Psicológica (2010), 31, 129-148.

Pruebas robustas para modelos ANOVA de dos factores con varianzas heterogéneas

Guillermo Vallejo[1], Paula Fernández[1] y Pablo Livacic-Rojas[2]

[1]*Universidad de Oviedo, España*
[2]*Universidad de Santiago, Universidad Mayor, Chile*

El objetivo de esta investigación fue comparar la robustez de dos estadísticos heteroscedásticos, Welch-James desarrollado por Johansen (WJ) y el estadístico Tipo Box desarrollado por Brunner, Dette y Munk (BDM), junto con el Modelo Lineal General (GLM), no heteroscedástico, de dos modos diferentes en función del cálculo del valor crítico. De una parte, cuando los valores críticos se basan en valores teóricos (WJ, BDM y GLM respectivamente), y de otra parte, cuando se obtienen mediante remuestreo bootstrap (WJB, BDMB y GLMB respectivamente). Para llevarlo a cabo se realizó un estudio de simulación sobre un diseño factorial carente de homogeneidad, normalidad y ortogonalidad. Los resultados muestran que cuando la relación entre el tamaño de las celdas y el tamaño de las varianzas fue positiva el procedimiento WJ fue el más robusto y que cuando la relación fue negativa el procedimiento más robusto fue WJB. Ambos procedimientos se comportaron de modo liberal cuando la forma de distribución fue sesgada, en mayor medida, cuanto mayor era el grado de desigualdad del tamaño de las celdas y la heterogeneidad de las varianzas.

Durante las tres últimas décadas ha surgido un extenso debate acerca de cuál es la mejor solución para contrastar hipótesis mediante modelos de análisis de la varianza (ANOVA) con dos o más factores (vías) de efectos fijos, los cuales si bien satisfacen los supuestos relativos a la distribución de probabilidad del término de error, carecen del adecuado equilibrio o balanceo. Tras decantarse por soluciones mínimo cuadráticas ejecutadas comparando modelos, gran parte del debate se ha centrado en la elección

[1] **Correspondencia**: Guillermo Vallejo, Universidad de Oviedo, Departamento de Psicología, Plaza Feijóo, s/n, 33003 Oviedo (Spain), E-mail (gvallejo@uniovi.es).
Agradecimientos: Este trabajo ha sido financiado mediante un proyecto de investigación concedido por el MEC (PSI2008-03624).

del modelo más adecuado para contrastar las hipótesis de interés. Una exposición detallada de la problemática implicada puede consultarse en MacNaughton (1998).

Cuando el tamaño de muestra de las combinaciones de tratamiento (celdas) es uniforme, todas las soluciones proporcionan idéntica descomposición de la suma de cuadrados (SC) del modelo. Sin embargo, cuando el tamaño de las celdas difiere, las diferentes soluciones proporcionan estimaciones de las SC que no son por lo general coincidentes, dependiendo del tipo de codificación empleado y del orden en el cual los efectos son introducidos en el modelo. Por ejemplo, mediante un ANOVA de dos vías se pueden obtener, al menos, tres descomposiciones diferentes de la SC correspondiente a las filas (factor *A*), a saber: SC Tipo I, Tipo II y Tipo III (el programa SAS utiliza estos tres nombres). El primer tipo implica computar la SC de *A* ignorando el efecto de las columnas (factor *B*) y de la interacción *AB*. El segundo tipo implica calcular la SC de *A* contemplando los efectos de *B* e ignorando la contribución de *AB*, mientras que el tercer tipo implica obtener la SC de *A* ajustándola, tanto para los efectos de *B* como para los efectos de *AB*. Cuando la descomposición de las SC es única, es posible interpretar los resultados de un experimento de manera clara, concisa y exacta. Desafortunadamente, esta simplicidad interpretativa desaparece cuando el diseño no está balanceado (Wang y Akritas, 2006); particularmente, cuando algún efecto es significativo utilizando un tipo de SC y no significativo usando otro u otros (Ato y Vallejo, 2007). Por consiguiente, cuando se utilice el modelo ANOVA factorial con datos no equilibrados puede resultar clave conocer las causas del desgaste de muestra.

Las razones que ocasionan las pérdidas, referidas frecuentemente como mecanismos generadores de las mismas, puede ser independientes de todas las variables y covariables presentes en el estudio o dependientes de las mismas. En el primer caso, cabe pensar que las observaciones registradas constituyan una muestra aleatoria, aunque restringida de la inicialmente prevista; por este motivo, es razonable contrastar las hipótesis sin hacerlas depender del número de réplicas realizadas. En el segundo caso no cabe efectuar la misma conjetura, puesto que el mecanismo responsable de la pérdida se relaciona sistemáticamente con los niveles de alguna de las variables manipuladas y/o con las características de los participantes observados en el estudio; por dicho motivo, lo razonable sería contrastar las hipótesis haciéndolas depender del tamaño de las celdas.

Cuando se emplea un modelo ANOVA factorial adoptando la solución mínimo cuadrática basada en el enfoque de comparación de modelos, se asume que los errores del modelo se distribuyen normal e independientemente con media cero y varianza constante. Sin embargo, Milligan, Wong y Thompson (1987) han puesto de relieve que el modelo ANOVA proporciona resultados severamente sesgados cuando los datos incumplen el supuesto previo de homogeneidad de las varianzas. En concreto, los autores citados encontraron que cuando los pesos asignados a las varianzas de la población se relacionaban negativamente con el tamaño de muestra de las celdas, las tasas de error empíricas excedían varias veces el umbral nominal. Por el contrario, cuando los pesos asignados a las varianzas de la población se relacionaban positivamente con el tamaño de las celdas las tasas de error tendían a aproximarse a cero. Zimmerman (2004) ha observado un desempeño similar en el contexto del ANOVA de una vía.

Hoy en día, se encuentran disponibles diversas soluciones para vencer el impacto negativo que la heterogeneidad de las varianzas ejerce sobre las tasas de error Tipo I. Por ejemplo, un método que no requiere la igualdad de las varianzas y que puede ser adoptado para contrastar hipótesis y comparar modelos en el contexto de los diseños factoriales con datos no equilibrados, es la prueba de Welch-James (WJ) desarrollada por Johansen (1980). Usando métodos numéricos, Keselman, Carriere y Lix (1995) encontraron que el desempeño del test WJ era muy razonable cuando los datos se distribuían normalmente, excepto cuando el tamaño de las celdas era reducido. Cuando los datos estaban sesgados, el enfoque WJ requería incrementar el tamaño de las celdas para controlar las tasas de error. Dicho incremento dependía de la magnitud del sesgo, de cuán desequilibradas estaban las celdas y de la relación existente entre el tamaño de las celdas y el tamaño de las varianzas. Posteriormente, Keselman, Kowalchuk y Lix (1998) sugirieron corregir los efectos ocasionados por la falta de normalidad, usando el enfoque WJ con estimadores robustos localización y escala. A saber: medias recortadas y varianzas winsorizadas. No obstante, además de la importancia de contar con un número mínimo de participantes por celda cuando los datos son recortados, cabe preguntarse sí las hipótesis contrastadas cuando se utilizan los usuales estimadores de tendencia central y variabilidad son las mismas que las probadas cuando se emplean estimadores robustos. Existen otras aproximaciones robustas basadas en SC ponderadas similares a la prueba de Welch (véase, Kulinskaya y Dollinger, 2007 y las referencias contenidas en este trabajo), sin embargo, no las consideramos en este artículo por el enorme esfuerzo computacional que supone su extensión a modelos ANOVA de dos o más vías.

Brunner, Dette y Munk (BDM, 1997), por su parte, también han propuesto una prueba estadística heteroscedástica para contrastar hipótesis y comparar modelos ANOVA no equilibrados. Su método consiste en una generalización del método de Box (1954), similar al desarrollado por Brown y Forsythe (BF, 1974) y a la posterior modificación propuesta por Vallejo, Fernández y Livacic-Rojas (2008). No obstante, el enfoque BDM es más robusto que el enfoque BF original y más potente que la versión modificada. Richter y Payton (2003) han evaluado las características operantes del enfoque BDM asumiendo pérdida de datos completamente al azar. Sus resultados ponen de relieve que el enfoque limitaba el número de errores al valor nominal cuando las varianzas eran heterogéneas y el tamaño de las celdas relativamente pequeño. Sin embargo, conviene reseñar que la ejecución del método tan sólo fue examinada comparando SC Tipo III bajo condiciones de normalidad.

Más recientemente, Vallejo, Fernández y Livacic-Rojas (2009) han investigado como afecta al desempeño de los procedimientos WJ y BDM la violación separada y conjunta de los supuestos de normalidad y homogeneidad de los datos. Los resultados pusieron de relieve que cuando la forma de la distribución era simétrica, ambos enfoques mantenían las tasas de error próximas al nivel nominal elegido. También se constató que cuando la forma de la distribución era asimétrica ninguno de los dos enfoques mantenía controlada la tasa de error. Este patrón de resultados fue consistente a través de las SC Tipo I, Tipo II y Tipo III empleadas para contrastar las hipótesis. Debido a lo expuesto, es imperativo que los investigadores elijan técnicas estadísticas que sean robustas al incumplimiento de los supuestos derivacionales en las que se basan.

Es muy probable que el enfoque usado con mayor frecuencia para intentar corregir la falta de normalidad de las variables, se base en transformar la escala de los datos mediante funciones no lineales. Cuando los investigadores conocen la distribución teórica de sus datos, pueden utilizar esta información para seleccionar la transformación adecuada. Cuando carecen de dicha información es útil estimar el parámetro de potencia a la que hay que elevar los datos usando las transformaciones de Box y Cox (1964). Una vez que la transformación adecuada ha sido hallada y los datos analizados en la nueva escala de medida, la interpretación de los resultados debe realizarse en sintonía con la escala utilizada. Conviene advertir, no obstante, que la interpretación de los datos cuando es posible localizar la transformación adecuada no está exenta de cierta controversia y debate.

Otro enfoque que puede ser viable para realizar un ANOVA factorial cuando las celdas tienen distinto tamaño y los datos se desvían del supuesto normalidad, se basa en derivar la distribución muestral empírica del estadístico de interés remuestreando repetidamente con reposición desde la muestra disponible. En el contexto específico de los diseños de medidas repetidas, Vallejo, Cuesta, Fernández y Herrero (2006) pusieron de manifiesto que el método percentil bootstrap-F constituía una alternativa robusta cuando los datos se desviaban de la normalidad y esfericidad multimuestral requeridas. Debido a que el comportamiento del método bootstrap-F fue sumamente bueno, inclusive con tamaños de muestra pequeños, Vallejo *et al.* (2006) recomiendan usarlo en la práctica. No obstante, se necesita efectuar investigación adicional, no sólo para determinar su robustez a las violaciones de sus supuestos derivacionales cuando se utilizan diseños no correlacionados, sino también para conocer su sensibilidad a la hora de detectar los efectos no nulos del diseño.

En consecuencia, en el presente trabajo emplearemos el método percentil bootstrap-F para combatir el sesgo de los datos y dos procedimientos estadísticos heteroscedásticos, a saber los enfoques WJ y BMD, para combatir la heterogeneidad de las varianzas. Además, con fines comparativos también investigaremos el comportamiento del modelo lineal clásico, tal y como está implementado en el módulo *Proc GLM* del programa SAS (SAS Institute, 2008, versión 9.2). La razón para incluir este modelo se basa en la demanda casi exclusiva que los usuarios de los diseños factoriales hacen de *Proc GLM*. Hasta la fecha, se desconoce el desempeño de evaluar los enfoques reseñados en conjunción con el método percentil bootstrap-F. En todos los casos las características operantes de los métodos propuestos serán evaluados ejecutando la SC Tipo III.

Definición de las pruebas estadísticas

Considérese un diseño en el cual los n_{jk} $(\Sigma_j \Sigma_k n_{jk} = N;\ j = 1,...,a;\ k = 1,...,b)$ participantes de cada una de las *JK* celdas sean medidos en una única ocasión. Usando el modelo de medias, la respuesta dada por el *i*-ésimo participante en el *j*-ésimo nivel de *A* y en el *k*-ésimo nivel de *B* es representada mediante la ecuación $Y_{ijk} = \mu_{jk} + e_{ijk}$, con $e_{ijk} \sim N(0,\sigma^2)$. De acuerdo con Searle (1987), las hipótesis nulas (H_0) asociadas con las SC Tipo III implican el esquema de ponderación siguiente:

$$H_0(A) = \sum_{k=1}^{b} \frac{\mu_{jk}}{b} - \sum_{k=1}^{b} \frac{\mu_{j'k}}{b} = 0,\ \forall\, j \neq j',$$

$$H_0(B) = \sum_{j=1}^{a} \frac{\mu_{jk}}{a} - \sum_{j=1}^{a} \frac{\mu_{jk'}}{a} = 0, \forall k \neq k',$$

$$H_0(AB) = \mu_{jk} - \mu_{j'k} - \mu_{jk'} + \mu_{j'k'} = 0, \forall j \neq j' \text{ y } k \neq k'. \quad (1)$$

Desde la perspectiva de comparación de modelos, con este enfoque cada efecto principal se prueba comparando la SC residual correspondiente al modelo no aditivo con la SC residual obtenida tras eliminar del modelo completo el efecto referido a la H_0 de interés. Este método de estimación es similar al análisis de medias no ponderadas descrito por Horst y Edwards (1982); de hecho, los resultados obtenidos con ambos enfoques son idénticos si los factores tienen sólo dos niveles. En nuestra opinión, siempre que el investigador esté interesado en poner a prueba aquellas hipótesis que surgen de su ámbito de trabajo y no aquellas otras que dependan del número de sujetos que aparecen en las celdas, este procedimiento representa la mejor opción.

El procedimiento Welch-James: Siguiendo a Johansen (1980), el estadístico WJ para probar las H_0 definidas en (1) puede expresarse como

$$T_{WJ} = (\mathbf{R}\hat{\boldsymbol{\mu}})'(\mathbf{R}\hat{\boldsymbol{\Sigma}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\mu}}), \quad (2)$$

donde $\mathbf{R}$ es una matriz de contrastes cuyo orden depende de la hipótesis a probar, $\hat{\boldsymbol{\mu}}$ es un vector de orden $JK \times 1$ obtenido tras concatenar verticalmente las medias de $\hat{\boldsymbol{\mu}}_j (= \hat{\mu}_{j1},...,\hat{\mu}_{jK})$ y $\hat{\boldsymbol{\Sigma}}$ es una matriz diagonal cuyas entradas son las varianzas dividido por el tamaño de las celdas. El estadístico T_{WJ} dividido por una constante, *c*, se distribuye aproximadamente como *F* con grados de libertad $\hat{\nu}_1 = R(\mathbf{R})$ y $\hat{\nu}_2 = \hat{\nu}_1(\hat{\nu}_1 + 2)/3\text{A}$, donde $\text{c} = \hat{\nu}_1 + 2\text{A} - (6\text{A})/(\hat{\nu}_1 + 2)$, $\text{A} = \sum_{jk}(1 - \mathbf{P}_{jk,jk})^2/(n_{jk} - 1)$, $\mathbf{P}_{jk,jk}$ es el elemento (*jk, jk*)-ésimo de la matriz $\mathbf{I} - \mathbf{R}\hat{\boldsymbol{\Sigma}}(\mathbf{R}\hat{\boldsymbol{\Sigma}}\mathbf{R}')^{-1}\mathbf{R}$ e $\mathbf{I}$ es la matriz de identidad.

Para probar el conjunto de hipótesis definido en (1) con el estadístico WJ dado en (2), $\mathbf{R} = \mathbf{C}_{AB}, \mathbf{C}_A$ y $\mathbf{C}_B$. Aquí, $\mathbf{C}_{AB} = \mathbf{C}_A \otimes \mathbf{C}_B$, donde $\mathbf{C}_A = \mathbf{P}_A \otimes \mathbf{1}'_B$, $\mathbf{C}_B = \mathbf{1}'_A \otimes \mathbf{P}_B$ y $\otimes$ denota el producto directo de matrices. Para cualquier entero *H,* $\mathbf{P}_H = (\mathbf{1}_{H-1} \,|\, \mathbf{I}_{H-1})$, donde $\mathbf{1}_H$ es un vector columna de orden $H \times 1$ e $\mathbf{I}_H$ es una matriz de identidad de orden *H*.

La aproximación Tipo-Box: De acuerdo con Brunner *et al*. (1997), la aproximación tipo Box para probar las H_0 definidas en (1) puede expresarse como

$$F_{TB} = \frac{N(\hat{\boldsymbol{\mu}}'\mathbf{M}\hat{\boldsymbol{\mu}})}{\mathrm{tr}(\mathbf{D}_M \hat{\boldsymbol{\Sigma}}^{\bullet})}, \tag{3}$$

donde $\mathbf{M} = \mathbf{R}'(\mathbf{R}\mathbf{R}')^{-1}\mathbf{R}$ es una matriz ortogonal de rango $\mathbf{R}'$, $\mathbf{D}_M$ denota la matriz diagonal de los elementos de $\mathbf{M}$ y $\hat{\boldsymbol{\Sigma}}^{\bullet} = N\hat{\boldsymbol{\Sigma}}$. Al igual que sucedía con el enfoque WJ definido anteriormente, $\mathbf{R} = \mathbf{C}_{AB}, \mathbf{C}_A$ y $\mathbf{C}_B$. Bajo H_0, el estadístico F_{TB} es aproximado por $F[\alpha; \nu_1, \nu_2]$, donde los estimadores de los grados de libertad correspondientes al numerador y denominador de la razón definida en (3) son

$$\hat{\nu}_1 = \frac{[\mathrm{tr}(\mathbf{D}_M \hat{\boldsymbol{\Sigma}}^{\bullet})]^2}{\mathrm{tr}(\mathbf{M}\hat{\boldsymbol{\Sigma}}^{\bullet}\mathbf{M}\hat{\boldsymbol{\Sigma}}^{\bullet})} \quad \text{y} \quad \hat{\nu}_2 = \frac{[\mathrm{tr}(\mathbf{D}_M \hat{\boldsymbol{\Sigma}}^{\bullet})]^2}{\mathrm{tr}(\mathbf{D}_M^2 \hat{\boldsymbol{\Sigma}}^{\bullet 2}\boldsymbol{\Lambda})}, \tag{4}$$

respectivamente, con $\boldsymbol{\Lambda} = \mathrm{diag}\{1/(n_{11}-1),...,1/(n_{ab}-1)\}$. El estimador del argumento ν_2 coincide puntualmente con el ofrecido por la generalización del enfoque BF.

Este enfoque también ha sido aplicado para analizar diseños factoriales que requieren una estructura del error no paramétrica (para detalles véase Akritas y Brunner, 2003).

El Enfoque bootstrap-F: De acuerdo con Vallejo *et al*. (2006), estimar los puntos críticos de los estadísticos WJ, BDM y GLM bajo H_0 con el enfoque bootstrap requiere efectuar los pasos siguientes:

1. Definir la unidad de muestreo a utilizar y desplazar las distribuciones empíricas de manera que las H_0 sean verdaderas. En nuestro caso, dicha unidad será las puntuaciones centradas en la media para cada una de las *jk* poblaciones de muestreo existentes en el diseño. La operación de centrado (cambio de origen en los datos) garantiza que cada uno de los *a* niveles de la variable filas y cada uno de los *b* niveles de la variable columnas no difieran entre sí. Además, los datos obtenidos por esta vía mantienen las mismas propiedades distribucionales que los datos originales, dado que no existe cambio en la escala de los mismos. De acuerdo con Westfall y Young (1993), el uso de la distribución empírica desplazada constituye la piedra angular en la que basa la técnica de remuestreo bootstrap para estimar los valores críticos.

2. Extraer múltiples muestras bootstrap de tamaño n_{jk} efectuando un muestreo aleatorio con reposición desde las puntuaciones centradas en la media. Debido a la existencia de varianzas heterogéneas y a la falta de equilibrio que presentan los diferentes grupos de celdas que configuran el diseño, el procedimiento de remuestreo se efectúa para cada una de las jk celdas del diseño. De este modo, cada muestra bootstrap es obtenida desde una distribución para la cual las H_0 son verdaderas.

3. Calcular F^* para cada uno de los efectos del diseño, el valor del estadístico F basado en la muestra bootstrap, y generar la distribución muestral empírica del estadístico (F_b^*) repitiendo el proceso b veces, siendo $\mathrm{b} = 1,...,\mathrm{B}$. Hall (1986) proporciona la justificación teórica para dicha elección. La distribución muestral bootstrap de F_b^* se utiliza para determinar la excepcionalidad del valor obtenido tras aplicar la prueba estadística F a la muestra de datos originales.

4. Determinar el valor p bootstrap. La significación de las razones F se estima directamente mediante el nivel de significación alcanzado (NSA) por el procedimiento de remuestreo bootstrap. En concreto, siguiendo el trabajo de Efron y Tibshirani (1993), $\mathrm{NSA} = \mathrm{B}^{-1}\sum_{b=1}^{B} \mathrm{I}_{[\mathrm{F}_b^* > \mathrm{F}]}$, donde $\mathrm{I}_{[\mathrm{F}_b^* > \mathrm{F}]}$, la usual función indicador, es igual a 1 si $\mathrm{F}_b^* > \mathrm{F}$ y a 0 si es menor. La proporción de valores F_b^* que supera al valor F observado representa el valor p bootstrap. Si el valor p es menor que la tasa de error nominal, entonces se rechaza la H_0.

Bajo este método, a diferencia de lo que sucedía con los tests WJ y BDM, los valores críticos derivados desde la teoría normal son innecesarios. Además de los citados Efron y Tibshirani (1993), Chernick (2007) y Good (2006) ilustran de manera relativamente sencilla cómo tratar con este método el contrate de hipótesis y otros problemas de estimación e inferencia estadística.

MÉTODO DE LA SIMULACIÓN

En orden a evaluar la robustez de los enfoques WJ, BDM y GLM cuando los valores críticos se obtienen mediante valores teóricos y mediante remuestreo bootstrap cuando se incumplen los supuestos de normalidad y homogeneidad, llevamos a cabo un estudio de simulación usando un ANOVA factorial de dos factores con J = 2 y K = 5. Para ello fueron manipuladas las cuatro variables siguientes:

1. Grado de desigualdad del tamaño de las celdas (C). Se usaron tres coeficientes de variación (*CV*) distintos para determinar el tamaño muestral de las celdas: pequeño (*CV* = .148), moderado (*CV* = .306) y grande (*CV* = .467), donde $\mathrm{CV} = \frac{1}{\bar{n}}[\Sigma_{jk}(n_{jk} - \bar{n})^2 /ab]^{1/2}$, siendo $\bar{n}$ el tamaño promedio de las celdas. Para el *CV* = .148 el tamaño de las celdas fue: $n_{1k} = (6, 7, 7, 7, 8)$ y $n_{2k} = (8, 9, 9, 9, 10)$; para el *CV* = .306 el tamaño de las celdas fue: $n_{1k} = (4, 5, 6, 7, 8)$ y $n_{2k} = (8, 9, 10, 11, 12)$, y para el *CV* = .467 el tamaño de las celdas fue: $n_{1k} = (3, 4, 5, 6, 7)$ y $n_{2k} = (7, 9, 11, 13, 15)$. El tamaño de muestra global fue siempre constante e igual a 80 (*N* = 80).

2. Grado de heterogeneidad de las varianzas (H). Se sometieron a prueba tres grados de heterogeneidad de las varianzas: pequeño, moderado y severo. Bajo la primera condición investigada la relación entre las varianzas de las diferentes celdas del diseño fue: 1:1:1:1:1:1:1:1:1:4. Bajo la segunda condición la relación fue: 1:1:1:1:1:1:1:1:1:16. La relación investigada bajo la tercera condición fue: 1:1:1:1:1:1:1:1:1:36.

3. Relación entre el tamaño de las celdas y el tamaño de las varianzas (*C*/H). Dado que la forma de emparejar los tamaños de las celdas y las varianzas tiene influencia en las pruebas estadísticas (Brown & Forsythe, 1974; Keselman *et al*., 1995), se investigó el comportamiento de los tres enfoques bajo relaciones positivas (la celda con mayor tamaño es emparejada con la varianza más grande) y negativas (la celda con menor tamaño es emparejada con la varianza más grande).

4. Forma de la distribución de la población. Si bien los procedimientos examinados se basan en el supuesto de normalidad, cuando se trabaja con datos reales es común que los índices de asimetría (γ_1) y curtosis (γ_2) se desvíen de cero (Micceri, 1989) e inducirnos a interpretar incorrectamente los resultados. Para investigar el efecto de la forma de la distribución en la robustez de los procedimientos, generamos datos desde distribuciones normales y no normales mediante las distribuciones *g* y *h* introducidas por Tukey (1977). Además de la distribución normal $(g = h = 0;\ \gamma_1 = \gamma_2 = 0)$, también investigamos otras tres: (a) *g* = 0 y *h* = .109, una distribución que tiene el mismo grado de sesgo y de curtosis que la exponencial doble o de Laplace $(\gamma_1 = 0$ y $\gamma_2 = 3)$; (b) *g* = .76 y *h* = -.098, una distribución que tiene el mismo grado de sesgo y de curtosis que la exponencial ($\gamma_1 = 2$ y $\gamma_2 = 6$); y (c) $g = 1$ y $h = 0$, una distribución que tiene el mismo grado de sesgo y de curtosis que la distribución lognormal ($\gamma_1 = 6.18$ y $\gamma_2 = 110.94$). Las distribuciones g y h fueron obtenidas utilizando la función RANNOR del SAS. Mediante ella generamos variables aleatorias normales estándar (Z_{ijk}) y transformamos cada una de ellas como $Z^*_{ijk} = g^{-1}[\exp(gZ_{ijk}) - 1]\exp(hZ^2_{ijk}/2)$, donde g y h son números reales que controlan el grado sesgo y de curtosis. Por último, para obtener una distribución con desviación estándar σ_{jk}, cada una de las puntuaciones que conforman la variable dependiente fue creada utilizando el modelo lineal $Y_{ijk} = \sigma_{jk} \times (Z^*_{ijk} - \mu_{gh})$, donde $\mu_{gh} = \{\exp[g^2/(2-2h)] - 1\}/[g(1-h)^{1/2}]$ es la media de la de la distribución g y h (para detalles véase Headrick, Kowalchuk & Sheng, 2008).

La razón que justifica la manipulación de valores tan extremos de las variables (1), (3) y (4) se basa en la premisa de que si un método funciona bien cuando los datos se desvían sustancialmente de los supuestos que subyacen al modelo ANOVA, entonces existe cierta seguridad de que dicho método se comportará satisfactoriamente en la mayor parte de las situaciones encontradas en la práctica. Para cada una de las (3×3×2×4) condiciones del estudio se crearon 1000 muestras bootstrap (B = 1000) y 5000 conjuntos de datos replicados. Los cálculos se obtuvieron mediante la programación de un MACRO en SAS/IML.

RESULTADOS

El procedimiento más directo para decidir si un determinado enfoque es o no robusto consiste en identificar todas aquellas tasas que excedan significativamente el valor nominal de alfa (α) en más/menos dos errores estándar. No obstante, utilizamos el criterio liberal de Bradley (1978) para facilitar la comparación entre nuestros resultados y los obtenidos por otros investigadores en estudios similares. De acuerdo con este criterio, aquellas pruebas cuya tasa de error empírica ($\hat{\alpha}$) se encuentre en el intervalo $.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$, serán consideradas robustas. Por consiguiente, para el nivel de significación nominal usado en esta investigación ($\alpha = .05$), el intervalo utilizado para definir la robustez de las pruebas fue $.025 \leq \hat{\alpha} \leq .075$. Se excusa decir que de haber utilizado otros criterios, diferentes interpretaciones de los resultados son posibles.

En las Tablas 1, 2 y 3 se presentan los resultados para los procedimientos WJ y WJB, BDM y BDMB, y, GLM y GLMB respectivamente. Se ha decidido esta presentación para poder observar con mayor claridad tanto la diferencia que existe en la capitalización del error de cada uno de los procedimientos en función del cálculo del valor crítico, como las diferencias que existen entre cada procedimiento y el resto de ellos. También, y con la misma intención, en la parte izquierda de las Tablas los resultados se refieren a las dos distribuciones simétricas (Normal y Laplace) sometidas a estudio y en la parte derecha se refieren a las dos distribuciones asimétricas (Exponencial y Lognormal). De otra parte, en la parte superior de las Tablas la relación que existe entre el tamaño de las celdas y el tamaño de las varianzas es positiva y en la parte inferior es negativa.

Los resultados obtenidos son los que siguen:

a. Con respecto a los procedimientos WJ y WJB: WJ manifiesta un excelente comportamiento en las tres fuentes de variación con independencia del grado de desigualdad del tamaño de las celdas (C), del valor del parámetro de heterogeneidad (H) y de la combinación entre los valores de ambas variables manipuladas (C/H) cuando la distribución es simétrica y la relación entre el tamaño de las celdas y el valor de las varianzas es positiva (en adelante nos referiremos a esta condición como relación CH positiva). Este comportamiento se altera cuando la relación entre ambas variables es negativa (en adelante, relación CH negativa), circunstancia en que experimenta una clara dependencia tanto de C como de H mostrándose liberal en las tres fuentes de variación cuando ambas variables toman valores severos y la distribución es Normal. Si la distribución es asimétrica y la relación CH positiva, WJ se muestra liberal en las combinaciones C1/H1 y H3 y C2/H3, casi en exclusividad en la distribución Lognormal y para B y A×B. Cuando la relación CH es negativa se manifiesta liberal en las tres fuentes de variación, en mayor medida conforme más severa es la asimetría y mayores son C y H.

El procedimiento WJB tiene un comportamiento excelente en las tres fuentes de variación (excepto para A en la relación C3/H3 que destaca ligeramente liberal) cuando la distribución es simétrica Normal, y tanto si la relación CH es positiva como si es negativa. Cuando la distribución es simétrica Laplace y para ambos tipos de relación CH ajusta la tasa de error en A. Para B y A×B se torna conservador, en más ocasiones cuando la relación CH es positiva. Donde la distribución es asimétrica y la relación CH es positiva ajusta la tasa de error para A. Sin embargo, para B y A×B se observa una tendencia en el error a situarse por debajo del nivel nominal que se agrava conforme mayores son la asimetría, C y H abandonando la robustez. Bajo la relación CH negativa la tasa de error tiende a incrementar respecto a la relación CH positiva, siendo así que, A abandona la robustez para mostrarse liberal y B y A×B la alcanzan en la mayoría de las combinaciones C/H.

b. Con respecto a los procedimientos BDM y BDMB: BDM tiene un excelente comportamiento en las tres fuentes de variación con independencia de C, H y de las combinaciones C/H cuando la distribución es simétrica y la relación CH es positiva. Este comportamiento se altera cuando la relación CH es negativa mostrándose liberal bajo distribución Normal en las combinaciones C2 y C3/H2 y H3, y bajo distribución Laplace en C3/H2 y H3. Si la distribución es asimétrica y la relación CH positiva se muestra liberal (en ambas, Exponencial y Lognormal) para B y A×B en las combinaciones C1 y C2/H3, y bajo distribución Lognormal también en C3/H3 y C1/H1. Bajo esta última distribución experimenta una posición conservadora en B y A×B en las combinaciones C1, C2 y C3/H1. Cuando la relación CH es negativa su comportamiento es muy similar al procedimiento WJ descrito en el punto anterior.

El procedimiento BDMB muestra un comportamiento prácticamente igual que BDM con el matiz de que es ligeramente más conservador que aquel, sin embargo, no se alteran los resultados con respecto a él de modo significativo.

c. Con respecto a los procedimientos GLM y GLMB: El procedimiento GLM tiende a estimar el error de Tipo I por debajo del nivel nominal, posicionándose de modo conservador en las tres fuentes de variación en las combinaciones C2/H2 y H3 y C3/H1, H2 y H3 cuando la forma de distribución es simétrica y la relación CH es positiva. Cuando la relación CH es negativa su estimación es muy liberal en las tres fuentes de variación. Donde la forma de distribución de los datos es asimétrica y la relación CH es positiva manifiesta un comportamiento similar a WJ con el añadido de que se muestra conservador para la fuente de variación A en las combinaciones C3/H1, H2 y H3, en mayor medida bajo distribución exponencial. Cuando la relación CH es negativa su estimación es muy liberal.

El procedimiento GLMB se muestra óptimo bajo las dos distribuciones simétricas y relación CH positiva en toda combinación de variables estudiada. Cuando la relación es negativa sobreestima el error mostrándose liberal en las combinaciones C2/H2 y H3 y C3/H1, H2 y H3. Sea cual sea la relación CH, si la distribución es asimétrica, su comportamiento es similar al de GLM antes descrito, aunque menos liberal que aquel, alcanzando la robustez en C1, C2 y C3/H1 cuando la relación CH es negativa.

Tabla 1.- Tasas de error asociadas con la suma de cuadrados Tipo III de los procedimientos WJ y WJB en un diseño factorial 2x5

Estadístico WJ. Relación Positiva												
	D.Normal			D.Laplace			D.Exponencial			D.Lognormal		
C/H	A	B	AB	A	B	AB	A	B	AB	A	B	AB
C_1H_1	4.92	5.00	4.20	4.28	4.00	4.18	4.08	6.06	4.16	3.58	5.20	3.08
C_1H_2	4.76	4.96	5.28	4.80	4.14	4.08	5.66	6.84	6.06	5.42	**8.12**	**8.00**
C_1H_3	4.90	4.94	5.12	4.78	4.54	3.74	7.10	**8.22**	7.02	**8.20**	**10.1**	**9.22**
C_2H_1	4.78	4.96	5.40	4.46	3.94	3.84	4.82	5.50	3.82	3.62	4.82	3.10
C_2H_2	4.94	4.66	4.96	5.08	4.20	3.74	4.94	6.28	5.86	4.68	7.40	6.50
C_2H_3	4.90	4.88	5.28	4.82	4.16	3.92	6.26	6.84	6.36	6.34	**8.70**	**8.34**
C_3H_1	5.04	4.98	4.96	4.90	3.98	3.84	4.66	5.94	3.20	4.36	5.02	2.50
C_3H_2	5.08	5.34	4.78	4.94	4.32	3.86	4.22	5.72	5.16	3.42	6.92	5.38
C_3H_3	4.68	5.08	4.54	4.70	4.24	3.88	5.14	5.98	5.04	4.56	7.40	6.94
Estadístico WJB. Relación Positiva												
C_1H_1	4.82	3.70	2.98	4.06	2.70	2.86	3.72	2.64	**2.08**	2.68	**0.78**	**0.64**
C_1H_2	4.44	3.58	3.88	4.26	2.56	**2.42**	5.36	2.92	3.42	4.62	2.82	3.16
C_1H_3	4.76	3.38	3.42	4.10	2.72	**2.40**	6.58	4.02	3.84	7.50	3.94	4.60
C_2H_1	4.94	3.80	3.86	4.36	**2.42**	**2.32**	4.50	**2.42**	**1.70**	2.78	**0.92**	**0.88**
C_2H_2	4.72	3.42	3.56	4.72	**2.34**	**2.42**	4.68	2.54	2.92	3.54	**2.18**	2.60
C_2H_3	4.44	3.26	3.28	4.54	2.60	2.64	5.80	2.88	3.34	5.64	2.54	3.74
C_3H_1	5.22	3.08	3.16	5.02	**2.08**	**2.06**	4.42	**1.96**	**1.18**	3.44	**0.90**	**0.60**
C_3H_2	5.24	3.32	2.92	4.60	**2.22**	**2.30**	4.04	**2.12**	**2.08**	2.78	**1.52**	**1.94**
C_3H_3	4.56	2.82	2.78	4.48	**2.16**	**2.02**	4.62	**2.12**	**2.24**	3.82	**1.74**	**2.44**
Estadístico WJ. Relación Negativa												
C_1H_1	5.12	5.24	5.64	4.78	3.78	3.86	5.08	7.34	5.30	5.14	7.18	3.94
C_1H_2	5.24	5.52	5.04	4.88	4.26	4.56	6.72	**9.12**	**7.78**	**8.58**	**11.3**	**10.9**
C_1H_3	5.98	5.18	5.32	4.72	4.16	4.72	**9.40**	**9.84**	**9.26**	**11.8**	**12.1**	**12.1**
C_2H_1	5.08	5.90	6.22	4.70	4.42	4.62	6.06	**7.92**	6.12	5.78	**8.94**	5.60
C_2H_2	6.14	6.34	5.54	5.38	5.06	5.16	**9.50**	**10.2**	**9.84**	**11.6**	**12.1**	**12.3**
C_2H_3	6.82	6.18	6.32	5.64	5.68	5.94	**12.3**	**11.3**	**11.4**	**14.9**	**14.9**	**15.3**
C_3H_1	5.88	6.32	6.08	4.66	5.08	5.10	6.70	**8.70**	7.46	**7.68**	**9.36**	5.80
C_3H_2	6.88	7.26	7.30	6.62	6.76	7.02	**11.2**	**12.4**	**12.0**	**12.4**	**15.0**	**15.3**
C_3H_3	**8.06**	**8.36**	**8.22**	6.56	7.08	6.96	**13.6**	**14.5**	**13.9**	**16.7**	**16.5**	**16.7**
Estadístico WJB. Relación Negativa												
C_1H_1	5.16	3.62	3.82	4.54	**2.28**	**2.34**	4.88	3.64	2.98	4.08	**1.90**	**1.18**
C_1H_2	4.70	3.44	3.08	4.48	2.66	**2.40**	6.30	4.48	4.84	**7.84**	4.42	5.16
C_1H_3	4.88	3.12	3.10	4.18	**2.14**	**2.30**	**8.90**	5.62	6.12	**10.9**	5.54	6.12
C_2H_1	5.18	3.54	3.68	4.58	2.64	2.68	5.90	3.66	3.12	4.98	2.70	**1.66**
C_2H_2	5.76	3.46	3.18	5.06	2.56	2.66	**9.00**	5.72	6.54	**10.6**	5.56	6.36
C_2H_3	5.96	2.96	3.38	5.08	**2.28**	**2.48**	**11.4**	6.70	**7.58**	**13.8**	7.16	**8.10**
C_3H_1	5.96	3.56	3.58	4.78	**2.38**	2.56	6.56	4.14	3.38	6.68	3.06	**1.72**
C_3H_2	6.66	3.86	3.88	6.46	3.48	3.36	**10.9**	7.16	**7.72**	**11.6**	6.96	**7.70**
C_3H_3	**7.66**	3.74	3.72	6.22	3.42	3.48	**13.1**	**7.86**	**8.90**	**16.0**	**8.96**	**9.60**

Leyenda: Distribución normal $(\gamma_1 = 0; \gamma_2 = 0)$; Distribución Laplace $(\gamma_1 = 0; \gamma_2 = 3)$; Distribución Exponencial $(\gamma_1 = 2; \gamma_2 = 6)$; Distribución Lognormal $(\gamma_1 = 6.18; \gamma_2 = 110.94)$; C_1, C_2 y C_3= Coeficiente de variación en el tamaño de las celdas pequeño, moderado y elevado respectivamente; H_1, H_2 y H_3= valor del parámetro de heterogeneidad leve, moderado y severo respectivamente. C/H= Combinaciones entre C y H; A, B y AB= son las respectivas fuentes de variación del diseño. La negrita denota valores que están fuera del intervalo 0.025-0.075.

Tabla 2.- Tasas de error asociadas con la suma de cuadrados Tipo III de los procedimientos BDM y BDMB en un diseño factorial 2x5

Estadístico BDM. Relación Positiva												
	D.Normal			D.Laplace			D.Exponencial			D.Lognormal		
C/H	A	B	AB	A	B	AB	A	B	AB	A	B	AB
C_1H_1	4.92	4.76	4.24	4.28	3.90	4.66	4.08	3.64	3.80	3.58	**2.08**	**2.24**
C_1H_2	4.76	5.12	5.16	4.48	4.98	4.50	5.66	6.92	6.76	5.42	**7.72**	**7.58**
C_1H_3	4.90	5.56	5.80	4.78	5.08	4.86	7.10	**9.72**	**9.30**	**8.20**	**13.3**	**12.4**
C_2H_1	4.78	4.58	4.46	4.46	4.02	3.58	4.82	2.52	3.26	3.62	**1.86**	**2.64**
C_2H_2	4.94	4.86	5.20	5.08	4.90	4.46	4.94	6.70	6.82	4.68	6.72	6.84
C_2H_3	4.90	5.70	5.74	4.82	5.18	4.84	6.26	**8.72**	**8.86**	6.34	**10.5**	**10.9**
C_3H_1	5.04	4.22	3.80	4.90	3.58	3.52	4.66	2.92	2.58	4.36	**1.76**	**2.14**
C_3H_2	5.08	4.88	4.56	4.94	3.94	4.08	4.22	5.06	5.00	3.42	5.64	6.02
C_3H_3	4.68	4.74	4.46	4.70	4.82	5.12	5.14	7.28	6.86	4.56	**8.88**	**9.04**
Estadístico BDMB. Relación Positiva												
C_1H_1	4.82	4.66	4.16	4.06	3.96	4.22	3.72	3.54	3.48	2.68	**1.62**	**1.86**
C_1H_2	4.44	4.90	4.78	4.26	4.68	3.94	5.36	6.78	6.46	4.62	6.96	6.76
C_1H_3	4.76	5.32	5.24	4.10	4.42	4.34	6.58	**8.90**	**8.62**	7.50	**11.9**	**11.4**
C_2H_1	4.94	4.46	4.44	4.36	3.96	3.42	4.50	2.54	2.82	2.78	**1.44**	**2.38**
C_2H_2	4.72	4.60	4.66	4.72	4.24	4.08	4.68	6.38	6.28	3.54	6.32	6.28
C_2H_3	4.44	5.32	5.20	4.54	4.80	4.28	5.80	**7.84**	**8.06**	5.64	**9.30**	**9.98**
C_3H_1	5.22	4.40	4.40	5.02	3.38	3.50	4.42	2.66	**2.48**	3.44	**1.52**	**1.86**
C_3H_2	5.24	4.46	4.44	4.60	3.78	4.00	4.04	4.72	4.68	2.78	5.16	4.44
C_3H_3	4.56	4.28	4.14	4.48	4.30	4.54	4.62	6.66	6.28	3.82	**7.86**	**8.08**
Estadístico BDM. Relación Negativa												
C_1H_1	5.12	5.14	5.80	4.78	4.00	4.00	5.08	4.44	4.60	5.14	**2.38**	**2.48**
C_1H_2	5.24	6.62	6.42	4.88	5.06	5.34	6.72	**9.38**	**9.10**	**8.58**	**10.6**	**10.7**
C_1H_3	5.98	6.56	6.26	4.72	5.54	5.50	**9.40**	**11.8**	**12.5**	**11.8**	**15.2**	**14.9**
C_2H_1	5.08	6.02	6.28	4.70	4.54	4.22	6.06	4.16	4.16	5.78	3.16	3.20
C_2H_2	6.14	**7.92**	7.40	5.38	6.70	6.18	**9.50**	**11.9**	**11.6**	**11.6**	**10.9**	**10.9**
C_2H_3	6.82	**7.70**	**7.78**	5.64	7.34	6.62	**12.3**	**15.2**	**14.8**	**14.9**	**17.3**	**17.3**
C_3H_1	5.88	5.64	5.70	4.66	4.60	5.14	6.70	4.88	5.20	**7.68**	2.78	2.82
C_3H_2	6.88	**8.70**	**8.48**	6.62	**7.60**	**7.62**	**11.2**	**13.5**	**13.7**	**12.4**	**12.3**	**11.7**
C_3H_3	**8.06**	**9.78**	**9.98**	6.56	**8.16**	**8.28**	**13.6**	**16.6**	**16.6**	**16.7**	**18.1**	**17.9**
Estadístico BDMB. Relación Negativa												
C_1H_1	5.16	5.18	5.52	4.54	3.92	3.86	4.88	4.00	4.44	4.08	**2.10**	**2.06**
C_1H_2	4.70	6.18	6.08	4.48	4.58	4.76	6.30	**8.88**	**8.74**	**7.84**	**9.68**	**9.62**
C_1H_3	4.88	5.66	5.48	4.18	4.84	4.50	**8.90**	**11.1**	**11.7**	**10.9**	**13.9**	**13.4**
C_2H_1	5.18	5.94	6.16	4.58	4.44	4.24	5.90	3.66	4.62	4.98	2.80	2.70
C_2H_2	5.76	**7.58**	7.20	5.06	6.24	6.10	**9.00**	**11.7**	**11.5**	**10.6**	**10.1**	**9.90**
C_2H_3	5.96	7.38	7.18	5.08	6.74	6.06	**11.4**	**14.7**	**14.4**	**13.8**	**15.9**	**15.9**
C_3H_1	5.96	6.00	5.88	4.78	4.62	4.49	6.56	4.74	5.06	6.68	**2.26**	**2.36**
C_3H_2	6.66	**8.60**	**8.54**	6.46	**7.54**	**7.60**	**10.9**	**13.0**	**13.4**	**11.6**	**11.2**	**10.4**
C_3H_3	**7.66**	**9.70**	**9.80**	6.22	**7.98**	**8.10**	**13.1**	**16.5**	**16.4**	**16.0**	**16.9**	**16.9**

Leyenda: ver Tabla 1.

Tabla 3.- Tasas de error asociadas con la suma de cuadrados Tipo III de los procedimientos GLM y GLMB en un diseño factorial 2x5

Estadístico GLM. Relación Positiva												
	D.Normal			D.Laplace			D.Exponencial			D.Lognormal		
C/H	A	B	AB	A	B	AB	A	B	AB	A	B	AB
C_1H_1	3.92	4.66	4.14	3.50	4.08	4.80	3.52	4.90	4.48	3.72	4.46	3.62
C_1H_2	2.98	5.42	5.48	2.98	5.46	4.96	4.74	**7.54**	7.38	5.44	**9.68**	**8.24**
C_1H_3	2.78	7.02	7.32	2.68	6.50	6.22	5.22	**10.4**	**10.2**	7.26	**14.7**	**13.7**
C_2H_1	2.86	3.22	3.16	2.94	3.56	2.92	3.48	3.80	3.50	3.32	4.30	4.10
C_2H_2	**1.52**	2.92	2.82	**1.56**	2.80	2.98	2.86	5.14	5.20	4.48	**8.06**	7.04
C_2H_3	**1.04**	3.96	3.62	**1.08**	3.32	3.26	2.78	6.88	7.06	4.80	**10.0**	**10.0**
C_3H_1	2.50	**1.68**	**1.80**	**2.44**	**2.02**	**2.02**	**2.46**	3.06	**2.26**	3.28	4.42	3.70
C_3H_2	**0.60**	**1.14**	**1.10**	**0.58**	**0.80**	**0.92**	**1.28**	2.70	2.66	2.52	5.16	4.34
C_3H_3	**0.14**	**0.96**	**0.72**	**0.24**	**0.90**	**0.88**	**0.96**	3.36	2.86	**2.16**	5.54	5.82
Estadístico GLMB. Relación Positiva												
C_1H_1	4.86	4.64	4.12	3.88	3.96	4.28	3.68	3.64	3.46	2.64	**1.72**	**1.86**
C_1H_2	4.26	4.46	4.54	4.40	4.24	3.66	5.24	6.70	6.18	4.88	7.04	6.82
C_1H_3	4.48	4.84	4.86	3.76	4.06	3.88	6.22	**8.48**	**8.22**	7.42	**11.5**	**10.9**
C_2H_1	5.24	4.90	4.94	4.80	4.54	3.76	4.64	3.18	3.30	2.92	**1.78**	2.54
C_2H_2	4.62	4.24	4.34	4.46	3.84	3.80	4.64	6.38	6.38	4.38	6.62	6.52
C_2H_3	4.38	4.82	4.68	4.22	4.38	3.92	5.42	7.20	7.42	5.86	**9.28**	**9.88**
C_3H_1	6.54	5.38	5.20	6.56	4.92	4.86	5.42	3.92	3.44	3.90	**2.04**	**2.32**
C_3H_2	5.40	4.54	4.40	5.06	3.76	4.06	4.40	5.28	4.96	3.70	5.90	6.10
C_3H_3	4.56	4.08	3.86	4.58	4.04	4.18	4.82	6.38	5.68	4.62	**8.00**	**8.26**
Estadístico GLM. Relación Negativa												
C_1H_1	7.18	**8.56**	**9.34**	6.92	**7.74**	**7.74**	6.44	**8.26**	**8.36**	6.10	6.32	6.88
C_1H_2	**11.0**	**17.3**	**16.9**	**10.6**	**15.8**	**15.9**	**11.0**	**17.8**	**17.6**	**11.0**	**18.1**	**17.7**
C_1H_3	**15.2**	**22.4**	**22.1**	**13.4**	**21.0**	**20.3**	**15.9**	**23.8**	**24.5**	**16.8**	**26.6**	**26.4**
C_2H_1	**8.94**	**12.1**	**12.5**	**9.26**	**11.0**	**11.0**	**8.82**	**10.2**	**11.1**	**8.10**	**9.92**	**10.1**
C_2H_2	**19.6**	**26.9**	**26.8**	**19.2**	**25.8**	**25.5**	**19.9**	**27.1**	**27.5**	**18.3**	**24.3**	**25.0**
C_2H_3	**27.6**	**33.7**	**34.8**	**26.4**	**34.1**	**34.4**	**28.0**	**37.1**	**37.1**	**26.7**	**36.6**	**36.0**
C_3H_1	**11.3**	**13.6**	**12.8**	**10.5**	**12.2**	**12.9**	**10.5**	**12.2**	**12.6**	**9.34**	**10.7**	**10.9**
C_3H_2	**24.4**	**31.4**	**31.5**	**25.1**	**31.8**	**31.4**	**24.1**	**33.6**	**32.4**	**20.9**	**29.2**	**28.9**
C_3H_3	**35.9**	**42.5**	**42.2**	**34.3**	**41.5**	**41.7**	**36.7**	**44.8**	**44.2**	**34.7**	**42.0**	**41.1**
Estadístico GLMB. Relación Negativa												
C_1H_1	5.42	5.60	6.02	4.88	4.22	4.20	5.12	4.28	4.56	3.82	**2.08**	**2.02**
C_1H_2	5.72	7.12	7.22	5.16	5.48	5.80	6.70	**9.82**	**9.52**	**7.64**	**9.86**	**9.74**
C_1H_3	6.32	6.86	6.74	5.04	5.78	5.64	**9.60**	**12.1**	**12.6**	**11.4**	**14.6**	**14.4**
C_2H_1	6.62	8.12	8.30	6.26	6.04	5.94	6.82	4.90	6.02	5.02	3.24	2.92
C_2H_2	**8.94**	**11.2**	**10.8**	**7.70**	**9.20**	**9.12**	**11.1**	**13.7**	**13.5**	**11.0**	**10.8**	**10.9**
C_2H_3	**9.96**	**11.0**	**11.1**	**8.38**	**10.4**	**9.92**	**14.6**	**17.9**	**18.1**	**15.6**	**18.2**	**17.9**
C_3H_1	**9.08**	**9.86**	**9.80**	**7.40**	**7.94**	**8.62**	**8.54**	6.76	7.32	6.82	3.30	3.38
C_3H_2	**12.2**	**15.2**	**15.1**	**11.8**	**14.1**	**14.4**	**14.4**	**18.0**	**18.2**	**13.1**	**13.8**	**12.6**
C_3H_3	**14.5**	**16.5**	**16.7**	**12.1**	**14.6**	**14.5**	**19.4**	**23.0**	**22.5**	**19.8**	**20.0**	**20.8**

Leyenda: ver Tabla 1.

CONCLUSIONES Y DISCUSIÓN

El objetivo de este trabajo ha sido examinar la robustez de los enfoques WJ, BDM y GLM con los valores críticos derivados desde la teoría normal y mediante remuestreo bootstrap (WJB, BDMB y GLMB) usando un ANOVA factorial de dos factores con J = 2 y K=5. Para ello se ha llevado a cabo una investigación Monte Carlo en la que se han simulado condiciones adversas debidas al desequilibrio en el tamaño de los grupos y a la heterogeneidad de las varianzas, con el añadido de la ausencia de normalidad en la distribución de los datos.

Los resultados hallados nos llevan a concluir que los procedimientos que más veces mantienen la tasa de error dentro de los límites del criterio de robustez de Bradley en las condiciones manipuladas son WJ y WJB. Cuando la relación entre el tamaño de las celdas y el tamaño de las varianzas (CH) fue positiva el procedimiento más robusto fue el test WJ con los valores críticos derivados desde la teoría normal, resultados que son consistentes con los hallados por Keselman et al. (1995) y Vallejo et al. (2009) bajo condiciones similares a las manipuladas por nosotros. Mientras que cuando la relación CH fue negativa el procedimiento más robusto fue el test WJB.

Así las cosas, ninguno de los procedimientos sometidos a prueba se sitúa entre los límites del criterio de robustez Bradley en todas las condiciones, siendo, cuando la distribución es asimétrica y la relación CH negativa, cuando más veces se exceden los límites del intervalo de robustez. Este punto merece dos puntualizaciones:

1.- Cuando la distribución es severamente asimétrica (Lognormal) todos los estadísticos se comportan de modo más liberal o más conservador (de este último modo es apreciablemente menos abundante) de lo que son en la distribución moderadamente asimétrica (Exponencial), siendo la tasa de error mayor conforme mayores son los niveles de las variables C y H (ejerciendo mayor influencia en el resultado esta última). También es destacable que cuando la relación CH es positiva, la fuente de variación A sólo resulta afectada en 6 ocasiones en el conjunto de los 6 estadísticos estudiados, siendo B y A×B las afectadas prácticamente en exclusividad. Cuando la relación CH es negativa las tres fuentes de variación se muestran afectadas aunque el error es inferior en A.

2.- Todos los estadísticos tienen un mejor comportamiento, con excepción de GLM en ambas relaciones CH (Milligan *et al*. 1987, Vallejo *et al*., 2008 y Vallejo *et al.*, 2009 obtuvieron resultados en concordancia con los encontrados por nosotros en condiciones similares), y GLMB en la relación CH negativa, cuando la distribución es no normal simétrica. En esta situación es preciso destacar que el efecto de la combinación C/H sólo se manifiesta cuando la relación CH es negativa, en mayor proporción cuando la distribución es normal que cuando es Laplace, y afecta por igual a las tres fuentes de variación. No ejerce ninguna influencia cuando la relación es positiva salvo una excepción. Esta es en el procedimiento GLM que, además en él la combinación C/H ejerce un efecto único que no aparece en el resto de estadísticos ni en otras condiciones de las aquí estudiadas, aunque sí han sido halladas por otros investigadores en condiciones similares (ver Milligan *et al.*, 1987, Vallejo *et al.*, 2008 y Vallejo *et al.*, 2009). En este caso, GLM experimenta un efecto conservador conforme incrementan C y H. De otra parte, también apreciamos que WJB bajo distribución Laplace también manifiesta un efecto conservador en B y A×B, pero probablemente sea debido al modo de cálculo de los valores críticos y no al efecto de las combinaciones C/H.

Uno de los objetivos más importantes en esta investigación ha sido estudiar el efecto que tiene sobre los estadísticos WJ, BDM y GLM el cálculo de los valores críticos mediante remuestreo bootstrap. De algún modo se podría decir que ejerce en ellos un efecto regulador, con los siguientes matices. A saber:

- WJB con respecto a WJ: en toda situación, sea cual sea la distribución, la relación CH, y la combinación de los valores que toman C y H, realiza una estimación del error inferior, pero en exclusividad sobre B y A×B, nunca sobre A. Este efecto provoca que en la distribución Laplace y bajo relación CH positiva adquiera un comportamiento conservador, alcanzando en las distribuciones asimétricas un efecto excesivamente conservador), y que cuando la relación CH es negativa mejore la estimación con respecto a WJ en la distribución normal y en ambas distribuciones asimétricas, y la empeore en la distribución Laplace.
- BDMB con respecto a BDM: ejerce el mismo efecto que el procedimiento WJB sobre WJ pero mucho más sutil, no logrando variar apenas ninguna estimación. Esto es, el comportamiento de los procedimientos BDM y BDMB es el que más se parece entre sí en el conjunto de condiciones estudiadas correspondiéndose prácticamente el uno con el otro en todas ellas, es por ello que son robustos o no robustos en las mismas condiciones experimentales.
- GLMB con respecto a GLM: observamos que cuando la relación CH es negativa ejerce el mismo efecto que WJB y BDMB sobre WJ y BDM respectivamente. GLMB es menos liberal que GLM, pero exageradamente, ya que reduce el error empírico prácticamente en un 50% respecto al GLM.

Cuando la relación CH es positiva y la distribución simétrica sitúa la tasa de error a nivel nominal para todas las fuentes de variación en todas las condiciones estudiadas. Si observamos la parte derecha de la Tabla da la impresión de que se parece más el comportamiento de uno y de otro, sin embargo, pese al buen comportamiento que ambos tienen cuando la relación CH es positiva, observamos que cuando la relación CH es negativa experimenta el mismo comportamiento que en su parte izquierda, siendo más moderado GLMB que GLM.

Cuando se trabaja en el ámbito experimental es difícil que nos topemos con datos cuyas características sean tan adversas como las sometidas a estudio en este trabajo. Sin embargo, en investigación aplicada, estas condiciones, parecidas, y aún peores, son posibles. Ningún estadístico paramétrico es útil en todas ellas, y la línea que se debe seguir es aquella en aras de encontrar un procedimiento para el análisis de datos proporcional, en el sentido de no capitalizar sobre el azar mucho más allá de lo deseable en toda situación, además de sensible en la detección de los efectos. De momento tranquiliza conocer, como en la introducción se puso de manifiesto, que son muchos los investigadores que con rigor están llevando a cabo esta tarea. También que, de momento, los procedimientos WJ y WJB se pueden utilizar con seguridad en las condiciones señaladas al comienzo de este apartado.

ABSTRACT

Robust tests for two-way ANOVA models under heteroscedasticity: The aim of this research was to compare the robustness of two heteroscedastic test statistics, the Welch-James statistic developed by Johansen (WJ) and the Type-Box statistic developed by Brunner, Dette and Munk (BDM), together with the General Linear Model (GLM), not heteroscedastic test statistic, in two different manners depending on the calculation of the critical value. On the one hand, when the critical values are based on theoretical values (WJ, BDM and GLM respectively), and on the other hand, when they are obtained by means of bootstrap resampling (WJB, BDMB and GLMB respectively). To carry out it a study of simulation was realized on a factorial design lacking in homogeneity, normality and orthogonally. The results show that when the relation between the size of the cells and the size of the variances was positive the procedure WJ was the most robust and that when the relation was negative the most robust procedure was WJB. Both procedures behaved in a liberal way when the shape of the distribution was skewed, in major measure major it were the degree of inequality of the size of the cells and the heterogeneity of the variances.

REFERENCIAS

Akritas, M. G. & Brunner, E. (2003). Nonparametric Models for ANOVA and ANCOVA: A Review. In M. G. Akritas & D. N. Politis (Eds.), *Recent Advances and Trends in Nonparametric Statistics* (pp. 79-91). Amsterdam: Elsevier

Ato, M. & Vallejo, G. (2007). *Diseños Experimentales en Psicología*. Madrid: Pirámide.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effects of inequality of variance in the one–way classification. *Annals of Mathematical Statistics, 25*, 290-403.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-246.

Bradley, J. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology,* 31, 144-152.

Brown, M.B. & Forsythe, A.B. (1974). The small sample behaviour of some statistics which test the equality of several means. *Technometrics,* 16, 129-132.

Brunner, E., Dette, H. & Munk, A. (1997). Box-type approximations in heteroscedastic factorial designs. *Journal of the American Statistical Association*, 92*, 1*494-1503.

Chernick, M. R. (2007). *Bootstrap Methods: A Guide for Practitioners and Researchers,* 2nd edition*.* New York: Wiley.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the Bootstrap.* New York: Chapman & Hall/CRC Press.

Good, P. (2006). *Resampling Methods: A Practical Guide to Data Analysis*, 3rd edition, Boston: Birkhäuser.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics, 14*, 1431-1452.

Horst, P. & Edwards, A. (1982). The k factorial experiment. *Psychological Bulletin, 91*, 190-192.

Headrick, T. C., Kowalchuk, R. K. & Sheng, Y. (2008). Parametric probability densities and distribution functions for Tukey *g*-and-*h* transformations and their use for fitting data. *Applied Mathematical Sciences*, *2*, 449-462.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika, 67*, 85-92.

Kulinskaya, E. & Dollinger, M. B. (2007). Robust weighted one-way ANOVA: Improved approximation and efficiency. *Journal of Statistical Planning and Inference 137*, 462–472.

Keselman, H. J., Carriere, K. C. & Lix, L. M. (1995). Robust and powerful nonorthogonal analyses. *Psychometrika, 60,* 395-418

Keselman, H. J., Kowalchuk, R. K. & Lix, L. M. (1998). Robust nonorthogonal analyses revisited: An update based on trimmed, means. *Psychometrika, 63*, 145-163.

MacNaughton, D. B. (1998). *Which Sums of Squares Are Best in Unbalanced Analysis of Variance*? Available at http://www.matstat.com/ss.htm.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 92*, 778-785.

Milligan, G. W., Wong, D. S. & Thompson, P. A. (1987). Robustness properties of nonorthogonal analysis of variance. *Psychological Bulletin, 101*, 464-470.

Richter, S. J. & Payton, M. E. (2003). Performing two-way analysis of variance. *Journal of Modern Applied Statistical Methods, 2*, 152-160.

SAS Institute (2008). The MIXED procedure, *SAS/STAT User's Guide,* SAS On-Line Documentation. Cary, NC: SAS Institute Inc.

Searle, S. R. (1987). *Linear Models for Unbalanced Data*. New York: Wiley.

Tukey, J.W. (1977). Modern techniques in data analysis. NSF-sponsored regional research conference at Southern Massachusetts University (North Dartmouth, MA).

Vallejo, G., Cuesta, M., Fernández, P. & Herrero, F. J. (2006). A comparison of the bootstrap-F, improved general approximation and Brown-Forsythe multivariate approaches in a mixed repeated measures design. *Educational and Psychological Measurement, 66*, 35-62.

Vallejo, G., Fernández, P. & Livacic-Rojas, P. E. (2008). Generalización del enfoque Brown-Forsythe a diseños factoriales. *Psicothema*, *20*, 969-973.

Vallejo, G., Fernández, P. & Livacic-Rojas, P. E. (2009). Analysis of unbalanced factorial designs with heteroscedastic data. *Journal of Statistical Computation and Simulation.* DOI: 10.1080./00949650802482386.

Wang, L. & Akritas, M. G. (2006). Two-way heteroscedastic ANOVA when the number of levels is large. *Statistica Sinica, 16*, 1387-1408

Westfall, P. H. & Young, S. S. (1993). *Resampling-Based Multiple Testing*. New York: Wiley.

Zimmerman, D.W. (2004). Inflaction of type I error rates by unequal variances associated with parametric, nonparametric, and rank-transformation tests. *Psicologica, 25*, 103-133.

(Manuscrito recibido: 15 Diciembre 2008; aceptado: 29 Enero 2009)