

Comparación de las pruebas de homocedasticidad para muestras de diferente tamaño con distribución normal y exponencial

Jerimy Arias Rojas¹; Karla Espeleta García¹; Alison Lobo Salas¹
jerimy.arias@ucr.ac.cr, karla.espeleta@ucr.ac.cr, alison.lobo@ucr.ac.cr

RESUMEN

Se realizó un estudio de simulación con un factor de tres niveles con el objetivo de determinar cuál es el efecto que pueden tener muestras de diferente tamaño con distribución normal y no normal sobre diferentes pruebas de homocedasticidad, asumiendo que se cumple el supuesto de homocedasticidad. Los resultados permiten tener criterios para escoger la prueba más adecuada de acuerdo con el escenario que se tenga, donde se llega que, con la distribución normal, en tamaños de muestra grandes, es indiferente la prueba de homocedasticidad que se utilice y en la distribución exponencial, se recomienda usar las pruebas que tenga en cuenta la mediana para comprobar la igualdad de varianzas.

PALABRAS CLAVE: simulación, normalidad, Levene, Bartlett, Fligner-Killeen.

ABSTRACT

A simulation study was carried out with a three-level factor with the objective of determining the effect that samples of different sizes with normal and non-normal distribution can have on different homoscedasticity tests, assuming that the assumption of homoscedasticity is met. The results allow having criteria to choose the most appropriate test according to the scenario, where it is found that, with the normal distribution, in large sample sizes, the homoscedasticity test to be used is indifferent, and in the exponential distribution, it is recommended to use the tests that consider the median to prove the equality of variances.

KEY WORDS: simulation, normality, Levene, Bartlett, Fligner-Killeen.

INTRODUCCIÓN

En este artículo se abordó el efecto de la homocedasticidad sobre diferentes tamaños de muestras en una simulación con un factor de tres niveles. En primer lugar, se compararon dos escenarios, uno en que el sí se cumple el supuesto de normalidad y otro en el que se utiliza una distribución exponencial, en ambos casos se aplicaron todas las pruebas que se mencionan más adelante.

En el análisis estadístico uno de los supuestos más relevantes es el de homogeneidad de varianzas, conocido como homocedasticidad. La validez de las estimaciones y pruebas de hipótesis se apoyan de suposiciones, por ejemplo, que los errores experimentales aleatorios son

¹ Estudiantes de Estadística de la Universidad de Costa Rica.

independientes, que siguen una distribución normal, tienen una media igual a cero e igualdad de varianza para todos los tratamientos. Cuando uno de estos supuestos no se puede probar, afecta las estimaciones de las pruebas de significancia obtenidas del análisis de varianza (ANOVA). Como mencionan Correa, Iral & Rojas (2006), este supuesto es crucial para garantizar la calidad de los procedimientos estadísticos utilizados tanto en pruebas de hipótesis como en la construcción de intervalos de confianza.

Existen pruebas para verificar si el supuesto de homocedasticidad es o no admisible. En esta simulación se quiere comparar las pruebas de Levene por el método de la media y la mediana, Bartlett, Brown-Forsythe, Breusch-Pagan, Fligner-Killeen. La prueba de Levene se usa para la suposición de varianzas iguales para las muestras con datos que provienen de una distribución normal. La prueba Bartlett se utiliza cuando hay seguridad de que los datos muestrales provienen de una distribución normal. La prueba Brown-Forsythe es una modificación de la Levene, pero difieren en que esta intenta corregir la asimetría mediante el uso de desviaciones de las medianas de los grupos. La Breusch-Pagan es una prueba formal para determinar si existe evidencia de heterocedasticidad para muestras grandes, la cual asume que los errores son independientes y normales. Por último, la prueba Fligner-Killeen se trata de una prueba no paramétrica que compara las varianzas basándose en la mediana, además, es una alternativa cuando no se cumple el supuesto de normalidad en las muestras.

Como parte de la importancia del cumplimiento de la homocedasticidad y normalidad, Jiménez, Barrantes & González (2020) encontraron que al cumplirse el supuesto de homocedasticidad existen ventajas en cuanto al incremento de la potencia de la prueba para un número idóneo de réplicas por tratamiento y, al incumplirse la homocedasticidad, no implica que sea imposible obtener una potencia deseable, sin embargo, se hace necesario un mayor número de réplicas por cada tratamiento. El cumplimiento del supuesto de normalidad tiene ventaja, pero su incumplimiento ocasiona una ineficiencia del estimador mínimo-cuadrático o estimador de la curva de mínima varianza y, por otro, los intervalos de confianza de los parámetros del modelo y los controles de significancia son solamente aproximados y no exactos (Tapia & Cellavos, 2021).

En estudios como el de Amat (2016) se encontró que si se tiene seguridad de que las muestras a comparar proceden de poblaciones que siguen una distribución normal, son recomendables la prueba de Bartlett, y si no se tiene la seguridad de que las poblaciones de origen son normales, se recomienda la prueba de Leven utilizando la mediana o la prueba no paramétrica Fligner-Killeen que también se basa en la mediana. Por el contrario, McDonald (s.f.) menciona que la prueba de Bartlett no es particularmente buena, porque es sensible a las desviaciones de la normalidad, así como a la de heterocedasticidad, sin embargo, concuerda con Amat (2016) mencionando que una alternativa de la Bartlett cuando no hay normalidad es utilizar la prueba de Levene, ya que es menos sensible a las desviaciones de la normalidad, pero si los datos son aproximadamente normales, es menos potente que la prueba de Bartlett.

El objetivo general de esta investigación es determinar cuál es el efecto que pueden tener muestras de diferente tamaño con distribución normal y no normal sobre diferentes pruebas de homocedasticidad, bajo condición de homocedasticidad.

METODOLOGÍA

Para esta simulación se utilizó un diseño factorial simple con un factor con tres niveles. El modelo de este se define como:

$$\mu_i = \mu + \alpha_i;$$

donde:

μ_i : media del tratamiento i

μ : media general

α_i : efecto del tratamiento i -ésimo

Para obtener los datos de este estudio se simularon seis muestras, tres de estas con distribución normal y tres distribuidas de forma exponencial. Para distinguir cada una, se utilizaron tres distintos tamaños de muestras por tratamiento, de 5, 15 y 35, y al ser tres tratamientos, se llega a que los tamaños de muestras son 15, 45 y 105, considerados como tamaño pequeño, grande y muy grande respectivamente.

Se asume que existe igualdad de varianzas para el caso de la muestra distribuida normalmente, por lo cual, se utilizó la misma varianza y media para cada nivel, y para la distribución exponencial también.

Para las pruebas de homocedasticidad aplicadas para la simulación se eligieron cinco, las cuales son: la prueba de Levene por el método de la media y la mediana, Bartlett, Brown-Forsythe, Breusch-Pagan y Fligner-Killeen.

La prueba de Levene se basa en el estadístico W , el cual se define como:

$$W = \frac{(n - k) \sum_{i=1}^k n_i (Z_{i.} - Z_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - Z_{i.})^2};$$

donde:

n : número de casos totales

k : número de grupos

n_i : número de casos en el grupo i

Z_{ij} : $|Y_{ij} - \underline{Y}_i|$, donde \underline{Y}_i es la media del i -ésimo grupo o $|Y_{ij} - Y_{\sim i}|$, donde $Y_{\sim i}$ es la mediana del i -ésimo grupo

Este estadístico se comparó contra una prueba $F_{k-1, n-k}$ con un nivel de significancia de α , definido como 0.05 para esta simulación.

Para la prueba de Brown-Forsythe se utilizó una modificación de la prueba de Levene, utilizando desviaciones de las medianas para intentar corregir la asimetría. Esta prueba es robusta ante desviaciones de normalidad y no requiere igualdad de tamaño de muestra.

Para el caso de la prueba de Bartlett, utiliza el estadístico T, que se define como:

$$T = \frac{(n - k) \ln s_p^2 - \sum_{i=1}^k (n_i - 1) \ln s_i^2}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{n - k} \right)}$$

donde:

n : número de casos totales

n_i : número de casos en el grupo i

k : número de grupos

s_p^2 : estimación combinada de la varianza

s_i^2 : varianza de la muestra

Este estadístico se comparó contra una prueba χ_{k-1}^2 con un nivel de significancia de α , definido como 0.05 para esta simulación. Si se conoce que no se cumple el supuesto de normalidad o se duda del mismo, se recomienda no utilizar esta prueba al ser sensible al incumplimiento de este supuesto.

Cuando se tienen covariables se usa esta prueba, pues la hipótesis nula cambia levemente, ya que se convierte en decir que la varianza condicional de la distribución de la respuesta o del error es la misma para cada posible combinación de los predictores.

La prueba de Breusch-Pagan se utiliza para modelos de regresión. En este caso, se utiliza:

$$\chi^2 = \frac{\frac{SCReg}{2}}{\left(\frac{SCRes}{n}\right)^2};$$

donde:

n : número de casos

$SCReg$: Suma de Cuadrados de Regresión de la regresión de los errores cuadráticos sobre X

$SCRes$: Suma de Cuadrados Residual de la Regresión de Y sobre X

Este estadístico se distribuye como una función Chi-cuadrado, χ_k^2 , donde k es el número de predictores. Además, tal como los otros casos, el nivel de significancia de α ha sido definido como 0.05 para esta simulación.

Por último, la prueba de Fligner-Killeen es otra prueba, que se caracteriza por ser no paramétrico, y esta compara las varianzas basándose en la mediana. Se obtiene de la siguiente manera:

Se realiza la diferencia absoluta entre el dato y la mediana de las observaciones, dado como $|X_{ij} - \tilde{X}_i|$, y se ordenan de mayor a menor. Además, se define la variable a , dado como:

$$a_{n,i} = \varphi^{-1} \left(\frac{1}{2} + \frac{i}{2(n+1)} \right);$$

donde:

φ^{-1} : percentil 100 de la distribución $N(0,1)$

n : número de casos totales

Además, se tiene \underline{a}_i , denotado como:

$$\underline{a}_i = \sum_{j \in H} a_{n,j};$$

donde:

H : muestra de la población

y se tiene la variable \underline{a} , obtenida como:

$$\underline{a} = \sum_{j=1}^n \frac{a_{n,j}}{n}$$

Por último, se denota el estadístico C como:

$$C = \frac{\sum_{i=1}^k n_i (\underline{a}_i - \underline{a})}{\sum_{j=1}^n \frac{(a_{n,j} - \underline{a})^2}{(n-1)}}$$

Este estadístico aproximadamente se distribuye χ_{k-1}^2 . Esta prueba es menos sensible a desviaciones de la normalidad.

A partir de todas las pruebas de homocedasticidad antes mencionadas, se extrajo el valor p mostrado y, con esto, se hizo el análisis respectivo. Se usaron 1000 réplicas por cada tamaño de muestra para obtener los resultados.

El análisis de los datos y las pruebas se realizaron con la ayuda del Software estadístico R versión 4.2.3 (R Core Team, 2023). Las librerías que se utilizan en este estudio son: car (Fox, J. & Weisberg, S., 2019), lmtest (Hothorn, T. & Zeileis, A., 2022) y HH (Heiberger, R., 2023).

RESULTADOS

Primeramente, se utilizaron distintos tamaños de muestra para ver el comportamiento de las diferentes pruebas de homocedasticidad en distribuciones normales y exponenciales, teniendo igualdad de varianzas. En el *Cuadro 1*, con una distribución normal, se observa que en todas las pruebas se cumple el supuesto de homocedasticidad, esto indica que la probabilidad real de cometer un error tipo I en los contrastes de hipótesis es menor, ya que la inferencia es por lo general más conservadora ante distribuciones normales, independientemente del tamaño de muestra. Se puede deducir que cuando se cumple normalidad en los datos no hay diferencias importantes entre estas pruebas de homocedasticidad, por lo que bajo este supuesto se podría utilizar cualquiera de las pruebas mencionadas en el trabajo.

Además de esto, se puede mencionar también que en las pruebas de Fligner-Killeen, Levene por el método de la mediana y la Brown – Forsythe, cuando la muestra es pequeña, identificaron la homocedasticidad de manera casi al 100%, algo que por lo general no es tan real, en caso de realizar un estudio.

Cuadro 1.

Porcentaje de comprobación de homocedasticidad en diferentes pruebas por cada tamaño de muestras con distribución normal

Pruebas	Tamaño de Muestra		
	n = 15	n = 45	n = 105
Fligner-Killeen	99.9	96.9	96.7
Levene - Método Mediana	99.1	97.6	96.5
Brown - Forsythe	99.1	97.6	96.5
Breusch-Pagan	94.9	95.5	95.3
Levene - Método Media	91.0	97.6	95.2
Bartlett	94.2	94.8	95.0

En el caso contrario, cuando los datos provienen de una distribución exponencial con tamaños de muestra pequeños, se puede notar que la prueba de Bartlett es principalmente más sensible, por lo cual, y como se puede observar en el *Cuadro 2*, con 1000 réplicas fijas y un tamaño de muestra 15, el 74% de las veces se comprueba homocedasticidad. De igual manera, en el caso de

un tamaño de muestra de 45, se obtiene la comprobación de homocedasticidad de la prueba de Bartlett en un 63% de las veces.

Por último, con tamaños de muestra muy grandes, por ejemplo, en este caso de 105, se observa que se comprueba homocedasticidad en el 59% de las ocasiones en la prueba de Bartlett, en consecuencia, se puede decir que para tamaños de muestra grandes que provienen de distribuciones exponenciales, la prueba de Bartlett es la menos adecuada, ya que lo ideal es utilizarla en distribuciones normales.

Cuadro 2.

Porcentaje de comprobación de homocedasticidad en diferentes pruebas por cada tamaño de muestras con distribución exponencial

Pruebas	Tamaño de Muestra		
	n = 15	n = 45	n = 105
Breusch-Pagan	90.0	92.8	95.4
Levene - Método Mediana	98.8	96.5	95.3
Brown - Forsythe	98.6	96.5	95.3
Fligner-Killeen	99.9	93.8	89.7
Levene - Método Media	80.7	80.0	81.9
Bartlett	73.9	62.7	59.1

Por último, la prueba de Levene se caracteriza por permitir elegir entre diferentes estadísticos de centralidad, como la mediana (por defecto) o la media. Es importante mencionar que, a la hora de contrastar la homocedasticidad, va a depender si los datos se distribuyen de forma normal o no. En el caso de los datos con distribuciones normales, no se ven diferencias entre la prueba de Levene por método de la media, mientras que, al comparar el tamaño de muestra en distribuciones exponenciales, se aprecia que la prueba de Levene por el método de la media es más sensible ante la falta de normalidad. Por lo tanto, se recomienda emplear la prueba de Levene por el método de la mediana cuando los datos no siguen una distribución normal o no se cumplen los supuestos de normalidad.

CONCLUSIONES

En el análisis estadístico uno de los supuestos que es necesario garantizar, es la homogeneidad de varianzas, para hacer dicho análisis se proponen distintas pruebas bajo distribuciones normales y no normales con distintos tamaños de muestra, con el fin de determinar cuál prueba es adecuada dependiendo del escenario que se tenga. Los diferentes métodos gráficos son ayudas visuales, por ejemplo, gráficos de cajas, que nos permiten visualizar de forma rápida ciertas características poblacionales y nos dan una idea de la variabilidad de los datos, sin embargo, no son suficientes por sí solos. A partir de este trabajo de simulación donde se tuvieron distintos tamaños de muestra para cada comparación, se hacen las siguientes recomendaciones:

En distribuciones normales se observó que no es relevante el tamaño de muestra, ya que las pruebas no detectaron problemas. Sin embargo, cuando se sabe de antemano que los datos se distribuyen de forma normal, se recomienda utilizar la prueba de Bartlett ya que, como destaca Amat (2016), es más potente bajo condiciones de normalidad.

En el caso de distribuciones no normales con tamaños de muestra pequeños, hay que tener precaución a la hora de hacer el análisis de homocedasticidad, ya que puede que no se identifique de manera correcta por tener pocos datos. Se observó que, con muestras pequeñas en distribuciones exponenciales, la prueba de Bartlett es sensible a desviaciones de normalidad por lo que no es recomendado utilizarla, una alternativa en casos donde se tienen datos que provienen de distribuciones no normales, es la prueba de Levene utilizando la mediana o la prueba no paramétrica de Fligner-Killeen que también se basa en la mediana (Amat, 2016).

Se sugiere llevar a cabo la misma simulación, con la particularidad de colocar diferentes medias por cada uno de los tratamientos en las muestras con distribución normal y exponencial, esto con el fin de abarcar situaciones y circunstancias que puedan ser diferentes a las observadas en la simulación planteada, y revisar si está particularidad afecta de manera positiva o negativa la aplicación de las pruebas de homocedasticidad.

BIBLIOGRAFÍA

- Amat, J. (2016). *Análisis de la homogeneidad de varianza (homocedasticidad)*. [https://cienciadedatos.net/documentos/9_homogeneidad_de_varianza_homocedasticidad#Test de Bartlett](https://cienciadedatos.net/documentos/9_homogeneidad_de_varianza_homocedasticidad#Test%20de%20Bartlett)
- Jiménez, P., Barrantes, S., & González, M. (2020) Efecto del uso de mínimos cuadrados ponderados en la potencia de la prueba de hipótesis para diferencias de medias, cuando se incumple el supuesto de homocedasticidad. *SERENQUETI*, 17-24. <https://www.kerwa.ucr.ac.cr/bitstream/handle/10669/82755/Revista%20Serengueti%20-%20Vol.2%20-%20No.1%20-%20Dic.%202020.pdf?sequence=1#page=17>
- Correa, J., Iral, R. & Rojas, L. (2006). Estudio de potencia de pruebas de homogeneidad de varianza. *Revista Colombiana de Estadística*. 29(1). 57-76. https://www.researchgate.net/publication/262736310_A_Study_of_the_Power_of_Tests_for_Homogeneity_of_Variance
- McDonald, J. (s.f.). Homocedasticidad y Heterocedasticidad. *Estadísticas Biológicas (McDonald)*. Universidad de Delaware [https://espanol.libretexts.org/Estadisticas/Estadistica_Aplicada/Libro%3A_Estadisticas_Biologicas_\(McDonald\)/04%3A_Pruebas_para_una_variable_de_medici%C3%B3n/4.05%3A_Homocedasticidad_y_Heterocedasticidad](https://espanol.libretexts.org/Estadisticas/Estadistica_Aplicada/Libro%3A_Estadisticas_Biologicas_(McDonald)/04%3A_Pruebas_para_una_variable_de_medici%C3%B3n/4.05%3A_Homocedasticidad_y_Heterocedasticidad)
- Tapia, C. & Cellavos, K. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: anderson-darling, ryan-joiner, shapiro-wilk y kolmogórov-smirnov. *Societas*, 23(2), 83-106. <http://portal.amelica.org/ameli/jatsRepo/341/3412237018/html/index.html>

ANEXOS

Código	Explicación
<pre>library(car) library(HH) library(lmtest)</pre>	<p>Librerías que se utilizaron para realizar la simulación.</p>
<pre>#-----MUESTRAS NORMALES-----# dnormal = function(n, mu, var) { x = rep(1:3, each=n) y1 = rnorm(n,mu,var[1]) y2 = rnorm(n,mu,var[2]) y3 = rnorm(n,mu,var[3]) y = c(y1,y2,y3) sim_normal = as.data.frame(cbind(y,x)) sim_normal\$x = factor(sim_normal\$x) return(sim_normal) }</pre>	<p>En este caso, se realizó una función llamada <i>dnormal</i> para poder obtener muestras con distribución normal. Donde <i>n</i> se refiere al tamaño de muestras por tratamiento, <i>mu</i> es la media y <i>var</i> es la varianza. Por lo cual, el tamaño de la muestra sería $n*3$, al ser 3 tratamientos.</p> <p>Se utiliza <i>x</i> como el factor con 3 niveles donde la función <i>rep</i> para asignar cada nivel, y <i>y1</i>, <i>y2</i>, <i>y3</i> son los datos que se tienen como variable respuesta, para este caso, se utilizó la función <i>rnorm</i> para tener datos aleatorios pero distribuidos normalmente con número de muestra, media y varianza definidas como <i>n</i>, <i>mu</i> y <i>var</i> respectivamente.</p> <p>Se unieron estas muestras en un vector llamado <i>y</i>, la cual se agregó junto a <i>x</i> en un <i>dataframe</i> y se convirtió el factor <i>x</i> como un factor. Se usó <i>return</i> para obtener los datos.</p>
<pre>#-----MUESTRAS EXPONENCIALES-----# dexpo = function(n, b){ x = rep(1:3, each=n) y1 = rexp(n,1/b[1]) y2 = rexp(n,1/b[2]) y3 = rexp(n,1/b[3]) y = c(y1,y2,y3) sim_exponencial = as.data.frame(cbind(y,x)) sim_exponencial\$x = factor(sim_exponencial\$x) return(sim_exponencial) }</pre>	<p>En este caso, se realizó una función llamada <i>dexpo</i> para poder obtener muestras con distribución exponencial. Donde <i>n</i> se refiere al tamaño de muestras por tratamiento y <i>b</i> es la media. Por lo cual, el tamaño de la muestra sería $n*3$, al ser 3 tratamientos.</p> <p>Se utiliza <i>x</i> como el factor con 3 niveles donde la función <i>rep</i> para asignar cada nivel, y <i>y1</i>, <i>y2</i>, <i>y3</i> son los datos que se tienen como variable respuesta, para este caso, se utilizó la función <i>rexp</i> para tener datos aleatorios pero distribuidos exponencialmente con número de muestra y media definidas como <i>n</i> y <i>b</i> respectivamente.</p> <p>Se unieron estas muestras en un vector llamado <i>y</i>, la cual se agregó junto a <i>x</i> en un</p>

	<p><i>dataframe</i> y se convirtió el factor <i>x</i> como un factor. Se usó <i>return</i> para obtener los datos.</p>
<pre>#-----MUESTRAS PEQUEÑAS-----# plevenemedian = c() plevenemean = c() pbf = c() pb = c() pfk = c() pbp = c() r=1000 for (i in 1:r) { datosnormal = dnormal(5,20,c(25,25,25)) #Aplicacion de pruebas lmean = leveneTest(y = datosnormal\$y, group = datosnormal\$x, center = "mean") lmedian = leveneTest(y = datosnormal\$y, group = datosnormal\$x, center = "median") bf = hov(y~x, data = datosnormal) b = bartlett.test(y~x, data = datosnormal) fk = fligner.test(y~x, data = datosnormal) bp = bptest(y~x, data = datosnormal) #Extracción del p-value plevenemean[i] = lmean\$`Pr(>F)`[1] plevenemedian[i] = lmedian\$`Pr(>F)`[1] pbf[i] = bf\$p.value pb[i] = b\$p.value pfk[i] = fk\$p.value pbp[i] = bp\$p.value } #Tabla Homocedasticidad-Heterocedasticidad #True = Hay homocedasticidad, False = Hay heterocedasticidad t1=table(plevenemean>0.05)*100/r t2=table(plevenemedian>0.05)*100/r t3=table(pbf>0.05)*100/r t4=table(pb>0.05)*100/r t5=table(pfk>0.05)*100/r t6=table(pbp>0.05)*100/r tt = rbind(t1,t2,t3,t4,t5,t6) rownames(tt)= c("Levene - Metodo Media", "Levene - Metodo Mediana", "Brown- Forsythe", "Bartlett", "Fligner-Killeen", "Breusch-Pagan")</pre>	<p>Se crearon vectores donde se van a acumular los valores <i>p</i> que se obtienen a partir de cada una de las pruebas de homocedasticidad.</p> <p>Se utilizaron 1000 repeticiones. Por eso <i>r</i> = 1000.</p> <p>Se hace una función <i>for</i> para 1000 repeticiones donde se obtiene los datos que se hicieron en la función <i>dnormal</i> donde se usa <i>n=5</i>, <i>mu=20</i> y <i>var=c(25,25,25)</i>; el tamaño de muestra es pequeña (15) y como se asume que existe homocedasticidad, se coloca la misma varianza para todos los casos.</p> <p>Se aplican las pruebas definidas como:</p> <ul style="list-style-type: none"> - Levene por método de media: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosnormales\$y</i>), <i>groups</i> como el factor (<i>datosnormales\$x</i>) y <i>center</i> como <i>mean</i> (media). Se guardan como <i>lmean</i>. - Levene por método de mediana: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosnormales\$y</i>), <i>groups</i> como el factor (<i>datosnormales\$x</i>) y <i>center</i> como <i>median</i> (mediana). Se guardan como <i>lmedian</i>. - Brown–Forsythe: se utiliza la función <i>hov</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>bf</i>. - Bartlett: se usa la función <i>bartlett.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>b</i>. - Fligner-Killeen: se utiliza la función <i>fligner.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>fk</i>.

<pre>colnames(tt)= c("Heterocedasticidad", "Homocedasticidad"); tt</pre>	<p>- Breusch-Pagan: se utiliza la función <i>bptest</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>bp</i>.</p> <p>A cada uno de estos se les extrae el valor p correspondiente y se acumulan en los vectores mencionados al principio.</p> <p>Esto se realiza por 1000 veces.</p> <p>Después, se comprueban cuales valores que se tienen en los vectores son mayores a 0.05 para saber si se cumple homocedasticidad según cada prueba y sus respectivos métodos de cálculo (se pone como <i>TRUE</i> para homocedasticidad Y <i>FALSE</i> si son heterocedásticos). Además, se multiplica por 100 y se divide entre <i>r</i> para tener su cantidad porcentual.</p> <p>Se hace una tabla con los datos recolectados.</p>
<pre>plevenemedianexp = c() plevenemeanexp = c() pbfxp = c() pbexp = c() pfxp = c() pbpxp = c() r=1000 for (i in 1:r) { datosexp = dexpo(5,c(20,20,20)) #Aplicacion de pruebas lmeanexp = leveneTest(y = datosexp\$y, group = datosexp\$x, center = "mean") lmedianexp = leveneTest(y = datosexp\$y, group = datosexp\$x, center = "median") bfxp = hov(y~x, data = datosexp) bexp = bartlett.test(y~x, data = datosexp) fexp = fligner.test(y~x, data = datosexp) bpxp = bptest(y~x, data = datosexp) #Extracción del p-value plevenemeanexp[i] = lmeanexp\$`Pr(>F)`[1] plevenemedianexp[i] = lmedianexp\$`Pr(>F)`[1]</pre>	<p>Se crearon vectores donde se van a acumular los valores p que se obtienen a partir de cada una de las pruebas de homocedasticidad.</p> <p>Se utilizaron 1000 repeticiones. Por eso $r = 1000$.</p> <p>Se hace una función <i>for</i> para 1000 repeticiones donde se obtiene los datos que se hicieron en la función <i>dexpo</i> donde se usa $n=5$ y $b=20$; el tamaño de muestra es pequeña (15) y como se asume que existe homocedasticidad, se coloca la misma <i>b</i> para todos los casos.</p> <p>Se aplican las pruebas definidas como:</p> <ul style="list-style-type: none"> - Levene por método de media: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosexp\$y</i>), <i>groups</i> como el factor (<i>datosexp\$x</i>) y <i>center</i> como <i>mean</i> (media). Se guardan como <i>lmean</i>. - Levene por método de mediana: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la

<pre> pbfexp[i] = bfexp\$p.value pbexp[i] = bexp\$p.value pfkexp[i] = fkexp\$p.value pbpexp[i] = bpexp\$p.value } #True = Hay homocedasticidad, False = Hay heterocedasticidad te1=table(plevenemeanexp>0.05)*100/r te2=table(plevenemedianexp>0.05)*100/r te3=table(pbfexp>0.05)*100/r te4=table(pbexp>0.05)*100/r te5=table(pfkexp>0.05)*100/r te6=table(pbpexp>0.05)*100/r tte = rbind(te1,te2,te3,te4,te5,te6) rownames(tte)= c("Levene - Metodo Media", "Levene - Metodo Mediana", "Brown- Forsythe", "Bartlett", "Fligner-Killeen", "Breusch-Pagan") colnames(tte)= c("Heterocedasticidad", "Homocedasticidad"); tte </pre>	<p>variable respuesta (<i>datosexp\$y</i>), <i>groups</i> como el factor (<i>datosexp\$x</i>) y <i>center</i> como <i>median</i> (mediana). Se guardan como <i>lmedian</i>.</p> <ul style="list-style-type: none"> - Brown–Forsythe: se utiliza la función <i>hov</i>, donde se coloca $y^{\sim}x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>bf</i>. - Bartlett: se usa la función <i>bartlett.test</i>, donde se coloca $y^{\sim}x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>b</i>. - Fligner-Killeen: se utiliza la función <i>fligner.test</i>, donde se coloca $y^{\sim}x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>fk</i>. - Breusch-Pagan: se utiliza la función <i>bptest</i>, donde se coloca $y^{\sim}x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>bp</i>. <p>A cada uno de estos se les extrae el valor <i>p</i> correspondiente y se acumulan en los vectores mencionados al principio.</p> <p>Esto se realiza por 1000 veces.</p> <p>Después, se comprueban cuales valores que se tienen en los vectores son mayores a 0.05 para saber si se cumple homocedasticidad según cada prueba y sus respectivos métodos de cálculo (se pone como <i>TRUE</i> para homocedasticidad Y <i>FALSE</i> si son heterocedásticos). Además, se multiplica por 100 y se divide entre <i>r</i> para tener su cantidad porcentual.</p> <p>Se hace una tabla con los datos recolectados.</p>
<pre> #-----MUESTRAS GRANDES-----# plevenemedian = c() plevenemean = c() pbf = c() pb = c() pfk = c() pbp = c() r=1000 for (i in 1:r) { </pre>	<p>Se crearon vectores donde se van a acumular los valores <i>p</i> que se obtienen a partir de cada una de las pruebas de homocedasticidad.</p> <p>Se utilizaron 1000 repeticiones. Por eso <i>r</i> = 1000.</p>

<pre> datosnormal = dnormal(15,20,c(25,25,25)) #Aplicacion de pruebas lmean = leveneTest(y = datosnormal\$y, group = datosnormal\$x, center = "mean") lmedian = leveneTest(y = datosnormal\$y, group = datosnormal\$x, center = "median") bf = hov(y~x, data = datosnormal) b = bartlett.test(y~x, data = datosnormal) fk = fligner.test(y~x, data = datosnormal) bp = bptest(y~x, data = datosnormal) #Extracción del p-value plevenemean[i] = lmean\$`Pr(>F)`[1] plevenemedian[i] = lmedian\$`Pr(>F)`[1] pbf[i] = bf\$p.value pb[i] = b\$p.value pfk[i] = fk\$p.value pbp[i] = bp\$p.value } #Tabla Homocedasticidad-Heterocedasticidad #True = Hay homocedasticidad, False = Hay heterocedasticidad t1=table(plevenemean>0.05)*100/r t2=table(plevenemedian>0.05)*100/r t3=table(pbf>0.05)*100/r t4=table(pb>0.05)*100/r t5=table(pfk>0.05)*100/r t6=table(pbp>0.05)*100/r tt = rbind(t1,t2,t3,t4,t5,t6) rownames(tt)= c("Levene - Metodo Media", "Levene - Metodo Mediana", "Brown- Forsythe", "Bartlett", "Fligner-Killeen", "Breusch-Pagan") colnames(tt)= c("Heterocedasticidad", "Homocedasticidad"); tt </pre>	<p>Se hace una función <i>for</i> para 1000 repeticiones donde se obtiene los datos que se hicieron en la función <i>dnormal</i> donde se usa $n=15$, $\mu=20$ y $var=c(25,25,25)$; el tamaño de muestra es grande (45) y como se asume que existe homocedasticidad, se coloca la misma varianza para todos los casos.</p> <p>Se aplican las pruebas definidas como:</p> <ul style="list-style-type: none"> - Levene por método de media: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosnormales\$y</i>), <i>groups</i> como el factor (<i>datosnormales\$x</i>) y <i>center</i> como <i>mean</i> (media). Se guardan como <i>lmean</i>. - Levene por método de mediana: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosnormales\$y</i>), <i>groups</i> como el factor (<i>datosnormales\$x</i>) y <i>center</i> como <i>median</i> (mediana). Se guardan como <i>lmedian</i>. - Brown–Forsythe: se utiliza la función <i>hov</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>bf</i>. - Bartlett: se usa la función <i>bartlett.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>b</i>. - Fligner-Killeen: se utiliza la función <i>fligner.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>fk</i>. - Breusch-Pagan: se utiliza la función <i>bptest</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>bp</i>. <p>A cada uno de estos se les extrae el valor p correspondiente y se acumulan en los vectores mencionados al principio.</p> <p>Esto se realiza por 1000 veces.</p> <p>Después, se comprueban cuales valores que se tienen en los vectores son mayores a 0.05 para saber si se cumple homocedasticidad según</p>
---	---

	<p>cada prueba y sus respectivos métodos de cálculo (se pone como <i>TRUE</i> para homocedasticidad Y <i>FALSE</i> si son heterocedásticos). Además, se multiplica por 100 y se divide entre <i>r</i> para tener su cantidad porcentual.</p> <p>Se hace una tabla con los datos recolectados.</p>
<pre>plevenemedianexp = c() plevenemeanexp = c() pbfxp = c() pbexp = c() pfxp = c() pbpxp = c() r=1000 for (i in 1:r) { datosexp = dexpo(15, c(20,20,20)) #Aplicacion de pruebas lmeanexp = leveneTest(y = datosexp\$y, group = datosexp\$x, center = "mean") lmedianexp = leveneTest(y = datosexp\$y, group = datosexp\$x, center = "median") bfxp = hov(y~x, data = datosexp) bexp = bartlett.test(y~x, data = datosexp) fexp = fligner.test(y~x, data = datosexp) bpxp = bptest(y~x, data = datosexp) #Extracción del p-value plevenemeanexp[i] = lmeanexp\$`Pr(>F)`[1] plevenemedianexp[i] = lmedianexp\$`Pr(>F)`[1] pbfxp[i] = bfxp\$p.value pbexp[i] = bexp\$p.value pfxp[i] = fexp\$p.value pbpxp[i] = bpxp\$p.value } #True = Hay homocedasticidad, False = Hay heterocedasticidad te1=table(plevenemeanexp>0.05)*100/r te2=table(plevenemedianexp>0.05)*100/r te3=table(pbfexp>0.05)*100/r te4=table(pbexp>0.05)*100/r te5=table(pfxp>0.05)*100/r te6=table(pbpxp>0.05)*100/r</pre>	<p>Se crearon vectores donde se van a acumular los valores p que se obtienen a partir de cada una de las pruebas de homocedasticidad.</p> <p>Se utilizaron 1000 repeticiones. Por eso <i>r</i> = 1000.</p> <p>Se hace una función <i>for</i> para 1000 repeticiones donde se obtiene los datos que se hicieron en la función <i>dexpo</i> donde se usa <i>n=15</i> y <i>b=20</i>; el tamaño de muestra es grande (45) y como se asume que existe homocedasticidad, se coloca la misma <i>b</i> para todos los casos.</p> <p>Se aplican las pruebas definidas como:</p> <ul style="list-style-type: none"> - Levene por método de media: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosexp\$y</i>), <i>groups</i> como el factor (<i>datosexp\$x</i>) y <i>center</i> como <i>mean</i> (media). Se guardan como <i>lmean</i>. - Levene por método de mediana: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosexp\$y</i>), <i>groups</i> como el factor (<i>datosexp\$x</i>) y <i>center</i> como <i>median</i> (mediana). Se guardan como <i>lmedian</i>. - Brown–Forsythe: se utiliza la función <i>hov</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>bf</i>. - Bartlett: se usa la función <i>bartlett.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>b</i>.

<pre> tte = rbind(te1,te2,te3,te4,te5,te6) rownames(tte)= c("Levene - Metodo Media", "Levene - Metodo Mediana", "Brown- Forsythe", "Bartlett", "Fligner-Killeen", "Breusch-Pagan") colnames(tte)= c("Heterocedasticidad", "Homocedasticidad"); tte </pre>	<p>- Fligner-Killeen: se utiliza la función <i>fligner.test</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>fk</i>.</p> <p>- Breusch-Pagan: se utiliza la función <i>bptest</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>bp</i>.</p> <p>A cada uno de estos se les extrae el valor p correspondiente y se acumulan en los vectores mencionados al principio.</p> <p>Esto se realiza por 1000 veces.</p> <p>Después, se comprueban cuales valores que se tienen en los vectores son mayores a 0.05 para saber si se cumple homocedasticidad según cada prueba y sus respectivos métodos de cálculo (se pone como <i>TRUE</i> para homocedasticidad Y <i>FALSE</i> si son heterocedásticos). Además, se multiplica por 100 y se divide entre <i>r</i> para tener su cantidad porcentual.</p> <p>Se hace una tabla con los datos recolectados.</p>
<pre> #-----MUESTRAS MUY GRANDES-----# plevenemedian = c() plevenemean = c() pbf = c() pb = c() pfk = c() pbp = c() r=1000 for (i in 1:r) { datosnormal = dnormal(35,20,c(25,25,25)) #Aplicacion de pruebas lmean = leveneTest(y = datosnormal\$y, group = datosnormal\$x, center = "mean") lmedian = leveneTest(y = datosnormal\$y, group = datosnormal\$x, center = "median") bf = hov(y~x, data = datosnormal) b = bartlett.test(y~x, data = datosnormal) fk = fligner.test(y~x, data = datosnormal) bp = bptest(y~x, data = datosnormal) </pre>	<p>Se crearon vectores donde se van a acumular los valores p que se obtienen a partir de cada una de las pruebas de homocedasticidad.</p> <p>Se utilizaron 1000 repeticiones. Por eso <i>r</i> = 1000.</p> <p>Se hace una función <i>for</i> para 1000 repeticiones donde se obtiene los datos que se hicieron en la función <i>dnormal</i> donde se usa <i>n=35</i>, <i>mu=20</i> y <i>var=c(25,25,25)</i>; el tamaño de muestra es muy grande (105) y como se asume que existe homocedasticidad, se coloca la misma varianza para todos los casos.</p> <p>Se aplican las pruebas definidas como:</p> <ul style="list-style-type: none"> - Levene por método de media: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosnormales\$y</i>), <i>groups</i>

<pre> #Extracción del p-value plevenemean[i] = lmean\$`Pr(>F)`[1] plevenemedian[i] = lmedian\$`Pr(>F)`[1] pbf[i] = bf\$p.value pb[i] = b\$p.value pfk[i] = fk\$p.value pbp[i] = bp\$p.value } #Tabla Homocedasticidad-Heterocedasticidad #True = Hay homocedasticidad, False = Hay heterocedasticidad t1=table(plevenemean>0.05)*100/r t2=table(plevenemedian>0.05)*100/r t3=table(pbf>0.05)*100/r t4=table(pb>0.05)*100/r t5=table(pfk>0.05)*100/r t6=table(pbp>0.05)*100/r tt = rbind(t1,t2,t3,t4,t5,t6) rownames(tt)= c("Levene - Metodo Media", "Levene - Metodo Mediana", "Brown- Forsythe", "Bartlett", "Fligner-Killeen", "Breusch-Pagan") colnames(tt)= c("Heterocedasticidad", "Homocedasticidad"); tt </pre>	<p>como el factor (<i>datosnormales\$x</i>) y <i>center</i> como <i>mean</i> (media). Se guardan como <i>lmean</i>.</p> <ul style="list-style-type: none"> - Levene por método de mediana: se usa la función <i>leveneTest</i>, donde se coloca <i>y</i> como la variable respuesta (<i>datosnormales\$y</i>), <i>groups</i> como el factor (<i>datosnormales\$x</i>) y <i>center</i> como <i>median</i> (mediana). Se guardan como <i>lmedian</i>. - Brown–Forsythe: se utiliza la función <i>hov</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>bf</i>. - Bartlett: se usa la función <i>bartlett.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>b</i>. - Fligner-Killeen: se utiliza la función <i>fligner.test</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>fk</i>. - Breusch-Pagan: se utiliza la función <i>bptest</i>, donde se coloca <i>y~x</i>, utilizando los datos llamados <i>datosnormales</i>. Se guardan como <i>bp</i>. <p>A cada uno de estos se les extrae el valor <i>p</i> correspondiente y se acumulan en los vectores mencionados al principio.</p> <p>Esto se realiza por 1000 veces.</p> <p>Después, se comprueban cuales valores que se tienen en los vectores son mayores a 0.05 para saber si se cumple homocedasticidad según cada prueba y sus respectivos métodos de cálculo (se pone como <i>TRUE</i> para homocedasticidad Y <i>FALSE</i> si son heterocedásticos). Además, se multiplica por 100 y se divide entre <i>r</i> para tener su cantidad porcentual.</p> <p>Se hace una tabla con los datos recolectados.</p>
<pre> plevenemedianexp = c() plevenemeanexp = c() pbfexp = c() </pre>	<p>Se crearon vectores donde se van a acumular los valores <i>p</i> que se obtienen a partir de cada una de las pruebas de homocedasticidad.</p>

<pre> pbexp = c() pfxexp = c() pbpexp = c() h = c() r=1000 for (i in 1:r) { datosexp = dexpo(35, c(20,20,20)) #Aplicacion de pruebas lmeanexp = leveneTest(y = datosexp\$y, group = datosexp\$x, center = "mean") lmedianexp = leveneTest(y = datosexp\$y, group = datosexp\$x, center = "median") bfexp = hov(y~x, data = datosexp) bexp = bartlett.test(y~x, data = datosexp) fkexp = fligner.test(y~x, data = datosexp) bpexp = bptest(y~x, data = datosexp) #Extracción del p-value plevenemeanexp[i] = lmeanexp\$`Pr(>F)`[1] plevenemedianexp[i] = lmedianexp\$`Pr(>F)`[1] pbfexp[i] = bfexp\$p.value pbexp[i] = bexp\$p.value pfxexp[i] = fkexp\$p.value pbpexp[i] = bpexp\$p.value } #True = Hay homocedasticidad, False = Hay heterocedasticidad te1=table(plevenemeanexp>0.05)*100/r te2=table(plevenemedianexp>0.05)*100/r te3=table(pbfexp>0.05)*100/r te4=table(pbexp>0.05)*100/r te5=table(pfxexp>0.05)*100/r te6=table(pbpexp>0.05)*100/r tte = rbind(te1,te2,te3,te4,te5,te6) rownames(tte)= c("Levene - Metodo Media", "Levene - Metodo Mediana", "Brown- Forsythe", "Bartlett", "Fligner-Killeen", "Breusch-Pagan") colnames(tte)= c("Heterocedasticidad", "Homocedasticidad"); tte </pre>	<p>Se utilizaron 1000 repeticiones. Por eso $r = 1000$.</p> <p>Se hace una función <i>for</i> para 1000 repeticiones donde se obtiene los datos que se hicieron en la función <i>dexpo</i> donde se usa $n=35$ y $b=20$; el tamaño de muestra es muy grande (105) y como se asume que existe homocedasticidad, se coloca la misma b para todos los casos.</p> <p>Se aplican las pruebas definidas como:</p> <ul style="list-style-type: none"> - Levene por método de media: se usa la función <i>leveneTest</i>, donde se coloca y como la variable respuesta (<i>datosexp\$y</i>), <i>groups</i> como el factor (<i>datosexp\$x</i>) y <i>center</i> como <i>mean</i> (media). Se guardan como <i>lmean</i>. - Levene por método de mediana: se usa la función <i>leveneTest</i>, donde se coloca y como la variable respuesta (<i>datosexp\$y</i>), <i>groups</i> como el factor (<i>datosexp\$x</i>) y <i>center</i> como <i>median</i> (mediana). Se guardan como <i>lmedian</i>. - Brown–Forsythe: se utiliza la función <i>hov</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>bf</i>. - Bartlett: se usa la función <i>bartlett.test</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>b</i>. - Fligner-Killeen: se utiliza la función <i>fligner.test</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>fk</i>. - Breusch-Pagan: se utiliza la función <i>bptest</i>, donde se coloca $y \sim x$, utilizando los datos llamados <i>datosexp</i>. Se guardan como <i>bp</i>. <p>A cada uno de estos se les extrae el valor p correspondiente y se acumulan en los vectores mencionados al principio.</p> <p>Esto se realiza por 1000 veces.</p>
--	---

	<p>Después, se comprueban cuales valores que se tienen en los vectores son mayores a 0.05 para saber si se cumple homocedasticidad según cada prueba y sus respectivos métodos de cálculo (se pone como <i>TRUE</i> para homocedasticidad Y <i>FALSE</i> si son heterocedásticos). Además, se multiplica por 100 y se divide entre <i>r</i> para tener su cantidad porcentual.</p> <p>Se hace una tabla con los datos recolectados.</p>
--	---