Psicothema 2007. Vol. 19, nº 4, pp. 673-678
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

METODOLOGÍA

Comparación de la potencia de nuevos enfoques para analizar datos de medidas repetidas

Pablo Livacic-Rojas, Guillermo Vallejo* y Paula Fernández*
Universidad de Santiago de Chile y * Universidad de Oviedo

Este trabajo compara la sensibilidad de cinco modernas técnicas analíticas para detectar los efectos de un diseño de medidas parcialmente repetidas cuando se incumplen los supuestos del tradicional enfoque ANOVA, a saber: el enfoque del modelo mixto ajustado mediante el módulo Proc Mixed del SAS, el enfoque Boostrap-F, el enfoque multivariado de Brown-Forsythe, el enfoque multivariado de Welch-James y el enfoque multivariado de Welch-James con estimadores robustos. Con anterioridad, Livacic-Rojas, Vallejo y Fernández habían descubierto que los métodos examinados aquí eran comparables en términos de sus tasas de error Tipo I. Los resultados obtenidos sugieren que tanto el enfoque del modelo mixto como los enfoques de Brown-Forsythe y Welch-James controlaban satisfactoriamente las tasas de error de Tipo II correspondientes a los efectos principales de las ocasiones de medida bajo la mayoría de las condiciones evaluadas.

Power comparison of new tests to analyze repeated measures data. This work compares the sensitivity of five modern analytical techniques for detecting the effects of a design with measures which are partially repeated when the assumptions of the traditional ANOVA approach are not met, namely: the approach of the mixed model adjusted by means of the SAS Proc Mixed module, the Bootstrap-F approach, the Brown-Forsythe multivariate approach, the Welch-James multivariate approach and Welch-James multivariate approach with robust estimators. Previously, Livacic-Rojas, Vallejo and Fernández found out that these methods are comparable in terms of their Type I error rates. The results obtained suggest that the mixed model approach, as well as the Brown-Forsythe and Welch-James approaches, satisfactorily controlled the Type II error rates corresponding to the main effects of the measurement occasions under most of the conditions assessed.

Los diseños de medidas repetidas se utilizan para evaluar el cambio del comportamiento de los sujetos a través del tiempo en función de los tratamientos u otras variables (biológicas, psicológicas o sociales) mediante el contraste de hipótesis en uno o más grupos. Dadas estas características, los diseños de medidas repetidas se han aplicado en diferentes áreas de la Psicología y en distintas ciencias. A su vez, estos diseños presentan una estructura analítica que requiere considerar el número y naturaleza de las fuentes de variación (tipo de tratamientos, efectos asociados al tiempo), la regla de asignación (aleatoria o no), el número de variables dependientes y las características de la muestra en uno o más grupos (Fernández, Livacic-Rojas y Vallejo, en prensa). La estructura más común de diseño de medidas repetidas es aquella que implica un factor entre-sujetos y otro factor intra-sujetos. En este contexto, las unidades experimentales (i= 1,…, n) se seleccionan aleatoriamente para cada nivel del factor entre-sujetos (j= 1,…, p) y se registran las respuestas en cada uno de los niveles del factor intra-sujetos (k= 1,…, q).

Los métodos de análisis más habitualmente utilizados para evaluar los efectos fijos del diseño han sido el Univariado (ANOVA) y el Multivariado (MANOVA). Sin embargo, el hecho de que los datos que se derivan de investigaciones en psicología no siempre se ajustan a los supuestos exigidos por ellos, esto es, la violación del supuesto de independencia de las puntuaciones, la desviación de la normalidad de las mismas y la ausencia de esfericidad y de homogeneidad de las matrices de dispersión, influyó en el desarrollo de métodos alternativos.

Por ejemplo, cuando los datos no cumplen con el supuesto de la esfericidad, se han desarrollado procedimientos que utilizan

Fecha recepción: 19-12-06 • Fecha aceptación: 22-3-07
Correspondencia: Pablo Livacic-Rojas
Facultad de Humanidades - Escuela de Psicología
Universidad de Santiago de Chile - Avenida Ecuador 3650 - Tercer piso
3650 Santiago (Chile)
E-mail: plivacic@usach.cl

grados de libertad modificados mediante procedimientos tipo Box, el análisis multivariado de la varianza (MANOVA), y también, una combinación de ambos (Blanca, 2004; Vallejo y Lozano, 2006). De otra parte, cuando se incumple el supuesto de igualdad de las matrices de dispersión y el diseño se encuentra desequilibrado, la literatura especializada sugiere desechar el uso de estos procedimientos debido a que exhiben dificultades para controlar las tasas de error de Tipo I (Vallejo, Cuesta, Fernández y Herrero, 2006). Para solventar esta dificultad se han desarrollado nuevos métodos, tales como el enfoque de la Aproximación General Mejorada desarrollado por Huynh (1978) y modificado por Algina y Oshima (1995), el procedimiento de Welch-James desarrollado por Johansen (1980) e implementado en el ámbito de los diseños de medidas repetidas por Keselman, Carriere y Lix (1993), la versión multivariada del procedimiento de Brown-Forsythe (1974) propuesta por Vallejo, Fidalgo y Fernández (2001), la técnica bootstrap desarrollada por Efron y colaboradores (Efron y Tibshirani, 1993) y el modelo lineal mixto.

Por su parte, los estudios relacionados con la sensibilidad de las pruebas estadísticas muestran que este tema tan sólo ha recibido la atención que se precisa en los últimos diez años (Botella, 2002; Maxwell, 2004). Según estos autores, el panorama de los estudios de potencia en Psicología no es del todo alentador debido a que son pocas las áreas donde se han logrado niveles de potencia superiores a 0,80 con tamaños del efecto moderados. Asociado a esto, se han observado bajas tasas de potencia cuando los grupos son de menos de 30 sujetos, con independencia del cumplimiento o no de los supuestos del modelo. Esta situación preocupa, porque es habitual que la investigación en Psicología se caracteriza en parte por tener muestras pequeñas, grupos desequilibrados y problemas multideterminados (Keselman, Algina, y Kowalchuck, 2002).

Dadas las consecuencias derivadas de los bajos niveles de potencia y al bajo control de las tasas de error de tipo I que exhiben los estudios en Psicología surge la necesidad de aplicar procedimientos que sean de alta flexibilidad para abordar estas dificultades y, con ello, realizar una estimación más rigurosa de los parámetros. Bajo estas condiciones, el modelo lineal general está siendo desplazado por el modelo lineal mixto y pruebas estadísticas alternativas que muestran una mejor potencia de prueba cuando se incumplen los supuestos del modelo clásico.

En el conjunto de los procedimientos disponibles, el MLM presenta diferencias con respecto de los otros procedimientos, a saber, permite analizar los datos con observaciones perdidas, determinar *a priori* la estructura de covarianza que subyace a los datos mediante los criterios AIC y BIC, manejar covariadas cambiantes a lo largo del tiempo e incrementar así la sensibilidad de los tratamientos (Kowalchuck, Keselman, Algina y Wolfinger 2004), aplicarse con independencia de que el número de mediciones exceda al número de unidades de observación, obtener errores estándar más adecuados de estos estimadores (Zimmerman y Núñez-Antón, 2001), también permite que el modelo contenga tanto variables de efectos fijos como de efectos aleatorios (Raudenbush y Bryk, 2002), extender su uso para el manejo de medidas repetidas de naturaleza categórica (Davis, 2002) y ajustarse a condiciones donde los sujetos tengan diferente número de observaciones y los intervalos de tiempo puedan ser específicos para cada uno de ellos, aspecto éste que posibilita modelar las variaciones entre e intra sujetos tanto para datos completos como incompletos (Vallejo, Fernández y Secades, 2004; Littell, 2002).

No obstante, conviene señalar que el MLM tiene el inconveniente de que los estimadores de precisión de la estructura de medias del modelo se basan en su distribución asintótica, lo cual puede provocar serios problemas cuando se trabaja con muestras reducidas (Chen y Wei, 2003; Wolfinger, 1996). Por lo tanto, cuando los tamaños de los grupos son pequeños, las inferencias que el investigador realice pueden ser incorrectas. Para solventar esta dificultad, Fai y Cornelius (1996) y Kenward y Roger (1997) han desarrollado dos soluciones basadas en modificar los grados de libertad del error y en ajustar el estadístico de Wald, ambas soluciones están implementadas en las versiones 8 y 9 del paquete estadístico SAS. Además, como señalan Keselman, Algina, Kowalchuk y Wolfinger (1999), el investigador debe estar atento al utilizar este enfoque debido a que en ocasiones exhibe dificultades con la identificación correcta de las estructuras de covarianza.

En cuanto a otros procedimientos de análisis, según Vallejo et al. (2006) la técnica bootstrap se ha aplicado escasamente en los diseños de medidas repetidas, aunque se ha implementado en otros ámbitos desde la década de los 80 a partir de los trabajos de Lunneborg y Tousignant (1985) y Wasserman y Böckenholt (1989). Estos autores se han referido a ella como una alternativa a las técnicas tradicionales (t, F) basadas en el muestreo de poblaciones.

Así las cosas, el objetivo del presente trabajo se centra en evaluar comparativamente la potencia de ocho recientes enfoques analíticos cuando se incumplen los supuestos de normalidad y homogeneidad, tanto conjuntamente como por separado. En concreto, los procedimientos estudiados han sido los siguientes: la Aproximación General Mejorada (AGM), la versión multivariada del procedimiento modificado de Brown Forsythe (BF), la extensión multivariada de los enfoques de Welch y James (WJ), el procedimiento de WJ con medias recortadas (MR), la técnica no paramétrica bootstrap (BOOT), el modelo mixto sin ajustar grados de libertad (F), el modelo mixto con grados de libertad ajustados mediante el método de Kenward y Roger (KR) y el método tipo Satterthwaite de Fai y Cornelius (S). Hasta la fecha no se ha examinado simultáneamente la sensibilidad de las ocho técnicas reseñadas. En orden a evitar redundancias, se omite la definición formal de los métodos examinados. No obstante, una descripción detallada de los mismos se encuentra en los trabajos de Livacic-Rojas et al. (2006), Vallejo y Livacic Rojas (2005) y Vallejo et al. (2006).

Método

En orden a evaluar la sensibilidad de los enfoques señalados, se llevó a cabo un estudio de simulación usando un diseño de medidas repetidas que tenía un factor entre sujetos ($p= 3$) y un factor intra sujetos ($k= 4$). En él se comparó la potencia de los enfoques propuestos cuando se incumplían los supuestos de esfericidad multimuestral y normalidad, tanto conjuntamente como por separado. Para tal fin se manipularon cinco variables: (a) el tamaño de muestra total, (b) la relación entre el tamaño de los grupos y el de las matrices de dispersión, (c) el tipo de matrices de dispersión, (d) la forma de la distribución de la población y (e) la permutación del vector de medias. Estas variables fueron seleccionadas por ser de uso habitual en investigaciones similares a la actual.

Se analizó el comportamiento de estos procedimientos utilizando tres distintos tamaños de muestras totales: Estos tamaños grupales fueron seleccionados por ser representativos de los encontrados en la práctica en Psicología. Dentro de cada condición, el coeficiente de variación muestral ($\triangle$) se fijó en 0.33, donde

$\Delta = \frac{1}{n}\left[\Sigma_j\left(n_j - \bar{n}\right)^2 / p\right]^{1/2}$, siendo $\bar{n}$ el tamaño promedio de los grupos. Para n= 30, los tamaños grupales fueron: 6, 10 y 14; para n= 45, los tamaños grupales fueron: 9, 15 y 21. Mientras que para n= 60, los tamaños grupales fueron: 12, 20 y 28.

Dado que las relaciones entre las matrices de dispersión y el tamaño de los grupos pueden tener diferentes efectos en las pruebas estadísticas, se investigó el comportamiento de los tres enfoques referidos utilizando diseños equilibrados y no equilibrados. Cuando los diseños estaban equilibrados, la relación entre las matrices de dispersión y el tamaño de los grupos era nula. Cuando los diseños carecían de equilibrio, la relación podía ser tanto positiva como negativa. Una relación positiva implica que el grupo de menor tamaño se asocia con la matriz de dispersión menor, mientras que una relación negativa implica que el grupo de menor tamaño se asocia con la matriz de dispersión mayor. Cuando los tamaños de muestra son iguales, Δ vale 0, mientras que el valor del coeficiente se incrementa cuanto más difiere el tamaño de los grupos entre sí. En este caso, el grado de desequilibrio existente entre los grupos fluctuó entre moderado y severo. El grado de heterogeneidad de las matrices de dispersión incluido en el presente trabajo fue $\sum_1 = \frac{1}{3}\sum_2$ y $\sum_3 = \frac{5}{3}\sum_2$.

El patrón de matrices de dispersión tenía cuatro niveles; en los tres primeros las matrices eran múltiplos unas de otras, mientras que en el cuarto no lo eran. Las estructuras de covarianza usadas para generar los datos simulados en el caso proporcional serán las tres que siguen: coeficientes aleatorios (CA), autorregresiva de primer orden heterogénea (ARH(1)) y no estructurada (NE). Para las tres estructuras reseñadas el valor del parámetro de esfericidad se mantuvo constante en .75.

Por lo que se refiere al caso en que no sean múltiplos unas de otras, las matrices de los diferentes grupos diferirán entre sí. El índice de esfericidad de las matrices de covarianza correspondiente a cada uno de los grupos se fijará igual a 1.00, 0.75 y 0.50, respectivamente. De este modo, aunque las matrices sean distintas, mantienen entre sí la relación siguiente: $tr\,(A'\sum_1 A) = tr\,(A'\sum_2 A) = tr\,(A'\sum_3 A)$. Como en el caso anterior, se manipulará el grado de heterogeneidad de las matrices mediante la relación $\sum_1 = \frac{1}{3}\Sigma_2$ y $\sum_3 = \frac{5}{3}\Sigma_2$. Así pues, se cumple que

$$tr\left(A'\textstyle\sum_1 A\right) = \frac{1}{3}tr\left(A'\textstyle\sum_2 A\right) \text{ y } tr\left(A'\textstyle\sum_3 A\right) = \frac{5}{3}tr\left(A'\textstyle\sum_2 A\right)$$

La cuarta variable investigada fue la forma de la distribución de la variable de medida. Este factor tenía cuatro niveles: normal, ligeramente sesgado, moderadamente sesgado y severamente sesgado. Al igual que en el estudio de Berkovits, Hancok y Nevitt (2000), los valores de los índices de asimetría (γ_1) y apuntamiento (γ_2) seleccionados para generar las distribuciones no normales multivariadas fueron: ligeramente sesgada (γ_1= 1, γ_2= 0.75), moderadamente sesgada (γ_1= 1.75, γ_2= 3.00) y fuertemente sesgada (γ_1= 3.00, γ_2= 21.00). Una de las razones de haber utilizado distribuciones sesgadas se debe a que éstas son representativas de la mayor parte de las investigaciones realizadas con datos aplicados (véase, por ejemplo, Micceri, 1989).

La última variable investigada fue la permutación de la configuración del patrón del vector de medias. Sobre la base del trabajo de Algina y Keselman (1998), la configuración seleccionada fue la de rango máximo. Dentro de cada uno de los tres grupos del diseño se incluyeron las permutaciones que siguen: (-μ,0,0, μ), (-μ,0,μ,0), (-μ,μ,0,0), (0,-μ,0, μ),(0, -μ,μ,0), (0,0,-μ,μ). Cuando el interés se centró en comparar la potencia de las diferentes pruebas para detectar los efectos de la interacción, únicamente el tercer grupo tuvo un vector de medias idéntico al descrito; para los otros dos grupos, el vector de medias fue nulo. Para cada uno de los tres tamaños de muestra manipulados y para cada uno de los efectos implicados en el diseño, se seleccionó un valor de μ que proporcionará un valor objetivo de potencia a priori de 0.80. No hay razón para pensar que el patrón de resultados se vea alterado al utilizar algún otro valor, por ejemplo 1-β= 0.40. El valor de potencia se obtuvo para el clásico enfoque ANOVA de medidas repetidas empleando las funciones FINV y FNONCT del SAS.

Para cada una de las condiciones del diseño, se crearon, con el lenguaje matricial IML del SAS (SAS Institute, 2001), 1000 conjuntos de datos replicados, utilizando la extensión multivariada que Vale y Maurelli (1983) desarrollaron del método de la potencia propuesto por Fleishman (1978). Para cada replicación se ejecutaron los enfoques bootstrap-F, BF, AGM, MLM y WJ y se examinaron las tasas de error de Tipo II usando un nivel de significación del 5%. Los valores críticos del enfoque bootstrap se determinaron mediante el procedimiento descrito por Vallejo et al. (2006), a partir de 599 remuestreos de tamaño obtenidos del conjunto de datos simulados. En la línea de lo aquí expuesto, es oportuno señalar que todos los cálculos se realizaron mediante un programa de simulación escrito en lenguaje de programación SAS/IML. Bajo hipótesis alternativa, las tasas de potencia empíricas se obtuvieron dividiendo el número de veces que cada estadístico excedía su correspondiente valor crítico cuando el vector de medias era distinto de cero por las 1000 réplicas efectuadas.

Resultados

La tabla 1 contiene los valores promedio de los seis vectores de medias para los efectos principales en cada una de las condiciones investigadas y que se corresponden con cada uno de los ocho procedimientos analizados para diferentes distribuciones y estructuras de covarianza. Los resultados muestran los promedios de los vectores de medias para la distribución normal proporcional, distribución no normal proporcional y no normal no proporcional. En estas distribuciones se observa:

1. Los niveles de potencia de 0.80 o más se alcanzan para casi todos los procedimientos y en las diferentes condiciones cuando la estructura de covarianza es CA.
2. Los procedimientos BF, BOOT, MR no alcanzan este nivel de potencia asociado a CA cuando la relación entre las matrices de dispersión y los grupos es neutra y negativa.
3. En términos generales, los niveles de potencia más bajos se alcanzan para casi todos los procedimientos y en las diferentes condiciones cuando la estructura de covarianza es NE.
4. A su vez, los niveles de potencia más bajos se alcanza en promedio para casi todos los procedimientos cuando relación entre la estructura de covarianza y el tamaño de los grupos es negativa.

5. Los procedimientos más potentes, en orden decreciente y obtenido sobre el total de condiciones analizadas, son F, S, KR, WJ, AGM, BF, BOOT y MR. Esta situación se verifica al considerar los resultados obtenidos para todas las condiciones analizadas.

La tabla 2 contiene los valores promedio de los seis vectores de medias para cada una de las condiciones para la distribución normal no proporcional. En esta distribución se puede observar que:

1. Ninguno de los procedimientos alcanza una potencia de 0.80 o más bajo ninguna de las condiciones evaluadas.
2. Los procedimientos más potentes, en orden decreciente y obtenido sobre el total de condiciones analizadas, son F, S, WJ, KR, BF, BOOT, MR y AGM. Es importante señalar que el procedimiento AGM no llega en ninguna de las condiciones evaluadas a una potencia de 0.80. Esta situación se puede verificar considerando los resultados obtenidos para todas las condiciones analizadas.

Tabla 1
Valores promediados de los vectores de media para la potencia en distribuciones normal proporcional, no normal proporcional y no no normal no proporcional

N	EC	F	S	KR	BF	WJ	BOOT	MR	AGM
10-10-10	NE	0.79	0.72	0.68	0.66	0.69	0.50	0.54	0.51
06-10-14	NE	0.79	0.74	0.69	0.70	0.71	0.56	0.60	0.52
14-10-06	NE	0.75	0.64	0.55	0.47	0.58	0.32	0.43	0.42
15-15-15	NE	0.79	0.72	0.70	0.69	0.70	0.62	0.63	0.51
09-15-21	NE	0.77	0.74	0.72	0.74	0.72	0.67	0.66	0.56
21-15-09	NE	0.71	0.64	0.59	0.56	0.60	0.46	0.52	0.40
20-20-20	NE	0.77	0.72	0.70	0.70	0.70	0.67	0.68	0.50
12-20-28	NE	0.76	0.74	0.72	0.73	0.72	0.70	0.69	0.56
28-20-12	NE	0.69	0.64	0.59	0.57	0.60	0.51	0.53	0.40
10-10-10	ARH	**0.80**	0.76	0.70	0.69	0.71	0.46	0.51	0.56
06-10-14	ARH	**0.82**	0.78	0.74	0.74	0.75	0.56	0.59	0.61
14-10-06	ARH	0.76	0.65	0.54	0.56	0.61	0.26	0.39	0.44
15-15-15	ARH	0.79	0.75	0.72	0.72	0.73	0.60	0.58	0.57
09-15-21	ARH	**0.81**	0.78	0.76	0.76	0.76	0.66	0.65	0.56
21-15-09	ARH	0.73	0.65	0.59	0.56	0.59	0.41	0.48	0.43
20-20-20	ARH	0.77	0.75	0.74	0.74	0.74	0.63	0.61	0.57
12-20-28	ARH	**0.80**	0.76	0.77	0.77	0.77	0.70	0.66	0.61
28-20-12	ARH	0.70	0.65	0.59	0.58	0.60	0.48	0.48	0.42
10-10-10	CA	**0.97**	**0.97**	**0.96**	**0.95**	**0.96**	0.72	0.80	**0.94**
06-10-14	CA	**0.96**	**0.95**	**0.95**	**0.96**	**0.97**	**0.82**	**0.83**	**0.96**
14-10-06	CA	**0.94**	**0.93**	**0.93**	0.77	**0.86**	0.49	0.71	**0.83**
15-15-15	CA	**0.97**	**0.95**	**0.97**	**0.96**	**0.96**	0.79	**0.87**	**0.95**
09-15-21	CA	**0.96**	**0.96**	**0.96**	**0.98**	**0.98**	**0.82**	**0.89**	**0.96**
21-15-09	CA	**0.95**	**0.93**	**0.96**	**0.85**	**0.88**	0.61	0.77	**0.84**
20-20-20	CA	**0.97**	**0.97**	**0.97**	**0.97**	**0.96**	**0.87**	**0.87**	**0.95**
12-20-28	CA	**0.96**	**0.96**	**0.96**	**0.98**	**0.98**	**0.91**	**0.85**	**0.97**
28-20-12	CA	**0.94**	**0.94**	**0.94**	**0.88**	**0.89**	0.70	0.75	**0.85**

Leyenda: EC= Estructura de Covarianza; NE= Matriz no estructurada; ARH= Matriz Autorregresiva heterogénea de primer orden; CA= Matriz de Coeficientes Aleatorios; F= Procedimiento F sin grados de libertad corregidos; S= Procedimiento de Fai y Cornelius; KR= Procedimiento de Kenward-Roger; BF= Procedimiento Brown-Forsythe; WJ= Procedimiento de Welch James; BOOT= Bootstap no paramétrico recortado; MR= Procedimiento WJ medias recortadas; AGM= Procedimiento de Aproximación General Mejorada; Negrita= Niveles de Potencia de 0.80 o superior

Tabla 2
Valores promediados de los vectores de media para evaluar la potencia en una distribución normal no proporcional

N	EC	D	F	S	K	BF	WJ	BOOT	MR	AGM
10-10-10	NE	FS	0.76	0.72	0.67	0.66	0.69	0.38	0.42	0.41
06-10-14	NE	FS	0.78	0.73	0.69	0.70	0.70	0.45	0.48	0.48
14-10-06	NE	FS	0.74	0.59	0.56	0.41	0.62	0.33	0.32	0.30
15-15-15	NE	FS	0.76	0.73	0.70	0.69	0.58	0.51	0.50	0.41
09-15-21	NE	FS	0.76	0.74	0.71	0.72	0.72	0.56	0.55	0.48
21-15-09	NE	FS	0.72	0.62	0.61	0.53	0.64	0.40	0.44	0.29
20-20-20	NE	FS	0.75	0.73	0.71	0.70	0.71	0.56	0.53	0.42
12-20-28	NE	FS	0.76	0.73	0.72	0.72	0.72	0.58	0.56	0.49
28-20-12	NE	FS	0.69	0.65	0.61	0.58	0.63	0.43	0.43	0.27

Leyenda: N= Tamaño de los Grupos; EC= Estructura de Covarianza; D= Tipo de Distribución; F= Procedimiento F sin grados de libertad corregidos; S= Procedimiento de Fai y Cornelius; KR= Procedimiento de Kenward-Roger; BF= Procedimiento Brown-Forsythe; WJ= Procedimiento de Welch James; BOOT= Bootstap no paramétrico recortado; MR= Procedimiento WJ medias recortadas; AGM= Procedimiento de Aproximación General Mejorada; NE= Matriz no estructurada.; FS= Distribución Fuertemente Sesgada; Negrita= Niveles de Potencia de 0.80 o superior

Discusión

El propósito de esta investigación ha sido evaluar la potencia de ocho procedimientos para analizar diseños de medidas repetidas en ausencia de esfericidad multimuestral. Específicamente, se ha examinado la potencia estadística de los procedimientos multivariado de Brown-Forsythe (BF, modificado por Vallejo et al., 2001), la técnica bootstrap (BOOT), el modelo mixto con los grados de libertad sin ajustar (prueba F), el procedimiento de Aproximación General Mejorada (AGM), el modelo mixto con los grados de libertad ajustados mediante el procedimiento desarrollado por Kenward-Roger (1997) (KR) y el procedimiento propuesto por Fai y Cornelius (S), el procedimiento de Welch-James con Medias Recortadas (MR), el procedimiento de Welch-James (WJ), cuando los supuestos de normalidad multivariada y homogeneidad de las matrices de dispersión eran violados independiente y simultáneamente. Este objetivo se llevó a cabo manipulando diversos tamaños de muestra generados desde distribuciones sesgadas y con diferentes estructuras de covarianza.

Los resultados de los valores promedio de los vectores en los diferentes procedimientos evaluados muestran que el procedimiento que exhibe mayor potencia es F. Cabe señalar, que este procedimiento es aquel con el nivel más bajo de robustez realizado por Livacic-Rojas et al. (2006). También resulta oportuno señalar que los procedimientos BOOT y MR (que también exhibían baja robustez en el mencionado estudio) presentan en promedio bajos niveles de potencia, lo cual podría relacionarse con el hecho de que son pruebas muy robustas cuando las distribuciones se encuentran fuertemente separadas de la distribución normal.

En cuanto a los procedimientos S, WJ, KR, BF y AGM, es importante destacar que en general presentan un rendimiento moderado respecto a la potencia, aunque son robustos, a excepción de la prueba S. Además, es necesario señalar que la prueba AGM es poco recomendable, dado que, aun siendo robusta, se adapta poco a condiciones propias de los estudios en ambientes aplicados (observaciones perdidas y correlaciones entre las observaciones).

La situación aquí expuesta se confirma al analizar el desempeño de los diferentes procedimientos tanto cuando las matrices de dispersión son y no son proporcionales unas de otras para todas las condiciones analizadas. Si bien el modelo mixto con grados de libertad ajustados (procedimientos KR y S) es el que exhibe mejores resultados en general al adaptarse a las condiciones reales de las investigaciones (AGM es una prueba que no es eficiente cuando existen observaciones perdidas), los procedimientos KR y S presentan el inconveniente de que están basados en el conocimiento previo de las estructuras de covarianza, lo cual es una situación que en la práctica no ocurre.

Finalmente, quisiéramos plantear de cara a investigaciones venideras que es necesario continuar los estudios referidos a la potencia dado que, si bien es cierto, varios procedimientos analizados tienen niveles altos en estudios de simulación, no es menos cierto que, en investigaciones reales, estos fluctúan entre bajos e intermedios.

Agradecimientos

Este trabajo ha sido financiado por la Universidad de Santiago de Chile (Ref.: DICYT 030593LR) y por el Ministerio de Educación y Ciencia de España (Ref.: SEJ-2005-01883).

Referencias

Algina, J., y Oshima (1995). An improved general approximation test for the main effect in a split-plot design. *British Journal of Mathematical and Statistical Psychology, 48,* 149-160.

Algina, J., y Keselman, H.J. (1998). A power comparison of the Welch-James and Improved General Approximation test. *Journal of Educational and Behavioral Statistics, 23,* 152-169.

Berkovits I., Hancock G., y Nevitt J.(2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement, 60,* 877-892.

Blanca, M.J. (2004). Alternativas de análisis estadístico en los diseños de medidas repetidas. *Psicothema, 16,* 509-518.

Botella, J., (2002). Potencia de pruebas alternativas para dos muestras relacionadas con datos perdidos. *Psicothema, 14,* 174-180.

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.

Chen, X., y Wei, L. (2003). A comparison of recent methods for the analysis of small-sample cross-over studies. *Statistics in Medicine, 22*, 2821-2833.

Davis, Ch. (2002). Statistical Methods for the Analysis of repeated measurements. Springer-Verlag. New York. U. S.A.

Efron, B., y Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fai, A., y Cornelius, P. (1996). Approximate F-test of multiple degree of freedom hypotheses in generalized least squares analyzes of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation, 54,* 363-378.

Fernández, P., Livacic-Rojas, P., y Vallejo, G. (en prensa). Cómo elegir la mejor prueba estadística para analizar diseños de medidas repetidas. *International Journal of Clinical and Health Psychology*.

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

Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika, 43*, 161-165.

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

Kenward, M.G., y Roger, H.J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53,* 983-997.

Keselman, H.J., Carriere, M.C., y Lix, L.M. (1993). Testing repeated measures hypotheses when covariance matrices are heterogeneous. *Journal of Educational Statistics, 18,* 305-319.

Keselman H.J., Algina J., Kowalchuck R., y Wolfinger, R. (1999). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics-Simulation and computation, 28 (12),* 2967-2999.

Keselman, H.J., Algina, J., y Kowalchuk, R.K. (2002). A comparison of data analysis strategies for testing omnibus effects in higher-order repeated measures designs. *Multivariate Behavioral Research*, *37*, 331-357.

Kowalchuck, R.K., Keselman, H.J., Algina, J., y Wolfinger, R. (2004). The analysis of repeated measurements with mixed-model adjusted *F* test. *Educational and Psychological Measurements, 64*, 224-242.

Littell, R.C. (2002). Analysis of unbalanced mixed model data: A case study comparison of ANOVA versus REML/GLS. *Journal of Agricultural, Biological, and Environmental Statistics, 7*, 472-490.

Livacic-Rojas, P., Vallejo, G., y Fernández, P. (2006). Diferentes procedimientos alternativos para evaluar la robustez mediante diseños de medidas repetidas. *Revista Latinoamericana de Psicología, 38*.

Lunneborg, C.E., y Tousignant, J.P. (1985). Efron's bootstrap with application to the repeated measures design. *Multivariate Behavioral Research, 20,* 161-178.

Maxwell, S. (2004). The persistence of underpowered studies in Psychological research: Causes, consequences, and remedies. *Psychological Methods, 9,* 147-163.

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

Raudenbush, S.W., y Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: SAGE .

SAS Institute Inc. (2001). *SAS/STAT Software: Version 8.2* (TS MO). Cary, NC: SAS Institute Inc.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Vale, C.D., y Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48,* 465-471.

Vallejo, G., Fidalgo, A.M., y Fernández, P. (2001). Effects of covariance heterogeneity on three procedures for analysing multivariate repeated measures designs. *Multivariate Behavioral Research*, *36*, 1-27.

Vallejo, G., Fernández, J.R., y Secades, R. (2004). Application of a mixed model approach for assessment of interventions and evaluation of programs. *Psychological Reports, 95,* 1095-1118.

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

Vallejo, G., Cuesta, M., Fernández, P.. y Herrero, 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., y Lozano, L. (2006). Modelos de análisis para diseños multivariados de medidas repetidas. *Psicothema, 18,* 293-299.

Wasserman, S., y Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology, 26,* 208-221.

Wolfinger, R. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1,* 205-230.

Zimmerman, D.L., y Núñez-Antón, V. (2001). Parametric modeling of growth curve data: An overview (with comments). *Test, 10,* 1-73.