PRUEBAS PARA LA VERIFICACIÓN DE NORMALIDAD: DETERMINACIÓN DE POTENCIAS CON MUESTRAS PEQUEÑAS

GABRIELA PILAR CABRERA
gabriela.pilar.cabrera@gmail.com
Facultad de Ciencias Exactas Físicas y Naturales. Universidad Nacional de Córdoba
Ciudad Universitaria. Córdoba. Argentina.

JOSÉ LUIS ZANAZZI
jl.zanazzi@gmail.com
Facultad de Ciencias Exactas Físicas y Naturales. Universidad Nacional de Córdoba
Ciudad Universitaria. Córdoba. Argentina.

LAURA BOAGLIO
lauraboaglio@gmail.com
Facultad de Ciencias Económicas. Universidad Nacional de Córdoba
Ciudad Universitaria. Córdoba. Argentina.

JOSE FRANCISCO ZANAZZI
jfzanazzi@gmail.com
Facultad de Ciencias Económicas. Universidad Nacional de Córdoba
Ciudad Universitaria. Córdoba. Argentina.

RESUMEN

La verificación de que los datos observados sobre un fenómeno aleatorio, pueden suponerse extraídos de una Distribución de Probabilidad Normal, es necesaria en el campo de la Investigación Operativa, 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 the field of Operational Research 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.

KEY WORDS

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

1. INTRODUCCIÓN

Muchas herramientas de Investigación Operativa y otras áreas del conocimiento asociadas, suelen requerir la verificación del supuesto de normalidad de los datos. En efecto, al estudiar políticas de inventarios, realizar simulaciones, o analizar fenómenos de espera que no responden al proceso Poisson, es frecuente que la distribución de Gauss se encuentre considerada entre los supuestos básicos.

También es común que en estas situaciones se tropiece con 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.

Otra situación similar se encuentra en algunos métodos multicriterio orientados a la toma de decisiones en pequeños grupos. Por ejemplo, los modelos SMAA (“Stochastic Multicriteria Acceptability Analysis”), admiten que las utilidades asignadas a los elementos de decisión comparados, pueden tener distribución gaussiana (Tervonen y Figueira, 2008).

Del mismo modo, el método denominado Procesos DRV (Decisión con Reducción de Variabilidad), supone que cuando los integrantes de un grupo de trabajo alcanzan un cierto nivel de consenso, en torno a los criterios a utilizar o a las utilidades asignadas a los elementos comparados, al valorar de manera independiente los elementos, la distribución resultante es la Normal (Zanazzi y Gomes, 2009). Con esta idea, la verificación del comportamiento gaussiano permite evaluar el nivel de acuerdo alcanzado por el grupo.

Ahora bien, en Estadística, 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.

Por otro lado, se denomina Potencia de la Prueba al complemento de la probabilidad del segundo tipo de error. En términos de las pruebas de normalidad, es la posibilidad que ante datos que provienen de poblaciones no normales, la prueba detecte esta cuestión (Montgomery y Runger, 2010)

La literatura especializada en Estadística ofrece una importante cantidad de aportes desarrollados en este sentido (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; Romão, Delgado y Costa, 2010; Yap y Sim, 2011; Razali, Shamsudin, Azid, Hadi & Ismail, 2012; Lafaye de Micheaux & Tran, 2014). Como es obvio, al seleccionar la prueba a utilizar, es importante que el especialista considere la potencia de la misma.

Algunos artículos han aportado estudios sobre la mayoría de las pruebas en cuestión, que incluyen la determinación de potencia. Lamentablemente, estos trabajos operan generalmente con muestras de veinticinco, cincuenta o más datos.

Por ese motivo, en este documento se realiza una selección de pruebas que pueden resultar apropiadas para verificar normalidad, cuando las muestras son pequeñas. Para las herramientas seleccionadas, se estima la potencia en situaciones donde la distribución original es simétrica, lo que puede considerarse como una de las peores condiciones. 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 estos tests.

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.

2. REVISIÓN BIBLIOGRÁFICA

En la literatura estadística se cuenta con al menos cuarenta pruebas que permiten evaluar el supuesto de normalidad (Dufour *et al.* 1998). La correcta aplicación de los métodos estadísticos paramétricos, requiere necesariamente del cumplimiento de dicho supuesto; es por ello que son muchos los expertos en estadística que se esfuerzan en la modificación y/o mejora de las pruebas de normalidad, así como del desarrollo de nuevas herramientas.

El interés por contar con métodos que permitan detectar desviaciones respecto a la Distribución Normal de probabilidades, se inicia con Pearson en 1895, con el estudio de los coeficientes de asimetría y curtosis (Razali y Wah, 2011). Actualmente se observa un creciente interés por estudiar el rendimiento de estos tests para diferentes alternativas.

En la actualidad, se entiende que no existe una única prueba de normalidad que merezca ser la 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.

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 encuentra 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 en Yazici y Yolacan (2007) y la sugerida en Romăo *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.

En tanto, ante distribuciones alternativas no normales y simétricas, aumenta sensiblemente la posibilidad de cometer el denominado Error Tipo II (Coin, 2007).

3. SUPUESTOS Y METODOLOGÍA ADOPTADA

En este trabajo se utiliza como hipótesis nula el supuesto de normalidad y como alternativa, 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 es simétrica. Dicho de otro modo, parece razonable que cualquier prueba obtenga mejores resultados cuando la distribución verdadera es notoriamente diferente de la gaussiana, como por ejemplo la Exponencial, y que en cambio evidencie dificultades cuando el comportamiento original es simétrico (Coin, 2007).

De todas formas, corresponde reconocer que distinguir entre la Normal y la Uniforme, es conveniente en una gran cantidad de situaciones prácticas. Por ejemplo en la producción, una característica de calidad de producto con distribución rectangular, se considera como evidencia de falta de control sobre el proceso productivo, donde en todo caso se realiza una selección posterior para descartar los artículos que no cumplen las especificaciones técnicas. En cambio, la normalidad se considera evidencia de control sobre el proceso (Montgomery, 1991).

Del mismo modo, si en un proceso de mantenimiento de máquinas o herramientas, se estudia la variabilidad del tiempo necesario para realizar una reparación, el hecho de que esta variable pueda suponerse extraída de una Normal, sugiere que se ha trabajado suficientemente sobre la tarea para hacerla predecible. En cambio, la distribución Uniforme resulta esperable cuando la duración de la actividad puede tanto ser muy breve como muy extensa. Se lo considera evidencia de falta de control.

En un problema multicriterio de toma de decisiones en grupo, donde los integrantes asignan utilidades de manera independiente, a una cierta cantidad de elementos comparados, es razonable esperar que cuando se alcanza un cierto nivel de consenso, cada persona realice asignaciones similares a las de sus compañeros y que la distribución resultante sea la gaussiana. Por ejemplo, si al valorar la prioridad de atención de una falla en una maquinaria, los miembros del equipo de trabajo efectúan valoraciones muy disímiles, esto puede considerarse como evidencia de falta de acuerdos básicos y por supuesto, es razonable que estas personas deban enfrentar situaciones de conflicto interno (Zanazzi, Gomes y Dimitroff, 2014; Zanazzi y Dimitroff, 2013).

Respecto a la modalidad de trabajo, es posible distinguir dos etapas bien diferenciadas:

a) Revisión de bibliografía y selección de pruebas convenientes. 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: lo cual implica disponibilidad de software o posibilidad para su implementación con hojas de cálculo.

b) Experimentación con las pruebas seleccionadas. Para determinar la potencia de las pruebas seleccionadas, para los tamaños de muestra diez y quince, se realizan experimentos de generación aleatoria. Esto es, se generan mil conjuntos de números con distribución rectangular. A continuación, se aplica cada una de las pruebas elegidas y se determina la proporción de veces que el test detecta que en realidad la distribución no es normal.

4. RESULTADOS OBTENIDOS

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

4.1. Pruebas basadas en medidas de los momentos

Estas pruebas se sostienen en el reconocimiento de las desviaciones de la normalidad mediante el análisis de algunos momentos de la muestra disponible. 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 que analizan los momentos

Estrategia	Nro	Nombre
Pruebas basadas en Momentos	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)*

4.2. Pruebas basadas en la Función de Distribución Empírica

En estas pruebas, 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 siguiente Tabla se enumeran estas aproximaciones.

TABLA 2. Pruebas que analizan la Distribución Empírica

Estrategia	**Nro**	**Nombre**
Pruebas basadas en la función de distribución empírica	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)*

4.3. Pruebas de correlación y regresión

Las pruebas de correlación y regresión, 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 siguiente lista estas propuestas.

TABLA 3. Pruebas de correlación y regresión

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

4.4. Otras pruebas

Se adopta esta denominación, porque los Tests agrupados en este apartado, no pueden ser encuadrados en los grupos anteriores. La siguiente Tabla identifica a estas aproximaciones.

TABLA 4. Pruebas no encuadradas en las estrategias anteriores

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

4.4. Potencia de las pruebas

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 Romăo *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 tamaños de muestra tamaño 25

Prueba de Normalidad	α=0,05	α=0,10
Kolmogorov-Smirnov (K-S)	39,3	46,3
Anderson-Darling (AD)*	45,1	51,8
Zhang y Wu (Z_C)	44,4	52
Zhang y Wu (Z_A)	43,6	50,8
Glen-Leemis-Barr (P_S)	45,2	51,9
D´Agostino-Pearson (K^2)	41,5	49,8
Hosking (T_{lmom})	47	53,6
Hosking (T_w)	45,3	52,3
Shapiro-Wilk (W)	45,5	52,6
Shapiro-Francia (W´)	43,5	50,4
Shaprio-Wilk modificado por Rahman y Govindaraju ($\tilde{W}$)	44,4	51,3
D´Agositno (Dag)	40	46,1
Fillibien ®	43,1	49,9
Chen-Shapiro (CS)	45,6	52,7
Barrio-Cuesta-Albertos-Matrán-Rodriguez (BCMR)	45	52,2
Coin (β_3^2)	48,5	55,6
Gel-Miao-Gastwirth (R)	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 siguiente Tabla resume la disponibilidad de cada una de estas pruebas, en las herramientas computacionales mencionadas.

TABLA 6. Disponibilidad de pruebas de normalidad en productos computacionales

Prueba de Normalidad	Spss 19	Infostat 2014	Stata 11	Minitab 17
Kolmogorov-Smirnov		x		
Kolmogorov-Smirnov modificado por Lilliefors	x	x		
Anderson-Darling (AD)*				x
D´Agostino-Pearson (según D´Agostino et al. 1990)			x	
Shapiro-Wilk	x		x	
Shapiro-Francia			x	
Shaprio-Wilk modificado por Rahman y Govindarajulu		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 tamaños de muestra de diez y quince datos, se reproducen a en la Tabla 7.

De dicha Tabla 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, se evidencia que en ningún caso es posible alcanzar una potencia superior al cincuenta por ciento.

TABLA 7. Potencia empírica obtenida mediante simulación con los productos computacionales indicados, α = 0,10, n=10 y n=15

Pruebas de Normalidad	Potencia empírica		Productos computacionales
	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.

5. 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 tests 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. Sin embargo, su aplicación con pocos datos debería ser evitada o considerada solo como un indicio.

Por último, los resultados obtenidos evidencian que el problema de verificar la normalidad en muestras pequeñas, cuando la hipótesis alternativa es una distribución simétrica, no puede considerarse resuelto. Por el contrario, debido a los múltiples y frecuentes requerimientos en ese sentido, se hace necesario desarrollar un estadístico de prueba que sea lo suficientemente potente para detectar desviaciones de la normalidad.

6. BIBLIOGRAFÍA

ANDERSON, T. W. y DARLING, D. A. (1954): “A Test of Goodness of Fit”. Journal of Statistical Associattion, vol. 49, 268, pp. 765-769.

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.

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.

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.

CRAMÉR, H. (1928): “On the composition of elementary errors: First paper: Mathematical deductions”. Scandinavian Actuarial Journal, vol. 1, pp. 13-74.

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.

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.

D'AGOSTINO, R. B.; STEPHENS, M. A.; D'AGOSTINO, R. B.; STEPHENS, M. A. (1986). Goodness-fo-fit-techniques. Statistics.

DARLING, D. A. (1957): “The Kolmogorov-Smirnov, Cramer-von Mises tests”. The Annals of Mathematical Statistics, pp. 823-838.

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, vol. 1, 1, pp. 154-173.

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.

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.

FILLIBEN, J. J. (1975): “The probability plot correlation coefficient test for normality”, Technometrics, vol. 17, 1, pp. 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.; GASTWIRTH, J. L. (2008): “A robust modification of the Jarque–Bera test of normality”. Econom. Lett, vol. 99, 1, pp. 30-32.

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.

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.

HENDERSON, A. R. (2006): “Testing experimental data for univariate normality, *Clinica chimica acta*, *366*(1), pp. 112-129.

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.

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

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.

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.

MONTGOMERY, D. (1991): “Control Estadístico de la Calidad”. México, Iberoamérica. 17-44.

MONTGOMERY, D. C., & RUNGER, G. C. (2010): Applied statistics and probability for engineers. John Wiley & Sons.

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.

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.

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

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.

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.

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.

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.

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.

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

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.

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

STEPHENS, M. A.; HARTLEY, H. O. (1972). Biometrika Tables for Statisticians, 2, New York: Cambridge University Press.

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

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.

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

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

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.

TANVEER-UL-ISLAM (2011): “Normality testing-A new direction”. International Journal of Business and Social Science, vol. 2, 3, pp. 115-118.

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

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.

URZÚA, C. M. (2007): Portable and powerful tests for normality. Tecnológico de Monterrey, Campus Ciudad de México.

VON MISES, R. (1931): Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik (vol. 1). F. Deuticke, Leipzig.

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

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.

ZANAZZI J.; DIMITROFF M (2013): “Métodos para Tomar Decisiones en Grupo. Comparación entre Procesos DRV y SMAA”. EPIO vol. 34, pp. 45-61.

ZANAZZI, J.; GOMES, L. (2009): “La Búsqueda de Acuerdos en Equipos de Trabajo: El Método Decisión con Reducción de la Variabilidad (DRV)”. Revista Pesquisa Operacional, vol. 29, pp. 195-221.

ZANAZZI, J. L., GOMES, L. F. A. M., DIMITROFF, M. (2014). “Group decision making applied to preventive maintenance systems”. Pesquisa Operacional, vol. 34, nro. 1, pp. 91-105.

ZHANG, J., WU, Y. (2005): “Likelihood-ratio tests for normality”. Computational statistics & data analysis, vol. 49, nro. 3, pp. 709-721.