

# Robust measures of central tendency: weighting as a possible alternative to trimming in response-time data analysis

Yury S. Dodonov\*, Yulia A. Dodonova

*Moscow City University of Psychology and Education, Moscow, Russia*

Received 30 June 2011; available online 23 October 2011

Suggested citation:

Dodonov, Y. S., & Dodonova, Y. A. (2011). Robust measures of central tendency: weighting as a possible alternative to trimming in response-time data analysis. *Psikhologicheskie Issledovaniya*, 5(19). Retrieved from <http://psystudy.ru>.

## ABSTRACT

In this paper, a problem with the robustness of measures of central tendency is analyzed in the context of studies on the speed of information processing. Within these studies, the raw data of each participant always includes a set of response times, and a single estimate of the location of individual response-time distribution has to be calculated as a measure of individual performance in a speed task. This paper consists of three sections. In the first section, the measures of central tendency are reviewed. Special attention is paid to the measures that are based on trimming, as they are commonly recommended in modern statistics as the most robust to distribution skewness and the presence of outliers. Another possible approach to obtaining a robust measure of central tendency, which is based on data weighting, is discussed; two weighted estimators of central tendency are introduced. The second section, based on the results of an empirical study and a computer simulation, demonstrates how the choice of which measure of central tendency is used for calculating individual response time in a speed task can affect conclusions on the significance of the association between response time and an external variable. In the third section, the measures of central tendency are compared in a computer simulation study, which mimics empirical response-time data. The R code for computing the measures of central tendency that are under consideration is presented in the appendix.

*Keywords:* measures of central tendency, trimmed mean, weighted mean, processing speed, response-time distribution

Every textbook on statistical analysis emphasizes that each method can be used if, and only if, certain requirements are met. The requirement for normality of data distribution is probably one of the most frequent in commonly performed analyses. Importantly, normality is not only required by those methods that involve analysis of associations between variables, such as regression, dispersion or factor analyses. Even when a single variable is analyzed, and a 'simple' estimate of location of a distribution is obtained, the accuracy of this estimate depends on the shape of distribution of empirical data, and the choice of measure of central tendency requires certain assumptions about the distribution to be met.

At the same time, an analysis of location of a distribution of empirical data is not rare for psychological studies. For instance, means of some variables in two groups can be compared, or the effect of some influence, for example educational, can be evaluated. However, in these cases an exact estimation of location of empirical distributions can be avoided, for example by employing non-parametric criteria, which do not require normality of data. Nevertheless, there are certain areas of psychological studies where an exact estimate of central tendency must be calculated and used in further analysis. For example, a problem in estimation of location of distributions is

especially urgent for studies on individual differences in the speed of information processing. Speed cognitive tasks that are used in these studies include a number of similar trials. Therefore, raw data for each participant include a set of registered response times. However, just a single index that represents participant's 'average' response time must be calculated for further analysis.

Although this task seems to be simple, it is not that trivial. First, individual distributions of response times have been shown to be positively skewed in many studies (Heathcote, Popiel, Mewhort, 1991; Hockley, 1984; Ratcliff, 1978, 1979, 1993; Ulrich, Miller, 1994). Different measures of central tendency are not equal for such distributions, as they are for symmetric distributions. Moreover, the skew of response-time distribution is not necessarily similar across participants. Therefore, choice of which measure of central tendency is used for calculating a single index of individual performance in a speed task is not indifferent for further analysis of individual differences.

Second, empirical response-time data are likely to include outliers. When completing a speed task, a participant can give anticipatory correct answers, or can otherwise delay a response for various reasons, including natural fluctuations of attention, random interference, and so on. In the case of both short and long response times of this kind, the results will be identified as outliers in further analysis, only if they are much smaller or larger than most

\* Corresponding author.

E-mail address: [ys.dodonov@gmail.com](mailto:ys.dodonov@gmail.com) (Yury S. Dodonov)

of the reasonable values. A problem with identification of outliers is a subject for special studies (e.g. Barnett, Lewis, 1994; Lovie, 1986) and will be touched upon in this paper only briefly. To this point, it is important to note that, in most cases, a researcher can never judge with confidence whether a particular value is an outlier or comes from the same distribution as the other values. Therefore, the presence of outliers is another issue that must be taken into account when a measure of central tendency is chosen for the analysis of location of individual response-time distributions.

Finally, an analysis of location of individual response-time distributions in studies on individual differences in processing speed is complicated by the fact that typical speed tasks commonly include a relatively small number of trials. Within experimental psychology, when a cognitive process is analyzed, a number of trials are commonly as large as a few hundred or even thousand. Ratcliff's studies are a good example of this kind (Ratcliff, 1978, 1979; Ratcliff, McKoon, 2008). However, in this case, testing each participant takes a few sessions of several hours each, which puts obvious limitations on the possible sample size, and makes this approach unsuitable for studies on individual differences. Therefore, speed tasks that include a relatively small number of trials, varying from 20 to 40, are commonly employed. Thus, a researcher deals with relatively small datasets in the analysis of individual response times, and the shape of the distribution of data can never be identified with confidence. Moreover, exact observations in this small dataset are likely to vary substantially when the same participant performs the same task, even though response times can be drawn from the same distribution. Therefore, a measure of central tendency that is used to calculate individual response times should not only be robust to skewness and the presence of outliers, but should also appropriately represent the location of an underlying distribution when a dataset is small and exact observations can vary.

It is important to emphasize the distinction between these two different sources of non-stability of a measure of central tendency (discussion on this can also be found in (Ratcliff, 1993)). The first possible source of non-stability is the appearance of outliers in empirical data, and changes in their quantity. Non-stability of this kind is mostly considered in modern literature on statistics, and just stability in the presence of outliers is commonly referred to as robustness. Different algorithms to analyze robustness have been developed. They include estimating a so-called 'breakdown point', which is a proportion of outliers that a measure can handle before being distorted; evaluating the influence function, and so on.

At the same time, there is a second possible source of non-stability of a measure of central tendency that has rarely attracted the attention of researchers. Non-stability of this kind originates from possible variations in the exact empirical values that form a dataset, although in fact they come from the same underlying distribution. Indeed, some of the parameters of individual response-time distribution can be expected from a researcher, for example, based on a theoretical model of processing. However, in reality, the researcher does not deal with a distribution function, but with a relatively small set of empirical response times. Even with the same underlying distribution, the exact values in a set could be different. Therefore, the researcher expects that a measure of central tendency will not be sensitive to fluctuations in

exact values and will adequately retrieve location of the underlying distribution. Does stability of this kind always go along with robustness to outliers? Can the robustness of a measure of central tendency be analyzed beyond the problem of its stability? These questions will be especially addressed in this paper.

The paper consists of three sections. In the first section, different measures of central tendency are reviewed. First, classical measures, which are also called Pythagorean means, are listed. Second, estimators of central tendency that are based on data trimming are reviewed in detail. Measures of this group are commonly recommended in contemporary literature as those the most robust to the skewness of a distribution and presence of outliers. Third, another possible approach to obtaining stable measures of central tendency, which is based on data weighting, is discussed; two measures of central tendency that are based on weighting are introduced. In the second section, which makes use of the empirical data and a computer simulation, the choice of which measure of central tendency is used to estimate individual response time in speed tasks is shown to be capable of influencing a conclusion concerning the significance of the relations between the response time and other variables. Finally, possible criteria for choosing a measure of central tendency are discussed in the third section, wherein the measures of central tendency are compared based on computer simulations of data that mimic the response times in a real empirical study. The code for R, which can be used for calculating the measures of central tendency that are discussed in this paper, is shown in Appendix.

## Measures of central tendency

The measures of central tendency that are analyzed in this paper are combined in three groups. For each measure, a brief description, the main formulae and a simple numeric example are presented.

### *Classical measures of central tendency*

The three measures of this group are sometimes called classical Pythagorean means. First, this is an arithmetic mean ( $M$ ), which is defined by:

$$M = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1).$$

The algorithm for calculating arithmetic means is simple and does not cause any difficulty, even for inexperienced researchers. So why are the other measures of central tendency needed? Textbooks on statistics commonly illustrate an answer to this question by using an obvious numeric example. For example, consider a data set containing the following values:

2, 3, 4, 4, 5, 5, 6, 6, 80.

The arithmetic mean in this example is 12.10. Does this value accurately reflect the location of the original distribution? In this example, the arithmetic mean is obviously distorted by a single outlier in the data set. Therefore, the adequacy of the estimate of central tendency that was obtained by simple averaging looks quite questionable.

The second classical measure of central tendency is the harmonic mean (*hM*), which is defined by:

$$hM = \frac{n}{1/x_1 + 1/x_2 + \dots + 1/x_n} = n \cdot \left[ \sum_{i=1}^n \frac{1}{x_i} \right]^{-1} \quad (2).$$

The harmonic mean is obviously most influenced by small values in a data set, while the effect of the values from the right tail of the distribution is much less. In the numeric example that was presented above, the harmonic mean is 4.45. In general, for a positive data set with non-equal values, the harmonic mean is always less than arithmetic mean.

Another classical mean that can be recommended when a distribution of values is positively skewed is geometric mean (*gM*), which is greater than the harmonic mean, but still less than the arithmetic mean. The geometric mean is defined as:

$$gM = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (3);$$

or,

$$gM = \exp \left[ \frac{1}{n} \sum_{i=1}^n \log(x_i) \right] \quad (4).$$

The formula (3) makes the meaning of the geometric mean obvious, but it is less convenient when datasets are large, because of high computational loading. Therefore, the formula (4) is preferable for computations. In our numerical example, the geometric mean is 5.78.

As can be seen, all values from the original dataset are employed in calculating classical measures of central tendency. Another approach to the analysis of location of a distribution, which is commonly recommended in statistics when a distribution is skewed and includes outliers, is based on preliminary trimming, or removing, some values from a dataset.

### Measures of central tendency based on data trimming

Measures of this group require removing some values from an original dataset. The basic idea of this approach is the following: if data are extremely likely to include values lying outside the main data array, removing several values from the tails of a distribution, and further averaging the remaining values, will result in a more stable estimate of central tendency than in the case of analysis of the entire dataset.

The median (*Md*) can be regarded as the first measure of this group. The median is given by the central value, if the number of data points is odd, and the mean of the two central values, if the number of data points is even. In other words, for the ordered data set  $x_1 \leq \dots \leq x_n$ :  $Md = x_{k+1}$ , if  $n = 2k + 1$ ; or,  $Md = (x_k + x_{k+1})/2$ , if  $n = 2k$ . In our numerical example the number of values is even, so the median is calculated by averaging the two central values, which gives a value of 5.00. In general, it's obvious from the definition that the median discards all data points except one or two, and only one or two values affect this measure of central tendency. Therefore, the median is much less sensitive to the presence of outliers in the tails of distribution of the data, than most of the other measures of central tendency. At the same time, the exact value of the median is completely determined by those values that appear to be central in the ordered data set,

which makes it problematic for the analysis of individual response times.

A measure of central tendency that is quite frequently used in studies on processing speed is the mean computed for values lying within two standard deviations. To calculate this measure of central tendency, the simple arithmetic mean and standard deviation are calculated on the entire dataset. After that, values that lie outside the two standard deviations from the original mean are regarded as outliers and are removed from analysis, and the remaining values are averaged. In our numerical example this measure of central tendency is equal to 4.56. Although two standard deviations is probably the most frequent threshold for removing the outlying values, obviously any other number of standard deviations can be used as a threshold within this algorithm. In any case, the values that must be removed are identified in relation to the arithmetic mean computed on the entire original dataset, which in turn can be distorted substantially by outliers, as was discussed above.

A problem of dependence of identification of values as outliers on the arithmetic mean is overcome with two other algorithms, developed for estimating central tendency: the trimmed mean and the Winsorized mean (Wilcox, 2001, 2003; Wilcox, Keselman, 2005). To obtain the *k*% trimmed mean, the *k*% lowest and *k*% highest values in the dataset are removed, and the mean is computed on the remaining data points. In our example of the dataset that consists of the observations 2, 3, 4, 4, 5, 5, 6, 6, 6, 80, the 20% trimmed mean is computed on the following values:

4, 4, 5, 5, 6, 6.

In this example the 20% trimmed mean is 5.00. In computing the Winsorized mean, the *k*% lowest and *k*% highest values in the dataset are not simply removed, but replaced by the lowest and highest untrimmed values, respectively. In our numerical example, the Winsorized mean is computed on the values:

4, 4, 4, 4, 5, 5, 6, 6, 6, 6.

In this example, the trimmed mean and the Winsorized mean are equal, but this is obviously not a common rule. In addition, both the trimmed and the Winsorized mean will change, depending on the percent of trimming. Depending on the purpose of the analysis, computer simulation studies recommend 10-15% trimming (Keselman et al., 2004), 15% trimming (Othman et al., 2004), 20% trimming (Keselman et al., 2008), or 20-25% trimming (Rocke, Downs, Rocke, 1982) (detailed discussion on this can be found in (Wilcox, 2005)). In each case, the amount of trimming is arbitrary and has to be determined by a researcher. In addition, the amount of trimming is commonly symmetrical and does not depend on the distribution of values in the empirical data and pattern of outliers.

An alternative approach to trimming is to determine outliers first, and then to adjust the amount of trimming based on the distribution of values in the dataset. There are several estimators of this kind in contemporary statistics (examples can be found in (Rousseeuw, Croux, 1993)). The present study will focus on two measures of this group, namely the M-estimator (also referred to as a Huber one-step M-estimator, based on the name of its author (Huber, 1981)), and a modified one-step estimator.

The first step in computing these measures is to identify outliers. To calculate the one-step M-estimator,

absolute deviation from the median is first computed for each observation. For a dataset  $x_1, x_2, \dots, x_n$  absolute deviations from the median are:  $|x_1 - Md(x)|, |x_2 - Md(x)|, \dots, |x_n - Md(x)|$ . Next, the median of these absolute values (median absolute deviation, *MAD*) is computed and scaled by a constant:

$$MADN = \frac{MAD}{.675} \quad (5).$$

An outlier is defined as any value, where the following statement is true:

$$\frac{|x_i - Md(x)|}{MADN} > K \quad (6),$$

where  $K$  is a constant, commonly chosen to be 1.28. Let  $a_1$  be a number of observations where

$$\frac{x_i - Md(x)}{MADN} < -1.28,$$

and  $a_2$  be a number of observations where

$$\frac{x_i - Md(x)}{MADN} > 1.28.$$

The one-step M-estimator (*OSE*) is:

$$OSE = \frac{1.28 \cdot MADN \cdot (a_2 - a_1) + \sum_{i=a_1+1}^{n-a_2} x_i}{n - (a_2 + a_1)} \quad (7).$$

The modified one-step M-estimator (*MOSE*) is easier to compute. It is calculated by removing the values that were identified as outliers and averaging the remaining values in the dataset. The threshold for considering a value to be an outlier is commonly defined by:

$$\frac{|x_i - Md(x)|}{MADN} > 2.24.$$

Any other value besides the commonly used value 2.24 can be used as a threshold, thus representing more rigorous or tolerant criterion for determining outliers.

In our numerical example, the one-step M-estimator is 4.87; the modified one-step M-estimator is 4.56. The latter value is equal to the arithmetic mean calculated on the values lying within two standard deviations in this example, because both procedures identified the same single data point as an outlier. However, this coincidence is obviously not a general rule.

### **Data weighting as a possible approach to obtaining stable measures of central tendency**

Two different approaches to the analysis of location of a distribution were discussed above. Within the first approach, all values from an original dataset were used in calculating the measures of central tendency; the second approach required removing some values from the analysis. Each approach has its advantages and disadvantages. Classical algorithms, which use entire datasets, are simpler, but the obtained estimates can be biased when some observations lie outside the main array. Algorithms that are based on trimming allow for calculating estimates that are more stable in the presence of outliers, but the amount of trimming is always arbitrary. In addition, the latter approach requires an assumption that all of the relatively high and low values must be removed in order to obtain an accurate estimate of location of a distribution.

However, there might be another approach to calculating measures of central tendency that does not require regarding some values in a dataset as indubitable accidental outliers. Indeed, data points in the tails of a distribution can still be used in an analysis of central tendency, but they can be downweighted, as compared to the values that are confined to the main data array. In other words, each value in the original dataset can be weighted with respect to its position in the array, and the weighted values can be averaged. Central observations in a dataset will then give the highest weights, while outlying values will get lower weighting coefficients.

The fact that weighting of observations with respect to their positions in a dataset is rarely employed in analysis of location of empirical response-time distributions seems remarkable. Indeed, a general formula of weighted mean is shown in almost every textbook on statistics:

$$WM = \frac{\sum_{i=1}^n (w_i \cdot x_i)}{\sum_{i=1}^n w_i} \quad (8).$$

However, weighting has still been extremely rarely used in the analysis of empirical data. A possible problem might be determining weighting coefficients for values from the original dataset. Indeed, any weighting coefficients  $w_i$  could be substituted into the above formula. However, finding an algorithm for computing these coefficients is itself a challenging task.

Two different algorithms for weighting will be introduced in this paper. In both algorithms, computing weighting coefficients does not require mean or other parameters of a distribution of data as input information, because each value in the dataset is weighted in relation to the entire data array.

The first algorithm is a calculation of a distance-weighted estimator (*DWE*). Consider a dataset  $x_1, x_2, \dots, x_n$ . The weighting coefficient for  $x_i$  is computed as the inverse mean distance between  $x_i$  and the other data points:

$$w_i = (n - 1) \cdot \left( \sum_{j=1}^n \text{abs}(x_i - x_j) \right)^{-1} \quad (9).$$

These coefficients are next substituted into the formula (8). Note that weighting coefficients could be calculated as a simple sum instead of an average of the distances, but in certain cases (e.g., large sums of distances) this would result in a lower accuracy in the calculations.

For illustrative clarity, consider a simple numerical example of a dataset consisting of 4 observations:  $x_1 = 2, x_2 = 3, x_3 = 5, x_4 = 12$  ( $n = 4$ ). Weighting coefficients for  $x_i$  are:

$$w_1 = 3/(|2 - 3| + |2 - 5| + |2 - 12|) = 3/14,$$

$$w_2 = 3/(|3 - 2| + |3 - 5| + |3 - 12|) = 3/12,$$

$$w_3 = 3/(|5 - 2| + |5 - 3| + |5 - 12|) = 3/12,$$

$$w_4 = 3/(|12 - 2| + |12 - 3| + |12 - 5|) = 3/26.$$

The distance-weighted estimator is:

$$DWE = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4}{w_1 + w_2 + w_3 + w_4} \approx 4.59.$$

In the numerical example that was analyzed earlier in this paper, the distance-weighted estimator is 5.69.

Another possible algorithm for weighting is employed in calculating the scalar-product weighted estimator (SPWE, (Dodonov, 2010)). Let  $x_1, x_2, \dots, x_n$  be an original dataset  $X$ . To calculate the SPWE, the original data array is transformed into a derived data array  $Y$  by pairwise averaging of the original data points:

$$y_k = \frac{x_i + x_j}{2} \Big|_{i < j} \quad (10).$$

To compute weighting coefficients for the derived array  $Y$ , data points from the original array  $X$  are regarded as unit vectors  $l$ , which make angles from 0 to  $\pi/2$  with the horizontal axis:

$$\theta(\hat{l}_i) = \frac{\pi}{2} \cdot \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (11).$$

Weighting coefficients for the derived array  $Y$  are computed as scalar products of the corresponding unit vectors  $l$ :

$$w_k = \hat{l}_i \cdot \hat{l}_j = \cos(\lambda[\hat{l}_i, \hat{l}_j]) = \cos\left[\frac{\pi}{2} \cdot \frac{x_i - x_j}{x_{max} - x_{min}}\right] \quad (12).$$

The SPWE is defined by:

$$SPWE = \frac{\sum_{i=1, j=1, i < j}^n [(x_i + x_j) \cdot \cos\left(\frac{\pi}{2} \cdot \frac{x_i - x_j}{x_{max} - x_{min}}\right)]}{\sum_{i=1, j=1, i < j}^n \cos\left(\frac{\pi}{2} \cdot \frac{x_i - x_j}{x_{max} - x_{min}}\right)} \quad (13).$$

Consider a numerical example:  $x_1 = 2, x_2 = 3, x_3 = 5$  and  $x_4 = 12$  ( $n = 4$ ). A derived data array  $Y$ , calculated by pairwise averaging of the original values, is:

$$y_1(x_1, x_2) = 2.5, y_2(x_1, x_3) = 3.5, y_3(x_1, x_4) = 7,$$

$$y_4(x_2, x_3) = 4, y_5(x_2, x_4) = 7.5, y_6(x_3, x_4) = 8.5.$$

The values of  $X$  are next transformed into unit vectors, which make angles from 0 to  $\pi/2$  with the horizontal axis, as shown in Figure 1.

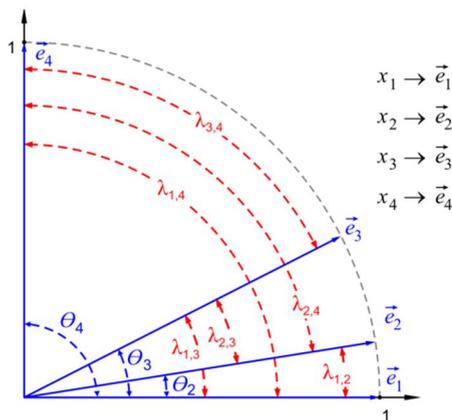


Figure 1. Original values transformed into the unit vectors

In this representation, an angle between two unit vectors can be regarded as a measure of the distance between the corresponding data points from the original array  $X$ . Therefore, the weighting coefficient for a pair of original values (for a corresponding value in the derived array  $Y$ ) is calculated as a scalar product of the corresponding unit vectors, which is a cosine of the angle between these vectors. In our example, the weighting coefficients  $w_k$  for  $y_k$  are:

$$w_1[y_1] \approx .988, w_2[y_2] \approx .891, w_3[y_3] = 0,$$

$$w_4[y_4] \approx .951, w_5[y_5] \approx .156, w_6[y_6] \approx .454.$$

This example illustrates that higher weights are ascribed to the values  $y_k$  that correspond to the pairs of close original data points. The closest data points in this example are  $x_1 = 2$  and  $x_2 = 3$ , so the value  $y_1$ , which was obtained by their averaging, gets the highest weight. The more distant two data points are, the less is the weighting coefficient for the value  $y_k$  that was obtained by their averaging. Substituting  $w_k$  and  $y_k$  into the formula (8) gives  $SPWE \approx 4.193$ .

In the above example of the dataset that consists of 10 observations (2, 3, 4, 4, 5, 5, 6, 6, 6, 80), the scalar-product weighted estimator is 5.04.

Computations that are needed for calculating the latter measure of central tendency look extensional, but the algorithm can simply be implemented to any program available for research, from standard versions of Excel to R code. In general, possible computational loadings that are allowed by contemporary programs do not place unreasonable limitations on the choice of a measure of central tendency in the analysis of location of empirical distributions. Conceptual issues, which will be discussed in the next sections, should be taken into account first in preferring one measure of central tendency to the others.

### The importance of choosing the measure of central tendency: an example of empirical response-time data

The simple numerical example that was used in the previous section clearly demonstrated that the estimate of location of the data distribution can vary substantially using different measures of central tendency. In the context of studies on processing speed in simple cognitive tasks, this means that different estimates of location of individual response-time distribution can be obtained for each participant depending on the chosen measure of central tendency. Can the choice of a measure of central tendency become crucial for the subsequent analysis of the association between response time in speed tasks and external variables? This problem is discussed in the following section, based on an example from a real empirical study.

### Methods

The results that were obtained in one of our studies (Dodonova, Dodonov, in press) are employed in this section. The sample size for the analyses that are presented below was 231 (58% female), and the mean age of the participants was 15.64 years (standard deviation .70 years).

The speed task required participants to determine, as quickly as possible, if the figure that appeared on the

screen was a triangle. The task consisted of twenty trials, wherein five different geometrical figures could appear on the screen. The approximate size of the figures was 35 × 35 mm. The participants had to press the key '1' for the triangle, and the key '0' for any other figure. The number of '1' and '0' responses was equalized throughout the test.

Error responses were eliminated from further analyses. Eleven possible estimates of individual response time were next computed for each participant, using eleven different measures of central tendency, which were described in the previous section. The R code that was used for calculations can be found in Appendix.

For illustrative purposes, the values of an external 'criterion' variable were generated in R so that a correlation between this new generated variable and empirical response times was marginal with respect to its significance. In other words, we intentionally modeled a situation where the choice of which measure of central tendency is used in calculating individual response times could become crucial for further conclusions on the significance of the association between response times and the external variable.

## Results

As expected, individual times of discrimination that were calculated based on different measures of central tendency were highly correlated. Correlations between individual response times computed by different algorithms varied from .949 (between the arithmetic mean and modified one-step M-estimator) to .999 (between the trimmed mean and distance-weighted estimator).

Correlations between the different measures of individual response time and the generated external variable are presented in Table 1.

The differences between the correlation coefficients that are presented in Table 1 are statistically non-significant (Fisher's z-transformation was employed for comparison:  $p > .05$  for each pair of coefficients). However, the association between response time and the external variable was significant for 6 measures out of 11, while for the other 5 measures of central tendency, the results clearly demonstrated non-significance of the association between processing speed and the external variable.

## Discussion

The speed task that was used in this example is typical for studies on individual differences in processing speed and their relations to other cognitive indexes. The task included a relatively small number of trials. Several response times were registered for each participant; a

single measure of individual performance had to be obtained for further analysis.

Individual response times that were obtained by using different measures of the location of individual response-time distributions were highly correlated. However, even small differences in the obtained indexes of individual performance in the speed task were crucial for the conclusion on the statistical significance of the association between response times and the external variable. Depending on the measure of central tendency preferred for the analysis of location of individual response-time distributions, response times appeared to be significantly related to the external variable (e.g. for the harmonic mean) or non-correlated with the same external variable (e.g. for the median).

The results clearly demonstrate that choosing a measure of central tendency must be a problem that is carefully considered, because the choice that is made at the very preliminary stage of analysis, when individual indexes of performance are calculated, can directly affect further results. Our example certainly represented borderline conditions, when the significance of the correlation was not obvious and was affected by the measure of central tendency. However, such 'borderline' correlation coefficients are not rare in empirical studies, including studies on speed of information processing.

For example, Jensen reports an average correlation  $-.10$  between simple reaction time and intelligence, and  $-.20$  between discrimination time and intelligence (Jensen, 1998, p.211). Based on a meta-analysis of 195 studies, Sheppard and Vernon (2008) report a correlation of  $-.17$  between reaction time and, for example, crystallized intelligence. The above-mentioned association between processing speed and intelligence, although it is relatively slight, has been replicated in a large number of studies. To demonstrate the significance of the correlation of  $-.17$  in their meta-analysis, Sheppard and Vernon employed a method proposed by Rosenberg (2005). The authors demonstrated that "a non-significant correlation from a sample of 58,158 participants would be required, or that over 58,000 missing reports with non-significant results would need to have been omitted from this review to reduce the significance of the average correlation to  $>.05$ " (Sheppard, Vernon, 2008, p.541). Both possibilities are unlikely, and the results obtained in the last decades provide clear evidence of the significance of this low association. However, a large body of previous studies is rarely available, and the significance of the observed association has to be judged based on the results of a single study. An example of this kind was analyzed in this paper.

The results that were presented above demonstrated

**Table 1**

Correlations between empirical individual response times calculated using different measures of central tendency and the simulated external variable

	<i>M</i>	<i>hM</i>	<i>gM</i>	<i>Md</i>	<i>M(2SD)</i>	<i>TrimM</i>	<i>WinsM</i>	<i>OSE</i>	<i>MOSE</i>	<i>DWE</i>	<i>SPWE</i>
<i>r</i>	.157	.174	.167	.072	.159	.125	.149	.126	.096	.129	.152
<i>p</i>	.017	.008	.011	.275	.015	.057	.023	.057	.145	.050	.021

*Note.* *r* – Pearson's correlation coefficient, *p* – significance level; *M* – arithmetic mean; *hM* – harmonic mean; *gM* – geometric mean; *Md* – median; *M(2SD)* – mean after eliminating values outside two standard deviations; *TrimM* – 20% trimmed mean, with the lowest and the highest 20% values removed; *WinsM* – Winsorized mean, with prior 20% trimming; *OSE* – one-step M-estimator; *MOSE* – modified one-step M-estimator; *DWE* – distance-weighted estimator; *SPWE* – scalar-product weighted estimator.

how the conclusion on the significance of associations in a single study can be affected by the choice of measure of central tendency in the analysis of location of individual response-time distributions. However, these results do not say anything about what measure of central tendency is preferable in data analysis. To address this question, the behaviors of the measures of central tendency were analyzed in a series of computer simulations that mimicked empirical response-time data. The results of the comparison are presented in the next section.

### Behavior of the measures of central tendency: computer simulation of response-time data

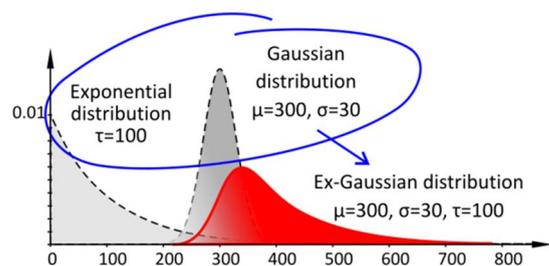
This section consists of three parts. The first part describes the methods of the computer simulation study, including the algorithm for generating data that mimic empirical response-time distribution, with and without outliers, as well as criteria for further comparison of the measures of central tendency. The second part shows the results obtained in the series of computer simulations. Discussion in the third part covers behaviors of the measures of central tendency under modeled conditions.

#### Methods

For the computer simulation study, the ex-Gaussian distribution function was used to mimic empirical response-time distribution. The ex-Gaussian distribution is the convolution of the normal and the exponential distributions (described as ex-Gaussian by Luce (1986)). Mathematically, the ex-Gaussian probability density function is:

$$f(x|\mu, \sigma, \tau) = \frac{1}{\tau\sqrt{2\pi}} \exp\left(\frac{\sigma^2}{2\tau^2} - \frac{x-\mu}{\tau}\right) \cdot \int_{-\infty}^{\left\{\left[\frac{x-\mu}{\sigma}\right] - \left(\frac{\sigma}{\tau}\right)\right\}} \exp\left(-\frac{y^2}{2}\right) dy \quad (14).$$

For illustrative purposes, Figure 2 shows the original exponential and Gaussian distributions, and an ex-Gaussian distribution that is a convolution of these two distributions.



**Figure 2.** The ex-Gaussian distribution as the convolution of the exponential and Gaussian distributions.

Thus,  $\mu$  and  $\sigma$  parameters of the ex-Gaussian distribution correspond to the mean and standard deviation of the Gaussian component. The third parameter  $\tau$  is the mean of the exponential component.

A tradition of fitting response-time data to the ex-Gaussian distribution function has a long history in studies on information processing. A conceptual interpretation of the parameters of the ex-Gaussian function in terms of the components of cognitive processing was first proposed by Hohle (1965), who suggested that the exponential

parameter represents the decision component of response times, while the Gaussian process can be conceptualized as the non-decision motor response component. Leaving aside the theoretical interpretation of the ex-Gaussian function, it is still important to note that several studies have demonstrated that it provides a very good fit to empirical response-time distributions in different speed tasks (Heathcote, Popiel, Mewhort, 1991; Hockley, 1982, 1984; Hohle, 1965; Ratcliff, 1978, 1979; Ratcliff, Murdock, 1976; Ulrich, Miller, 1994). Therefore, the ex-Gaussian distribution was chosen in the present study for the computer simulation of the data that mimics empirical response times. The parameters of this distribution were set to be similar to those that could be obtained in simple speed tasks, such as choice reaction time or discrimination time tasks:  $\mu = 400$ ,  $\sigma = 20$ ,  $\tau = 100$ .

Thirty values from the ex-Gaussian distribution with the above-mentioned parameters were generated in R. The number of values was small in order to mimic a number of trials in a speed task, which is commonly kept to between 20 and 40 trials. Therefore, we simulated datasets of individual response times that are typical in size for empirical studies on individual differences in processing speed. For a generated dataset, each measure of central tendency under consideration was calculated. The procedure was repeated 50,000 times.

Therefore, data points were drawn from the same distribution using the parameters that were fixed across the generations, although the exact values in a dataset obviously varied across the generations. Clearly, if a measure of central tendency is sensitive to the fluctuations of exact observations in a dataset, its value will vary across the generations, and across the data originating from the same distribution. On the other hand, if a measure of central tendency is stable in retrieving the location of the underlying distribution, variability of its values across generations will be low. In other words, the standard deviation of a measure across 50,000 generations can be regarded as a measure of its stability relative to variations of observations in a dataset.

The procedure that was described above was used to generate response times under an 'ideal' condition with a complete absence of outliers. However, real empirical data generally include outlying values that are not confined to a distribution of the other observations. Theoretically, these random values do not necessarily appear in the tails of a distribution. Random delays of responses can be short, and anticipatory responses are not always implausible. In other words, 'noise' can appear in any part of a distribution. In general, different algorithms can be used to model outliers in a computer simulation study. First, some values can be added to several randomly chosen values from the main distribution. For example, in one of his studies, Ratcliff (1993) generated response times from an ex-Gaussian distribution, and then added random numbers from a uniform distribution from 0 to 2,000 to some of the original values. With this approach, however, only random delays of response are simulated, while the real data can also include responses that are faster than most of the others. To model data with different types of outliers, several values from an original distribution can be replaced by values drawn from some other distribution. For example, the presence of outliers in only the left or only the right tail of a distribution was demonstrated by Ratcliff (1993) by generating 80% of the data points from the basic ex-Gaussian distribution, and generating the other 20% of

the data points from the ex-Gaussian distribution with a noticeably smaller or larger  $\mu$  parameter.

In other words, the choice of an algorithm for simulating data with outliers is always arbitrary and obviously depends on purposes of a study. As we aimed to simulate any outlying values that produce 'noise' in the main distribution, outliers were drawn randomly from the uniform distribution of values from 0 to 2000, as shown in Figure 3.

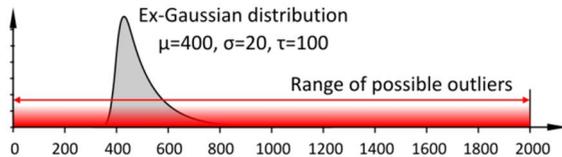


Figure 3. Basic distribution and the range of possible outliers.

Therefore, three more conditions were simulated, in addition to the basic condition with no outliers, wherein the data could include 1, 2 or 3 outliers, respectively. In each condition with outliers, as in the basic condition with no outliers, 30 values were generated, and the measures of central tendency were calculated within each generation. For each condition, 50,000 generations were completed. The mean and standard deviation of each measure of central tendency across 50,000 generations were computed. The R code for the generations and further computations under a 2-outliers condition is shown in Appendix.

As the means for the measures of central tendency were calculated separately for each condition with fixed number of outliers, another index that could be calculated was a standard deviation of these 4 means. This index represented the stability of a measure of central tendency across the conditions with different numbers of outliers. As

discussed above, only stability of this kind is in fact referred to as the robustness of the measure of central tendency, or its insensitivity to outliers. However, this is not the only source of stability this study will analyze.

Finally, one more series of 50,000 generations was completed, wherein the number of outliers was not constant. Within each generation, the 30 values could include 0 to 3 outliers. As for the previous conditions, all the measures of central tendency were calculated in each generation; the mean and standard deviation of each measure across 50,000 generations were next computed. Under these conditions, the variability of a measure of central tendency originates from the both sources of non-stability that were discussed above: sensitivity to fluctuations in a dataset, and non-robustness to outliers.

## Results

For each modeled condition, Table 2 shows the mean and standard deviation of each measure of central tendency across 50,000 generations. Means were rounded off for simplicity. In each condition, boldface font is used for the three measures that are the most stable across 50,000 generations.

Column SD\* shows standard deviations of the 4 means of the measures of central tendency that were computed across the 4 conditions with a fixed number of outliers. Again, boldface font is used to mark the three measures that are the most robust to presence of outliers and changes in their number.

The last two columns of Table 2 show the means and standard deviations of the measures of central tendency across 50,000 generations, wherein the number of outliers in the data varied from 0 to 10%. Boldface font is again used for the three most stable measures.

Table 2  
Stability of the measures of central tendency

Measure of central tendency	No outliers		1 outlier		2 outliers		3 outliers		SD*	0 to 3 outliers	
	M	SD	M	SD	M	SD	M	SD		M	SD
<i>M</i>	500	18.580	517	26.609	533	32.598	550	37.787	21.57	525	35.151
<i>hM</i>	<b>484</b>	<b>14.699</b>	481	43.451	478	58.925	474	71.390	4.20	479	52.197
<i>gM</i>	<b>491</b>	<b>16.319</b>	498	23.043	505	28.160	513	32.991	9.13	502	27.072
<i>Md</i>	473	18.109	475	19.062	477	19.877	<b>480</b>	<b>20.922</b>	<b>2.80</b>	476	19.600
<i>M(2SD)</i>	484	17.280	492	19.022	497	20.410	501	23.164	7.11	494	21.008
<i>TrimM</i>	479	16.581	<b>482</b>	<b>17.573</b>	<b>485</b>	<b>18.509</b>	<b>489</b>	<b>19.756</b>	<b>4.03</b>	<b>484</b>	<b>18.389</b>
<i>WinsM</i>	484	17.164	<b>488</b>	<b>18.356</b>	<b>491</b>	<b>19.587</b>	496	21.253	5.07	<b>490</b>	<b>19.593</b>
<i>OSE</i>	484	17.543	487	18.740	490	19.851	494	21.375	4.35	488	19.700
<i>MOSE</i>	475	19.369	477	19.971	478	20.484	480	21.259	<b>2.20</b>	478	20.319
<i>DWE</i>	<b>483</b>	<b>16.501</b>	<b>487</b>	<b>17.715</b>	<b>492</b>	<b>19.006</b>	<b>498</b>	<b>20.901</b>	6.71	<b>490</b>	<b>19.407</b>
<i>SPWE</i>	490	17.066	498	19.407	507	21.882	517	25.079	11.61	503	23.275

Note. The measures of central tendency are denoted as in Table 1; SD – standard deviation of a measure across 50,000 generations; SD\* – standard deviation of the mean of a measure across 4 conditions with fixed number of outliers. In each column, boldface font is used for the three measures with the fewest standard deviations.

## Discussion

As expected, the arithmetic mean was one of the least stable measures of central tendency, both under the no-outlier condition and in presence of outliers. The behavior of this measure of central tendency worsened as the number of outliers increased. The opposite trend was observed for the median. When the number of outliers was as large as 10%, the median was one of the most stable measures of central tendency. However, when there were no outliers, or their number varied within a series of generations, the median was less stable than most of the other measures of central tendency.

In the absence of outliers, the measures, which are the most stable in the case of analyzed skewed distribution, are the harmonic mean, the geometric mean and the distance-weighted estimator that was introduced in this paper. However, the behaviors of the harmonic mean and geometric mean worsened substantially when outliers appeared in a dataset. When a distribution included outliers, the harmonic mean, geometric mean and arithmetic mean were the three least stable measures, highly sensitive to the exact values in a dataset. At the same time, the distance-weighted estimator remained one of the most stable measures of central tendency within each condition with outliers.

The mean computed on the values within two standard deviations was more stable than the simple arithmetic mean. However, this measure was not preferable in any modeled condition. This result looks predictable because of the procedure of calculating this measure of central tendency, which required removing values that lay outside the two standard deviations, from the arithmetic mean calculated on all the values. Since outliers can substantially distort the latter measure, the entire procedure can hardly be effective. However, this measure of central tendency is still one of the most commonly used in the analysis of empirical response times. In this context, our results are a clear reminder to researchers that there are algorithms available for the analysis of location of response-time distributions that are much more effective than the habitual procedure of removing values lying outside two standard deviations.

In general, our results confirm that the 20% trimmed mean can be regarded as one of the most preferable measures of central tendency. This measure was not sensitive to the fluctuations in a dataset in any of the conditions with outliers. In addition, the trimmed mean was one of the measures most robust to presence of outliers in the data and to changes in their number.

The Winsorized mean was also one of the most preferable measures of central tendency in the 2 of 4 modeled conditions with a constant number of outliers, as well as in the condition with the varying number of outliers. Note, however, that the algorithm for calculating the 20% Winsorized mean requires 20% trimming first, and also a replacement of the removed values. At the same time, the Winsorized mean was a less stable measure of central tendency than the trimmed mean. Therefore, the results of this study do not provide any evidence for the necessity of Winsorizing to obtain a stable measure of central tendency.

Finally, the results that concern the one-step M-estimator and the modified one-step M-estimator are quite surprising. These measures of central tendency, as well as several other estimators of this type, are commonly

recommended in the literature as statistics that are robust to the presence of outliers and the skewness of a distribution. Indeed, the modified one-step M-estimator was a measure that was the most stable across the modeled conditions. However, for each condition with a constant number of outliers, neither M-estimator fell within the three most stable measures of central tendency. Moreover, neither M-estimator could be regarded as a stable measure within the condition with the varying number of outliers.

In this context, we should come back to the discussion of the two possible sources of the non-stability of a measure of central tendency that were mentioned above. In contemporary literature on statistics, the stability of a measure of central tendency is mostly analyzed by its robustness to the appearance of outliers in a dataset. Different approaches to estimating the robustness of a measure of central tendency were mentioned above. In this study, the stability of a measure across conditions with different constant numbers of outliers was regarded as an index of robustness. The results confirm that the median and the M-estimators are the most stable in this sense. However, another problem is obvious from the results. Although the means of these measures were relatively stable across the conditions with different numbers of outliers, within each condition these measures were highly sensitive to fluctuations in the values in a given dataset, even though the values were drawn from the same distribution. The results of the simulation of data with varying number of outliers clearly demonstrate that the median and the M-estimators cannot be recommended as stable measures of central tendency, if both possible sources of non-stability are considered in the analyses.

To summarize, two measures of central tendency, the 20% trimmed mean and the distance-weighted estimator, were the most preferable for the conditions that were modeled in this study. It is worthwhile to emphasize that this simulation study aimed to mimic response-time data that are typical for a real speed task. The behaviors of the measures of central tendency will obviously change with changes in the skewness of a distribution, the number of outliers and their range. In other words, the accurate choice of a measure of central tendency should ideally be based on a preliminary comparative analysis of the behaviors of the measures under conditions that best describe real empirical data.

## General discussion

This paper focused on the problem of estimating central tendency for empirical data, distributions of which are skewed and include outliers. This problem was primarily analyzed in the context of studies on processing speed, where a single index of individual performance in a speed task is obtained based on a set of response times that are positively skewed and are likely to include outliers.

Results of the computer simulation study demonstrate that two measures of central tendency can be regarded as preferable in these conditions: the 20% trimmed mean and the distance-weighted estimator that was introduced in this paper. The conclusion concerning the trimmed mean is quite expected, as contemporary statistics commonly recommends this measure for the analysis of central tendency on data that are not normally distributed and are likely to include outliers. However, it is important to note that only 60% of the empirically obtained values are

employed in calculating the 20% trimmed mean. Researchers should be clearly aware of this property of the trimmed mean. As noted by Wilcox (2001), the idea that removing observations can lead to a more accurate estimate than an analysis of the entire dataset is counterintuitive. However, the problem is somewhat deeper, because the question of whether a researcher can eliminate any values that were observed empirically is open. When demonstrating the effect of outliers on the measures of central tendency, textbooks on statistics commonly use clear examples, similar to the set of observations 2, 3, 4, 4, 5, 5, 6, 6, 6, and 80, which was employed in the first section of this paper. However, can we regard a value as an outlier when its position in relation to a data array is not so obvious? For instance, what values must be regarded as outliers in a dataset 260, 291, 303, 343, 403, 468, 494, 536, 548, 821 (the response times of a real participant, ms)? Importantly, the problem is not just in the accurate identification of outlying values, but also in the appropriateness of ignoring these values when calculating a measure of central tendency. Indeed, is a long response time completely due to the effects of extrinsic factors, or do long response times also reflect some feature of the underlying process that is analyzed by a researcher?

To our mind, the main advantage of the measures of central tendency that are based on data weighting is that they do not require definite judging of whether or not some values must be deleted as outliers. For instance, the distance-weighted estimator that was proposed in this paper is calculated on the entire dataset, where each value is weighted according to its distance from the other observations in an array. The computer simulation study showed that the behavior of this measure was comparable to that of the 20% trimmed mean. Clearly, for the positively skewed distribution that was used in this study, the distance-weighted estimator was always somewhat higher than the trimmed mean, because high values in the right tail of the distribution were not simply ignored as they would be in trimming, but were accounted for in the analysis of location of distributions.

Finally, it is important to emphasize that the problem of choosing a measure of central tendency is not specific to the empirical data obtained in studies on individual differences in processing speed. Indeed, an analysis of individual response times is an area where accurately estimating the location of a distribution is extremely important. The present paper demonstrated how the behaviors of the measures of central tendency differed across conditions, and how a chosen measure of central tendency can become crucial for further analysis and conclusions. This problem is, however, by no means limited to response-time data. In fact, it should be considered when any empirical data have to be averaged. In other words, a researcher can calculate a single score based on the results obtained from a questionnaire, or summarize the observations or quantitative clinical data, and in all of these cases, 'simple averaging' is the procedure that requires special attention and preliminary analysis of empirical data.

To summarize, the following conclusions can be formulated:

1. Researchers mostly deal with data that are not normally, or even symmetrically, distributed, and the number of observations is commonly small. Under these conditions, the arithmetic mean cannot be regarded as a

reliable measure of central tendency because of its non-robustness to the skewness of a distribution and the presence of outliers.

2. The approach to the analysis of location of distribution of empirical data that is commonly recommended by contemporary statistics is based on removing values from the tails of a distribution and a further averaging of the remaining values. Although different algorithms are proposed for identifying which values have to be removed, a stable measure of central tendency can be obtained even by simple trimming of a fixed amount of the smallest and largest values. The amount of trimming is arbitrary, and has to be determined depending on the features of a study, but generally as much as 20% trimming seems to give satisfactory results.

3. An approach that is based on weighting can be regarded as an alternative to data trimming. With this approach, a measure of central tendency is calculated on all values in a dataset, but their relative weights depend on their distance from the other points in the data array. Of the two weighted estimators of central tendency, at least one is preferable to most of the other measures and is comparable in its behavior to the 20% trimmed mean. At the same time, the obvious advantage of an approach based on weighting is that it does not require removing any values, which is extremely important for empirical studies when no data point can be identified as an outlier with confidence.

## Appendix

### *R code for computing the measures of central tendency*

```
# arithmetic mean:
mean(x)

# harmonic mean: hM
hM=function(x,n) {y=x[!is.na(x)];n=length(x); n/sum(1/y)}

# geometric mean: gM
gM=function(x) {y=x[!is.na(x)];exp(mean(log(y)))}

# median:
median(x)

# Mean within 2 standard deviations
m2sd=function(x) {
y=x[!is.na(x)];a1=mean(y)+2*sd(y)
a2=mean(y)-2*sd(y);z=y[y>a2 & y<a1]; mean(z)}

# 20% trimmed mean:
mean(x,trim=0.2)

# Winsorized mean: wins
wins=function(x,tr){y=sort(x);n=length(x)
ibot=floor(tr*n)+1;itop=length(x)-ibot+1
xbot=y[ibot];xtop=y[itop];y=ifelse(y<=xbot,xbot,y)
y=ifelse(y>=xtop,xtop,y);mean(y)}

# one-step M-estimator: ose
ose=function(x) {y=x[!is.na(x)];m=median(y)
z=abs(y-m);md=median(z)/0.6745
a1=m-1.28*md;a2=m+1.28*md;d=y[y>a1 & y<a2]
i1=length(y[y<a1]);i2=length(y[y>a2])
(1.28*md*(i2-i1)+sum(d))/length(d)}
```

```
# modified one-step M-estimator: mose
mose=function(x) { y=x[!is.na(x)];m=median(y)
z=abs(y-m);md=median(z)/0.6745
a1=m-2.24*md;a2=m+2.24*md
d=y[y>a1 & y<a2];mean(d)}

# scalar-product weighted estimator: spwe
spwe=function(x) {y=x[!is.na(x)]
a=pi/2*(y-min(y))/(max(y)-min(y));b=a
p=outer(b, a, function(b, a) abs(cos(b-a)))
c=y;q=outer(c, y, function(c, y) (c+y)/2)
m1=p-diag(diag(p));m2=q-diag(diag(q))
sum(m1*m2)/sum(m1)}

# distance-weighted estimator: dwe
dwe=function(x) {y=x[!is.na(x)];a=y;b=y
p=outer(b, a, function(b, a) abs((b-a)))
n=colSums(p)/length(y);w=1/n;sum(y*w)/sum(w)}
```

### Example of R code for the computer simulation study

2-outliers condition, description of the parameters can be found in the text

```
Mu=400; Sigma=20; Nu=100
N=30;out=2
y=1:50000;a1=y;a2=y;a3=y;a4=y;a5=y
a6=y;a7=y;a8=y;a9=y;a10=y;a11=y
for(i in 1:50000){
t=rnorm(N-out, Mu, Sigma)+Nu*rexp(N-out)
p=runif(out, 0, 2000)
x=c(t,p)
a1[i]=mean(x);a2[i]=median(x);a3[i]=hM(x);a4[i]=gM(x)
a5[i]=m2sd(x);a6[i]=ose(x);a7[i]=mose(x);a8[i]=dwe(x)
a9[i]=spwe(x);a10[i]=mean(x,trim=0.2); a11[i]=wins(x,0.2)
}
mean(a1);mean(a2);mean(a3);mean(a4);mean(a5)
mean(a6);mean(a7);mean(a8);mean(a9);mean(a10)
mean(a11)
sd(a1);sd(a2);sd(a3);sd(a4);sd(a5);sd(a6)
sd(a7);sd(a8);sd(a9);sd(a10);sd(a11)
```

### References

- Barnett, V., Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). New York: Wiley.
- Dodonov, Y. S. (2010). Response time and intelligence: problems of data weighting and averaging. *Poster presented on the Eleventh Annual Conference of International Society for Intelligence Research*. Alexandria, USA.
- Dodonova, Y. A., Dodonov, Y. S. (in press). Speed of emotional information processing and emotional Intelligence. *International Journal of Psychology*.
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340–347.
- Hockley, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 8, 497–512.
- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 598–615.

- Hohle, R. H. (1965). Inferred components of reaction times as a function of foreperiod duration. *Journal of Experimental Psychology*, 69, 382–386.
- Huber P. J. (1981). *Robust statistics*. New York: Wiley.
- Jensen A. (1998). *The g factor*. London: Praeger.
- Keselman, H. J., Othman, A. R., Wilcox, R. R., & Fradette, K. (2004). The new and Improved two-sample t test. *American Psychological Society*, 15(1), 47–51.
- Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, 13, 110–129.
- Lovie, P. (1986). Identifying outliers. In A. D. Lovie (Ed.), *New developments in statistics for psychology and the social sciences* (pp. 44–69). British Psychological Society: London.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary organization*. New York: Oxford University Press.
- Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette, K. (2004). Comparing measures of the ‘typical’ score across treatment groups. *British Journal of Mathematical and Statistical Psychology*, 57, 215–234.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446–461.
- Ratcliff R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510–532.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Ratcliff, R. & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190–214.
- Rocke, D. M., Downs, G. W., & Rocke, A. J. (1982). *Are robust estimators really necessary?* *Technometrics*, 24, 95–101.
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, 59, 464–468.
- Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273–1283.
- Sheppard, L. D., Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, 44(3), 535–551.
- Ulrich, R., & Miller, J. (1994). Effects of outlier exclusion on reaction time analysis. *Journal of Experimental Psychology: General*, 123, 34–80.
- Wilcox, R. R. (2001). *Fundamentals of modern statistical methods*. New York: Springer.
- 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., & Keselman, H. J. (2005). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8, 254–274.