Evaluación de la potencia de la prueba con el método de mínimos cuadrados ponderados en tratamientos balanceados y no balanceados

Stephanie Vega López[1], Francisco Calderón Fonseca[1], Luis Alonso Madrigal Ramírez[1].

stephanie.vegalopez@ucr.ac.cr, jose.calderonfonseca@ucr.ac.cr, luisalonso.madrigal@ucr.ac.cr

RESUMEN

Las estimaciones generadas a partir de un modelo estadístico se pueden considerar fiables cuando dicho modelo cumple con los supuestos necesarios como la homocedasticidad (igualdad de varianzas). En el presente artículo se evalúa el impacto que posee el uso de técnicas que permiten asumir la ausencia de homocedasticidad (mínimos cuadrados ponderados) sobre la capacidad del modelo de encontrar diferencias cuando existen, es decir la potencia de la prueba. Se plantea una simulación que permite comparar qué sucede con la potencia de la prueba cuando se poseen modelos balanceados y no balanceados, número de repeticiones por tratamiento grandes y pequeñas y diferentes niveles de varianza. Como resultados se obtiene que la potencia de la prueba es mayor cuando se utiliza mínimos cuadrados ponderados, depende a su vez del tamaño de las repeticiones por tratamiento, ya que conforme aumenta, la potencia también lo hace. En cuanto a la variabilidad si bien al ser mayor, se observa que la potencia de la prueba disminuye, pero el efecto de la variabilidad se compensa con un número de repeticiones por tratamiento adecuado.

PALABRAS CLAVE: Homocedasticidad, variabilidad, tamaño de repeticiones, supuestos.

INTRODUCCIÓN

La potencia de la prueba es un concepto fundamental en la inferencia estadística, además desempeña un papel crucial en la toma de decisiones basadas en pruebas de hipótesis. En este trabajo, exploraremos en detalle el efecto que poseen las técnicas de mínimos cuadrados ordinarios y mínimos cuadrados ponderados sobre la potencia de la prueba de hipótesis para medias diferentes, considerando tres niveles de variancia y la cantidad de observaciones por tratamiento, teniendo en cuenta si se asume un experimento balanceado o no balanceado.

En las investigaciones estadísticas, es de suma importancia comprobar el cumplimiento de los supuestos para que las estimaciones a realizar sean fiables, entre estos supuestos se encuentra la homocedasticidad, ya que el tener una varianza homogénea para los residuos de los valores de la respuesta utilizada, permite que se posea una mayor validez de los modelos, además que se puedan interpretar los resultados y de esta manera que la inferencia estadística y los estimadores sean más eficaces.

[1] Estudiantes, Bachillerato en Estadística. Universidad de Costa Rica.

González, Modroño y Regúlez (2011) mencionan que el no cumplir con el supuesto de homocedasticidad, es decir el presentar varianzas desiguales conlleva a problemas importantes como estimaciones sesgadas, estimadores no eficaces y pruebas de significancia que no son aceptadas, por lo cual es de suma importancia considerar el cumplimiento del supuesto y en caso necesario aplicar técnicas como mínimos cuadrados ponderados u ordinarios para asumir la presencia de heterocedasticidad.

De acuerdo con Quesada y Figuerola (2010) la potencia de la prueba permite establecer cuándo una prueba llevada a cabo nos conduce a efectos reales, es decir que nos indica la probabilidad de concluir que las medias son diferentes cuando en realidad sí lo son, entre mayor sea la potencia, la confianza de los resultados obtenidos será mayor y con ello el error tipo II disminuye; por lo cual la potencia de la prueba permite realizar comparaciones entre los métodos utilizados.

Teniendo presente la información consultada, se puede esperar que la potencia de la prueba de igualdad de medias sea menor cuando se posee heterocedasticidad, por lo tanto, el objetivo principal del presente estudio es comparar el impacto que posee la utilización de las medidas como mínimos cuadrados ponderados y mínimos cuadrados ordinarios sobre la potencia de la prueba en modelos balanceados y no balanceados con diferentes niveles de varianza entre tratamientos.

METODOLOGÍA

Primeramente, se definen dos variables distribuidas como:

$$Y \sim N(\mu, \sigma) \quad y \quad X;$$

Donde Y: representa la variable respuesta con una distribución normal y X: representa el factor utilizado

Se plantea un modelo teórico como:

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

donde: μ es la media general, μ_i: es la media del tratamiento i, donde i toma valores de 1 a 4 ya que se poseen 4 tratamientos y α_i: corresponde al coeficiente asociado al factor X, donde i toma valores de 1 a 4 ya que se poseen 4 tratamientos. Se plantea un modelo con suma nula por tanto se posee la restricción: $\sum_{i=1}^{k-1} \alpha_i = 0$.

El modelo extendido utilizado es:

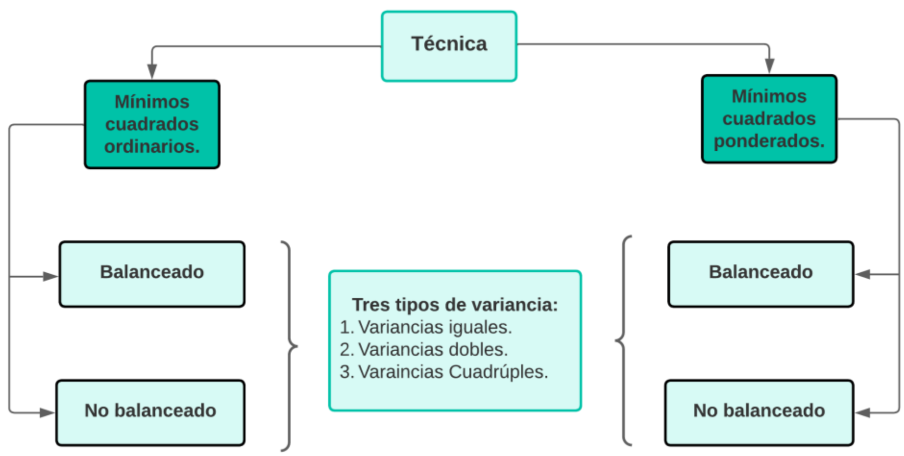
$$\mu_i = \mu + \alpha_2 T_2 + \alpha_3 T_3 + \alpha_4 T_4;$$

Donde : μ es la media general, μ_i: es la media del tratamiento i, donde i toma valores de 1 a 4 ya que se poseen 4 tratamientos, α_2: corresponde al coeficiente asociado al nivel 2 del factor, α_3: corresponde al coeficiente asociado al nivel 3 del factor, , α_4: corresponde al coeficiente asociado al nivel 4 del factor, T_2: Variable auxiliar del tratamiento 2, T_3: Variable auxiliar del tratamiento 3 y T_4: Variable auxiliar del tratamiento 4. Se plantea un modelo con suma nula por tanto se posee la restricción: $\alpha_1 = -(\alpha_2 + \alpha_3 + \alpha_4)$.

Se plantea además diferentes situaciones para la comparación a realizar tal y como se muestra en el esquema presentado en la Figura 1, para cada técnica se plantean dos escenarios posibles un modelo balanceado donde sus repeticiones son iguales dentro de cada tratamiento y un modelo no balanceado donde sus repeticiones no son iguales dentro de cada tratamiento. Asimismo dentro de cada tipo de modelo se plantean tres tipos de varianza, la primera con las cuatro varianzas iguales, la segunda con el doble entre la primera y segunda variancia y la tercera con el cuádruple entre la primera y segunda variancia.

Figura 1

Esquema de simulaciones a realizar



En la primer técnica presentada, se utiliza mínimos cuadrados ordinarios (MCO) la cual de acuerdo con Novales (2010) consiste en obtener los coeficientes que minimicen la suma de cuadrados de la distancia entre las observaciones y la recta estimada es decir el valor de $SCR = \sum_{i=1}^{n} e_i^2$, en el cual se suman los residuos al cuadrado, lo cual permite medir de mejor manera qué tan cerca pasa la recta estimada de las observaciones, es decir los residuos, para poder estimar de una manera más certera los coeficientes del modelo mediante la fórmula $\beta_\omega = (X'X)^{-1}(X'Y)$, donde X es la matriz de estructura utilizada y Y es la variable respuesta.

En la segunda técnica, se utiliza mínimos cuadrados ponderados (MCP) la cual según (González et al., 2011) utiliza los pesos para cada observación en función de su varianza estimada $w_i = \frac{1}{(s_i)^2}$, lo que conlleva a que las observaciones con menor varianza reciban un mayor peso, mientras que las observaciones con mayor varianza reciban un menor peso, con lo cual se obtienen los coeficientes mediante la fórmula $\beta_\omega = (X'WX)^{-1}(X'WY)$, donde X es la matriz de estructura utilizada y Y es la variable respuesta.

Se procede a realizar las ocho simulaciones pertinentes, se realizan cuatro simulaciones con mínimos cuadrados ponderados (ver Anexos, Tabla 3) y cuatro simulaciones con mínimos cuadrados ordinarios (ver Anexos, Tabla 4), en las cuales se generan datos muestrales provenientes de una distribución normal tal y como se mencionó anteriormente, además de la asignación del tamaño muestral, las medias por tratamiento y las varianzas presentadas en la Tabla 1.

Tabla 1

Asignación de los valores necesarios para las simulaciones (tamaño, réplicas, medias y varianzas).

Tamaño	Réplicas por tratamiento	Medias por tratamiento	Variancias
Balanceado Simulaciones 1,2 ,3 y 4.	n=(35, 35, 35, 35) n= c(10, 10, 10, 10)	μ= (10, 11 , 10.5, 10.5)	1) Iguales: σ2 = (1 ,1 ,1 , 1) 2) Diferentes: σ2 = (1, 2, 1, 2) 3) Diferentes: σ2 = (1, 4, 2, 3)
No balanceado Simulaciones 5,6 ,7 y 8.	n=(8, 12, 10, 10) n=(30, 40, 35, 35)	μ= (10, 11 , 10.5, 10.5)	1) Iguales: σ2 = (1, 1, 1, 1) 2) Diferentes: σ2 = (1, 2 , 1, 2) 3) Diferentes: σ2 = (1, 4, 2, 3)

Finalmente, se calcula la potencia de la prueba la cual indica la probabilidad de encontrar diferencias de ±1 unidades entre los promedios de la respuesta, en caso de que dichas diferencias existan. Para probar la potencia de la prueba se establecieron medias poblacionales hipotéticas de manera que se tuviera en cuenta la diferencia relevante planteada. Se realizaron 1000 repeticiones de los modelos y del análisis de varianza, se extrajo el valor F asociado y a partir del mismo se calculan las probabilidades asociadas, dicha probabilidad fue almacenada al finalizar cada interacción para calcular la proporción de veces en las cuales está probabilidad fue menor al nivel de significancia establecido en 5%. Para efectos de las simulaciones realizadas, se utilizó el lenguaje de programación R versión 4.3.0. con su interfaz RStudio R Core Team (2023).

RESULTADOS

A continuación, se presentan los resultados obtenidos en las simulaciones realizadas. Primeramente, en la Tabla 2 se puede observar que, al utilizar la técnica de mínimos cuadrados ordinarios (MCO) la potencia de la prueba es alta (0.94) cuando se tiene un modelo balanceado con tamaño de repeticiones grande de 35 observaciones por tratamiento y varianzas iguales, de igual manera, en el caso del modelo no balanceado cuando el tamaño de repeticiones por tratamiento es grande (30, 40, 35, 35) y varianzas iguales, la potencia de la prueba es de 0.91.

Al contrario de lo sucedido en la primera simulación realizada, aunque se cuenta con un modelo balanceado, al tener un tamaño de repeticiones pequeño de 10 observaciones por tratamiento y varianzas iguales, la potencia de la prueba obtenida es de 0.39, de igual manera en el caso del modelo no balanceado donde se tienen tamaños de repeticiones por tratamiento pequeño (8,12,10,10) y varianzas iguales la potencia de la prueba obtenida es de 0.35.

En cuanto al cambio de la potencia de la prueba respecto a las varianzas que poseen los datos, se observa en la Tabla 2 que cuando la varianza es igual, la potencia es alta; conforme la varianza aumenta la potencia disminuye en todos los casos, siendo más baja 0.16 cuando es balanceado y se poseen varianzas de 1, 4, 2, 3 por tratamiento respectivamente, con un tamaño de repeticiones pequeño de 10 observaciones por tratamiento; de igual manera sucede en el modelo no balanceado donde se tienen tamaños de repeticiones por tratamiento pequeños (8, 12, 10 ,10).

Tabla 2

Resultados de la prueba de potencia para la técnica mínimos cuadrados ordinarios (MCO) y mínimos cuadrados ponderados (MCP)

Tamaño y réplicas	Variancias	Potencia	
		MCO	**MCP**
Balanceado: n=(10, 10, 10, 10)	iguales: σ2 = (1 ,1 ,1 , 1)	0.39	0.39
	Diferentes: σ2 = (1, 2, 1, 2)	0.26	0.28
	Diferentes: σ2 = (1, 4, 2, 3)	0.16	0.21
Balanceado: n=(35, 35, 35, 35)	iguales: σ2 = (1 ,1 ,1 , 1)	0.94	0.94
	Diferentes: σ2 = (1, 2, 1, 2)	0.83	0.86
	Diferentes: σ2 = (1, 4, 2, 3)	0.56	0.68
No balanceado: n=(8, 12, 10, 10)	iguales: σ2 = (1 ,1 ,1 , 1)	0.35	0.35
	Diferentes: σ2 = (1, 2, 1, 2)	0.23	0.25
	Diferentes: σ2 = (1, 4, 2, 3)	0.16	0.20
No balanceado: n=(30, 40, 35, 35)	iguales: σ2 = (1 ,1 ,1 , 1)	0.91	0.91
	Diferentes: σ2 = (1, 2, 1, 2)	0.71	0.78
	Diferentes: σ2 = (1, 4, 2, 3)	0.47	0.61

Por otra parte, en la Tabla 2 se observa que, cuando se utiliza la técnica de mínimos cuadrados ponderados (MCP) la potencia de la prueba es alta (0.94) cuando se tiene un modelo balanceado con tamaño de repeticiones grande de 35 observaciones por tratamiento y varianzas iguales, de igual manera en el caso del modelo no balanceado cuando el tamaño de repeticiones por tratamiento también es grande (30, 40, 35, 35) y varianzas iguales, la potencia de la prueba es de 0.91; en ambos casos son iguales a los obtenidos con MCO.

De igual manera, aunque se cuenta con un modelo balanceado al tener un tamaño de repeticiones pequeño de 10 observaciones por tratamiento y varianzas iguales, la potencia de la prueba obtenida es de 0.39 igual al obtenido con MCO, de igual manera en el caso del modelo no balanceado donde se tienen tamaños de repeticiones por tratamiento pequeño (8,12,10,10) y varianzas iguales la potencia de la prueba obtenida es de 0.25.

En cuanto al cambio de la potencia de la prueba respecto a las varianzas que poseen los datos se observa en la Tabla 2 que cuando la varianza es igual, la potencia es alta; conforme la varianza aumenta la potencia disminuye en todos los casos, siendo más baja 0.20 cuando es no balanceado, se poseen varianzas de 1, 4, 2, 3 y donde se tienen tamaños de repeticiones por tratamiento pequeño (8, 12, 10 ,10).

En cuanto a las técnicas utilizadas para tomar en cuenta la heterocedasticidad presente en los modelos, se observa que al utilizar mínimos cuadrados ponderados la potencia de la prueba es mayor respecto a los mínimos cuadrados ordinarios, por ejemplo, cuando la variancia es poca (1,2,1,2) la potencia de la prueba en MCP es de 0.86 cuando es balanceado y de tamaño de repeticiones por tratamiento de 35; mientras que en MCO para el mismo tamaño de repeticiones y mismas varianzas es de 0.83.

De igual manera cuando el modelo es no balanceado de tamaño de repeticiones por tratamiento de 30, 40, 35 y 35, en la técnica MCP el valor de la potencia de la prueba es de 0.78; mientras que en MCO para el mismo tamaño de repeticiones y mismas varianzas es de 0.71.

CONCLUSIONES

Primeramente, se logra evidenciar que el cumplir con el supuesto de homocedasticidad, es decir tener varianzas iguales en los tratamientos conlleva a un tamaño de potencia de prueba mayor en todos los casos, ya que se observa que, aunque el modelo sea desbalanceado si las varianzas son iguales la potencia tiende a ser superior, sin embargo se observa que la calidad de la potencia de la prueba depende a su vez de la cantidad de réplicas que se posee por tratamiento siendo mayor en réplicas de mayor tamaño.

De igual manera cuando existe heterocedasticidad, ante la disminución en la potencia de la prueba cuando el tamaño de repeticiones por tratamiento es alto se observa que la potencia de la prueba decrece en menor cantidad.

Se observa que los mínimos cuadrados ponderados producen una potencia más alta respecto a los mínimos cuadrados ordinarios, lo que implica que aunque se posea una variabilidad alta entre los tratamientos, al ponderar estas diferencias no sean tan determinantes en cuanto a las estimaciones a realizar.

Conocer el efecto producido sobre la potencia de la prueba de las diferentes combinaciones planteadas entre tamaño de muestra y variabilidad presentados en la simulación permite que se pueda conocer de antemano la posible efectividad del modelo para encontrar diferencias si en verdad existen

Además se logra evidenciar que las soluciones propuestas para considerar la heterocedasticidad son efectivas para mantener una potencia alta, siendo más efectivo cuando las varianzas son muy diferentes el uso de mínimos cuadrados ponderados.

Se observa que a nivel estadístico es de gran importancia mantener un tamaño de réplicas por tratamiento alto, cuando se posee variabilidad ya que esto genera que la potencia de la prueba sea superior a 0.60 lo cual lleva a que sea más probable encontrar diferencias entre los promedios cuando realmente existen.

Finalmente, en el presente estudio se utiliza una diferencia estadística relevante fija de ±1, además de un solo factor en las ocho simulaciones realizadas, por lo tanto, se recomienda para trabajos futuros observar el impacto sobre la potencia de la prueba realizando una variación en la diferencia esperada, así como tomar en cuenta mayor cantidad de factores para poder incluir interacciones entre los mismos.

BIBLIOGRAFÍA

González, E., Modroño, J., y Regúlez, M. (2011). *Métodos Econométricos y Análisis de Datos*. OCW. https://ocw.ehu.eus/course/view.php?id=23

Novales, A. (2010). *análisis de Regresión*. Ucm.es. https://www.ucm.es/data/cont/docs/518-2013-11-13-Analisis%20de%20Regresion.pdf

Oddi, J., Miguez, F., Benedetti, G y Garibaldi, L (2020) Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza. Edu.ar. https://rid.unrn.edu.ar/bitstream/20.500.12049/6396/1/438-453%20diciembre20-final.pdf

Quesada, J., y Figuerola, J. (2010). Potencia de una prueba estadística: aplicación e interpretación en ecología del comportamiento. Csic.es. http://www.ebd.csic.es/jordiplataforma/subidas/Etologuia2010.pdf

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

ANEXOS

Tabla 3

Código utilizado en las cuatro simulaciones con mínimos cuadrados ordinarios

Simulación potencia de la prueba con mínimo cuadrados ordinarios	
Código	Explicación

med=function(r,mu,v,M){
  k=length(mu)
  r1=8
  r2=12
  r3=10
  r4=10
  x=as.factor(c(rep(1,each=r1),c(rep(2,each=r2)),
  c(rep(3,each=r3)),c(rep(1,each=r4))))
  muj=c(rep(mu[1],each=r1),rep(mu[2],each=r2),
  rep(mu[3],each=r3),rep(mu[4],each=r4))
 	
  p=c()
  for(i in 1:M){
    Y1=muj[1:r1]+rnorm(r1,0,sqrt(v[1]))
    Y2=muj[(r1+1):(r1+r2)]+rnorm(r2,0,sqrt(v[2]))
    Y3=muj[(r1+r2+1):(r1+r2+r3)]+rnorm(r3,0,sqrt(v[3]))
    Y4=muj[(r1+r2+r3+1):
(r1+r2+r3+r4)]+rnorm(r4,0,sqrt(v[4]))
    Y=c(Y1,Y2,Y3,Y4)
 	
    pi=anova(lm(Y~x))[1,5]
    p=c(p,pi)
  }
return(mean(p<0.05))
}
potencia1=round(med(r=c(r1,r2,r3,r4),
mu=c(10,11,10.5,10.5),v=c(1,1,1,1),M=1000),3)
potencia2=round(med(r=c(r1,r2,r3,r4),
mu=c(10,11,10.5,10.5),v=c(1,2,1,2),M=1000),3)
potencia3=round(med(r=c(r1,r2,r3,r4)
,mu=c(10,11,10.5,10.5),v=c(1,4,2,3),M=1000),3)
 	
ppd2=cbind(potencia1,potencia2,potencia3)
ppd2
	Se definen las características principales del diseño, numero de tratamientos (k), replicas por tratamiento (r), factor de diseño (x), media en cada nivel del factor (muj), luego se procede con la simulación, en la cual se realizan 1000 iteraciones, donde en cada iteración se calcula los valores de nuestra variable respuesta en cada tratamiento , esto con el fin de realizar un modelo con dichos valores como variable respuesta, de este modelo se extrae la probabilidad asociada a nuestro factor de diseño (pi) para guardarla en un vector vacío creado inicialmente (p), este proceso se repite durante la cantidad de iteraciones deseadas, para finalmente tener un vector con i probabilidades y así obtener la media de aquellas probabilidades menores a 0.05, lo que nos indicaría la potencia de la prueba.
	
	Por ultimo y según el objetivo planteado, se realizan tres simulaciones definiendo en el vector v varianzas iguales, varianzas dobles y varianzas cuádruples.
	Por lo que se obtienen tres valores finales los cuales nos indican la potencia de la prueba en diferentes escenarios de heterocedasticidad.
	
	Todo este proceso se realiza primero con un numero de replicas por tratamiento balanceado y luego con un numero de replicas por tratamiento desbalanceado, para un total de cuatro simulaciones.

Tabla 4

Código utilizado en las cuatro simulaciones con mínimos cuadrados ponderados

<table>
<tr><th colspan="2">Simulación potencia de la prueba con mínimo cuadrados ponderados</th></tr>
<tr><th>Código</th><th>Explicación</th></tr>
<tr><td>med=function(r,mu,v,M){
k=length(mu)
n=r*k
x=as.factor(rep(1:k,each=r))
muj=rep(mu,each=r)
w=c(rep(1/v[1],each=r),rep(1/v[2],each=r),
rep(1/v[3],each=r),rep(1/v[4],each=r))
p=c()
for(i in 1:M){
Y1=muj[1:r]+rnorm(r,0,sqrt(v[1]))
Y2=muj[(r+1):(r+r)]+rnorm(r,0,sqrt(v[2]))
Y3=muj[(r+r+1):(r+r+r)]+rnorm(r,0,sqrt(v[3]))
Y4=muj[(r+r+r+1):(r+r+r+r)]+rnorm(r,0,sqrt(v[4]))
Y=c(Y1,Y2,Y3,Y4)

pi=anova(lm(Y~x,weights = w))[1,5]
p=c(p,pi)
}
return(mean(p<0.05))
}
potencia1=round(med(r=35,mu=c(10,11,10.5,10.5),
v=c(1,1,1,1),M=1000),3)
potencia2=round(med(r=35,mu=c(10,11,10.5,10.5),
v=c(1,2,1,2),M=1000),3)
potencia3=round(med(r=35,mu=c(10,11,10.5,10.5),
v=c(1,4,2,3),M=1000),3)

ppp1=cbind(potencia1,potencia2,potencia3)</td><td>Se definen las características principales del diseño, numero de tratamientos (k), factor de diseño (x), media en cada nivel del factor (muj) y los pesos para cada observación (w), luego se procede con la simulación, en la cual se realizan 1000 iteraciones, donde en cada iteración se calcula los valores de nuestra variable respuesta en cada tratamiento , esto con el fin de realizar un modelo con dichos valores como variable respuesta incluyendo el vector de los pesos, de este modelo se extrae la probabilidad asociada a nuestro factor de diseño (pi) para guardarla en un vector vacío creado inicialmente (p), este proceso se repite durante la cantidad de iteraciones deseadas, para finalmente tener un vector con i probabilidades y así obtener la media de aquellas probabilidades menores a 0.05, lo que nos indicaría la potencia de la prueba.

Por ultimo y según el objetivo planteado, se realizan tres simulaciones definiendo en el vector v, varianzas iguales, varianzas dobles y varianzas cuádruples.
Por lo que se obtienen tres valores finales (ppp1) los cuales nos indican la potencia de la prueba en diferentes escenarios de heterocedasticidad.</td></tr>
</table>

	Todo este proceso se realiza primero con un numero de replicas por tratamiento balanceado y luego con un numero de replicas por tratamiento desbalanceado, que se define en la variable r, para un total de cuatro simulaciones.