



On the efficacy of procedures to normalize Ex-Gaussian distributions

Fernando Marmolejo-Ramos^{1*}, Denis Cousineau², Luis Benites³ and Rocío Maehara³

¹ Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Stockholm, Sweden

² School of Psychology, University of Ottawa, Ottawa, ON, Canada

³ Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

Edited by:

Holmes Finch, Ball State University, USA

Reviewed by:

Jocelyn Holden Bolin, Ball State University, USA

Julianne M. Edwards, Ball State University, USA

*Correspondence:

Fernando Marmolejo-Ramos, Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Frescati Hagväg 9A, Stockholm, SE-106 91, Sweden
e-mail: fernando.marmolejo.ramos@psychology.su.se;
Website: <http://sites.google.com/site/fernandomarmolejoramos/>

Reaction time (RT) is one of the most common types of measure used in experimental psychology. Its distribution is not normal (Gaussian) but resembles a convolution of normal and exponential distributions (Ex-Gaussian). One of the major assumptions in parametric tests (such as ANOVAs) is that variables are normally distributed. Hence, it is acknowledged by many that the normality assumption is not met. This paper presents different procedures to normalize data sampled from an Ex-Gaussian distribution in such a way that they are suitable for parametric tests based on the normality assumption. Using simulation studies, various outlier elimination and transformation procedures were tested against the level of normality they provide. The results suggest that the transformation methods are better than elimination methods in normalizing positively skewed data and the more skewed the distribution then the transformation methods are more effective in normalizing such data. Specifically, transformation with parameter $\lambda = -1$ leads to the best results.

Keywords: Ex-Gaussian, reaction times, normality tests, outliers

INTRODUCTION

Reaction times (RTs) have been a privileged measure of behavior in experimental psychology allowing an estimation of the duration of cognitive processes and inference of the likely cognitive process (see Donders et al., 1969). Hence, their understanding and proper analysis is essential. It is known that reaction time data are positively skewed, and therefore are not normally distributed. As Olivier and Norberg (2010) argue, commonly used statistical tests are not appropriate for the analysis of RT data since RTs are (in most cases) non-normally distributed. Yet, most researchers rely on parametric tests (primarily ANOVA) to analyze reaction times data despite these tests assumptions are not met with RT data. More specifically, they require variables to be normally distributed within conditions and have homogeneous variances between conditions in order to give unbiased results (e.g., Calkins, 1974; Marmolejo-Ramos and González-Burgos, 2013). Even small violations of those assumptions can lead to biased results from the tests (see Wilcox, 1998). To meet these conditions, many researchers transform the data and/or search for maverick data points. The aim of this paper is to compare various procedures that assist in normalizing data via power transformations and outlier elimination procedures.

REACTION TIME DISTRIBUTIONS AND METHODS TO DEAL WITH OUTLIERS

Reaction time distributions are characterized by a positive skew. Many explanations have been proposed to explain this near universal finding (with one exception; Hopkins and Kristofferson, 1980, who found symmetrical RT distributions). The first explanation from McGill (1963), argued that observable RTs are caused by two

processes operating in succession. The first is a central decision mechanism whose distribution is highly skewed (exponential distribution). This mechanism is related to an accumulation of information processes whose activation times have a rate of accumulation $\tau \text{ ms}^{-1}$. These assumptions are based on neurological studies of single cell firing patterns (see e.g., Langlois et al., 2014) showing that for a small threshold, the resulting distribution is very skewed and well-described by an exponential distribution. The second process is responsible for response selection and motor execution. This second process is presumably affected by many factors and therefore (owing to the central limit theorem) results in a normal distribution. The sum of these two sets of time has a distribution described by a convolution of a Gaussian distribution and an exponential distribution known as an Ex-Gaussian distribution. Hohle (1965), Ratcliff and Murdock (1976), Hockley (1984), and Heathcote et al. (1991), among others, fitted this distribution to RTs and found a generally good fit.

In other words, a simple cognitive task starts at the level of the perceptual processes. Light travels to the retina (negligible time), activating the cones and rods on the retinae and transducing the signal through the visual brain areas (V1, V2, etc.). Following perception and up to a semantically meaningful percept, there is the decision process presumably occurring in the frontal lobes. A decision is then followed by activations sent for response selection and down to the motor areas and spinal cord triggering a muscular response, which puts pressure on a response key. Overall, a simple decision involves a chain of signal transformation through a dozen specialized brain areas each adding to the observed latency. The total processing chain can be subdivided into three stages: perception, decision and response selection, and motor response.

Based on the assumption that the time taken by each brain area adds up to the total response time observed by the apparatus, and knowing that manipulating the difficulty of the decision without altering the perceptibility of the stimuli and without altering the motor response complexity can affect skew, it can be hypothesized that (1) perception processes add up to a total perception time; (2) response selection and motor response processes add up to a total response time, and (3) the balance leads to a decision time. Although the time taken by the perceptual processes are unknown, owing to the Central limit theorem, if multiple processes with unknown times are added to obtain a total time then the resulting perception time should be normally distributed. This same principle applies to the response selection and motor response stage.

In recent years, though, some authors have questioned the additivity assumptions. Under an alternative view of the chain of processing, the brain areas send activations continuously and related areas react when a critical amount of activations have been received. Hence, each area is not operating in isolation. Thus, violating the independence of operation of each sub-process implicit in the additivity assumption. Theorems analogous to the Central limit theorem suggest that the resulting perception time should be log-normally distributed in this scenario (Ulrich and Miller, 1993; Mouri, 2013). Finally, as the decision processes are based on just a few sub-processes, asymptotic theorems cannot be invoked and this stage should preserve its highly skewed characteristic at the level of latency.

Other explanations have subsequently been proposed to explain the skew in RTs. Ulrich and Miller (1993) suggested that response times may be caused by a cascade of events (following McClelland, 1979, cascade model). This model predicts a distribution called the Log-Normal (also see West and Shlesinger, 1990)

whose shape is indistinguishable from the EGd (Chechile, personal communication). Raab (1962), followed by LaBerge (1962) and Pike (1973), instead proposed a race model where brain signals compete with each other to be the first to trigger a response (recent documentation includes Rouder, 2000; Miller and Ulrich, 2003; Cousineau, 2004). These models all suggest that the Weibull distribution should be the distribution of RTs (see also Schwarz, 2001 for a variation of this idea). The Weibull and Ex-Gaussian/Log-Normal can be in principle distinguished, but this requires a lot of RTs per conditions (more than 100), uncontaminated by practice effects, fatigue effects, etc. Nevertheless, the true distribution of RT may also (more likely) be none of the above.

In what follows, we assume that the EGd is a convenient way to characterize RTs much like statisticians assume the normal distribution. Furthermore, the literature indicates that the EGd is the distribution most broadly explored. A comprehensive characterization of EGds can be found in Marmolejo-Ramos and González-Burgos (2013). The parameters of an EGd are represented by a mean (μ), a SD (σ), and an exponential factor (τ). The mean and the SD represent times from the normally distributed stage of processing, whereas the exponential factor represents times from the exponentially distributed stage of processing (see McAuley et al., 2006). The mean of an EGd can be inferred from its parameters, as well as its standard variation. The EGd's SD and exponential factor can be used to estimate its third (skewness) and fourth (kurtosis) central moments (see **Figure 1**).

It is a common practice in psychological research that measures RT data to deal with observations which fall very far from a group's mean. Such observations are possibly the product of participants' lack of attention (very long RTs) or overly fast guesswork (very fast RTs). However, it may also be caused by random fluctuations in

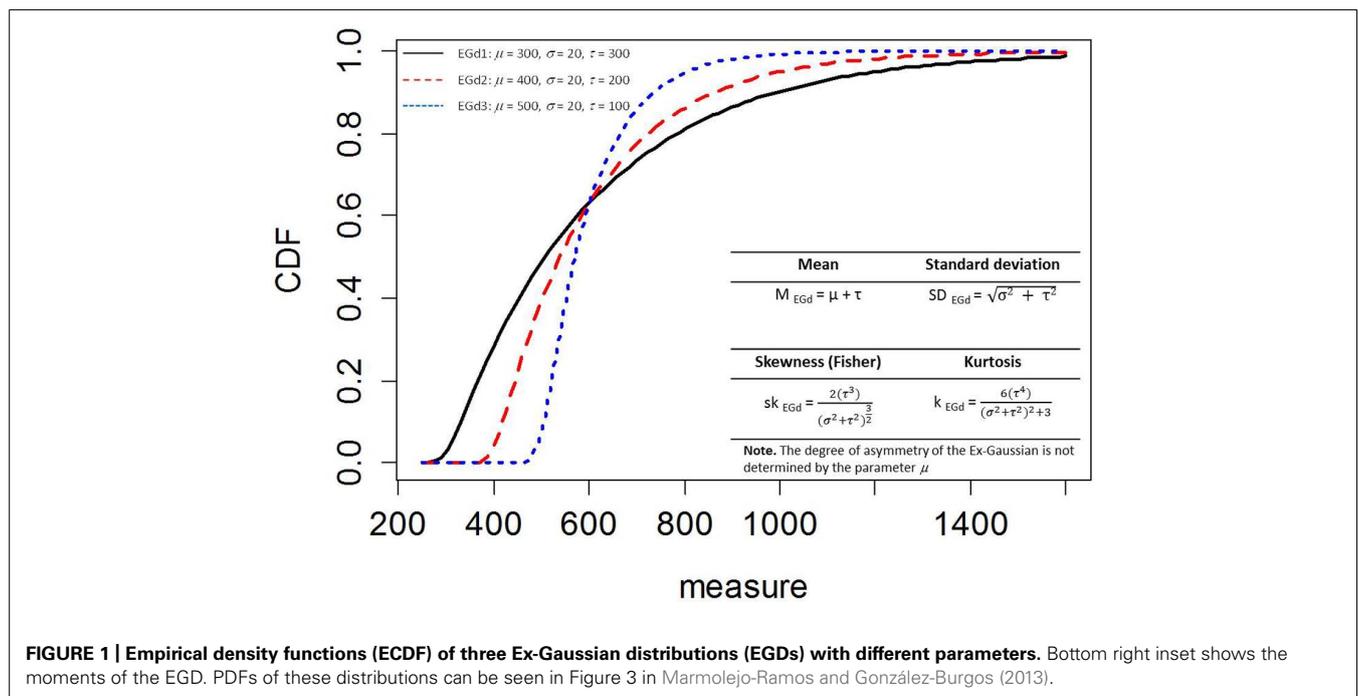


FIGURE 1 | Empirical density functions (ECDF) of three Ex-Gaussian distributions (EGDs) with different parameters. Bottom right inset shows the moments of the EGd. PDFs of these distributions can be seen in Figure 3 in Marmolejo-Ramos and González-Burgos (2013).

internal thresholds. These outliers deeply affect the estimation of the data's central tendency (see Whelan, 2008; Baayen and Milin, 2010; Cousineau and Chartier, 2010). Although for some researchers, outliers are not influential in data sets (see Orr et al., 1991; Lance et al., 1996), most psychologists (e.g., Judd et al., 1995) and statisticians (e.g., Beckman and Cook, 1983) agree that outliers affect parameter estimation. Outliers are usually dealt with by using pre-determined criteria; mainly via SD cut-offs (e.g., eliminating observations above and/or below 2 SD, see Cowles and Davis, 1982) or data re-expressions (e.g., transforming the data into logarithms, see Bland and Altman, 1996). Additionally, a combination of these approaches has been proposed (e.g., transforming the data and then removing outliers or vice versa, see Marmolejo-Ramos and Matsunaga, 2009).

Under the 2 SD procedures, researchers remove observations which fall ± 2 SD from a participant's mean in a particular condition (see Ratcliff, 1993; for an example of this application see Bertels et al., 2010). This procedure trims long tails in RT distributions on a subject per condition basis, but at the cost of leaving experimental conditions with an unequal number of trials. Although SD cut-off leads to an underestimation of the population's RT, such a biased estimation appears to depend on sample size (Perea, 1999). In addition, the cut-off values are symmetrical about the mean but the data are not. Hence, it is more likely that high outliers will be removed, resulting in a systematic bias to reduce the observed mean. To minimize overestimation bias, some researchers have proposed adjustments for a number of SD according to sample size (see Van Selst and Jolicoeur, 1994; Thompson, 2006; see also Table 2 in Cousineau and Chartier, 2010). Nevertheless, in practice most researchers rarely adjust SD cut-off according to sample size using a 2 SD, 2.5 SD, or even a 3 SD cut-off criterion instead (for examples, of each see Havas et al., 2007; Bertels et al., 2010; Otte et al., 2011; respectively, see Leys et al., 2013; for a review).

The SD cut-off is the most commonly used procedure to deal with outliers in RT research. However, advances in statistics suggest the use of more robust methods to deal with outliers. One such approach derives from multivariate outlier detection methods and is called the minimum covariance determinant (MCD) method. This method aims to estimate the best subset of normally distributed points in a data set which are clustered in an ellipsoid with the smallest volume (or minimum covariance matrix). The computations of the MCD rely on Mahalanobis distances and robust estimators of multivariate location (see Rousseeuw and van Driessen, 1999). Although the MCD method is primarily designed to deal with multivariate data, it does not preclude it from being applied to univariate data.

Another approach to deal with non-normality is data transformations. With this procedure all observations are retained but they are re-expressed using a different, non-linear, scale that improves normality of the data (see Osborne, 2002; Olivier and Norberg, 2010). RT data can be re-expressed into logarithms (for an example of this application see Markman and Brendl, 2005), square-roots (for an example see Moran and Schwartz, 1999), and inverse (for an example see Moss et al., 1997). A well-known transformation that achieves all these re-expressions is the Box-Cox transformation. In this transformation, the selection of a particular parameter

known as lambda, is accompanied by a (restricted) log-likelihood statistic that signals the best parameter needed in order to achieve the highest normality (see Olivier and Norberg, 2010). Thus, specific lambda parameters have been associated with the inverse ($\lambda = -1$), logarithmic ($\lambda = 0$), and square-root ($\lambda = 0.5$) transformations and previous studies have suggested that the inverse transformation has a strong normalization effect (see Ratcliff, 1993).

A simulation study in which the normalization power of the Box-Cox transformations and elimination procedures are tested against a particular type of skewed distributions is yet to be done. As to the Box-Cox method, it would be useful to see how other transformation parameters could improve the normality of EGds. Thus, the intermediate parameters -0.5 are worth testing since it can be seen as a trade-off between an inverse (i.e., -1) and a logarithmic (i.e., 0) transformation. The present simulation study aims to test the power of these outlier elimination and transformation methods to normalize EGds of different parameters and sample sizes. The results will indicate the most effective methods when dealing with positively skewed distributions.

MATERIALS AND METHODS

VALIDATION OF AN ALTERNATIVE SIMULATION METHOD AND A COMPREHENSIVE APPROACH TO THE ASSESSMENT OF NORMALITY IN NON-NORMAL DISTRIBUTIONS

In order to determine how various outlier elimination and transformation methods can improve the normality of data sampled from EGds, it is necessary to first check the normality of the EGds before applying these methods. A typical approach is to estimate the power of normality tests against non-normal distributions. Under this approach it is traditional to (i) Compute the Critical values (CVs) of one or more normality tests against a $N(0,1)$ of different sample sizes, and to (ii) Use those CVs as cut-off points to reject normality in non-normal distributions of the same sample sizes used in the simulations (see Marmolejo-Ramos and González-Burgos, 2013, for a detailed explanation).

The power of a normality test relies on the number of times the test correctly rejects normality. We note hereafter the proportion of rejection of normality as PoR. A high PoR (e.g., a PoR close to 1) signifies that the distribution being tested is highly non-normal, whereas a low PoR indicates otherwise (e.g., a PoR close to 0). On the other hand, all tests should show PoRs hovering around 0.05 (as $\alpha = 0.05$) when tested against a normally distributed set of data regardless of the sample size (e.g., Romão et al., 2010). Such a situation is to be expected since normality tests should have a low probability of incorrectly rejecting the hypothesis that a $N(0,1)$ is normal, and that probability should be close to the nominal level used in the study (the α level). In sum, normality tests should have a low PoR, against normal distributions and have a high power, or a high PoR, against non-normal distributions.

In a recent study, Marmolejo-Ramos and González-Burgos (2013) studied the power of various normality tests against three EGds and other non-normal distributions such as the Weibull (2,1) and Log-Normal (0,1). Their results not only replicated the comprehensive results reported by Romão et al. (2010) regarding the

Weibull and Log-Normal distributions, but also found that of all the normality tests studied, the Shapiro–Wilk (SW) was the test with the highest power against EGds. For instance, using the CVs approach described above, these researchers found that the SW test has a power of approximately 0.45 when dealing with EGds with parameters $\mu = 300$, $\sigma = 20$, and $\tau = 300$ when the sample size was 10. When the sample size was 10 and the parameters of the EGds were $\mu = 400$, $\sigma = 20$, and $\tau = 200$ and $\mu = 500$, $\sigma = 20$, and $\tau = 100$, the powers of the SW test were just below 0.45 and between 0.35 and 0.40, respectively (see Figure 4, in their study for more detailed results).

The present study features an alternative approach in which the p -value associated with a normality test is used. Marmolejo-Ramos and González-Burgos (2013, footnote 5) argue that in simulation studies, p -values are not mandatorily calculated as they are calculated in statistical packages. That is, statistical packages rely on theoretical distributions, while simulation studies rely on empirical distributions. The present study uses the p -value associated with a normality test, as given in statistical packages, to validate its usage as an alternative way to measure the PoR given by a normality test for a certain distribution. The PoR results obtained via the p -value are simply the proportion of times a normality test gives p -values below, and is not equal to, a chosen alpha level, e.g., $\alpha = 0.05$, when tested against a certain distribution of a particular sample size. Thus, if the p -value approach proposed herein is effective, it should be able to replicate or approximate the results found by Marmolejo-Ramos and González-Burgos (2013) regarding the power of SW test against various EGds.

To validate the p -value approach, a simulation study was performed to test the power of the SW against the same EGds described above, when sample size was 10, and with an alpha level of 0.05. Additionally, the present simulations implement the method proposed by Marmolejo-Ramos and González-Burgos (2013) consisting of iterating (i) each simulation (s) a set number of times and estimating measures of central tendency (a) and dispersion (SD) across iterations. Thus, these parameters were: $i = 30$, $s = 20'000$, and $a =$ the Mean (and its ± 1 SD). That is, each run of 20'000 simulations was iterated 30 times and for each vector containing 30 iterations, the mean PoR, the mean p -value and their associated SDs were estimated. More importantly, the results of the iterations are amenable to formal statistical analysis in order to determine the main effects of and interactions between the variables included in the simulations. The results indicate that the proposed p -value method does replicate the results obtained by Marmolejo-Ramos and González-Burgos (2013); PoR: $M_{EGd1} = 0.427$ (SD = 0.003), $M_{EGd2} = 0.413$ (SD = 0.003), and $M_{EGd3} = 0.354$ (SD = 0.003); p -value: $M_{EGd1} = 0.178$ (SD = 0.001), $M_{EGd2} = 0.188$ (SD = 0.001), and $M_{EGd3} = 0.232$ (SD = 0.001), as determined by the SW normality test when $n = 10$.

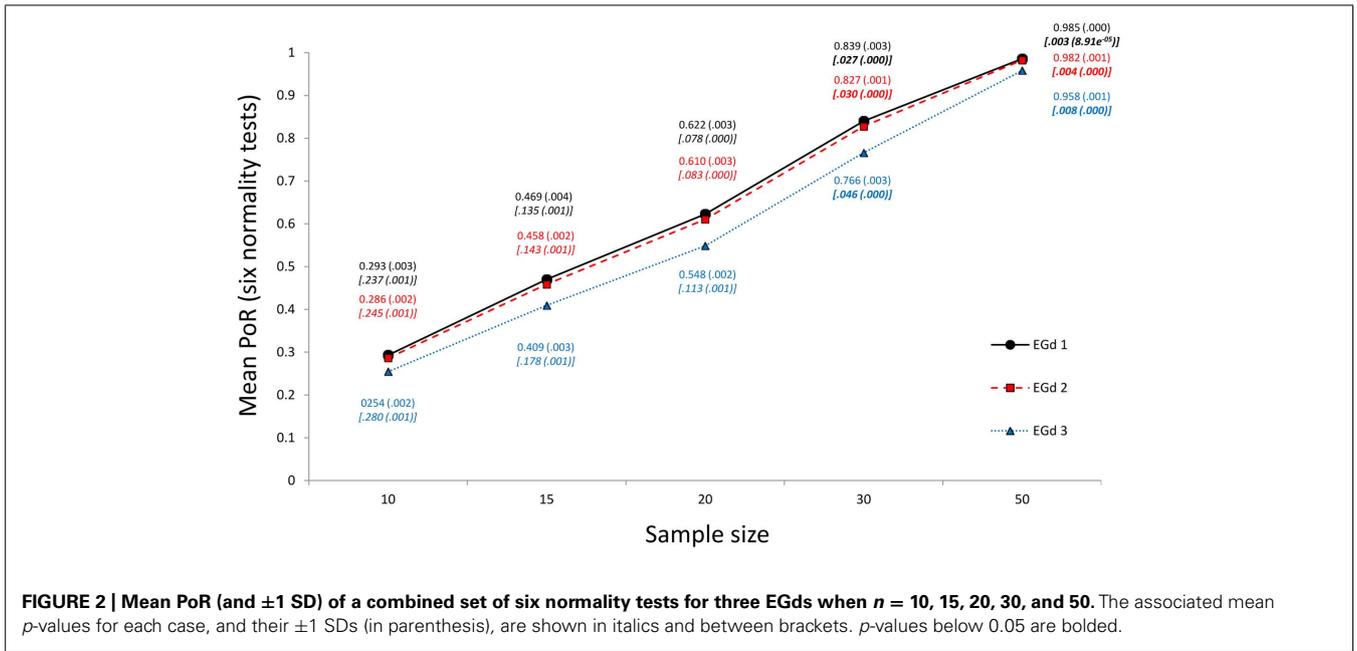
As suggested above, it is essential to determine the status of non-normal EGds before any outlier elimination or transformation method is applied. As is traditional in simulation studies of normality tests (see Romão et al., 2010; Alizadeh Noughabi and Arghami, 2011; Yap and Sim, 2011), the PoR of a particular normality test is computed for a certain distribution

of a certain sample size. Indeed, this is the usual means of evaluating the normality of a data set, i.e., usually researchers use the SW or the Kolmogorov–Smirnov test (KS) to determine what the normality of a data set is (see Marmolejo-Ramos and González-Burgos, 2013). However, as has been shown through simulation studies of normality tests, some tests have higher power than others in determining normality, and the type of distribution being tested plays a role in this (Engmann and Cousineau, 2011). Thus, relying on a sole normality test could be problematic in that it is difficult to determine the parent distribution of the data set in advance. Marmolejo-Ramos and González-Burgos (2013) proposed a method that can assist in ameliorating this issue. These researchers recommend fitting the data with a set of potential parent distributions, and estimating which parent distribution gives the best fit. Once a parent distribution is identified, it could be possible to select an appropriate fitting normality test that is powerful against the type of distribution.

An alternative method in which the combined results of various normality tests are used is proposed herein. There are approximately 40 different types of normality tests (see Razali and Wah, 2010) that can be categorized as regression/correlations, empirical distribution functions, measure of moments, or a combination of these (see Romão et al., 2010; Marmolejo-Ramos and González-Burgos, 2013). New normality tests are still being proposed (e.g., Akbilgiç and Howe, 2011; Harri and Coble, 2011; He and Xu, 2013), which may lead to new categorizations. Thus, it seems rather inadvisable to rely solely on one test, especially when considering that tests also differ based on the different characteristics of the normal distribution on which they focus (Romão et al., 2010). Therefore, a comprehensive assessment of normality would require the combination of results given by normality tests from different categories. That is, an average of the p -values given by normality tests belonging to the categories mentioned above should give an educated approximation of the normality of a given distribution. **Figure 2** shows the results of applying a normality-tests-combination method to the three EGds mentioned above when sample sizes are 10, 15, 20, 30, and 50 via the Marmolejo-Ramos and González-Burgos' simulation method described above. The normality tests used were the SW, Shapiro–Francia (SF); these are regression/correlation-based tests), KS, Anderson–Darling (AD); these are empirical distribution function-based tests), Doornik–Hansen (DH), and the robust Jarque–Bera (rJB); these are measure of moments-based tests; details in relation to these tests can be found in Romão et al., 2010)¹.

This section aims to determine the degree of normality achieved by the outlier elimination and transformation methods described above, on various EGds. The six normality tests

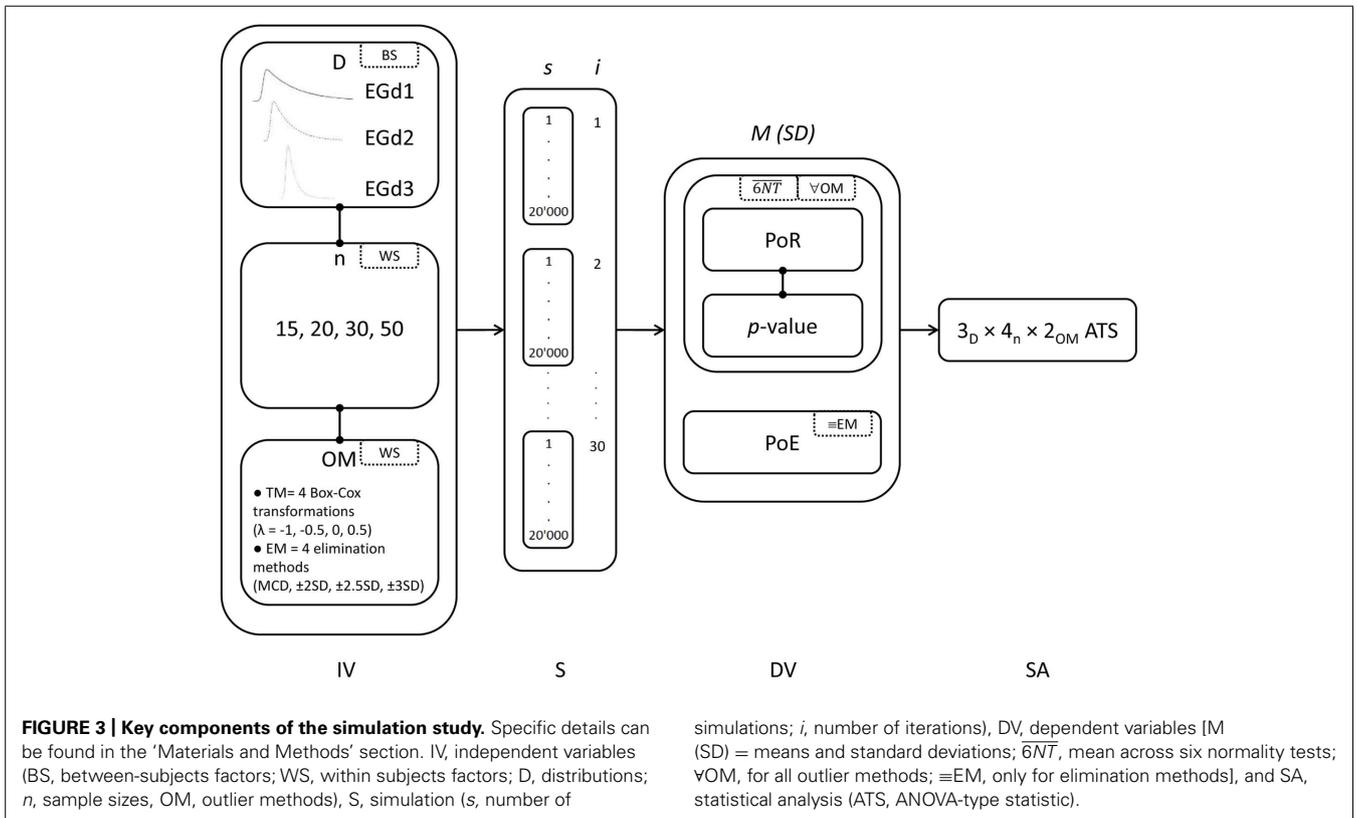
¹All of these tests are implemented in R in the following packages: “stats” package (SW test), “nortest” package (AD, KS, and SF tests), “lawstat” package (rJB test), and “normwhn.test” package (DH test). Note that the p -values computed for the SW test are based on the Royston method (Royston, 1982a,b, 1995), the version of the KS test in the “nortest” package is the Lilliefors, the critical values used to compute the rJB are obtained by an approximation to a χ^2 distribution and with zero number of Monte Carlo simulations for the empirical critical values, and the p -value registered for the DH test was that labeled “sig.Ep” in the output.



described above were used to provide a gage of the average level of normality achieved by the outlier methods. The simulation approach described above, which uses iterations of simulations and estimation of an average, was used in the study. In addition, the p -value approach described above was used to

determine the PoRs after the outlier methods were applied to the EGds.

Three different sets of EGds were generated. The parameters were those described above: $\mu = 300, \sigma = 20,$ and $\tau = 300$ (EGd1), $\mu = 400, \sigma = 20,$ and $\tau = 200$ (EGd2), and $\mu = 500, \sigma = 20,$ and



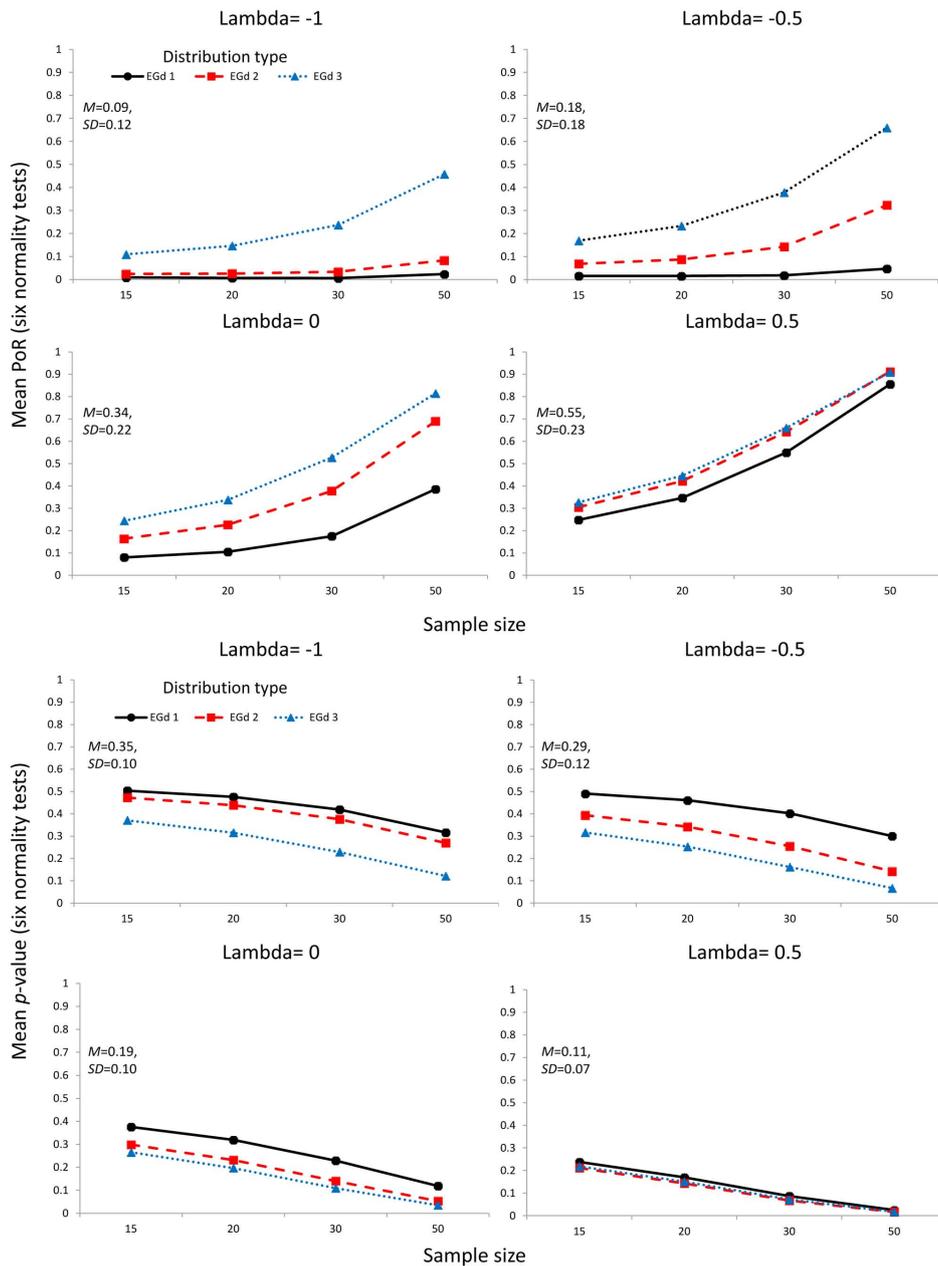
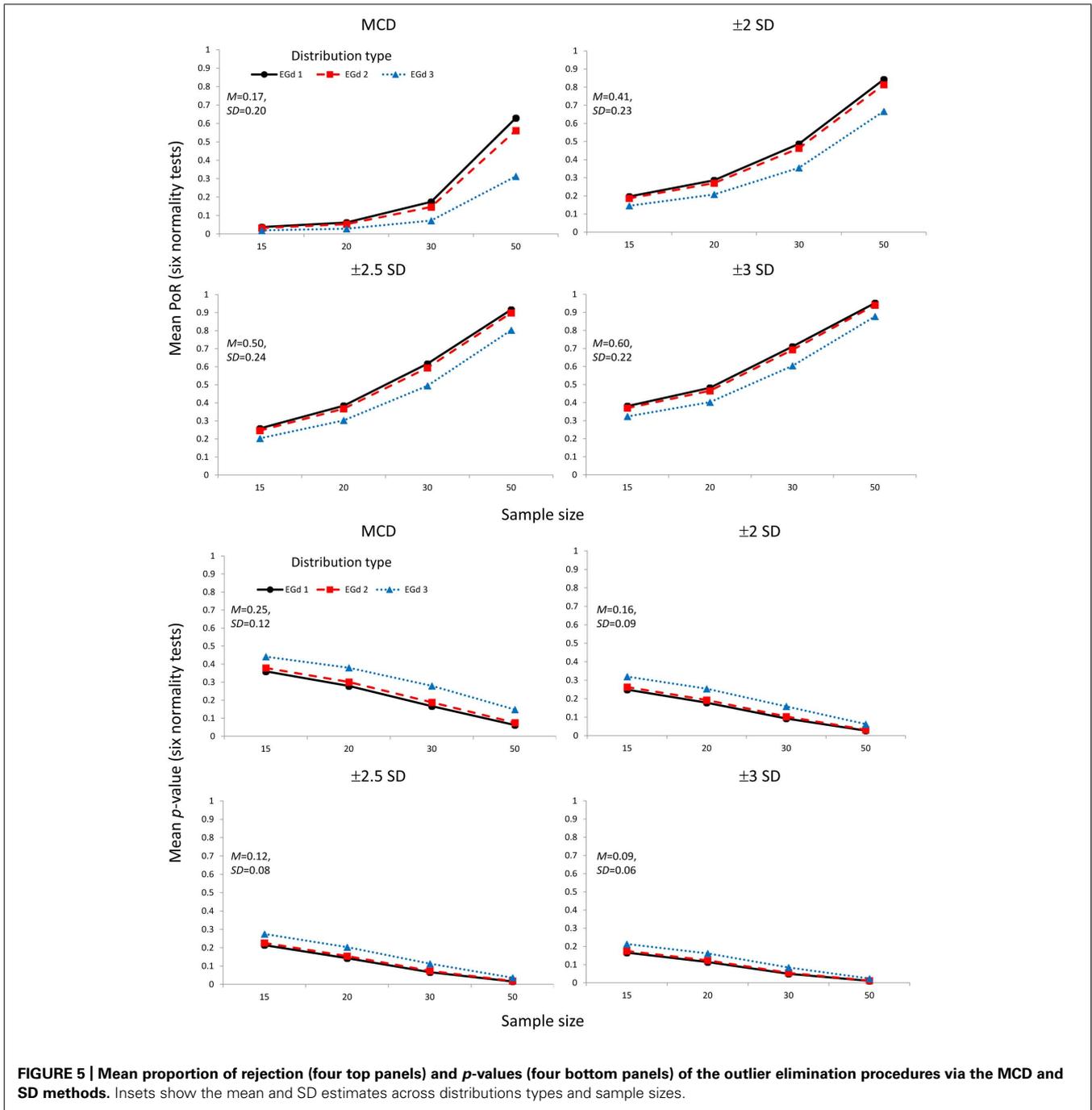


FIGURE 4 | Mean proportion of rejection (four top panels) and *p*-values (four bottom panels) of the outlier accommodation procedure via data transformation. Lambda represents the parameter used to perform the transformation. Insets show the mean and SD estimates across distributions types and sample sizes.

$\tau = 100$ (EGd3). Each EGd was generated in four sample sizes: 15, 20, 30, and 50². These parameters represent actual RT data and are taken from the 12 different EGDs reported originally by Miller (1989). The simulation was carried out using the method

proposed by Marmolejo-Ramos and González-Burgos (2013) and with the following parameters: $i = 30$, $s = 20'000$, and $a =$ the Mean (and its ± 1 SD). The mean *p*-value of the six normality tests described earlier was computed for each *s* product of the combination of EGd type, sample size, and outlier method. Across simulations, the PoR was the proportion of times the average *p*-values fell below, and were not equal to, an alpha level of 0.05. Finally, the mean PoR and *p*-value were estimated for each vector containing *i*.

²A pilot simulation showed that EGds of sample size 10 were in some cases trimmed down to seven when the elimination methods were used. Because such a low sample size affected the results of the AD and the DH tests, the sample size 10 was excluded from the final simulation study.



Thus, for each combination of EGd, sample size, and outlier method, i number of PoR and p -value results were available. The mean results of the PoRs and p -values were submitted to a 3 (types of EGd = EGd1, EGd2, and EGd3) \times 4 (sample sizes = 15, 20, 30, and 50) \times 2 (outlier methods = transformation and elimination) ANOVA-type statistic (ATS; see Noguchi et al., 2012 for details) in order to determine main effects and interactions. The “type of EGds” was entered in the analysis as the between-subjects factor, while the other factors were entered as the within-subjects factors. The items for the outlier transformation method were the

four transformation parameters of the Box–Cox transformations described above, i.e., -1 , -0.5 , 0 , and 0.5 and the items for the outlier elimination method were the four methods discussed above, i.e., the MCD, ± 2 SD, ± 2.5 SD, and ± 3 SD methods. Comparisons of two dependent groups were performed via the Yuen test (T_y ; see Wilcox, 2012 for details).

In the particular case of the outlier elimination methods, the proportion of data eliminated (PoE) was estimated in the same way as the PoRs. That is, for each distribution to which an outlier elimination method was applied, the proportion of observations

Table 1 | Results of the ANOVA-type statistic (ATS).

DV	Main effect	Interaction
PoR	$F_D(1.94, 82.53) = 4165.33, p < 0.001$ $F_O(1, \infty) = 31857.81, p < 0.001$ $F_S(2.54, \infty) = 78415.56, p < 0.001$	$F_{D \times O}(1.96, \infty) = 22335.40, p < 0.001$ $F_{S \times O}(2.72, \infty) = 883.00, p < 0.001$ $F_{D \times S}(4.90, \infty) = 525.59, p < 0.001$ $F_{D \times O \times S}(5.17, \infty) = 26756, p < 0.001$
pV	$F_D(1.93, 81.51) = 1681.45, p < 0.001$ $F_O(1, \infty) = 74198.60, p < 0.001$ $F_S(2.82, \infty) = 82610.79, p < 0.001$	$F_{D \times O}(1.96, \infty) = 34226.12, p < 0.001$ $F_{S \times O}(2.85, \infty) = 502.25, p < 0.001$ $F_{D \times S}(5.50, \infty) = 498.73, p < 0.001$ $F_{D \times O \times S}(5.57, \infty) = 255.15, p < 0.001$
PoE	$F_D(1.96, 83.66) = 4382.35, p < 0.001$ $F_E(2.66, \infty) = 267304.12, p < 0.001$ $F_S(2.82, \infty) = 186.08, p < 0.001$	$F_{D \times E}(5.02, \infty) = 165.09, p < 0.001$ $F_{S \times E}(6.15, \infty) = 4529.77, p < 0.001$ $F_{D \times S}(5.42, \infty) = 12.11, p < 0.001$ $F_{D \times E \times S}(11.18, \infty) = 35.62, p < 0.001$

DV, dependent variable; PoR, proportion of rejection data; pV, p-value data; PoE, proportion of elimination data; D, distribution type (EGd1, EGd2, and EGd3); O, outlier method (elimination and transformation); S, sample size (15, 20, 30, and 50), and E, elimination method (MCD, ± 2 SD, ± 2.5 SD, and ± 3 SD).

removed by a specific method was computed. Then, an average of PoE was estimated for the total number of simulations and each simulation run was iterated i times. Finally, the mean PoEs across iterations were computed.

Also, for both outlier methods, the mean p -value is reported in order to render more obvious the direct relationship between PoR and p -values. That is, the higher the mean p -value, the lower the PoR, and the lower the mean p -value, the higher the PoR. That is, more chances of normality rejection are paired with mean p -values, signaling non-normality. **Figure 3** illustrates the key features of the simulation study.

A NOTE ON THE CHARACTERISTICS OF THE PRESENT SIMULATIONS

All throughout this article the simulation method used in the present study has been depicted so it is worthwhile emphasizing the value of the simulation approach used herein. Canonical simulation studies on normality test report tables of a unique number that represents the power of the test under study, i.e., the proportion of times the test rejected normality (here PoR) at the alpha level chosen. For instance, in an ideal simulation study in which 20'000 simulations are run, it is expected that in 1'000 of them the assumption of normality is incorrectly rejected when tested against a $N(0,1)$, which in turn, gives a power of 0.05 (or a PoR of 0.05). However, if such simulation is run a second, third, or an x number of times, it is likely that the number of $N(0,1)$ distributions flagged as non-normal, varies from 1000 every time. Therefore, giving a PoR of approximately 0.05. Such variation in the outcome can be due to several factors such as the type of RNG used, the use of seeding in the simulations, the statistical package used, and/or simply chance.

There is in fact another issue associated with the study of normality tests that can play a part in the simulation process. When a normality test is used against a $N(0,1)$, a distribution of x number of observations, i.e., the number of simulations, containing the results of the test statistic is formed. CVs are then obtained by calculating the key quantiles of the test statistic's distribution

(e.g., the 95% quantile in positively skewed distributions when alpha is 0.05). However, the CVs found are directly dependent on the computation used to estimate the quantiles and there are various computations involved [for instance, the R software implements nine different quantile computations (see Hyndman and Fan, 1996)].

RESULTS

PROPORTION OF REJECTION AND P-VALUES

The mean PoR and mean p -values corresponding to the transformation of outliers via the Box–Cox transformation parameters are shown in **Figure 4**. **Figure 5** shows the mean PoR and mean p -values for the case of elimination of outliers via the MCD and the $\pm n$ SD methods.

The ATS analyses suggested significant main effects of distribution type; sample size and outlier method and their two and three way interactions in both the PoR and p -value analyses (see **Table 1**). The effect sizes of the three way interactions are shown in **Figure 6**.

As shown in the first row in **Figure 6**, the likelihood of the rejection of normality increases as sample size (i.e., from 15 to 50) and distribution skewness increase (i.e., from EGd3 to EGd1). This result is corroborated by the p -values analyses in that as sample size and distribution skewness increase, the p -values decrease (second row in **Figure 6**). This is a phenomenon recognized in research on the power of normality tests (see Marmolejo-Ramos and González-Burgos, 2013) and is replicated here by the main effects of sample size and distribution type (see **Table 1**). An interesting result from the relative treatment effects plots, however, is that while the likelihood of rejection of normality increases from EGd3 to EGd1 in the case of elimination methods, an opposite pattern occurs to the transformation methods. This result indicates that transformation methods have greater normalization power than elimination methods as the distribution becomes more skewed. The relative treatment effects plots for the p -value data corroborate this by showing that transformation methods lead

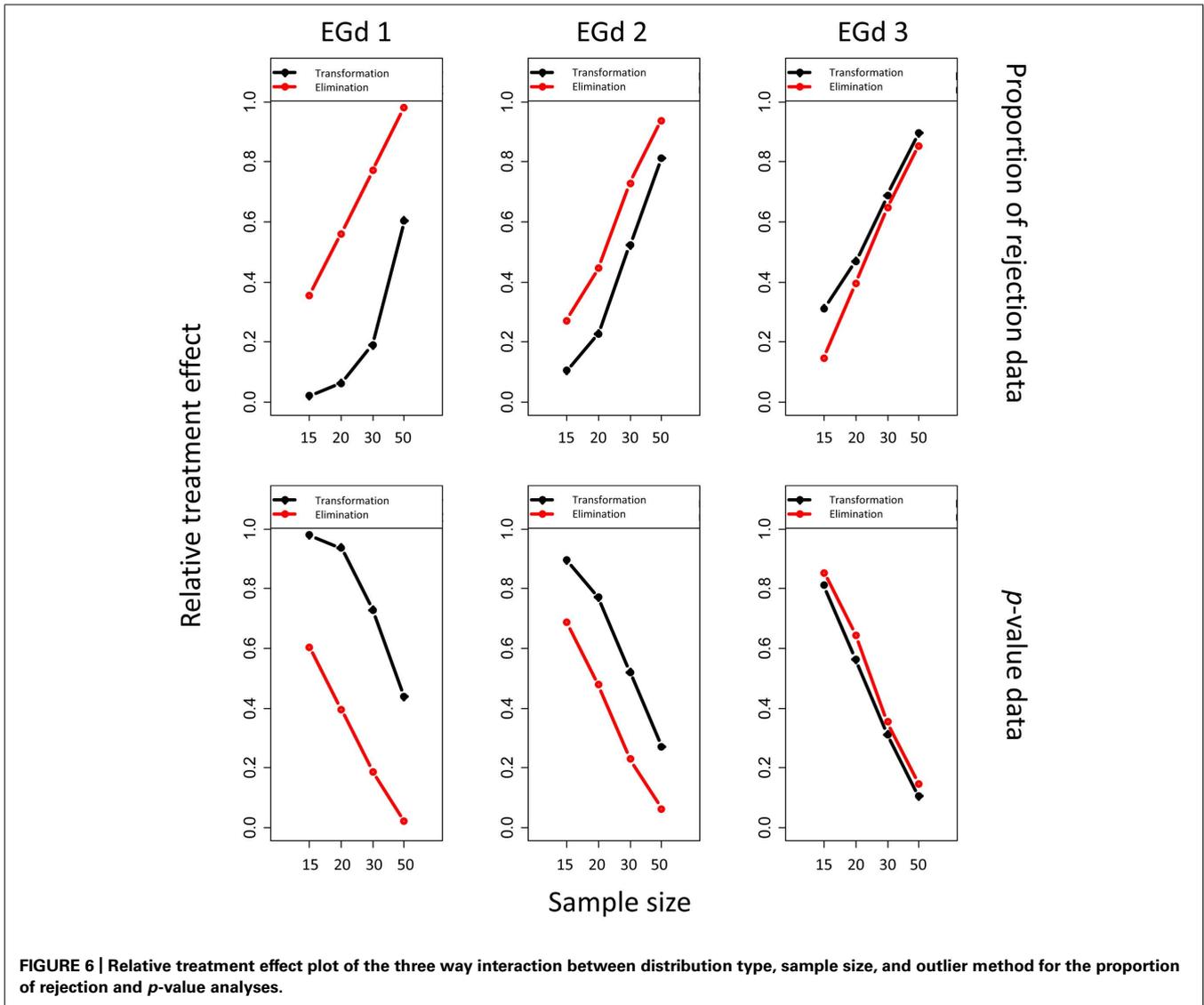


FIGURE 6 | Relative treatment effect plot of the three way interaction between distribution type, sample size, and outlier method for the proportion of rejection and *p*-value analyses.

to higher *p*-values than elimination methods as the distribution becomes more skewed.

As **Figures 4 and 6** indicate, the transformation methods seem to lead to decreased normality rejection as compared with elimination methods. The larger effects of the former methods over the latter are summarized in **Figure 6**. As shown in **Figure 4**, the transformation methods seem to be particularly useful when dealing with highly skewed distributions (i.e., EGd1) in that, across sample sizes, low PoRs, and high *p*-values were obtained for these distributions after transformation. Specifically, the results indicate that transformations with $\lambda = -1$ would seem to provide the best results (see insets in **Figure 4**). These results are in agreement with past research suggesting that the inverse transformation has a strong normalization effect (see Ratcliff, 1993).

PROPORTION OF ELIMINATION

The mean PoE corresponding to the elimination of outliers via the MCD and $\pm n$ SD methods is shown in **Figure 7**.

The ATS analyses suggested significant effects of distribution type; sample size, and outlier elimination method and their two and three way interactions in the PoE analyses (see **Table 1**). The effect sizes of the three way interactions are shown in **Figure 8**.

Figure 5 reports the mean proportion of rejection achieved by each method for each distribution in different sample sizes and their associated *p*-values. The results suggest that the MCD method seems to lead to lower PoR and higher *p*-values than the SD methods. However, by comparing **Figures 5 and 7**, a trade-off between the PoRs (and associated *p*-values) and the POE associated with each of these methods becomes clear. Thus, while the MCD method leads to the lowest PoRs, it does have the highest POE. On the contrary, the ± 3 SD method has low PoE but at the cost of leading to a rather high proportion of normality rejection.

As the relative treatment effect plot in **Figure 8** indicates, for all methods, the likelihood of eliminating more data increases as the distribution becomes more skewed; i.e., from EGd 3 to EGd

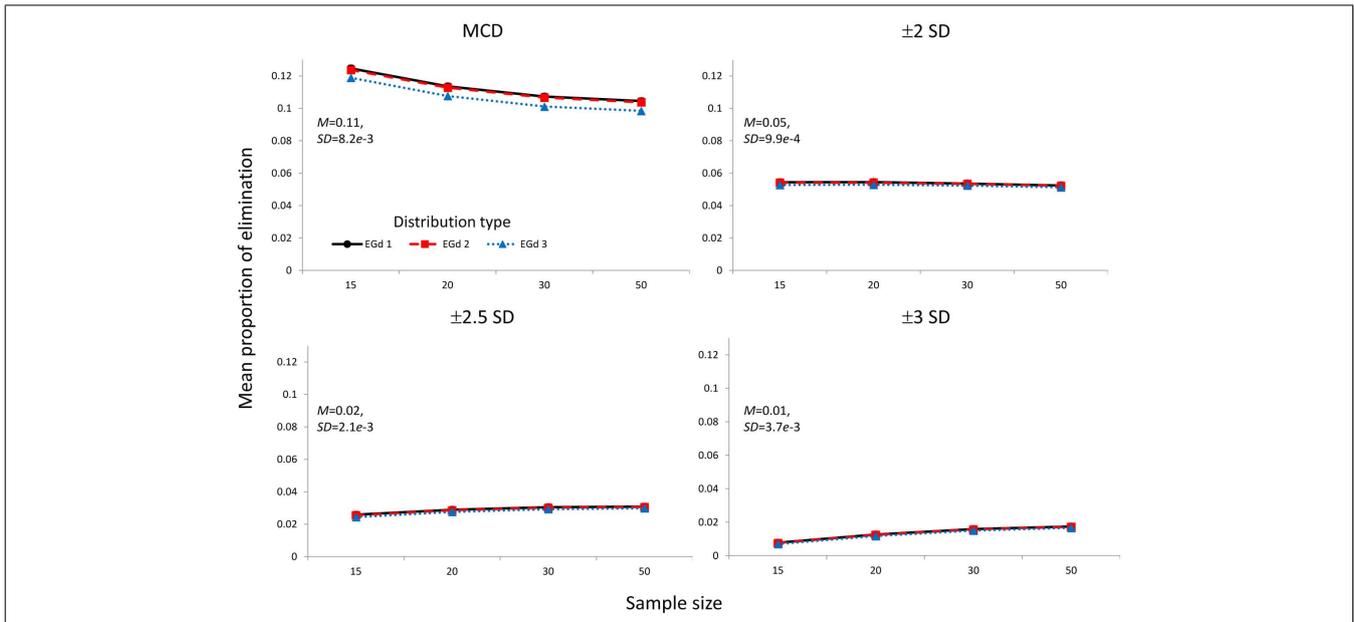


FIGURE 7 | Mean proportion of elimination of the outlier elimination procedures via the MCD and SD methods. Insets show the mean and SD estimates across distributions types and sample sizes.

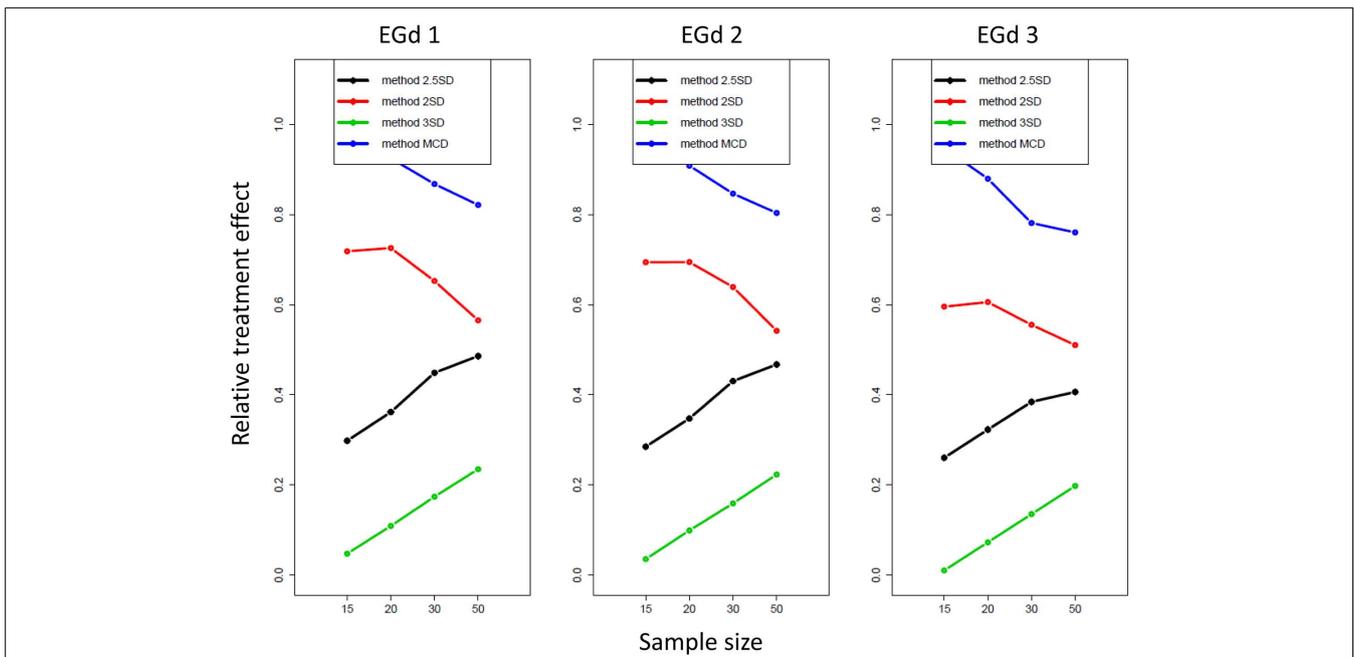


FIGURE 8 | Relative treatment effect plot of the three way interaction between distribution type, sample size, and outlier method for the proportion of elimination analyses.

1. However, while the MCD and ± 2 SD's likelihood of rejecting data appears to reduce as sample size increases, for the remaining methods such a likelihood increases as sample size increases.

In summary, the key result is that the transformation methods are more effective than the elimination methods at normalizing positively skewed distributions. That is, the outlier method had

a main effect. Thus, indicating that across distributions types and sample sizes the transformation methods led to lower PoR ($M_{PoR} = 0.29$, $SD = 0.17$) and higher p -values ($M_{p-value} = 0.24$, $SD = 0.09$) than the elimination methods ($M_{PoR} = 0.42$, $SD = 0.22$; $M_{p-value} = 0.16$, $SD = 0.09$) [PoR: $T_y(359) = -14.54$, $p < 0.001$; p -value: $T_y(359) = 20.26$, $p < 0.001$].

DISCUSSION

The results of this simulation study suggest that the Box–Cox transformation methods outperform the elimination methods in normalizing positively skewed data and the more skewed the distribution, the more effective the transformation methods in normalizing such type of data. Various ideas need to be discussed in relation to this finding.

The difference between transformations and elimination procedures is that transformations seek to stabilize variance and skewness (see Bartlett, 1947) whereas the other procedures are devised to eliminate extreme observations; as a by-product, both methods help in improving normality. However, these methods ultimately aim to determine the best estimator of central tendency. One could argue that these procedures simply distort the original data sets in order to render them suitable for a parametric test. Although there are arguments in favor of using parametric tests regardless (see Schmider et al., 2010), others advocate the use of other statistical methods (e.g., Wilcox, 2012). For instance, Lachaud and Renaud (2011) indicate that either the data should be filtered (e.g., via the $\pm n$ SD approach) before analysis using general linear modeling (e.g., ANOVA, quasi-*F*, and multilevel modeling) or analyzed using robust methods (e.g., ATS, bootstrap, and permutation methods). These authors also give useful recommendations as to how to analyze data when using general linear modeling approaches. Hence, a combination of the methods studied here with robust techniques could also be productive (Rashid et al., 2013, term this approach ‘side-by-side analyses’). Some researchers have taken these methods further. For instance, Ulrich and Maienborn (2010) took the mean RT of correct trials for each subject in each condition and compared the results with those obtained when the median RT and the 10% trimmed mean of correct trials for each subject in each condition were taken. That is, the researchers compared the results of analysis using a 0 (arithmetic mean), 10, and 50% (median) trimmed means. Finally, they performed analyses on the means of the trimmed means. Other researchers opt for taking the median RT of correct trials for each subject in each condition and perform analyses on the means of those medians (see for example Ansoorge et al., 2010; Rein and Markman, 2010). Trimmed means, and other robust estimators of central tendency (e.g., Rosenberg and Gasko, 1983; Wilcox and Keselman, 2003; Bickel and Frühwirth, 2006; Vélez and Correa, 2014), can therefore be seen as non-invasive forms of data elimination in that outlying observations are temporarily canceled out in order to estimate an average.

Applying the methods studied herein to data believed to be non-normal, does not automatically guarantee that the data has met parametric assumptions. That is, it is important to corroborate, via graphical and formal tests, that these assumptions have been reached. Although this is a well-known recommendation, it is rare to find published papers reporting normality or homogeneity tests in order to justify the use of parametric analyses. It is therefore important that whatever method is used to filter data, formal normality, and homogeneity tests are reported in order to substantiate the use of parametric tests.

METHODOLOGICAL CONSIDERATIONS AND FUTURE STUDIES

Every research study has room for improvement and this study is no exception. Admittedly, the estimation of PoR and *p*-values used here is rather liberal and may have some degree of Type I error attached to it. Thus, a replication study could estimate CVs for each normality test employed via quantiles as is traditionally done (although see section 2.2) and *p*-values could be combined via conservative methods such as the Stouffer method (see Vélez et al., under review). Also, other normality tests that are particularly robust to the distributions being studied could be considered. Equally important, other distributions that are a good fit for real data should be included in the simulation study. For instance, in the case of RT, data distributions such as Weibull and Log-Normal need to be studied. Another type of data commonly encountered in experimental research but less studied, is that of discrete *n*-point Likert-type data. Distributions that fit this type of data could be studied in the context of outliers and normality research as well. Yet, the studies suggested here should be preceded by research demonstrating how well various potential candidate distributions fit RT and Likert-type data (e.g., via AIC measures) and showing which distributions seem to give the best fits in both real and simulated RT and Likert-type data. Indeed, there should be research aimed at grounding the parameters of distributions fitting RT and Likert-type data into psychological processes of interest (e.g., McAuley et al., 2006 explained the parameters of the EGd in terms of cognitive processes tapped via RTs). To the best of our knowledge such research is yet to be done.

Although some of the most commonly employed outlier elimination and transformation methods were addressed herein, other methods should also be studied. For instance, data truncation and outlier replacement are procedures also found in papers reporting experimental results in cognitive science (an example of data truncation can be found in Bub and Masson, 2010; an example of outlier replacement can be seen in Pylyshyn and Storm, 1988). The performance of newer methods such as the Ueda (1996/2009), van der Loo (2010), and ± 2.5 MAD (Leys et al., 2013) should be studied in the context of RT data.

Finally, it can be contended that in principle, the procedures studied here should not only improve data’s normality but also their homogeneity. Thus, future studies should test the effects of the procedures studied here on the homogeneity of two or more batches of data. Canonical tests such as the Levene and the Brown-Forsythe test and recent robust versions of them (e.g., De Almeida et al., 2008) should be used to verify this claim.

CONCLUSION

This paper sets out to offer an educated consensus on the recommended approach in cases where data need to be treated in order to submit to a parametric test. The results indicate that methods that transform the data in order to accommodate outliers lead to higher chances of normalization than methods that eliminate data points. Although some of the most commonly used elimination and transformation methods were studied herein, further methods need to be considered. Other distributions that can be used to model reaction time and Likert-type data should also be addressed in future studies.

AUTHOR CONTRIBUTIONS

Fernando Marmolejo-Ramos thanks Delphine Courvoisier and Firat Özdemir for their help with earlier versions of this project. Fernando Marmolejo-Ramos also thanks Kimihiro Noguchi for his help with his 'nparLD' R package and Xavier Romão for facilitating access to HPC facilities. Finally, the authors thank Robyn Groves and Rosie Gronthos for proreading this manuscript.

REFERENCES

- Akbiğiç, O., and Howe, A. J. (2011). A novel normality test using an identity transformation of the Gaussian function. *Eur. J. Pure Appl. Math.* 4, 448–454.
- Alizadeh Noughabi, H., and Arghami, N. R. (2011). Monte Carlo comparison of seven normality tests. *J. Stat. Comput. Simul.* 81, 965–972. doi: 10.1080/00949650903580047
- Ansorge, U., Kiefer, M., Khalid, S., Grassl, S., and König, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition* 116, 303–320. doi: 10.1016/j.cognition.2010.05.010
- Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28.
- Bartlett, M. S. (1947). The use of transformations. *Biometrics* 3, 39–52. doi: 10.2307/3001536
- Beckman, R. J., and Cook, R. D. (1983). Outliers. *Technometrics* 25, 119–149.
- Bertels, J., Kollinsky, R., and Morais, J. (2010). Emotional valence of spoken words influences the spatial orienting of attention. *Acta Psychol.* 134, 264–278. doi: 10.1016/j.actpsy.2010.02.008
- Bickel, D. R., and Frühwirth, R. (2006). On a fast, robust estimator of the mode: comparisons to other robust estimators with applications. *Comput. Stat. Data Anal.* 50, 3500–3530. doi: 10.1016/j.csda.2005.07.011
- Bland, J. M., and Altman, D. G. (1996). Transforming data. *Br. Med. J.* 312:770. doi: 10.1136/bmj.312.7033.770
- Bub, D. N., and Masson, M. E. J. (2010). Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 341–358. doi: 10.1037/a0017606
- Calkins, D. (1974). Some effects of non-normal distribution shape on the magnitude of the Pearson product moment correlation coefficient. *Interam. J. Psychol.* 8, 261–287.
- Cousineau, D. (2004). Merging race models and adaptive networks: a parallel race network. *Psychon. Bull. Rev.* 11, 807–825. doi: 10.3758/BF03196707
- Cousineau, D., and Chartier, S. (2010). Outlier detection and treatment: a review. *Int. J. Psychol. Res.* 3, 58–67.
- Cowles, M., and Davis, C. (1982). On the origins of the 0.05 level of statistical significance. *Am. Psychol.* 37, 553–558. doi: 10.1037/0003-066X.37.5.553
- De Almeida, A., Elian, S., and Nobre, J. S. (2008). Modificações e alternativas aos testes de Levene e de Brown e Forsythe para igualdade de variâncias e médias. *Rev. Colomb. Estad.* 31, 241–260.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychol.* 30, 412–431. (Original published in 1869).
- Engmann, S., and Cousineau, D. (2011). Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnov test. *J. Appl. Quant. Methods* 6, 1–17.
- Harri, A., and Coble, K. H. (2011). Normality testing: two new tests using L-moments. *J. Appl. Stat.* 38, 1369–1379. doi: 10.1080/02664763.2010.498508
- Havas, D. A., Glenberg, A. M., and Rinck, M. (2007). Emotion simulation during language comprehension. *Psychon. Bull. Rev.* 14, 436–441. doi: 10.3758/BF03194085
- He, D., and Xu, X. (2013). A goodness-of-fit testing approach for normality based on the posterior predictive distribution. *Test* 22, 1–18. doi: 10.1007/s11749-012-0282-6
- Heathcote, A., Popiel, S. J., and Mewhort, D. J. K. (1991). Analysis of response time distributions: an example using the Stroop task. *Psychol. Bull.* 109, 340–347. doi: 10.1037/0033-2909.109.2.340
- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 598–615. doi: 10.1037/0278-7393.10.4.598
- Hohle, R. (1965). Inferred components of reaction times as functions of foreperiod duration. *J. Exp. Psychol.* 69, 382–386. doi: 10.1037/h0021740
- Hopkins, G. W., and Kristofferson, A. B. (1980). Ultrastable stimulus-reponse latencies: acquisition and stimulus control. *Percept. Psychophys.* 27, 241–250. doi: 10.3758/BF03204261
- Hyndman, R. J., and Fan, Y. (1996). Sample quantiles in statistical packages. *Am. Stat.* 50, 361–365.
- Judd, C. M., McClelland, G. H., and Culhane, S. E. (1995). Data analysis: continuing issues in the everyday analysis of psychological data. *Annu. Rev. Psychol.* 46, 433–465. doi: 10.1146/annurev.ps.46.020195.002245
- LaBerge, D. A. (1962). A recruitment theory of simple behavior. *Psychometrika* 27, 375–396. doi: 10.1007/BF02289645
- Lachaud, C. M., and Renaud, O. (2011). A tutorial for analysing human reaction times: how to filter data, manage missing values, and choose a statistical model. *Appl. Psychol.* 32, 389–416. doi: 10.1017/S0142716410000457
- Lance, C. E., Stewart, A. M., and Carretta, T. R. (1996). On the treatment of outliers in cognitive and psychomotor test data. *Mil. Psychol.* 8, 43–58. doi: 10.1207/s15327876mp0801_4
- Langlois, D., Cousineau, D., and Thivierge, J.-P. (2014). Maximum likelihood estimators for truncated and censored power-law distributions show how neuronal avalanches may be misvaluated. *Phys. Rev. E Stat Nonlin. Soft Matter Phys.* 89, 12709. doi: 10.1103/PhysRevE.89.012709
- Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013
- Markman, A. B., and Brendl, M. (2005). Constraining theories of embodied cognition. *Psychol. Sci.* 16, 6–10. doi: 10.1111/j.0956-7976.2005.00772.x
- Marmolejo-Ramos, F., and González-Burgos, J. (2013). A power comparison of various tests of univariate normality on Ex-Gaussian distributions. *Methodology* 9, 137–149.
- Marmolejo-Ramos, F., and Matsunaga, M. (2009). Getting the most from your curves: exploring and reporting data using informative graphical techniques. *Tutor. Quant. Methods Psychol.* 5, 40–50.
- McAuley, T., Yap, M., Christ, S. E., and White, D. A. (2006). Revisiting inhibitory control across the life span: insights from the ex-Gaussian distribution. *Dev. Neuropsychol.* 29, 447–458. doi: 10.1207/s15326942dn2903_4
- McClelland, J. L. (1979). On the time relations of mental processes: a framework for analyzing processes in cascade. *Psychol. Rev.* 86, 287–330. doi: 10.1037/0033-295X.86.4.287
- McGill, W. J. (1963). “Stochastic latency mechanisms,” in *Handbook of mathematical psychology*, eds R. D. Luce, R. R. Busch, and E. Galanter (New York: John Wiley and Sons), 309–360.
- Miller, J. (1989). A warning about median reaction time. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 539–543. doi: 10.1037/0096-1523.14.3.539
- Miller, J. O., and Ulrich, R. (2003). Simple reaction time and statistical facilitation: a parallel grains model. *Cogn. Psychol.* 46, 101–115. doi: 10.1016/S0010-0285(02)00517-0
- Moran, D. W., and Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *J. Neurophysiol.* 82, 2676–2692.
- Moss, H. E., McCormick, S., and Tyler, L. K. (1997). The time course of activation of semantic information during spoken word recognition. *Lang. Cogn. Process.* 12, 695–731. doi: 10.1080/016909697386664
- Mouri, H. (2013). Log-normal distribution from a process that is not multiplicative but is additive. *Phys. Rev. E* 88:042124. doi: 10.1103/PhysRevE.88.042124
- Noguchi, K., Gel, Y. R., Brunner, E., and Konietzschke, F. (2012). nparLD: an R software package for the nonparametric analysis of longitudinal data in factorial experiments. *J. Stat. Softw.* 50, 1–23.
- Olivier, J., and Norberg, M. M. (2010). Positively skewed data: revisiting the Box-Cox power transformation. *Int. J. Psychol. Res.* 3, 68–75.
- Orr, J. M., Sackett, P. R., and DuBois, C. L. Z. (1991). Outlier detection and treatment in I/O psychology: a survey of researcher beliefs and an empirical illustration. *Pers. Psychol.* 44, 473–486. doi: 10.1111/j.1744-6570.1991.tb02401.x
- Osborne, J. (2002). *Notes on the Use of Data Transformation. Practical Assessment, Research and Evaluation*. Available at: <http://pareonline.net/getvn.asp?v=8&n=6> [accessed May 18, 2009].
- Otte, E., Habel, U., Schulte-Rüther, M., Konrad, K., and Koch, I. (2011). Interference in simultaneously perceiving and producing facial expressions – evidence from electromyography. *Neuropsychologia* 49, 124–130. doi: 10.1016/j.neuropsychologia.2010.11.005

- Perea, M. (1999). Tiempos de reacción y psicología cognitiva: dos procedimientos para evitar el sesgo debido al tamaño muestral. *Psicológica* 20, 13–21.
- Pike, R. (1973). Response latency models for signal detection. *Psychol. Rev.* 80, 53–68. doi: 10.1037/h0033871
- Pylyshyn, Z. W., and Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197. doi: 10.1163/156856888X00122
- Raab, D. H. (1962). Effects of stimulus-duration on auditory reaction-time. *J. Am. Psychol.* 75, 298–301. doi: 10.2307/1419616
- Rashid, M. M., McKean, J. W., and Kloke, J. D. (2013). Review of rank-based procedures for multicenter clinical trials. *J. Biopharm. Stat.* 23, 1207–1227. doi: 10.1080/10543406.2013.834919
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510
- Ratcliff, R., and Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychol. Rev.* 86, 190–214. doi: 10.1037/0033-295X.83.3.190
- Razali, N. M., and Wah, Y. B. (2010). “Power comparison of some selected normality tests,” in *Proceedings of the Regional Conference on Statistical Sciences, (RCSS’10)*, Kota Bharu, 126–138.
- Rein, J. R., and Markman, A. B. (2010). Assessing the concreteness of relational representation. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1452–1465. doi: 10.1037/a0021040
- Romão, X., Delgado, R., and Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *J. Stat. Comput. Simul.* 80, 545–591. doi: 10.1080/00949650902740824
- Rosenberg, J. L., and Gasko, M. (1983). “Comparing location estimators: trimmed means, medians, and trimean,” in *Understanding Robust and Exploratory Data Analysis*, eds D. Hoaglin, F. Mosteller, and J. Tukey (New York, NY: Wiley), 297–336.
- Rouder, J. N. (2000). Assessing the roles of change discrimination and luminance integration: evidence for hybrid race model of perceptual decision making in luminance discrimination. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 359–368. doi: 10.1037/0096-1523.26.1.359
- Rousseeuw, P. J., and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223. doi: 10.1080/00401706.1999.10485670
- Royston, P. (1982a). Algorithm AS 181: the W test for normality. *Appl. Stat.* 31, 176–180. doi: 10.2307/2347986
- Royston, P. (1982b). An extension of Shapiro and Wilk’s W test for normality to large samples. *Appl. Stat.* 31, 115–124. doi: 10.2307/2347973
- Royston, P. (1995). Remark AS R94: a remark on algorithm AS 181: the W test for normality. *Appl. Stat.* 44, 547–551. doi: 10.2307/2986146
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., and Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology* 6, 147–151.
- Schwarz, W. (2001). The ex-wald distribution as a descriptive model of response times. *Behav. Res. Methods Instrum. Comput.* 33, 457–469. doi: 10.3758/BF03195403
- Thompson, G. L. (2006). An SPSS implementation of the non-recursive outlier deletion procedure with shifting z-score criterion (Van Selst and Jolicoeur, 1994). *Behav. Res. Methods* 38, 344–352. doi: 10.3758/BRM.38.2.344
- Ueda, T. (2009). A simple method for the detection of outliers (trans. F. Marmolejo-Ramos and S. Kinoshita). *Electron. J. Appl. Stat. Anal.* 2, 67–76. (Original work published in 1996).
- Ulrich, R., and Maienborn, C. (2010). Left-right coding of past and future in language: the mental timeline during sentence processing. *Cognition* 117, 126–138. doi: 10.1016/j.cognition.2010.08.001
- Ulrich, R., and Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *J. Math. Psychol.* 37, 513–525. doi: 10.1006/jmps.1993.1032
- van der Loo, M. P. J. (2010). *Distribution Based Outlier Detection for Univariate Data*. The Hague: Statistics Netherlands.
- Van Selst, M., and Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Q. J. Exp. Psychol.* 47A, 631–650. doi: 10.1080/14640749408401131
- Vélez, J. I., and Correa, J. C. (2014). Should we think of a different median estimator? *Comun. Estadística* 7, 11–17.
- West, J., and Shlesinger, M. (1990). The noise in natural phenomena. *Am. Sci.* 78, 40–45.
- Whelan, R. (2008). Effective analysis of reaction time data. *Psychol. Rec.* 58, 475–482.
- Wilcox, R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. Amsterdam: Elsevier.
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *Am. Psychol.* 53, 300–314. doi: 10.1037/0003-066X.53.3.300
- Wilcox, R. R., and Keselman, H. J. (2003). Modern robust data analysis methods: measures of central tendency. *Psychol. Methods* 8, 254–274. doi: 10.1037/1082-989X.8.3.254
- Yap, B. W., and Sim, C. H. (2011). Comparison of various types of normality tests. *J. Stat. Comput. Simul.* 81, 2141–2155. doi: 10.1080/00949655.2010.520163

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 July 2014; accepted: 14 December 2014; published online: 07 January 2015.

Citation: Marmolejo-Ramos F, Cousineau D, Benites L and Maehara R (2015) On the efficacy of procedures to normalize Ex-Gaussian distributions. *Front. Psychol.* 5:1548. doi: 10.3389/fpsyg.2014.01548

This article was submitted to *Quantitative Psychology and Measurement*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Marmolejo-Ramos, Cousineau, Benites and Maehara. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.