Standardized Distance from the Mean to the Median as a Measure of Skewness

José Moral de la Rubia

School of Psychology, Universidad Autónoma de Nuevo León, Monterrey, México
Email: jose.morald@uanl.edu.mx, jose_moral@hotmail.com

Abstract

The normal distribution, which has a symmetric and middle-tailed profile, is one of the most important distributions in probability theory, parametric inference, and description of quantitative variables. However, there are many non-normal distributions and knowledge of a non-zero bias allows their identification and decision making regarding the use of techniques and corrections. Pearson's skewness coefficient defined as the standardized signed distance from the arithmetic mean to the median is very simple to calculate and clear to interpret from the normal distribution model, making it an excellent measure to evaluate this assumption, complemented with the visual inspection by means of a histogram and a box-and-whisker plot. From its variant without tripling the numerator or Yule's skewness coefficient, the objective of this methodological article is to facilitate the use of this latter measure, presenting how to obtain asymptotic and bootstrap confidence intervals for its interpretation. Not only are the formulas shown, but they are applied with an example using R program. A general rule of interpretation of ∓ 0.1 has been suggested, but this can only become relevant when contextualized in relation to sample size and a measure of skewness with a population or parametric value of zero. For this purpose, intervals with confidence levels of 90%, 95% and 99% were estimated with 10,000 draws at random with replacement from 57 normally distributed samples-population with different sample sizes. The article closes with suggestions for the use of this measure of skewness.

1. Introduction

The skewness of the distribution of a random variable is an important property

for choosing the techniques for parameter estimation and hypothesis testing. Above all, it is of interest to check whether the distribution is symmetrical because of its possible deviation from the normality and the effect of skewness on the parametric tests, such as the t-test for difference of means or analysis of variance [1]. Normality takes on special relevance because of its central role in probability theory, asymptotic estimation, and is the distribution model of many quantitative variables in different fields of science, such as measures of intelligence, temperament, and expressive attitudes in open societies within psychology.

Many measures of skewness have been developed since the late 19th century, but one of the first measurement proposals, made by Karl Pearson [2], still constitutes one of the best choices: the standardized signed distance from the arithmetic mean to the median [3] [4] [5] [6]. Its interpretation is intuitive and clear, especially in comparison with the unimodal profile of a normal distribution. There are no general cut-off points to establish when the distribution is skewed and therefore not normal. Taking up the suggestion of Doane and Seward [4], this article aims to facilitate the implementation of this measure by showing operationally how to establish when there is deviation from a symmetric profile and estimating bootstrap confidence intervals using the program R. In addition, bootstrap confidence intervals at 90%, 95% and 99% centered at 0 for normally distributed samples of different sizes are shown as interpretative guide of deviation from symmetry and in order to qualify the general interpretative rule of ∓ 0.1. It is worth noting that the normal distribution is characterized by its symmetry and middle tails, hence it is the best pattern for generating interpretive symmetry guidelines for distributions with finite moments.

The article begins by presenting the concept of skewness and its measurement to focus on Pearson's [2] measure and its variant with −1 to 1 bounding, when the numerator is not multiplied by three, named Yule's skewness coefficient [4] [7] [8] [9]. The formulas for calculating the asymptotic error and confidence interval are shown [8] and scripts for computing bootstrap intervals using the program R are provided. As interpretative guide, confidence intervals at 90%, 95% and 99% are estimated by bootstrap resampling for different sample sizes under the assumption of the mesokurtic symmetry of a normal distribution [10] [11], which makes it possible to calibrate the interpretive rule of ∓ 0.1. The manuscript concludes by drawing some conclusions and making suggestions for the use of this measure of the shape of the distribution. It should be noted that the article has essentially a practical significance by showing the use of a skewness statistic with the application of a free software program and giving guidelines for interpretation based on the symmetrical pattern of the normal distribution.

2. Concept of Skewness

The concept of skewness was introduced at the beginning of scientific statistics at the end of the 19th century by the English biologist and mathematician Francis Galton [12] in the study of human faculties and by the English mathemati-

cian and philosopher Karl Pearson [2] [13] [14] in the development of his system of 12 continuous distributions [15]. Skewness and kurtosis as shape parameters together with a location parameter and a scale parameter allow defining the different distributions of the Pearsonian system. Since the influential work of this English mathematician, measures of skewness have been linked to continuous distributions, such as uniform, triangular, beta, exponential, gamma, Chi-squared, logistic, normal or lognormal and, in turn, applied to discrete distributions, such as Bernoulli, binomial, multinomial, geometric, hypergeometric or Poisson. Thus, the concept is not only used to specify continuous quantitative variables, such as reaction time or years of age, but also to describe discrete quantitative variables, such as the frequency of a behavior, number of successes on a test or number of children, and ordinal variables, such as the degree of agreement with an attitudinal statement on a Likert-type scale or subjective socioeconomic stratum [16]. A measure of skewness has even been proposed to characterize qualitative variables, such as categorical classifications [17].

Skewness can be understood as a property of the shape of the empirical distribution when represented by a bar chart, in the case of an ordinal or discrete quantitative variable with few values, or by a histogram, in the case of a continuous or discrete quantitative variable with many values, as well as of the theoretical or probability distribution when represented by a diagram of the probability mass function (discrete variable) or density function (continuous variable). A measure of central tendency, such as the arithmetic mean, median, mode or mid-rank, is taken as an axis of symmetry to divide the distribution into two parts. If both parts of the distribution are equal, that is, one is the reflection of the other, and there is symmetry. If both parts are disparate, there is asymmetry. For example, if the arithmetic mean (μ) is taken as the axis of symmetry, a distribution would have symmetry if $f_X(x - \mu) = f_X(x + \mu)$, where $f_X(x)$ is the probability mass function in a discrete distribution or the probability density function in a continuous distribution [18].

When describing the shape of a distribution from the diagram of a probability density or mass function, the following elements are distinguished: a peak, two (left and right) shoulders, and two (left and right) tails. The *peak* is the modal or most frequent value of the distribution [19]. It corresponds to the point on the abscissa axis (X values) where the curve reaches its maximum on the ordinate axis (point densities or probabilities). The *left shoulder* can be defined as the area between the 25th percentile and the 50th percentile or between the score at one standard deviation below the arithmetic mean and the arithmetic mean. The *right shoulder*, on the other hand, can be defined as the area between the 50th percentile and the 75th percentile or between the arithmetic mean and the score at one standard deviation above the arithmetic mean [20]. The *left tail* is the area between the left shoulder and the minimum value (threshold parameter a) or $-\infty$; likewise, the *right tail* is the area between the right shoulder and the maximum value (threshold parameter b) or $+\infty$. If the minimum value falls within the left shoulder area, there is no left tail; similarly, if the maximum value falls within

the right shoulder area, there is no right tail. These concepts would also apply to the histogram, including the bar chart.

Relative skewness measures (free of unit of measurement or free of location and scale parameters) are defined as ratios, proportions or averages centered at 0 [21]. A value of 0 indicates null skewness, *i.e.*, symmetry. In a continuous unimodal distribution, it reveals that the two shoulders and the two tails on both sides of the symmetry axis are identical, that is, one side is the reflection of the other (Figure 1, narrower curve in both diagrams). A positive value in the skewness measure usually indicates that the right tail is longer than the left, which causes the arithmetic mean to be shifted to the right and there are fewer values above the arithmetic mean than below. In continuous unimodal distributions with positive skewness, the mode (peak) falls below the median and the median below the arithmetic mean (Figure 1, wider curve in the right diagram). However, its generalization to other types of distribution depends on how heavy or light the tails are. Thus, a shortened tail at one end can compensate for a lengthened tail at the other end and give rise to symmetry (null value) when, in reality, both parts of the distribution are disparate. In continuous unimodal distributions, a negative value in the skewness measure shows that the left tail is longer than the right, which causes the arithmetic mean to be shifted to the left and there are fewer cases below the arithmetic mean than above (Figure 1, wider curve in the left diagram). In these distributions with negative skewness, the mode (peak) is above the median and the median is above the arithmetic mean [22]. This regularity holds well in continuous unimodal distributions, but in discrete distributions it has many counterexamples [6], hence it is very important to represent the distribution by a bar chart or histogram when assessing symmetry and interpreting its different measures [23].

There are several measures of skewness [20] [24]. Some are based on the standardized third central moment or the third standardized cumulant. Others are based on quantiles or expectiles. These latter can be considered a smoothed version of quantiles and, like the quantiles, also measure non-central location [7]. There are a third mixed measures that are based on moments and quantiles,
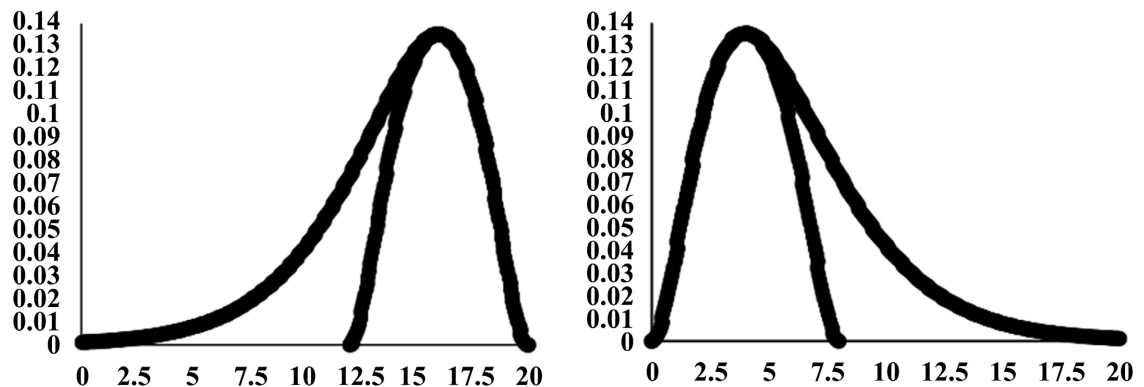

Figure 1. Probability density functions showing two examples of asymmetric curves and the corresponding symmetric curves for four continuous random variables.

such as the and standardized signed distance from the arithmetic mean to the mode [13] or to the median [2] [25], the standardized signed distance from the semi-range to the mode, median or arithmetic mean [21] or the area of skewness [6]. Not to mention a fourth type of robust asymmetry measures [5] [26].

3. Skewness Coefficients Based on Pearson's Standardized Distance

Karl Pearson [13] took as a model the normal distribution, which is a unimodal, symmetric, and bell-shaped distribution. Being unimodal symmetric, the arithmetic mean, median, and mode coincide, consequently, the difference between the arithmetic mean and mode or median is 0. If there is asymmetry due to elongation of the right tail due to atypical cases, the value of the arithmetic mean would be greater than that of the median and mode (unique), so the difference between the arithmetic mean and mode or median would be positive. Conversely, if there were asymmetry due to left tail elongation due to outlier cases, the value of the arithmetic mean would be smaller than that of the median and mode, so the difference between the arithmetic mean and (single) mode or median would be negative. With the intention of standardizing this measure, Karl Pearson decided to divide the difference or distance between the arithmetic mean and mode, $\mu_X - Mo(X)$, or median, $\mu_X - Mdn(X)$, by the standard deviation (σ_X), giving rise to the skewness coefficients based on the standardized distance of the arithmetic mean to the mode or median, which can be denoted by Sk_{P1} and Sk_{P2}, respectively.

3.1. Distance from Arithmetic Mean to Median

Pearson's first proposal had a clear drawback, its dependence on mode. The estimation of this measure of central tendency can be problematic with samples of strictly continuous variables, in which the data are not repeated, so it is required to tabulate by class intervals and use the class mark (midpoint of class interval) or a linear interpolation formula; in addition, it is very unstable with small samples [27]. Although nowadays, this pitfall has been overcome from modal peak density estimators [28] and a robust skewness measure based on the estimation of the mode has been proposed [26]. On the other hand, the distribution needs to be unimodal for its calculation, as is the case for the normal distribution, but not for all distributions, such as the beta distribution with shape parameters α and β less than 1 or the arcsine distribution. Moreover, it is not a clearly bounded measure.

To overcome these disadvantages, Karl Pearson [2] proposed a second formula, replacing the mode by the median and multiplying the numerator by three. In this way, a more stable and robust measure of central tendency is used, which can be calculated with any type of distribution. Moreover, with this change, the measure of skewness is bounded between −3 and 3. Consequently, the numerator of the new coefficient is three times the difference between the arithmetic

mean and the median and its denominator is the standard deviation.

$$A_{P_2} = \frac{3\left[E(X) - Mdn(X)\right]}{\sqrt{E\left[(x - E(X))^2\right]}} = \frac{3\left[\mu_X - Mdn(X)\right]}{\sqrt{E\left[(x - \mu_X)^2\right]}} = \frac{3\left[\mu_X - Mdn(X)\right]}{\sigma_X} \qquad (1)$$

$E(X) = \mu_X$ = arithmetic mean or mathematical expectation of X.

$Mdn(X) = Q_X(p = 0.5)$ = median or quantile of order 0.5 of X.

If X is a continuous random variable with probability density function $f_X(x)$, its median is a value x within the support of X that satisfies the following condition:

$$Mdn(X) = \{x \in X \mid P(X \leq x) = 0.5\}$$

If X is a discrete random variable with probability mass function $f_X(x) = P(X = x)$, its median is a real value x (between the maximum and minimum of the support of X) that satisfies the following double condition:

$$x = \{x_i\}_{i=1}^{n} = \{x_1, x_2, \cdots, x_n\} \subseteq X$$

$$mdn(x) = \{\tilde{x} \in \mathbb{R} \mid P(X \leq \tilde{x}) \geq 0.5 \land P(X \geq \tilde{x}) \geq 0.5\}$$

In case more than one value of X meets the condition, the average of these values is taken as the median, which may result in a number with decimals outside the support of X. However, the median value will be greater than or equal to the minimum value of X and less than or equal to the maximum value of X.

$$\sigma_X = \sqrt{E\left[(X - E(X))^2\right]} = \text{standard deviation of } X.$$

when skewness is calculated on a sample of n data of X, the sample mean ($\bar{x}$), sample median (mdn), and the sample standard deviation (s_{n-1}) are used.

$$Sk_{P_2} = \frac{3\left[\bar{x} - mdn(x)\right]}{s_{n-1}(x)} = \frac{3\left[\dfrac{\sum_{i=1}^{n} x_i}{n} - mdn(x)\right]}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}} \qquad (2)$$

If the number of data is odd and the values are ordered in ascending order, $x_{(i)}$, the sample median is the value in the order $i = n/2$, that is, in the central position.

$$mdn(x) = x_{\left(i = \frac{n}{2}\right)}$$

If the number of data is even, the median is the average of the two values in the center.

$$mdn(x) = \frac{x_{\left(i = \frac{n-1}{2}\right)} + x_{\left(i = \frac{n+1}{2}\right)}}{2}$$

In the first case, the median is an x-value belonging to the support of X. In the second case, being an average of two values, the median can be an integer or a number with decimal places; if X is an ordinal variable or a discrete quantitative

variable with support in the set of natural numbers N or integers Z, the median value would be a rational number (with 0.5 as decimal) outside the support of X.

The median can also be defined as follows, using the indicator function ($1X$) or the ratio of the number of elements meeting a condition in the sample to the total number of elements in the sample.

$$x = \left\{ x_i \right\}_{i=1}^{n} = \left\{ x_1, x_2, \cdots, x_n \right\} \subseteq X$$

$$mdn(x) = \left\{ \tilde{x} \in \mathbb{R} \left| 1_{(x \leq \tilde{x})} \geq 0.5 \wedge 1_{(x \geq \tilde{x})} \geq 0.5 \right. \right\}$$

$$= \left\{ \tilde{x} \in \mathbb{R} \left| \frac{\#(x \leq \tilde{x})}{n} \geq 0.5 \wedge \frac{\#(x \geq \tilde{x})}{n} \geq 0.5 \right. \right\}$$

cardinality or number of elements in the sample that meet the condition.

Statistical manuals point out that there is no rule of thumb or cut-off point for interpreting this coefficient [4]. However, it is suggested that values of 0 plus minus three tenths in Sk_{P2} may reflect symmetry with medium sample sizes (100 to 200). Values below −0.3 may indicate negative skewness and values greater than 0.3 may reflect positive skewness. With small sample sizes ($n < 100$), an estimation error greater than three tenths should be considered and, with large sample sizes ($n > 200$), a smaller estimation error should be contemplated, because the larger the sample size, the more accurate the estimate [29].

This measure of skewness is bounded between −3 and 3 at the population level, but can take values outside this interval with sample data [9]. Singh *et al.* [6] estimated the Sk_{P2} statistic with bootstrap confidence intervals at 90%, 95%, and 99%. They made 10,000 draws at random with replacement from a normal distribution for four different sample sizes: 25, 50, 75, and 100. Consequently, their estimates are valid for unimodal distributions, especially with a bell-shaped profile in the histogram or bar chart. Table 1 shows their results. Values of Sk_{P2} within the interval reflect symmetry, below negative skewness, and above positive skewness. For its use, the enlisted n value closest to the empirical sample size should be sought or an approximation by linear interpolation should be made. Since the enlisted samples are less than or equal to 100, the interval should be

Table 1. Bootstrap confidence intervals at 90%, 95% and 99% of Sk_{P2} statistic for samples with four different sizes drawn from a normal distribution.

n	90% *CI*		95% *CI*		99% *CI*	
	LL	*UL*	*LL*	*UL*	*LL*	*UL*
25	−0.746	0.750	−0.884	0.862	−1.131	1.093
50	−0.522	0.506	−0.614	0.608	−0.784	0.792
75	−0.433	0.427	−0.522	0.503	−0.678	0.661
100	−0.373	0.365	−0.440	0.440	−0.569	0.573

Note. n = sample size, LL = lower limit and UL = upper limit of the confidence interval (*CI*). Estimates based on quantiles of 10,000 samples of size n drawn with replacement from a normal distribution. Source: Singh *et al.* [6].

used with a 90% confidence level to compensate for the conservative nature towards the null hypothesis of non-significance of the bootstrap resampling method with small sample [30]. It should be noted that the width of these intervals is larger than six, which coincides with the previous statement that an estimation error larger than 0.3 should be considered with samples smaller than 100, that is, with small samples.

One can always opt for interval estimation using the programming of the R package for bootstrapping [10] [11]. It is recommended to specify a confidence level of 90% with small and medium empirical data samples, and 95% with large samples, use a large number of draws with replacement, such as 1000, and apply nonparametric methods, such as bias-corrected accelerated bootstrap (type = "bca"), when the distribution of the variable is unknown or non-normal. In case of normality, a more efficient option is the parametric normal method: type = "norm" [31]. Furthermore, it is suggested to complement the evaluation of the symmetry of the sample data with graphical representations by means of a frequency histogram and a box-and-whisker plot [23] [32]). In case the bootstrap confidence interval includes 0, it could be affirmed that the distribution is symmetric with an alpha level of significance (0.1 or 0.05).

```
library(modeest)
x<-c(numeric sample data separated by commas)
boxplot(x, horizontal = TRUE, col = " lightgrey")
hist(x, breaks = "freedman-diaconis", col = " lightgrey", border = "black")
m = mean(x)
mdn = median(x)
s = sd(x)
skpII = 3*(m-mdn)/s
cat("Pearson's skewness coefficient II = ",skpII,"\n")
library(boot)
b<-boot(data = x, function(x, i) {3*(mean(x[i])-median(x[i]))/sd(x[i])}, R = 1000)
b
plot(b)
boot.ci(b, conf = 0.90, type = c("bca", "norm"), digits = 4)
```

3.2. Yule's Skewness Coefficient

There is a version of this measure without multiplying the numerator by three that is usually named Yule's skewness coefficient. Although Pearson justified tripling the numerator to have a population-level bounded measure, the only thing this factor does is to generate a change of scale. At the population level, the range of Yule's measure is −1 to 1 [9], like the interquartile and percentile skewness coefficients, which is a more convenient range than −3 to 3. From this new bounding, without multiplying the numerator by three, the suggested interpretation for symmetry would be a value in the interval [−0, 1, 0.1] with intermediate samples [4].

Cabilio and Masaro [8] showed that the asymptotic distribution of the Sk_Y statistic under the null hypothesis of symmetry is a normal distribution when the

variable X has a known distribution with finite means and variances.

$$H_0 : \tau = \mu_X - Mdn(X) = 0$$

$$\sqrt{n}\left[\mu_X - Mdn(X)\right] = \tau\sqrt{n} \xrightarrow{d} N\left(\mu_{\tau\sqrt{n}} = 0, \sigma^2_{\tau\sqrt{n}} = \sigma^2_{H_0}(\tau, F)\right)$$

$$\tau = \mu_X - Mdn(X) \xrightarrow{d} N\left(\mu_\tau = 0, \sigma^2_\tau = \frac{\sigma^2_{H_0}(\tau, F)}{n}\right)$$

n = number of independent samples of X

$$X \sim F(\theta), F_X(x|\theta) = \int_{-\infty}^{x} f_X(x|\theta)\,\mathrm{d}x$$

$$Mdn(X) = \tilde{x}; \sigma^2_{H_0}(\tau, F) = \sigma^2_X + \frac{1}{4f_X^2(\tilde{x})} - \frac{\tau}{f_X(\tilde{x})}; \ \tau = \mu_X - 2\int_{-\infty}^{\tilde{x}} x\,\mathrm{d}F_X(x)$$

$$H_0 \rightarrow \mu_X = Mdn(X)$$

$$\sigma^2_{H_0}(\tau, F) = 1 + \frac{1}{4\sigma_X^2 f_X^2(0)} - \frac{1}{\sigma_X f_X(0)} E\left(\left|\frac{X - \mu_X}{\sigma_X}\right|\right)$$

In the case of a standard normal distribution, the asymptotic error is approximately $\sqrt{0.5708/n}$.

$$f_X^2(0) = \sqrt{1/(2\pi)}$$

$$\sigma^2_{H_0}(\tau, F) = 1 + \frac{1}{4 \times 1 \times 1/(2\pi)} - \frac{\sqrt{2/\pi}}{1 \times \sqrt{1/(2\pi)}} = \frac{\pi}{2} - 1 = \frac{\pi - 2}{2} \approx 0.5708$$

$$P\left(\overline{x} - mdn - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi - 2}{2n}} \leq \mu_X - Mdn(X) \leq \overline{x} - mdn - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi - 2}{2n}}\right) = 1 - \alpha$$

In the case of a non-standard normal distribution, the asymptotic error and interval can be calculated using the following formulas:

$$\sigma^2_{H_0}(\tau, F) = 1 + \frac{1}{4 \times \sigma_X^2 \times 1/(2\pi)} - \frac{\sqrt{2/\pi}\sigma_X}{\sigma_X \times \sqrt{1/(2\pi)}} = 1 + \frac{2}{\pi\sigma_X^2} - 2 = \frac{2}{\pi\sigma_X^2} - 1$$

$$P\left(\overline{x} - mdn - z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{n}\left(\frac{2}{\pi\sigma_X^2} - 1\right)} < \mu_X - Mdn(X) < \overline{x} - mdn - z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{n}\left(\frac{2}{\pi\sigma_X^2} - 1\right)}\right) = 1 - \alpha$$

For any type of finite moment distribution, the asymptotic error estimator is given by the following mathematical expression:

$$\hat{\sigma}_{H_0} = \sqrt{\frac{1}{n}\left(s_{n-1}^2 + \frac{1}{4f_n^2(mdn)} - \frac{\sum_{i=1}^{n}|x_i - mdn|/n}{f_n(mdn)}\right)} \tag{3}$$

f_n = Kernel density estimator from the n sample data. Cabilio and Masaro [8] recommend using a uniform Kernel function and a bandwidth equal to $n^{-0.2}$:

$$f_n(mdn) = \frac{1}{2nh}\max\left(1, \sum_{i=1}^{n} I\left[mdn - h \leq x_i \leq mdn - h\right]\right), h = \frac{1}{\sqrt[5]{n}}$$

$$P\left(\overline{x} - mdn - z_{1-\frac{\alpha}{2}}\hat{\sigma}_{H_0} < \mu_X - Mdn(X) < \overline{x} - mdn - z_{1-\frac{\alpha}{2}}\hat{\sigma}_{H_0}\right) = 1 - \alpha \tag{4}$$

The script for calculating this asymptotic confidence interval with the R program is as follows with a concrete example:

library("kdensity")

```
x<- c(93, 60, 55, 91, 107, 86, 113, 90, 97, 92, 74, 144, 93, 116, 90, 87, 125, 89, 84, 104, 92,
109, 77, 77, 81, 124, 112, 125, 105, 128, 103, 96, 98, 91)
length(x)
mean(x)
median(x)
sd(x)
library("DescTools")
MeanAD(x, center = median)
bw <-(34**-0.2)
bw
kde = kdensity(x, bw, kernel = "uniform", normalized = FALSE)
kde(93)
```

Output in the order given by the instructions in the script: $n = 34$, $m = 97.2941$, $mdn = 93$, $s_{n-1} = 19.1684$, $mad\_mdn = 14.4118$, $h = 0.4940$, and $f_n(mdn) = 0.0344$. The Yule's skewness measure is computed with Equation (2), its asymptotic standard error with Equation (3), and the asymptotic confidence interval with Equation (4).

$$\bar{x} - mdn = 97.2941 - 93 = 4.2941$$

$$Sk_Y = \frac{\bar{x} - mdn}{s_{n-1}} = \frac{97.2941 - 93}{19.1684} = \frac{4.2941}{19.1684} = 0.2240$$

$$\hat{\sigma}_{H_0} = \sqrt{\frac{1}{n}\left(s_{n-1}^2 + \frac{1}{4 f_n^2(mdn)} - \frac{\sum_{i=1}^n |x_i - mdn|/n}{f_n(mdn)}\right)}$$

$$= \sqrt{\frac{1}{34}\left(19.1684^2 + \frac{1}{4 \times 0.0344^2} - \frac{14.4118}{0.0344}\right)} = 2.1676$$

$$P\left(\bar{x} - mdn - z_{1-\frac{\alpha}{2}}\hat{\sigma}_{H_0} < \mu_X - Mdn(X) < \bar{x} - mdn - z_{1-\frac{\alpha}{2}}\hat{\sigma}_{H_0}\right) = 1 - \alpha$$

Due to the small sample size, an alpha value of 0.1 is used for the asymptotic confidence interval. It is assumed that the sample was drawn randomly from a probability distribution with finite moments.

$$P\left(4.2941 - 1.6449 \times 2.1676 < \mu_X - Mdn(X) < 4.2941 - 1.6449 \times 2.1676\right) = 0.90$$

$$P\left[\mu_X - Mdn(X) \in (0.0458, 8.5425)\right] = 0.90$$

$$0 \notin (0.7288, 7.8595) \Rightarrow \mu_X \neq Mdn(X) \text{ with } \alpha = 0.1$$

$$P\left[\frac{\mu_X - Mdn(X)}{\sigma} = Sk_Y \in \left(\frac{0.7288}{19.1684}, \frac{7.8595}{19.1684}\right) = (0.0380, 0.4100)\right] = 0.90$$

$$0 \notin (0.0380, 0.4100) \Rightarrow Sk_Y \neq 0 \text{ with } \alpha = 0.1$$

The asymptotic confidence interval requires a large sample. Another option to

interpret whether or not there is symmetry is to generate a bootstrap confidence interval at 90% (with small and medium samples) or 95% (with large samples) with 1000 draws with replacement by the bias-corrected and accelerated percentile method. Even the normal method is more appropriate and efficient if the sample data are normally distributed. The instructions for the R program are the following, in which the evaluation with graphs is added [10] [33]. If the bootstrap confidence interval includes 0 and the plots show symmetric shoulders and tails, the null hypothesis of symmetry would hold with a significance level of alpha (0.1 or 0.05). This is not the case for the sample in this example, which show positive skewness with an outlier in the right tail, as the box-and-whisker plot reveals (Figure 2).

```
library(modeest)
x<- c(93, 60, 55, 91, 107, 86, 113, 90, 97, 92, 74, 144, 93, 116, 90, 87, 125, 89, 84, 104, 92,
109, 77, 77, 81, 124, 112, 125, 105, 128, 103, 96, 98, 91)
boxplot(x, horizontal = TRUE, col = "lightgrey")
hist(x, breaks = "freedman-diaconis", col = "lightgrey", border = "black")
m = mean(x)
mdn = median(x)
s = sd(x)
skY = (m-mdn)/s
cat("Yule's skewness coefficient = ",skY,"\n")
library(boot)
b<-boot(data = x, function(x, i) { (mean(x[i])-median(x[i]))/sd(x[i])}, R = 1000)
b
plot(b)
boot.ci(b, conf = 0.90, type = c("bca", "norm"), digits = 4)
```
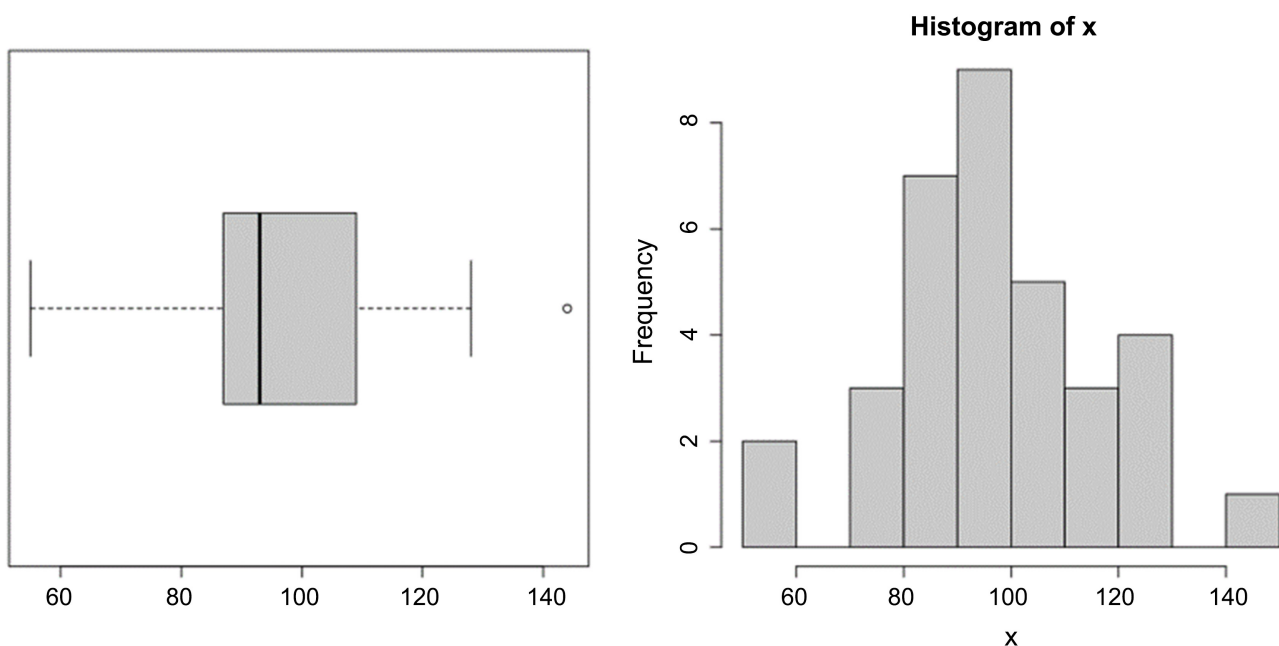

Figure 2. Box-plot and histogram of frequencies of the sample of 34 data.

Output in the order given by the instructions in the script:

Yule's skewness coefficient = 0.2240.

Bootstrap statistics: bias = −0.0647 and standard error = 0.1266.

Bootstrap confidence interval at 90% using bias-corrected and accelerated percentile method: (0.0651, 0.4699).

Bootstrap confidence interval at 90% using normal method: (0.0806, 0.4970).

4. Bootstrap Confidence Intervals for the Sk_Y Statistic from Normally Distributed Samples-Population of Different Sizes

As reiterated in various scientific publications, there are no general cut-off points to establish whether symmetry is present when using this measure [3] [4]. Bootstrap confidence intervals at 90%, 95% and 99% using normal method are shown in Table 2. Fifty-seven normally distributed samples-population were generated with finite sizes and a range from $z = -3.5$ ($F_Z[z = -3.5] = \Phi(-3.5) = 0.000232629$), to 3.5 ($F_Z[z = 3.5] = \Phi(3.5) = 0.999767371$). This range was chosen to ensure mesokurtosis and increase the efficiency of the bootstrap confidence intervals, since it is rare for data to appear more than three standard deviations away from the mean in a normal distribution [27]. The range corresponds

Table 2. Parametric bootstrap confidence intervals for the Sk_Y statistics.

N	Sk_Y	Bias	SE	90% CI		95% CI		99% CI	
				LL	UL	LL	UL	LL	UL
10	0	−0.0035	0.2180	−0.3551	0.3622	−0.4238	0.4309	−0.5581	0.5652
15	0	0.0005	0.1902	−0.3134	0.3124	−0.3734	0.3723	−0.4905	0.4895
20	0	0.0019	0.1556	−0.2578	0.2540	−0.3069	0.3031	−0.4027	0.3989
25	0	−0.0016	0.1481	−0.2420	0.2452	−0.2886	0.2919	−0.3798	0.3831
30	0	−0.0008	0.1297	−0.2125	0.2140	−0.2534	0.2549	−0.3332	0.3347
35	0	0.0009	0.1255	−0.2072	0.2055	−0.2468	0.2450	−0.3240	0.3223
40	0	0.0000	0.1136	−0.1868	0.1869	−0.2226	0.2227	−0.2926	0.2926
45	0	0.0002	0.1122	−0.1848	0.1843	−0.2201	0.2197	−0.2892	0.2888
50	0	−0.0025	0.1033	−0.1674	0.1725	−0.2000	0.2050	−0.2636	0.2687
55	0	0.0003	0.1011	−0.1665	0.1660	−0.1983	0.1978	−0.2606	0.2600
60	0	0.0003	0.0943	−0.1554	0.1547	−0.1851	0.1844	−0.2431	0.2425
65	0	−0.0012	0.0921	−0.1503	0.1528	−0.1794	0.1818	−0.2361	0.2386
70	0	−0.0001	0.0876	−0.1439	0.1442	−0.1715	0.1718	−0.2254	0.2257
75	0	−0.0004	0.0857	−0.1407	0.1414	−0.1677	0.1685	−0.2205	0.2213
80	0	−0.0002	0.0820	−0.1347	0.1350	−0.1605	0.1608	−0.2110	0.2113
85	0	−0.0006	0.0812	−0.1330	0.1342	−0.1586	0.1598	−0.2086	0.2099
90	0	−0.0004	0.0775	−0.1270	0.1278	−0.1514	0.1522	−0.1991	0.1999
95	0	0.0000	0.0765	−0.1258	0.1258	−0.1499	0.1499	−0.1970	0.1970
100	0	−0.0006	0.0739	−0.1210	0.1222	−0.1443	0.1454	−0.1898	0.1910

Continued

105	0	0.0000	0.0732	−0.1204	0.1204	−0.1435	0.1434	−0.1885	0.1885
110	0	−0.0002	0.0717	−0.1177	0.1181	−0.1403	0.1407)	−0.1844	0.1848
115	0	0.0001	0.0703	−0.1157	0.1154	−0.1378	0.1376	−0.1811	0.1808
120	0	−0.0008	0.0679	−0.1109	0.1125	−0.1323	0.1339	−0.1742	0.1758
125	0	0.0006	0.0668	−0.1105	0.1092	−0.1315	0.1303	−0.1727	0.1714
130	0	0.0006	0.0650	−0.1076	0.1063	−0.1281	0.1268	−0.1681	0.1668
135	0	−0.0006	0.0647	−0.1059	0.1071	−0.1263	0.1275	−0.1662	0.1674
140	0	0.0010	0.0619	−0.1028	0.1008	−0.1223	0.1203	−0.1604	0.1584
145	0	−0.0003	0.0633	−0.1038	0.1044	−0.1237	0.1243	−0.1627	0.163
150	0	−0.0002	0.0606	−0.0994	0.0998	−0.1185	0.1189	−0.1558	0.1562
155	0	−0.0009	0.0600	−0.0978	0.0995	−0.1167	0.1185	−0.1537	0.1550
160	0	0.0004	0.0585	−0.0966	0.0958	−0.1150	0.1142	−0.151	0.1502
165	0	−0.0008	0.0591	−0.0963	0.098	−0.1149	0.1166	−0.1513	0.1529
170	0	−0.0006	0.0577	−0.0943	0.0955	−0.1124	0.1136	−0.1479	0.1491
175	0	0.0003	0.0566	−0.0933	0.0927	−0.1111	0.1106	−0.1459	0.1454
180	0	0.0002	0.0559	−0.0922	0.0917	−0.1098	0.1093	−0.1442	0.1438
185	0	−0.0004	0.0553	−0.0906	0.0914	−0.1080	0.1088	−0.1421	0.1429
190	0	0.0002	0.0541	−0.0892	0.0888	−0.1062	0.1059	−0.1395	0.1392
195	0	0.0008	0.0540	−0.0897	0.0881	−0.1067	0.1051	−0.14	0.1384
200	0	−0.0007	0.0528	−0.0862	0.0875	−0.1028	0.1042	−0.1353	0.1367
210	0	0.0004	0.0517	−0.0854	0.0845	−0.1017	0.1008	−0.1335	0.1326
220	0	0.0007	0.0501	−0.083	0.0817	−0.0988	0.0975	−0.1297	0.1283
230	0	0.0006	0.0496	−0.0822	0.0810	−0.0978	0.0966	−0.1284	0.1272
240	0	−0.0004	0.0484	−0.0792	0.0800	−0.0944	0.0952	−0.1242	0.1250
250	0	0.0008	0.0477	−0.0794	0.0777	−0.0944	0.0928	−0.1238	0.1222
260	0	−0.0007	0.0466	−0.0760	0.0774	−0.0906	0.0921	−0.1194	0.1208
270	0	0.0002	0.0458	−0.0755	0.0751	−0.0899	0.0895	−0.1181	0.1177
280	0	<0.0001	0.0443	−0.0729	0.0730	−0.0869	0.087	−0.1142	0.1143
290	0	−0.0001	0.0440	−0.0723	0.0724	−0.0861	0.0862	−0.1132	0.1133
300	0	0.0004	0.0430	−0.0711	0.0703	−0.0846	0.0838	−0.1111	0.1103
320	0	<0.0001	0.0419	−0.0689	0.0689	−0.0821	0.0821	−0.1079	0.1079
340	0	0.0005	0.0404	−0.0669	0.0659	−0.0796	0.0786	−0.1045	0.1034
360	0	−0.0002	0.0394	−0.0646	0.0650	−0.0770	0.0774	−0.1012	0.1017
380	0	0.0001	0.0388	−0.0640	0.0638	−0.0762	0.076	−0.1001	0.0999
400	0	−0.0006	0.0370	−0.0603	0.0614	−0.0720	0.0731	−0.0947	0.0959
450	0	0.0007	0.0355	−0.0591	0.0577	−0.0703	0.0689	−0.0922	0.0908
500	0	−0.0002	0.0334	−0.0547	0.0551	−0.0652	0.0656	−0.0858	0.0862
1000	0	0.0001	0.0238	−0.0392	0.0390	−0.0467	0.0465	−0.0613	0.0611

Note. N = population size, Sk_Y = Yule's skewness coefficient, SE = standard error, CI = confidence interval, LL = lower limit and UL = upper limit of the bootstrap confidence interval using normal method.

to eight times the interquartile range and about seven times the standard deviation, as considered by the Rice University rule [34] for determining the number of class interval in the histogram ($k = \lceil 2 \times n^{1/3} \rceil$) in relation to the Freedman-Diaconis and Scott rules. This approach is especially relevant for a variable with normal distribution:

If there is symmetry, the interquartile range is twice the semi-interquartile range: $R_{IQ} = P_{75} - P_{25} = 2 \times R_{SIQ} = (P_{75} - P_{25})/2$. The number of class intervals (k) is obtained by dividing the range or difference of the maximum and minimum ($R = \max - \min$) by the width of the intervals (h). Here the widths are obtained using the Freedman-Diaconis (h_{FD}) and Scott (h_{Scott}) rules:

$$x = \{x_i\}_{i=1}^n = \{x_1, x_2, \cdots, x_n\}$$

$$k = \frac{R(x)}{h_{FD}} = \frac{\max(x) - \min(x)}{\dfrac{2[P_{75}(x) - P_{25}(x)]}{\sqrt[3]{n}}} \approx \frac{4[P_{75}(x) - P_{25}(x)]}{\dfrac{2[P_{75}(x) - P_{25}(x)]}{\sqrt[3]{n}}}$$

$$= \frac{8\left[\dfrac{P_{75}(x) - P_{25}(x)}{2}\right]}{\dfrac{2[P_{75}(x) - P_{25}(x)]}{\sqrt[3]{n}}} = 2\sqrt[3]{n}$$

$$k = \frac{R(x)}{h_{SCott}} = \frac{\max(x) - \min(x)}{\dfrac{3.49 \times s_{n-1}(x)}{\sqrt[3]{n}}} \approx \frac{6.98 \times s_{n-1}(x)}{\dfrac{3.49 \times s_{n-1}(x)}{\sqrt[3]{n}}} = 2\sqrt[3]{n}$$

The population sizes N ranged from 10 to 200 in 5-data increments, from 210 to 300 in 10-data increments, from 320 to 400 in 20-data increments, from 450 to 500 in 50-data increments, and ending with a size of 1000. Population data were obtained with the probit function: $z_k = \Phi^{-1}\left(P_k = 0.000232629 + (k-1) \times \Delta P\right)$, where $k = 1, 2, \cdots, N$ and $\Delta P = (0.999767371 - 0.000232629)/(N-1)$. For example, for population size 10: $\Delta P = 0.999534742/9 = 0.111059416$.

$Z(N = 10) = \{z_1 = \Phi^{-1}\left(P_1 = 0.000232629\right) = -3.5,$

$z_2 = \Phi^{-1}\left(P_2 = 0.000232629 + 0.111059416 = 0.111292045\right) = -1.219685581,$

$z_3 = \Phi^{-1}\left(P_3 = 0.000232629 + 2 \times 0.111059416 = 0.222351461\right) = -0.76427577,$

$z_4 = \Phi^{-1}\left(P_4 = 0.000232629 + 3 \times 0.111059416 = 0.333410876\right) = -0.430514044,$

$z_5 = \Phi^{-1}\left(P_5 = 0.000232629 + 4 \times 0.111059416 = 0.444470292\right) = -0.139644873,$

$z_6 = \Phi^{-1}\left(P_6 = 0.000232629 + 5 \times 0.111059416 = 0.555529708\right) = 0.139644873,$

$z_7 = \Phi^{-1}\left(P_7 = 0.000232629 + 6 \times 0.111059416 = 0.666589124\right) = 0.430514044,$

$z_8 = \Phi^{-1}\left(P_8 = 0.000232629 + 7 \times 0.111059416 = 0.777648539\right) = 0.76427577,$

$z_9 = \Phi^{-1}\left(P_9 = 0.000232629 + 8 \times 0.111059416 = 0.888707955\right) = 1.219685581,$

$z_{10} = \Phi^{-1}\left(P_{10} = 0.000232629 + 9 \times 0.111059416 = 0.999767371\right) = 3.5\}.$

From these 57 samples-population, 10,000 bootstrap samples were drawn with replacement to calculate the Yule's skewness coefficient, and under a model of a

normal distribution were computed confidence intervals at 90%, 95%, and 99%. The instructions for the R program are as follows and the result is shown in **Table 2**. As an example, **Figure 3** shows that the fit to normality of the distribution of the Sk_Y statistics calculated in 10,000 different random samples drawn from the 10-data population-sample is good. Its histogram has the expected symmetric, mesokurtic, bell-shaped profile and the coordinate points are aligned at 45 degrees on the normal Q-Q plot.

```
x10<-c(−3.5, −1.219685581, −0.76427577, −0.430514044, −0.139644873, 0.139644873,
0.430514044, 0.76427577, 1.219685581, 3.5)
m = mean(x10)
mdn = median(x10)
s = sd(x10)
AY = (m-mdn)/s
cat("Coefficient de Yule = ",AY,"\n")
length(x10)
library(boot)
b<-boot(data = x10, function(x10, i) {(mean(x10[i])-median(x10[i]))/sd(x10[i])}, R =
10000)
b
boot.ci(b, conf = c(0.90, 0.95, 0.99), type = "norm", digits = 4)
plot(b)
```
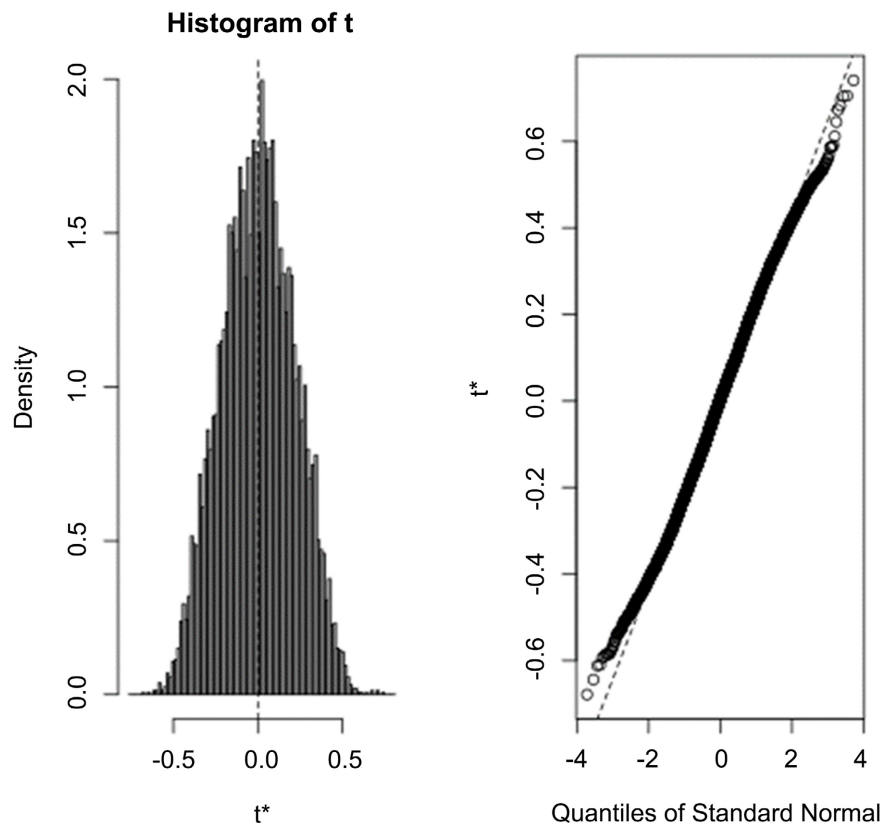

Figure 3. Histogram and normal Q-Q plot of Sk_Y statistics in the 10,000 different random samples drawn from the sample-population of 10 normally distributed data.

Table 3. Comparison of the widths of the 90% confidence intervals between Singh *et al.*'s study and the present study.

n	Singh *et al.*'s study [6]					Present study			$h_P - h_Y$
	LL_P	UL_P	$LL_P/3$	$UL_P/3$	h_P	LL_Y	UL_Y	h_Y	
25	−0.746	0.75	−0.249	0.250	0.499	−0.242	0.245	0.487	0.011
50	−0.522	0.506	−0.174	0.169	0.343	−0.167	0.173	0.340	0.003
75	−0.433	0.427	−0.144	0.142	0.287	−0.141	0.141	0.282	0.005
100	−0.373	0.365	−0.124	0.122	0.246	−0.121	0.122	0.243	0.003

Note. n = sample size, LI_P and UL_P = lower and upper limit of the 90% confidence interval for Sk_{P2} from the study of Singh *et al.* (2019), $h_P = LL_P/3 - UL_P/3$, LL_Y and UL_Y = lower and upper limit of the 90% confidence interval for Sk_Y from the present study $h_Y = LL_Y - UL_Y$.

Returning to the previous example with the sample of 34 data, the value of Yule's skewness coefficient ($Sk_Y = 0.2240$) falls outside the 90% bootstrap confidence interval (−0.2083, 0.2072) corresponding to 34 data. This confidence interval is obtained by linear interpolation between the 90% bootstrap confidence interval (−0.2125, 0.2140) corresponding to 30 data and the 90% bootstrap confidence interval (−0.2072, 0.2055) corresponding to 35 data shown in **Table 2**. Therefore, the skewness coefficient is significantly greater than 0 at the 10% significance level, indicating a right-skewed distribution.

$$LI = -0.2125 + \frac{34 - 30}{35 - 30}\left(-0.2072 - \left(-0.2125\right)\right) = -0.2083$$

$$LS = 0.2140 + \frac{34 - 30}{35 - 30}\left(0.2055 - 0.2140\right) = 0.2072$$

$$Sk_Y = 0.2240 \notin \left(-0.2083, 0.2072\right) \Rightarrow Sk_Y > 0$$

If the confidence intervals reported by Singh *et al.* [6] for the Sk_{P2} statistic were divided by three, their estimates could be compared with those of the present study for the Sk_Y statistic. **Table 3** shows the difference in width for the 90% bootstrap confidence intervals between the two studies. It is observed that they are very similar, although the width is slightly smaller in the present study.

5. Conclusions

Interval estimation methods can be classified into three categories [35]: *exact* methods are used when the sampling distribution of the statistic can be determined; *asymptotic* methods require large samples and are based on certain theorems, such as the additive or multiplicative Central Limit Theorem; and bootstrap methods are nonparametric approaches that do not rely on specific assumptions about the population distribution. They are suitable for small sample sizes or when population distributions are unknown. The key idea is to resample the original data with replacement to create multiple bootstrap samples. Confidence intervals are then estimated from the distribution of statistics calculated

from these samples. Bootstrap methods allow more flexibility, but require more computational power. They are often used in situations where other methods may be unreliable or when assessing the stability and variability of the estimates. The latter two methods have been seen in this article for estimating the Sk_Y statistic, but the study is based on the latter. It ensures that bootstrap confidence intervals are centered at 0 and serve as an interpretive guide to symmetry, as each bootstrap sample is generated from a strictly symmetric sample-population with a characteristically normal range (-3.5 to 3.5), which may be the most innovative aspect of the work.

There are many types of continuous distributions. One subtype is symmetric; within these, another subtype is unimodal with its axis of symmetry at the mode or peak. A specific case of this last subtype is the family of normal distributions, which is characterized not only by symmetry, but also by mesokurtosis or mean tails [36]. From the computed bootstrap confidence intervals, it can be seen that the ∓ 0.1 rule as an interval for null skewness ($Sk_Y = 0$) applies for samples of at least 150 data with a significance level of 10% and at least 220 data with a significance level of 5%. With a sample size as small as 10, the confidence interval at 90% is (-0.355, 0.362) and at 95% is (-0.424, 0.431), and the interval at 95% is less than ∓ 0.05 with a large sample of 1000. It is important to remark that these intervals complemented by the histogram allow us to assess whether the data are symmetrical compared to a normal distribution. They can be extended by linear interpolation for intermediate sizes or better run the confidence interval calculation using bootstrap sampling with the R program, which is a freely available program that has been developed by the mathematical community since 1997 [37]. This program is available for online calculations at https://rdrr.io/snippets or can be downloaded for installation on a personal computer at https://cran.r-project.org. It should also be noted that, with the samples-population generated with data n normally distributed and range -3 to 3, all bootstrap confidence intervals of the Sk_Y are perfectly centered at 0, and the estimation efficiency seems to be slightly better than without this constraint.

The Yule's coefficient is very simple to calculate and applies to any type of quantitative data, with which normality can be assessed. In addition, it is a clear measure of asymmetry to interpret in relation to the proximity to normality or deviation from normality by asymmetry. We speak of quantitative data and not of any type of distribution, since there is the case of the Cauchy's distribution which is symmetrical, but with very heavy tails and without finite moments, that is, without arithmetic mean and moments of higher order. For this distribution, which is clearly far from normality, the Yule's skewness coefficient is inadequate, and the interquartile coefficient would be the alternative. With ordinal data, if they are assumed to be points of a continuous bipolar distribution, as is done in the polychoric correlation, especially if they have a wide range, this assessment of normality would also be possible [38]. With qualitative data, another approach is applied to assess skewness and kurtosis [17].

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Orcan, F. (2020) Parametric or Non-Parametric: Skewness to Test Normality for Mean Comparison. *International Journal of Assessment Tools in Education*, **7**, 255-265. https://doi.org/10.21449/ijate.656077

[2] Pearson, K. (1895) X. Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London A*, **186**, 343-414. https://doi.org/10.1098/rsta.1895.0010

[3] Bruni, V., and Vitulano, D. (2020) SSIM Based Signature of Facial Micro-Expressions. *Proceedings of the Image Analysis and Recognition*: 17*th International Conference*, Póvoa de Varzim, 24-26 June 2020, 267-279. https://doi.org/10.1007/978-3-030-50347-5_24

[4] Doane, D.P. and Seward, L.E. (2011) Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*, **19**, Article No. 18. https://doi.org/10.1080/10691898.2011.11889611

[5] Mohammed, M.B., Adam, M.B., Ali, N. and Zulkafli, H.S. (2022) Improved Frequency Table's Measures of Skewness and Kurtosis with Application to Weather Data. *Communications in Statistics—Theory and Methods*, **51**, 581-598. https://doi.org/10.1080/03610926.2020.1752386

[6] Singh, A., Gewali, L. and Khatiwada, J. (2019) New Measures of Skewness of a Probability Distribution. *Open Journal of Statistics*, **9**, 601-621. https://doi.org/10.4236/ojs.2019.95039

[7] Eberl, A. and Klar, B. (2020) Asymptotic Distributions and Performance of Empirical Skewness Measures. *Computational Statistics & Data Analysis*, **146**, Article ID: 106939. https://doi.org/10.1016/j.csda.2020.106939

[8] Cabilio, P. and Masaro, J. (1996) A Simple Test of Symmetry about an Unknown Median. *Canadian Journal of Statistics*, **24**, 349-361. https://doi.org/10.2307/3315744

[9] Majindar, K.N. (1962) Improved Bounds on a Measure of Skewness. *Annals of Mathematical Statistics*, **33**, 1192-1194. https://doi.org/10.1214/aoms/1177704482

[10] Canty, A. and Ripley, B. (2022) Boot: Bootstrap R (S-Plus) Functions. R Package Version 1.3-28.1. https://cran.r-project.org/web/packages/boot/boot.pdf

[11] Tibshirani, R., Leisch, F. and Kostyshak, S. (2022) Package "Bootstrap". https://cran.r-project.org/web/packages/bootstrap/bootstrap.pdf

[12] Galton, F. (1883) Enquiries into Human Faculty and Its Development. Macmillan and Company, London. https://doi.org/10.1037/14178-000

[13] Pearson, K. (1894) Contributions to the Mathematical Theory of Evolution. I. On the Dissection of Asymmetrical Frequency Curves. *Philosophical Transactions of the Royal Society of London A*, **185**, 71-110. https://doi.org/10.1098/rsta.1894.0003

[14] Pearson, K. (1916) Mathematical Contributions to the Theory of Evolution. XIX.

Second Supplement to a Memoir on Skew Variation. *Philosophical Transactions of the Royal Society of London A*, **216**, 429-457. https://doi.org/10.1098/rsta.1916.0009

[15] Srivastava, R. (2023) Karl Pearson and "Applied" Statistics. *Resonance*, **28**, 183-189. https://doi.org/10.1007/s12045-023-1542-3

[16] DeVellis, R.F. and Thorpe, C.T. (2021) Scale Development: Theory and Applications. Sage Publications, Thousand Oaks.

[17] Moral de la Rubia, J. (2022) A Measure of One-Dimensional Asymmetry for Qualitative Variables. *Revista de Psicología* (*PUCP*), **40**, 519-551. https://dx.doi.org/10.18800/psico.202201.017

[18] Shi, J., Luo, D., Wan, X., Liu, Y., Liu, J., Bian, Z. and Tong, T. (2020) Detecting the Skewness of Data from the Sample Size and the Five-Number Summary.

[19] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C. and Keshri, A. (2019) Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, **22**, 67-72. https://doi.org/10.4103/aca.ACA_157_18

[20] Gupta, S.C. and Kapoor, V.K. (2020) Descriptive Measures. In: *Fundamentals of Mathematical Statistics*, 12th Edition, Sultan Chand & Sons, New Delhi, Section 2, 1-78.

[21] Altinay, G. (2016) A Simple Class of Measures of Skewness. Munich Personal RePEc Archive, Paper No. 72353, 1-13. https://mpra.ub.uni-muenchen.de/72353

[22] Sarka, D. (2021) Descriptive Statistics. In: *Advanced Analytics with Transact-SQL*, Apress, Berkeley, 3-29. https://doi.org/10.1007/978-1-4842-7173-5_1

[23] Hatem, G., Zeidan, J., Goossens, M. and Moreira, C. (2022) Normality Testing Methods and the Importance of Skewness and Kurtosis in Statistical Analysis. *BAU Journal—Science and Technology*, **3**, Article No. 7. https://doi.org/10.54729/KTPE9512

[24] Aytaçoğlu, B. and Sazak, H.S. (2017) A Comparative Study on the Estimators of Skewness and Kurtosis. *Ege University Journal of the Faculty of Science*, **41**, 1-13.

[25] Yule, G.U. (1912) An Introduction to the Theory of Statistics. Charles Griffin and Company Limited, London.

[26] Bickel, D.R. (2002) Robust Estimators of the Mode and Skewness of Continuous Data. *Computational Statistics & Data Analysis*, **39**, 153-163. https://doi.org/10.1016/S0167-9473(01)00057-3

[27] Kaliyadan, F. and Kulkarni, V. (2019) Types of Variables, Descriptive Statistics, and Sample Size. *Indian Dermatology Online Journal*, **10**, 82-86. https://doi.org/10.4103/idoj.IDOJ_468_18

[28] Chacón, J.E. (2020) The Modal Age of Statistics. *International Statistical Review*, **88**, 122-141. https://doi.org/10.1111/insr.12340

[29] Upton, G.J. and Cook, I. (2014) Pearson's Coefficient of Skewness. In: *Oxford Dictionary of Statistics*, 3rd Edition, Oxford University Press, Cambridge, 81-82.

[30] Efron, B. (2003) Second Thoughts on the Bootstrap. *Statistical Science*, **18**, 135-140. https://doi.org/10.1214/ss/1063994968

[31] Manly, B.F.J. and Navarro-Alberto, J.A. (2022) Randomization, Bootstrap and Monte Carlo Methods in Biology. 4th Edition, Chapman & Hall, Boca Raton.

[32] Rizzo, M. (2019) Statistical Computing with R. 2nd Edition, Chapman & Hall/CRC Press, Boca Raton.

[33] Braun, W.J. and Murdoch, D.J. (2021) A First Course in Statistical Programming with R. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108993456

[34] Lane, D.M. (2021) Histograms. In: *Online Statistics Education: A Multimedia Course of Study*, Department of Statistics, Rice University, Houston.
https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Lane)/02%3A_Graphing_Distributions/2.04%3A_Histograms

[35] DiCiccio, T.J., Ritzwoller, D.M., Romano, J.P. and Shaikh, A.M. (2022) Confidence Intervals for Seroprevalence. *Statistical Science*, **37**, 306-321.
https://doi.org/10.1214/21-STS844

[36] Mukhopadhyay, N. (2020) Probability and Statistical Inference. CRC Press, Boca Raton.

[37] Giorgi, F.M., Ceraolo, C. and Mercatelli, D. (2022) The R Language: An Engine for Bioinformatics and Data Science. *Life*, **12**, Article No. 648.
https://doi.org/10.3390/life12050648

[38] Lyhagen, J. and Ornstein, P. (2023) Robust Polychoric Correlation. *Communications in Statistics—Theory and Methods*, **52**, 3241-3261.
https://doi.org/10.1080/03610926.2021.1970770