

ANÁLISIS COMPARATIVO DE PRUEBAS DE NORMALIDAD CON MUESTRAS PEQUEÑAS Y DISTRIBUCIONES ALTERNATIVAS SIMÉTRICAS

CABRERA GABRIELA PILAR¹, ZANAZZI JOSÉ LUIS¹, LAURA ALICIA GONZÁLEZ²,
JOSÉ FRANCISCO ZANAZZI¹

¹*Facultad de Ciencias Exactas Físicas y Naturales. Universidad Nacional de Córdoba.*

²*Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba.*

gabriela.pilar.cabrera@gmail.com

RESUMEN

El estudio de pruebas más potentes de ajuste al Modelo Normal Univariante, en los últimos años cobra importancia, tanto en la investigación teórica como empírica. Evidencia de esto, es el desarrollo de un gran número de estas pruebas; y la atención puesta en proporcionar al analista ayuda en la elección de las pruebas más adecuadas para las necesidades particulares. Numerosos artículos se orientan a evaluar la potencia de estas pruebas. Lamentablemente, la mayoría de las verificaciones existentes operan con muestras de cincuenta o más datos. En cambio, en este trabajo se estima la potencia de diversos tests, con muestras de diez y quince datos. Además se comprueba la potencia en situaciones donde la distribución original es simétrica, lo cual es sin duda la peor condición para la prueba. Para estas determinaciones, se realizan experimentos de simulación. Finalmente se concluye con una valoración cualitativa sobre la conveniencia de las pruebas analizadas, recomendándose la implementación de la prueba Shapiro-Wilk modificada por Rahman y Govindarajulu (1997). Esta prueba de normalidad evidencia una mayor potencia empírica en términos generales, sin embargo con muestras pequeñas la potencia de ese test es reducida, por lo que se considera imprescindible complementar el análisis con otras herramientas. Cabe señalar que, el análisis de las pruebas de normalidad realizado coincide con otros estudios, dado que las potencias para diez y quince datos, parecen compatibles con las reportadas por otros autores, para tamaños de muestra mayores.

Palabras clave: *Pruebas de ajuste al modelo Normal Univariante – Muestras pequeñas – Distribuciones alternativas simétricas.*

Introducción

La verificación de que los datos observados sobre un fenómeno aleatorio, pueden suponerse extraídos de una Distribución de Probabilidad Normal, es necesaria en una variedad de situaciones prácticas para las que frecuentemente se cuenta con pocas observaciones. Si bien en la literatura estadística actual se cuenta con alrededor de 40 pruebas de ajuste al Modelo Normal Univariante (Dufour *et al.* 1998); no existe una única prueba de normalidad que merezca ser “la recomendable” (Tanveer-UI-Islam, 2011). En otras palabras, debido a la gran variedad de alternativas a la normalidad, no existe “una prueba” más potente en términos generales.

En el presente documento, se realizó una revisión bibliográfica sobre pruebas de ajuste al Modelo Normal Univariante o Univariado y se propuso la selección de un conjunto de éstas, con base en los siguientes criterios: potencia para muestras de diez y quince datos en situaciones donde la distribución original es simétrica y disponibilidad de estas pruebas en los programas de computadora que se

utilizan frecuentemente en nuestro país. Se concluyó con una valoración cualitativa sobre la conveniencia de las pruebas analizadas.

Desarrollo

Cabe recordar que el inicio del estudio para detectar desviaciones de la normalidad tuvo sus orígenes en las investigaciones de Pearson (1895), en relación a los coeficientes de asimetría y curtosis (Althouse *et al.* 1998). A partir de esto, comenzaron a surgir varias pruebas de normalidad, llamadas también pruebas de bondad de ajuste. Estas pruebas son procedimientos de inferencia estadística, diseñados para evaluar si la distribución de probabilidad de una variable aleatoria univariante, es la distribución de probabilidades Normal.

En este texto, la presentación de estas pruebas se estructuró en cuatro grupos, con base en la clasificación propuesta por Seier (2002). Las pruebas basadas en los C-momentos y L-momentos de la muestra; las pruebas que se basan en la comparación de la función de distribución empírica con la función de distribución acumulada normal -a partir de diferentes modos de calcular estas distancias-; las pruebas de correlación y regresión y un conjunto de pruebas que no se sitúan en la clasificación anterior.

Ahora bien, para el cálculo de la potencia de estas pruebas, se consideró como hipótesis nula el supuesto de normalidad y como alternativa, la posibilidad de que la distribución verdadera sea Uniforme. Ésta es una de las peores condiciones posibles para los *tests* de normalidad, debido a que la distribución es simétrica. Mediante experimentos de generación aleatoria, se determinó la potencia de cada una de ellas, para la detección de la distribución Uniforme para los tamaños de muestra de diez y quince datos.

El estudio de la disponibilidad de estas pruebas, se enfocó en programas de computadora que se utilizan frecuentemente en nuestro país por usuarios no especializados en estadística. A los fines de este trabajo, se analizaron los *software* estadísticos: Infostat 2014, SPSS 19, Stata 11 y Minitab 17.

Resultados y Conclusiones

Como producto de la investigación de bibliografía, se concretó una primera selección de unas veinte pruebas, divididas en cuatro grupos según la estrategia utilizada. Se enumeran estos tests en la tabla siguiente:

Tabla 1. Pruebas seleccionadas

Estrategia	Nro.	Nombre
Pruebas basadas en Momentos	1	<i>D'Agostino-Pearson (1973)</i>
	2	<i>Jarque-Bera (1980)</i>
	3	<i>Prueba robusta de Jarque-Bera (Gel y Gastwirth, 2008)</i>
	4	<i>Bonett-Seier (2002)</i>
	5	<i>Hosking (1990)</i>
Pruebas basadas en la función de distribución empírica	6	<i>Kolmogorov-Smirnov (1933)</i>
	7	<i>Kolmogorov-Smirnov modificado por Lilliefors (1967)</i>
	8	<i>Kolmogorov-Smirnov modificado por Stephens y Harley (1972)</i>
	9	<i>Anderson y Darling (1954)</i>
	10	<i>Zhang y Wu (2005)</i>
	11	<i>Glen, Leemis y Barr (2001)</i>
Pruebas de correlación y regresión	12	<i>Shapiro-Wilk (1965)</i>
	13	<i>Shapiro-Francia (1972)</i>
	14	<i>Chen-Shapiro(1995)</i>
	15	<i>Modificación del Shapiro-Wilk sugerida por Rahman y Govindarajulu (1997)</i>
	16	<i>Modificación de Shapiro-Wilk propuesta por D'Agostino(1971)</i>
	17	<i>Filliben (1975)</i>
Otras pruebas	18	<i>Prueba de correlación de cuantiles de Del Barrio et al. (1999)</i>
	19	<i>Prueba de Coin (2007)</i>
	20	<i>Gel, Miao y Gastwirth (2007)</i>

Cabe destacar además, que los estudios más tempranos (Shapiro y Wilk, 1965; Shapiro *et al.* 1968; Pearson *et al.* 1977; Gan y Koehler, 1990 y D'Agostino y Stephens, 1986) sugieren, que la mayoría de los procedimientos analizados funcionan bien cuando las distribuciones alternativas de no normalidad resultan fuertemente sesgadas. En tanto, ante distribuciones alternativas no normales y simétricas, surgen notables diferencias (Coin, 2007). En la tabla 2, tabla 3, tabla 4, tabla 5 y tabla 6 se sintetizaron los resultados de varios estudios realizados en este sentido:

Tabla 1- Potencia empírica para distribución $U [0,1]$ -Rahman y Govindarajulu (1997)-

Prueba de Normalidad	Potencia Empírica	
	n=10; $\alpha=0,10$	n=20; $\alpha=0,10$
Shapiro-Wilk (W)	16%	36%
Shapiro-Wilk modificado por Rahman y Govindarajulu (\tilde{W})	27%	54%

Tabla 2 - Potencia empírica para distribución $U [0,1]$ -Coin (2007)-

Prueba de Normalidad	n=20; $\alpha=0,05$
Kolmogorov-Smirnov ($K-S$)	10%
Anderson-Darling (AD)	17%
Zhang y Wu (Z_A)	6%
Jarque-Bera (JB)	1%
Bonett-Seier (T_w)	21%
Shapiro-Wilk (W)	20%
D'Agostino (Dag)	10%
Coin (β_3^2)	30%

Tabla 3 - Potencia empírica para distribuciones simétricas -Romão *et al.* (2010)-

Prueba de Normalidad	n=25; $\alpha=0,05$	n=25; $\alpha=0,10$
Kolmogorov-Smirnov modificado por Lilliefors (D)	39%	46%
Anderson-Darling (AD^*)	45%	52%
Zhang y Wu (Z_C)	44%	52%
Zhang y Wu (Z_A)	44%	51%
Glen-Leemis-Barr (P_S)	45%	52%
D'Agostino-Pearson combinado (K^2)	42%	50%
Hosking (T_{mom})	47%	54%
Bonett-Seier (T_w)	45%	52%
Shapiro-Wilk (W)	46%	53%
Shapiro-Francia (W')	44%	50%
Shapiro-Wilk modificado por Rahman y Govindarajulu (\tilde{W})	44%	51%
D'Agostino (Dag)	40%	46%
Filliben (r)	43%	50%
Chen-Shapiro (CS)	46%	53%
Barrio-Cuesta-Albertos-Matrán-Rodríguez ($BCMR$)	45%	52%
Coin (β_3^2)	49%	56%
Gel-Miao-Gastwirth (R)	46%	52%

Tabla 4 - Potencia empírica para distribución $U [0,1]$ -Razali y Wah (2011)-

Prueba de Normalidad	$\alpha=0,05$		$\alpha=0,10$	
	n=10	n=20	n=10	n=20
<i>Shapiro-Wilk (W)</i>	9%	20%	18%	36%
<i>Kolmogorov-Smirnov (K_S)</i>	9%	11%	16%	17%
<i>Kolmogorov-Smirnov modificado por Lilliefors (D)</i>	7%	10%	13%	19%
<i>Anderson-Darling (AD*)</i>	9%	17%	17%	29%

Tabla 5 - Potencia empírica para distribución $U [0,1]$ -Yap y Sim (2011)-

Prueba de Normalidad	n=10; $\alpha = 0,05$	n=20; $\alpha = 0,05$
<i>Kolmogorov-Smirnov (K-S)</i>		9%
<i>Kolmogorov-Smirnov modificado por Lilliefors (D)</i>		7%
<i>Anderson-Darling (AD)</i>		9%
<i>Shapiro-Wilk (W)</i>		10%
<i>Jarque-Bera (JB)</i>		2%
<i>Cramer-von Mises (CVM)</i>		8%
<i>D'Agostino y Pearson (K²)</i>		2%
<i>Chi cuadrado de Pearson</i>		0%

Tabla 6 - Potencia empírica para distribución $U [0,1]$ - Brzezinski (2012)-

Prueba de Normalidad	Potencia Empírica	
	n=20; $\alpha=0,05$	n= 20; $\alpha=0,10$
<i>D'Agostino-Pearson (según D'Agostino et al. 1990)</i>	13%	27%
<i>Shapiro-Wilk (W)</i>	20%	36%
<i>Shapiro-Francia (W')</i>	8%	19%
<i>Chen-Shapiro (CS)</i>	22%	37%

Por otra parte, en la tabla 7 se resumió la disponibilidad de las pruebas de bondad de ajuste en software estadísticos antes indicados. Cabe destacar, que pese a que muchos estudios plantean el escaso poder empírico de la prueba de Kolmogorov-Smirnov y de su versión modificada por Lilliefors (estadísticos de prueba K-S y D), éstos se encuentran disponibles en los *software* estadísticos (Romão *et al.* 2010). Se resalta además, la presencia casi nula en la literatura estadística de la prueba de Ryan-Joiner presente en Minitab 17.

Tabla 7 - Disponibilidad de pruebas de normalidad en *software* estadísticos

Prueba de Normalidad	Spss 19	Infostat 2014	Stata 11	Minitab 17
<i>Kolmogorov-Smirnov</i>		x		
<i>Kolmogorov-Smirnov modificado por Lilliefors</i>	x	x		
<i>Anderson-Darling (AD*)</i>				x
<i>D'Agostino-Pearson (según D'Agostino et al. 1990)</i>			x	
<i>Shapiro-Wilk</i>	x		x	
<i>Shapiro-Francia</i>			x	
<i>Shapiro-Wilk modificado por Rahman y Govindarajulu</i>		x		
<i>Chen-Shapiro</i>			x	
<i>Ryan-Joiner</i>				x

Ahora bien, en las tablas 8 y 9 se muestran los resultados de la experimentación con las pruebas disponibles en los paquetes estadísticos mencionados. Se incluyó además, el cálculo mediante programación computacional de la potencia empírica de la prueba de Gel-Miao-Gastwirth (R) - estadístico recomendado para la detección de distribuciones simétricas no normales en muestras de tamaño de 25 datos o más y de fácil programación computacional-.

Tabla 8 - Potencia empírica para distribución $U [0,1]$ obtenida experimentalmente

Pruebas de Normalidad	n=10; $\alpha=0,10$	Productos computacionales
<i>Shapiro-Wilk modificado por Rahman y Govindarajulu</i>	25%	Infostat
<i>Shapiro-Wilk</i>	18%	Spss
<i>Anderson-Darling (AD*)</i>	16%	Minitab
<i>D'Agostino-Pearson (según D'Agostino et al. 1990)</i>	15%	Stata
<i>Kolmogorov-Smirnov modificado por Lilliefors</i>	13%	Infostat
<i>Shapiro-Francia</i>	11%	Stata
<i>Ryan-Joiner</i>	11%	Minitab
<i>Gel-Miao-Gastwirth</i>	10%	No programado

Tabla 9 - Potencia empírica para distribución $U [0,1]$ obtenida experimentalmente

Pruebas de Normalidad	n=15; $\alpha=0,10$	Productos computacionales
<i>Shapiro-Wilk modificado por Rahman y Govindarajulu</i>	45%	Infostat
<i>Shapiro-Wilk</i>	29%	Spss
<i>Anderson-Darling (AD*)</i>	26%	Minitab
<i>D'Agostino-Pearson (según D'Agostino et al. 1990)</i>	15%	Stata
<i>Kolmogorov-Smirnov modificado por Lilliefors</i>	17%	Infostat
<i>Shapiro-Francia</i>	18%	Stata
<i>Ryan-Joiner</i>	17%	Minitab

En coincidencia con los resultados que fueron obtenidos por Rahman y Govindarajulu (1997), la prueba de Shapiro-Wilk modificada por estos autores; resultó la de mayor potencia empírica para la detección de la distribución Uniforme como alternativa a la Normal, en muestras de tamaño 10 y 15 datos y para un nivel de significancia del 10%. Es de destacar que ésta prueba se encuentra disponible en el *software* estadístico *Infostat 2014*, desarrollado en la Universidad Nacional de Córdoba.

De todos modos, parece claro que en tren de verificar la normalidad de los datos cuando se cuenta con muestras pequeñas, es conveniente complementar la aplicación de estas pruebas con otros estudios que ayuden a validar sus conclusiones. En particular, las herramientas gráficas y los coeficientes de asimetría y curtosis pueden ser de gran ayuda para determinar posibles causas de la desviación de la normalidad.

Bibliografía

- Althouse, L. A., Ware, W. B., Ferron, J. M. (1998). "Detecting Departures from Normality: A Monte Carlo Simulation of a New Omnibus Test Based on Moments". *Annual Meeting of the American Educational Association Research*, San Diego, CA.
- Anderson, T. W., Darling, D. A. (1954). "A Test of Goodness of Fit". *Journal of Statistical Association*, 49(268), 765-769.
- Bonett, D. G., Seier, E. (2002). "A test of normality with high uniform power". *Computational statistics & data analysis*, 40(3), 435-445.
- Brzezinski, M. (2012). "The Chen-Shapiro test for normality". *Stata Journal*, 12 (3), 368-374.
- Chen, L., Shapiro S. S (1995). "An alernative test for normality based on normalized spacings". *Journal of Statistical Computation and Simulation*, 53, 269-287.
- Coin, D. (2007). "A goodness-of-fit test for normality based on polynomial regression". *Computational statistics & data analysis*, 52(4), 2185-2198.
- D'Agostino R., Pearson, E. S. (1973). "Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$ ". *Biometrika*, 60, 613-622.

- D'Agostino, R. B. (1971). "An omnibus test of normality for moderate and large sample sizes". *Biometrika*, 58, pp. 341-348.
- D'Agostino, R., Stephens, M. (1986). Goodness-of-fit-techniques. Statistics.
- Del Barrio, E.; Cuesta-Albertos, J.A., Matrán, C., Rodríguez-Rodríguez, J. M. (1999). "Tests of goodness of fit based on the L2-Wasserstein distance". *Ann. Stat.*, 27(4), pp. 1230–1239.
- Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., Robledo C.W. InfoStat versión 2014. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>
- Dufour, J. M., Farhat, A., Gardiol, L., Khalaf, L. (1998). "Simulation-based Finite Sample Normality Tests in Linear Regressions". *The Econometrics Journal*, 1(1), 154-173.
- Filliben, J. J. (1975). "The probability plot correlation coefficient test for normality". *Technometrics* 17(1), 111–117
- Gan, F. F., Koehler, K. J. (1990). "Goodness-of-Fit Tests Based on P-P Probability Plots". *Technometrics*, 32(3), 289-303.
- Gel, Y. R., Miao, W., Gastwirth, J. L (2007). "Robust directed tests of normality against heavy-tailed alternatives". *Comput. Stat. Data Anal.* 51(5), 2734–2746.
- Gel, Y.R., Gastwirth, J.L. (2008). "A robust modification of the Jarque–Bera test of normality". *Economics Lett*, 99(1), 30–32.
- Glen, A. G., Leemis, L. M., Barr, D. R. (2001). "Order statistics in goodness-of-fit testing". *IEEE Trans. Reliab.* 50(2), 209–213.
- Henderson, A. R. (2006). "Testing experimental data for univariate normality". *Clin. Chim. Acta* 366(1/2), 112–129
- Hosking, J.R.M (1990). "L-moments: Analysis and estimation of distributions using linear combinations of order statistics". *J. R. Statist. Soc. Ser. B* 52(1), pp. 105–124.
- Jarque, C. M., Bera, A. K. (1980). "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". *Economics Letters*, 6(3), 255-259.
- Kolmogorov, A. N. (1933). "Sulla determinazione empirica di una legge di distribuzione". *Giornale dell' Instituto Italiado degli Attuari*, 4, 83-91.
- Lilliefors, H. (1967). "On the Kolmogorov–Smirnov test for normality with mean and variance unknown". *J. Amer. Statist. Assoc.*, 62(318), 399–402.
- Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill International.
- Pearson, E.S.; D'Agostino, R.B.; Bowman, K.O. (1977). "Tests for departure from normality: comparison of powers". *Biometrika*, 64 (2), pp. 231–246.
- Rahman, M. M., Govindarajulu, Z. (1997). "A modification of the test of Shapiro and Wilk for normality". *Journal of Applied Statistics*, 24 (2), 219-236.
- Razali, N. M.; Wah, Y. B. (2011). "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests". *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.
- Romão, X., Delgado, R., Costa, A. (2010). "An empirical power comparison of univariate goodness-of-fit tests for normality". *Journal of Statistical Computation and Simulation*, 80(5), 545-591.
- Seier, E. (2002). "Comparison of Tests for Univariate Normality". *InterStat Statistical Journal*, 1, 1-17.
- Shapiro, S. S., Wilk, M. B., Chen, H. J. (1968). "A Comparative Study of Various Tests of Normality". *Journal of American Statistical Association*, 63(324), 1343-1372.
- Shapiro, S.S.; Francia, R. (1972). "An approximation analysis of variance test for normality". *Journal of the American Statistical Association*, 67, 215–216.
- StataCorp, L. P. (2007). "Stata data analysis and statistical Software". *Special Edition Release*, 10.
- Tanveer-UI-Islam (2011). "Normality testing—A new direction. International". *Journal of Business and Social Science*, 2(3), 115–118.
- Yap, B. W., Sim C. H. (2011). "Comparisons of various types of normality tests". *Journal of Statistical Computation and Simulation*, 81(12), 2141-2155.
- Zhang, J.; Wu, Y. (2005). "Likelihood-ratio tests for normality". *Comput. Stat. Data Anal.* 49(3), 709–721.