Psicothema 2012. Vol. 24, nº 3, pp. 449-454
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

Efecto de la violación de la normalidad y esfericidad en el modelo lineal mixto en diseños split-plot

Jaime Arnau[1], Rebecca Bendayan[2], María J. Blanca[2] y Roser Bono[1]
[1] Universidad de Barcelona y [2] Universidad de Málaga

El objetivo del presente trabajo fue evaluar la robustez del Modelo Lineal Mixto, mediante el PROC MIXED del SAS, con la corrección de los grados de libertad por el procedimiento Kenward-Roger (K-R) y con diseños split-plot. El estudio se realizó mediante simulación Monte Carlo con un diseño de tres grupos y cuatro medidas repetidas, asumiendo una matriz de covarianza no estructurada para generar los datos. Las variables del estudio fueron la esfericidad, con valores de épsilon de 0,75 y 0,57; el tamaño de grupo, igual o desigual, y la forma de la distribución. Referente a esta última, se han tenido en cuenta distribuciones no normales, combinando distintos valores de curtosis en cada grupo. Con diseños no balanceados, se analizó también el efecto del emparejamiento (positivo o negativo) del grado de curtosis con el tamaño de grupo. Los resultados muestran que el procedimiento K-R es liberal, sobre todo para el efecto de interacción, bajo determinadas condiciones de violación de la normalidad. Al aplicar este procedimiento con muestras pequeñas se debe considerar la relación entre los valores de curtosis de los grupos y el emparejamiento de la curtosis con el tamaño de grupo como variables relevantes.

Effect of the violation of normality and sphericity in the linear mixed model in split-plot designs. This study aimed to evaluate the robustness of the linear mixed model, with the Kenward-Roger correction for degrees of freedom, when implemented in SAS PROC MIXED, using split-plot designs with small sample sizes. A Monte Carlo simulation design involving three groups and four repeated measures was used, assuming an unstructured covariance matrix to generate the data. The study variables were: sphericity, with epsilon values of 0.75 and 0.57; group sizes, equal or unequal; and shape of the distribution. As regards the latter, non-normal distributions were introduced, combining different values of kurtosis in each group. In the case of unbalanced designs, the effect of pairing (positive or negative) the degree of kurtosis with group size was also analysed. The results show that the Kenward-Roger procedure is liberal, particularly for the interaction effect, under certain conditions in which normality is violated. The relationship between the values of kurtosis in the groups and the pairing of kurtosis with group size are found to be relevant variables to take into account when applying this procedure.

En la investigación actual existe un creciente interés por describir, pronosticar y explicar procesos biológicos, psicológicos y sociales que se producen como consecuencia del paso del tiempo (Fernández, Livacic-Rojas y Vallejo, 2007; Keselman, Algina y Kowalchuk, 2001). En estos casos, los diseños de medidas repetidas se muestran útiles, siendo los diseños *split-plot* o mixtos los más comúnmente utilizados en la investigación en Ciencias Sociales (Fernández, Vallejo, Livacic-Rojas y Tuero, 2010; Keselman, Huberty et al., 1998) y de la Salud (Bono, Arnau y Balluerka, 2007).

El análisis estadístico tradicional, a la vez que el más usado, de los datos provenientes de los diseños de medidas repetidas es el análisis de varianza (ANOVA) mixto. Su aplicación requiere satisfacer los supuestos de normalidad, independencia y esfericidad. El primer supuesto asume que las observaciones de cada unidad de análisis sean extraídas de una población con distribución normal multivariada, el segundo supone la independencia entre las observaciones correspondientes a los distintos sujetos, y el tercero implica que la matriz de covarianza tenga varianza constante de la diferencia entre todos los pares de puntuaciones (Huynh y Feldt,

Fecha recepción: 28-7-11 • Fecha aceptación: 16-11-11
Correspondencia: Rebecca Bendayan
Facultad de Psicología
Universidad de Málaga
29071 Málaga (Spain)
e-mail: bendayan@uma.es

1970; Rouanet y Lepine, 1970). En el caso de diseños *split-plot* se debe satisfacer la esfericidad multimuestra (Huynh, 1978; Mendoza, Toothaker y Crain, 1976). Ésta implica la homogeneidad de las matrices de covarianza dentro de cada nivel del factor intersujeto y esfericidad en la matriz de covarianza común. Cuando las asunciones del modelo se satisfacen, el ANOVA mixto proporciona una prueba adecuada para los efectos del diseño. Sin embargo, en la investigación en Ciencias Sociales y de la Salud, la violación de estos supuestos es frecuente. Los resultados de estudios de simulación Monte Carlo muestran que el estadístico *F* es liberal ante las violaciones de esfericidad (Berkovits, Hancock y Nevitt, 2000; Keselman, Lix y Keselman, 1996; Rogan, Keselman y Mendoza, 1979). Cuando se viola la esfericidad multimuestra, es liberal o conservador según la asociación entre el tamaño de los grupos y la matriz de covarianza sea negativa o positiva, respectivamente (Keselman et al., 1996). Se han propuesto diferentes alternativas de análisis estadístico para los diseños de medidas repetidas cuando se violan los supuestos asociados como las pruebas *F* ajustadas, el análisis multivariante de la varianza, la prueba de la Aproximación General y de Aproximación General Mejorada, procedimiento de Welch-James, y el enfoque del modelo mixto (MLM), entre otras.

En las últimas tres décadas, el MLM ha sido ampliamente desarrollado desde un punto de vista teórico (Cnaan, Laird y Slasor, 1997; Laird y Ware, 1982; Littell, Milliken, Stroup y Wolfinger, 1996). Los MLM permiten modelar la estructura de la matriz de covarianza de los errores a la vez que introducir factores aleatorios y modelar la estructura de la matriz, por lo que resulta de utilidad cuando se violan los supuestos de independencia y varianza constante de los errores. El MLM permite asumir la estructura de la matriz de covarianza que mejor describa los datos. Algunas de las estructuras de matrices de covarianza más comunes son la no-estructurada, de simetría compuesta y autorregresiva de primer orden (Bono, Arnau y Vallejo, 2010; Keselman, Algina et al., 1998). Para seleccionar la estructura de covarianza real se utilizan diferentes índices que representan el ajuste del modelo en términos de discrepancias entre los valores observados y estimados a través del modelo, como el criterio de información de Akaike (Akaike, 1973) y el criterio de información bayesiano de Schwarz (Schwarz, 1978). Una vez seleccionada la matriz de covarianza, se utilizan estadísticos tipo Wald para estimar los efectos fijos de medidas repetidas y de su interacción con el factor intersujeto (Schaalje, McBride y Fellingham, 2002). Estos estadísticos son válidos con tamaños muestrales grandes, aunque con muestras pequeñas o estructuras de covarianza incorrectas es posible que el error Tipo I presente un sesgo positivo (Wright y Wolfinger, 1996). Entre los procedimientos de ajuste de los grados de libertad, los métodos de Satterthwaite (Satterthwaite, 1941; 1946) y Kenward-Roger (K-R; Kenward y Roger, 1997) son una buena alternativa ya que mejoran las propiedades en muestras pequeñas, siendo el procedimiento K-R el que proporciona resultados más precisos (Alnosaier, 2007; Arnau, Bono y Vallejo, 2009; Kowalchuk, Keselman, Algina y Wolfinger, 2004; Schaalje et al., 2002).

Desde que el procedimiento K-R ha sido implementado en el módulo *Proc Mixed* del SAS se han conseguido grandes avances y su robustez ha sido estudiada manipulando distintas condiciones, como el número de niveles de los factores de agrupamiento y de medidas repetidas, tamaño muestral, número de sujetos por grupo, forma de la distribución de los datos, homogeneidad o no de las matrices de covarianza, estructura de las matrices de covarianza, grado de violación de la esfericidad y asociación entre el tamaño de los grupos y matrices de covarianza. Adquiere especial importancia porque se encuentran frecuentemente en la investigación real en Ciencias Sociales y de la Salud las violaciones de la normalidad (Micceri, 1989; Blanca, Arnau, Bono, López-Montiel y Bendayan, en prensa) y de la esfericidad (Huynh, 1978; Rogan et al., 1979), al igual que el tamaño muestral pequeño (Keselman et al., 1998; Fernández et al., 2010).

Los estudios de simulación muestran que el procedimiento K-R es robusto para el efecto principal de medidas repetidas, con esfericidad asumida, ante violaciones de la homogeneidad de varianzas y distintas violaciones de la normalidad, como la distribución lognormal (Kowalchuk et al., 2004), chi-cuadrado con tres grados de libertad (Vallejo et al., 2004) o distribuciones no conocidas con violaciones leves (asimetría= 1; curtosis= 0,75), moderadas (asimetría= 1,75; curtosis= 3) y severa (asimetría= 3; curtosis= 21) (Livacic-Rojas, Vallejo y Fernández, 2006, 2010).

En cuanto al efecto de interacción, los resultados no tienen un patrón muy claro, encontrándose en algunos casos resultados contradictorios. Con esfericidad asumida, se ha encontrado que K-R es robusto con distribución lognormal (Kowalchuk et al., 2004) y conservador con distribución chi-cuadrado con tres grados de libertad (Vallejo et al., 2004). No obstante, con distribuciones no conocidas y con los grados de violación citados anteriormente, se ha encontrado que es robusto (Livacic-Rojas et al., 2010), conservador (Livacic-Rojas et al., 2006, 2010) o liberal (Vallejo y Ato, 2006), principalmente asociados a determinadas condiciones de heterogeneidad de varianzas.

Por otro lado, escasean los estudios que analizan el efecto de la violación conjunta de la esfericidad y normalidad. Vallejo y Ato (2006) analizan este caso solo para el efecto de interacción, encontrando que K-R es a menudo liberal en diferentes condiciones de tamaño muestral y de heterogeneidad de varianzas.

En los estudios de simulación se encuentra una gran variabilidad en cuanto a las variables manipuladas, en general, y a los valores de asimetría y curtosis utilizados para estudiar las violaciones de la normalidad, en particular. Asimismo, se suele suponer que los datos de cada grupo tienen exactamente la misma distribución, lo cual puede no responder a la realidad. La evidencia empírica muestra que el estadístico *F* en el ANOVA intersujeto es robusto ante violaciones moderadas de la normalidad en diseños balanceados cuando los datos presentan la misma distribución en los grupos (Sawilowsky y Blair, 1992; Wu, 2007). Sin embargo, cuando las distribuciones difieren y con diseños no balanceados, el error Tipo I incrementa y la potencia disminuye (Glass, Peckham y Sanders, 1972; Harwell, Rubinstein, Hayes y Old, 1992; Lix, Keselman y Keselman, 1996). No obstante, no se ha investigado esta condición en los diseños *split-plot*, por lo que se considera en este estudio como una de las principales variables de interés.

El objetivo del presente trabajo es evaluar la robustez, mediante la tasa de error Tipo I, del MLM con la corrección de los grados de libertad por el procedimiento K-R con diseños *split-plot* de muestras pequeñas ante violaciones de la normalidad en diferente grado en los distintos grupos. Se introduce una violación leve de la asimetría en todos los grupos y se varía el valor de la curtosis en cada uno de ellos, ya que investigaciones previas han revelado que los efectos de la curtosis son mayores que los de la asimetría (Hopkins y Weeks, 1990). El estudio se ha llevado a cabo mediante simulación Monte Carlo con un diseño de tres grupos y cuatro medidas repetidas, asumiendo una matriz de covarianza no estructurada para generar y ajustar los datos. Las variables del estudio han sido: esfericidad, con valores de épsilon de 0,75 y 0,57; tamaño de grupo, igual o desigual; forma de la distribución, con asimetría de 0,8 y valores distintos de curtosis en los tres grupos, manteniendo una relación de 1:2, 1:3 o 1:4 entre ellos; y emparejamiento curtosis y tamaño de grupo, pudiendo ser positiva si el grupo de mayor tamaño muestral tiene mayor grado de curtosis, o negativa si éste tiene menor grado.

Método

Se ha realizado un estudio de simulación Monte Carlo considerando un diseño *split-plot* con un factor intersujetos con tres niveles (J= 3) y un factor intrasujetos con cuatro ocasiones de medida (K= 4). Los datos fueron generados mediante el método de Fleishman (1978) y generalizado al caso multivariado por Vale y Maurelli (1983) utilizando el sistema SAS/IML. Para generar los datos se asumió homogeneidad de varianzas entre grupos y se utilizó una matriz de covarianza desestructurada (UN) por ser la más común en Ciencias Sociales y de la Salud. Algunos estudios recomiendan asumir este tipo de estructura cuando el número de observaciones es moderado y el tamaño de muestra es pequeño (Chen y Wei, 2003; Kowalchuk et al., 2004). Se manipularon las siguientes variables:

Tamaño muestral de los grupos. El tamaño total de la muestra fue de 30 ya que es uno de los tamaños más frecuentemente usado en las investigaciones psicológicas (Arnau, Bono y Vallejo., 2009; Vallejo, Fernández, Tuero y Livacic-Rojas, 2010). Se manipuló el tamaño de grupo, simulando diseños balanceados y no balanceados, siendo el coeficiente de variación del tamaño de grupo, Δn_j, de 0 cuando los tamaños de grupos eran iguales, y de Δn_j = 0,41 cuando los tamaños de grupo eran diferentes.

Forma de la distribución de la variable medida. Además de la distribución normal se incluyeron distribuciones con asimetría fija de 0,8 combinada con diversos valores de curtosis. Se han utilizado distribuciones con índices de asimetría y curtosis que no se ajustan a ninguna distribución conocida como es frecuente en muchos estudios de simulación de diseños de medidas repetidas (Bauer y Curran, 2003; Berkovits et al., 2000; Vallejo, Ato y Valdés, 2008). En situaciones reales es probable que las variables no se ajusten a la distribución normal, ni presenten la misma distribución en cada grupo, por lo que en este estudio se ha generado una distribución no normal distinta para cada grupo. Los valores de curtosis se han establecido de modo que la relación entre la curtosis de los distintos grupos fuera de 1:2, 1:3 y 1:4.

Emparejamiento curtosis y tamaño de grupo. Se refiere a la forma de relacionar el tamaño de los grupos y el grado de curtosis de la distribución de cada uno de ellos. Cuando el diseño es balanceado, la relación entre el tamaño de los grupos y el grado de curtosis es nula; mientras que en diseños no balanceados, la relación puede ser positiva o negativa. La relación es positiva cuando se empareja el grupo de mayor tamaño muestral con el mayor grado de curtosis y el grupo de menor tamaño muestral con el menor grado de curtosis. La relación es negativa cuando se empareja el grupo de mayor tamaño muestral con el menor grado de curtosis y el grupo de menor tamaño muestral con el mayor grado de curtosis.

Esfericidad. Se consideraron valores de ε de 0,57 y 0,75. La mayoría de los estudios de simulación utilizan como buena aproximación a la esfericidad el valor de 0,75, mientras que para la no-esfericidad usan el valor 0,57 (Arnau et al., 2009; Keselman, Algina, Kowalchuk y Wolfinger, 1999; Lix, Algina y Keselman, 2003).

Se analizan las tasas de error Tipo I asociadas a los efectos de medidas repetidas e interacción del MLM de cada una de las condiciones del estudio, que se resumen en la tabla 1. Para cada combinación se han realizado 10.000 réplicas, ya que diversos estudios plantean esta cantidad de réplicas como la necesaria para obtener resultados adecuados y fiables (Rasch y Guiard, 2004; Robey y Barcikowsky, 1992)

Resultados

Para evaluar la robustez del procedimiento se utilizó el criterio de robustez de Bradley (1978), especialmente recomendado cuando los procedimientos deben asumir una serie de condiciones de violación (Keselman et al., 1999; Kowalchuk et al., 2001). Según

Tabla 1
Distintas combinaciones de las variables examinadas en el estudio

								No normal γ_1= 0,8					
								Relación entre la curtosis de los grupos					
N	**n_1**	**n_2**	**n_3**	**Δn_j**	**Normal γ_1= 0**		**Emparejamiento curtosis y tamaño grupo**	**1:2**		**1:3**		**1:4**	
	Esfericidad				**0,57**	**0,75**		**0,57**	**0,75**	**0,57**	**0,75**	**0,57**	**0,75**
30	10	10	10	0,00	g1: γ_2= 0 g2: γ_2= 0 g3: γ_2= 0		Nulo	g1: γ_2= 0,4 g2: γ_2= 0,8 g3: γ_2= 1,6		g1: γ_2= 0,8 g2: γ_2= 2,4 g3: γ_2= 7,2		g1: γ_2= 0,8 g2: γ_2= 3,2 g3: γ_2= 12	
30	5	10	15	0,41	g1: γ_2= 0 g2: γ_2= 0 g3: γ_2= 0		+	g1: γ_2= 0,4 g2: γ_2= 0,8 g3: γ_2= 1,6		g1: γ_2= 0,8 g2: γ_2= 2,4 g3: γ_2= 7,2		g1: γ_2= 0,8 g2: γ_2= 3,2 g3: γ_2= 12	
30	15	10	5	0,41	g1: γ_2= 0 g2: γ_2= 0 g3: γ_2= 0		–	g1: γ_2= 0,4 g2: γ_2= 0,8 g3: γ_2= 1,6		g1: γ_2= 0,8 g2: γ_2= 2,4 g3: γ_2= 7,2		g1: γ_2= 0,8 g2: γ_2= 3,2 g3: γ_2= 12	

* Nota: N: tamaño muestral total; n_j: tamaño muestral de cada grupo; Δn_j: coeficiente de variación; g_j: grupo; γ_1: asimetría; γ_2: curtosis; + -: emparejamiento curtosis-tamaño grupo positivo y negativo

este criterio una prueba es robusta si las tasas empíricas de error Tipo I se encuentran comprendidas entre el intervalo 0,025 y 0,075. Se considera que el procedimiento es liberal cuando la tasa error Tipo I empírica es mayor que el límite superior y que el procedimiento es conservador cuando está por debajo del límite inferior. La tabla 2 muestra las tasas empíricas de error Tipo I estimadas para los efectos de medidas repetidas y su interacción con el factor intersujeto, según la combinación de las distintas condiciones estudiadas. Los valores en negrita corresponden a valores por encima del intervalo establecido por Bradley (1978).

Condiciones de normalidad

Para el *efecto de medidas repetidas (tiempo)* el procedimiento K-R es robusto ante violaciones de la esfericidad, tanto con diseños balanceados como no balanceados. Sin embargo, para el *efecto de interacción* el procedimiento se muestra liberal en todas las condiciones de tamaño muestral y esfericidad.

Condiciones de no normalidad

Para el *efecto de medidas repetidas (tiempo)* el procedimiento K-R es liberal ante violaciones de la esfericidad cuando la relación entre la curtosis de los grupos es 1:2. Asimismo ocurre cuando la relación entre la curtosis de los grupos es 1:3 y 1:4 y el emparejamiento curtosis-tamaño grupo es positivo. Es decir, cuando el grupo de mayor tamaño es el que presenta mayor grado de curtosis, el procedimiento K-R es liberal, independientemente de la esfericidad. Sin embargo, cuando el emparejamiento es nulo o negativo, este procedimiento es robusto ante violaciones de la esfericidad cuando la relación entre la curtosis de los grupos es 1:3 y 1:4.

Para el *efecto de interacción* el procedimiento K-R es liberal ante violaciones de la esfericidad cuando la relación entre la curtosis de los grupos es 1:2; 1:3 y 1:4, y el emparejamiento curtosis-tamaño grupo es nulo o positivo, mostrándose robusto ante violaciones de la esfericidad únicamente cuando el emparejamiento curtosis-tamaño grupo es negativo. Es decir, el procedimiento K-R es robusto cuando el grupo de menor tamaño (n= 5) es el que presenta mayor grado de curtosis.

Discusión

El objetivo del presente trabajo fue evaluar la robustez del MLM con la corrección de los grados de libertad por el método K-R con diseños *split-plot* de tamaños de muestra pequeños cuando los grupos tienen un grado leve de asimetría y distintos grados de curtosis, con y sin violación de la esfericidad.

Los resultados obtenidos muestran que cuando la distribución es normal en los tres grupos el efecto de medidas repetidas (tiempo) con el procedimiento K-R es robusto ante violaciones de la esfericidad, tanto con diseños balanceados como no balanceados. Estos resultados son consistentes con los de Arnau et al. (2009), quienes encontraron robustez ante violaciones de la esfericidad con datos normales. Sin embargo, para el efecto de interacción, el procedimiento se muestra liberal en todas las condiciones de tamaño muestral y esfericidad.

En investigaciones reales es probable que las variables no presenten la misma distribución, por lo que en este estudio se ha generado una distribución no normal distinta para cada grupo. Este tipo de situación ha sido explorada en estudios de simulación sobre el ANOVA intersujeto, hallándose que cuando los grupos tienen distinta asimetría y curtosis, el error Tipo I incrementa, mientras que cuando los grupos tienen la misma distribución e igual tamaño, la F es más robusta (Keselman et al., 1996). El presente estudio muestra el mismo patrón de resultados cuando se utiliza el MLM con procedimiento K-R, ya que si bien es robusto para el efecto de medidas repetidas en diseños balanceados y cuando el emparejamiento curtosis-tamaño de grupo es negativo, es liberal

Tabla 2

Tasas de error Tipo I estimadas para los efectos tiempo e interacción (valor nominal 0,05)

								No normal γ_1= 0,8					
								Relación entre la curtosis de los grupos					
N	**n_1**	**n_2**	**n_3**	**Δn_j**	**Normal γ_1= 0**		**Emparejamiento curtosis-tamaño grupo**	**1:2**		**1:3**		**1:4**	
Valores de curtosis					g1: γ_2= 0 g2: γ_2= 0 g3: γ_2= 0			g1: γ_2= 0,4 g2: γ_2= 0,8 g3: γ_2= 1,6		g1: γ_2= 0,8 g2: γ_2= 2,4 g3: γ_2= 7,2		g1: γ_2= 0,8 g2: γ_2= 3,2 g3: γ_2= 12	
Esfericidad					0,57	0,75		0,57	0,75	0,57	0,75	0,57	0,75
Efecto de medidas repetidas (tiempo)													
30	10	10	10	0,00	0,074	0,070	Nulo	**0,079**	**0,080**	0,073	0,071	0,071	0,074
30	5	10	15	0,41	0,071	0,071	**+**	**0,082**	**0,076**	**0,083**	**0,084**	**0,092**	**0,089**
30	15	10	5	0,41	0,074	0,072	**-**	**0,076**	**0,075**	0,071	0,065	0,062	0,060
Efecto de interacción													
30	10	10	10	0,00	**0,079**	**0,076**	Nulo	**0,076**	**0,076**	**0,077**	**0,076**	**0,078**	**0,075**
30	5	10	15	0,41	**0,077**	**0,078**	**+**	**0,080**	**0,076**	**0,083**	**0,090**	**0,093**	**0,095**
30	15	10	5	0,41	**0,082**	**0,079**	**-**	**0,077**	**0,076**	0,070	0,070	0,067	0,063

* Nota: N: tamaño muestral total; n_j : tamaño muestral de cada grupo; Δn_j: coeficiente de variación; g_j: grupo; γ_1: asimetría; γ_2: curtosis; + -: emparejamiento curtosis-tamaño grupo positivo y negativo. En **negrita**: liberal

cuando el emparejamiento es positivo. Asimismo, es robusto para la interacción con el factor intersujeto cuando el emparejamiento curtosis-tamaño de grupo es negativo y es liberal, en diseños balanceados como no balanceados, cuando el emparejamiento es positivo o nulo.

En general, los resultados apuntan a que el procedimiento K-R no es robusto ante violaciones de la normalidad con diferentes distribuciones en los grupos con un diseño *split-plot*, sobre todo para el efecto de la interacción. Estos resultados no son congruentes con otros autores que encontraron una tendencia a la robustez ante violaciones de la normalidad con esfericidad asumida (Kowalchuk et al., 2004; Livacic-Rojas et al., 2010; Vallejo et al., 2004). No obstante, van en la línea de los encontrados por Vallejo y Ato (2006), quienes estudiaron el efecto conjunto de normalidad y esfericidad, encontrando que K-R es liberal para el efecto de interacción. En cualquier caso es difícil comparar los resultados puesto que en los estudios citados todos los grupos seguían la misma distribución, a diferencia del presente que incluye diferente curtosis en cada uno de ellos.

En conclusión, los resultados sugieren que la relación entre los valores de curtosis de los grupos y el emparejamiento de la curtosis con el tamaño de grupo son variables importantes que hay que tener en cuenta cuando se aplique el enfoque del MLM con la corrección de los grados de libertad por el procedimiento K-R. Asimismo, futuros estudios de simulación Monte Carlo deberían profundizar en la influencia de estas variables en otras condiciones de tamaño muestral, valores de asimetría y curtosis, así como heterogeneidad de covarianzas entre grupos.

Agradecimientos

Este trabajo ha sido financiado mediante el proyecto de investigación PSI2009-11136 concedido por el Ministerio de Ciencia e Innovación.

Referencias

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. En Second International Symposium on Information Theory. B.N. Petrov y F. Csaki (Eds.), Akademiai Kiado, Budapest, 267-281.

Alnosaier, W.S. (2007). Kenward-Roger approximate *F* test for fixed effects in mixed linear models. Tesis doctoral publicada en Proquest Information and Learning Company.

Arnau, J., Bono, R., y Vallejo, G. (2009). Analyzing small samples of repeated measures data with the mixed-model adjusted F test. *Communications in Statistics. Simulation and Computations, 38*, 1083-1103.

Bauer, D.J., y Curran, P.J. (2003). Distributional assumptions of growth mixture models: Implications for over extraction of latent trajectory classes. *Psychological Methods, 8,* 338-363.

Berkovits, I., Hancock, G.R., y Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement, 60*(6), 877-892.

Blanca, M.J., Arnau, J., Bono, R., López-Montiel, D., y Bendayan, R. (en prensa). Skewness and kurtosis in real data samples. Methodology. *European Journal of Research Methods for the Behavioral and Social Sciences*.

Bono, R., Arnau, J., y Balluerka, N. (2007). Using linear mixed models in longitudinal studies: Application of SAS PROC MIXED. *Revista Electrónica de Metodología Aplicada, 12*(2), 15-31.

Bono, R., Arnau, J., y Vallejo, G. (2010). Modelización de diseños split-plot y estructuras de covarianza no estacionarias: un estudio de simulación. *Escritos de Psicología, 3*(3), 1-7.

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

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.

Cnaan, A., Laird, N.M., y Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*, 2349-2380.

Fernández, P., Livacic-Rojas, P., y Vallejo, G. (2007). Cómo elegir la mejor prueba estadística para analizar un diseño de medidas repetidas. *International Journal of Clinical and Health Psychology, 7*(1), 153-175.

Fernández, P., Vallejo, G., Livacic-Rojas, P., y Tuero, E. (2010). Características y análisis de los diseños de medidas repetidas en la investigación experimental en España en los últimos diez años. *Actas del XI Congreso de Metodología de las Ciencias del Comportamiento*. Málaga.

Fleishman, A. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521-531.

Glass, G.V., Peckham, P.D., y Sanders, J.R. (1972). Consequences of failure to meets assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research, 42,* 237-288.

Harwell, M.R., Rubinstein, E.N., Hayes, W.S., y Olds, C.C. (1992). Summarizing Monte Carlo results in methodological research: The one and two factors fixed effect ANOVA case. *Journal of Educational Statistics, 17,* 315-339.

Hopkins, K.D., y Weeks, D.L. (1990). Test for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement, 50*, 717-729.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika, 43*(2), 161-175.

Huynh, H., y Feldt, L.S. (1970). Conditions under which mean square ratios in repeated measurement designs have exact F-Distribution. *Journal of the American Statistical Association, 65*, 1582-1589.

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

Keselman, H.J., Algina, J., y Kowalchuk, R.K. (2001). The analysis of the repeated measures design: A review. *British Journal of Mathematical and Statistical Psychology, 54,* 1-20.

Keselman, H.J., Algina, J., Kowalchuk, R.K., y Wolfinger, R.D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics - Computation and Simulation, 27*, 591-604.

Keselman, H.J., Algina, J., Kowalchuk, R.K., y Wolfinger, R.D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 52*, 63-78.

Kowalchuk, R.K., Keselman, H.J., Algina, J., y Wolfinger, R.D. (2004). The analysis of repeated measurements with mixed-model adjusted *F* tests. *Educational and Psychological Measurement, 64*(2), 224-242.

Keselman, H.J., Huberty, C.J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donahue, B., Kowalchuk, R.K., Lowman, L.L., Petoskey, M.D., Keselman, J.C., y Levin, J.R. (1998). Statistical practices of education researchers: An analysis of the ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350-386.

Keselman, J.C., Lix, L.M., y Keselman, H.J. (1996). The analysis of repeated measurements: A quantitative research synthesis. *British Journal of Mathematical and Statistical Psychology, 49*, 275-298.

Laird, N.M., y Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics, 38,* 963-974.

Littell, R.C., Milliken, G.A., Stroup, W.W., y Wolfinger, R.D. (1996). *SAS System for mixed models*. Caary, NC: SAS Institute Inc.

Livacic-Rojas, P., Vallejo, G., y Fernández, P. (2006). Procedimientos estadísticos alternativos para evaluar la robustez mediante diseños de medidas repetidas. *Revista Latinoamericana de Psicología, 38*(3), 579-598.

Livacic-Rojas, P., Vallejo, G., y Vallejo, P. (2010). Analysis of Type I error rate of univariable and multivariate procedures in repeated measures designs. *Communications in Statistics - Simulation and Computation, 39*, 624-640.

Lix, L., Keselman, J.C., y Keselman, H.J. (1996). Consequences of assumptions violations revisited: A quantitative review of alternatives to the one-way analysis of variance *F* test. *Review of Educational Research, 66*, 579-620.

Lix, L., Algina, J., y Keselman, H.J. (2003). Analyzing multivariate repeated measures designs: A comparison of two approximate degrees of freedom procedures. *Multivariate Behavioral Research, 38*, 403-431.

Mendoza, J.L., Toothtaker, L.E., y Crain, B.R. (1976). Necessary and sufficient conditions for *F* ratios in the $L \times J \times K$ factorial design with two repeated factors. *Journal of the American Statistical Associations, 71,* 992-993.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Rasch, D., y Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science, 46*, 175-208.

Rogan, J.C., Keselman, H.J., y Mendoza, J.L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 32*, 269-286.

Robey, R.R., y Barcikowski, R.S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology, 45,* 283-288.

Rouanet, H., y Lepine, D. (1970). Comparison between treatments in a repeated measures design: ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology, 23*, 147-163.

Sawilowsky, S.S., y Blair, R.C. (1992). A more realistic look at the robustness and Type II error properties of the *t* test to departures from normality. *Psychological Bulletin, 111,* 356-360.

Schaalje, G.B., McBride, J.B., y Fellingham, G.W. (2002). Adequacy of approximations to distributions of test statistics in complex missed linear models. *Journal of Agricultural, Biological, and Environmental Statistics, 7,* 512-524.

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

Satterthwaite, F. (1941). Synthesis of variance. *Psychometrika, 6,* 309-316.

Satterthwaite, F. (1946). Approximate distribution of estimates of variance components. *Biometrics, 2,* 110-114.

Vale, C.D., y Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*(3), 465-471.

Vallejo, G., Arnau, J., y Ato, M. (2007). Comparative robustness of recent methods for analyzing multivariate repeated measures designs. *Educational and Psychological Measurement, 67*(3), 410-432.

Vallejo, G., Ato, M., y Valdés, T. (2008). Consequences of misspecifying the error covariance structure in linear mixed models for longitudinal data. *Methodology. European Journal of Research for the Behavioral and Social Sciences, 4*, 10-21.

Vallejo, G., Fernández, P., Herrero F.J., y Conejo, N.M. (2004). Alternative procedures for testing fixed effects in repeated measures designs when assumptions are violated. *Psicothema, 16*(3), 498-508.

Vallejo, G., Fernández, P., Tuero, E., y Livacic-Rojas, P. (2010). Análisis de un diseño de medidas repetidas usando remuestreo *bootstrap* y permutación aleatoria. *Actas del XI Congreso de Metodología de las Ciencias del Comportamiento*. Málaga.

Wright, S.P., y Wolfinger, R.D. (1996). Repeated measures analysis using mixed models: Some simulation results. *Conference on Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions*. Nantucket, MA.

Wu, P.C. (2007). Modern one-way ANOVA *F* methods: Trimmed means, one step *M*-estimators and bootstrap methods. *Journal of Quantitative Research, 1*, 155-173.