Comportamiento de la potencia para la prueba de igualdad de medias de área de líquenes foliosos entre la Sede Rodrigo Facio y el Monte de la Cruz, considerando distintos tamaños de muestra y en presencia de homocedasticidad y heterocedasticidad modelada con mínimos cuadrados ponderados

Carlos Leandro Aguilar[6], Isaí Ugalde Araya[6], Khersttyn Darcin Barahona[6]

carlos.leandroaguilar@ucr.ac.cr, isai.ugalde@ucr.ac.cr, khersttyn.darcin@ucr.ac.cr

RESUMEN

Al realizar estimaciones con modelos de regresión, debe verificarse el cumplimiento de los supuestos como primera medida. Si estos no se cumplen, entonces se deben aplicar estrategias remediales para continuar con el análisis estadístico. El presente trabajo pretende determinar qué sucede con la probabilidad de que la prueba de hipótesis se rechace correctamente, es decir, la potencia, calculada bajo un diseño factorial de parcelas divididas, cuando se tiene homocedasticidad o heterocedasticidad, así como igual o distinto tamaño de muestra. En otras palabras, la simulación propuesta brinda la potencia de la prueba cuando se tiene un diseño balanceado y uno desbalanceado, con distintas magnitudes en las variancias. Se obtiene como resultado que el efecto de encontrar una diferencia relevante entre medias es poco significativo cuando se incumple el supuesto de homocedasticidad a medida que se incrementa el número de réplicas, sin importar si el diseño es balanceado o desbalanceado. Se puede observar que, al aplicar mínimos cuadrados ponderados, el comportamiento de la potencia es muy parecido respecto a cuando se tienen variancias iguales. Se recomienda aumentar el tamaño de muestra, especialmente en los escenarios donde el diseño presenta heterocedasticidad, y donde se pierde mayor número de parcelas, con el objetivo de obtener potencias altas.

PALABRAS CLAVE: modelos de regresión lineal, diseño factorial, prueba de hipótesis, simulación, parcelas divididas.

INTRODUCCIÓN

En análisis estadísticos uno de los supuestos más importantes es el de homogeneidad de varianzas, también conocido como homocedasticidad, el cual es ideal en aplicaciones estadísticas de suma importancia como el análisis de variancia (ANOVA por sus siglas en inglés). Como destacan Correa, Iral y 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 impactos desfavorables cuando este supuesto se incumple, entre los cuales destaca el porcentaje en que se rechaza la hipótesis nula, la diferencia mínima que se puede detectar con suficiente sensibilidad entre los promedios de los tratamientos y, por ende, la potencia de la prueba observada en el análisis de variancia.

[6]Estudiantes de la Universidad de Costa Rica

A su vez, como mencionan Cárdenas Castro y Arancibia Martini (2014), se entiende por sensibilidad de una prueba la capacidad para detectar diferencias o efectos donde los haya. Por otro lado, se entiende por potencia estadística el grado de probabilidad de rechazar una hipótesis nula cuando esta es realmente falsa, es decir, a la capacidad de una prueba para detectar diferencias entre grupos o tratamientos cuando las hay.

Badii et al. (2007) señalan que el diseño de parcelas divididas es un tipo especial de diseño de bloques completos que suele utilizarse en experimentos factoriales. Se aplican niveles de uno o más factores a las parcelas completas que se dividen en subparcelas, a las cuales se aplican los niveles de uno o más factores adicionales.

Aunado a esto, existen situaciones en las que el número de observaciones en cada tratamiento es distinto, por lo que la ortogonalidad entre efectos e interacciones se incumple. Lo anterior conlleva a que no se puedan aplicar las técnicas del análisis de variancia usuales. Por lo tanto, según Romero (2020) las sumas de cuadrados utilizadas en este caso pueden tener grandes efectos en las estimaciones de la potencia de la prueba.

Ahora bien, Gujarati (2004), Greene (1999), Tu et al. (2005) y Mela et al. (2002), citados en Ramírez Valverde y Ramírez Valverde (2006), establecen que el incumplimiento del supuesto de homocedasticidad, implica que los estimadores mínimos cuadrados ordinales (MCO) tengan ciertos problemas, debido a que, a pesar de que continúan siendo insesgados y consistentes, dejan de ser eficientes. En consecuencia, es menos probable rechazar la hipótesis, lo que conlleva a que se obtenga una potencia de la prueba baja. Es decir, si existe mucha variabilidad, una potencia pequeña no permite determinar si realmente hay o no diferencias entre los estadísticos de interés.

Además, como estrategia remedial ante el incumplimiento del supuesto de homogeneidad de variancias se utiliza el método de mínimos cuadrados ponderados porque como dice Alvarado (2020) las observaciones se ponderan por alguna función de su variancia, es un modelo que supone heterocedasticidad y asume que los errores son independientes y con distribución normal con media cero y desviación estándar poblacional diferente para cada error. Sin embargo, si se tienen valores de las variancias muy grandes, se obtiene una potencia de la prueba baja, así como inestabilidad en los signos de las estimaciones de los parámetros (Ramírez Valverde y Ramírez Valverde, 2006).

En este contexto, en el presente artículo se pretende determinar el impacto que puede tener sobre la potencia de la prueba, la presencia de igualdad de número de réplicas y de variancias, así como la presencia de distinto tamaño de muestra y de variancias (utilizando mínimos cuadrados ponderados. Se comparan, por lo tanto, distintos escenarios, en los cuales se tiene homocedasticidad y heterocedasticidad, así como réplicas iguales o diferentes. Además, se pretende identificar qué sucede con la potencia al duplicar el cuadrado medio residual y las variancias.

[6]Estudiantes de la Universidad de Costa Rica

METODOLOGÍA

Se toma como base el estudio llevado a cabo por Ugalde Araya et al. (2021), el cual consistió en un cuasi-experimento con dos factores: el de diseño, correspondiente a la ubicación geográfica, y el tipo de liquen. Los niveles del primer factor fueron Monte de la Cruz, La Sabana y la Sede Rodrigo Facio de la Universidad de Costa Rica; y los del segundo correspondieron a costroso y folioso. La unidad de observación fue el árbol, equivalente a la parcela, y la variable respuesta fue el área de cobertura del liquen, en cm^2. Por la naturaleza del estudio, el diseño utilizado fue el de parcelas divididas, el cual es análogo al de la presente investigación:

$$Y_{ijls} = \mu + \alpha_i + \varepsilon_{il}^{(1)} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijs}^{(2)}$$

donde:

- μ representa la media general del área del liquen.
- α_i simboliza el efecto de la i-ésima ubicación geográfica, i puede tomar valores de 1 a 3, porque se tiene tres lugares.
- $\varepsilon_{il}^{(1)}$ corresponde al error aleatorio del l-ésimo árbol de cada sitio, es decir, al error de parcela.
- β_j es el efecto del j-ésimo tipo de líquen, j puede ser 1 o 2 porque se cuenta con dos tipos de liquen.
- $(\alpha\beta)_{ij}$ representa el efecto de la interacción entre los factores ubicación geográfica y tipo de liquen.
- $\varepsilon_{ijs}^{(2)}$ es el error aleatorio de la s-ésima zona del árbol en donde está cada liquen, es decir, el error de subparcela.

Este modelo puede escribirse como un modelo de suma nula, haciendo uso de variables auxiliares, por lo que α_i y β_j cumplen con la restricción $\sum_{i=1}^{k} \alpha_i = 0$ y $\sum_{j=1}^{k} \beta_j = 0$, respectivamente, lo cual implica que $\alpha_k = -\sum_{i=1}^{k-1} \alpha_i$, y $\beta_j = -\sum_{j=1}^{k-1} \beta_j$. Se tienen seis tratamientos, dado que el factor ubicación geográfica posee tres niveles, y el factor tipo de liquen tiene dos. Por lo tanto, el modelo de suma nula resultante corresponde a:

$$\mu_{ij} = \mu + \alpha_1 L_1 + \alpha_2 L_2 + \beta_1 T_1 + (\alpha\beta)_{11} L_1 T_1 + (\alpha\beta)_{21} L_2 T_1$$

donde L_1 y L_2 son variables auxiliares asociadas a los primeros dos niveles del factor ubicación geográfica, T_1 es la variable auxiliar asociada al primer nivel del factor tipo de liquen, y $L_1 T_1$ y $L_2 T_1$ son variables auxiliares relacionadas con las interacciones entre los niveles de los factores mencionados anteriormente. Estas variables toman valores de -1, 0 y 1, según el tratamiento en el que se encuentre la observación.

A partir del análisis de supuestos, los autores encontraron que se incumplía el de homocedasticidad de las variancias, por lo que utilizaron Mínimos Cuadrados Ponderados para

[6]Estudiantes de la Universidad de Costa Rica

controlar la heterocedasticidad. Asimismo, determinaron que existía interacción entre los factores, por lo que plantearon las siguientes hipótesis[1] [2]:

- Liquen costroso

1) $H_0: \mu_{11} = \mu_{21} \qquad H_1: \mu_{11} > \mu_{21}$

2) $H_0: \mu_{11} = \mu_{31} \qquad H_1: \mu_{11} > \mu_{31}$

3) $H_0: \mu_{31} = \mu_{21} \qquad H_1: \mu_{31} > \mu_{21}$

- Liquen folioso

4) $H_0: \mu_{32} = \mu_{12} \qquad H_1: \mu_{32} > \mu_{12}$

5) $H_0: \mu_{32} = \mu_{22} \qquad H_1: \mu_{32} > \mu_{22}$

6) $H_0: \mu_{22} = \mu_{12} \qquad H_1: \mu_{22} > \mu_{12}$

Ugalde Araya et. al (2021) encontraron diferencias en el área de cobertura promedio entre la Sede Rodrigo Facio y el Monte de la Cruz, para el tipo de liquen folioso. Por lo tanto, la hipótesis nula y alternativa para el análisis corresponden a:

$$H_0: \mu_{Facio(folioso)} = \mu_{Cruz(folioso)} \qquad H_1: \mu_{Facio(folioso)} > \mu_{Cruz(folioso)}$$

El objetivo del presente estudio consiste en analizar el efecto del tamaño de muestra, de la homocedasticidad, y de la heterocedasticidad controlada por el método de Mínimos Cuadrados Ponderados, sobre la potencia de la prueba de igualdad de medias en la que se hallaron diferencias. Asimismo, se desea identificar cómo esta cambia al duplicar el CMR (Cuadrado Medio Residual) y las variancias, en el caso homocedástico y heterocedástico, respectivamente. Por lo tanto, se plantean los siguientes escenarios:

Tabla 1

Escenarios de análisis según número de réplicas y variancias para los tratamientos

Número de réplicas	Cuadrado medio residual	Variancias
Iguales	Original	Originales
Iguales	Duplicado	Duplicadas
Diferentes	Original	Originales
Diferentes	Duplicado	Duplicadas

[1] Ubicación: Cruz = 1, Sabana = 2, Facio = 3

[2] Liquen: Costroso = 1, Folioso = 2

Para la simulación de los datos de área en el caso de la heterocedasticidad, se establece que los datos provienen de distribuciones normales, con variancias distintas, es decir, las fijadas para cada tratamiento del estudio de Ugalde Araya et. al (2021). Además, según Ramírez Valverde y Ramírez Valverde (2006), para obtener los estimadores Mínimos Cuadrados Ponderados (MCP), se multiplican tanto la variable respuesta como las variables auxiliares por

el valor inverso de las variancias de cada tratamiento, conocido como pesos. Por lo tanto, los coeficientes MCP se obtienen de aplicar la fórmula $\hat{\beta} = (X^t WX)^{-1} X^t WY$, donde W es la matriz de pesos. Cabe mencionar que la diagonal de esta matriz corresponde a $\frac{1}{\hat{s}_i^2}$, donde i toma valores de 1 a 6, por lo que se tienen seis pesos, uno por cada tratamiento, que se aplican a cada observación (Campos Jiménez et al, 2020). Como es relevante identificar qué sucede con la potencia al duplicar las variancias, estas se multiplican por dos y se compara con las originales.

Asimismo, para el caso de la homocedasticidad, se establece que los datos proceden de distribuciones normales, con variancias iguales ($\sigma_i^2 = \hat{\sigma}^2$), siendo $\hat{\sigma}^2$ el Cuadrado Medio Residual (CMRes), el cual corresponde a la estimación de la variabilidad residual dentro de cada tratamiento. Se obtiene la fórmula:

$$CMRes = \frac{\sum_{i=1}^{k}(r_i - 1)s_i^2}{n - k}$$

donde k es el número de tratamientos, r_i es el número de réplicas en el i-ésimo tratamiento, s_i^2 es la variancia del i-ésimo tratamiento, y n es el tamaño de muestra, correspondiente a la suma del número de réplicas de los tratamientos. Además, las estimaciones de los coeficientes se consiguen a través del método de mínimos cuadrados ordinarios. Aplicando operaciones matriciales, se tiene que $\hat{\beta} = (X^t X)^{-1} X^t Y$, en donde Y representa a los valores de área, y X es la matriz de estructura, con los valores que toman las variables auxiliares (Ramírez Valverde y Ramírez Valverde, 2006). En términos prácticos, el CMRes se obtiene de realizar un promedio simple o ponderado de las variancias de cada tratamiento.

En relación con el número de réplicas, se generan muestras para un número de parcelas entre 8 y 40, las cuales se ajustan según tipo de escenario. Cabe destacar que solamente para Rodrigo Facio se plantean casos de pérdida de parcelas o árboles, que van desde 1 a 6 réplicas perdidas, debido a que para Monte de la Cruz y La Sabana se establecen iguales tamaños de muestra, de manera análoga al estudio de Ugalde Araya et al. (2021). Asimismo, se añaden efectos a las observaciones de área una vez simuladas, con el fin de que exista interacción entre los factores.

Expuesto lo anterior, una vez que se han obtenido los datos simulados para cada escenario, a partir de distribuciones normales con las variancias originales y duplicadas, así como con CMRes original y CMRes duplicado; se construye el contraste entre las medias de la Sede Rodrigo Facio y Monte de la Cruz, para el liquen folioso, con base en el modelo de suma nula. De este modo, se obtiene una estimación del valor de la diferencia de medias, el cual se estandariza dividiéndolo entre el error estándar del contraste, obtenido a través de operaciones matriciales entre el vector de coeficientes y la matriz de variancia-covariancia del modelo. Posteriormente, se obtiene la probabilidad asociada a esta prueba de hipótesis mediante la

[6]Estudiantes de la Universidad de Costa Rica

distribución t de Student, y se compara contra un nivel de significancia de 0.0084, lo que permite identificar si se rechaza que haya diferencias o no entre los promedios.

Este proceso se repite en 1000 ocasiones, y se contabiliza el número en el que se encuentran diferencias entre las medias. De esta forma, la potencia de la prueba se obtiene realizando un promedio de las veces que se rechaza la hipótesis de igualdad de medias de áreas entre los sitios de interés, para el liquen folioso. Una potencia alta implica que se lograron identificar diferencias entre las medias, en un número elevado de las mil repeticiones.

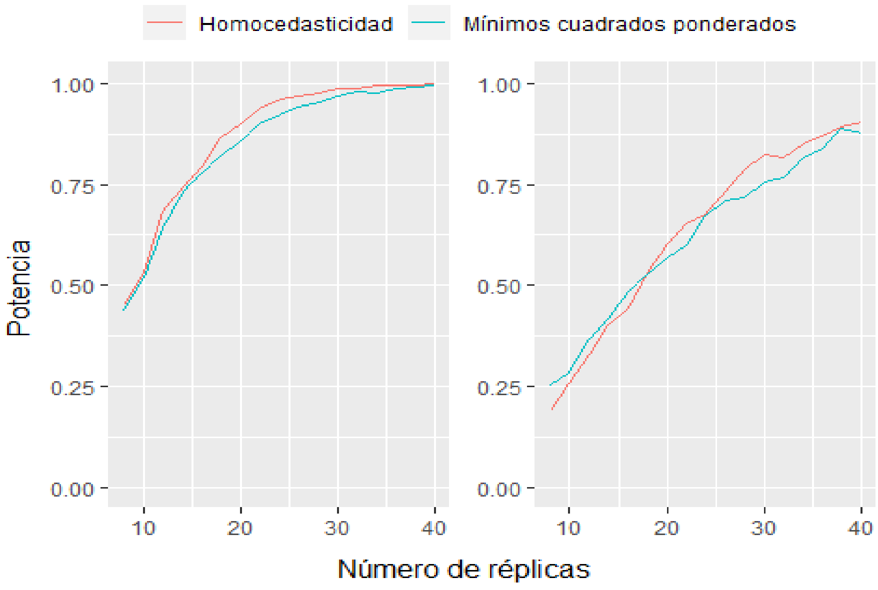
Para la simulación de los datos y el análisis de los mismos, se utiliza la versión 4.1.1 del software estadístico R (R Core Team, 2021), y las librerías nlme (Pinheiro et al., 2021), car (Fox y Weisberg, 2011), ggplot2 (Wickham, 2016), y ggpubr (Kassambara, 2020).

RESULTADOS

Como puede notarse a partir de la Figura 1, en un diseño balanceado de variancias y CMRes originales, la potencia es creciente y considerablemente alta luego de treinta réplicas, y tiende a mantenerse estable conforme aumenta el número de estas. Esto aplica tanto para el escenario homocedástico como heterocedástico con mínimos cuadrados ponderados. Además, se aprecia una tendencia a que la homocedasticidad sea la que brinde la mayor potencia según cantidad de réplicas. Asimismo, en un diseño balanceado, pero bajo el caso de variancias y CMRes duplicados, se percibe que el aumento en la variabilidad provoca una menor potencia en general, la cual, tanto para homocedasticidad como mínimos cuadrados, no se acerca a 1. Esto implicaría necesitar una mayor cantidad de réplicas para alcanzar este valor. Para réplicas mayores, la homocedasticidad tiende a producir mayor potencia.

Figura 1

Potencia de la prueba de igualdad de medias, asumiendo homocedasticidad y heterocedasticidad controlada por mínimos cuadrados ponderados, según número de réplicas, con cuadrado medio residual y variancias originales y duplicados.

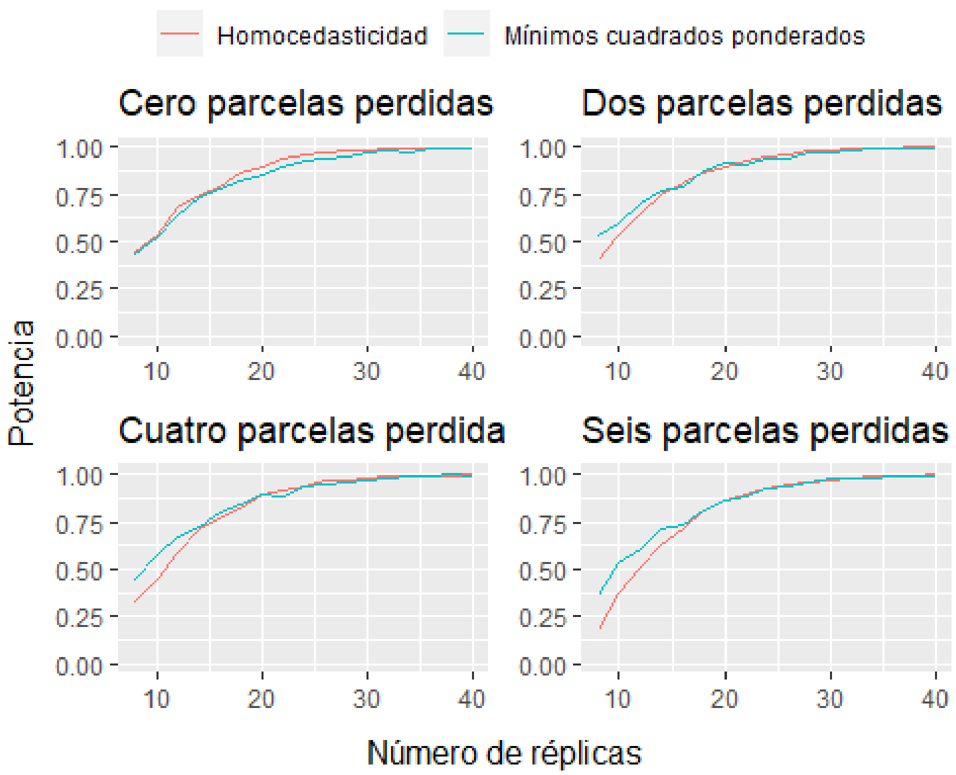


[6]Estudiantes de la Universidad de Costa Rica

Ahora bien, específicamente con el caso de variancias y CMRes originales, a partir de la Figura 2 se puede observar que la diferencia en las potencias cuando hay presencia de heterocedasticidad con mínimos cuadrados y homocedasticidad es casi nula para tamaños de muestra grandes, dado que, a partir de veinte réplicas hay una tendencia a que el comportamiento sea el mismo para cada uno de los escenarios planteados. Sin embargo, se observa claramente que conforme se pierden parcelas, la potencia para el escenario homocedástico decrece, y la heterocedasticidad bajo mínimos cuadrados produce potencias mayores. Esto se relaciona con el hecho de que, a menores tamaños de muestra, se espera que haya una mayor presencia de variabilidad y que aun cuando los tratamientos posean igual variancia, si esta es elevada la potencia será baja con un número de réplicas pequeño.

Figura 2

Potencia de la prueba de igualdad de medias, asumiendo homocedasticidad y heterocedasticidad controlada por mínimos cuadrados ponderados, según número de réplicas, con cierto número de parcelas perdidas, y considerando cuadrado medio residual y variancias originales



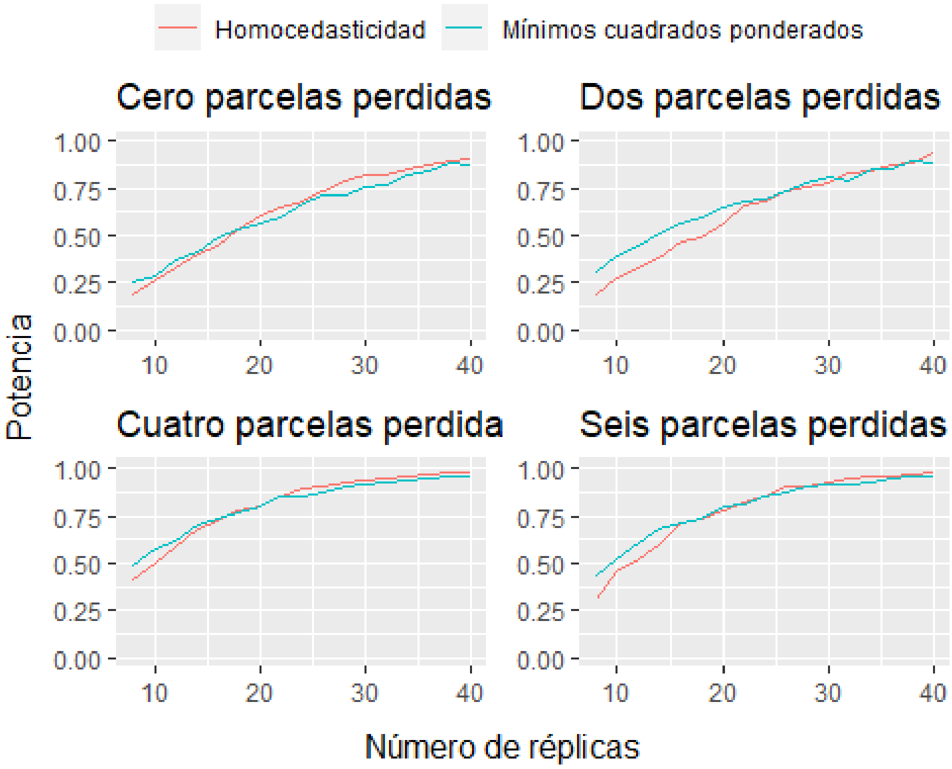
Analizando el caso de variancias y CMRes duplicados, para ambos casos, al tener un número alto de parcelas, la potencia tiende a tener un comportamiento exponencial. Se percibe que en las réplicas iniciales el procedimiento de mínimos cuadrados ponderados tiende a dar una mayor potencia en comparación con homocedasticidad. Un comportamiento similar al encontrado con las variancias y CMRes originales. De este modo, si se comparan variancias y CMRes originales además de duplicadas, en presencia de mayor variabilidad, se requiere un número de réplicas más alto para alcanzar potencias elevadas; evidencia de esto es que en la Figura 3, se nota que el valor máximo que se alcanza es de aproximadamente 0.9, y en tamaños

[6]Estudiantes de la Universidad de Costa Rica

de muestra pequeños, se tiene potencias muy bajas de casi 0.25; situación que no ocurría en la Figura 2, donde había potencias que alcanzaban valores muy cercanos a 1.

Figura 3

Potencia de la prueba de igualdad de medias, asumiendo homocedasticidad y heterocedasticidad controlada por mínimos cuadrados ponderados, según número de réplicas, con cierto número de parcelas perdidas, y considerando cuadrado medio residual y variancias duplicadas

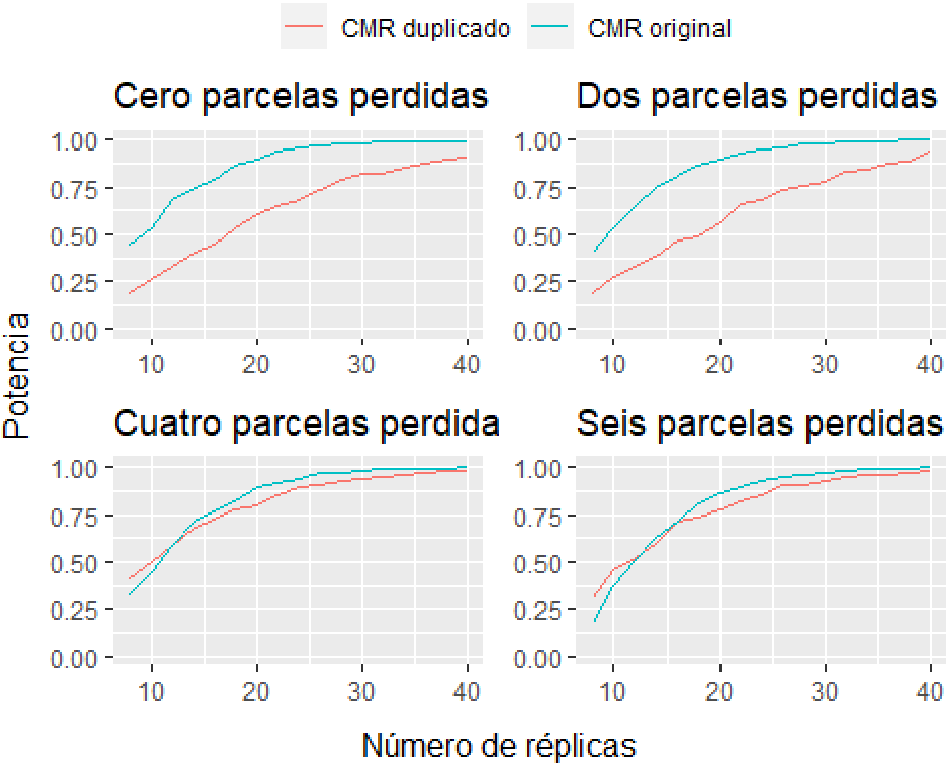


A su vez, si se analiza solamente la homocedasticidad, en la Figura 4 se puede apreciar que cuando el cuadrado medio residual original se compara con uno que es mayor (dos veces más grande), los valores que toma la potencia son considerablemente mayores para la pérdida de parcelas entre 0 y 2, y para cualquier tamaño de muestra, debido a que se tiene menor variabilidad de la respuesta. Sin embargo, a partir de cuatro parcelas perdidas, el CMRes original tiende a tener potencias más bajas en réplicas pequeñas, dado que el tamaño de muestra es menor, y se acerca a los valores que daría el CMRes duplicado.

[6]Estudiantes de la Universidad de Costa Rica

Figura 4

Potencia de la prueba de igualdad de medias, asumiendo homocedasticidad, según número de réplicas, con cierto número de parcelas perdidas, y considerando cuadrado medio residual original y duplicado

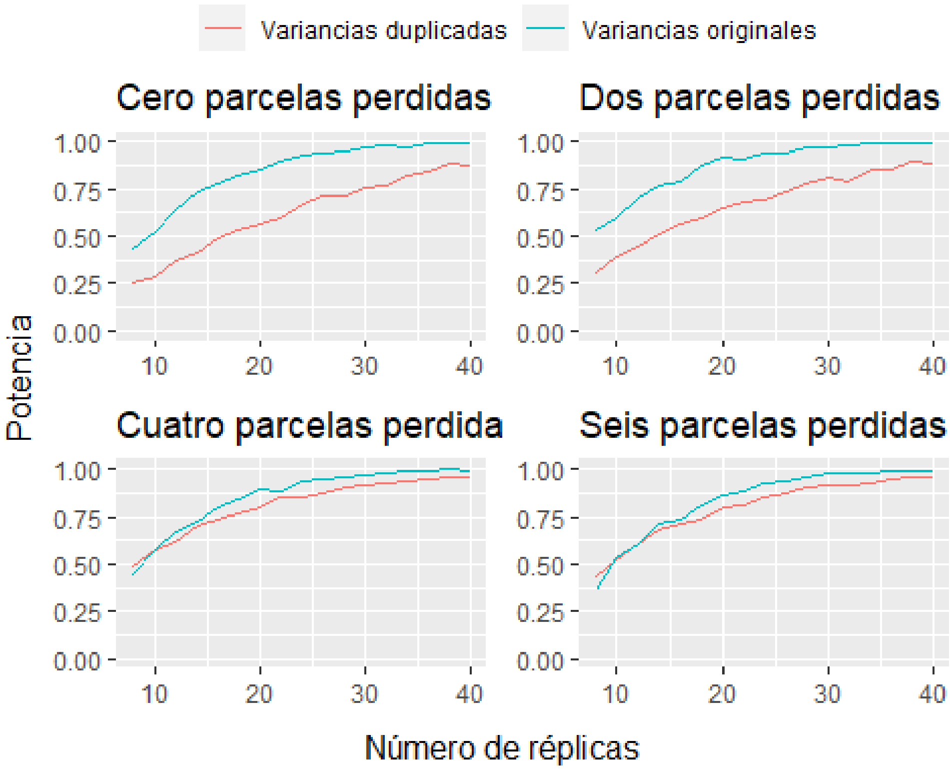


A partir de la Figura 5, bajo el escenario heterocedástico, controlado con mínimos cuadrados ponderados, se determina que cuando las variancias originales se comparan contra las duplicadas, en los casos en los que hay nula o poca pérdida de parcelas, la potencia es considerablemente superior para cualquier número de réplicas. Sin embargo, para el caso en que se pierden cuatro o más parcelas, no se aprecia una diferencia marcada entre las potencias calculadas a partir de variancias originales y duplicadas. En este escenario como en el anterior (ver Figura 4) se observa la importancia de un tamaño de muestra grande para una obtención de potencias altas.

[6]Estudiantes de la Universidad de Costa Rica

Figura 5

Potencia de la prueba de igualdad de medias, asumiendo heterocedasticidad controlada por mínimos cuadrados ponderados, según número de réplicas, con cierto número de parcelas perdidas, y considerando variancias originales y duplicadas



CONCLUSIONES

A partir del presente estudio, se logra identificar que a partir de treinta réplicas se tiende a obtener mayores potencias según cada caso. Asimismo, se nota que, a menor tamaño de muestra, la potencia disminuye, lo cual implicaría que no se pueda determinar si existen diferencias o no, es decir, a mayor número de réplicas por cada tratamiento en el experimento, se pueden obtener mayores valores en la potencia.

En relación con la variabilidad, el escenario donde se incumple el supuesto de homocedasticidad, y la distribución de la respuesta en cada población presenta mucha dispersión, es menos probable llegar a detectar diferencias entre los promedios aun cuando en realidad estos sí disten entre sí con una magnitud definida previamente como relevante. En otras palabras, con heterocedasticidad las potencias siguen manteniéndose relativamente bajas, pero al incrementarse el tamaño de muestra, estas tienden a tener valores altos; por lo que tener mucha variabilidad no es necesariamente excluyente de tener una alta probabilidad de que se rechace correctamente la hipótesis.

Asimismo, al aplicarse mínimos cuadrados ponderados para corregir el problema de la falta de homocedasticidad, se logra controlar el incumplimiento de este supuesto, lo cual se evidencia en la cercanía entre las curvas de homocedasticidad y mínimos cuadrados, según caso.

[6]Estudiantes de la Universidad de Costa Rica

Cabe destacar que el diseño de parcelas divididas permite disminuir la variabilidad presente, dado que cada árbol funciona como una parcela. Además, como mencionan Oddi et al. (2020), la flexibilidad de las funciones nlme() del software estadístico R, permiten modelar escenarios como los planteados, en donde las variancias no son constantes, por lo que el tomar en cuenta este hecho, es decir, el no ignorar el incumplimiento del supuesto de homocedasticidad al momento de establecer un modelo, permite obtener mejores estimaciones de la potencia, tanto para cuando se tenga un diseño balanceado como desbalanceado.

Finalmente, si se desea obtener potencias elevadas, se recomienda aumentar el tamaño de muestra, especialmente en los escenarios donde el diseño presenta heterocedasticidad para que se pueda presentar una disminución en la variabilidad, y además de no perder un número elevado de parcelas ya que como se vio anteriormente provoca un declive en la potencia.

BIBLIOGRAFÍA

Alvarado, R. (2020) *Modelos de Regresión Aplicados: IV. Construcción del modelo. Universidad de Costa Rica.* Presentaciones del curso.

Badii, M.H., J. Castillo, R. Rositas y G. Ponce. (2007). Experimental designs. Pp. 335-348. In: M.H. Badii & J. Castillo (eds.). *Técnicas Cuantitativas en la Investigación*. UANL, Monterrey.

Campos , P., Quirós , S. Q., y Venegas , 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. *SERENGUETI*, 17.

Cárdenas Castro, M. y Arancibia Martini, H. (2014). POTENCIA ESTADÍSTICA Y CÁLCULO DEL TAMAÑO DEL EFECTO EN G*POWER: COMPLEMENTOS A LAS PRUEBAS DE SIGNIFICACIÓN ESTADÍSTICA Y SU APLICACIÓN EN PSICOLOGÍA. *Salud & Sociedad*, 5(2),210-224. [fecha de Consulta 30 de Septiembre de 2021]. ISSN:. Disponible en: https://www.redalyc.org/articulo.oa?id=439742475006

Correa, J., Iral, R. y Rojas, L. (2006). Estudio de potencia de pruebas de homogeneidad de varianza. *Revista Colombiana de Estadística*, 29(1), 57-76. https://www.redalyc.org/pdf/899/89929104.pdf

Fox, J. y Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Inédita. Ugalde , I., Leandro , C., Darcin , K. (2021). *Efecto de la ubicación geográfica sobre el área de ocupación de líquenes costrosos y foliosos en los árboles del Monte de la Cruz, La Sabana y la Sede Rodrigo Facio.* [Artículo no publicado]. Universidad de Costa Rica.

[6]Estudiantes de la Universidad de Costa Rica

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Oddi, F., Miguez, F., Benedetti, G. y Garibaldi, L. (2020). Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza. *Ecología Austral: Asociación Argentina de Ecología, 30(3), 331-496. https://rid.unrn.edu.ar/bitstream/20.500.12049/6396/1/438-453%20diciembre20-final.pdf*

Pinheiro J., Bates D., DebRoy S., Sarkar D. R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152, <URL: https://CRAN.R-project.org/package=nlme>.

Quezada,C. (2007). Potencia estadística, sensibilidad y tamaño de efecto: ¿Un nuevo canon para la investigación?. *Onomázein, 2(16), 159-170. https://www.redalyc.org/pdf/1345/134516684004.pdf*

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramírez Valverde, G., y Ramírez Valverde, B. (2006). Colinealidad y mínimos cuadrados ponderados. *Revista Venezolana de Análisis de Coyuntura*, XII(1),283-296. https://www.redalyc.org/articulo.oa?id=36412113

Romero, P. (2020). Métodos de Diseño y Análisis de Experimentos [Diapositiva de PowerPoint]. Presentado en Departamento de Probabilidad y Estadística. IIMAS UNAM. *Métodos de Diseño y Análisis de Experimentos* (unam.mx) http://sigma.iimas.unam.mx/patricia/disenos/notas/Factoriales_desbalanceados.pdf

Wickham, H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

[6]Estudiantes de la Universidad de Costa Rica