Psicothema 2008. Vol. 20, nº 4, pp. 969-973
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

Generalización del enfoque Brown-Forsythe a diseños factoriales

Guillermo Vallejo Seco, María Paula Fernández García y Pablo Esteban Livacic Rojas*
Universidad de Oviedo y * Universidad de Santiago de Chile

El presente trabajo propone una solución basada en generalizar las ideas de Brown y Forsythe al problema de contrastar hipótesis en diseños factoriales carentes de homogeneidad. A diferencia del tradicional modelo de análisis de la varianza, el enfoque propuesto no requiere satisfacer el supuesto de homogeneidad de las varianzas. Un comprensivo estudio de simulación, en el cual se manipuló sistemáticamente el tamaño de muestra de las celdas, la relación entre el tamaño de las celdas y el tamaño de las varianzas, el grado de heterogeneidad y la forma de la distribución de la población, soporta la robustez de la aproximación propuesta para contrastar los efectos del diseño factorial en ausencia de heterogeneidad y también bajo no normalidad.

Generalization of the Brown-Forsythe approach to factorial designs. The current paper proposes a solution that generalizes ideas of Brown and Forsythe to the problem of comparing hypotheses in two-way classification designs with heteroscedastic error structure. Unlike the standard analysis of variance, the proposed approach does not require the homogeneity assumption. A comprehensive simulation study, in which sample size of the cells, relationship between the cell sizes and unequal variance, degree of variance heterogeneity, and population distribution shape were systematically manipulated, shows that the proposed approximation was generally robust when normality and heterogeneity were jointly violated.

Los diseños factoriales hacen referencia a disposiciones experimentales en las que se estudia simultáneamente la acción de dos o más variables independientes, con los niveles de cada variable combinado con los niveles de las variables restantes. Aunque resulta frecuente leer que estos diseños se debieron al genio creador de Fisher, la idea no se originó con él (Yates, 1964). Sin embargo, Fisher (1926) sí que fue el primero en ofrecer argumentos convincentes en torno a los beneficios que los diseños factoriales proporcionan cuando son utilizados en lugar de sucesivos diseños unifactoriales. De acuerdo con Fisher, el uso de estos diseños reduce los costos, incrementa la información y potencia la base inductiva de las conclusiones. Además, si las fuentes de variación son estadísticamente independientes, la simplicidad analítica e interpretativa de los diseños factoriales es equivalente a la de los diseños unifactoriales. Desafortunadamente, la situación cambia radicalmente cuando las fuentes de variación no son ortogonales y se incumplen los supuestos derivacionales del tradicional modelo de análisis de la varianza (ANOVA).

En relación con la primera cuestión, resaltar que durante las tres últimas décadas ha surgido un extenso debate acerca de cuál es la mejor solución para analizar diseños factoriales que, aunque satisfacen los supuestos relativos a la distribución de probabilidad del término de error, carecen del adecuado balanceo. Tras decantarse por soluciones mínimo cuadráticas ejecutadas comparando modelos, gran parte del debate se ha centrado en la elección del modelo más adecuado para contrastar las hipótesis de interés (véase Appelbaum y Cramer, 1974; Carlson y Timm, 1974; Herr y Gaebelein, 1978; Overall y Spiegel, 1969, entre otros). Cuando el tamaño de muestra de las combinaciones de tratamiento (celdas) es uniforme, todas las soluciones proporcionan idéntica descomposición de la suma de cuadrados del modelo. Sin embargo, cuando el tamaño de las celdas difiere, las diferentes soluciones proporcionan estimaciones de las sumas de cuadrados que, generalmente, dependen del tipo de codificación empleado y del orden en el cual los efectos son introducidos en el modelo.

Por lo que respecta a la segunda, señalar que comúnmente se asume que los errores o perturbaciones aleatorias 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 lineal general proporciona resultados severamente sesgados cuando las varianzas son heterogéneas. 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, no era raro encontrarse con tasas de error empíricas que 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 empíricas tendían a aproximarse a cero. Así pues, dependiendo de la relación existente entre el tamaño de las celdas y el tamaño de las varianzas, el enfoque clásico puede exhibir un comportamiento en exceso liberal o conservador por incumplirse los supuestos derivacionales en los que se fundamenta.

Fecha recepción: 30-5-08 • Fecha aceptación: 5-6-08
Correspondencia: Guillermo Vallejo Seco
Facultad de Psicología
Universidad de Oviedo
33003 Oviedo (Spain)
E-mail: gvallejo@uniovi.es

Para vencer el impacto negativo que la heterogeneidad de las varianzas ejerce sobre las tasas de error Tipo I, diversas soluciones se hallan disponibles hoy en día. Sirva de botón de muestra las cinco que siguen: (a) las encaminadas a estabilizar las varianzas utilizando alguna transformación de la familia Box-Cox (por ejemplo, Budescu y Appelbaum, 1981); (b) las basadas en corregir los valores críticos ajustando los grados de libertad (Kleijnen, Cremers y van Belle, 1985); (c) las basadas en utilizar procedimientos no paramétricos (por ejemplo, Wang y Akritas, 2006); (d) las basadas en imputar los datos faltantes; y (e) las caracterizadas por utilizar métodos de estimación basados en el principio de máxima verosimilitud. Se excusa decir que ninguna solución permite mantener controladas las tasas de error al nivel de significación nominal en todas las situaciones, ni todas las soluciones presentan el mismo grado de robustez.

Una solución que no requiere la igualdad de las varianzas, y que puede ser adoptado para contrastar hipótesis y comparar modelos en diseños factoriales, se basa en generalizar las ideas desarrolladas por Brown y Forsythe (BF, 1974) en el contexto del ANOVA de una sola vía. El estadístico BF utiliza el mismo numerador que el contraste *F*-ANOVA, pero altera el denominador explotando la propiedad según la cual tanto el numerador como el denominador tienen el mismo valor esperado bajo hipótesis nula. A pesar de la popularidad de la prueba, la aproximación propuesta originalmente por Brown y Forsythe puede resultar inadecuada (Mehrotra, 1997). El problema radica en que ellos asumen que el numerador se comporta como una variable ji-cuadrado con grados de libertad igual al número de tratamientos menos uno, χ^2(a-1). Sin embargo, una combinación lineal de variables $\chi^2(1)$ puede que no resulte bien aproximada por una variable χ^2(a-1); particularmente, cuando una varianza domina al resto. En consecuencia, el propósito del presente trabajo radica en generalizar la prueba BF a diseños factoriales siguiendo un procedimiento similar al propuesto por Vallejo y Ato (2006) y Vallejo, Arnau y Ato (2007) en el ámbito de los diseños univariados y multivariados de medidas repetidas.

Definición de la prueba estadística

Considérese un diseño en el cual los n_{jk} $(\Sigma_j\Sigma_k n_{jk}= N;\ j= 1,\ldots,J;\ k= 1,\ldots,K)$ 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 $e_{ijk} \sim N(0,\sigma^2)$. Bajo este modelo, el investigador suele estar interesado en determinar los efectos de *A, B* y *AB,* cuyas sumas de cuadrados pueden ser expresadas como sigue:

$$
\begin{aligned}
Q_A &= N\bar{\mathbf{y}}'\left[\mathbf{C}_A\left(\mathbf{C'}_A \times \mathbf{C}_A\right)^{-1}\mathbf{C'}_A\right]^{-1}\bar{\mathbf{y}},\\
Q_B &= N\bar{\mathbf{y}}'\left[\mathbf{C}_B\left(\mathbf{C'}_B \times \mathbf{C}_B\right)^{-1}\mathbf{C'}_B\right]^{-1}\bar{\mathbf{y}},\\
Q_{AB} &= N\bar{\mathbf{y}}'\left[\mathbf{C}_{AB}\left(\mathbf{C'}_{AB} \times \mathbf{C}_{AB}\right)^{-1}\mathbf{C'}_{AB}\right]^{-1}\bar{\mathbf{y}},\\
Q_R &= Tr(\mathbf{M\Sigma})
\end{aligned}
\tag{1}
$$

donde $\mathbf{C}_{[.]}$ es una matriz de contrastes cuyo orden depende de la hipótesis factorial a probar, $\bar{\mathbf{y}}$ es un vector de orden *JK*×1 obtenido tras concatenar verticalmente las medias de $\bar{\mathbf{y}}_j(= \hat{\mu}_{j1},\ldots,\hat{\mu}_{jK})$, $\mathbf{M}(= \mathbf{C'}\,(\mathbf{CDC'})^{-1}\mathbf{CD})$ es una matriz cuadrada cuya traza es igual al rango de la matriz $\mathbf{C}_{[.]}$ de interés, $\mathbf{D}$ es una matriz diagonal con $1/n_{11},\ldots,1/n_{jk}$ en su diagonal y $\mathbf{\Sigma}= +_{j=1}+_{k=1}\,\sigma^2_{jk}N/n_{jk}$.

Las esperanzas matemáticas de estas formas cuadráticas son:

$$
\begin{aligned}
E(\mathrm{Q_A}) &= \sum_{j=1}^{a}\sum_{k=1}^{b}\tilde{n}_j\sigma^2_{jk} + \left(\sum_{j=1}^{a}\tilde{n}_j - \frac{\sum_j \tilde{n}_j^2}{N}\right)\alpha_j^2\\
E(\mathrm{Q_B}) &= \sum_{j=1}^{a}\sum_{k=1}^{b}\tilde{n}_k\sigma^2_{jk} + \left(\sum_{k=1}^{b}\tilde{n}_j - \frac{\sum_k \tilde{n}_k^2}{N}\right)\beta_k^2\\
E(\mathrm{Q_{AB}}) &= \sum_{j=1}^{a}\sum_{k=1}^{b}\tilde{n}_{jk}\sigma^2_{jk} + \left(\sum_{j=1}^{a}\sum_{k=1}^{b}\tilde{n}_{jk} - \frac{\sum_j\sum_k \tilde{n}_{jk}^2}{N}\right)(\alpha\beta)^2_{jk}\\
E(\mathrm{Q_R}) &= \sum_{j=1}^{a}\sum_{k=1}^{b}\left(n_{jk}-1\right)\sigma^2_{jk}.
\end{aligned}
\tag{2}
$$

De hallarse el diseño balanceado, los multiplicadores $\tilde{n}_j$, $\tilde{n}_k$ y $\tilde{n}_{jk}$ se simplifican a $n_{jk}(a\text{-}1)$, $n_{jk}(b\text{-}1)$ y $n_{jk}(a\text{-}1)(b\text{-}1)$, respectivamente.

Bajo la apropiada hipótesis nula, cada una de las formas cuadráticas anteriores puede ser satisfactoriamente aproximada mediante una distribución ji-cuadrado escalada con adecuados grados de libertad $[g\chi^2_{(\nu)}]$. Por tanto, la razón de dos formas cuadráticas independientes, por ejemplo Q_A y Q_R, será aproximada ajustando el numerador y el denominador de la razón mediante la distribución χ^2 como sigue:

$$
Q_A = b\sum_{j=1}^{a} n_{jk}\left(\bar{Y}_{\cdot j\cdot} - \bar{Y}\right)^2 \sim g_1\chi^2(\nu_1)
\tag{3}
$$

y

$$
Q_R = \sum_{j=1}^{a}\sum_{k=1}^{b}\left(n_{jk}-1\right)^2 s^2_{jk} \sim g_2\chi^2(\nu_2)
\tag{4}
$$

donde *g* es un factor de escala y $\chi^2(\nu)$ es una distribución ji-cuadrado central con ν grados de libertad. Tanto *g* como ν pueden ser determinados resolviendo simultáneamente las ecuaciones $g_i\nu_i= E[\mathrm{Q}_i]$ y $2g_i^2\nu_i= V[\mathrm{Q}_i]$, para *i*= 1, 2. Bajo hipótesis nula, normalidad y heterogeneidad, simple pero tediosa álgebra pone de relieve que $g_1\nu_1= g_2\nu_2$, así como:

$$
\nu_{A1} = \frac{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m(A)_{ii}\,\sigma^2_{jk}\right]^2}{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m^{\bullet}(A)_{ii}\,\sigma^2_{jk}\right]^2 \sum_{j=1}^{a}\sum_{k=1}^{b}\sigma^4_{jk}\left[m(A)_{ii} - m^{\bullet}(A)_{ii}\right]}
\tag{5}
$$

y

$$
\nu_{A2} = \frac{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m(A)_{ii}\,\sigma^2_{jk}\right]^2}{\sum_{j=1}^{a}\sum_{k=1}^{b}\left(\frac{1}{n_{jk}-1}\right)\left[\sigma^4_{jk}\, m(A)^2_{ii}\right]}
\tag{6}
$$

donde m_{ii} es el *i*-ésimo elemento diagonal de la matriz $\mathbf{M}_A(=\mathbf{C'}_A(\mathbf{C}_A\mathbf{DC'}_A)^{-1}\mathbf{C}_A\mathbf{D})$ y $m^{\bullet}_{ii}$ es su complementario. Aquí $\mathbf{C}_A$[=

$(\mathbf{1}_{a\text{-}1} \mid \mathbf{I}_{a\text{-}1}) \otimes (1/b)\mathbf{1}_b^{\prime}]$ es una matriz de contrastes y $\mathbf{D}$ es una matriz diagonal definida con anterioridad. Procediendo de manera similar los grados de libertad (gl) correspondientes al efecto principal del factor B pueden ser aproximados como:

$$\nu_{B1} = \frac{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m(B)_{ii}\sigma_{jk}^{2}\right]^{2}}{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m^{\bullet}(B)_{ii}\sigma_{jk}^{2}\right]^{2}\sum_{j=1}^{a}\sum_{k=1}^{b}\sigma_{jk}^{4}\left[m(B)_{ii} - m^{\bullet}(B)_{ii}\right]} \tag{7}$$

y

$$\nu_{B2} = \frac{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m(B)_{ii}\sigma_{jk}^{2}\right]^{2}}{\sum_{j=1}^{a}\sum_{k=1}^{b}\left(\frac{1}{n_{jk}-1}\right)\left[\sigma_{jk}^{4} m(B)_{ii}^{2}\right]} \tag{8}$$

A su vez, *gl* correspondientes a la interacción de A × B pueden ser aproximados como:

$$\nu_{AB1} = \frac{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m(AB)_{ii}\sigma_{jk}^{2}\right]^{2}}{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m^{\bullet}(AB)_{ii}\sigma_{jk}^{2}\right]^{2}\sum_{j=1}^{a}\sum_{k=1}^{b}\sigma_{jk}^{4}\left[m(AB)_{ii} - m^{\bullet}(AB)_{ii}\right]} \tag{9}$$

y

$$\nu_{AB2} = \frac{\left[\sum_{j=1}^{a}\sum_{k=1}^{b} m(AB)_{ii}\sigma_{jk}^{2}\right]^{2}}{\sum_{j=1}^{a}\sum_{k=1}^{b}\left(\frac{1}{n_{jk}-1}\right)\left[\sigma_{jk}^{4} m(AB)_{ii}^{2}\right]} \tag{10}$$

Finalmente, aplicando el teorema 6.1 de Box (1954) la distribución central de la razón F^* puede ser aproximada como sigue:

$$F^{*} = \frac{Q_i}{Tr(\mathbf{M\Sigma})} \sim \left(\frac{g_1\nu_1}{g_2\nu_2}\right) F(\nu_1, \nu_2) \tag{11}$$

donde Q_i denota la suma de cuadrados de los efectos principales e interacción definidos en (1) y $F(\nu_1,\nu_2)$ la distribución F central con ν_1 y ν_2 grados de libertad.

Método

En orden a evaluar la robustez del enfoque modificado de Brown y Forsythe (1974) llevamos a cabo un estudio de simulación usando un diseño factorial de dos factores con J= 2 y K= 5. El desempeño del estadístico propuesto fue comparado con el del modelo lineal clásico, tal y como está implementado en el módulo *Proc GLM* del programa SAS (SAS Institute, 2005, versión 9.1.3). Para tal fin fueron manipuladas las cuatro variables siguientes:

1. *Tamaño de muestra de las celdas*. El comportamiento de ambos estadísticos fue investigado usando dos tamaños de muestra globales distintos: N= 60 y N= 120. Para N= 60, el tamaño de las celdas fue: n_{1k}= (4,7,3,6,5) y n_{2k}= (3,5,7,9,11). Mientras que para N= 120, el tamaño fue: n_{1k}= (8,14,6,12,10) y n_{2k}= (6,10,14,18,22).
2. *Relación entre el tamaño de las celdas y el tamaño de las varianzas*. Se investigó el comportamiento del enfoque bajo relaciones positivas, negativas y nulas. Una relación positiva implica que la combinación de tratamiento de menor tamaño se asocia con la varianza menor, mientras que una relación negativa implica que la combinación de tratamiento de menor tamaño se asocia con la varianza mayor. Si el tamaño de las varianzas permanece constante a través de las distintas celdas la relación se considera nula.
3. *Valor del parámetro de heterogeneidad*. El grado de heterogeneidad de las varianzas incluido en el presente trabajo tenía dos niveles. Bajo la primera condición las varianzas no se desviaban del patrón de homogeneidad requerido por el modelo lineal general (σ_{jk}^2= σ^2= 1), mientras que bajo la segunda condición las varianzas se obtuvieron como sigue: σ_{1k}^2= $[(1+1) / 2]^2$ y σ_{2k}^2= $[(1+jk) / 2]^2$, con k= 1,…, K.
4. *Forma de la distribución de la población.* Específicamente, además de la distribución normal (γ_1= 0 y γ_2= 0), también fue investigada una distribución exponencial doble o de Laplace (γ_1= 0 y γ_2= 3) y otra exponencial (γ_1= 2 y γ_2= 6).

Las muestras de datos utilizadas para generar las distribuciones normales y no normales fueron obtenidas con la función RANNOR del SAS. Bajo la primera condición, las puntuaciones de la variable dependiente fueron creadas usando el modelo lineal siguiente: Y_{ijk}= σ_{jk}×Z_{ijk}, donde Z_{ijk} es una variable normal estandarizada. Las distribuciones no normales fueron obtenidas transformando las puntuaciones Z_{ijk} mediante el método de potencia de Fleishman (1978). Específicamente, los valores de la variable dependiente fueron obtenidos usando el modelo lineal que sigue: Y_{ijk}= σ_{jk}×Z^*_{ijk}, donde Z^*_{ijk}= a+[(dZ_{ijk} + c)Z_{ijk} + b]Z_{ijk} y a, b, c y d son las constantes de Fleishman que permiten generar datos con el deseado grado de sesgo y curtosis.

Para cada una de las condiciones del diseño se crearon cinco mil conjuntos de datos replicados. Para realizar los cálculos se utilizó un MACRO escrito en lenguaje 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 (α= .05), el intervalo utilizado para definir la robustez de las pruebas fue $.025 \leq \hat{\alpha} \leq 0.75$.

La tabla 1 contiene las tasas de error halladas cuando cada efecto principal se contrasta comparando la suma de cuadrados residual del modelo interactivo, con la suma de cuadrados residual obtenida tras eliminar del modelo interactivo el efecto referido a la hipótesis nula de interés. Básicamente, los resultados ponen de relieve los dos aspectos siguientes:

Tabla 1
Tasas de error asociadas con la suma de cuadrados tipo III del diseño factorial 2×5

PH	N	R(N/V)	Normal H_F	Normal H_C	Normal H_I	Laplace H_F	Laplace H_C	Laplace H_I	Exponencial H_F	Exponencial H_C	Exponencial H_I
Procedimiento BFM											
PH_1	60	=	4.95	3.67	3.51	4.71	3.85	3.55	4.46	2.68	2.86
PH_1	120	=	4.93	4.81	4.66	4.92	3.69	3.73	4.61	2.59	2.71
PH_2	60	+	5.01	2.50	2.88	4.63	**1.95**	**2.19**	4.55	**1.60**	**1.67**
PH_2	60	–	6.34	5.00	4.89	4.50	3.86	3.88	7.39	4.09	4.13
PH_2	120	+	4.90	2.83	2.95	4.66	2.67	2.73	5.35	**2.05**	**2.10**
PH_2	120	–	5.52	4.59	4.67	5.24	3.70	3.87	6.28	3.97	4.25
Procedimiento GLM											
PH_1	60	=	5.08	5.07	4.97	4.89	4.76	4.82	5.06	5.14	3.06
PH_1	120	=	5.00	5.11	4.77	5.24	4.82	4.70	4.82	4.81	5.01
PH_2	60	+	**1.03**	**0.63**	**0.59**	**0.95**	**0.78**	**0.64**	**1.78**	**1.02**	**1.14**
PH_2	60	–	**9.83**	**12.77**	**12.71**	**9.36**	**11.79**	**11.84**	**11.12**	**11.53**	**11.09**
PH_2	120	+	**0.63**	**0.63**	**0.83**	**0.94**	**0.58**	**0.45**	**1.36**	**0.84**	**0.73**
PH_2	120	–	**10.05**	**12.59**	**11.90**	**9.33**	**11.61**	**12.06**	**9.79**	**12.01**	**12.02**

Distribución normal = (γ_1= 0 & γ_2= 0); distribución Laplace= (γ_1= 0 & γ_2= 3); distribución exponencial= (γ_1= 2 & γ_2= 6); γ_1= índice de simetría; γ_2= índice de curtosis. N= tamaño de muestra; PH_1= patrón homogéneo; PH_2= patrón heterogéneo; H_F= hipótesis filas; H_C= hipótesis columnas; H_I= hipótesis interacción; R(N/V)= relación entre el tamaño de las celdas y el tamaño de las varianzas. La negrita denota valores que están fuera del intervalo 0.025 – 0.075

1. Cuando la relación entre el tamaño de las celdas y el tamaño de las varianzas era nula o negativa, las tasas de error obtenidas con la generalización del test BF se encontraban siempre dentro de las bandas del criterio de robustez de Bradley. Este patrón de resultados se produjo con independencia del tamaño de muestra, grado de heterogeneidad y forma de la distribución de la población. *Proc GLM*, por su parte, controlaba las tasas de error cuando la relación entre el tamaño de las celdas y el tamaño de las varianzas era nula. Sin embargo, las tasas de error resultaron sustancialmente distorsionadas cuando las varianzas fueron heterogéneas. En concreto, *Proc GLM* se volvía excesivamente liberal cuando la relación reseñada era negativa.
2. Cuando la relación entre el tamaño de las celdas y el tamaño de las varianzas era positiva ninguno de los dos procedimientos examinados controlaba adecuadamente las tasas de error. Las tasas de error obtenidas con el enfoque BF resultaban inferiores al alfa teórico para el factor *B* y la interacción *AB*, mientras que el patrón de resultados obtenido con *Proc GLM* fue conservador bajo buena parte de los efectos examinados. Por lo que respecta al enfoque *Proc GLM*, los resultados de la tabla 1 indican un sustancial efecto de la falta de homogeneidad y un ligerísimo efecto de la ausencia de normalidad.

Discusión

El propósito de la presente investigación ha sido doble. Por un lado, nos hemos centrado en corregir y generalizar el procedimiento de Brown y Forsythe a diseños factoriales de dos vías y, por otro lado, hemos comparado la robustez del enfoque BF con la *Proc GLM* del SAS en diseños carentes de ortogonalidad, homogeneidad y normalidad. Estudios previos han descubierto que el enfoque *GLM* no es robusto al incumplimiento de los supuestos distribucionales (Keselman, Carriere y Lix, 1995). Por su parte, el enfoque BF modificado se ha mostrado robusto cuando ha sido usado para analizar datos longitudinales con matrices de dispersión heterogéneas (Vallejo, Fernández, Herrero y Conejo, 2004; Vallejo y Livacic-Rojas, 2005; Vallejo y Ato, 2006). Nuestros resultados, además de ser consistentes con los obtenidos en los estudios reseñados, ofrecen nuevos descubrimientos para ayudar a los investigadores a seleccionar alternativas viables para contrastar efectos factoriales.

Para concluir queremos efectuar dos breves comentarios. En primer lugar, los resultados son limitados a las condiciones examinadas en el estudio, aunque pueden ser generalizadas a un rango de condiciones más amplio. Por ejemplo, los estadísticos desarrollados en el presente trabajo pueden ser extendidos para contrastar sumas de cuadrados Tipo I (enfoque secuencial), Tipo II (enfoque de ajuste de constantes de Yates) y Tipo III (método de cuadrados de medias ponderadas de Yates) a diseños con dos o más variables, tanto univariados como multivariados. Para ello tan sólo se requiere formular adecuadamente las hipótesis nulas y construir convenientemente las matrices de proyección ortogonal definidas en la ecuación (1). En segundo lugar, cuando la forma de la distribución esté fuertemente sesgada sugerimos considerar la posibilidad de usar otras pruebas alternativas (por ejemplo, pruebas basadas en computación intensiva), pues es muy probable que el enfoque se vuelva excesivamente conservador.

Agradecimientos

Este trabajo ha sido financiado mediante un proyecto de investigación concedido por el Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica del Ministerio de Educación y Ciencia (Ref.: SEJ2005-01883).

Referencias

Appelbaum, M.I., y Cramer, E.M. (1974). Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin, 81*, 335-343.

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.

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

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

Budescu, D.V., y Appelbaum, W.G. (1981). Variance stabilizing transformations and the power of the F test. *Journal of Educational Statistics, 6,* 55-74.

Carlson, J.E., y Timm, N.H. (1974). Analysis of nonorthogonal fixed effect designs. *Psychological Bulletin, 81,* 563-570.

Fisher, R.A. (1926). The arrangement of field experiments. *Journal of Ministry Agriculture of Great Britain*, *33*, 503-513.

Fleishman, A.I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43,* 521-532.

Herr, D.G., y Gaebelein, J. (1978). Nonorthogonal two-way analysis of variance. *Psychological Bulletin, 85*, 207-216.

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

Kleijnen, J.P.C., Cremers, P., y van Belle, F. (1985). The power of weighted and ordinary least squares with estimated unequal variances in experimental designs. *Communications in Statistics. Simulation and Computation 14,* 85-102.

Mehrotra, D.V. (1997). Improving the Brown-Forsythe solution to the generalized Behrens-Fisher problem. *Communication in Statistics-Simulation and Computation, 26*, 1139-1145.

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

Overall, J.E., y Spiegel, D.K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin, 72,* 311-322.

Vallejo, G., Arnau, J., y Ato, M. (2007). Comparative robustness of recent methods for the analysis of multivariate repeated measures designs. *Educational & Psychological Measurement*, *67*, 1-23.

Vallejo, G., y Ato, M. (2006). Modified Brown-Forsythe procedure for testing interaction effects in split-plot designs. *Multivariate Behavioral Research*, *41,* 549-578.

Vallejo, G., Fernández, P., Herrero, J., y Conejo, N.M. (2004). Alternatives procedures for testing fixed effects in repeated measures designs when assumptions are violated. *Psicothema, 16,* 498-508.

Vallejo, G., y Livacic-Rojas, P.E. (2005). A comparison of two procedures for analyzing small sets of repeated measures data. *Multivariate Behavioral Research, 40,* 179-205.

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

Yates, F. (1964). Sir Ronald Fisher and the design of experiments. *Biometrics, 20,* 307-321.