Impacto de diferentes pruebas de normalidad ante la existencia de heterocedasticidad

Jose Cascante[1], Fatima Saborío[1], Karen Acuña[1]
jose.cascantesolis@ucr.ac.cr, fatima.saborio@ucr.ac.cr, karen.acunapoveda@ucr.ac.cr

RESUMEN

El cumplimiento de todos los pasos en una investigación científica es crucial si se quiere garantizar obtener conclusiones con autenticidad, por esto, verificar los supuestos teóricos es una parte inevitable. Entre los supuestos a probar se destacan el de normalidad y el de homocedasticidad, por su parte, es importante tomar en cuenta el impacto que tiene incumplir el supuesto de homocedasticidad al estudiar la normalidad, debido a que este afecta la veracidad de las probabilidades obtenidas al inflar los errores tipo I. Con el fin de cuantificar este impacto, se realizó una simulación para conocer la proporción de rechazos erróneos de la hipótesis nula (error tipo I) al evaluar normalidad con las pruebas Kolmogórov-Smirnov, Shapiro-Wilk y Jarque-Bera. Estableciendo diferentes escenarios se obtuvo como resultado que las pruebas de Shapiro-Wilk y Jarque-Bera tienden a tener una proporción de veces en las que se comete error tipo I más alta que la prueba de Kolmogorov-Smirnov. Asimismo, al aumentar la cantidad de repeticiones por tratamiento, la proporción de veces en las que se comete error tipo I también aumenta, principalmente en los casos de heterocedasticidad.

PALABRAS CLAVES: Supuestos, error tipo I, simulación, robustez, réplicas.

INTRODUCCIÓN

Ante la realización de una investigación científica se debe de seguir un proceso multifacético donde una parte muy importante es comprobar que se cumplan los supuestos teóricos. Por esto, Gandica de Roa (2020) menciona que muchas veces esta verificación de los supuestos se pasa por alto por lo engorroso que puede llegar a ser, sin embargo, este seguimiento paso a paso es capaz de generar conclusiones con validez y una alta confiabilidad. Asimismo, uno de estos supuestos es el de la normalidad, el cual es necesario ya que es la base para tomar la decisión de cuáles pruebas utilizar en un determinado estudio.

Además, otro supuesto fundamental que se debe de comprobar su cumplimento es la homocedasticidad, el cual, como menciona Reacher y Schaalje (2000) en los modelos lineales generales el tener homocedasticidad asegura la precisión de los errores estándar y asintóticos, y covarianzas entre los parámetros estimados. A su vez, en aquellos casos donde se tiene heterocedasticidad compromete la estimación de la incertidumbre de la medición, a pesar de que los parámetros estimados continúen insesgados y consistentes, las estimaciones de la matriz de covarianza estimada entre los parámetros serán incorrectas, esto puede generar menor poder estadístico (menor potencia) o inflar los errores tipo I (Sánchez et al., 2011; Rencher y Schaalje, 2000).

[1]Estudiantes de Bachillerato en Estadística de la Universidad de Costa Rica

También, como menciona Oddi et al. (2020) la existencia de heterocedasticidad implica un problema en el análisis, debido a que esto afecta la veracidad de las probabilidades obtenidas en las pruebas de significancia, los errores estándar y los estadísticos no seguirán las distribuciones teóricas, por lo que las inferencias realizadas por un modelo que viola los supuestos no serán válidas.

Con el fin de evaluar el supuesto de normalidad se utilizarán 3 diferentes pruebas de bondad de ajuste, esto porque cada una implica una forma diferente de analizar este supuesto, lo que genera un criterio más amplio de los factores que pueden influir en la toma de decisiones. Cada una de estas pruebas permiten realizar un contraste de hipótesis en el cual se toma como hipótesis nula que los datos siguen una distribución normal, en el caso en el que se encuentre que la probabilidad asociada a la prueba es menor a la significancia, se rechaza esta hipótesis.

Flores y Flores (2021) mencionan en primer lugar que la prueba Shapiro-Wilk contrasta la normalidad cuando el tamaño de muestra es menor a 50 observaciones, ya que con muestras muy grandes es equivalente a la prueba Kolmogorov-Smirnov, asimismo, la prueba Kolmogorov-Smirnov es verdaderamente útil cuando se utilizan procesos no lineales e iterativos que llevan a distribuciones no gaussianas. Por último, la prueba Jarque-Bera, evalúa la normalidad a partir del sesgo y la curtosis, es una de las pruebas más populares en aplicaciones de regresión (Carmona y Carrión, 2015).

Dada la importancia del cumplimento del supuesto de normalidad, se plantea como objetivo analizar el impacto de infracción del supuesto de homocedasticidad sobre la probabilidad de cometer error tipo I en las pruebas de normalidad de Shapiro-Wilk, Jarque-Bera y Kolmogórov-Smirnov mediante la utilización de un modelo factorial con interacción, además analizar el efecto del aumento de réplicas en la probabilidad de cometer error tipo I en las pruebas de normalidad.

Según lo mencionado por Gandica de Roa (2020) de que a pesar de que las tres pruebas de normalidad anteriormente mencionadas cumplen con evaluar la normalidad, no existe evidencia de cuál prueba es más conveniente que la otra, dado que no se encontró evidencia de cuál de las pruebas es superior a las otras en robustez y potencia. Por este motivo, se plantea como hipótesis que los resultados de la proporción de veces que se comete error tipo I van a ser similares en todas las pruebas de normalidad aplicadas.

METODOLOGÍA

Para llevar a cabo el objetivo de la investigación se creará una simulación que cuantifique la proporción de veces que se comete error tipo I en cada una de las pruebas seleccionadas, obteniendo los datos a partir de una distribución normal con medias establecidas y diferentes varianzas por lo que no se desea rechazar la hipótesis nula, esto suponiendo un experimento con un diseño factorial de 2 factores que se identificarán como factor A y factor B, ambos factores tendrán 2 niveles cada uno, por lo que se cuenta con 4 tratamientos. Se utilizará un modelo con interacción entre los factores anteriormente mencionados.

A partir de lo anterior, se proponen 6 escenarios, de estos 5 presentan heterocedasticidad y 1 de ellos presenta homocedasticidad, con el fin de realizar una comparación. Es importante mencionar que para todos estos escenarios los tratamientos

compartirán las mismas medias seleccionadas aleatoriamente las cuales corresponden a 98, 86, 14, 29 respectivamente, además, se supone que el experimento es balanceado por lo que cada escenario se probará con 7 cantidades de repeticiones distintas.

Los primeros 5 escenarios presentan heterocedasticidad, por lo que el primero proviene de una distribución normal con las medias seleccionadas, y diferentes varianzas para cada tratamiento sin algún patrón exacto, estas fueron: 80, 57, 4, 3. Para el segundo escenario se seleccionaron 2 varianzas altas para los primeros 2 tratamientos y 2 varianzas muy bajas para los últimos 2 tratamientos, las cuales son: 73, 80, 10, 15; En el tercer escenario se seleccionaron 4 varianzas muy bajas y parecidas entre sí, estas corresponden a: 5, 10, 9, 11; Para el cuarto escenario se establecieron las varianzas con el mismo patrón que en el escenario 2, sin embargo, en este caso las varianzas de los 2 primeros tratamientos son iguales, y las de los últimos 2 igual, de la siguiente forma: 67, 67, 6, 6; Por otro lado, en el quinto escenario se eligió una varianza alta para el primer tratamiento y para los demás, 3 iguales y bajas, las cuales son: 25, 8, 8, 8. Por último, el sexto escenario presenta homocedasticidad, por lo que se establecen las mismas varianzas a los 4 tratamientos: 8, 8, 8, 8.

Para analizar el impacto de la homocedasticidad en el supuesto de normalidad se seleccionaron 3 pruebas, esto con el fin de cuantificar la proporción de veces que se comete error tipo I en cada una de ellas. Las pruebas de bondad de ajuste para analizar la normalidad que se eligieron son Shapiro-Wilk, Jaque-Bera y Kolmogórov-Smirnov.

En primer lugar, la prueba de Shapiro-Wilk, esta considera que el gráfico de probabilidad normal, conocido como gráfico cuantil-cuantil o QQ-Plot, al examinar el ajuste de un grupo de datos para la distribución normal sea parecido a una línea de regresión, es decir, al examinar estos datos, el gráfico debe presentar una tendencia diagonal continúa. Esta prueba utiliza el estadístico W (Flores y Flores, 2021):

$$W = \frac{\left(\sum_{i=1}^{n} \alpha_i X_{(i)}\right)^2}{\sum_{i=1}^{n} \left(X_i - \underline{X}\right)}$$

Donde $X_{(i)}$ es el número de la posición que ocupa en la muestra, X_i corresponde a los datos de la muestra ordenados por tamaño, $\underline{X}$ a la media muestral de los mismos, n es la cantidad de observaciones en la muestra, además, α_i es un conjunto de variables ordenadas de azar con una distribución normal estándar.

Asimismo, la prueba de Jarque-Bera es parte de la clase de pruebas generales de momentos, es decir, esta prueba analiza la asimetría y la curtosis al mismo tiempo de los datos que vienen de un modelo Gaussiano. La misma utiliza los coeficientes de la asimetría y la curtosis respectivamente, así como el estadístico JB (Lo et al., 2025).

$$b_{n,2} = \frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(X_i - \underline{X}\right)^3}{\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(X_i - \underline{X}\right)^2\right]^{\frac{3}{2}}}$$

$$a_{n,2} = \frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(X_i - \underline{X}\right)^4}{\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(X_i - \underline{X}\right)^2\right]^2}$$

$$JB = \frac{n}{6}\left[b_{n,2}^2 + \frac{1}{4}\left(a_{n,2} - 3\right)^2\right]$$

Donde $a_{n,2}$ es la asimetría y $b_{n,2}$ es la curtosis, además, n es la cantidad de observaciones en la muestra.

Por último, la prueba Kolmogórov-Smirnov, la cual consiste en comparar la función de distribución acumulada empírica de los datos de la muestra con la distribución esperada si los datos fueran normales, por lo que se toma en cuenta al tomar la decisión de rechazar la hipótesis o no es la diferencia que existe entre estas funciones de distribución. Esta prueba contrasta la hipótesis de que existe normalidad en la población por lo que usa 2 contrastes, para 2 colas, y una ecuación (Flores y Flores, 2021).

La ecuación corresponde a:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \{\frac{1{:}si, y_i \leq x_i}{0{:}alternativa}$$

Mientras que los contrastes son:

$$D_n^+ = max\left(F_n(x) - F(x)\right)$$
$$D_n^- = max\left(F(x) - F_n(x)\right)$$

Donde $F(x)$ es la función de distribución y n la cantidad de observaciones en la muestra.

Para el desarrollo de la simulación; gráficos y diferentes pruebas se utilizó el lenguaje de R con la versión 4.2.2 (R Core Team, 2024), utilizando como principales paquetes; *shiny* (Chang et al., 2024), *ggplot2* (Wickham, 2016), *tseries* (Trapletti y Hornik, 2024), *plotly* (Sievert, 2020), *stats* (R core team, 2023) y *reshape2* (Wickham, 2007), además, la significancia utilizada fue de 0.05 y la cantidad de iteraciones fue de 5000 en todos los escenarios.

RESULTADOS

Primeramente, se presentan las tablas con los resultados de los 6 escenarios planteados, estos resultados corresponden a la proporción de veces que se comete error tipo I para cada escenario con las 7 cantidades de repeticiones.

Cuadro 1.

Proporción de veces que se comete error tipo I con la prueba de Shapiro-Wilk.

Número de réplicas	3	4	12	25	30	40	50
Escenario 1	0.11	0.25	0.77	0.98	1	1	1
Escenario 2	0.03	0.11	0.32	0.59	0.66	0.80	0.90
Escenario 3	0.02	0.04	0.06	0.07	0.06	0.08	0.06
Escenario 4	0.05	0.14	0.57	0.91	0.95	0.99	1
Escenario 5	0.03	0.06	0.16	0.22	0.31	0.36	0.42
Escenario 6	0.02	0.04	0.05	0.06	0.04	0.05	0.05

En el cuadro 1 se presentan los resultados para la prueba Shapiro-Wilk, en esta se puede apreciar como en el primer escenario que se estableció con 4 varianzas muy diferentes sin ningún patrón exacto, a partir de las 30 repeticiones se comete error tipo I en todos los casos, sin embargo, en pocas cantidades de réplicas esta proporción es baja, aunque en el resto de los escenarios la proporción de veces de cometer error tipo I es más alta; además, para el sexto y último escenario, el cual posee homocedasticidad es posible apreciar que es el que tiene una proporción menor a la de los demás escenarios.

Por otro lado, en el cuarto escenario para el cual se establecieron 2 varianzas altas e iguales para los primeros 2 tratamientos y 2 varianzas bajas e iguales para los últimos 2 tratamientos, se observa que a partir de 25 repeticiones se comete error tipo I en la mayoría de los casos, en contraste con el tercer escenario el cual corresponde al caso donde todas las varianzas de los tratamientos son distintas, pero estas varían muy poco entre cada una, esta posee una baja proporción aunque se utilicen muchas repeticiones, similar a los resultados obtenidos en el sexto escenario.

Cuadro 2.

Proporción de veces que se comete error tipo I con la prueba de Jarque-Bera.

Número de réplicas	3	4	12	25	30	40	50
Escenario 1	0.01	0.14	0.59	0.90	0.95	0.99	1
Escenario 2	0.00	0.06	0.28	0.57	0.67	0.80	0.86
Escenario 3	0.00	0.01	0.06	0.07	0.07	0.09	0.11
Escenario 4	0.00	0.10	0.45	0.80	0.86	0.96	0.97
Escenario 5	0.00	0.03	0.22	0.36	0.41	0.46	0.54
Escenario 6	0.00	0.01	0.04	0.04	0.03	0.03	0.05

Asimismo, en el Cuadro 2 se muestran los resultados para la prueba de Jarque-Bera, es importante notar que estos resultados siguen un comportamiento similar a los encontrados con la prueba de Shapiro-Wilk, sin embargo, en el primer escenario, hasta que se llega a 50 réplicas es que se comete error tipo I en todos los casos, esto puede ser evidencia de que la prueba Jarque-Bera aunque produce resultados parecidos, tiende a rechazar menos veces que la prueba de Shapiro-Wilk.

Cuadro 3.

Proporción de veces que se comete error tipo I con la prueba de Kolmogórov-Smirnov.

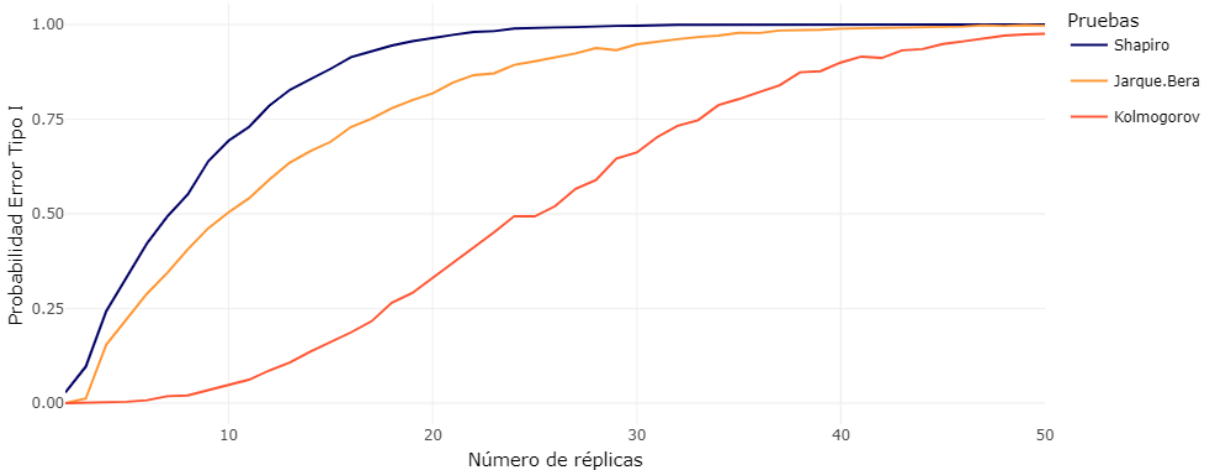
Número de réplicas	3	4	12	25	30	40	50
Escenario 1	0.00	0.00	0.07	0.48	0.67	0.91	0.98
Escenario 2	0.00	0.00	0.00	0.01	0.02	0.04	0.08
Escenario 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Escenario 4	0.00	0.00	0.45	0.12	0.20	0.36	0.57
Escenario 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Escenario 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Por consiguiente, en el Cuadro 3 se encuentran los resultados para la prueba de Kolmogórov-Smirnov, estos aunque tienen comportamientos como los de las pruebas de Jarque-Bera y Shapiro-Wilk, la proporción de veces en las que se comete error tipo I es muy baja en los escenarios en los que se presenta un nivel de heterocedasticidad más bajo, incluso, sin importar la cantidad de réplicas en los escenarios 3, 5 y 6, los cuales se establecieron con varianzas muy similares no se comete error tipo I en ninguna ocasión, pero en los escenarios donde se presenta una mayor heterocedasticidad (1 y 4) aumentó la proporción de veces de cometer error tipo I comienza a aumentar especialmente cuando se tiene réplicas mayores a 12.

A continuación, se presentan los gráficos de dos de los escenarios para hacer una comparación entre las pruebas y además, evidenciar la diferencia que existe entre los casos con heterocedasticidad y homocedasticidad.

Figura 1.

Proporción de veces que se comete error tipo I comparando las pruebas de Shapiro-Wilk, Jaque-Bera y Kolmogorov-Smirnov con heterocedasticidad.

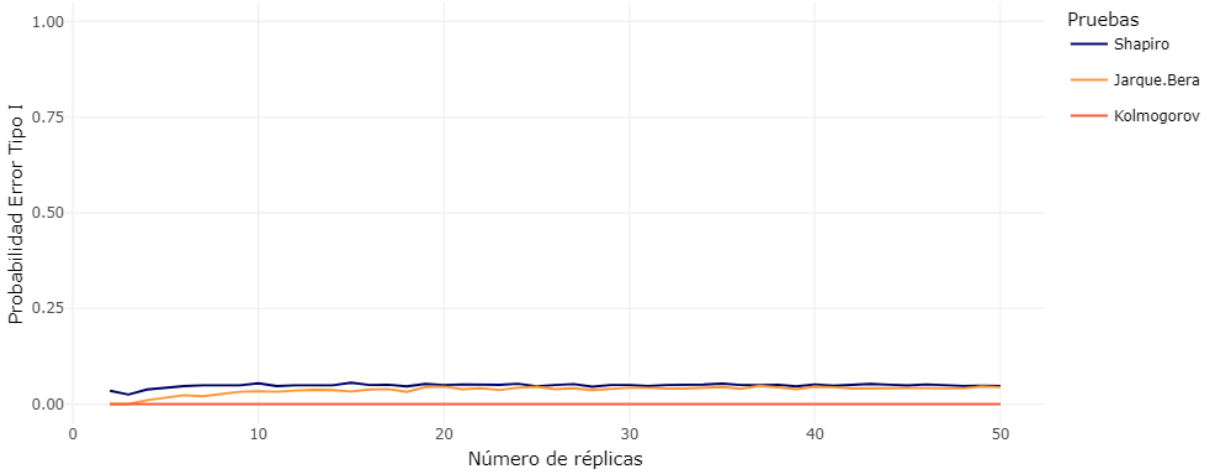


En primer lugar, en la figura 1 se muestra la comparación de las pruebas Shapiro-Wilk, Jaque-Bera y Kolmogórov-Smirnov en el primer escenario, el cual presenta heterocedasticidad, se aprecia que al tener pocas réplicas por tratamiento las pruebas tienen una proporción de cometer error tipo I más baja, conforme estas réplicas van aumentando la proporción también lo hace; Además, aunque pruebas de Shapiro-Wilk y Jarque-Bera tienen una tendencia similar,

se nota que la prueba Shapiro-Wilk presenta proporciones más altas, por otro lado, la prueba de Kolmogórov-Smirnov como se mencionó en el cuadro 3 presenta proporciones bajas en comparación a las otras dos pruebas, pero se nota un aumento en la proporción de veces de cometer error tipo I cuando se tiene más de 12 repeticiones.

Figura 2.

Proporción de veces que se comete error tipo I comparando las pruebas de Shapiro-Wilk, Jaque-Bera y Kolmogórov-Smirnov con homocedasticidad.



Por otra parte, en la figura 2 se evidencia la comparación de la proporción de veces que se comete error tipo I en las pruebas Shapiro-Wilk, Jaque-Bera y Kolmogórov-Smirnov para el escenario 6, este corresponde al escenario con homocedasticidad. Es posible notar que todas las pruebas presentan proporciones de cometer error tipo I muy bajas, principalmente la prueba de Kolmogórov-Smirnov, la cual permanece en 0 para todos los casos. Asimismo, las pruebas Shapiro-Wilk y Jarque-Bera presentan proporciones que además de ser bajas, son extremadamente similares.

CONCLUSIONES

Es importante mencionar que los datos al provenir de distribuciones normales, se espera que la proporción de veces de cometer error tipo I sea mínima, ya que, al evaluar pruebas de normalidad no se desea rechazar la hipótesis nula de que los datos provienen de una distribución normal, por lo que es notorio la gran distinción obtenida en los escenarios con homocedasticidad y con heterocedasticidad, como en los casos con homocedasticidad las pruebas de normalidad obtuvieron una menor proporción de casos de cometer error tipo I en todas las réplicas, opuesto a los casos que se tienen heterocedasticidad, esta proporción aumenta especialmente cuando se incrementa los tamaños de muestra, esto confirma lo mencionado por Sánchez et al. (2011) de que con heterocedasticidad se compromete la estimación de la incertidumbre de la medición, además de que la heterocedasticidad infla los errores tipo I (Rencher, Schaalje, 2000).

Con respecto a las pruebas para evaluar normalidad Shapiro-Wilk y Jarque-Bera, estas obtuvieron resultados muy similares en la mayoría de los escenarios especialmente, en los que se tenía homocedasticidad o poca alteración de las varianzas de cada tratamiento, en estos

casos la prueba de Kolmogórov-Smirnov tuvo una proporción muy baja de prácticamente cero en todas las repeticiones. Pero, en los casos que se presentó heterocedasticidad, la proporción de veces de cometer error tipo I fue más alta en las pruebas de Shapiro y Jarque-Bera respecto a la prueba de Kolmogórov-Smirnov, la cual aumentó respecto se incrementó el tamaño de muestra, esto incumple con la hipótesis planteada de que la proporción de veces que se comete error tipo I va a ser similar en las pruebas de normalidad, ya que si se presentaron alteraciones muy evidentes entre cada prueba.

Dado que la prueba de Kolmogórov-Smirnov fue la que tuvo un mejor comportamiento en los escenarios con heterocedasticidad especialmente con repeticiones menores a 12, esta prueba puede ser una buena opción es los casos que se incumple el supuesto de homocedasticidad y se desea analizar la normalidad, como lo mencionado por Pedrosa et al. (2015) que cuando se incumple los supuestos de homocedasticidad se puede utilizar pruebas no paramétricas por ser más robustas cuando se violan los supuestos. Pero, también es evidente como fuera de esta condición, todas las pruebas en los escenarios heterocedásticos dan resultados erróneos, debido a que al inflarse el error tipo I aumenta la proporción de veces en las que se comete, esto indica lo susceptibles que fueron las pruebas de normalidad al violar el supuesto de homocedasticidad, dado que al ser pruebas de normalidad y los datos son normales se espera que la proporción de veces de cometer error tipo I sea mínima, contrario a lo que se obtuvo, por lo que no se recomienda la utilización de estas pruebas en casos con heterocedasticidad.

Finalmente, a modo de recomendación se debe recordar que no es óptimo solo basarse en los resultados de una prueba al realizar el análisis de supuestos, además, para analizar la normalidad en caso de que no se lograse corregir la invalidación de homocedasticidad se podría optar en casos donde se ha logrado recolectar muchas réplicas analizar este supuesto para los datos de cada tratamiento, además, otra opción es analizar la normalidad mediante un gráfico de cuantil-cuantil (gráfico de normalidad o QQ-Plot).

BIBLIOGRAFÍA

Carmona, M., & Carrión, H. (2015). Potencia de la prueba estadística de normalidad Jarque-Bera frente a las pruebas de Anderson-Darling, Jarque-Bera robusta, Chi cuadrada, Chen-Shapiro y Shapiro-Wilk [Universidad Autónoma del Estado de México]. https://core.ac.uk/download/pdf/159384191.pdf

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B. (2024). shiny: Web Application Framework for R. R package version 1.8.1.1, <https://CRAN.R-project.org/package=shiny>.

Flores, C., & Flores, K. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. Societas, 23(2), 83-97. http://portal.amelica.org/ameli/jatsRepo/341/3412237018/3412237018.pdf

Gandica de Roa, E. (2020). Potencia y Robustez en Pruebas de Normalidad con Simulación Montecarlo. Revista Scientific, 5(18), 108-119. http://www.indteca.com/ojs/index.php/Revista_Scientific/article/view/468

Lo, G., S., Thiam, O., and Haidara, M.C. (2015) High Moments Jarque-Bera Tests for Arbitrary Distribution Functions. *Applied Mathematics. 6*, 707-716. http://dx.doi.org/10.4236/am.2015.64066

Oddi, F., J., Miguez, F., E., Benedetti, G. G., & Garibaldi, L. A. (2020). Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza. Ecología Austral; 30; 438-453. https://doi.org/10.25260/10.25260/EA.20.30.3.0.1131

Pedrosa, I., Basterretxea, J., Fernández, A., Basteiro, J., García, E. (2015). Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar?. *Universitas Psychologica 14*(1), 245-254. https://doi.org/10.11144/Javeriana.upsy13-5.pbad

R Core Team. (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rencher, A., Schaalje, B. (2000). *Linear models in statistics.* Editorial John Wiley & Sons, Inc., Hoboken.

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Sanchez, J., Cano, R., Corral, S., Fuentes, X. (2011). Heteroscedasticity and homoscedasticity, and precision profiles in clinical laboratory sciences. *Clinica Chimica Acta. 412*(23), 2351-2352. https://doi.org/10.1016/j.cca.2011.08.005

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida.

Trapletti, A, Hornik, K. (2024). Tseries: Time Series Analysis and Computational Finance. R package version 0.10-56, <https://CRAN.R-project.org/package=tseries>.

Wickham, H. (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

ANEXOS

Simulación pruebas de normalidad bajo heterocedasticidad.	
Código	**Explicación**

library(shiny); library(ggplot2); library(plotly); library(tseries); library(reshape2)
 	
# Interfaz 	
ui <- fluidPage(
  titlePanel("Análisis de Pruebas de Normalidad bajo Heterocedasticidad"),
 	
  sidebarLayout(
  sidebarPanel(
  textInput("mu", "Medias (mu):", value = "98,86,14,29"),
  textInput("v", "Varianzas (v):", value = "98,86,14,29"),
  sliderInput("r", "Número de réplicas:", min = 2, max = 80, value = 50),
  actionButton("run", "Correr Simulación")
  ),
 	
  mainPanel(
  plotlyOutput("plot"),
  tableOutput("results")
)
)
)
	En primer lugar, se cargan las librerías necesarias
	
	Se establece la entrada de los datos que establecerá el usuario, en este caso las medias y las varianzas son vectores de 4 valores, mientras que la cantidad de réplicas se establecen con un deslizador; Asimismo, para correr la simulación se usa un “Action button”
	
	Además, se establecen las entradas del gráfico y la cuadro con los resultados de la simulación

# Servidor 	
server <- function(input, output) {
 	
  sh <- function (r1, mu, v) {
  k <- length(mu)
  n <- r1 * k
  y <- rnorm(n = n, mean = rep(mu, each = r1), sd = sqrt(rep(v, each = r1)))
  factor.a <- factor(c(rep(1, each = n/2), rep(2, each = n/2)))
  factor.b <- factor(rep(c("A", "B"), each = r1,
	En el servidor en primer lugar se crean 3 funciones, una para cada prueba de bondad de ajuste a utilizar
	La primera obtiene el valor p con la prueba de Shapiro-Wilks, para esto usa un experimento planteado con 2 factores y 2 niveles cada uno, los parámetros para esta función son la cantidad de réplicas, las medias y las varianzas.

times = 2))
  mod <- lm(y ~ factor.a * factor.b)
  p <- shapiro.test(mod$res)$p.value
  return(p)
 }	

kol <- function (r1, mu, v) {
  k <- length(mu)
  n <- r1 * k
  y <- rnorm(n = n, mean = rep(mu, each = r1), sd = sqrt(rep(v, each = r1)))
  factor.a <- factor(c(rep(1, each = n/2), rep(2, each = n/2)))
  factor.b <- factor(rep(c("A", "B"), each = r1, times = 2))
  mod <- lm(y ~ factor.a * factor.b)
  p <- ks.test(mod$res, "pnorm", mean = mean(mod$res), sd = sd(mod$res))$p.value
  return(p)
 }
	La segunda función encuentra el valor p con la prueba Kolmogórov-Smirnov, de igual forma, los parámetros de la función son la cantidad de repeticiones, las medias y las varianzas

jb <- function (r1, mu, v) {
  k <- length(mu)
  n <- r1 * k
  y <- rnorm(n = n, mean = rep(mu, each = r1), sd = sqrt(rep(v, each = r1)))
  factor.a <- factor(c(rep(1, each = n/2), rep(2, each = n/2)))
  factor.b <- factor(rep(c("A", "B"), each = r1, times = 2))
  mod <- lm(y ~ factor.a * factor.b)
  p <- jarque.bera.test(mod$res)$p.value
  return(p)
 }
	Por último, la tercera función encuentra el valor p con la prueba Jaque-Bera, de igual forma, los parámetros de la función son la cantidad de repeticiones, las medias y las varianzas

observeEvent(input$run, {
  mu <- as.numeric(unlist(strsplit(input$mu, ",")))
  v <- as.numeric(unlist(strsplit(input$v, ",")))
  r_values <- seq(2, input$r)
	Establece la entrada de los datos que seleccionará el usuario, a las medias y a las varianzas se les aplican las funciones strsplit para que separe los valores por “,” y unlist para que los datos dejen de ser una lista.

results <- data.frame(Réplicas = integer(), Shapiro = double(), `Jarque Bera` = double(), Kolmogorov = double())	Se crea un dataframe el cual almacena la cantidad de réplicas y las 3 diferentes pruebas como columnas vacías, la cantidad de réplicas

for (r1 in r_values) {
M <- 1000
sha <- c()
jab <- c()
kolv <- c()
 	
for (j in 1:M) {
p1 <- sh(r1, mu, v)
p2 <- jb(r1, mu, v)
p3 <- kol(r1, mu, v)
sha <- c(sha, p1)
jab <- c(jab, p2)
kolv <- c(kolv, p3)
}
	permite datos tipo entero y las pruebas tipo numérico.
	Ahora se crean 2 ciclos for, en el primero se define la cantidad de iteraciones en 1000, luego se crean 3 objetos vacíos con los nombres de las pruebas, los cuales van a almacenar todos los valores p de las diferentes iteraciones. El segundo almacena los valores p de las pruebas utilizando las 3 funciones creadas anteriormente, para puedo usar los mismos 3 objetos del primer ciclo for, pero, ahora asignándole un vector con los objetos vacíos y los valores p de las funciones

results <- rbind(results, data.frame(Réplicas = r1, Shapiro = mean(sha < 0.05), `Jarque Bera` = mean(jab < 0.05), Kolmogorov = mean(kolv < 0.05))) }	Se genera un dataframe el cual almacena los resultados de la cantidad de réplicas y el promedio de las veces que se comete error tipo I, es decir, como el resultado de la condición “sha < 0.05” por ejemplo, son 0 y 1 se encuentra la proporción de veces que la prueba se rechaza.
output$results <- renderTable({ results })	Se establece la salida de la cuadro con los resultados, esta hace una primera columna con la cantidad de repeticiones que selecciona el usuario, luego, crea 3 columnas más en las que se encuentra la proporción de veces que se comete error tipo I en cada prueba. Por otro lado, se utiliza la función melt con el fin de presentar de forma “larga” los resultados.
output$plot <- renderPlotly({ tab2 <- melt(results, id.vars = "Réplicas", variable.name = "Pruebas", value.name = "Probabilidad") g1 <- ggplot(tab2, aes(x = Réplicas, y = Probabilidad, colour = Pruebas)) + geom_line() + ylab("Probabilidad Error Tipo I") + xlab("Número de réplicas") + theme_minimal() + scale_color_manual(values = c("#000066", "#FF9933", "#FF5733")) ggplotly(g1) })	Se genera el gráfico con los resultados de la cuadro anterior utilizando la función ggplot, sin embargo, luego se usa la función ggplotly con el fin de hacer un gráfico interactivo.

```
  })
}

# Correr la app
shinyApp(ui = ui, server = server)
```