación Z de Fisher tiende a sobreestimar los límites del intervalo.

Apuntamiento

En las figuras 2A y 2B hemos representado gráficamente la proporción de aciertos para las condiciones de violación del supuesto de normalidad de los residuos por apuntamiento. Se puede observar cómo la proporción de aciertos disminuye radicalmente cuando la distribución es leptocúrtica, siendo esta disminución mayor en la medida en que lo sean el valor del coeficiente y el tamaño muestral. En el peor de los casos estudiados (ρ_{xy}= 0,9 y n= 50) la proporción de aciertos llega a valores cercanos al 83%.

Por el contrario, cuando la distribución es platocúrtica la proporción de aciertos se incrementa, siendo este incremento mayor en la medida en que lo sea el valor del coeficiente y el tamaño de la muestra, si bien este último efecto es de menor importancia.

Por otra parte, también en este caso (figuras 3A y 3B) se da el efecto que ya hemos comentado de que la transformación Z de Fisher sobreestima los límites del intervalo de confianza, si bien ahora este efecto es algo más importante (máximo del 80% frente al 70% en el caso de normalidad de los residuos) cuando la distribución de los residuos es leptocúrtica y bastante menos importante (hasta el punto de ser prácticamente inexistente para tamaños muestrales iguales o superiores a 20), si la distribución de los residuos es platicúrtica.

Asimetría

En las figuras 4A y 4B reproducimos el porcentaje de aciertos obtenido al violar el supuesto de normalidad de los residuos generando la distribución de éstos con asimetría no nula. Como puede observarse, estos resultados son cercanos al 95% para un nivel de confianza de 0,95, tal como sucedía con la condición de normalidad.

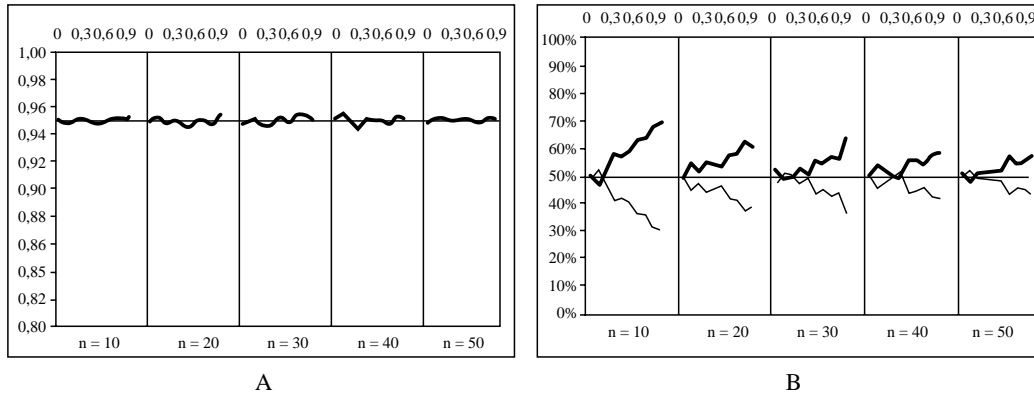


Figura 1. Proporción de aciertos en la estimación mediante intervalo de confianza (n.c. del 95%) del valor poblacional del coeficiente de correlación de Pearson (A) y porcentaje (sobre el total de errores cometidos en la estimación) de veces que el verdadero valor del parámetro estaba por encima del nivel superior del intervalo de confianza (línea fina) y por debajo del nivel inferior del mismo (línea gruesa) (B), con la transformación Z de Fisher para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de normalidad de los residuos

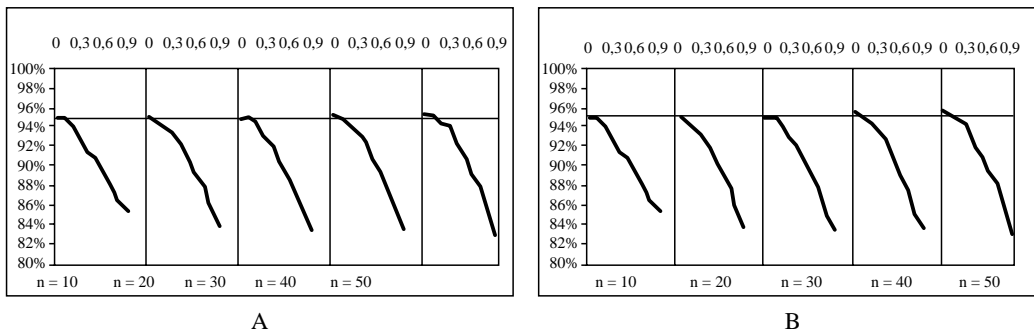


Figura 2. Proporción de aciertos en la estimación mediante intervalo de confianza (n.c. del 95%) del valor poblacional del coeficiente de correlación de Pearson con la transformación Z de Fisher para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de no normalidad de los residuos (A: $a_3=0$ y $a_4=5,2$; B: $a_3=0$ y $a_4=-1,15$)

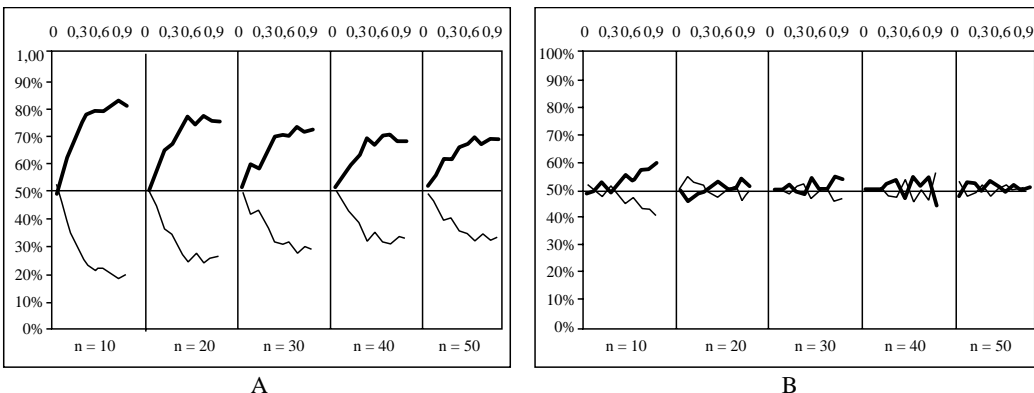


Figura 3. Porcentaje (sobre el total de errores cometidos en la estimación) de veces que el verdadero valor del parámetro estaba por encima del nivel superior del intervalo de confianza (línea fina) y por debajo del nivel inferior del mismo (línea gruesa) para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de no normalidad de los residuos (A: $a_3=0$ y $a_4=5,2$; B: $a_3=0$ y $a_4=-1,15$)

En cuanto a la proporción de errores por encima y por debajo de los límites del intervalo de confianza, en las figuras 5A y 5B se aprecia una vez más que la transformación Z de Fisher tiende a sobreestimar los límites del intervalo, si bien esta sobreestimación no es superior en este caso a la que se da bajo la condición de normalidad.

Anormalidad extrema

En las figuras 6A y 6B hemos representado gráficamente la proporción de aciertos para las dos condiciones de asimetría y apuntamiento combinados. En las figuras 7A y 7B hemos representado la proporción de errores por encima y por debajo de los límites del intervalo de confianza.

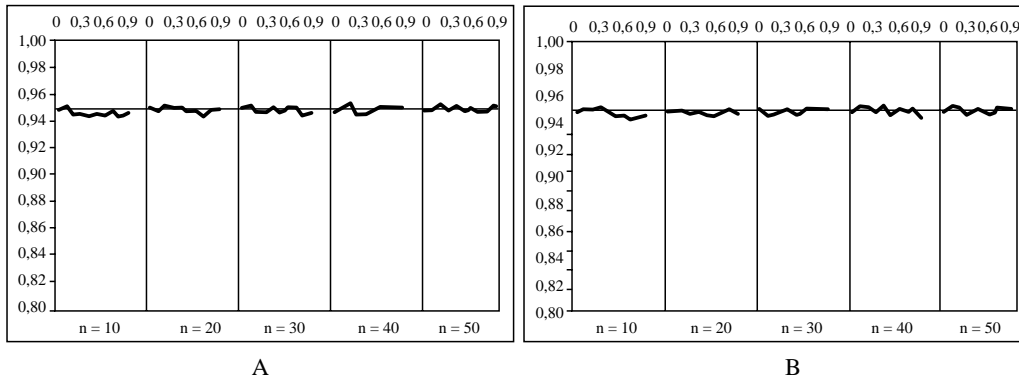


Figura 4. Proporción de aciertos en la estimación mediante intervalo de confianza (n.c. del 95%) del valor poblacional del coeficiente de correlación de Pearson con la transformación Z de Fisher para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de no normalidad de los residuos (A: $a_3 = -0,85$ y $a_4 = 0$; B: $a_3 = 0,85$ y $a_4 = 0$)

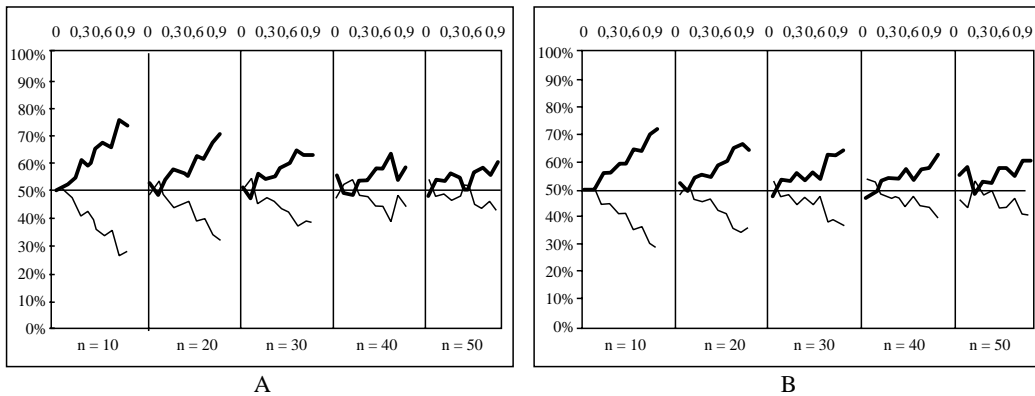


Figura 5. Porcentaje (sobre el total de errores cometidos en la estimación) de veces que el verdadero valor del parámetro estaba por encima del nivel superior del intervalo de confianza (línea fina) y por debajo del nivel inferior del mismo (línea gruesa) para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de no normalidad de los residuos (A: $a_3 = -0,85$ y $a_4 = 0$; B: $a_3 = 0,85$ y $a_4 = 0$)

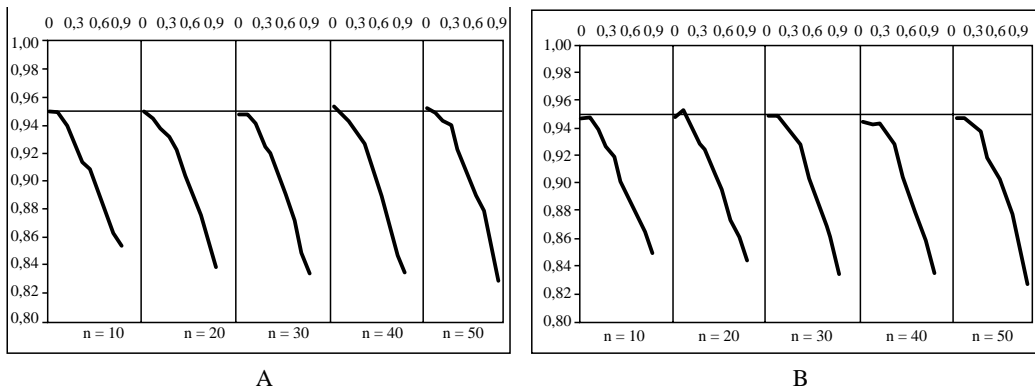


Figura 6. Proporción de aciertos en la estimación mediante intervalo de confianza (n.c. del 95%) del valor poblacional del coeficiente de correlación de Pearson con la transformación Z de Fisher para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de no normalidad de los residuos (A: $a_3 = -2$ y $a_4 = 5,2$; B: $a_3 = 2$ y $a_4 = 5,2$)

Los resultados son ligeramente peores a los obtenidos cuando la distribución es simétrica y leptocúrtica (figura 2A), lo que se puede observar mejor en la figura 8, donde hemos representado los resultados obtenidos con $r_{xy}=0,9$ y $n=50$ en cada una de las condiciones de normalidad-ausencia de normalidad estudiadas.

Estos resultados, junto con los obtenidos al violar el supuesto de normalidad de los residuos sólo por asimetría no nula (figuras 4 y 5), nos hace pensar que la transformación Z de Fisher es resistente a pequeñas violaciones del supuesto de normalidad por asimetría, pero se resiente cuando ésta es alta y cuando su apuntamiento es positivo.

Discusión y conclusiones

Dada la no centralidad de la distribución muestral del coeficiente de correlación de Pearson, parece una buena idea utilizar la transformación Z de Fisher para determinar los intervalos de confianza de r_{xy} .

La proporción de aciertos obtenida mediante esta transformación, entendiendo como tales aciertos que el valor del coeficiente de correlación poblacional esté entre los límites del intervalo de confianza es, siendo la distribución de los residuos normal y para un nivel de confianza de 0,95, prácticamente igual al nivel de confianza para cualquier valor del coeficiente de correlación y cualquier ta-

maño muestral igual o superior a 10 sujetos. No obstante, existe un cierto desequilibrio entre la proporción de errores por encima y por debajo del intervalo de confianza, con un claro desequilibrio hacia estos últimos, lo que indica que en general este procedimiento sobreestima los límites del intervalo; este problema se incrementa con el valor del coeficiente y disminuye con el tamaño muestral.

En ausencia de normalidad, se obtienen resultados diferentes según la violación del supuesto de normalidad, sea por un valor no nulo del coeficiente de apuntamiento o del de asimetría.

Las violaciones del supuesto de normalidad de los residuos por apuntamiento positivo (distribución leptocúrtica) provocan una gran disminución en la proporción de aciertos, de forma que ésta cae a niveles de alrededor del 83%. En lo que se refiere al desequilibrio entre la proporción de errores por encima y por debajo de los límites del intervalo, el problema permanece e incluso se agrava ligeramente.

Por el contrario, el efecto del apuntamiento negativo (distribución platicúrtica) es justamente el opuesto: incrementa la proporción de aciertos por encima del nivel de confianza nominal y corrige el desequilibrio entre la proporción de errores por encima y por debajo de los límites del intervalo, al menos para tamaños muestrales iguales o superiores a 20. En cuanto a la asimetría de la distribución de los residuos, sólo parece afectar a los resultados a partir de valores altos de a_3 , y no parece tener efectos importan-

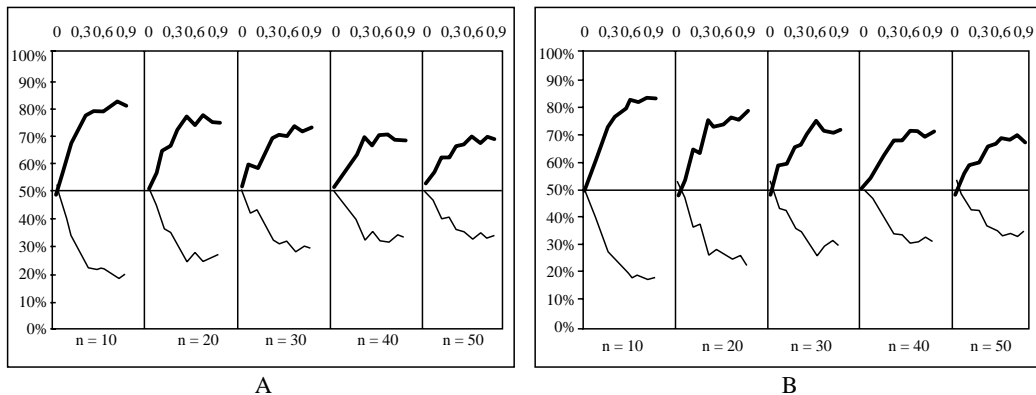


Figura 7. Porcentaje (sobre el total de errores cometidos en la estimación) de veces que el verdadero valor del parámetro estaba por encima del nivel superior del intervalo de confianza (línea fina) y por debajo del nivel inferior del mismo (línea gruesa) para distintos tamaños muestrales y valores de ρ_{xy} (eje de abscisas) en condiciones de no normalidad de los residuos (A: $a_3 = -2$ y $a_4 = 5,2$; B: $a_3 = 2$ y $a_4 = 5,2$)

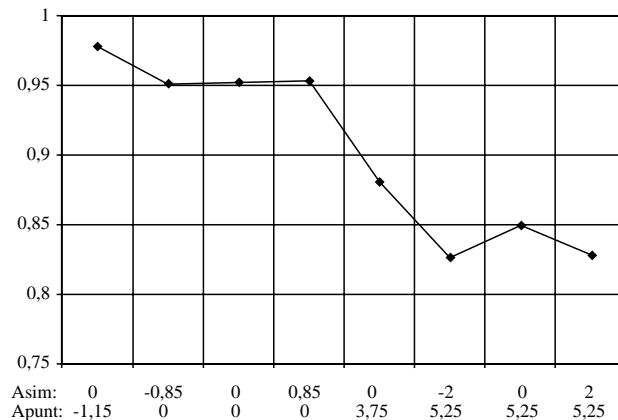


Figura 8. Proporción de aciertos en la estimación mediante intervalo de confianza (n.c. del 95%) del valor poblacional del coeficiente de correlación de Pearson con la transformación Z de Fisher para distintas combinaciones de apuntamiento ($\rho_{xy} = 0,90$ y $n = 50$)

tes sobre el ya citado desequilibrio entre la proporción de errores por encima y por debajo de los límites del intervalo.

Por tanto, nuestros resultados nos llevan a concluir que, asegurando la normalidad de los residuos y dado un número suficiente de sujetos, el aproximar los intervalos de confianza del coeficiente de correlación de Pearson mediante la transformación Z de Fisher puede ser una solución fácil y sencilla para los investigadores, permitiendo además que se incluya una información que contribuirá a un mayor conocimiento de las variables que se manejan en Psicología y sus relaciones.

Sin embargo, cabe recordar que el coeficiente de correlación producto-momento mide independencia estadística sólo en condiciones de normalidad, y que la transformación Z se comporta razonablemente bien para inferir ρ_{xy} (Fisher, 1921) cuando las variables X e Y sigan una distribución normal divariada. Por tanto, el investigador tiene que asegurarse que se respetan las condiciones de normalidad. En caso contrario, tanto porque dicha transformación no está diseñada en otras condiciones, como porque los resultados de la simulación lo demuestran, no debe emplearse.

Referencias

- Borges del Rosal, A., San Luis Costas, C., Sánchez Bruno, A. y Cañadas Ossinski, I. (2001). El juicio contra la hipótesis nula. Muchos testigos y una sentencia virtuosa. *Psicothema* 13(1), 173-178.
- Cebollero, M. y Guardia, J. (2002). Indicadores epidemiológicos descriptivos para el estudio de trastornos que cursan con recaídas. *Psicothema*, 14(1), 63-70.
- Díaz, J.F. y Sánchez-López, P. (2002). Relaciones de estilos de personalidad y satisfacción autopercebida en diferentes áreas vitales. *Psicothema*, 14(1), 100-105.
- Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507-521.
- Fisher, R.A. (1921). On the «probable error» of a coefficient of correlation deduced from small sample. *Metron*, 1, 3-32.
- Fleishman, A.I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532.
- Kennedy, W. y Gentle, J.E. (1980). *Statistical Computing*. Marcel Dekker.
- Martínez, I.M., Cifre, E., Llorens, S. y Salanova, M. (2002). Efectos de la tecnología asistida por ordenador en el bienestar psicológico afectivo. *Psicothema*, 14(1), 118-123.
- Prieto, A. y Sorbal, J. (2003). Impacto persuasivo del testimonio seguro e inseguro: ¿dos caras de un mismo fenómeno? *Psicothema*, 15(2), 167-171.
- Sánchez-Bruno, A. y Borges del Rosal, A. (en prensa). Estudio de la distribución muestral de la transformación Z de Fisher. *Metodología de las Ciencias del Comportamiento*.
- Schmidt, F. y Hunter, J. (2002). Are there benefits from NHST? *American Psychologist*, 57 (1), 65-71.
- Wilkinson, L. y Task Force on Statistical Inference, APA Board of Scientific Affairs (1999). Statistical Methods in Psychology journals: Guidelines and explanation. *American Psychologist*, 54, 594-604.
- Zimmerman, D.W., Zumbo, B.D. y Williams, R.H. (2003). Bias estimation and hypothesis testing of correlation. *Psicológica*, 24, 133-158.