

Efecto del número de réplicas y número de tratamientos con distribución no normal en la potencia de pruebas de normalidad

Manrique Camacho Pochet¹, Amanda Cedeño Guzmán¹, Iván Daniel Rodríguez Cruz^{1,2} y Marie Sofia Villalobos Martínez¹

manrique.camacho@ucr.ac.cr, amanda.cedeno@ucr.ac.cr, ivan.rodriguezacruz@ucr.ac.cr y marie.villalobos@ucr.ac.cr

RESUMEN

En el contexto de la regresión lineal y los modelos experimentales, uno de los supuestos fundamentales radica en que los residuos se distribuyan de forma normal. De esta forma, se llevó a cabo una simulación con el propósito de analizar el impacto de las distribuciones asimétricas positivas en los valores residuales de un modelo experimental. Por consiguiente, se evaluó la potencia de las pruebas de bondad de ajuste para verificar el supuesto de normalidad de los residuos. Al realizar el presente análisis se emplearon tanto pruebas paramétricas como no paramétricas (Shapiro-Wilks, Jarque-Bera y Kolmogorov-Smirnov-Lilliefors, respectivamente) con distinta cantidad de réplicas y haciendo uso de diversos números de distribuciones asimétricas exponenciales con un determinado parámetro β y mediante diez mil iteraciones. Los resultados muestran que de forma íntegra la prueba de Shapiro-Wilks en los múltiples escenarios obtuvo una potencia superior a las demás. En cuanto a los tamaños muestrales, se encontró que, a mayor tamaño de muestra, mayor potencia; de la misma forma, entre más cantidad de tratamientos con distribución exponencial, se obtiene una potencia de la prueba más alta. Además, la prueba no paramétrica de Kolmogorov-Smirnov con corrección de Lilliefors es la que denotó un menor rendimiento en comparación con las otras dos pruebas a través del estudio.

PALABRAS CLAVE: Simulación, valores residuales, distribución exponencial, pruebas de bondad de ajuste.

INTRODUCCIÓN

Al momento de ajustar un modelo lineal uno de los inconvenientes que generalmente se presenta es la falta de normalidad de los residuos, es decir, ocurre un incumplimiento del supuesto de normalidad. Dicha suposición es considerada por muchos críticos en el área como una de las premisas más relevantes debido a su importancia al interpretar y hacer inferencias válidas (estimación puntual, intervalos de confianza, etc.) sobre los datos que se analizan (Thadewald y Büning, 2007).

Por consiguiente, las pruebas de bondad de ajuste han sido una herramienta primordial, junto al análisis gráfico, por ejemplo, los gráficos de *Q-Q plot* o los histogramas, que funcionan para

¹ Estudiante de Estadística de la Universidad de Costa Rica.

² Estudiante de Estadística y Sociología de la Universidad de Costa Rica.

poder contrastar la hipótesis nula de que la distribución de los residuos de los datos proviene de una distribución normal (esta hipótesis está contenida en la expresión (2)).

En consecuencia, una práctica común para verificar el cumplimiento de este supuesto es efectuar estudios de simulación con el fin de analizar cómo afecta, p. ej., el comportamiento de distintas distribuciones asimétricas positivas (como la distribución Exponencial, Gamma, Weibull, entre otras) al cálculo de la potencia de la prueba (Nosakhare y Bright, 2017). Para contrastar las hipótesis se llevarán a cabo diferentes pruebas de normalidad. Asimismo, se pueden encontrar estudios de simulación que se centran en probar si determinantes como los valores extremos (Coin, 2008; Ventura-León et al., 2022), la falta de homocedasticidad (Mak, 2000), la comparación de pruebas de bondad ajuste (paramétricas y no paramétricas) (Thadewald y Büning, 2007) o el tamaño de la muestra (Dufour et al., 1998) tienen influencia al tomar la decisión de rechazar o no la hipótesis del supuesto de normalidad de los residuos previamente mencionado.

De esta forma, para poder analizar el cumplimiento del supuesto de normalidad existen una diversidad de prueba que desde distintos marcos metodológicos (paramétricos y no paramétricos) logran, en menor o mayor medida, una respuesta para precisar si los residuos de un determinado conjunto de datos se distribuyen o no de forma normal. Entre las pruebas de bondad de ajuste más comunes se encuentran las pruebas Shapiro-Wilks (SW) (Shapiro y Wilk, 1965), la prueba de Kolmogorov-Smirnov con la corrección Lilliefors (KS_L) (Gross y Ligges, 2015) y la prueba robusta de Jarque-Bera (JB) (Jarque y Bera, 1980), entre muchas otras más. Sin embargo, las pruebas de bondad de ajuste varían en términos de las características que utilizan para analizar los conjuntos de datos. Algunas pruebas consideran la asimetría y la curtosis, mientras que otras se centran en la función de distribución o la relación lineal entre la variable y la variable normal estándar. Esta diversidad permite evaluar diferentes aspectos de la distribución (Mohd Razali y Yap, 2011, pp. 21–22).

En suma, este estudio hará hincapié en analizar el impacto de las distribuciones asimétricas positivas en los valores residuales de un modelo experimental, el cual contará con dos factores y un efecto de interacción entre ellos al evaluar la potencia de las pruebas de bondad de ajuste para verificar el supuesto de normalidad de los residuos. En concreto, se indagará cómo la distribución exponencial con parámetro β en cada tratamiento afecta la capacidad de la prueba para detectar la falta de normalidad en los residuos.

METODOLOGÍA

Para poder llevar a cabo el estudio de simulación vinculado al tema propuesto, es necesario establecer previamente una serie de conceptos básicos relacionados con los estadísticos utilizados en las pruebas de bondad de ajuste que fueron empleadas, así como otros términos relevantes para el objetivo de este trabajo. En la sección de Simulación se explicará más al por menor cómo se creó y se trabajó la simulación propuesta.

Supuesto de normalidad

En diseños experimentales un modelo lineal se puede expresar tal que

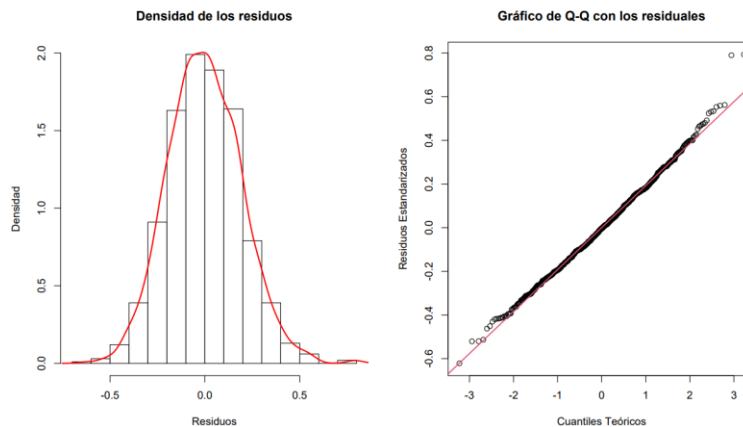
$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (1)$$

Donde $\varepsilon_{ij} \sim N(0, \sigma^2)$ y $Y_{ij} \sim N(\mu, \sigma^2)$. Es decir, la distribución condicional de la variable respuesta bajo cada tratamiento sigue una distribución normal, y todas las distribuciones condicionales de la respuesta bajo los tratamientos cuentan con la misma varianza. Cabe denotar que este supuesto se asocia teóricamente al Teorema del Límite Central.

Al tener en cuenta que ε_{ij} y Y_{ij} deben distribuirse normalmente, entonces el supuesto de normalidad de los residuos es fundamental debido a que cualquier función lineal de variables distribuidas normalmente también seguirá una distribución normal (Gujarati y Porter, 2010, p. 99). Igualmente, al suponer normalidad los estimadores de Mínimos Cuadrados Ordinarios o MCO se van a comportar como EIVM (Estimador Insegado de Varianza Mínima) y son consistentes (cuando $n \rightarrow \infty$ los estimadores cada vez se acercan más al valor verdadero del parámetro).

Figura 1

Densidad (izquierda) y gráfico Q-Q (derecha) de los residuos de un modelo lineal, estos datos se distribuyen aproximadamente normal.



Nota. Elaboración propia con datos de la base *quakes* presente en el paquete *datasets* (versión 3.6.2) en R (4.3.1).

Sin embargo, suponer normalidad de los residuos muchas veces es muy complicado debido a la naturaleza de los datos, en resultado existen ciertas pruebas de bondad de ajuste que analizan el supuesto asociado a que los residuos se distribuyen normalmente. Además, para (Dufour et al., 1998) dicho supuesto supone una seguridad estadística al momento de inferir y generar intervalos de confianza para responder a los objetivos de investigación.

Hipótesis de normalidad

Al momento de evaluar el cumplimiento de un *test* que analice el supuesto de normalidad se debe tener en cuenta que se está contrastando la siguiente hipótesis:

H_0 : Los residuos se distribuyen normal.

(2)

H_1 : Los residuos no se distribuyen normal.

Con este contraste, las pruebas de bondad de ajuste que se describirán en el siguiente apartado se centran para poder determinar un estadístico y una probabilidad asociada a él, la cual hará válida o no esta hipótesis referente a la distribución de los datos.

Pruebas de bondad de ajuste

Shapiro-Wilks

El test de normalidad de Shapiro-Wilks se emplea para evaluar la hipótesis de si una muestra específica n proviene de una distribución normal (Shapiro y Wilk, 1965). El estadístico asociado W está expresado por la siguiente fórmula:

$$W = \frac{b^2}{S^2} = \frac{\left(\sum_{i=1}^{n/2} a_i x_i^*\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

Donde a_i es una serie de pesos establecidos en el trabajo original de Shapiro-Wilks. Los pesos se calcularon a partir de los cuantiles de una distribución teórica estadística. Además, el denominador del estadístico W corresponde a la fórmula que toma la varianza, y los x_i^* , según los autores originales, deben estar ordenado de forma descendente. El test de Shapiro-Wilk es más sensible a tamaños de muestra más grandes ($3 \leq n \leq 5000$) (Mohd Razali y Yap, 2011, p. 25), pero también se comporta bien para muestras de tamaño $n < 20$ (Shapiro y Wilk, 1965, p. 602).

Jarque-Bera

La prueba de bondad de ajuste Jarque-Bera se basa en una estimación conjunta entre los coeficientes de asimetría y curtosis. Según (Thadewald y Büning, 2007) el estadístico JB está dado por la siguiente expresión:

$$JB = \frac{n}{6} \left[S^2 + \frac{(K - 3)^2}{4} \right] \quad (4)$$

donde la asimetría de la muestra utilizada está dada por $S^2 = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}$ que es un estimador de $\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$ y la curtosis de la muestra es $K = \frac{\hat{\mu}_4}{\hat{\mu}_2^2}$ corresponde a un estimador de $\beta_2 = \frac{\mu_4}{\mu_2^2}$; los valores de, donde μ_2, μ_3 y μ_4 son el segundo, tercer y cuarto momento central estimados de la siguiente forma:

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^j \quad j = 2, 3, 4.$$

Por otro lado, el estadístico $JB \sim X_{(2)}^2$ (Thadewald y Büning, 2007, p. 91). Esto implica que la hipótesis H_0 contenida en la ecuación (2) se rechaza si el estadístico JB es mayor o igual a un cuantil Chi-cuadrado con 2 grados de libertad:

$$JB \geq X_{(2,1-\alpha)}^2$$

Lilliefors (Kolmogorov-Smirnov)

El test está basado en la prueba de Kolmogorov-Smirnov, pero depende de conocer los valores de los parámetros de la distribución. Esto puede ser complejo, ya que a menudo es necesario estimar la media y la varianza a partir de una muestra. En casos así se pueden estimar desde una muestra que sea representativa y aleatoria (Mohd Razali y Yap, 2011, p. 23). De esta forma, el estadístico de prueba ($KS_L = D$) se define como la máxima diferencia absoluta entre la CDF (función de distribución acumulada) empírica y la CDF hipotética. En otras palabras, el estadístico D tiene la siguiente expresión:

$$D = \max(D^+, D^-) \quad (5)$$

Donde,

$$D^+ = \max_{i=1, \dots, n} \left[\frac{i}{n} - p_{(i)} \right] \quad \text{y} \quad D^- = \max_{i=1, \dots, n} \left[p_{(i)} - \frac{(i-1)}{n} \right]$$

En este caso $p_{(i)} = \Phi \left[\left(x_{(i)} - \frac{\bar{x}}{s} \right) \right]$. Siendo Φ la función de distribución acumulada de una distribución normal estándar en donde \bar{x} y s son la media y la varianza, respectivamente (Gross y Ligges, 2015). Además, n representa el tamaño de la muestra.

Potencia de la prueba

Según (Chou-Chen, 2023) la potencia de la prueba se refiere a la probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa es cierta. Es decir, en un espacio paramétrico Ω segmentado en dos subconjuntos Ω_0 y Ω_1 de un parámetro θ . El Error Tipo I $\alpha(\delta)$ se define tal que:

$$\alpha(\delta) = P(\text{rechazar } H_0 \text{ cuando es cierta}) = P(X \in RC_\delta | \theta \in \Omega_0) = \text{Potencia}(\theta | \theta \in \Omega_0)$$

Donde RC_δ es una zona de rechazo del contraste dado un $\theta \in \Omega_0$.

Y, por ende, también se puede definir el Error Tipo II que corresponde a no aceptar H_0 cuando en verdad es falsa. En otras palabras:

$$\beta(\delta) = P(X \in RC_\delta^c | \theta \in \Omega_0) = 1 - \text{Potencia}(\theta | \theta \in \Omega_1)$$

De esta forma, este enfoque de simulación (realizado en la versión 4.3.1 de R)³ basado en múltiples iteraciones permitió evaluar la potencia de las pruebas de normalidad en diferentes escenarios y obtener una estimación fehaciente de su desempeño. Al almacenar los resultados en vectores separados, se pudo analizar y comparar la potencia de cada prueba de manera individual en función de los diferentes casos de distribución analizados.

³ Se usaron ciertas librerías como: tidyverse (2.0.0), stats (4.4.0), tsoutliers (0.6-8), nortest (1.0-4), entre otras.

SIMULACIÓN

Primeramente, en este estudio se utilizó el método de Montecarlo para evaluar la potencia de la prueba de Shapiro-Wilks, Jarque-Bera y Kolmogorov-Smirnov con corrección de Lilliefors. Además, para poder examinar el efecto del número de réplicas y número de tratamientos con distribución no normal en la potencia de pruebas de normalidad, se debió tener en cuenta los siguientes aspectos.

Por ende, se planteó un modelo experimental con dos factores A (factor de diseño) y B (factor de no diseño) cada uno con tres niveles y un efecto de interacción entre ellos dos. Por ende, la expresión que tomó este modelo fue la siguiente:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad (6)$$

Donde,

1. μ : media general.
2. α_i : efecto del factor de diseño, $i = 1,2,3$.
3. β_j : efecto del factor de no diseño, $j = 1,2,3$.
4. $(\alpha\beta)_{ij}$: efecto de interacción entre el factor de diseño y el factor de no diseño.

Este modelo se tomó como sugerencia del modelo experimental realizado por (Rodríguez Cruz et al., 2023).

Después de plantear el modelo, es de interés analizar el efecto del número de réplicas y la cantidad de tratamientos con distribución no normal en la potencia de las pruebas de normalidad. Para ello, se generaron números aleatorios para cada tratamiento, donde la cantidad de números generados corresponde al número de réplicas especificado para ese tratamiento. Los números dentro de cada tratamiento se generaron siguiendo una distribución normal con media μ y varianza fija de 4, o una distribución exponencial con parámetro $\beta = 2$ (para poseer presencia de homocedasticidad), dependiendo del tratamiento.

Además, se realizaron diferentes casos donde se modificó de forma aleatoria la cantidad de tratamientos que en primera instancia se distribuían normal por tratamientos que se distribuían de forma exponencial, hasta llegar a un total de 9 tratamientos con distribución no normal. Los datos utilizados en el análisis se obtienen de estos casos generados aleatoriamente. Cada tratamiento que provenía de una distribución normal tenía su propio valor de μ (Rodríguez Cruz et al., 2023). Estos valores fueron los siguientes:

$$\mu_{ij} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}, \mu_{31}, \mu_{32}, \mu_{33}) = (4, 6, 2, 3, 4, 3, 6, 5, 4)$$

Seguidamente, se prefirió un enfoque balanceado en el cual todos los tratamientos cuentan con la misma cantidad de réplicas. Esto implica que se asigna la misma cantidad de observaciones a cada tratamiento, independientemente de la distribución del tratamiento. Al mantener un diseño balanceado, se busca asegurar una comparación equitativa entre los tratamientos y minimizar cualquier sesgo potencial introducido por una distribución desigual de réplicas y solo enfocarse en

el t3pico de normalidad de los residuos. En consecuencia, para analizar la influencia de los tama1os de r3plica, se tomaron en cuenta cuatro valores:⁴

$$r = 5, 10, 15, 20$$

Por lo tanto, se analizaron 36 escenarios por prueba de normalidad donde se calcul3 la potencia de la prueba con el mismo escenario para las tres (Shapiro-Wilks, Jarque-Bera y KS-Lilliefors); considere la siguiente tabla que muestra los casos.

Tabla 1

Cantidad de escenarios para las pruebas de bondad de ajuste (SW, JB, KS_L).

Distribuciones de tratamientos	N3mero de r3plicas (<i>r</i>)			
8 tratamientos con distribuci3n $N(\mu, 4)$ y 1 tratamiento $Exp(2)$	5	10	15	20
7 tratamientos con distribuci3n $N(\mu, 4)$ y 2 tratamiento $Exp(2)$	5	10	15	20
6 tratamientos con distribuci3n $N(\mu, 4)$ y 3 tratamiento $Exp(2)$	5	10	15	20
5 tratamientos con distribuci3n $N(\mu, 4)$ y 4 tratamiento $Exp(2)$	5	10	15	20
4 tratamientos con distribuci3n $N(\mu, 4)$ y 5 tratamiento $Exp(2)$	5	10	15	20
3 tratamientos con distribuci3n $N(\mu, 4)$ y 6 tratamiento $Exp(2)$	5	10	15	20
2 tratamientos con distribuci3n $N(\mu, 4)$ y 7 tratamiento $Exp(2)$	5	10	15	20
1 tratamientos con distribuci3n $N(\mu, 4)$ y 8 tratamiento $Exp(2)$	5	10	15	20
0 tratamientos con distribuci3n $N(\mu, 4)$ y 9 tratamiento $Exp(2)$	5	10	15	20

Nota. La expresi3n $Exp(2)$ hace referencia a una distribuci3n Exponencial con par3metro $\beta = 2$.

Teniendo los escenarios definidos, se cre3 una funci3n con el objetivo de calcular los valores residuales del modelo ajustado, ecuaci3n (6). En ella se generan los valores aleatorios de cada tratamiento dependiendo del caso en el cual se encontraba. Luego, se crearon los factores *A* y *B*, cada uno con tres niveles. Posteriormente, se ajust3 el modelo citado anteriormente, y se calcularon los valores residuales de 3l.

Dado esto, se estim3 la proporci3n de veces que cada prueba de normalidad rechaz3 correctamente la hip3tesis nula. Es decir, se evalu3 si los residuos no seguían una distribuci3n normal, considerando que no fueran normales. Adem3s, para las tres pruebas de bondad de ajuste se tuvo en cuenta un nivel de significancia de $\alpha = 0.05$. Es importante destacar que este valor de α implica que se rechazará H_0 si el valor *p* es menor al nivel de significancia.

Seguidamente, se implement3 una funci3n que abarc3 la l3gica de simulaci3n para calcular la potencia de las tres pruebas de normalidad en cada escenario posible. Por lo cual, se realizaron un total de diez mil iteraciones para cada escenario, y los resultados se almacenaron en vectores por separado seg3n la prueba de bondad de ajuste cuando se rechaz3 la hip3tesis nula.

⁴ Estos valores generalmente son usados en los estudios de investigaci3n como los de (Mohd Razali y Yap, 2011; Thadewald y B3ning, 2007). Adem3s, el valor asignado a cada *n* en funci3n de *r* es igual a $n = r \times 9$.

En cada iteración, se generaron las observaciones correspondientes a cada tratamiento y se aplicaron las pruebas de normalidad. Asimismo, se calcularon los valores residuales con la función mencionada anteriormente y de esta manera se comprobó si alguna de las pruebas rechazaba la hipótesis nula de normalidad. Si ocurría esto, se registraba el resultado en el vector correspondiente a esa prueba con un 1, y en caso contrario, con un 0. Además, esta función recopiló todos los cálculos en una base de datos para tener posterior control sobre las tablas y gráficos de los resultados.

RESULTADOS

Una vez realizadas las diez mil iteraciones para cada uno de los escenarios propuestos en esta simulación, se encontraron aspectos que demostraron la influencia que tiene el número de réplicas y el número de tratamientos con distribución Exponencial con parámetro $\beta = 2$ sobre la potencia de las pruebas de normalidad de los residuos descrita anteriormente.

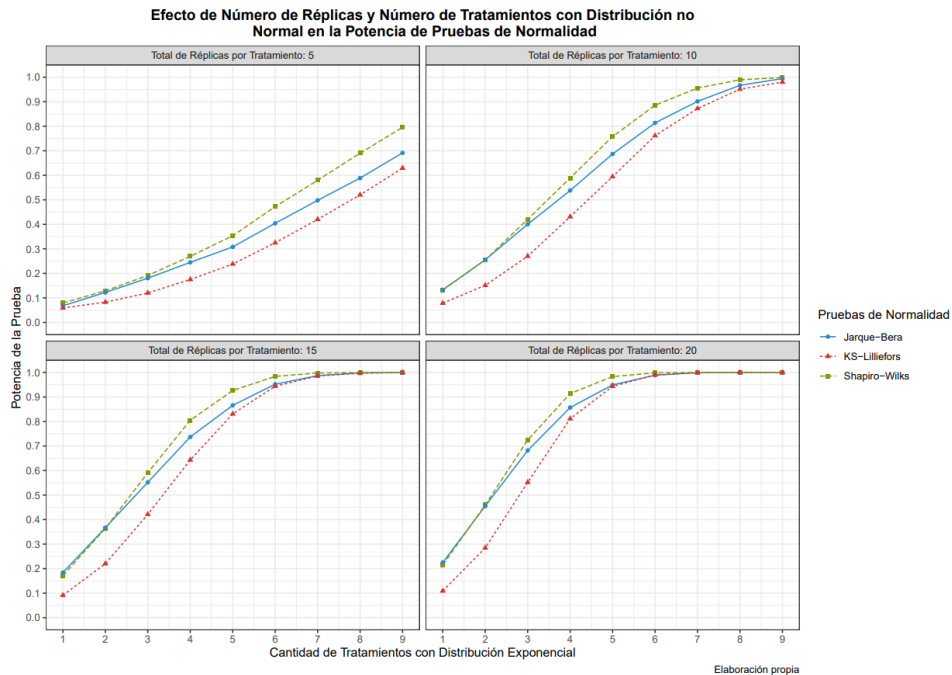
A continuación, se procedió a elaborar la Tabla 2, que resume la potencia simulada en los distintos escenarios. Dicha tabla muestra las diferentes combinaciones de tratamientos y sus respectivos números de réplicas, considerando un nivel de significancia $\alpha = 5\%$. Por otro lado, se realizó la Figura 2, la cual está compuesta por una serie de sub-gráficos, donde cada uno representa una cantidad fija de réplicas. Dentro de cada sub-gráficos, se incrementa gradualmente la cantidad de tratamientos con distribución exponencial. En el eje y de cada sub-gráficos se muestra la potencia de la prueba. En resumen, la Figura 2 muestra de manera visual cómo cambia la potencia de la prueba al aumentar el número de tratamientos dado diferentes números de réplicas.

Primeramente, se detectó que las tres pruebas analizadas en un escenario de 5 réplicas por tratamiento presentaron valores de la potencia relativamente bajos, en especial la prueba no paramétrica de Lilliefors (Kolmogorov-Smirnov), inclusive cuando los nueve tratamientos provenían de una distribución asimétrica positiva como lo fue la distribución exponencial. Este aspecto, es posible ligarlo en saber que en tamaños de muestras pequeños todas las pruebas tienen una sensibilidad a no encontrar esas diferencias cuando en verdad sí las hay (Noughabi, 2018). No obstante, la prueba paramétrica Shapiro-Wilks fue la que logró llegar a un umbral cercano de 80% de la potencia de la prueba, exactamente de 0.7962 (ver la Tabla 2).

Por otro lado, hubo similitudes entre las pruebas paramétricas de Jarque-Bera y Shapiro-Wilks en casos donde se contaba con una o dos distribuciones exponenciales, y las restantes siete distribuciones eran normales, esto para cada número de réplicas empleado. Este patrón se observó en los dos primeros puntos (azules y verdes) de las cuatro rejillas (total de réplicas por tratamiento) que se muestran en la Figura 2, para cada número de réplicas utilizado.

Figura 2

Resultados de la potencia de la prueba por tipo de prueba de bondad de ajuste.



Igualmente, un hecho importante de mencionar es como la prueba no paramétrica de Lilliefors (Kolmogorov-Smirnov) fue, en todos los escenarios posibles, el test que obtuvo las potencias del supuesto de normalidad de los residuos más bajas en comparación de sus homólogas paramétricas Shapiro-Wilks y Jarque-Bera. Parte de los resultados obtenidos con KS_L se basan, como menciona (Pedrosa et al., 2014), a su: «tendencia excesivamente conservadora» (p. 250). La prueba de Lilliefors (Kolmogorov-Smirnov) es más robusta que la prueba original de Kolmogorov-Smirnov. Esto significa que tiene más probabilidades de rechazar la hipótesis nula del supuesto de normalidad, incluso cuando la distribución de los datos es diferente de la normal.

En cuanto a las diferencias que mostró la prueba de Shapiro-Wilks en comparación con de Jarque-Bera y Lilliefors (Kolmogorov-Smirnov) se ven explícitamente cuando se tienen más de tres distribuciones asimétricas positivas y seis normales. En este caso, en la Figura 2 (y en la Tabla 2) la línea verde (Shapiro-Wilks) llega a aventajar a la azul (Jarque-Bera) y a la roja (Kolmogorov-Smirnov con corrección de Lilliefors) en el aspecto de poseer valores de potencia del cumplimiento del supuesto de normalidad de residuos más altos. Asimismo, se ve una diferencia más pronunciada si solo se toma en cuenta la línea verde y roja. La prueba de Jarque-Bera se llega a alejar un poco de los resultados de la prueba de Shapiro-Wilks, pero no tan marcada como la prueba de Lilliefors (Kolmogorov-Smirnov).

Finalmente, esta simulación logró evidenciar que al estar en escenarios con números de réplicas grandes (por ejemplo, con 15 y 20) y que poseen al menos un total de 6 distribuciones exponenciales por tratamiento, las potencias calculadas fueron muy similares entre ellas. Este hecho, (Farrell y Rogers-Stewart, 2006) lo trabaja en su estudio de simulación y argumenta que, al incrementar paulatinamente el tamaño de muestra, los resultados de las potencias de diversas

pruebas de bondad de ajuste llegan a niveles similares. En este caso, como se observa en la Figura 2 en la rejilla de la esquina inferior derecha, en su mayor parte todas las pruebas analizadas a partir de 6 distribuciones asimétricas por tratamiento llegan a obtener una potencia de la prueba mayor a 0.90. Asimismo, en la Tabla 2 se pueden observar todos los resultados obtenidos de esta simulación.

CONCLUSIONES

Mediante el análisis de los resultados obtenidos en esta simulación se entrelazó con otros estudios previos que trabajaron el análisis de la potencia del cumplimiento del supuesto de normalidad desde diversas metodologías y enfoques que dan ruta a seguir mejorando y analizando estudios como este.

Primero, la prueba de Shapiro-Wilks (paramétrica) con el estadístico W logró obtener resultados semejantes a los que (Mendes y Pala, 2003; Mohd Razali y Yap, 2011) mostraron en su estudio de simulación; independientemente del número de réplicas y de la cantidad específica de tratamientos con distribución exponencial, se observó que esta prueba era consistentemente la más efectiva en términos de potencia en comparación con las demás pruebas utilizadas. No obstante, en trabajos como los de (Nosakhare y Bright, 2017) donde se trabajó con distintos tipos de distribuciones (simétricas y asimétricas positivas) se sugiere reemplazar la prueba de Shapiro-Wilks por su homóloga con corrección Shapiro-Francia, tanto para muestras pequeñas como grandes.

Como se evidenció en la sección anterior, esta simulación también siguió el mismo hilo de resultados que autores como (Farrell y Rogers-Stewart, 2006) argumentan sobre la influencia que llega a tener el tamaño de la muestra (número de réplicas por tratamiento) sobre el cálculo de la potencia, ya que, a mayor tamaño de muestra, mayor potencia, independientemente de la cantidad de distribuciones exponenciales para este caso.

Igualmente, según (Mohd Razali y Yap, 2011; Thadewald y Büning, 2007) la prueba no paramétrica de Kolmogorov-Smirnov con corrección de Lilliefors es la que tiene menor rendimiento en comparación con las otras dos pruebas. Esto se debe a que se le dificulta rechazar que los residuales de un modelo se distribuyen normal cuando en realidad esto pasa, en parte se debe a su metodología más robusta y conservadora que sus homólogas paramétricas Shapiro-Wilks y Jarque-Bera.

En resumen, esta simulación se convierte en un punto de referencia para analizar otras distribuciones asimétricas positivas, como la distribución Gamma, Log-Normal, Pareto, entre otras (Nosakhare y Bright, 2017). Esto se debe a que, dependiendo de las características de cada una de estas distribuciones, los resultados obtenidos mediante el enfoque propuesto en este trabajo podrían variar. De esta manera, se podría obtener un panorama más completo sobre el efecto que tienen el número de réplicas y el número de tratamientos con distribuciones no normales en los resultados de la potencia de pruebas del cumplimiento del supuesto de normalidad de los residuos. Esto es especialmente relevante en presencia de distintos estadísticos que analizan la hipótesis mencionada en la expresión (2). Los investigadores deberían estudiar cómo otros factores, como los valores extremos, la heterocedasticidad y la complejidad del modelo (más tipos de distribuciones), pueden influir en el cumplimiento del supuesto de normalidad de los residuos.

ANEXOS

Anexo I: Tabla de resultados

Tabla 2

Comparación de la potencia de la prueba para las diferentes pruebas de normalidad contra las combinaciones de distribuciones en los tratamientos.

Distribución de tratamientos	Número de réplicas de cada tratamiento (r)	Potencia de la prueba		
		$\alpha = 0.05$		
		Shapiro-Wilks	Jarque-Bera	KS-Lilliefors
8 tratamientos con distribución $N(\mu, 4)$ y 1 tratamiento $Exp(2)$	5	0.0787	0.0699	0.0586
	10	0.1230	0.1310	0.0708
	15	0.1694	0.1804	0.0989
	20	0.2085	0.2248	0.1026
7 tratamientos con distribución $N(\mu, 4)$ y 2 tratamiento $Exp(2)$	5	0.1268	0.1202	0.0851
	10	0.2516	0.2529	0.1519
	15	0.3582	0.3607	0.2165
	20	0.4566	0.4556	0.2774
6 tratamientos con distribución $N(\mu, 4)$ y 3 tratamiento $Exp(2)$	5	0.1929	0.1851	0.1203
	10	0.4066	0.3849	0.2696
	15	0.5858	0.5544	0.4170
	20	0.7289	0.6824	0.5566
5 tratamientos con distribución $N(\mu, 4)$ y 4 tratamiento $Exp(2)$	5	0.2676	0.2441	0.1696
	10	0.5941	0.5430	0.4375
	15	0.7926	0.7349	0.6429
	20	0.9060	0.8443	0.7963
4 tratamientos con distribución $N(\mu, 4)$ y 5 tratamiento $Exp(2)$	5	0.3663	0.3221	0.2469
	10	0.7588	0.6818	0.6116
	15	0.9275	0.8689	0.8288
	20	0.9827	0.9517	0.9402
3 tratamientos con distribución $N(\mu, 4)$ y 6 tratamiento $Exp(2)$	5	0.4711	0.4060	0.3318
	10	0.8821	0.8143	0.7628
	15	0.9841	0.9509	0.9364
	20	0.9986	0.9877	0.9891
2 tratamientos con distribución $N(\mu, 4)$ y 7 tratamiento $Exp(2)$	5	0.5810	0.4959	0.4190
	10	0.9583	0.9058	0.8807
	15	0.9981	0.9896	0.9832
	20	0.9998	0.9983	0.9980
1 tratamientos con distribución $N(\mu, 4)$ y 8 tratamiento $Exp(2)$	5	0.6935	0.5936	0.5270
	10	0.9880	0.9652	0.9410
	15	1.0000	0.9982	0.9972
	20	1.0000	0.9999	0.9998
0 tratamientos con distribución $N(\mu, 4)$ y 9 tratamiento $Exp(2)$	5	0.7962	0.6894	0.6236
	10	0.9987	0.9941	0.9816
	15	1.0000	0.9999	0.9994
	20	1.0000	1.0000	1.0000

Anexo II: Código de la simulación

En la siguiente tabla se proporciona el código utilizado para la realización de la simulación propuesta en la plataforma de R en su versión 4.3.1. Además, se adjuntan una serie de comentarios que hacen posible seguir el hilo conductor del enfoque de esta simulación.

Tabla 3

Código de la simulación comentado.

Efecto del número de réplicas y número de tratamientos con distribución no normal en la potencia de pruebas de normalidad.	
Primera parte	
Código	Explicación
<pre>trat.distribucion = funcion(n, mu, v = 4, iteracion){</pre>	<p>Creamos una función la cual utiliza de parámetros un vector de n, un vector de mu, las cuales van a ser los mu de cada tratamiento para las distribuciones normales, una variancia fijada en 4 que es la variancia de todas las distribuciones de los tratamientos y la iteración es solo utilizada para la siguiente función de la simulación, dentro de esta función esa iteración lo que va a hacer es calcular la cantidad de n-(r*iteracion) que se ocupa generar con distribuciones normales y r*iteracion la cantidad de valores de la distribución exponencial. Se va a estar iterando la cantidad de distribuciones exponenciales que tienen que haber es decir que ese parámetro denominado interacción va a subir hasta llegar a 9 puesto que hay un total de 9 tratamientos. Esta función nos genera los escenarios y nos va a dar los residuales para cada escenario para poder calcular las potencias de la prueba de normalidad para cada caso.</p>
<pre>k = 9 # La simulación se basa en tener 10 tratamientos siempre r = n/k # Cantidad de réplicas por tratamiento, la simulación es con un diseño experimental balanceado if ((n-r*iteracion) == 0){</pre>	<p>En caso de que ya vayamos por la última iteración de las distribuciones, es decir que ya se removieron todas las distribuciones normales y ahora se ocupan generar todos los valores de cada factor con distribución Exponencial(2) para poder hacer la potencia con todos los tratamientos generados con distribución exponencial.</p>
<pre>x1 = factor(rep(1:k/2, each = r))</pre>	<p>Se crea el factor de A con 3 niveles con r repeticiones por nivel.</p>

<code>x2 = factor(rep(1:k/2, each = r))</code>	Se crea el factor B con 3 niveles con r repeticiones por nivel.
<code>y = rexp(n = n, rate = rep(1/2,each = r))</code> <code>}</code>	Se genera la variable respuesta para cada factor con una distribución exponencial de tamaño n, es decir el total de observaciones, por cada tratamiento se hizo dentro de la función rexp con un rate = rep(1/2, each=r) para que a cada factor se le asigne correctamente la varianza de 4
<code>if ((n-r*iteracion)!= 0){</code>	Este caso se cumple siempre cuando no se vaya por la última interacción de las distribuciones exponenciales es decir que va a generar una cantidad de tratamientos k-interacciones de distribución normal con los parámetros que queden del vector de mu.
<code>y_norm = rnorm(n-(r*iteracion),</code> <code>rep(mu,each=r), sqrt(v))</code>	Se generan n-(r*iteracion) valores de distribución normal en la cual se distribuye a cada uno de los tratamientos dependiendo del mu asignado en el vector del parámetro de la función.
<code>x1 = factor(rep(1:k/2, each = r))</code>	Se crea el factor de A con 3 niveles con r repeticiones por nivel.
<code>x2 = factor(rep(1:k/2, each = r))</code>	Se crea el factor B con 3 niveles con r repeticiones por nivel.
<code>y = c(y_norm, rexp(n = r*iteracion, rate =</code> <code>rep(1/2,each = r)))</code> <code>}</code>	Se genera un vector combinado de la variable y_norm con valores de una distribución exponencial de rexp tamaño r*iteracion, para que de esta manera tener el n que sale de parámetro en función.
<code>mod = aov(y ~ x1+x2+x1:x2)</code>	Se genera un modelo para poder obtener los residuales del modelo y poder evaluar la normalidad.
<code>return(mod\$res)</code> <code>}</code>	Devolvemos los residuales del modelo.
Segunda parte	
<code>simulacion.completa = function(n.ast =</code> <code>c(5*9,10*9,15*9,20*9), mu_fijo =</code> <code>c(4,6,2,3,4,3,6,5,4)){</code>	Esta función utiliza los parámetros, n.ast en donde va a ser un vector de los valores de n los cuales van a ser los casos de análisis de la cantidad de repeticiones por tratamiento, después tenemos el mu_fijo el cual es un vector de mu el cual son los promedios que se utilizaran para generar las distribuciones normales. Esta función calcula la proporción de veces que se rechaza la hipótesis nula para cada prueba de normalidad y cada escenario y las guarda en una lista para poder después devolver una base de datos la cual contenga al final todos los valores necesarios para saber la potencia de cada prueba para cada valor

	de n y cada combinación de tratamiento de distribuciones normales con exponenciales.
aleatorio = sample(mu_fijo, length(mu_fijo))	Generamos un <i>sample</i> el cual siempre se va a mantener, y este sirve para poder eliminar aleatoriamente un mu del vector mu_fijo al azar. Además, de poder insertar de esta manera una cantidad de 5 valores de distribución exponencial con parámetro lambda = ½.
data = list()	Lista en la cual vamos a guardar las potencias de las pruebas de normalidad, nombres de las pruebas, tamaño de n e la cantidad de tratamientos con distribuciones exponenciales de ½ a la hora de realizar la simulación.
m = 1	Este m nos ayudará a ir insertando en la lista.
pruebas = c("Shapiro-Wilks", "Jarque-Bera", "KS-Lilliefors")	Nombres de las pruebas para mayor facilidad de insertar el nombre en la lista donde guardamos las potencias y la cantidad de n.
for (k in 1:length(n.ast)){	Ciclo for el cual va a permitir correr a través de todas las n para poder iterar a través de todos los casos de los valores de la n nombrados anteriormente en el vector de n.ast, es decir que se itera a través de todos los casos de las n para poder ver las potencias para cada caso de n.
mu = mu_fijo	Definimos mu = mu_fijo puesto que se va a estar actualizando este parámetro para de esta manera ir quitándole al azar el mu de un tratamiento.
for (j in 1:length(mu_fijo)){	Este ciclo for sirve para poder iterar toda la lista de mu_fijo es decir del vector de mues, este es el valor que metemos en la iteración de la función anterior puesto que con esto vamos a saber cuántas distribuciones exponenciales hay. Cada vez que avance el j significa que hay una distribución exponencial más.
almacen.shap = rep(NA,10000)	Almacén en la cual vamos a guardar los valores de la potencia de la prueba de Shapiro-Wilks.
almacen.jb = rep(NA,10000)	Almacén en la cual vamos a guardar los valores de la potencia de la prueba de Jarque-Bera.
almacen.ksl = rep(NA,10000)	Almacén en la cual vamos a guardar los valores de la potencia de la prueba de KS-Lilliefors.
for (i in 1:10000){	Simulación de 10.000 iteraciones para saber cuántas veces se rechaza la hipótesis nula.
residuales = trat.distribucion(n.ast[k],mu=mu,iteracion = j)	Generamos los residuales utilizando la función hecha anteriormente.
almacen.shap[i] = 1*(shapiro.test(residuales)\$p.value < 0.05)	Almacenamos en nuestros almacenes nombrados en este caso almacen.shap anteriormente, cuando se rechaza la hipótesis nula, cada vez que se

	rechace se le asigna un 1 y cuando no logra rechazar un 0.
almacen.jb[i] = 1*(jarque.bera.test(residuales)\$p.value < 0.05)	Almacenamos en nuestros almacenes nombrados anteriormente, cuando se rechaza la hipótesis nula.
almacen.ksl[i] = 1*(lillie.test(residuales)\$p.value < 0.05) }	Almacenamos en nuestros almacenes nombrados anteriormente en este caso almacen.ksl, cuando se rechaza la hipótesis nula.
data[[m]] = c(pruebas[1],mean(almacen.shap),n.ast[k],j)	Ahora cuando salimos de la simulación de 10.000 veces guardamos la potencia de la prueba en la lista, nombrada al principio de la función, dentro de esta se guarda; el nombre de la prueba en este caso es Shapiro-Wilks; el promedio del almacén de Shapiro, es decir cuantas veces pudo rechazar H0 cuando H1 es cierta; el n es decir la cantidad de réplicas totales, en la cual se hizo el análisis de la potencia; y el j que es la cantidad de tratamientos que hay con distribución exponencial en la cual se utilizó para calcular la potencia de la prueba.
m = m + 1	Le sumamos un 1 al índice de m para poder insertar el siguiente valor en la lista.
data[[m]] = c(pruebas[2],mean(almacen.jb),n.ast[k],j)	Guardamos la potencia de la prueba en la lista, nombrada al principio de la función, dentro de esta se guarda; el nombre de la prueba en este caso es Jarque-Bera; el promedio del almacén de JB, es decir cuantas veces pudo rechazar H0 cuando H1 es cierta; el n es decir la cantidad de réplicas totales, en la cual se hizo el análisis de la potencia; y el j que es la cantidad de tratamientos que hay con distribución exponencial en la cual se utilizó para calcular la potencia de la prueba.
m = m + 1	Le sumamos un 1 al índice de m para poder insertar el siguiente valor en la lista.
data[[m]] = c(pruebas[3],mean(almacen.ksl),n.ast[k],j)	Guardamos la potencia de la prueba en la lista, nombrada al principio de la función, dentro de esta se guarda; el nombre de la prueba en este caso es K-S Lilliefors; el promedio del almacén de ksl, es decir cuantas veces pudo rechazar H0 cuando H1 es cierta; el n es decir la cantidad de réplicas totales, en la cual se hizo el análisis de la potencia; y el j que es la cantidad de tratamientos que hay con distribución exponencial en la cual se utilizó para calcular la potencia de la prueba.
m = m + 1	Le sumamos un 1 al índice de m para poder insertar el siguiente valor en la lista.

<pre>mu = mu[-aleatorio[j]] } }</pre>	<p>Aquí es donde usamos la información que nos dio el <i>sample</i> definido al principio de la función para que de esta manera se remueva al azar un valor del vector de mues de distribución normal e insertar en lugar de ese tratamiento con distribución normal una distribución exponencial, ya que el j va a avanzar, hasta llegar a tener todos los tratamientos generados con distribuciones exponenciales. Y cuando esto sucede se prosigue al primer for loop hecho es decir se repite todo, pero con otro tamaño de n.</p>
<p>Tercera parte</p>	
<pre>data = as.data.frame(do.call("rbind",data))</pre>	<p>Reestructuramos la lista llamada data, es decir toda la información conseguida de todos los escenarios posibles con las tres pruebas, en un dataframe para tener mejor visualización de los datos.</p>
<pre>colnames(data) = c("Normalidad", "Potencia", "N", "Distribuciones")</pre>	<p>Definimos los nombres de las columnas del dataframe de data.</p>
<pre>return(data) }</pre>	<p>Por último, devolvemos el data frame creado llamado data para poder tener visualización de los resultados y poder hacer análisis gráfico.</p>
<pre>data = simulacion.completa() # Visualizar la base: data</pre>	<p>Obtenemos los resultados y los visualizamos.</p>

Nota. Este código consta de tres partes generales. La primera de ellas corresponde al cálculo de los valores residuales con las pautas mencionadas en la METODOLOGÍA, la segunda sección está asociada al cálculo de la potencia de la prueba para cada una de las tres pruebas de bondad de ajuste que se utilizaron y la última parte corresponde a la reestructuración de los datos para la generación de una base de datos que ayude a mostrar los datos de forma óptima para el análisis y la visualización de los resultados, como las tablas o los gráficos presentes en este documento.

BIBLIOGRAFÍA

- Chou-Chen, S. W. (2023, May). *Clase 15: Contrastes de hipótesis*. Teoría Estadística. https://shuwei325.github.io/XS3310-I23/XS3310-I23_15.html#3
- Coin, D. (2008). Testing normality in the presence of outliers. *Statistical Methods and Applications*, 17, 3–12. <https://doi.org/https://doi.org/10.1007/s10260-007-0046-8>
- Dufour, J., Farhat, A., Gardiol, L., y Khalaf, L. (1998). Simulation-based finite sample normality tests in linear regressions. *The Econometrics Journal*, 1(1), C154–C173. <https://doi.org/10.1111/1368-423X.11009>
- Farrell, P. J., y Rogers-Stewart, K. (2006). Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *Journal of Statistical Computation and Simulation*, 76(9), 803–816. <https://doi.org/10.1080/10629360500109023>
- Gross, J., y Ligges, U. (2015). *lillie.test: Lilliefors (Kolmogorov-Smirnov) test for normality*. Nortest: Tests for Normality. <https://www.rdocumentation.org/packages/nortest/versions/1.0-4/topics/lillie.test>
- Gujarati, D. N., y Porter, D. C. (2010). *Econometría* (5th ed.). McGraw-Hill/Irwin, Inc.
- Jarque, C. M., y Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255–259. [https://doi.org/https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/https://doi.org/10.1016/0165-1765(80)90024-5)
- Mak, T. K. (2000). Heteroscedastic regression models with non-normally distributed errors. *Journal of Statistical Computation and Simulation*, 67(1), 21–36. <https://doi.org/10.1080/00949650008812034>
- Mendes, M., y Pala, A. (2003). Type I Error Rate and Power of Three Normality Tests. *Information Technology Journal*, 2(2), 135–139. <https://doi.org/10.3923/itj.2003.135.139>
- Mohd Razali, N., y Yap, B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analytics*, 2. https://www.researchgate.net/publication/267205556_Power_Comparisons_of_Shapiro-Wilk_Kolmogorov-Smirnov_Lilliefors_and_Anderson-Darling_Tests
- Nosakhare, U. H., y Bright, A. F. (2017). Evaluation of Techniques for Univariate Normality Test Using Monte Carlo Simulation. *American Journal of Theoretical and Applied Statistics*, 6(5–1), 51–61. <https://www.sciencepublishinggroup.com/journal/paperinfo?journalid=146&doi=10.11648/j.ajtas.s.2017060501.18>
- Noughabi, H. A. (2018). A Comprehensive Study on Power of Tests for Normality. *Journal of Statistical Theory and Applications*, 17(4), 647–660. <https://doi.org/10.2991/jsta.2018.17.4.7>
- Pedrosa, I., Juarros-Basterretxea, J., Robles-Fernández, A., Basteiro, J., y García-Cueto, E. (2014). Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar? *Universitas Psychologica*, 14(1), 245–254. <https://doi.org/10.11144/Javeriana.upsy14-1.pbad>

- Rodríguez Cruz, I. D., Camacho Pochet, M., Cedeño Guzmán, A., y Villalobos Martínez, M. S. (2023). *Efecto del grado de acidez del suelo (pH) y la temperatura sobre el crecimiento del largo promedio de las raíces de semillas de lechuga Great Lakes* 118.
- Shapiro, S. S., y Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>
- Thadewald, T., y Büning, H. (2007). Jarque–Bera Test and its Competitors for Testing Normality – A Power Comparison. *Journal of Applied Statistics*, 34(1), 87–105. <https://doi.org/10.1080/02664760600994539>
- Ventura-León, J., Peña-Calero, B. N., y Burga-León, A. (2022). The effect of normality and outliers on bivariate correlation coefficients in psychology: A Monte Carlo simulation. *The Journal of General Psychology*, 1–18. <https://doi.org/10.1080/00221309.2022.2094310>