Psicothema

Psicothema 2014, Vol. 26, No. 2, 279-285
doi: 10.7334/psicothema2013.174

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

The effect of skewness and kurtosis on the Kenward-Roger approximation when group distributions differ

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

Abstract

Background: This study examined the independent effect of skewness and kurtosis on the robustness of the linear mixed model (LMM), with the Kenward-Roger (KR) procedure, when group distributions are different, sample sizes are small, and sphericity cannot be assumed. **Methods:** A Monte Carlo simulation study considering a split-plot design involving three groups and four repeated measures was performed. **Results:** The results showed that when group distributions are different, the effect of skewness on KR robustness is greater than that of kurtosis for the corresponding values. Furthermore, the pairings of skewness and kurtosis with group size were found to be relevant variables when applying this procedure. **Conclusions:** With sample sizes of 45 and 60, KR is a suitable option for analyzing data when the distributions are: (a) mesokurtic and not highly or extremely skewed, and (b) symmetric with different degrees of kurtosis. With total sample sizes of 30, it is adequate when group sizes are equal and the distributions are: (a) mesokurtic and slightly or moderately skewed, and sphericity is assumed; and (b) symmetric with a moderate or high/extreme violation of kurtosis. Alternative analyses should be considered when the distributions are highly or extremely skewed and samples sizes are small.

Keywords: Linear mixed model, Kenward-Roger, small samples, skewness, sphericity.

Resumen

El efecto de la violación de simetría y curtosis en la aproximación Kenward-Roger cuando las distribuciones de los grupos difieren. **Antecedentes:** este estudio examina el efecto independiente de la violación de la simetría y de la curtosis en la robustez del modelo lineal mixto, con la corrección Kenward-Roger de los grados de libertad, cuando las distribuciones de los grupos difieren, los tamaños muestrales son pequeños y se viola el supuesto de esfericidad. **Método:** se realizó un estudio de simulación Monte Carlo con un diseño de tres grupos y cuatro medidas repetidas. **Resultados:** cuando las distribuciones de los grupos son diferentes, el efecto de la violación de la simetría es mayor que el de la curtosis. Además, el emparejamiento de asimetría y curtosis con el tamaño de grupo se constatan como variables a considerar cuando se utiliza este procedimiento. **Conclusiones:** KR constituye una buena opción cuando el diseño es equilibrado, y (a) los tamaños muestrales totales son iguales a 45 o 60, y las distribuciones son mesocúrticas y no extremadamente asimétricas, o bien, simétricas con distintos grados de violación de curtosis; o (b) con tamaños muestrales de 30 y distribuciones mesocúrticas y leve/moderadamente asimétricas, o bien, simétricas con una violación moderada/extrema de la curtosis. Con estos tamaños muestrales y distribuciones severa o extremadamente asimétricas no es recomendable utilizar KR.

Palabras clave: Modelo Lineal Mixto, Kenward-Roger, muestras pequeñas, asimetría, esfericidad.

Exploring how a variable changes over time in different groups is a frequent aim in educational and psychological research (e.g., exploring differences in cognitive development between boys and girls), and in these cases, the study is usually based on a split-plot design. Indeed, split-plot designs, namely those in which there are one or more grouping factors on which individuals are repeatedly measured on two or more occasions, are the most commonly used in educational and psychological research (Bono, Arnau, & Balluerka, 2007; Keselman et al., 1998). The traditional analytic method for these designs is the classical linear model, which is valid under certain distributional assumptions (Huynh & Feldt, 1970; Rouanet & Lepine, 1970): normality, independence of the observations, variance homogeneity, and sphericity. However, the data from these studies are often not normally distributed (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013; Micceri, 1989), sphericity cannot always be assumed (Huynh, 1978; Jaccard & Ackerman, 1985), and, moreover, samples are often small (Fernández, Vallejo, Livacic-Rojas, & Tuero, 2010; Keselman et al., 1998). In addition to this, Monte Carlo simulation studies have demonstrated that when these assumptions are not satisfied, this approach is mainly not robust. For example, the analysis of variance (ANOVA) F-test tends to be liberal (Berkovits, Hancock, & Nevitt, 2000; Box, 1954; Collier, Baker, Mandeville, & Hayes, 1967; Imhof, 1962; Keselman, Lix, & Keselman, 1996; Keselman & Rogan, 1980; Rasmussen, 1989; Rogan, Keselman, & Mendoza, 1979) when sphericity assumption is not satisfied; and when multisample sphericity is not satisfied, F-tests may be liberal or

Received: May 29, 2013 • Accepted: January 20, 2014
Corresponding author: Rebecca Bendayan
Facultad de Psicología
Universidad de Málaga
29071 Málaga (Spain)
e-mail: bendayan@uma.es

conservative, depending on the type of pairing of the covariance matrices and group sizes (Keselman et al., 1996).

The linear mixed model (LMM; Cnaan, Laird, & Slasor, 1997; Laird & Ware, 1982; Littell, Milliken, Stroup, & Wolfinger, 1996) is one of the most suitable approaches for analyzing data from split-plot designs when the assumptions of ANOVA with within-subject and between-subject factors are not met. The LMM approach uses statistics that have good large-sample properties, but this approach is not appropriate when samples are small (Wright & Wolfinger, 1996). Small sample properties can be improved by procedures that adjust the degrees of freedom, for example, the method developed by Kenward and Roger (KR; 1997).

Over the last decade, several simulation studies have explored the robustness of the LMM with the KR procedure when the assumptions of this model are not met. In this context, robustness is usually assessed by applying Bradley's (1978) liberal criterion, according to which a test is robust when the empirical Type I error rate is between .025 and .075 for $\alpha = .05$. Monte Carlo simulation studies have found that, for the repeated measures effect, the KR procedure is robust to variance heterogeneity with assumed sphericity and different violations of normality (Kowalchuk, Keselman, Algina, & Wolfinger, 2004; Livacic-Rojas, Vallejo, & Fernández, 2006, 2010; Vallejo, Fernández, Herrero, & Conejo, 2004). With respect to the interaction effect, the results are inconsistent. On the one hand, with known non-normal distributions and sphericity assumed, Kowalchuk et al. (2004) reported that the procedure was robust when the distribution was log-normal, whereas Vallejo et al. (2004) found it to be conservative with chi-square distributions with three degrees of freedom. On the other hand, with unknown non-normal distributions, KR has been found to be robust (Livacic-Rojas et al., 2010), conservative (Livacic-Rojas et al., 2006, 2010) or liberal (Vallejo & Ato, 2006).

With regard to the effect of skewness and kurtosis, previous research has highlighted that the two phenomena have a different effect on the robustness of several statistical tests (Lei & Lomax, 2005). Some studies have found that the effect of kurtosis is greater than that of skewness (Harwell, Rubinstein, Hayes, & Olds, 1992; Hopkins & Weeks, 1990), whereas other studies have reported the opposite (Chaffin & Rhiel, 1993; Scheffé, 1959). Recent studies have also pointed out the differential effect of skewness and kurtosis on KR robustness (Arnau, Bono, Blanca, & Bendayan, 2012; Arnau, Bendayan, Blanca, & Bono, 2013). Arnau, Bono et al. (2012) explored KR robustness with log-normal, exponential, and double exponential distributions when the assumptions of sphericity and variance homogeneity were not jointly met. They found that, for both the repeated measures and interaction effects, KR was less robust when the distribution was log-normal, in which case it was nearly always liberal when the sphericity assumption was not met. Their findings also suggested that skewness could have a greater effect than kurtosis on KR robustness. Arnau et al. (2013) explored the independent effect of skewness and kurtosis on KR robustness and found, in general, that the effect of skewness was greater than the corresponding effect of kurtosis. More specifically, their findings highlighted that KR was a good option for analyzing total sample sizes of 45 or larger when distributions are normal or slightly or moderately skewed, whereas the procedure was mainly liberal with sample sizes of 30.

All the cited studies have considered that the distributions of the groups were exactly the same, although this may not be the case with real data (Harwell et al., 1992; Olson, 1974; Tiku, 1964). To date, only one study on KR robustness has simulated data with different distributions in each group (Arnau, Bendayan, Blanca, & Bono, 2012). These authors examined KR robustness with sample sizes of 30 and moderately skewed distributions, finding that KR was mainly liberal for both the repeated measures and interaction effects. Whereas this study examined the effect of skewness and kurtosis jointly on KR robustness, the independent effect of skewness and kurtosis when group distributions differ has yet to be explored. Consequently, the purpose of the present study was to examine the independent effect of skewness and kurtosis on KR robustness when group distributions are different, sample sizes are small, and sphericity cannot be assumed.

Linear mixed model and the Kenward-Roger procedure

The LMM allows researchers to include random factors and to model the covariance structure of their data prior to testing the treatment effects. In general, the LMM described by Laird and Ware (1982) can be written as in (1):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \boldsymbol{e} \tag{1}$$

where $\mathbf{y}$ is the observations vector, $\mathbf{X}$ is the matrix for the fixed effects model, $\boldsymbol{\beta}$ is the vector of the fixed effects parameters, $\mathbf{Z}$ is the matrix for the random effects model, $\mathbf{u}$ is the vector of the random effects parameters, and $\boldsymbol{e}$ is the vector of random errors.

The distribution assumptions of this model are that $\mathbf{u}$ and $\boldsymbol{e}$ are independent random vectors distributed as $\mathbf{u} \sim N(\mathbf{0},\mathbf{G})$ and $\boldsymbol{e} \sim N(\mathbf{0},\mathbf{R})$, respectively, where $\mathbf{G}$ is a matrix of unknown covariance parameters for the between-subjects random effects and $\mathbf{R}$ is a covariance matrix for the within-subjects errors. As $\mathbf{u}$ and $\boldsymbol{e}$ are independent vectors, their covariance is equal to 0 and the covariance matrix of $\mathbf{y}$ is $\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$.

The matrices $\mathbf{G}$ and $\mathbf{R}$ are usually unknown and, consequently, an estimate of $\mathbf{V}$ must be used. The residual maximum likelihood estimation is often used to estimate $\mathbf{V}$ (Zimmerman & Núñez-Antón, 2001), as in (2):

$$\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\beta}}) = (\mathbf{X'V^{-1}X})^{-} \tag{2}$$

Once the covariance matrix has been selected and its parameters estimated, $\boldsymbol{\beta}$ is estimated through the generalized least squares estimator, as in (3):

$$\hat{\boldsymbol{\beta}} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y \tag{3}$$

However, the true variance of $\hat{\boldsymbol{\beta}}$ is not $(\mathbf{X'V^{-1}X})^{-}$ because $\hat{\boldsymbol{\beta}}$ contains variation due to $\hat{\mathbf{V}}$, so it is not always a good estimate of $\mathbf{V}$ (Littell, 2002). As Vallejo et al. (2004) highlighted, this means that the likelihood-based inference should be interpreted with caution when the sample size is not large enough. As mentioned, small sample properties can be improved by procedures that adjust the degrees of freedom, for example, the KR procedure. This procedure provides an adjusted estimator of the covariance matrix of $\boldsymbol{\beta}$ that reduces the bias for small sample inference when the asymptotic covariance matrix underestimates $\hat{\mathbf{V}}$.

Here, the LMM uses Wald-type statistics that can be defined as in (4):

$$W = (C\beta)'(C(X'V^{-1}X)^{-1}C')^{-1}(C\hat{\beta}) \quad (4)$$

where **C** is a contrast matrix with range q, and the Wald F for the hypothesis H_0: $C\beta = 0$ is $F = \mathbf{W}/q$.

If we calculate a scale factor δ and an approximate value for the degrees of freedom ν, then the F statistic for the KR method is given by (5):

$$F^* = \delta F_{KR} = \frac{\delta}{q}(C\hat{\beta})'(C(X'V^{-1}X)^{-1}C')^{-1}(C\hat{\beta})' \quad (5)$$

The moments of F^* are generated and matched to the moments of the distribution F so as to solve δ and ν. Under the null hypothesis, it is assumed that F^* is approximately distributed in the same way as F, with q degrees of freedom in the numerator and ν degrees of freedom in the denominator. This means that two values have to be calculated from the data: the degrees of freedom in the denominator ν and a scale factor δ following (6), (7) and (8). Thus,

$$\nu = 4 + \frac{q+2}{qy-1} \quad (6)$$

where,

$$y = \frac{V[F_{KR}]}{2E[F_{KR}]^2} \quad (7)$$

and,

$$\delta = \frac{\nu}{E[F_{KR}](\nu-2)} \quad (8)$$

Method

A Monte Carlo simulation study was designed to examine the independent effect of skewness and kurtosis on KR robustness when group distributions are different, sample sizes are small, and sphericity cannot be assumed. The variables manipulated were as follows: (a) total sample size and group size; (b) distributional shape of the response variable; (c) sphericity; (d) pairing of skewness with group size; and (e) pairing of kurtosis with group size.

(a) Total sample size and group size. Total sample sizes of N = 30, 45, and 60 were considered because these are the most frequently used in behavioral and educational research (Keselman et al., 1998; Livacic-Rojas et al., 2006; Fernández et al., 2010). For each value of N, both equal and unequal group sizes were considered. Unequal group sizes, in which the number of individuals decreases, were considered because unbalanced data due to experimental mortality is very common in longitudinal studies (Keselman et al., 1998). Specifically, with unequal group size, the coefficient of sample size variation, Δn_j, was. 33, while the group sizes were as follows: 14, 10, 6 (N = 30); 21, 15, 9 (N = 45); and 28, 20, 12 (N = 60). When the group sizes were equal, Δn_j = 0, the group sizes were 10, 10, 10 (N = 30); 15, 15, 15 (N = 45); and 20, 20, 20 (N = 60).

(b) Distributional shape of the response variable. In order to explore the differential effect of skewness (γ_1) and kurtosis (γ_2) on KR robustness, several distributional shapes of the response variable were considered. Different values of the γ_1 and γ_2 coefficients were chosen based on the results of a recent study that assessed the distributional shape of real data by examining the values of γ_1 and γ_2 in small samples of educational and behavioral research data (Blanca et al., 2013). This study revealed that γ_1 usually ranges between -2.49 and 2.33, while γ_2 usually ranges between -1.92 and 7.41. The values of the γ_1 and γ_2 coefficients were chosen according to the cut-off points for the typical degree of contamination found in this type of data, as proposed by Blanca et al. (2013). The studied conditions were labeled as slight, moderate, and high/extreme contamination when at least two of the distributions had values representing slight, moderate, and high/extreme contamination according to the criteria of Blanca et al. (2013). The values used to explore the effect of skewness and kurtosis are shown in Tables 1 and 2, respectively.

(c) Sphericity. In order to analyze KR robustness to violations of normality and sphericity jointly, two indices of sphericity were used: a value of ε = .75 was taken as a good approximation to sphericity, while ε = .57 was used to represent non-sphericity.

(d) Pairing of kurtosis with group size. The type of pairing between kurtosis and group size has been shown to be a relevant variable to consider when using this procedure (Arnau, Bendayan, et al., 2012). The type of pairing between kurtosis and group size was defined as one of the following: null, positive, or negative. Pairing was null when group sizes were equal. Pairing was positive when the largest group was associated with the largest value of the γ_2 coefficient and the smallest group was associated with the smallest value of the γ_2 coefficient. Pairing was negative when the largest group was associated with the smallest value of the γ_2 coefficient and the smallest group was associated with the largest value of the γ_2 coefficient.

(e) Pairing of skewness with group size. The type of pairing between skewness and group size was defined as one of the following: null (when group sizes were equal), positive (when the largest group was associated with the largest value of the γ_1 coefficient and the smallest group was associated with the

Table 1
Values of γ_1 used to explore the effect of skewness ($\gamma_2 = 0$)

Slight	Moderate	High/Extreme
g1: $\gamma_1 = 0.2$	g1: $\gamma_1 = 0.4$	g1: $\gamma_1 = 0.8$
g2: $\gamma_1 = 0.3$	g2: $\gamma_1 = 0.9$	g2: $\gamma_1 = 1.8$
g3: $\gamma_1 = 0.4$	g3: $\gamma_1 = 1.2$	g3: $\gamma_1 = 2.4$

Note: g_j: group; γ_1:skewness; γ_2: kurtosis

Table 2
Values of γ_2 used to explore the effect of kurtosis ($\gamma_1 = 0$)

Slight	Moderate	High/Extreme
g1: $\gamma_2 = 0.4$	g1: $\gamma_2 = 0.8$	g1: $\gamma_2 = 0.8$
g2: $\gamma_2 = 0.8$	g2: $\gamma_2 = 2.4$	g2: $\gamma_2 = 3.2$
g3: $\gamma_2 = 1.6$	g3: $\gamma_2 = 7.2$	g3: $\gamma_2 = 12$

Note: g_j: group; γ_1:skewness; γ_2: kurtosis smallest value of the γ_1 coefficient), or negative (when the largest group was associated with the smallest value of the γ_1 coefficient and the smallest group was associated with the largest value of the γ_1 coefficient).

Data were generated using a series of macros created ad hoc in SAS 9.2 (SAS Institute, 2008), following the same procedure used in Arnau, Bono, et al. (2012). Moreover, all data were generated assuming variance homogeneity and using the unstructured (UN) covariance structure, as several studies recommend this approach when the number of observations is moderate or sample sizes are small (Chen & Wei, 2003; Kowalchuk et al., 2004). Recent research has also indicated that some of the standard criteria used to select the covariance structure in linear mixed models through PROC MIXED (e.g., the Akaike Information Criterion) fail to select the true covariance structure when sample sizes are small (Vallejo, Fernández, Livacic-Rojas, & Tuero-Herrero, 2011). Here, ten thousand replications were performed for each combination at a significance level of .05 (Bendayan, Blanca, Arnau, & Bono, 2014; Robey & Barcikowsky, 1992).

Results

Robustness was determined according to Bradley's criterion, whereby the effect estimate is robust when the empirical Type I error rate is between .025 and .075 for $\alpha = .05$. A test is considered to be liberal when the empirical Type I error rate is above the upper limit, and conservative when it is below the lower limit.

General results

In general, the results show that the effect of skewness on KR robustness seems to be greater than the corresponding effect of kurtosis. For the repeated measures effect, the overall percentage of KR robustness when the data are skewed (59.25%) is lower than the overall percentage of KR robustness when the data have different degrees of kurtosis violation (94.44%). For the interaction effect, the overall percentage of KR robustness when the data are skewed (51.85%) is also lower than the overall percentage of KR robustness when the data have different degrees of kurtosis violation (79.63%).

With regards to the effect of the violation of the sphericity, for the repeated measures effects, the overall percentage of KR robustness when sphericity cannot be assumed (70.37%) is lower than the overall percentage of KR robustness when sphericity can be assumed (83.33%). For the interaction effect, the overall percentage of KR robustness when sphericity cannot be assumed (64.81%) is slightly lower than the overall percentage of KR robustness when sphericity can be assumed (66.66%). As can be seen in Tables 3 and 4, the slight effect of sphericity on KR robustness appears to be associated with the smaller total sample size (i.e., $N = 30$). In fact, the results highlight the effect of the total sample size on KR robustness. For the repeated measures effects, the overall percentage of robustness of KR when the total sample size is equal to 30 (61.11%) is lower than the overall percentage of robustness of KR when the total sample size is equal to 45 or 60 (84.72%). For the interaction effect, the overall percentage of robustness of KR when the total sample size is equal to 30 (22.22%) is lower than the overall percentage of robustness of KR when the total sample size is equal to 45 or 60 (87.5%).

Specific results

Skewness effect. Table 3 shows that, for the repeated measures effect, KR is mainly robust with total sample sizes of 45 and 60 when the distributions are slightly or moderately skewed, independently of whether or not sphericity can be assumed. With a total sample size of 30 and slightly skewed distributions, KR is robust when sphericity is assumed, but tends to be liberal when sphericity cannot be assumed, especially when the group sizes are not equal. With a total sample size of 30 and moderately skewed distributions, KR is robust when sphericity is assumed, except when the pairing of skewness with group size is negative (the largest group is associated with the smallest value of the γ_1 coefficient and the smallest group is associated with the largest value of the γ_1 coefficient). A similar pattern is found when sphericity is not assumed. A great decrease in KR robustness is found when the distributions are highly or extremely skewed, with the procedure being liberal for all the total sample sizes considered and, in general, independently of sphericity. Under these conditions, the procedure is only robust with total sample sizes of 45 and 60 when sphericity is assumed and the pairing of skewness with group size is positive (the largest group was associated with the largest value of the γ_1 coefficient and the smallest group was associated with the smallest value of the γ_1 coefficient).

For the interaction effect, KR is robust with total sample sizes of 45 and 60 when the distributions are slightly or moderately skewed, regardless of whether or not sphericity can be assumed. However, with a total sample size of 30, KR is liberal when the distributions are slightly or moderately skewed, independently of whether or not sphericity can be assumed or not. As in the case of the repeated measures effect, a great decrease in KR robustness is found when the distributions are highly or extremely skewed, with the procedure being liberal for all the total sample sizes considered and, in general, independently of sphericity, except when the pairing of skewness with group size is positive with a total sample size of 60.

Kurtosis effect. Table 4 shows that, for the repeated measures effect, KR is robust in nearly all the studied conditions. The procedure is liberal only when the total sample size is 30, the pairing of skewness with group size is positive, and the distributions have either a moderate or a high/extreme degree of kurtosis. For the interaction effect, KR is mainly robust in all the studied conditions with total sample sizes of 45 and 60. However, it tends to be liberal with a total sample size of 30.

Discussion

The purpose of this study was to examine the independent effect of skewness and kurtosis on KR robustness when group distributions are different, samples sizes are small, and sphericity cannot be assumed.

In general, the effect of skewness on KR robustness when group distributions differ appears to be greater than that of kurtosis for the corresponding values. This is consistent with the results of previous studies in which all the group distributions were the same (Arnau, Bono, et al., 2012; Arnau et al., 2013). Our results also suggest that the effect of skewness on KR may depend on whether or not sphericity can be assumed when the total sample size is equal to 30, the procedure being less robust when this assumption was not satisfied. As the present study is the first to explore the independent effect of skewness and kurtosis when group distributions differ, its results cannot be directly compared to other studies with different conditions. Nevertheless, our findings can be regarded as partially coincident with those of other studies that considered the violation of sphericity and normality jointly, and which examined either equal group distributions (Arnau, Bono, et al., 2012; Arnau et al., 2013; Vallejo & Ato, 2006) or different ones (Arnau, Bendayan, et al., 2012). Further studies are needed to explore the differential effect of sphericity and normality with other total sample sizes than were examined here.

With regard to the effect of skewness, and for both the repeated measures and interaction effects, KR was robust with total sample sizes of 45 and 60 when the distributions were slightly or moderately skewed, independently of whether or not the sphericity assumption was violated. However, with total sample sizes of 30, KR was often liberal when the distributions were slightly or moderately skewed. When the distributions were highly or extremely skewed, KR was mainly liberal for both the repeated measures and interaction effects, independently of total sample size or violation of the sphericity assumption. The pairing of skewness with group size only appeared to be a relevant variable when total sample size was 60, with the KR procedure being more robust when pairing was positive and sphericity was assumed.

With regard to the effect of kurtosis, and for both the repeated measures and interaction effects, KR was robust with total sample sizes of 45 and 60, independently of the degree of kurtosis, the assumption of sphericity, or whether or not the groups were balanced. However, with a total sample size of 30, KR tends to be liberal for both effects.

Although the shape of the distribution, the sphericity assumption, and the total sample size are relevant variables to consider when applying the LMM with the KR procedure to educational and psychological research data, the results suggest, in line with previous studies (Arnau, Bendayan, et al., 2012; 2013), that it is also necessary to take into account the pairing of skewness and kurtosis with group size. It would be interesting in future research to explore the effect of both these types of pairings.

Furthermore, the percentage of KR robustness is lower for the interaction effect than for the repeated measures effect, which is especially relevant in the case of repeated measures designs (Livacic et al., 2010). It should also be noted that although the KR procedure is usually proposed as a valid alternative for analyzing data from split-plot designs with small samples, the present results suggest that it may not be the best option for analyzing this type of data.

Table 3

Empirical type I error rates for the repeated measures and interaction effects (nominal value 0.05) with respect to skewed data

						Degree of skewness contamination					
					Pairing of skewness with group size	**Slight**		**Moderate**		**High/Extreme**	
						g1: $\gamma_1 = 0.2$ g2: $\gamma_1 = 0.3$ g3: $\gamma_1 = 0.4$		**g1: $\gamma_1 = 0.4$ g2: $\gamma_1 = 0.9$ g3: $\gamma_1 = 1.2$**		**g1: $\gamma_1 = 0.8$ g2: $\gamma_1 = 1.8$ g3: $\gamma_1 = 2.4$**	
						ε = 0.57	**ε = 0.75**	**ε = 0.57**	**ε = 0.75**	**ε = 0.57**	**ε = 0.75**
N	n_1	n_2	n_3	Δn_j							
						Repeated measures effect					
30	10	10	10	0.00		0.074	0.070	**0.077**	0.072	**0.138**	**0.123**
45	15	15	15	0.00		0.070	0.064	0.063	0.065	**0.111**	**0.103**
60	20	20	20	0.00		0.067	0.059	0.059	0.061	**0.103**	**0.104**
30	6	10	14	0.33	+	**0.077**	0.072	0.074	0.070	**0.110**	**0.099**
45	9	15	21	0.33	+	0.066	0.068	0.062	0.062	**0.081**	0.071
60	12	20	28	0.33	+	0.064	0.063	0.054	0.059	**0.076**	0.057
30	14	10	6	0.33	–	**0.078**	0.074	**0.082**	**0.081**	**0.158**	**0.154**
45	21	15	9	0.33	–	0.070	0.065	**0.076**	0.070	**0.164**	**0.146**
60	28	20	12	0.33	–	0.061	0.061	0.067	0.070	**0.154**	**0.143**
						Interaction effect					
30	10	10	10	0.00		**0.087**	**0.080**	**0.081**	**0.079**	**0.105**	**0.102**
45	15	15	15	0.00		0.067	0.067	0.072	0.068	**0.089**	**0.093**
60	20	20	20	0.00		0.063	0.060	0.061	0.063	**0.086**	**0.081**
30	6	10	14	0.33	+	**0.079**	**0.079**	**0.075**	**0.077**	**0.086**	**0.081**
45	9	15	21	0.33	+	0.064	0.066	0.068	0.067	0.070	0.061
60	12	20	28	0.33	+	0.065	0.062	0.062	0.062	0.062	0.058
30	14	10	6	0.33	–	**0.080**	**0.079**	**0.079**	**0.080**	**0.119**	**0.125**
45	21	15	9	0.33	–	0.069	0.064	0.071	0.072	**0.123**	**0.125**
60	28	20	12	0.33	–	0.065	0.064	0.072	0.066	**0.124**	**0.123**

Note: *N*: total sample size; n_j:group sample size; Δn_j: coefficient of sample size variation; Δ: sphericity; g_j:group; γ_1: skewness; γ_2: kurtosis; + -: positive and null group size-skewness pairing. In bold: liberal

Table 4
Empirical type I error rates for the repeated measures and interaction effects (nominal value 0.05), using data with different kurtosis coefficients

						Degree of kurtosis contamination					
						Slight		**Moderate**		**High/Extreme**	
					Pairing of kurtosis with group size	**g1: $\gamma_2 = 0.4$ g2: $\gamma_2 = 0.8$ g3: $\gamma_2 = 1.6$**		**g1: $\gamma_2 = 0.8$ g2: $\gamma_2 = 2.4$ g3: $\gamma_2 = 7.2$**		**g1: $\gamma_2 = 0.8$ g2: $\gamma_2 = 3.2$ g3: $\gamma_2 = 12$**	
						$\varepsilon = 0.57$	**$\varepsilon = 0.75$**	**$\varepsilon = 0.57$**	**$\varepsilon = 0.75$**	**$\varepsilon = 0.57$**	**$\varepsilon = 0.75$**
N	**n_1**	**n_2**	**n_3**	**Δn_j**							
						Repeated measures effect					
30	10	10	10	0.00		0.068	0.069	0.067	0.069	0.063	0.067
45	15	15	15	0.00		0.061	0.062	0.064	0.062	0.063	0.063
60	20	20	20	0.00		0.060	0.060	0.055	0.058	0.061	0.053
30	6	10	14	0.33	+	0.072	0.073	**0.075**	0.071	**0.080**	**0.080**
45	9	15	21	0.33	+	0.069	0.068	0.070	0.067	0.074	0.072
60	12	20	28	0.33	+	0.060	0.059	0.064	0.060	0.067	0.065
30	14	10	6	0.33	–	0.071	0.072	0.063	0.064	0.051	0.058
45	21	15	9	0.33	–	0.065	0.066	0.060	0.056	0.055	0.058
60	28	20	12	0.33	–	0.058	0.059	0.054	0.057	0.051	0.054
						Interaction effect					
30	10	10	10	0.00		**0.078**	**0.077**	0.074	0.074	0.074	0.073
45	15	15	15	0.00		0.068	0.069	0.064	0.064	0.065	0.064
60	20	20	20	0.00		0.056	0.051	0.055	0.054	0.058	0.056
30	6	10	14	0.33	+	**0.078**	**0.083**	**0.086**	**0.089**	**0.085**	**0.085**
45	9	15	21	0.33	+	0.068	0.069	0.068	0.066	**0.075**	0.071
60	12	20	28	0.33	+	0.061	0.061	0.065	0.060	0.064	0.063
30	14	10	6	0.33	–	**0.076**	**0.075**	0.074	0.070	0.067	0.064
45	21	15	9	0.33	–	0.066	0.064	0.063	0.063	0.057	0.058
60	28	20	12	0.33	–	0.058	0.058	0.049	0.060	0.050	0.050

Note: *N*: total sample size; n_j:group sample size; Δn_j: coefficient of sample size variation; Δ: sphericity; g_j:group; γ_1: skewness; γ_2: kurtosis; + -: positive and null group size-kurtosis pairing. In bold: liberal

In summary, the results enable several recommendations to be made regarding application of the LMM with the KR procedure to explore how a variable changes over time in different groups (e.g., exploring the differences in cognitive development between boys and girls). Real data often show different distributions in each group (Harwell et al., 1992; Olson, 1974; Tiku, 1964), and if this is the case, the KR procedure seems to be adequate when the distributions are not highly or extremely skewed and total sample sizes are 45 or larger. Specifically, with total sample sizes of 45 or larger, the LMM with the KR procedure is a suitable option for analyzing data when the distributions are: (a) mesokurtic and not highly or extremely skewed; and (b) symmetric, with different degrees of kurtosis. With total sample sizes of 30, the KR procedure is adequate when group sizes are equal and the distributions are: (a) slightly or moderately skewed and mesokurtic, with the sphericity assumption being met; and (b) symmetric with a moderate or high/extreme violation of kurtosis. Alternative analyses should be considered when the distributions are highly or extremely skewed and the total sample size is 60 or smaller, as well as when the total sample size is 30 and group sizes are unequal.

Acknowledgements

This study has been funded by Spain's Ministry of Economy and Competitiveness, project PSI2012-32662.

References

Arnau, J., Bendayan, R., Blanca, M.J., & Bono, R. (2012). Effect of the violation of normality and sphericity in the linear mixed model in split-plot designs. *Psicothema, 24*(3), 449-454.

Arnau, J., Bendayan, R., Blanca, M.J., & Bono, R. (2013). The effect of skewness and kurtosis on the robustness of mixed linear models. *Behavior Research Methods, 45*(3), 873-879.

Arnau, J., Bono, R., Blanca, M.J., & Bendayan, R. (2012). Using the linear mixed model to analyze non-normal data distributions in longitudinal designs. *Behavior Research Methods, 44*(4), 1224-1238.

Blanca, M.J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology. European Journal of Research Methods for the Behavioral and Social Sciences, 9*(2), 78-84.

Bendayan, R., Arnau, J., Blanca, M.J., & Bono, R. (2014). Comparison of the procedures of Fleishman and Ramberg et al., for generating non normal data in Simulation Studies. *Anales de Psicología, 30*(1), 364-371.

Berkovits, I., Hancock, G.R., & 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.

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

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems II. Effect of inequality of variance and of correlation of error in the two way classification. *Annals of Mathematical Statistics, 25*, 484-498.

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

Chafin, W.W., & Rhiel, S.G. (1993). The effect of skewness and kurtosis on the one-sample *T* test and the impact of knowledge of the population standard deviation. *Journal of Statistical Computation and Simulation, 46*(1-2), 79-90.

Chen, X., & 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., & Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*, 2349-2380.

Collier, R.O., Baker, F.B., Mandeville, G.K., & Hayes, T.F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. *Psychometrika, 32*, 339-353.

Fernández, P., Vallejo, G., Livacic-Rojas, P., & 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.

Harwell, M.R., Rubinstein, E.N., Hayes, W.S., & 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., & 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., & 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.

Imhof, J.P. (1962). Testing de hypothesis of o fixed main-effects in Scheffé mixed model. *The Annals of Mathematical Statistics, 33*, 1085-1095.

Jaccard, J., & Ackerman, L. (1985). Repeated measures analysis of means in clinical research. *Journal of Consulting and Clinical Psychology, 53,* 426-428.

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

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

Keselman, H.J., & Rogan, J. (1980). Repeated measures F test and psychophysiological research: Controlling the number of false positives. *Psychophysiology, 17,* 499-503.

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

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

Lei, M., & Lomax, R.G. (2005). The effect of varying degrees on nonnormality in structural equation modeling. *Structural Equation Modeling, 12*, 1-27.

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

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

Livacic-Rojas, P., Vallejo, G., & 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., & Fernández, 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.

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

Olson, C.L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69,* 348, 894-908.

Rasmussen, J.L. (1989). Parametric and non-parametric analysis of groups by trials design under variance-covariance inhomogeneity. *British Journal of Mathematical and Statistical Psychology, 42*, 91-102.

Robey, R.R., & 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.

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

Rouanet, H., & 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.

SAS Institute Inc. (2008). SAS/STAT 9.2 user's guide. Cary, NC:Author.

Scheffé, H. (1959). *The analyses of variance*. New York: Wiley.

Tiku, M.L. (1964). Approximating the general non-normal variance ratio sampling distributions. *Biometrika, 51*, 83-95.

Vallejo, G., & 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, F.J., & 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., Livacic-Rojas, P., & Tuero-Herrero., E. (2011). Selecting the best unbalanced repeated measures model. *Behavior Research Methods, 43,* 18-36.

Wright, S.P., & 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.

Zimmerman, D.L., & Núñez-Antón, V. (2001). Parametric modeling of growth curve data: An overview. *Test, 10,* 111-186.