Original Article

Skewness and Kurtosis in Real Data Samples

María J. Blanca,[1] Jaume Arnau,[2] Dolores López-Montiel,[1] Roser Bono,[2] and Rebecca Bendayan[1]

[1]Faculty of Psychology, Department of Psychobiology and Methodology, University of Malaga, Spain, [2]Faculty of Psychology, Department of Behavioural Sciences Methodology, University of Barcelona, Spain

Abstract. Parametric statistics are based on the assumption of normality. Recent findings suggest that Type I error and power can be adversely affected when data are non-normal. This paper aims to assess the distributional shape of real data by examining the values of the third and fourth central moments as a measurement of skewness and kurtosis in small samples. The analysis concerned 693 distributions with a sample size ranging from 10 to 30. Measures of cognitive ability and of other psychological variables were included. The results showed that skewness ranged between −2.49 and 2.33. The values of kurtosis ranged between −1.92 and 7.41. Considering skewness and kurtosis together the results indicated that only 5.5% of distributions were close to expected values under normality. Although extreme contamination does not seem to be very frequent, the findings are consistent with previous research suggesting that normality is not the rule with real data.

Keywords: skewness, kurtosis, shape distribution, normality

Monte Carlo computer simulation studies are used in a wide variety of conditions to identify the effect that the violation of assumptions, such as independence, normality, and homoscedasticity, may have on Type I error and power. Although earlier studies indicated that analysis of variance is robust to normal distribution violations with large samples (Cochran, 1947; Pearson, 1931; Scheffé, 1959; Srivastava, 1959), more recent research has reported a substantial effect on the power and Type I error rates of parametric techniques in these situations (Bradley, 1978; Clinch & Keselman, 1982; Levine & Dunlap, 1982; Rassmussen, 1985). For example, although several studies have shown that the *F* statistic is robust when groups have the same distribution with a balanced design (Sawilowsky & Blair, 1992; Schmider, Ziegler, Danay, Beyer, & Bühner, 2010; Wu, 2007), the Type I error rate increases and power diminishes when distributions differ in shape (Glass, Peckham, & Sanders, 1972; Harwell, 2003; Lix, Keselman, & Keselman, 1996; Tiku, 1964, 1971; Wilcox, 1995).

These findings indicate that a normal distribution of data cannot be assumed simply on the basis of the robustness of parametric statistics, and that it needs to be checked prior to proceeding with the selected statistical test. Furthermore, there is evidence to suggest that real data are often not normally distributed. Micceri (1989) analyzed the distributional characteristics of over 400 large-sample achievement and psychometric measures and found several classes of contamination in addition to asymmetry and tail weight. Other researchers have also found a variety of non-normal distributions in social and health sciences data, with different shapes and degrees of skewness and kurtosis (Brown, Weatherholt, & Burns, 2010; Harvey & Siddique, 1999, 2000; Hwang & Satchell, 1999; Kobayashi, 2005; Kondo, 1977; Qazi, DuMez, & Uckun, 2007; Shang-Wen & Ming-Hua, 2010; Van Der Linder, 2006).

One of the most widely used procedures for assessing distributional shape is Fisher's measure of skewness (γ_1) and kurtosis or the coefficient of excess (γ_2), based on the third and fourth central moments. Values of $\gamma_1 = 0$ indicate a symmetrical shape, positive values mean that the curve is skewed to the right (right-tail), and negative values suggest skewing to the left (left-tail). The γ_2 coefficient has frequently been considered in the literature as a measure of peakedness and flatness, although other interpretations have also been proposed (Balanda & MacGillivray, 1988, 1990; DeCarlo, 1997; Ruppert, 1987). Values of $\gamma_2 = 0$ mean that the data show the same kurtosis as a normal distribution, $N(0,1)$, whereas positive values are interpreted as being more peaked and negative ones as flatter than the normal. However, it has been argued that the information obtained from these coefficients can be misleading, above all with small sample sizes (An & Ahmed, 2008; Bonato, 2011; Henderson, 2006; Hill & Dixon, 1982; Micceri, 1989), and alternative robust measures have been proposed (Brys, Hubert, & Struyf, 2006; Groeneveld, 1998; Groeneveld & Meeden, 1984; Hill & Dixon, 1982; Hogg, 1974, 1982; Hogg, Fisher, & Randles, 1975; Reed & Stark, 1996). Nevertheless, the majority of simulation studies are based on the modification of γ_1 and γ_2, with two algorithms widely used for simulating the non-normality distribution

Table 1. Descriptive statistics related to sample size as a function of type of measurement (N = 693)

Variables	Mean	Median	Mode	Standard deviations	Minimum	Maximum
Ability	20.57	21	24	5.09	10	30
Other psychological variables	20.02	20	21	4.66	10	29

condition: Fleishman's power transformation method (Fleishman, 1978), extended to the multivariate situation by Vale and Maurelli (1983), and the generalized lambda distribution system (Ramberg, Dudewicz, Tadikamalla, & Mykytka, 1979).

In these simulation studies, researchers usually select either values of γ_1 and γ_2 that represent a well-known distribution (e.g., exponential, double exponential, or lognormal distributions) or several values that define non-known distributions that are considered to represent the real-world situation. In these cases, absolute values of γ_1 and γ_2 less than 1.0 tend to be categorized as slight non-normality, values between 1.0 and about 2.3 are regarded as moderate non-normality, and values beyond 2.3 correspond to severe non-normality (Lei & Lomax, 2005).

The aim of this paper is to assess the distributional shape of real data by examining the values of γ_1 and γ_2 in small samples and thereby obtain a criterion for selecting their proper values in Monte Carlo studies. Small samples are considered because they correspond to what is commonly found in social science publications (Fernández, Vallejo, Livacic-Rojas, & Tuero, 2010; Keselman et al., 1998). Specifically, the analysis concerned 693 distributions corresponding to measures of cognitive ability and other psychological variables that were derived from 130 different populations, with sample size ranging from 10 to 30.

Method

Sample

The analysis focused on 693 distributions derived from natural groups formed in institutions and corresponding to 130 different populations, with sample size ranging from 10 to 30. Of these distributions, 192 were obtained from archive data of high school pupils, 175 were from psychometric studies, and 326 were measures from correlational studies regarding several variables. Measures of cognitive ability (N = 323) and other psychological variables (N = 370) were considered. The measures of ability included scores on the Dominoes Test (D-48), the Differential Aptitudes Test, Primary Mental Aptitudes, Letter Squares, the Identical Forms Test, the Differences Perception Test, Situation-1, the Toulouse-Piéron Test, the Global and Local Attention Test, the Magallanes Visual Attention Scale, and the General Intelligence Factor Test. The measures of psychological variables included scores on the Revised Eysenck Personality Questionnaire, the State-Trait Anxiety Inventory, the Family Environment Scale, the Big-Five Questionnaire, the Beck Depression Inventory, the State-Trait Anger Expression Inventory, the Self-Report Altruism Scale, and the Spanish Psychosocial Scale. Table 1 presents the descriptive statistics related to sample size as a function of the type of measurement.

Procedure

The data were obtained by request to several research groups from Spanish universities and had to fulfill the following requirements: (a) sample size between 10 and 30; (b) they were derived from groups formed in institutions such as classrooms, hospitals, etc.; (c) they had not undergone any data treatment; and (d) they represented measurements of a psychological variable. Requests for data were also made to several high schools for archive data that met the same conditions.

Data Analysis

For each distribution, γ_1 and γ_2 values were calculated as measurements of skewness and kurtosis based on the third and fourth central moments, respectively, as follows:

$$\gamma_1 = \frac{\sqrt{n(n-1)}}{n-2}\frac{m_3}{m_2^{3/2}} \tag{1}$$

$$\gamma_2 = \frac{n-1}{(n-2)(n-3)}\left\{(n+1)\left(\frac{m_4}{m_2^2}-3\right)+6\right\} \tag{2}$$

where n is the sample size, $m_k = \sum_{i=1}^{n}(x_i - \bar{x})^k/n$, and $\bar{x}$ is the sample mean.

The results are presented in the form of descriptive statistics of γ_1 and γ_2, box plots, values as a function of sample size, and frequency of contamination from normal distributions. In order to determine the degree of contamination, 11 cut-off points were arbitrarily established to define contamination in skewness and kurtosis (see Table 2). As regards

Table 2. Arbitrary cut-off points to define contamination

Interval	Skewness/kurtosis
< −2.25	Very extreme negative
−2.25, −1.76	Extreme negative
−1.75, −1.26	High negative
−1.25, −0.76	Moderate negative
−0.75, −0.26	Slight negative
−0.25, 0.25	Near normal
0.26, 0.75	Slight positive
0.76, 1.25	Moderate positive
1.26, 1.75	High positive
1.76, 2.25	Extreme positive
> 2.25	Very extreme positive

Table 3. Descriptive statistics of skewness and kurtosis

	Mean	Median	Mode	Standard deviations	Minimum	Maximum	Range
			Ability (*N* = 323)				
Skewness	−0.09	−0.04	−0.15	0.58	−2.49	1.80	4.43
Skewness (abs)	0.45	0.37	0.05	0.39	0	2.49	2.49
Kurtosis	−0.05	−0.26	−0.71	1.17	−1.62	7.41	9.02
Kurtosis (abs)	0.83	0.70	0.60	0.82	0.01	7.41	7.40
			Other psychological variables (*N* = 370)				
Skewness	0.12	0.11	0.15	0.75	−2.43	2.33	4.76
Skewness (abs)	0.58	0.41	0.30	0.49	0	2.43	2.43
Kurtosis	0.31	−0.03	−0.30	1.41	−1.92	6.82	8.74
Kurtosis (abs)	1.00	0.73	0.30	1.05	0	6.82	6.82
			All distributions (*N* = 693)				
Skewness	0.02	0.02	0.00	0.69	−2.49	2.33	4.83
Skewness (abs)	0.52	0.39	0.05	0.45	0	2.49	2.49
Kurtosis	0.14	−0.17	−0.30	1.32	−1.92	7.41	9.33
Kurtosis (abs)	0.92	0.71	0.30	0.95	0	7.41	7.41

Note. abs = absolute value.

absolute values, six cut-off points were also established to define contamination with respect to combined skewness and kurtosis, from near normal to very extreme contamination. A chi-square test was applied to compare the contamination between the ability measures and the measures of other psychological variables.

Results

Table 3 shows the descriptive statistics of skewness and kurtosis for each type of variable and for all distributions. The values of skewness range between −2.49 and 2.33, with a mean of 0.02 and 0.52 in absolute values. The values for kurtosis range between −1.92 and 7.41, with a mean of 0.14 and 0.92 in absolute values.

Figures 1 and 2 show the box plots of skewness and kurtosis. Values of skewness greater than 1.6 and less than −1.5 are considered outliers for all distributions. For kurtosis, values greater than 2.7 appear as outliers. These box plots suggest a relatively symmetric distribution of skewness and asymmetric one of kurtosis (right-tail).

Figure 3 shows the values of skewness and kurtosis in absolute values as a function of sample size. Both are independent of sample size, with correlation coefficients near zero: .03 and −.02, respectively. These results indicate that values of skewness and kurtosis are similar across the samples with between 10 and 30 individuals.

Table 4 shows the percentage of contamination according to the arbitrary cut-off points. In relation to skewness the results show that 30.9% of the distributions present negative values, 34.1% values close to a symmetrical distribution, and 35% a positive value. As regards kurtosis,

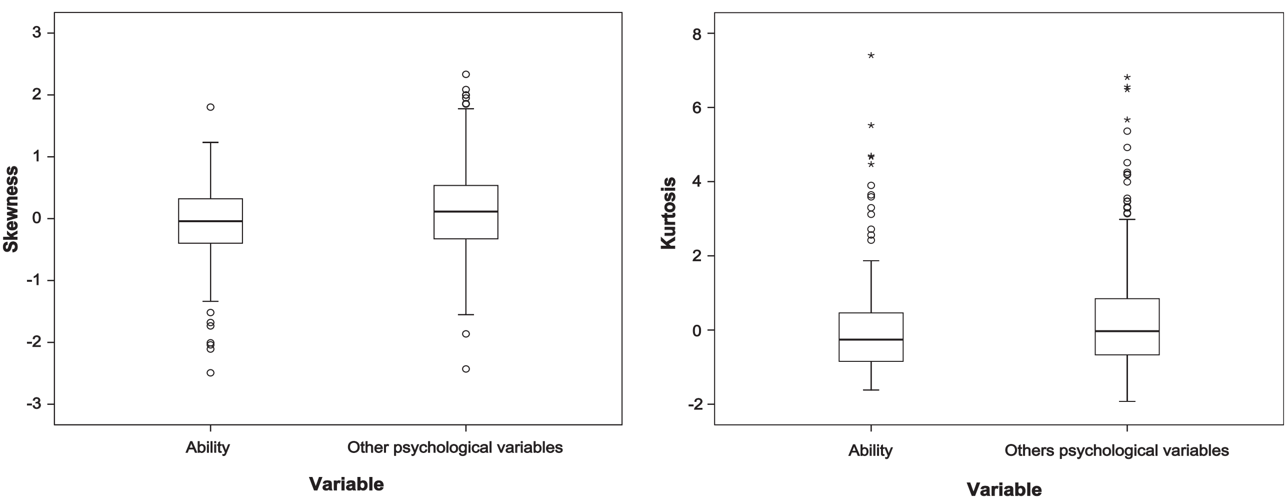

Figure 1. Box plot of skewness and kurtosis as a function of type of variable.

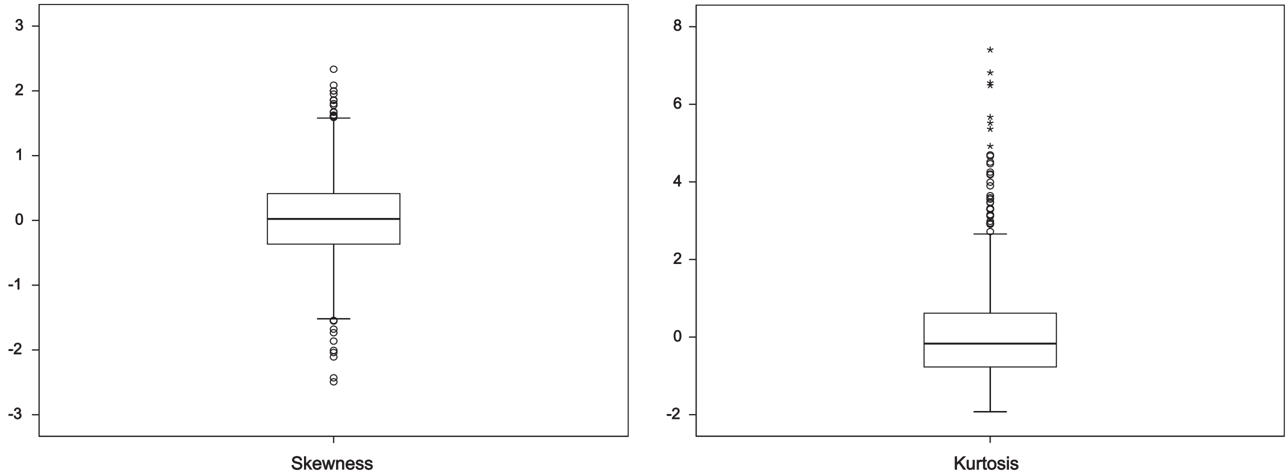

Figure 2. Box plot of skewness and kurtosis for all distributions.

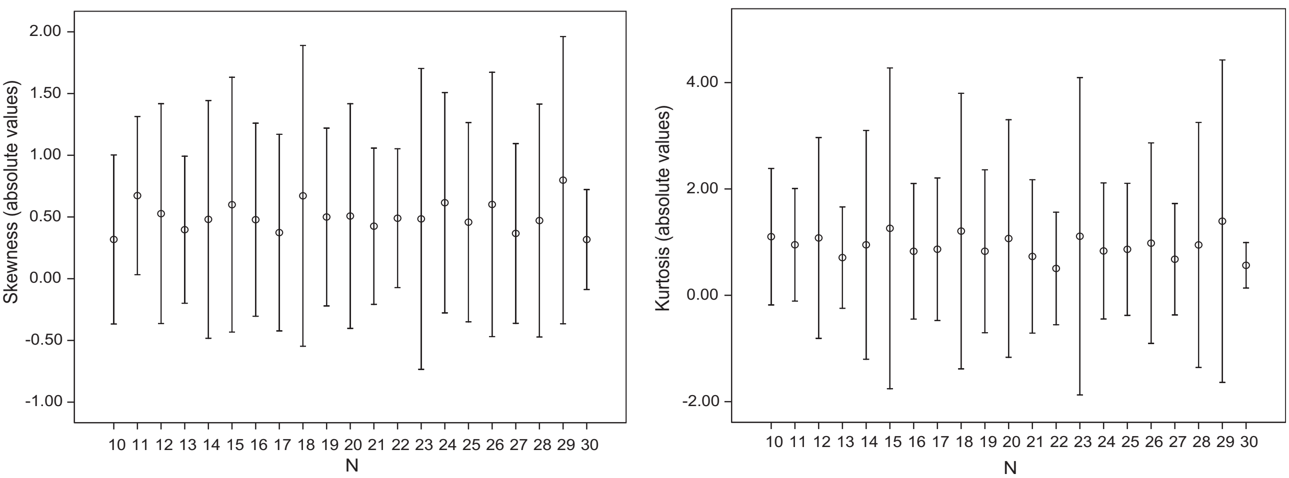

Figure 3. Values of skewness and kurtosis in absolute values as a function of sample size (*N*) (vertical bars represent ±2 standard deviations).

Table 4. Percentage of contamination according to the arbitrary cut-off points of skewness and kurtosis as a function of type of variable

		Skewness			Kurtosis		
Interval	Label	A	PV	All	A	PV	All
< −2.25	Very extreme negative	0.3	0.3	0.3	–	–	–
−2.25, −1.76	Extreme negative	0.9	0.3	0.6	–	0.3	0.1
−1.75, −1.26	High negative	1.9	4.1	3.0	4.0	5.1	4.7
−1.25, −0.76	Moderate negative	9.0	5.4	6.9	24.8	17.0	20.6
−0.75, −0.26	Slight negative	21.1	18.9	20.1	21.1	19.7	20.3
−0.25, 0.25	Near normal	38.1	30.5	34.1	20.4	18.1	19.2
0.26, 0.75	Slight positive	25.1	22.2	23.5	12.1	14.1	13.2
0.76, 1.25	Moderate positive	3.4	11.1	7.5	8.0	7.0	7.6
1.26, 1.75	High positive	–	5.1	2.8	5.3	4.3	4.7
1.76, 2.25	Extreme positive	0.3	1.9	1.1	0.3	4.6	2.5
> 2.25	Very extreme positive	–	0.3	0.1	4.0	9.7	7.0

Note. A = Ability; PV = Other psychological variables; All = All distributions.

Table 5. Percentage of distributions as a function of the arbitrary cut-off points for contamination by skewness and kurtosis

	Skewness										
Kurtosis	< -2.25	-2.25, -1.76	-1.75, -1.26	-1.25, - 0.76	-0.75, -0.26	-0.25, 0.25	0.26, 0.75	0.76, 1.25	1.26, 1.75	1.76, 2.25	> 2.25
< -2.25	-	-	-	-	-	-	-	-	-	-	-
-2.25, -1.76	-	-	-	-	-	0.1					
-1.75, -1.26	-	-	-	-	0.6	3.2	0.9	-	-	-	-
-1.25, - 0.76	-	-	-	0.4	3.8	12.3	4	0.1	-	-	-
-0.75, -0.26	-	-	-	0.7	4.8	8.9	4.9	1	-	-	-
-0.25, 0.25	-	-	-	0.4	5.5	5.5	5.8	2	-	-	-
0.26, 0.75	-	-	-	1.6	2.6	2.5	4.9	1.3	0.3	-	-
0.76, 1.25	-	-	-	1.7	1.2	1.2	2	0.9	0.6	-	-
1.26, 1.75	-	-	0.3	1.6	0.4	0.3	0.9	0.7	0.6	-	-
1.76, 2.25	-	-	0.4	0.4	0.6	-	0.1	0.3	0.6	0.1	-
> 2.25	0.3	0.6	2.3	0.1	0.6	0.1	-	1.2	0.7	1	0.1

45.7% of the distributions present a negative value, 19.2% values close to a normal distribution, and 35.1% a positive value.

Table 5 shows the joint distribution of skewness and kurtosis across the arbitrary cut-off points of contamination. Only 5.5% of the distributions were close to expected values under normality. The highest percentage of distributions found, 12.3%, corresponds to values of skewness between −0.25 and 0.25 and of kurtosis between −1.25 and −0.76.

Of the 11 cut-off points, six were established, independently of the sign, to define contamination with combined skewness and kurtosis values. These six points correspond to the squares indicated in Table 5. The maximum values of skewness and kurtosis are shown in Table 6 with the percentage of distributions as a function of the type of variable. The results show that ability measures are less contaminated than are measures of other psychological variables, $\chi^2(5) = 25.394$, $p < .01$.

Table 6. Percentage of distributions as a function of the arbitrary cut-off points of contamination

Maximum values of skewness and/or kurtosis	Label	A	PV	All
0, |0.25|	Near normal	6.2	4.9	5.5
|0.26, 0.75|	Slight	41.5	38.4	39.9
|0.76, 1.25|	Moderate	38.7	31.2	34.5
|1.26, 1.75|	High	9.3	11.1	10.4
|1.76, 2.25|	Extreme	0.3	4.9	2.6
|> 2.25|	Very extreme	4.0	9.7	7.0

Note. A = Ability; PV = Other psychological variables; All = All distributions.

Discussion

The aim of this paper was to assess the distributional shape of real data by examining the values of skewness and kurtosis in small samples. A total of 693 distributions, including measures of cognitive ability and other psychological variables, were analyzed. For each distribution, γ_1 and γ_2 values were calculated as measurements of skewness and kurtosis based on the third and fourth central moments, respectively.

The results indicate that values of γ_1 ranged between −2.49 and 2.33, with 34.1% presenting values close to a symmetrical distribution. Values of γ_2 ranged between −1.92 and 7.41, and only 19.2% presented values close to zero. Furthermore, kurtosis values far from zero were more frequent than were values of skewness. Both coefficients were independent of sample size.

Considering γ_1 and γ_2 jointly, only 5.5% of the distributions were close to expected values under normality. Overall, 39.9% of distributions were slightly non-normal, with maximum values (in absolute terms) of both coefficients being in the range 0.26–0.75. A further 34.5% of distributions were moderately non-normal, with values in the range 0.76–1.25. Finally, 10.4% of distributions showed high contamination (range 1.26–1.75), while 2.6% and 7% can be considered as presenting extreme (range 1.76–2.25) and very extreme (> 2.25) contamination, respectively. Thus, 74.4% of distributions presented either slight or moderate contamination, while 20% showed a more extreme contamination. These results indicate that normality is not the rule with small samples and are consistent with the conclusions of other researchers who have found a variety of non-normal distributions in social and health sciences data (Brown et al., 2010; Harvey & Siddique, 1999, 2000; Hwang & Satchell, 1999; Kobayashi, 2005; Kondo, 1977; Micceri, 1989; Qazi et al., 2007; Shang-Wen & Ming-Hua, 2010; Van Der Linder, 2006). However, extreme departures from the normal distribution do not seem to be very frequent in the distributions analyzed here. The present results also indicate that ability measures are less contaminated than are measures of other psychological variables such as personality, anxiety, depression, etc., being this finding consistent with Micceri (1989).

The real data analyzed here did not represent values of skewness and kurtosis as those used in many Monte Carlo studies of statistical robustness. This suggests that researchers might improve the relevance of their robustness findings by using a range of typical, for their discipline, empirical rather than theoretical distributions. At all events, researchers should check rather than assume that data are normally distributed, and should consider using the nonparametric statistics and tests with robust estimators that have been proposed as alternatives to parametric tests for independent groups and repeated measures if the power and Type I and Type II error rates are distorted (e.g., Akritas & Brunner, 1997a, 1997b, 2003; Brunner, Domhof, & Langer, 2002; Brunner & Puri, 2002; Heritier, Cantoni, Copt, & Victoria-Feser, 2009; Keselman et al., 2008; Luh & Guo, 2001, 2004; Rauf, Werner, & Brunner, 2008; Shah & Madden, 2004; Wilcox, 1993, 2001, 2002, 2003, 2005, 2009; Wilcox & Keselman, 2001).

Acknowledgment

This research was supported by Grant No. PSI2009-11136 from Spanish Ministry of Science and Innovation.

References

Akritas, M. G., & Brunner, E. (1997a). Nonparametric methods for factorial design with censored data. *Journal of the American Statistical Association, 92*, 568–576.

Akritas, M. G., & Brunner, E. (1997b). A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference, 61*, 249–277.

Akritas, M. G., & Brunner, E. (2003). Nonparametric models for ANOVA and ANCOVA: A review. In M. G. Akritas & D. N. Politis (Eds.), *Recent advances and trends in nonparametric statistics* (pp. 79–91). Amsterdam, The Netherlands: Elsevier.

An, L., & Ahmed, E. (2008). Improving the performance of kurtosis estimator. *Computational Statistic & Data Analysis, 52*, 2669–2681.

Balanda, K. P., & MacGillivray, H. L. (1988). Kurtosis: A critical review. *The American Statistician, 42*, 111–119.

Balanda, K. P., & MacGillivray, H. L. (1990). Kurtosis and spread. *Canadian Journal of Statistics, 18*, 17–30.

Bonato, M. (2011). Robust estimation of skewness and kurtosis in distributions with infinite higher moments. *Finance Research Letters, 8*, 77–87.

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

Brown, D. D., Weatherholt, T. N., & Burns, B. M. (2010). Attention skills and looking to television in children from low income families. *Journal of Applied Developmental Psychology, 31*, 330–338.

Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*. New York, NY: Wiley.

Brunner, E., & Puri, M. L. (2002). A class of rank-score tests in factorial designs. *Journal of Statistical Planning Inference, 103*, 331–360.

Brys, G., Hubert, M., & Struyf, A. (2006). Robust measures of tail weight. *Computational Statistics & Data Analysis, 50*, 733–759.

Clinch, J. J., & Keselman, H. J. (1982). Parametric alternatives to the analysis of variance. *Journal of Educational Statistics, 7*, 207–214.

Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics, 33*, 22–38.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*, 292–307.

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 10 años [Characteristics and analyses of the repeated measures designs in experimental research in Spain during the last ten years]. In *Actas del XI Congreso de Metodología de las Ciencias Sociales y de la Salud* (pp. 193–198). Málaga: UMA-Tecnolex.

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

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

Groeneveld, R. (1998). A class of quantile measures for kurtosis. *The American Statistician, 51*, 325–329.

Groeneveld, R., & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician, 33*, 391–399.

Harvey, C., & Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis, 34*, 465–487.

Harvey, C., & Siddique, A. (2000). Conditional skewness in asset pricing test. *Journal of Finance, 55*, 1263–1295.

Harwell, M. R. (2003). Summarizing Monte Carlo results in methodological research: The single-factor, fixed-effects ANOVA case. *Journal of Educational Statistics, 28*, 45–70.

Henderson, A. R. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta, 366*, 112–129.

Heritier, S., Cantoni, E., Copt, S., & Victoria-Feser, M. P. (2009). *Robust methods in biostatistics*. West Sussex, UK: Wiley.

Hill, M. A., & Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrics, 38*, 377–396.

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and

theory. *Journal of the American Statistical Association, 69*, 909–927.
Hogg, R. V. (1982). On adaptive statistical inferences. *Communications in Statistics: Theory and Methods, 11*, 2531–2542.
Hogg, R. V., Fisher, D. M., & Randles, D. H. (1975). A two-sample adaptive distribution-free test. *Journal of the American Statistical Association, 70*, 656–661.
Hwang, S., & Satchell, S. (1999). Modeling emerging market risk premia using higher moments. *International Journal of Finance and Economics, 4*, 271–296.
Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods, 13*, 110–129.
Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., & 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.
Kobayashi, K. (2005). Analysis of quantitative data obtained from toxicity studies showing non-normal distribution. *Journal of Toxicological Science, 30*, 127–134.
Kondo, K. (1977). The log-normal distribution of the incubation time of exogenous diseases. *Japanese Journal of Human Genetics, 21*, 217–237.
Lei, M., & Lomax, R. G. (2005). The effect of varying degrees on nonnormality in structural equation modeling. *Structural Equation Modeling, 12*, 1–27.
Levine, D. W., & Dunlap, W. P. (1982). Power of the *F* test with skewed data: Should one transform or not? *Psychological Bulletin, 9*, 22–80.
Lix, L. M., Keselman, J. C., & 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.
Luh, W. M., & Guo, J. H. (2001). Using Johnson's transformation and robust estimators with heteroscedastic test statistics: An examination of the effects of nonnormality and heterogeneity in the nonorthogonal two-way ANOVA design. *British Journal of Mathematical and Statistical Psychology, 54*, 79–94.
Luh, W. M., & Guo, J. H. (2004). Improved robust test statistic base on trimmed means and Hall's transformation for two-way ANOVA models under non-normality. *Journal of Applied Statistics, 31*, 623–643.
Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156–166.
Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika, 23*, 114–133.
Qazi, S., DuMez, D., & Uckun, F. M. (2007). Meta analysis of advanced cancer survival data using lognormal parametric fitting: A statistical method to identify effective treatment protocols. *Current Pharmaceutical Design, 13*, 1533–1544.
Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., & Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics, 21*, 201–214.
Rassmussen, J. L. (1985). The power of Student's *t* and Wilcoxon statistics. *Evaluation Review, 9*, 505–510.
Rauf, M., Wener, C., & Brunner, E. (2008). Analysis of high-dimensional repeated measures designs: The one sample case. *Computational Statistics and Data Analysis, 53*, 416–427.
Reed, J. F., & Stark, D. B. (1996). Hinge estimators of location: Robust to asymmetry. *Computer Methods and Programs in Biomedicine, 49*, 11–17.
Ruppert, D. (1987). What is kurtosis? An influence function approach. *The American Statistician, 41*, 1–5.
Sawilowsky, S. S., & 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*, 353–360.
Scheffé, H. (1959). *The analysis of variance*. New York, NY: Wiley.
Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology, 6*, 147–151.
Shah, D. A., & Madden, L. V. (2004). Nonparametric analysis of ordinal data in design factorial experiment. *Phytopathology, 94*, 33–43.
Shang-Wen, Y., & Ming-Hua, H. (2010). Estimation of air traffic longitudinal conflict probability based on the reaction time of controllers. *Safety Science, 48*, 926–930.
Srivastava, A. B. L. (1959). Effect of non-normality on the power function of *t*-test. *Biometrika, 46*, 114–122.
Tiku, M. L. (1964). Approximating the general nonnormal variance-ratio sampling distribution. *Biometrika, 51*, 83–95.
Tiku, M. L. (1971). Power function of the *F*-test under non-normal situations. *Journal of the American Statistical Association, 66*, 913–916.
Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*, 451–464.
Van Der Linder, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*, 181–204.
Wilcox, R. R. (1993). Analysing repeated measures or randomized block design using trimmed means. *British Journal of Mathematical and Statistical Psychology, 46*, 63–76.
Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect sizes? *Review of Educational Research, 65*, 51–77.
Wilcox, R. R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy.* New York, NY: Springer.
Wilcox, R. R. (2002). Understanding the practical advantages of modern ANOVA methods. *Journal of Clinical and Adolescent Psychology, 31*, 399–412.
Wilcox, R. R. (2003). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.
Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.
Wilcox, R. R. (2009). *Understanding conventional methods and modern insights*. New York, NY: Oxford University Press.
Wilcox, R. R., & Keselman, H. J. (2001). Using trimmed means to compare *K* measures corresponding to two independent groups. *Multivariate Behavioral Research, 36*, 421–444.
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.

Received September 13, 2011
Accepted December 29, 2011
Published online May 10, 2012

María J. Blanca

Facultad de Psicología
Dpto. Psicobiología y Metodología
de las Ciencias del Comportamiento
Campus Universitario de Teatinos, s/n
29079 Málaga
Spain
Tel. +34 95 213-1088
Fax +34 95 213-2621
E-mail blamen@uma.es

