Comparación de potencias en pruebas estadísticas de normalidad, con datos escasos

Gabriela Cabrera[1], José Francisco Zanazzi[1], José Luis Zanazzi[1], Laura Boaglio[1]

[1]*Laboratorio de Ingeniería y Mantenimiento Industrial. FCEFyN, UNC, Argentina.*

Fecha de recepción del manuscrito: 31/03/2017
Fecha de aceptación del manuscrito: 10/09/2017
Fecha de publicación: 15/09/2017

Resumen— 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. Frecuentemente en estos problemas, se cuenta con pocas observaciones. En el ámbito de la estadística se proponen muchas pruebas de hipótesis para verificar normalidad. 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.

Palabras clave— Estadística. Pruebas para verificar normalidad. Muestras pequeñas. Potencia de las pruebas.

Abstract— Verification that the observed data on a random phenomenon may be assumed to have been drawn from a normal probability distribution is necessary in a variety of practical situations. Frequently in these problems, there are few observations. In the field of statistics many hypothesis tests are proposed to verify normality. Some researchers have conducted studies on most of the tests in question, including the determination of power. Unfortunately, verifications, generally operate with samples of fifty or more data. Therefore, in this work the power of various tests is estimated with samples of ten, fifteen and twenty data. Besides, power is determined in situations where the original distribution is symmetrical, which undoubtedly represents the worst condition for the test. For these operations simulation experiments are performed. The paper concludes with a qualitative assessment of the appropriateness of the analyzed tests.

Keywords— Statistics– Normal Probability test – Small Samples – Power of Tests

INTRODUCCIÓN

En el ámbito de la ingeniería y áreas del conocimiento asociadas, es frecuente la necesidad de verificar si un pequeño conjunto de datos, puede considerarse observado sobre una población con Distribución Normal (Montgomery y Runger, 2010). Con esa finalidad, la literatura especializada en estadística ha propuesto una interesante variedad de procedimientos, denominados pruebas de hipótesis; pero lamentablemente, no existe una aproximación que supere con holgura a las restantes (Tanveer, 2010). Ante esa problemática, el presente artículo se aboca a comparar diferentes metodologías y procura identificar características distintivas de cada una.

Al analizar el supuesto de normalidad con datos estadísticos, frecuentemente surge la dificultad de que las muestras tienen longitud reducida. Tal situación se produce, por ejemplo, cuando se realizan estudios de duración de actividades en procesos productivos o cuando en estudios de confiabilidad de sistemas, se desarrollan determinaciones de la vida útil de ciertos equipamientos.

Por otra parte, desde el punto de vista conceptual se asume que existen dos comportamientos extremos para estas variables, que representan realidades opuestas. Por un lado, se considera deseable que la variable tenga un comportamiento compatible con la Distribución Normal, dado que se lo considera representativo de un proceso trabajado, donde se han adoptado las acciones necesarias para reducir la dispersión. Por otro lado, la distribución uniforme se considera evidencia de falta de control y de existencia de variaciones excesivas (Mystica y otros, 2015).

Por ejemplo, sea el caso de un proceso en el que se fabrican pistones para motores, donde se considera que el diámetro de las piezas es una característica de calidad clave. Si el diámetro tiene distribución rectangular, es tan posible que la medida se ubique en el valor nominal, como en valores que van más allá de las tolerancias de ingeniería. En cambio, si tiene un comportamiento gaussiano, los pistones resultan muy similares y verifican con facilidad las especificaciones técnicas. En general, la normalidad de una variable que representa los resultados de un proceso productivo, se considera evidencia de control sobre el proceso estudiado (Deming, 2000).

Ahora bien, al aplicar una prueba de hipótesis, es posible cometer dos tipos de errores. El denominado error tipo I, consiste en rechazar una suposición que es correcta. Por su parte, el error de tipo II se produce cuando no se rechaza una hipótesis falsa. Se denomina potencia de la prueba al complemento de la probabilidad del segundo tipo de error. En términos de pruebas de normalidad, es la posibilidad que ante datos que provienen de poblaciones no normales, la prueba detecte esta cuestión (Montgomery y Runger, 2010).

En el ámbito de la estadística, se han propuesto una importante cantidad de pruebas de normalidad (Seier, 2002; Farrel y Stewars, 2006; Henderson, 2006; Öztuna, Elhan y Tüccar, 2006; Yazici y Yolacan, 2007, Gel, Miao y Gastwirth, 2007; Coin, 2007; Tanveer, 2011; Yap y Sim, 2011; Razali, Shamsudin, Azid, Hadi & Ismail, 2012; Lafaye de Micheaux & Tran, 2014). Al seleccionar el procedimiento a utilizar, es importante que el usuario considere la probabilidad de cometer un error de tipo II.

Algunos artículos se han orientado a estudiar y comparar estas pruebas, donde uno de los aspectos considerados es precisamente la potencia del test (Romão, Delgado y Costa, 2010). Lamentablemente, estos trabajos operan generalmente con muestras de veinticinco, cincuenta o más datos. Además, no prestan especial interés a la distinción entre las distribuciones normal y uniforme.

Ante esa problemática, el presente trabajo realiza una revisión de diferentes aproximaciones y selecciona pruebas que pueden resultar apropiadas para verificar normalidad, cuando las muestras son pequeñas. Para las herramientas escogidas, se estima la potencia de cada prueba en situaciones donde la distribución original es uniforme; las estimaciones se obtienen mediante experimentos de simulación. Finalmente, se presentan una suerte de comentarios y sugerencias, acerca de la conveniencia de utilizar algunos de los tests considerados.

En cuanto a la organización del documento, después de la introducción se realiza una revisión bibliográfica, tanto de posibles pruebas, como de trabajos dirigidos a inferir la potencia de las mismas. Luego se presenta la metodología utilizada y a continuación se discuten los principales resultados.

REVISIÓN DE ANTECEDENTES BIBLIOGRÁFICOS

En el área del conocimiento vinculada con estadística, se han propuesto más de cuarenta pruebas que permiten evaluar el supuesto de normalidad (Dufour *et al.* 1998). Actualmente se observa un creciente interés por estudiar el rendimiento de estos tests para diferentes alternativas.

El interés por contar con métodos que permitan detectar desviaciones respecto a la distribución normal de probabilidades, se inicia con Pearson en 1895, con un estudio de los coeficientes de asimetría y curtosis (Razali y Wah, 2011). En la actualidad, se entiende que no existe una única prueba de normalidad que merezca ser considerada como la más indicada (Tanveer, 2011). En otras palabras, debido a la gran variedad de alternativas a la normalidad, no existe una prueba más potente en términos generales.

Por el contrario, hay pruebas que son más potentes para ciertos objetivos, en tanto que pierden validez para otros. En esta dirección enfocan sus investigaciones: Shapiro, Wilk & Chen (1968), Chen y Shapiro (1995), Seier (2002), Thadewald y Büning (2007), Poitras (2006), Farrel y Stewart (2006), Öztuna *et al.* (2006), Yazici y Yolacan (2007), Úrzua (2007), Gel y Gastwirth (2008), Razali y Wah (2011), Quessy y Mailhot (2011), Yap y Sim (2011), Razali *et al.* (2012), Lafaye de Micheaux y Tran (2014), entre otros.

Entre las pruebas de normalidad más conocidas, se pueden citar cuatro grupos (Arshad, Rasool y Ahmad, 2003). El primero se encuentra formado por aquellas en las que se mide el grado de discrepancia entre las distribuciones empíricas y la función de distribución acumulada normal; en esta línea se encuentran: Kolmogorov-Smirnov (Kolmogorov, 1933), Lilliefors (Lilliefors, 1967), Anderson-Darling (Anderson y Darling, 1954), (Darling, 1957), Cramer-von Mises (Cramer, 1928), (Von Mises, 1931), (Smirnov, 1936).

El segundo grupo tiene como estrategia común el análisis de la correlación entre la distribución teórica y la experimental; se basan en la relación de dos estimaciones por mínimos cuadrados, ponderados por una escala obtenida de las estadísticas de orden (Dufour, Farhat, Gardiol & Khalaf, 1998). En este conjunto se destacan las pruebas de Shapiro-Wilk (Shapiro y Wilk, 1965; Royston, 1982, 1995); Shapiro-Francia (Shapiro y Francia, 1972; Sarkadi, 1975), la modificación del estadístico de prueba de Shapiro-Wilk propuesta por Rahman y Govindarajulu (Rahman, Govindarajulu, 1997) y el test de Chen-Shapiro (Chen y Shapiro, 1995) entre otros.

En el tercer grupo se consideran aquellas metodologías que se sustentan en la idea de que las desviaciones de normalidad pueden ser detectadas por dos momentos de la muestra: la asimetría y la kurtosis. Este enfoque se adopta en D'Agostino y Pearson (1973), D´Agostino, Belanger, & D'Agostino Jr., (1990), Jarque-Bera (1987) y Gel y Gastwirth (2008).

Por último, el cuarto grupo se encuentra formado por pruebas especiales que no pueden ser encuadradas en la clasificación anterior, como es el caso de la propuesta formulada en Yazici y Yolacan (2007) y la sugerida en Romão *et al.* (2010). Cabe precisar que en general, estas aproximaciones no se encuentran disponibles en los programas orientados al análisis estadístico.

Respecto a la conveniencia de aplicar una u otra prueba, Romao *et al.* (2010) analiza las potencias de estas herramientas, ante diferentes tamaños de muestra. Concluye que no es posible identificar a uno de estos tests como cercano al ideal, dado que el resultado depende de las condiciones del problema y de la verdadera distribución de los datos.

Al respecto, los estudios más tempranos (Shapiro y Wilk, 1965; Shapiro *et al.* 1968; Pearson *et al.* 1977; Gan y Koehler, 1990; D'Agostino *et al.* 1986), sugieren, que la mayoría de los procedimientos analizados funcionan bien cuando las distribuciones alternativas de no normalidad resultan fuertemente sesgadas.

Dicho de otro modo, una estrategia aceptada como primera aproximación, para determinar si los datos de una muestra pueden considerarse extraídos de una población Normal, es comparar el coeficiente de asimetría muestral con cero. Esa posibilidad carece de sentido cuando se sospecha que la distribución original en realidad es simétrica.

Como regla general, puede aceptarse que si la distribución original no es normal, las pruebas tienen mejores posibilidades de detectar esa desviación, si la distribución verdadera es notoriamente diferente de la gaussiana, como es el caso, por ejemplo, de la exponencial. En cambio, cualquier procedimiento de verificación va a evidenciar mayores dificultades cuando el comportamiento original es simétrico (Coin, 2007).

SUPUESTOS Y METODOLOGÍA ADOPTADA

En el presente documento se adopta como hipótesis nula el supuesto de que la variable analizada tiene en realidad distribución de probabilidad normal. Como alternativa, interesa la posibilidad de que la distribución verdadera sea uniforme. Se supone como punto de partida, que esta es una de las peores condiciones posibles para los tests de normalidad, debido a que la distribución rectangular es simétrica.

Respecto a la modalidad de trabajo, en la realización es posible distinguir las siguientes etapas:

a) Investigación bibliográfica: se revisan los artículos orientados a proponer o estudiar pruebas de verificación de la distribución normal. En esos trabajos, se presta atención a las características de los métodos y a experimentos orientados a valorar la potencia de los procedimientos propuestos. Como producto de esta revisión, se identifica un conjunto de procedimientos que resulten apropiados para distinguir entre las distribuciones normal y uniforme, cuando se opera con muestras pequeñas. Las condiciones que hacen que una prueba pueda ser seleccionada, son las siguientes:

- Potencia del test: se eligen aquellas que obtienen buenos resultados para tamaños de muestras de veinte o más datos.
- Amigabilidad del procedimiento: se considera preferible que la estrategia y especialmente el estadístico, resulten comprensibles para los usuarios, aunque no tengan una fuerte formación en estadística;
- Facilidad de implementación: se considera conveniente que el procedimiento seleccionado, se encuentre disponible en un programa de software, o que resulte sencilla su implementación con hojas de cálculo.

b) Experimentación con las pruebas seleccionadas. El conjunto de pruebas seleccionadas en la fase anterior, se somete a una serie de experimentos de simulación que permitan estimar su potencia, cuando se trabaja con muestras de diez y quince datos. Para cada tamaño de muestra, se generan mil conjuntos de números con distribución rectangular. A continuación, se aplica cada una de las pruebas elegidas con las mil muestras y se determina la proporción de veces que el test detecta que en realidad la distribución no es normal.

RESULTADOS Y DISCUSIÓN

Como producto de la investigación de bibliografía, se concretó una primera selección de veinte pruebas, divididas en cuatro grupos según la estrategia utilizada. A continuación se enumeran los tests seleccionados.

Se definió un primer grupo con pruebas que se apoyan en el análisis de los primeros momentos estadísticos de las muestras. La idea es comparar los valores obtenidos a partir de los datos, con los esperables cuando la distribución es normal. En la siguiente Tabla se hace referencia a los artículos donde se proponen estas aproximaciones.

TABLA 1. PRUEBAS BASADAS EN MOMENTOS

Nro	Nombre
1	*D´Agostino-Pearson (1973)*
2	*Jarque-Bera (1980)*
3	*Prueba robusta de Jarque-Bera (Gel y Gastwirth, 2008)*
4	*Bonett-Seier (2002)*
5	*Hosking (1990)*

Un segundo grupo se orienta a realizar el análisis de la función empírica acumulada. En efecto, se comparan la función de distribución empírica (estimada con base en los datos de la muestra), con la función de distribución acumulada de la normal. La estrategia consiste en analizar el grado de similitud o diferencia entre las dos funciones.

Dufour *et al.* (1998) entiende este grupo de pruebas, como basado en una medida de la discrepancia entre las distribución empírica y la distribución que se propone en la hipótesis nula. A su vez, este conjunto de pruebas puede subdividirse en dos sub-grupos: las que utilizan el supremo de discrepancias y las que trabajan con el cuadrado de las mismas. En la Tabla 2 se enumeran estas aproximaciones.

TABLA 2. PRUEBAS SOBRE LA DISTRIBUCIÓN EMPÍRICA

Nro	Nombre
6	*Kolmogorov-Smirnov (1933)*
7	*Kolmogorov-Smirnov modificado por Lilliefors (1967)*
8	*Kolmogorov-Smirnov modificado por Stephens y Harley (1972)*
9	*Anderson y Darling (1954)*
10	*Zhang y Wu (2005)*
11	*Glen, Leemis y Barr (2001)*

El tercer grupo, comprende a las pruebas de correlación y regresión, que se basan en el cociente de dos estimaciones de escala, obtenidos por el método de mínimos cuadrados de los estadísticos de orden. Las dos estimaciones se distribuyen normalmente, en el numerador se propone una estimación por mínimos cuadrados ponderados y en el denominador la varianza de la muestra de otra población. La Tabla 3 lista estas propuestas.

TABLA 3. PRUEBAS DE CORRELACIÓN Y REGRESIÓN

Nro.	Nombre
12	*Shapiro-Wilk (1965)*
13	*Shapiro-Francia (1972)*
14	*Chen-Shapiro(1995)*
15	*Modificación del Shapiro-Wilk (Rahman y Govindarajulu, 1997)*
16	*Modificación de Shapiro-Wilk propuesta por D´Agostino(1971)*
17	*Filliben (1975)*

Además, se identifica un cuarto grupo integrado por aproximaciones que no pueden encuadrarse en los conjuntos anteriores. La Tabla 4 identifica a estas aproximaciones.

TABLA 4. PRUEBAS NO ENCUADRADAS EN LAS ANTERIORES

Nro	Nombre
18	*Prueba de correlación de cuantiles (Del Barrio et al., 1999)*
19	*Prueba de Coin (2007)*
20	*Gel, Miao y Gastwirth (2007)*

Como producto adicional de la revisión de bibliografía, es posible obtener diferentes evaluaciones de las potencias ofrecidas por estas pruebas. En la Tabla 5 se lista la potencia de aquellas pruebas de normalidad que en el trabajo de Romão *et al.* (2010) tienen una potencia empírica mayor al 40%; para distribuciones alternativas simétricas no normales y n=25.

TABLA 5. EVALUACIONES DE POTENCIAS CON TAMAÑO 25

Prueba de Normalidad	**α**	
	0,05	**0,1**
Kolmogorov - Smirnov	39,3	46,3
Anderson - Darling	45,1	51,8
Zhang y Wu	44,4	52
Glen-Leemis-Barr	45,2	51,9
D` Agostino-Pearson	41,5	49,8
Hosking	47	53,6
Shapiro - Wilk	45,5	52,6
Shapiro - Francia	43,5	50,4
Shapiro - Wilk modificado R y G	44,4	51,3
D`Agostino	40	46,1
Fillibien	43,1	49,9
Chen - Shapiro	45,6	52,7
Coin	48,5	55,6
Gel - Miao - Gastwirth	45,8	52,4

En la Tabla 5, α es la probabilidad asignada a la zona de rechazo de la hipótesis, esto es, la posibilidad de cometer un error del primer tipo. Cabe recordar que a medida que el nivel de significación aumenta, disminuye la probabilidad de cometer un error tipo II y con ello, se incrementa la potencia de la prueba (Montgomery y Runger, 2010).

Con ese razonamiento, si el analista se preocupa por detectar desviaciones respecto al comportamiento gaussiano, parece recomendable adoptar valores grandes de nivel de significación, en este caso 0,10. Desde ese punto de vista, el resultado es alentador, dado que varias de las pruebas tienen potencias mayores al 50%.

En cuanto a la disponibilidad de estas herramientas en los programas de computadora que ofrecen soporte estadístico y que se utilizan frecuentemente en nuestro país, a los fines de este trabajo se analizaron los paquetes Infostat (desarrollado en la Universidad Nacional de Córdoba), SPSS, Stata y Minitab. La Tabla 6 resume la disponibilidad de cada una de estas pruebas, en las herramientas computacionales mencionadas.

TABLA 6. PRUEBAS Y SOFTWARE

Pruebas	**SPSS**	**Infostat**	**Stata**	**Minitab**
Kolmogorov - Smirnov		**x**		
K -S de Lilliefors	**x**	**x**		
Anderson - Darling				**x**
D` Agostino-Pearson			**x**	
Shapiro - Wilk	**x**		**x**	
Shapiro - Francia			**x**	
Shapiro - Wilk modificado R y G		**x**		
Chen - Shapiro			**x**	
Ryan - Joiner				**x**

Ante esta evidencia, se realizó una selección de ocho pruebas, para las cuales se aproximó experimentalmente la potencia con la metodología antes planteada. Corresponde destacar que todas estas pruebas son consideradas como “No paramétricas”, en la literatura especializada en estadística. Los resultados obtenidos para muestras de diez y quince datos, se reproducen en Tabla 7.

De la Tabla 7 se desprende que la prueba de Shapiro-Wilk modificada por Rahman y Govindarajulu, resulta la de mayor potencia para la detección de la distribución uniforme como alternativa a la normal, en muestras de tamaño diez y quince, para un nivel de significancia del diez por ciento. Estos resultados son coincidentes con los presentados por Rahman y Govindarajulu (1997), para muestras más extensas. A la mencionada prueba le siguen, de la mayor a la menor potencia, el test de Shapiro-Wilk (W) y la prueba Anderson-Darling (AD).

Un detalle interesante es que el test de Shapiro-Wilk modificado por Rahman y Govindarajulu, evidencia una leve disminución de potencia al pasar de veinticinco a quince datos. De todos modos, en ningún caso es posible alcanzar una potencia superior al cincuenta por ciento.

TABLA 7. POTENCIA EMPÍRICA OBTENIDA A = 0,10

Pruebas de Normalidad	Potencia empírica		Software
	n = 15	n=10	
Shaprio-Wilk modificado por Rahman y Govindarajulu	45%	25%	Infostat
Shapiro-Wilk	29%	18%	Spss
Anderson-Darling (AD)*	26%	16%	Minitab
D´Agostino-Pearson (según D´Agostino et al. 1990)	15%	15%	Stata
Kolmogorov-Smirnov modificado por Lilliefors	17%	13%	Infostat
Shapiro-Francia	18%	11%	Stata
Ryan-Joiner	17%	11%	Minitab
Gel-Miao-Gastwirth	14%	10%	No disponible

Respecto a la prueba de Gel-Miao-Gastwirth, fue incluida en la simulación debido a que resulta fácil de interpretar y de calcular. Además el mencionado test obtiene resultados muy interesantes con muestras de más de cincuenta datos. Sin embargo, para las cantidades de datos analizados en la simulación, las potencias obtenidas son bajas.

CONCLUSIONES

En el presente trabajo se estudia el problema de verificar normalidad con muestras pequeñas. Con esa finalidad se identifican veinte pruebas diferentes y se realiza un análisis comparativo bajo la consideración de tres criterios: potencia empírica para la detección de la distribución uniforme, como alternativa a la normal; amigabilidad del procedimiento y disponibilidad de software.

La investigación bibliográfica ha permitido seleccionar ocho pruebas de tipo no paramétrico, que brindan resultados interesantes para muestras grandes. Se experimentó con esos procedimientos para muestras pequeñas y se encontró que ninguno de los procedimientos analizados alcanza una potencia superior al cincuenta por ciento.

Dentro de las aproximaciones estudiadas, la prueba de Shapiro-Wilk, modificada por Rahman y Govindarajulu (1997), alcanza los mejores niveles de potencia, tiene una lógica amigable y se encuentra disponible en un software estadístico de fácil acceso.

Por último, los resultados obtenidos evidencian que el problema de verificar la normalidad en muestras pequeñas, cuando la hipótesis alternativa es una distribución simétrica, merece mayor atención. En efecto, debido a los múltiples y frecuentes requerimientos en ese sentido, debería ser factible el desarrollo de nuevos estadísticos que aseguren una mayor potencia.

REFERENCIAS

[1] Anderson, T. W. y Darling, D. A. (1954): "A Test of Goodness of Fit". Journal of Statistical Associattion, vol. 49, 268, pp. 765-769.

[2] Arshad, M.; Rasool, M. T.; Ahmad, M. I. (2003): "Anderson Darling and Modified Anderson Darling Test for Generalized Pareto Distribution". Pakistan Journal of Applied Sciences, vol. 3, 2, pp. 85-88.

[3] Chen, l.; Shapiro S. S (1995): "An alternative test for normality based on normalized spacings". Journal of Statistical Computation and Simulation, vol. 53, pp. 269-287.

[4] Coin, D. (2007). "A goodness-of-fit test for normality based on polynomial regression". Computational statistics & data analysis, vol. 52, nro. 1, pp. 2185-2198.

[5] Cramér, H. (1928): "On the composition of elementary errors: First paper: Mathematical deductions". Scandinavian Actuarial Journal, vol. 1, pp. 13-74.

[6] D'agostino, R. B., Belanger, A., & D'agostino Jr., R. B. (1990): "A suggestion for using powerful and informative tests of normality". The American Statistician, vol. 44, nro. 4, pp. 316-321.

[7] D'agostino R.; Pearson, E. S. (1973): "Tests for departure from normality. Empirical results for the distributions of b2 and √b1". Biometrika, vol. 60, pp. 613-622.

[8] D'agostino, R. B.; Stephens, M. A.; D'agostino, R. B.;Stephens, M. A. (1986). Goodness-fo-fit-techniques. Statistics.

[9] Darling, D. A. (1957): "The Kolmogorov-Smirnov, Cramer-von Mises tests". The Annals of Mathematical Statistics, pp. 823-838.

[10] Deming, W. E. (2000). Out of the Crisis. MIT press.

[11] 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.

[12] Dufour, J. M.; Farhat, A.; Gardiol, l.; Khalaf, l.(1998): Simulation- based Finite Sample Normality Tests in Linear Regressions". The Econometrics Journal, vol. 1, 1, pp. 154-173.

[13] 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, vol. 27, 4, pp. 1230-1239.

[14] Farrel, P. J.; Rogers-Stewart, K.R. (2006): "Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test". Journal of Statistical Computation and Simulation, vol. 76, 9, pp. 803-816.

[15] Filliben, J. J. (1975): "The probability plot correlation coefficient test for normality", Technometrics, vol. 17, 1, pp. 111-117.

[16] Gan, F. F., & Koehler, K. J. (1990). Goodness-of-Fit Tests Based on P-P Probability Plots. Technometrics, 32(3), 289-303.

[17] Gel, Y. R.; Gastwirth, J. L. (2008): "A robust modification of the Jarque–Bera test of normality". Econom. Lett, vol. 99, 1, pp. 30-32.

[18] Gel, Y. R., Miao, W., Gastwirth, J. L (2007): "Robust directed tests of normality against heavy-tailed alternatives", Comput. Stat. Data Anal, vol. 51, 5, pp. 2734-2746.

[19] Glen, A. G., Leemis, l. M., Barr, D. R. (2001): “Order statistics in goodness-of-fit testing”, IEEE Trans. Reliab, vol. 50, 2, pp. 209-213.

[20] Henderson, A. R. (2006): “Testing experimental data for univariate normality”, Clinica chimica acta, 366(1), pp. 112-129.

[21] Jarque, C. M.; Bera, A. K. (1987): “A Test for Normality of Observations and Regression Residuals”. International Statistical Review, vol. 55, 2, pp. 163-172.

[22] Kolmogorov, A. N. (1933): “Sulla determinazione empírica di uma legge di distribuzione”. Giornalle dell´Instituto Italiano degli Attuari, vol. 4, pp. 83-91.

[23] Lafaye De Micheaux, P.; Tran, V. A. (2014): “Power R: Reproducible Research Toll to ease Carlo Power Simulation Studies for Goodness-of-fit Test R”. Journal of Statistical Software, vol. 27, pp. 1230-1239.

[24] Lilliefors, H. (1967): “On the Kolmogorov-Smirnov test for normality with mean and variance unknown”. J. Amer. Statist. Assoc, vol. 62, 318, pp. 399-402.

[25] Montgomery, D. (1991): “Control Estadístico de la Calidad”. México, Iberoamérica. 17-44.

[26] Montgomery, D. C., & Runger, G. C. (2010): “Applied statistics and probability for engineers”. John Wiley & Sons.

[27] Mystica, A., Bai, J., & Suganthi, M. (2015). “Statistical process control”. Clear International Journal of Research in Commerce & Management, 6(1).

[28] Öztuna, D., Elhan, A. H., & Tüccar, E. (2006): “Investigation of four different normality tests in terms of type 1 error rate and power under different distributions”. Turkish Journal of Medical Sciences, 36(3), 171-176.

[29] Pearson, E.S.; D’agostino, R.B.; Bowman, K.O. (1977): “Tests for departure from normality: comparison of powers”. Biometrika, vol. 64, 2, pp. 231-246.

[30] Poitras, G. (2006). “More on the correct use of omnibus tests for normality”. Economic Letters, vol. 90, pp. 304-309.

[31] Quessy, J. F.; Mailhot, M. (2011). “Asymptotic power of tests of normality under local alternatives. Journal of Statistical” Planning and Inference, vol. 141, nro. 8, pp. 2787-2802.

[32] 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, vol. 2, 1, pp. 21-33.

[33] Razali, N. M.; Shamsudin, N. R.; Azid, N. N. N.; Hadi, A. A.; Ismail, A. (2012): “A comparison of normality tests using SPSS, SAS and MINITAB: An application to Health Related Quality of Life data”. Statistics in Science, Business, and Engineering (ICSSBE), International Conference on (pp. 1-6), IEEE.

[34] Rahman, M. M., Govindarajulu, Z. (1997): “A modification of the test of Shapiro and Wilk for normality”. Journal of Applied Statistics, vol. 24, 2, pp. 219-236.

[35] 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, vol. 80, 5, pp. 545-591.

[36] Royston, J. P. (1982): “An extension of Shapiro and Wilk's W test for normality to large samples”. Applied Statistics, pp. 115-124.

[37] Royston, P. (1995): “Remark AS R94: A Remark on Algortihm AS181: The W-test for Normality”. Journal of the Royal Statistical, vol. 44, 4, pp. 547-551.

[38] Sarkadi, K. (1975): “The consistency of the Shapiro-Francia test”. Biometrika, vol. 62, 2, pp. 445-450.

[39] Stephens, M. A.; Hartley, H. O. (1972). Biometrika Tables for Statisticians, 2, New York: Cambridge University Press.

[40] Seier, E. (2002): “Comparison of Tests for Univariate Normality”. InterStat Statistical Journal, vol.1, pp.1-17.

[41] Shapiro, S. S., Wilk, M. B.; Chen, H. J. (1968): “A Comparative Study of Various Tests of Normality”. Journal of American Statist. Assoc., vol. 63, pp. 1343-1372.

[42] Shapiro, S.S.; Wilk, M.B. (1965): “An analysis of variance test for normality: complete samples”. Biometrika, vol. 52, pp. 591-611.

[43] Smirnov, N. V. (1936): “Sui la distributtion de w2 (Criterium de M.R.v Mises)”. Comptes Rendus (Paris), vol. 202, pp. 449-452.

[44] Shapiro, S.S.; Francia, R. (1972): “An approximation analysis of variance test for normality”. Journal of the American Statistical Association, vol. 67, pp. 215-216.

[45] Tanveer-Ul-Islam (2011): “Normality testing-A new direction”. International Journal of Business and Social Science, vol. 2, 3, pp. 115-118.

[46] Tervonen, T., & Figueira, J. R. (2008): “A survey on stochastic multicriteria acceptability analysis methods”. Journal of Multi Criteria Decision Analysis,15 (12), 1-14.

[47] Thadewald, T.; Büning, H. (2007): “Jarque-Bera test and its competitors for testing normality-a power comparison”. Journal of Applied Statistics, vol. 34, 1, pp. 87-105.

[48] Urzúa, C. M. (2007): “Portable and powerful tests for normality”. Tecnológico de Monterrey, Campus Ciudad de México.

[49] Von Mises, R. (1931): “Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik (vol. 1)”. F. Deuticke, Leipzig.

[50] Yazici, B.; Yolacan, S. (2007): “A Comparison of Various Tests of Normality”. Journal of Statistical Computation and Simulation, vol. 77, 2, pp. 175-183.

[51] Yap, B. W.; Sim, C. H. (2011): “Comparisons of various types of normality tests”. Journal of Statistical Computation and Simulation, vol. 81, 12, pp. 2141-2155.

[52] Zhang, J., Wu, Y. (2005): “Likelihood-ratio tests for normality”. Computational statistics & data analysis, vol. 49, 3, pp. 709-721.