



## Potencia y Robustez en Pruebas de Normalidad con Simulación Montecarlo

**Autora:** Elizabeth María Gandica de Roa  
Universidad Nacional Experimental del Táchira, **UNET**  
[elizabethgandica@gmail.com](mailto:elizabethgandica@gmail.com)  
San Cristóbal, Venezuela  
<https://orcid.org/0000-0001-8797-4198>

### Resumen

En esta investigación se planteó como objetivo general, examinar la potencia y robustez de las pruebas de normalidad en muestras grandes y pequeñas, generadas con simulación Montecarlo. Se aplicaron pruebas de hipótesis no paramétricas que miden el grado de discrepancia entre las distribuciones empíricas y la función de distribución acumulada normal, que analizan la correlación entre la distribución teórica y la experimental y las que se sustentan en el estudio de la asimetría y curtosis. La comparación se hizo en dos grupos con tamaño de muestras distintas. En las muestras grandes se compararon las pruebas de Kolmogorov-Smirnov; Chi-Cuadrado de Pearson; Jarque-Bera y Geary; en las muestras pequeñas Shapiro-Wilk; Cramér-von Mises; Lilliefors y Watson. Los contrastes se realizaron con el Programa informático RStudio y el criterio de rechazo para las hipótesis nulas se hizo a través del p-value. Como conclusión, la prueba de mayor robustez en muestras grandes es Kolmogorov estimándose que su probabilidad es menor a 0,11. En muestras pequeñas este resultado corresponde a Shapiro-Wilk con una estimación menor a 0,14. Con relación a la potencia en las pruebas de normalidad para muestras grandes se demostró que la más potente de ellas es la prueba Jarque Bera, con un intervalo de confianza entre 0,86 y 1. Para las muestras pequeñas ninguna de las pruebas sometidas a estudio resultó potente.

**Palabras clave:** simulación Montecarlo; potencia; robustez; pruebas de normalidad.

#### Cómo citar este artículo:

Gandica de Roa, E. (2020). **Potencia y Robustez en Pruebas de Normalidad con Simulación Montecarlo.** *Revista Scientific*, 5(18), 108-119, e-ISSN: 2542-2987. Recuperado de: <https://doi.org/10.29394/Scientific.issn.2542-2987.2020.5.18.5.108-119>

**Fecha de Recepción:**  
20-05-2020

**Fecha de Aceptación:**  
25-09-2020

**Fecha de Publicación:**  
05-11-2020



## Power and Robustness in Normality Tests with Montecarlo Simulation

### Abstract

The general objective of this research was to examine the power and robustness of normality tests in large and small samples, generated with Monte Carlo simulation. Non-parametric hypothesis tests were applied to measure the degree of discrepancy between the empirical distributions and the normal cumulative distribution function, which analyze the correlation between the theoretical and experimental distribution and those that are supported by the study of asymmetry and kurtosis. The comparison was made in two groups with different sample sizes. In large samples, the Kolmogorov-Smirnov tests were compared; Pearson's Chi-Square; Jarque-Bera and Geary; in the small samples Shapiro-Wilk; Cramér-von Mises; Lilliefors and Watson. The contrasts were carried out with the RStudio computer program and the rejection criterion for the null hypotheses was made through the p-value. In conclusion, the most robust test in large samples is Kolmogorov, estimating that its probability is less than 0.11. In small samples this result corresponds to Shapiro-Wilk with an estimate of less than 0.14. Regarding the power in the normality tests for large samples, it was shown that the most powerful of them is the Jarque Bera test, with a confidence interval between 0.86 and 1. For small samples, none of the tests under study it turned out powerful.

**Keywords:** Montecarlo simulation; power; sturdiness; normality tests.

#### How to cite this article:

Gandica de Roa, E. (2020). **Power and Robustness in Normality Tests with Montecarlo Simulation.** *Revista Científica*, 5(18), 108-119, e-ISSN: 2542-2987. Recovered from: <https://doi.org/10.29394/Scientific.issn.2542-2987.2020.5.18.5.108-119>

**Date Received:**  
20-05-2020

**Date Acceptance:**  
25-09-2020

**Date Publication:**  
05-11-2020



## 1. Introducción

El éxito de todo proceso investigativo supone el seguimiento y cumplimiento de todas las fases del mismo. Comprobar los supuestos teóricos requeridos en cada fase se torna para muchos un obstáculo, y algunas veces es tan engorroso que se pasan por alto detalles, que repercuten en las conclusiones y por ende en la toma de decisiones. En el proceso investigativo los procedimientos deben seguirse paso a paso. Frente a esto, Hurtado (2020): sostiene que este seguimiento riguroso generará conclusiones con estándares de validez y confiabilidad altos, para dar lugar a nuevos conocimientos.

En concordancia con esta aseveración, se encuentra que cuando se recopilan datos en el ámbito investigativo de cualquier área de conocimiento, es frecuente la necesidad de verificar si estos pueden considerarse provenientes de una población con Distribución Normal, lo que sugiere que los datos tienen medidas de media, mediana y moda similares. Con esta finalidad, en la literatura estadística se presenta una variedad de procedimientos denominados Pruebas de Hipótesis, sustentadas en la Estadística Inferencial, con las que se contrasta la suposición de normalidad de los datos frente a la no normalidad. Sin embargo, no hay evidencia de cual prueba es más conveniente, tal como afirma Islam (2011): no existe una aproximación que supere con holgura a las restantes, es decir, no se ha podido demostrar cuál de las pruebas es superior en robustez y en potencia.

Es importante chequear el supuesto de normalidad en los datos, pues este se requiere para tomar la decisión en cuanto a la prueba que debe ser utilizada. Si los datos son normales se seguirá el camino de pruebas de hipótesis paramétricas, de lo contrario se utilizarán pruebas no paramétricas. En la presente investigación se comparan diferentes pruebas para determinar la potencia y robustez en cada una de ellas, propiedades deseables en cualquier método estadístico.

Cuando se aplican pruebas de hipótesis, se deben tener presentes dos



tipos de errores, llamados Error Tipo I y Error Tipo II. El primero se comete cuando se rechaza una hipótesis nula que es correcta. Al calcular la probabilidad del error tipo I, se estará calculando la robustez del Test. El segundo error se produce cuando no se rechaza una suposición falsa. El complemento de la probabilidad de cometer el error tipo II, será considerado como la potencia del Test estudiado. En términos de normalidad, que es lo que corresponde a este estudio, la hipótesis nula será definida como la afirmación de que los datos de la población son distribuidos normales.

De acuerdo con Arshad, Rasool y Ahmad (2003): proponen cuatro grupos de clasificación para las pruebas de normalidad: en el primero se encuentran aquellas pruebas que miden el grado de diferencia entre las frecuencias observadas y las frecuencias esperadas del modelo asumido; el segundo grupo corresponde a las pruebas que miden la correlación entre las frecuencias observadas y las frecuencias esperadas; en el tercer grupo se encuentran los métodos que miden la asimetría y la curtosis de los datos sometidos a prueba; y en un cuarto grupo están las pruebas especiales que no entran en ninguno de los grupos anteriores.

## 2. Metodología

Para la generación de muestras aleatorias se utilizó simulación Montecarlo, el cual es un método estadístico numérico que permite la generación de una gran cantidad de variables aleatorias para llegar a resultados parecidos a los reales. En esta investigación se generaron muestras aleatorias de tamaño 30 y de tamaño 1000. Para las muestras pequeñas  $n=30$  se estudiaron las pruebas de normalidad de Cramér-Von Mises, Lilliefors y Watson del grupo 1 y Shapiro-Wilk y Durbin-Watson del grupo 2, y en las muestras grandes  $n=1000$ , Kolmogorov-Smirnov, Chi cuadrado de Pearson y Anderson-Darling del grupo 1, Geary del grupo 2 y la prueba Jarque-Bera del grupo.



Para determinar la longitud de la simulación se estableció un nivel de confianza del 95% y un error para la estimación del 10%. La longitud de la simulación se calculó a través de la siguiente formula:

$$n \geq \frac{Z^2 * P * Q}{E^2} = 96 \text{ corridas}$$

Donde Z= 1,96, E= 0,1 y P=Q=0,5.

Con este resultado se efectuaron 100 corridas o muestras en cada grupo, el estudio se llevó a cabo en tres fases:

En la fase 1, se calculó la robustez aplicando las pruebas de normalidad a 100 muestras normales, de tamaño  $n=1000$  y  $n=30$ . Para ello se utilizó el programa estadístico RStudio para la aplicación de los métodos. La hipótesis nula se estableció con la suposición de normalidad y se aplicó como criterio de decisión el p-valor con un nivel de significancia de 0,05.

En esta fase se valoró la probabilidad de cometer el error tipo I, es decir, la probabilidad de rechazar la hipótesis nula. Para determinar la robustez de la prueba el criterio liberal de Bradley (1978): establece que si la probabilidad de cometer el error tipo I ( $\hat{\alpha}$ ) se encuentra en el intervalo  $0,5\alpha \leq \hat{\alpha} \leq 1,5\alpha$ , el test será considerado robusto. Por consiguiente, para el nivel de significación establecido en esta investigación ( $\alpha = 0.05$ ) el intervalo utilizado para definir la robustez de las pruebas fue:  $0,025 \leq \hat{\alpha} \leq 0,075$ .

En la fase 2, se calculó la potencia, aplicando las pruebas de normalidad a 100 muestras no normales de tamaño  $n=1000$  y  $n=30$ . Se probaron 25 Exponenciales, 25 Weibull, 25 Chi cuadrado y 25 Uniformes, con parámetros distintos para cada una. Se evaluó la probabilidad de cometer el error tipo II y se dedujo su complemento, es decir, se midió la probabilidad de rechazar la hipótesis de suposición de normalidad. Para determinar si la prueba es potente se aplicó el criterio expuesto por Cabrera, Zanazzi, Zanazzi y Boaglio (2017): el cual establece que, si la probabilidad de rechazar la hipótesis nula supera

el 50%, se puede concluir que la prueba es potente, por lo que, si se aumenta el nivel de significancia, se incrementará la potencia de la prueba.

En la fase 3, se realizaron contrastes de proporciones entre las pruebas que resultaron potentes y robustas en las fases anteriores, esto con el fin de determinar la significancia estadística de las diferencias en las probabilidades encontradas.

### 3. Resultados

#### 3.1. Fase 1

En la tabla 1 y tabla 2 se muestran los resultados de la simulación para 100 muestras de tamaño 1000 y 30, a las que se les aplicó las pruebas de normalidad escogidas. El rechazo de la hipótesis nula se realizó comparando el p-valor, con el nivel de significancia. Se sumaron las veces que la prueba dio rechazo y se calculó la tasa de error empírica ( $\hat{\alpha}$ ), para luego compararla con el intervalo de robustez establecido. Como resultado se obtuvo que para muestras grandes las pruebas robustas son: Kolmogorov-Smirnov, Geary y Jarque-Bera y en las muestras pequeñas Shapiro-Wilk, Cramér-Von Mises y Lilliefors.

**Tabla 1.** Robustez en muestras grandes.

ROBUSTEZ n=30 $\alpha=0.05$					
Método	Anderson-Dalín	X <sup>2</sup> Pearson	Kolmogorov	Geary	Jarque-Bera
$\hat{\alpha}$	0,08	0,11	0,01	0,06	0,05
Robustez	NO	No	SI	SI	SI

Fuente: La Autora (2020).

**Tabla 2.** Robustez en muestras pequeñas.

ROBUSTEZ n=1000 $\alpha=0.05$				
Criterio de Robustez $0,025 \leq \hat{\alpha} \leq 0,075$				
Método	Shapiro-Wilk	Cramér-von Mises	Lilliefors	Watson
$\hat{\alpha}$	0,04	0,03	0,03	0,38
Robustez	SI	SI	SI	NO

Fuente: La Autora (2020).

### 3.2. Fase 2

La tabla 3 y tabla 4, muestran los resultados de las potencias en las pruebas aplicadas a 100 muestras no normales de tamaños 1000 y 30. Para muestras grandes las pruebas que resultaron potentes fueron: Anderson-Darling, Chi cuadrado; Kolmogorov-Smirnov y Jarque-Bera. En las muestras pequeñas ninguna prueba dio potente, pues ninguna superó el 50% del criterio establecido.

**Tabla 3.** Potencias en muestras grandes.

POTENCIA EMPÍRICA n=1000 $\alpha=0.05$					
Criterio de la Potencia $1-\beta \geq 0,5$					
Método	Anderson-Darling	X <sup>2</sup> Pearson	Kolmogorov	Gary	Jarque-Bera
1 - $\beta$	0,87	0,74	0,55	0,33	0,96
Potencia	SI	SI	SI	NO	SI

Fuente: La Autora (2020).

**Tabla 4.** Potencia en muestras pequeñas.

POTENCIA EMPÍRICA n=30 $\alpha=0.05$				
Criterio de la Potencia $1-\beta \geq 0,5$				
Método	Shapiro-Wilk	Cramér-von Mises	Lilliefors	Watson
1 - $\beta$	0,43	0,32	0,30	0,27
Potencia	NO	NO	NO	NO

Fuente: La Autora (2020).

En la tabla 5, se muestran los intervalos de confianza para los valores encontrados de robustez y potencia en las pruebas analizadas. Esta estimación está calculada con un error del 10% y un nivel de significancia del 5%.

**Tabla 5.** Estimación de las probabilidades de robustez y potencia.

Estimación de la potencia y robustez de los métodos				
Prueba	Robustez		Potencia	
	Li	Ls	Li	Ls
Anderson-Darling	0	0,18	0,77	0,97
X <sup>2</sup> Pearson	0	0,21	0,64	0,84
Kolmogorov	0	0,11	0,45	0,65
Geary	0	0,16	0,23	0,43
Jarque-Bera	0	0,15	0,86	1



Artículo Original / Original Article

<b>Shapiro-Wilk</b>	0	0,14	0,33	0,53
<b>Cramér-von Mises</b>	0	0,13	0,22	0,42
<b>Lilliefors</b>	0	0,13	0,2	0,4
<b>Watson</b>	0,21	0,48	0,17	0,37

Fuente: La Autora (2020).

### 3.3. Fase 3

En esta fase se compararon las pruebas que resultaron potentes y robustas en las fases anteriores, con el fin de determinar que prueba es la más robusta y la más potente a un nivel de significancia del 5%. Se realizó un contraste de hipótesis de proporciones. La hipótesis nula para la robustez se planteó de la siguiente manera:

**Ho:** La probabilidad de rechazo de la prueba A es menor que la probabilidad de rechazo de la prueba B.

Mientras que, para contrastar la potencia, la hipótesis nula fue:

**Ho:** La probabilidad de no rechazo de la prueba A es mayor que la probabilidad de no rechazo de la prueba B.

La tabla 6 y tabla 7 muestran los resultados del contraste para las pruebas robustas. El estadístico de prueba se comparó con  $Z_t = 1,64$ , valor de la tabla normal correspondiente a un nivel de significancia de 5% y un ensayo unilateral.

**Tabla 6.** Contraste de pruebas robustas para muestras grandes.

<b>Ho: La probabilidad de rechazo de A es menor que la probabilidad de rechazo de B</b>				
<b>Prueba A</b>	<b>Prueba B</b>	<b>Estadístico</b>	<b>Decisión</b>	<b>Conclusión</b>
<b>Kolmogorov</b>	<b>Geary</b>	-1,94	No rechazar	K es más robusto que G
<b>Kolmogorov</b>	<b>Jarque-Bera</b>	-1,67	No rechazar	k es más robusto que J
<b>Geary</b>	<b>Jarque-Bera</b>	0,31	No rechazar	G es más robusto que J

Fuente: La Autora (2020).

La tabla 7 muestra como resultado, que la prueba Kolmogorov es la más robusta para muestras grandes. Seguida de la prueba de Geary y Jarque-Bera.



**Tabla 7.** Contraste de pruebas robustas para muestras pequeñas.

CONTRASTE PARA LAS PRUEBAS ROBUSTAS-MUESTRAS PEQUEÑAS				
Ho: La probabilidad de rechazo de A es menor que la probabilidad de rechazo de B				
Prueba A	Prueba B	Estadístico	Decisión	Conclusión
Shapiro-Wilk	Cramér-von Mises	0,38	No rechazar	S es más robusto que C
Shapiro-Wilk	Lilliefors	0,38	No rechazar	S es más robusto que L
Cramér-von Mises	Lilliefors	0	No rechazar	C es más robusto que L

**Fuente:** La Autora (2020).

En las muestras pequeñas la prueba que resultó más robusta es Shapiro-Wilk seguida de Cramér-von Mises y Lilliefors. En la tabla 8 se presentan los resultados del contraste para las pruebas potentes. El estadístico de prueba se comparó con  $Z_t = -1,64$ , valor de la tabla normal correspondiente a un nivel de significancia de 0,05 y un ensayo unilateral.

**Tabla 8.** Contraste de pruebas potentes para muestras grandes.

CONTRASTE PARA LAS PRUEBAS POTENTES-MUESTRAS GRANDES				
Ho: La probabilidad de rechazo de A es mayor que la probabilidad de rechazo de B				
Prueba A	Prueba B	Estadístico	Decisión	Conclusión
Anderson-Darling	Chi Cuadrado	2,35	No rechazar	A es más potente que C
Anderson-Darling	Kolmogorov	5,33	No rechazar	A es más potente que K
Anderson-Darling	Jarque-Bera	-2,31	Rechazar	A no es más potente que J
Chi Cuadrado	Kolmogorov	2,86	No rechazar	C es más potente que K
Chi Cuadrado	Jarque-Bera	-4,58	Rechazar	C no es más potente que J
Kolmogorov	Jarque-Bera	-7,67	Rechazar	K no es más potente que J

**Fuente:** La Autora (2020).

Para las muestras grandes se pudo determinar que la prueba más potente es la de Jarque-Bera.

#### 4. Conclusiones

Los resultados arrojan que en las muestras grandes la prueba de Kolmogorov es la más robusta estimándose que su probabilidad de robustez es menor a 0,11. En muestras pequeñas este resultado corresponde a: Shapiro-Wilk, Cramér-von Mises, Lilliefors y Watson, resultando como prueba más robusta Shapiro-Wilk, con una estimación menor a 0,14. Con relación a la potencia en las pruebas de normalidad para muestras grandes, se encontró



que las pruebas de Anderson-Darling, Chi cuadrado de Pearson, Kolmogorov-Smirnov y Jarque-Bera alcanzaron un porcentaje superior al 50%, y se demostró que la más potente de ellas es la prueba de Jarque-Bera, con un intervalo de confianza entre 0,86 y 1. Para las muestras pequeñas ninguna de las pruebas sometidas a estudio resultó potente.

Cabe mencionar como resultado paralelo, que en la simulación para la potencia de las pruebas en muestras grandes, se pudo observar que cuando la distribución sometida al contraste es simétrica, como es el caso de la distribución uniforme, la potencia de la prueba tiende a disminuir, mientras que en el caso de distribuciones no simétricas como la exponencial, la potencia aumenta.

Para muestras pequeñas el resultado es menos alentador, pues ninguna prueba resultó potente, de lo cual se pudiera inferir que las pruebas sometidas a estudio en esta investigación son sensibles a la simetría de la distribución y al tamaño de la muestra, esto evidencia que cuando se desea verificar normalidad en muestras pequeñas, se debe tener especial cuidado y más atención.

En las pruebas para muestras grandes si se desea una prueba más potente que robusta, se recomienda utilizar la prueba Jarque-Bera. Si por el contrario se desea una prueba más robusta que potente utilizar Kolmogorov-Smirnov, esto quedara a criterio del investigador y de las características del contraste.

Para muestras pequeñas la prueba recomendada es la de Shapiro-Wilk, que resultó ser una prueba robusta y aunque no alcanzo ser potente en la muestra estudiada, la estimación arrojó un límite superior de 53%, lo que permite inferir que en algunos casos con un nivel de confianza del 95% la prueba pudiera considerarse potente.

Finalmente, la continuidad de este tipo de estudios con respecto a las pruebas de normalidad genera grandes aportes a la investigación, debido a



que esta verificación siempre será necesaria para poder determinar los métodos estadísticos a ser aplicados y sostener la solidez de las inferencias realizadas.

## 5. Referencias

- Arshad, M., Rasool, M., & Ahmad, M. (2003). **Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution**. *Journal of Applied Sciences*, 3(2), 85-88, e-ISSN: 1812-5662. Recovered from: <https://dx.doi.org/10.3923/jas.2003.85.88>
- Bradley, J. (1978). **Robustness?**. *British Journal of Mathematical and Statistical Psychology*, 31, 144-152, e-ISSN: 0007-1102. Recovered from: <https://doi.org/10.1111/j.2044-8317.1978.tb00581.x>
- Cabrera, G., Zanazzi, J., Zanazzi, J. & Boaglio, L. (2017). **Comparación de potencias en pruebas estadísticas de normalidad, con datos escasos**. *Revista Facultad de Ciencias Exactas, Físicas y Naturales*, 4(2), 47-52, e-ISSN: 2362-2539. Recuperado de: <https://revistas.unc.edu.ar/index.php/FCEFyN/article/view/16775>
- Hurtado, F. (2020). **Fundamentos Metodológicos de la Investigación: El Génesis del Nuevo Conocimiento**. *Revista Cientific*, 5(16), 99-119, e-ISSN: 2542-2987. Recuperado de: <https://doi.org/10.29394/Scientific.issn.2542-2987.2020.5.16.5.99-119>
- Islam, T. (2011). **Normality Testing- A New Direction**. *International Journal of Business and Social Science*, 2(3), 115-118, e-ISSN: 2219-6021. Recovered from: <http://www.ijbssnet.com/journal/index/169>

**Elizabeth María Gandica de Roa**  
e-mail: [elizabethgandica@gmail.com](mailto:elizabethgandica@gmail.com)



Nacida en San Cristóbal, estado Táchira, Venezuela. Licenciada en Matemáticas de la Universidad Simón Bolívar (USB, 1991) de Caracas, Venezuela; Magister en matemáticas de la Universidad Simón Bolívar (USB, 1993); Especialista en Administración de Empresas; y Magister en Dirección de Empresas, egresada de la Universidad Católica del Uruguay “Dámaso Antonio Larrañaga” (UCU, 1998); Doctora en Educación, egresada de la Universidad Pedagógica Experimental Libertador (UPEL, 2016); Profesora Titular a dedicación exclusiva en la Universidad Nacional Experimental del Táchira (UNET); Línea de Investigación estadística y representaciones sociales; Ponente en congresos nacionales e internacionales sobre matemática, estadística, filosofía y lógica.