Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

Research Article

Received: date:07.04.2025
Accepted: date:04.06.2025
Published: date:30.06.2025

Refined Normality Test Based on the Parametric Seven-Number Summary

Jose Moral de la Rubia[1*]

[1]School of Psychology, Universidad Autónoma de Nuevo León, Monterrey Mexico; jose.morald@uanl.edu.mx
Orcid: 0000-0003-1856-1458[1]
*Correspondence: jose.morald@uanl.edu.mx

Abstract: In 2022, a normality test based on the parametric seven-number summary was proposed. Its test statistic is the sum of squared standardized quantiles and was initially approximated by a chi-square distribution with seven degrees of freedom, without accounting for correlation among quantiles. Objective: To improve the test by incorporating these correlations. Two alternatives were proposed: (1) estimating the sampling distribution of the Q-statistic via bootstrap (Q_B), and (2) using a quadratic form with a correlation matrix of quantiles under normality (Q_T). Methods: The three variants (Q, Q_B, Q_T) were compared with the Shapiro-Wilk W-test in terms of accuracy (hit ratio) and statistical power. A total of 372 random samples were generated across 31 sample sizes from twelve continuous distributions. Correct classifications were compared using Cochran's Q test, and power was assessed via repeated-measures ANOVA. Results: Q_B was significantly the most accurate and showed the highest average power compared to Q and Q_T. Its accuracy was equivalent to that of the Shapiro–Wilk W-test, although the latter outperformed all three Q variants in average power. Conclusions: Q_B is a suitable inferential extension of the seven-number summary for testing normality.

Keywords: Testing normality, normal distribution, quadratic forms; bootstrap, generalized chi-square distribution, inference statistics.

1. Introduction

Bowley [1] proposed the five-number summary to describe the distribution of a variable based on five positional statistics: the minimum, the three quartiles, and the maximum. In 1910, he expanded this summary to seven numbers by incorporating the first and last deciles, allowing for the definition of two extreme deviation ranges. Very low values fall below the 10th percentile, while very high values exceed the 90th percentile. The central tendency, or shoulder area, extends across the interquartile range, spanning from the first to the third quartile. Low values lie between the 10th and 25th percentiles, while high values fall between the 75th and 90th percentiles [2]. Thus, the seven-number summary provides a structured description of the distribution and facilitates score interpretation using six class intervals: $[min, p_{10}]$ very low scores, $(p_{10}, p_{25}]$ low scores, $(p_{25}, p_{50}]$ low-medium scores, $(p_{50}, p_{75}]$ medium-high scores, $(p_{75}, p_{90}]$ high scores, and $(p_{90}, max]$ very high scores [3-4].

In recent years, Moral [5] proposed a method for testing whether a sample has been drawn from a normal distribution using seven quantiles that are equispaced at two-thirds from the central point of a standard normal distribution. This approach is an adaptation of the seven-number summary for assessing normality. Since this summary is defined with respect to the normal distribution, it is referred to as parametric. Previously, it had only been used for descriptive and indicative purposes in normality assessment. However, Moral [5] proposed its application as an omnibus statistical test. Notably, in a similar vein, Avdović and Jevremović [6] also suggested assessing normality by comparing empirical and theoretical quantiles, though their approach is based on the entire set of n sample data rather than on seven specific quantiles.

Citation: J. M. de la Rubia, "Refined Normality Test Based on the Parametric Seven-Number Summary," *Journal of Statistics and Applied Sciences*, no. 11, pp. 1–39, 2025, doi: 10.52693/jsas.1670945.

In the proposal of the test based on the Parametric Seven-Number Summary (PSNS), a parametric approximation using the chi-square distribution with seven degrees of freedom was employed. This approximation assumes independence among the standardized quantiles under the assumption of normality, with their sum of squares serving as the test statistic (Q-statistic). Although this convenient approach simplifies the calculations and appears to perform well in practice [5], it does not hold up theoretically.

The first objective of this study is to propose two variants of the test to assess the sampling distribution of the test statistic. In the first variant, the sampling distribution of the Q-statistic—along with the calculation of the p-value, type II error, and statistical power—is determined using the bootstrap procedure. This variant is denoted as Q_B, to distinguish it from the original proposal (Q). An R script was developed for this purpose [7]. See Appendices A (Q-statistic), B (bootstrap version of PSNS Q test), and C (statistical power of bootstrap version of PSNS Q test).

Specifically, 10,000 resampled datasets of the same size as the original sample are generated by either (a) resampling with replacement from the observed data (nonparametric bootstrap) or (b) drawing 2,000 random samples from a normal distribution where the location parameter is estimated by the sample mean and the scale parameter is estimated by the sample standard deviation with Bessel's correction (parametric bootstrap). The first approach produces the empirical bootstrap distribution, while the second generates the normative bootstrap distribution. The latter is used both to calculate the p-value (the proportion of values in the normative bootstrap distribution that exceeds the test statistic value in the original sample) and to determine the critical value (the 0.95th quantile of normative bootstrap distribution). This critical value is then used to compute statistical power via rejection rate in the empirical bootstrap distribution.

A sample size of 2,000 was chosen for the normative bootstrap distribution, as it is considered sufficiently large for the estimation of p-values and critical values [8]. However, for the calculation of statistical power, the number of bootstrap samples was increased to 10,000. This modification allowed for a single simulation to be used, thus avoiding the need to define a rejection threshold (e.g., 5% or 50% rejection across 1,000 simulations) for classifying each simulation as a rejection when computing the rejection rate [9].

The second variant introduces a new statistic based on a quadratic form, where the symmetric matrix corresponds to the expected correlations between the seven quantiles under the null hypothesis of normality [10]. This statistic is denoted as Q_T, with the subscript T referring to the theoretical values under the null hypothesis of a standard normal distribution used in its computation. The p-value and statistical power of Q_T are obtained using the generalized chi-square distribution, implemented via the CompQuadForm package in R [11]. See Appendix D (quadratic form version of PSNS Q test). The original version of the Q Test can be found in Appendix E (chi-square approximation to PSNS Q test).

The second objective of this study is to determine which of the proposed methods (Q, Q_B, and Q_T) demonstrates the best performance. To this end, the three variants of the normality test based on the PSNS, along with the Shapiro-Wilk W test [12], are compared in terms of accuracy (proportion of correct classifications or hit rate) and statistical power. The comparisons are conducted across 31 sample sizes (ranging from 50 to 1,500 in increments of 50, plus 2,000) and 12 distributions (one normal and 11 non-normal). The W test is included using Royston's standardization procedure [13], as it is regarded as the most powerful and accurate test for sample sizes ranging from 3 to 2,000 observations [14-15]. Specifically, the largest original sample size considered in this study was 2,000, which is unusually large for an empirical study in the social and health sciences. For larger samples sizes, the Shapiro-Francia W' test [16] using Royston's standardization procedure [17] could have been applied.

This is a methodological study, in which the use of simulated data is considered the most appropriate. This type of data allows us to determine whether the null hypothesis of normality should be retained or rejected—something that is not possible with real scale-based data, where the distribution is assumed and ultimately unknown. However, since the purpose of the study is to apply the normality test to real-world datasets, a third objective has been added: to demonstrate the practical effectiveness of the proposed test variants using R scripts applied to real-world data.

To achieve this, a personality measurement instrument was administered to a sample of 385 students who applied for admission to a psychology program at a public university in northeastern Mexico. The study assessed alexithymia, a trait characterized by difficulty in identifying and expressing feelings, as well as a cognitively externalized style disconnected from the internal experiential world of emotions, dreams, and fantasies. This personality trait is evaluated using the Toronto Alexithymia Scale (TAS-20) [18], which was validated in Mexico by Moral [19].

Alexithymia—measured on an interval scale such as the total TAS-20 score—is expected to follow a normal distribution, given that this personality trait is influenced by both environmental and biological factors [18]. It is hypothesized that TAS-20 total scores within the central range of the distribution reflect adaptive or functional expressions of the trait, whereas scores in the tails reflect maladaptive manifestations (right tail) and exceptional emotional competence and awareness (left tail). Specifically, individuals with alexithymic personality disorder are expected to score in the right tail of the distribution. Empirical evidence supports this model [20].

2. The Summary of the Seven Numbers to Evaluate Normality

In the test based on the PSNS, the test statistic, symbolized by Q, is defined as the sum of the squares of the quantiles corresponding to the orders -2, -4/3, -3/4, 0, 2/3, 4/3, and 2 of a standard normal distribution. Before squaring and summing the seven sample quantiles, they are standardized [5]. Initially, the sample mean and standard deviation are used for standardization. Subsequently, they are re-standardized using the expected mean and standard deviation under the null hypothesis of a standard normal distribution. Since these variables are correlated, the sampling distribution of this quadratic form follows a generalized chi-square distribution [21]. However, Moral [5] suggested a simplified approach for calculating the p-value by approximating it with a chi-square distribution with seven degrees of freedom.

It is worth noting that the 2nd, 8th, 25th, 50th, 75th, 92nd, and 98th percentiles of a standard normal distribution are approximately evenly spaced. When rounded to two decimal places, the second percentile corresponds to -2.05, the eighth to -1.34, the twenty-fifth to -0.67, the fiftieth to 0, the seventy-fifth to 0.67, the ninety-second to 1.34, and the ninety-eighth to 2.05. Within this sequence, the quantiles are separated by a distance of 0.67 along the Z-score axis, except at the two extremes, where a slight discrepancy of 0.04 is observed (-1.34 - (-2.05) = 0.71 = 0.67 + 0.04 and 2.05 - 1.34 = 0.71 = 0.67 + 0.04).

To achieve a more precise alignment with evenly spaced quantiles, the following orders must be used: 0.0228, 0.0912, 0.2525, 0.5, 0.7475, 0.9088, and 0.9772. These yield the quantiles -2, -4/3, -2/3, 0, 2/3, 4/3, and 2, which are evenly spaced at two-thirds intervals within the range [-2, 2], centered at 0. These seven quantiles closely approximate the percentiles used in the seven-number summary, namely the minimum (0th percentile), 10th percentile, 25th percentile, 50th percentile, 75th percentile, 90th percentile, and maximum (100th percentile) of a standard normal distribution. Therefore, they are considered a summary of seven key values for a normal distribution [5], also referred to as the PSNS [22-23].

It should be noted that, since the normal distribution has infinite support over the interval $(-\infty, +\infty)$, it does not inherently have a minimum or maximum. However, any finite sample x of size n drawn from a normal distribution will have a limited range of scores, bounded by the sample minimum and maximum, $[min(\mathrm{x}), max(\mathrm{x})]$. Moreover, approximately 95.44% of the distribution lies within the interval $[-2, 2]$. For this reason, in the PSNS [5], the minimum is replaced by the 0.0228 quantile and the maximum by the 0.9772 quantile.

In asymptotic sampling, the distribution of the quantiles of a distribution with finite moments follows a normal distribution [10]. If X is a random variable with a density function $f_X(x)$ and finite moments, the sample quantile of order p, denoted as $q_x(p)$, follows a normal distribution. Its arithmetic mean or expected value, denoted as $\mu[q_x(p)]$, corresponds to the population quantile $Q_X(p)$. The variance is given by the ratio of the product of the quantile order and its complement (numerator) to the product of the sample size and the square of the density function evaluated at the quantile (denominator). See Equation 1. The standard deviation or standard error, denoted as $\sigma[q_x(p)]$, is the square root of this ratio (Equation 2).

$$x = \{x_i\}_{i=1}^{n} = \{x_1, x_2, \dots, x_n\} \subseteq X$$
$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \tag{1}$$
$$q_x(p) = x_{(i)} \sim N\left(\mu[q_x(p)] = Q_X(p), \sigma^2[q_x(p)] = \frac{p \times (1-p)}{n \times f_X^2[Q_X(p)]}\right)$$

$$\sigma[q_x(p)] = \sqrt{\frac{p \times (1-p)}{n \times f_X^2[Q_X(p)]}} \tag{2}$$

Equation 2 can be used to obtain the asymptotic standard errors of the seven quantiles that make up the PSNS, based on the orders of the quantiles (p = 0.02275, 0.09121, 0.25249, 0.5, 0.74751, 0.90879, and 0.97725) of a standard normal distribution [2, 5]. Table 1 presents these quantiles, specifically for the standard normal distribution N(0, 1).

Table 1. Value in the probit function, cumulative distribution function, density function, and asymptotic standard errors of the PSNS in a standard normal distribution

$Q_Z(p) = z_p$	$F_Z(z_p) = p$	$f_Z(z_p)$	$\sigma(z_p)$
-2	0.02275	0.05399	$2.76168/\sqrt{n}$
-4/3	0.09121	0.16401	$1.75544/\sqrt{n}$
-2/3	0.25249	0.31945	$1.35998/\sqrt{n}$
0	0.5	0.39894	$1.25331/\sqrt{n}$
2/3	0.74751	0.31945	$1.35998/\sqrt{n}$
4/3	0.90879	0.16401	$1.75544/\sqrt{n}$
2	0.97725	0.05399	$2.76168/\sqrt{n}$

Note. $Q_Z(p) = \Phi^{-1}(p) = z_p$ = probit function, quantile function, or inverse function of the cumulative distribution function of a variable following a standard normal distribution $N(0, 1)$, p = order or cumulative probability associated with the quantile z_p in the standard normal distribution, $f_Z(z_p) = \varphi(z_p)$ = density function of the standard normal distribution, $\sigma(z_p)$ = asymptotic standard error of the sample p-th quantile calculated from a random sample of size n drawn from a standard normal distribution.

The standard errors in Table 1 ($\sigma[z_p]$) allow the definition of asymptotic confidence intervals within which sample quantiles are expected to fall when computed from random samples of size n drawn from a standard normal distribution, with a probability of 1 - α [10]. See Equation 3.

$$Z \sim N(0, 1)$$
$$P\left(\hat{z}_p \in \left[z_p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p \times (1-p)}{n \times f_Z^2(z_p)}}, z_p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p \times (1-p)}{n \times f_Z^2(z_p)}}\right]\right) = 1 - \alpha \tag{3}$$

z_p = population p-th quantile of the standard normal distribution.
$\hat{z}_p$= sample quantile or point estimate of the population p-th quantile of a variable X following a standard normal distribution.
n = sample size.
$f_Z^2(z_p)$ = squared height or density of the z_p in a standard normal distribution.
$z_{1-\alpha/2}$ = (1-α/2)-th quantile of a standard normal distribution, where α is the significance level, typically 0.05 ($z_{0.975}$ = 1.96).

From this asymptotic confidence interval, it is possible to derive a normally distributed statistic with a mean of 0 and a variance of 1, which allows testing whether a standardized sample quantile (standardized using its sample mean and variance) corresponds to the expected quantile of a standard normal distribution [5]. See Equation 4.

$$Z = \frac{\hat{z}_p - z_p}{\sqrt{\dfrac{p \times (1-p)}{n \times f_Z^2(z_p)}}} \sim N(0,1) \tag{4}$$

3. Testing the Expected Value Under Normality for a Number of the PSNS

Based on the Z statistic in Equation 4, a statistical test can be formulated to assess whether the sample p-th quantile is equivalent to the expected value under a normal distribution, and this test can then be generalized to the seven values of the PSNS.

Equation 5 shows the statistical hypotheses of the two-tailed test.

$$\begin{aligned} H_0&: Z[Q_X(p)] = \frac{Q_X(p) - \mu_X}{\sigma_X} = Q_Z(p) = z_p, where\ Z \sim N(0,1) \\ H_1&: Z[Q_X(p)] = \frac{Q_X(p) - \mu_X}{\sigma_X} \neq Q_Z(p) = z_p, where\ Z \sim N(0,1) \end{aligned} \tag{5}$$

$Z[Q_X(p)]$ = population p-th quantile of variable X, standardized by population mean and variance.
z_p = hypothetical value of the p-th quantile in a standard normal distribution N(0, 1).

Assumptions: A random sample of large size drawn from a normal distribution.

Test statistic: Equation 6 shows the double standardization (first by the sample mean and standard deviation and second by the mean and expected standard deviation under the null hypothesis of normality) experienced by the sample quantile $q_x(p)$.

$$Z\big[z\big(q_x(p)\big)\big] = \frac{z[q_x(p)] - z_p}{\sqrt{\dfrac{p(1-p)}{nf_Z^2(z_p)}}} = \frac{\dfrac{q_x(p) - \bar{x}}{s_{n-1}(x)} - z_p}{\sqrt{\dfrac{p(1-p)}{nf_Z^2(z_p)}}} \tag{6}$$

The theoretical quantile z_p is obtained by evaluating the probit function or quantile function of a standard normal distribution in p (Equation 7). In the context of the PSNS, p takes the following values: $F_Z(-2) = 0.0228$, $F_Z(-2/3) = 0.0912$, $F_Z(-1/3) = 0.2525$, $F_Z(0) = 0.5$, $F_Z(1/3) = 0.7475$, $F_Z(2/3) = 0.9088$, and $F_Z(2) = 0.9772$, where Z ~ $N(0, 1)$ and F_Z is its cumulative distribution function, usually expressed as $\Phi(z_p) = p$.

$$Q_Z(p) = \Phi^{-1}(p) = z_p, where\ X \sim N(0,1) \tag{7}$$

To obtain the standardized sample p-th quantile, denoted by $z[c_x(p)]$, we start by calculating this quantile. To do this, the n sample data of X are ordered in ascending direction: $x_{(i)}$, $i = 1, 2, \cdots, n$. For more details refer to Equation 8.

$$\begin{gathered} x = \{x_i\}_{i=1}^{n} = \{x_1, x_2. \ldots, x_n\} \subseteq X \\ x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \end{gathered} \tag{8}$$

The position of the quantile q_p among the n sample observations is calculated using R's type-8 rule, which provides greater accuracy in estimating quantiles across different distributions compared to other linear interpolation methods [4, 24-25]. Refer to Equation 9 for more details. The type-6 rule is also recommended for general use with very different distributions [26], while the type-9 rule is specifically suggested for the normal distribution [27].

$$i = 1/3 + p \times (n + 1/3) = i + (i - \lfloor i \rfloor) \tag{9}$$

If the value of i is an integer, the data is searched in that order; whereas if it is a number with decimal places, the quantile is obtained by linear interpolation (Equation 10).

$$\hat{Q}_X(p) = q(p) = x_{(\lfloor i \rfloor)} + (i - \lfloor i \rfloor)\big(x_{(\lfloor i \rfloor + 1)} - x_{(\lfloor i \rfloor)}\big) \tag{10}$$

The mean (Equation 11) and sample standard deviation (Equation 12) are further calculated.

$$\hat{\mu}_X = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{11}$$

$$\hat{\sigma}_X = s_{n-1}(x) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{12}$$

Finally, the sample quantile is standardized, as shown in Equation 13.

$$z[q_x(p)] = \frac{\hat{Q}_X(p) - \hat{\mu}_X}{\hat{\sigma}_X} = \frac{q_x(p) - \bar{x}}{s_{n-1}(x)} \tag{13}$$

Sampling distribution: The random variable $Z[z(c_x(p))]$, proposed as a test statistic, follows a standard normal distribution under the null assumption or hypothesis that variable X follows a normal distribution (Equation 14).

$$Z\left[z\left(q_x(p)\right)\right] \sim N(0, 1) \tag{14}$$

Decision based on probability value or critical level: Let $P(Z \leq |z|)$ the probability of obtaining a value less than or equal to the absolute value of the test statistic in a standard normal distribution. If 2-tailed p-value = 2 × (1 - $P(Z \leq |z|) \geq \alpha$, the null hypothesis is maintained in a two-tailed test with a α significance level. Conversely, if 2-tailed p-value = 2 × (1 - $P(Z \leq |z|) < \alpha$, it is rejected.

4. The Omnibus Test of Normality Based on the PSNS

4.1. Approach

If the seven standardized values $z[z(c_x(p))]$ are squared and added together, the resulting statistic, denoted by Q, follows a generalized chi-square distribution, since it is the sum of squares of seven dependent random variables with a standard normal distribution [28-29]. The chi-square distribution with seven degrees of freedom is a concrete case of this generalized chi-square distribution. This equality occurs when quantiles, considered as random variables, have very small or no correlations. Moral [5] proposed using the chi-square distribution with seven degrees of freedom as an approximation to the generalized chi-square distribution and, in this way, obtain an omnibus test of normality with simple and well-known calculations of the probability, power, and effect size value [30]. Analogous to the variance formula of a quantile (Equation 1), the formula of covariance between two quantiles is a quotient, the numerator of which is the product of the order of one quantile and the complement of the order of the other quantile, and its denominator is the product of the sample size and densities of both quantiles [10]. See Equation 15. This covariance converges to zero when the sample size (n) tends to infinity, which is favored by a greater distance between the two quantiles. Except in the uniform distribution, the more distant two quantiles are, the lower their correlation. Also, the variance and standard deviation of each quantile tend to zero as the sample size increases. However, the correlation or quotient between the covariance and the geometric mean of the variances (Equation 16) is independent of the sample size. The n in the numerator of the covariance is simplified with the two square roots of n of each standard deviation in the denominator. On the other hand, this correlation is always positive, since there is no term that could be negative in the formula. Precisely, the definition of quantile or order statistic implies a monotonic non-decreasing order [31].

$$\sigma\left(q_{p_1}, q_{p_2}\right) = \frac{p_1 \times (1 - p_2)}{n \times f_Z(z_{p_1}) \times f_Z(z_{p_2})}, where\ p_1 < p_2 \tag{15}$$

$$\rho\left(q_{p_1}, q_{p_2}\right) = \frac{\sigma\left(q_{p_1}, q_{p_2}\right)}{\sqrt{\sigma^2\left(q_{p_1}\right) \times \sigma^2\left(q_{p_2}\right)}} = \frac{\sigma\left(q_{p_1}, q_{p_2}\right)}{\sigma\left(q_{p_1}\right) \times \sigma\left(q_{p_2}\right)} \tag{16}$$

The correlations between the quantiles of the PSNS of a normal distribution vary from a minimum of 0.023 to a maximum of 0.581 with a mean of 0.275, 12 out of the 21 correlations are less than 0.3, with 5 out of these 12 correlations being less than 0.1. In turn, 5 out of the 21 correlations take values between 0.3 and 0.49 and 4 between 0.5 and 0.59. The four highest correlations are between the third and fourth numbers and between the fourth and fifth numbers, as well as between the third and fifth numbers with their values contiguous towards the corresponding tail (Table 2).

Table 2. Matrix of correlations between PSNS quantiles of a normal distribution

$\rho(z_{[i,]}, z_{[,j]})$	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	1	0.4816	0.2625	0.1526	0.0887	0.0483	0.0233
[2,]	0.4816	1	0.5451	0.3168	0.1841	0.1004	0.0483
[3,]	0.2625	0.5451	1	0.5812	0.3378	0.1841	0.0887
[4,]	0.1526	0.3168	0.5812	1	0.5812	0.3168	0.1526
[5,]	0.0887	0.1841	0.3378	0.5812	1	0.5451	0.2625
[6,]	0.0483	0.1004	0.1841	0.3168	0.5451	1	0.4816
[7,]	0.0233	0.0483	0.0887	0.1526	0.2625	0.4816	1

Note. $\rho(z_{[i,]}, z_{[,j]})$ = correlation between quantiles of a standard normal distribution: $1 = z(p = F_Z(-2))$, $2 = z(p = F_Z(-2/3))$, $3 = z(p = F_Z(-1/3))$, $4 = z(p = F_Z(0))$, $5 = z(p = F_Z(1/3))$, $6 = z(p = F_Z(2/3))$, and $7 = z(p = F_Z(2))$, where F_Z is the cumulative distribution function of a standard normal distribution.

The formula shown in Equation 16, which requires Equations 2 and 15 for calculation, provides the population correlation, i.e., when the sample size tends to infinity. Appendix F contains an R script for computing this correlation matrix using a bootstrap procedure.

By extracting a large sample of 1,000 data points from a standard normal distribution using a single seed (original reproducible sample), generating 10,000 bootstrap samples, computing the seven quantiles in each bootstrap sample, calculating the 21 correlations between these quantiles, and averaging the 10,000 estimates for each correlation, the resulting correlation matrix closely resembles the one obtained using the formula above (Table 3). The absolute residuals, or differences, between the corresponding correlations (asymptotic vs. bootstrap) range from 0.0003 to 0.0237, with a mean of 0.0101. The Root Mean Square Error (RMSE) is 0.0090. Therefore, both matrices can be considered equivalent.

Table 3. Correlation matrix obtained by bootstrap (10,000 replications) from a random sample of 1000 data extracted from a standard normal distribution.

$r(q_{x[i,]}, q_{[,j]})$	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	1	0.4677	0.2628	0.1465	0.0933	0.0323	0.0206
[2,]	0.4677	1	0.5301	0.2965	0.1751	0.0913	0.0581
[3,]	0.2628	0.5301	1	0.5575	0.3212	0.1758	0.0934
[4,]	0.1465	0.2965	0.5575	1	0.5647	0.3064	0.1507
[5,]	0.0933	0.1751	0.3212	0.5647	1	0.5374	0.2555
[6,]	0.0323	0.0913	0.1758	0.3064	0.5374	1	0.4732
[7,]	0.0206	0.0581	0.0934	0.1507	0.2555	0.4732	1

Note. $q_{x[i,]}$ = i-th quantile and $c_{[,j]}$ = j-th quantile of a sample x of size 1000 randomly drawn from a standard normal distribution using a fixed seed (123) to ensure reproducibility. Quantile order (i for the quantile per row and j for the quantile per column): $1 = \Phi(-2)$, $2 = \Phi(-4/3)$, $3 = \Phi(-2/3)$, $4 = \Phi(0)$, $5 = \Phi(2/3)$, $6 = \Phi(4/3)$, and $7 = \Phi(2)$, where Φ is the cumulative distribution function of a standard normal distribution, and $r(q_{x[i,]}, q_{[,j]})$ = Pearson's product-moment correlation coefficient between the sample quantiles of order i and order j (computed from 10,000 bootstrap samples using R's type-9 rule).

It should be noted that the high correlations between the terms (random variables) in the sum of squares of the Q statistic cause the generalized chi-square distribution to deviate from the standard chi-square distribution due to pronounced leptokurtosis. This results in a greater concentration of values around the center and a shorter right tail. This effect occurs because, when variables are correlated, they tend to move together, reducing the overall dispersion of the sum of squares. Consequently, leptokurtosis increases (more values near the mean), and positive skewness decreases, with fewer extreme values appearing in the right tail.

One of the challenges of the generalized chi-square distribution is the absence of tables for its cumulative densities and probabilities, as it does not have a standard form, unlike the chi-square distribution. If this distribution is used as an approximation to compute the p-value, the significance level could be increased to compensate for the shortening of the right tail [5].

Another, much better option is to compute the p-value, statistical power, and effect size using the bootstrap procedure [32]. This empirical method allows for the generation of the generalized chi-square distribution underlying the test statistic, which is a sum of squares of dependent standard normal variables, thereby preserving the original formulation. A third alternative is to express the statistic as a quadratic form–specifically, the product of the transposed vector of standardized quantiles, the correlation matrix of the quantiles, and the vector of standardized quantiles–and use R's CompQuadForm package [11]. In this approach, 21 duplicate summands are incorporated, corresponding to the cross-products of the seven quantiles weighted by their correlations.

4.2. Statistical Hypotheses

Equation 17 presents the statistical hypotheses of the Q test. They are formulated as a two-tailed test and involve a random vector consisting of seven standardized normal variables (quantiles).

$$\begin{gathered} H_0\colon X \sim N(\mu_X, \sigma_X) \Rightarrow Z = \frac{X-\mu_X}{\sigma_X} \sim N(0,1) \Rightarrow \begin{pmatrix} z_{Q_Z(p=0.023)} \\ z_{Q_Z(p=0.091)} \\ z_{Q_Z(p=0.252)} \\ z_{Q_Z(p=0.5)} \\ z_{Q_Z(p=0.748)} \\ z_{Q_Z(p=0.977)} \end{pmatrix} = \begin{pmatrix} -2 \\ -4/3 \\ -2/3 \\ 0 \\ 2/3 \\ 4/3 \\ 2 \end{pmatrix} \\ H_1\colon X \nsim N(\mu_X, \sigma_X) \Rightarrow Z = \frac{X-\mu_X}{\sigma_X} \nsim N(0,1) \Rightarrow \begin{pmatrix} z_{Q_Z(p=0.023)} \\ z_{Q_Z(p=0.091)} \\ z_{Q_Z(p=0.252)} \\ z_{Q_Z(p=0.5)} \\ z_{Q_Z(p=0.748)} \\ z_{Q_Z(p=0.977)} \end{pmatrix} \neq \begin{pmatrix} -2 \\ -4/3 \\ -2/3 \\ 0 \\ 2/3 \\ 4/3 \\ 2 \end{pmatrix} \end{gathered} \tag{17}$$

4.3. Assumptions

It is assumed that the sample is random and consists of n scores or values from a continuous quantitative variable that follows a normal distribution. A large sample size is required.

Based on a simulation study, Moral [5] recommended a minimum sample size of 150 participants when approximating the p-value using the chi-square distribution with seven degrees of freedom (original formulation). This sample size ensures high accuracy (i.e., a proportion of correct decisions—retaining or rejecting the null hypothesis of normality—of at least 0.8), as well as high statistical power for detecting non-normality and a high complement of power (i.e., the probability of correctly retaining the null hypothesis of normality when the alternative is false) for the normal distribution, with both values (Φ when H_0 should be retained, or 1 − Φ when H_0 should be rejected) being ≥ 0.8. However, the test achieved perfect accuracy (i.e., 1) in detecting normality with a sample size as small as 20, although with a power below 0.2. The test's limited power to reject the null hypothesis of normality was observed in the presence of symmetric mesokurtic distributions, such as the uniform and semicircular distributions, as well as symmetric distributions with slight leptokurtosis, such as the logistic distribution. This limitation can be attributed to the high correlations among the four central values of the PSNS (Table 2).

4.4. Test Statistics (Three Variants)

4.4.1 Statistics Q and Q_B

The test statistic is defined as the sum of squared standardized quantiles from the PSNS, using the expected values under the assumed distribution (standard normal), and excluding the influence of the sample mean ($\bar{x} = \sum_{i=1}^{n} x_i/n$) and standard deviation ($s_{n-1}(x) = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$). This new random variable is denoted as Q, in reference to the quadratic forms described in Cochran's theorem [28]. See Equation 18, where x is a random sample of n data from a variable X with unknown distribution F_X that is hypothesized to be normal and Z is a variable defined over the domain (-∞, +∞) that follows a standard normal distribution. The R's implementation is provided in Appendix A.

$$\begin{gathered} x = \{x_i\}_{i=1}^{n} \sim F_n(x_i); \subseteq X \sim F_X(x_i); z_i \in Z \sim F_Z(z_i) \equiv N(0,1) \\ Q = \sum_{i=1}^{7} \left(\frac{\frac{q_X(p_i) - \bar{x}}{s_{n-1}(x)} - q_Z(p_i)}{\sigma\big(q_Z(p_i)\big)} \right)^2 = \sum_{i=1}^{7} \left(\frac{\frac{q_X(p_i) - \bar{x}}{s_{n-1}(x)} - z_{p_i}}{\sigma\big(z_{p_i}\big)} \right)^2 = \end{gathered} \tag{18}$$

$$= \left(\frac{\frac{q_x(p_1 = 0.02275) - \bar{x}}{s_{n-1}(x)} + 2}{2.76168/\sqrt{n}}\right)^2 + \left(\frac{\frac{q_x(p_2 = 0.09121) - \bar{x}}{s_{n-1}(x)} + \frac{4}{3}}{1.75544/\sqrt{n}}\right)^2$$

$$+ \left(\frac{\frac{q_x(p_3 = 0.25249) - \bar{x}}{s_{n-1}(x)} + \frac{2}{3}}{1.35998/\sqrt{n}}\right)^2 + \left(\frac{\frac{q_x(p_4 = 0.5) - \bar{x}}{s_{n-1}(x)}}{1.25331/\sqrt{n}}\right)^2$$

$$+ \left(\frac{z\,\frac{q_x(p_5 = 0.74751) - \bar{x}}{s_{n-1}(x)} - \frac{2}{3}}{1.35998/\sqrt{n}}\right)^2 + \left(\frac{\frac{q_x(p_6 = 0.90879) - \bar{x}}{s_{n-1}(x)} - \frac{4}{3}}{1.75544/\sqrt{n}}\right)^2$$

$$+ \left(\frac{\frac{q_x(p_7 = 0.97725) - \bar{x}}{s_{n-1}(x)} - 2}{2.76168/\sqrt{n}}\right)^2$$

It should be noted that this sum of squares, derived from seven correlated Gaussian random variables with a mean of 0 and variance of 1, does not account for the standardized covariance matrix of the quadratic form. Strictly speaking, it is not a true quadratic form, as it does not assume that this matrix is an identity matrix. Therefore, the symbol Q serves as a referential allusion to Cochran's theorem [28]. In the Q_B variant, the distribution of the test statistic in the sampling process—used for computing the p-value (see Appendix B for the R implementation) and statistical power (see Appendix C for the R implementation)—is determined using the empirical resampling method with replacement (bootstrap), which provides a more accurate approximation than the original approach.

4.4.2 Q_T Statistic

Another option is to express the statistic as a quadratic form without ignoring the correlation matrix and use R's CompQuadForm package [11] to compute the probability value and power. This option is denoted Q_T and is properly a quadratic form. The subscript T refers to the fact that the matrix of covariances is theoretically established from the null hypothesis of normal distribution.

A quadratic form is a linear combination involving a transposed vector (of order 1 × k), a symmetric matrix of order $k \times k$, and the original vector of order $k \times 1$. To obtain the quadratic form, the values $z[z(q_x(p)]$ are used as the vector, denoted by z (Equation 19). The symmetric matrix corresponds to the matrix of standardized covariances or correlations between the quantiles of the PSNS of a normal distribution. This matrix is denoted by R as the correlation matrix (Equation 20).

The quadratic form z^T R z results in a second-degree polynomial with $k \times k$ terms, comprising the sum of the squared variables and the cross-products of the variables weighted by their corresponding correlations (Equation 21). If the vector and the matrix contain only numerical values, without variables, the quadratic form yields a scalar. See Appendix D for the calculation of the Q_T statistic in R.

$$z^t = (z_1 \quad z_2 \quad z_3 \quad z_4 \quad z_5 \quad z_6 \quad z_7) \tag{19}$$

$$\mathrm{R} = \begin{pmatrix} 1 & 0.2819 & 0.2625 & 0.1526 & 0.0887 & 0.0826 & 0.0233 \\ 0.2819 & 1 & 0.9314 & 0.5413 & 0.3146 & 0.2930 & 0.0826 \\ 0.2625 & 0.9314 & 1 & 0.5812 & 0.3378 & 0.3146 & 0.0887 \\ 0.1526 & 0.5413 & 0.5812 & 1 & 0.5812 & 0.5413 & 0.1526 \\ 0.0887 & 0.3146 & 0.3378 & 0.5812 & 1 & 0.9314 & 0.2625 \\ 0.0826 & 0.2930 & 0.3146 & 0.5413 & 0.9314 & 1 & 0.2819 \\ 0.0233 & 0.0826 & 0.0887 & 0.1526 & 0.2625 & 0.2819 & 1 \end{pmatrix} \tag{20}$$

$$\begin{aligned} Q_T = z^t \times \mathrm{R} \times z = {} & z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^2 + z_6^2 + z_7^2 + 2 \times 0.4816 \times z_1 \times z_2 \\ & + 2 \times 0.2625 \times z_1 \times z_3 + 2 \times 0.1526 \times z_1 \times z_4 + 2 \times 0.0887 \times z_1 \times z_5 \\ & + 2 \times 0.0483 \times z_1 \times z_6 + 2 \times 0.0233 \times z_1 \times z_7 + 2 \times 0.5451 \times z_2 \times z_3 \\ & + 2 \times 0.3168 \times z_2 \times z_4 + 2 \times 0.1841 \times z_2 \times z_5 + 2 \times 0.1004 \times z_2 \times z_6 \\ & + 2 \times 0.0483 \times z_2 \times z_7 + 2 \times 0.5812 \times z_3 \times z_4 + 2 \times 0.3378 \times z_3 \times z_5 \\ & + 2 \times 0.1841 \times z_3 \times z_6 + 2 \times 0.0887 \times z_3 \times z_7 + 2 \times 0.5812 \times z_4 \times z_5 \\ & + 2 \times 0.3168 \times z_4 \times z_6 + 2 \times 0.1526 \times z_4 \times z_7 + 2 \times 0.5451 \times z_5 \times z_6 \\ & + 2 \times 0.2625 \times z_5 \times z_7 + 2 \times 0.4816 \times z_6 \times z_7 \end{aligned} \quad (21)$$

4.5. Sampling Distribution

The statistic or random variable Q or Q_B (sum of squares) follows a generalized chi-square distribution, $\chi^2_g(7, \Sigma)$, as does the quadratic form Q_T (Equation 22). This distribution deviates from the chi-square distribution with seven degrees of freedom due to leptokurtosis and a shortened right tail, a consequence of the positive correlations among the seven random variables whose sum of squares defines it [21, 28-29].

The Q statistic would follow a chi-square distribution if the seven variables were independent, and the Q_T statistic would do so if the R matrix were an identity matrix. However, the seven quantiles are dependent because they originate from the same distribution, with an average correlation of 0.275.

If the chi-square distribution with seven degrees of freedom is used as an approximation, the significance level could be increased to 0.1 to compensate for the shortened right tail (see Appendix E for the calculation in R).

However, an approximation to the generalized chi-square distribution can be obtained via the bootstrap method. On the one hand, the empirical bootstrap distribution is generated from 10,000 replications of the original sample. On the other hand, the normative bootstrap distribution is obtained from 2,000 replications of a random sample drawn from a normal distribution whose location parameter is the arithmetic mean of the original sample and whose scale parameter is the standard deviation of the original sample, corrected using Bessel's correction. The 0.95 quantile (using R's type-8 method) of normative bootstrap distribution serves as the critical value. The bootstrap right-tailed p-value is computed, defined as the proportion of values (Q estimates) in the normative bootstrap distribution that exceeds the test statistic (from original sample). See Appendix B for the R code used in this calculation. This empirical approach provides a more accurate estimation of both the p-value (Appendix B) and the statistical power (see the R script in Appendix C) compared to the chi-square distribution with seven degrees of freedom (see the R script in Appendix E). This approach replaces the chi-square distribution with a normative bootstrap distribution that has a similar form, as it constitutes a generalized chi-square distribution. Consequently, the probability value is computed based on the right tail of the distribution.

$$Q \sim \chi_g^2(7, \Sigma) \ and \ Q_T \sim \chi_g^2(7, \mathrm{R}) \quad (22)$$

4.6. Decision

Let $P(\chi^2\mathrm{g}(7, \Sigma) \geq Q)$, be the probability of obtaining a value greater than or equal to the test statistic Q in a generalized chi-square distribution with seven degrees of freedom. If $P(\chi^2_g(7, \Sigma) \geq Q) \geq \alpha$, the null hypothesis of normality is retained in a two-tailed test with a significance level of α. Conversely, if $P(\chi^2_g(7, \Sigma) < \alpha$, the null hypothesis is rejected.

The same procedure applies to the quadratic form Q_T and the initial approximation of the Q statistic using the chi-square distribution with seven degrees of freedom [5].

4.7. Statistical Power

4.7.1 With Q_B Statistic

Type II error is defined as the probability of retaining the null hypothesis given that the alternative hypothesis is true (Q > critical value). Therefore, it represents the error of failing to reject the null hypothesis. It is denoted by β. Its complement, denoted by φ, represents the statistical power, that is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true [33].

Using a resampling procedure with replacement from the original sample, 10,000 bootstrap samples are generated. For each bootstrap sample, the Q statistic is calculated, and the proportion of null hypothesis rejections is determined. The null hypothesis of normality is rejected if the Q statistic exceeds the critical value obtained from 2,000 samples generated under normality (normative bootstrap distribution). The rejection rate represents the bootstrap power, reflecting the test's ability to detect deviations from normality in the original sample. See Appendix E for the R implementation.

4.7.2 With the Q_T Statistic

The statistical power of the quadratic form $Q_T = z^T R z$ can be computed using Davies' AS 155 algorithm (1980) [34]. The initial evaluation of the generalized non-central chi-square distribution (which corresponds to the distribution of a quadratic form of correlated variables) is performed at a critical level associated with a specific alternative hypothesis. This level corresponds to the $(1 - \alpha)$th quantile of a chi-square distribution, where the degrees of freedom equal the number of eigenvalues of the correlation matrix. Since this is an iterative procedure, the critical value is adjusted at each step.

In Davies' model, the λ parameter (degrees of freedom) is the eigenvalue vector of the correlation matrix R, while the δ parameter (non-centrality parameter) is defined as the element-wise product of the eigenvalues and the squared standardized quantiles. See Appendix D for the calculation in R.

4.7.3 With the Q Statistic

If we consider the approximation to the chi-square distribution with seven degrees of freedom of the original proposal, the Type II error (β) can be computed using the cumulative distribution function of a non-central chi-square distribution with seven degrees of freedom. Its non-centrality parameter is set to the value of the test statistic Q and is evaluated at the critical level corresponding to the true alternative hypothesis (see Equation 23). The statistical power (φ) is the complement of this cumulative probability (Equation 24). See Appendix E for the calculation in R.

$$\beta = \chi^2_{df=7\,,NCP=Q}\left({}_{1-\alpha}\chi^2_{df}\right) \tag{23}$$

$$\phi = 1 - \beta = 1 - \chi^2_{df\,=\,7,NCP=\,Q}\left({}_{0.90}\chi^2_{7}\right) \tag{24}$$

4.8. Effect Size

4.8.1 With the Q_B Statistic

The effect size of the deviation from normality for the Q statistic can be calculated as a standardized distance, analogous to Cohen's *d* statistic [35]. The mean of Q in the normative bootstrap sampling distribution (obtained from 2,000 bootstrap samples of size n drawn from a standard normal distribution) serves as the expected value of Q, while the sample standard deviation of Q in that distribution represents the bootstrap standard error.

Using these two values, the standardization, denoted by z, is performed. This value can be interpreted in relation to the dispersion of a normal distribution: a z value below 1.7 indicates a trivial effect size, below 2.5 a small effect, below 3 a medium effect, and above 3 a very large effect (Equation 25). See Appendix B for the calculation in R.

$$Z = \frac{\left|Q - \hat{Q}^{N}_{bootstrap}\right|}{se^{N}_{bootstrap}} = \begin{cases} < 1.7 & trivial \\ < 2.5 & small \\ < 3 & median \\ \geq 3 & large \end{cases} \tag{25}$$

4.8.2 With the Q_T Statistic

For the quadratic form Q_T, whose vector consists of correlated standard normal variables and whose symmetric matrix is the correlation matrix of these variables, the mathematical expectation of Q_T is given by the trace of the correlation matrix R, i.e., $E[QT] = \mathrm{tr}(R) = 7$. The variance of Q_T is equal to twice the sum of the squared eigenvalues of R [36-37].

The seven eigenvalues of the correlation matrix of the quantiles from the PSNS of the normal distribution are: 3.281666, 1.449412, 0.8942277, 0.8293143, 0.4113026, 0.06717703, and 0.06690025.

Consequently, the standard deviation of Q_T is the square root of twice the sum of the squares of these eigenvalues, resulting in $\sigma(Q_T) = 5.3918$.

By standardizing Q_T using its mean and standard deviation, a measure of effect size in terms of standardized distance is obtained. This measure is analogous to Cohen's d [35], though applied in the context of tests based on quadratic forms rather than mean differences. It can be interpreted relative to the dispersion of a normal distribution: a z-value below 1.7 indicates a trivial effect size, below 2.5 a small effect, below 3 a medium effect, and above 3 a very large effect (Equation 26). See Appendix D for the calculation in R.

$$Q_T = z^t \times \mathrm{R} \times z \sim \chi_g^2(7, \mathrm{R})$$

$$Z = \frac{|Q_T - E(Q_T)|}{\sigma(Q_T)} = \frac{|Q_T - tr(R)|}{\sqrt{2 \times \sum_{i=1}^{7} \lambda_i^2}} = \frac{|Q_T - 7|}{5.3918} = \begin{cases} < 1.7 & trivial \\ < 2.5 & small \\ < 3 & median \\ \geq 3 & large \end{cases} \tag{26}$$

4.8.3 With Q Statistic

If we consider the approximation to the chi-square distribution with seven degrees of freedom of the original proposal, the effect size of the deviation from normality in the sample distribution can be measured using Cramér's V coefficient [38], which ranges from 0 to 1 [39]. Following Cohen [35], V values below 0.1 indicate a trivial effect size, values between 0.1 and 0.29 indicate a small effect, between 0.3 and 0.49 a medium effect, between 0.5 and 0.69 a large effect, and values of 0.7 or higher indicate a very large effect (Equation 27). See Appendix E for the calculation in R.

$$V = \sqrt{\frac{Q}{df \times n}} = \sqrt{\frac{Q}{7 \times n}} = \begin{cases} < 0.1 & trivial \\ < 0.3 & small \\ < 0.5 & median \\ < 0.7 & large \\ \geq 0.7 & huge \end{cases} \tag{27}$$

5. Materials and Methods

5.1. Sample Generation

Samples were generated using the reverse transform procedure [40] in the R program for 31 sample sizes (ranging from 50 to 1500, with increments of 50 and 2000) and 12 continuous distributions: Laplace (mu = 50, sigma = 10), normal (mean = 50, sd = 10), Student's t-distribution with 4 degrees of freedom, linearly transformed to have a mean of 50 and a standard deviation of 10, chi-square with 5 degrees of freedom, arcsine or beta (shape1 = 0.5, shape2 = 0.5), uniform (min = 0, max = 100), triangular (min = 0, max = 3, mode = 2.99), semicircular Wigner or beta (shape1 = 1.5, shape2 = 1.5), linearly transformed to have a radius of 3, exponential (rate = 2), hyperbolic drying linearly transformed to have a mean of 50 and a standard deviation of 10, logistic (location = 50, scale = 10), and Rayleigh (sigma = 12). A seed was set for result reproduction (123, except for 12 with the normal distribution). You can visualize the samples used with the following script [41]. The result obtained from running the script in Appendix G with the first listed distribution—without a hash symbol, which is the Laplace distribution—is shown in Table 4. For presentation purposes, each one of 31 original samples (of different size) drawn from Laplace distribution in the Table 4 was reduced to the minimum, median and maximum values, with the other values being indicated by ellipses (…).

Table 4. Original sample drawn from a Laplace's distribution

Sample size	Minimum	...	Median	...	Maximum
50	9.5898829	...	47.61486	...	77.26236
100	20.4659908	...	50.09903	...	93.65911
150	-6.1995893	...	48.75828	...	83.52246
200	-14.2188404	...	49.47437	...	99.03407
250	-2.2532879	...	50.11989	...	103.48888
300	-1.2977199	...	49.56351	...	97.06442

350	-14.1836689	...	50.73552	...	100.83601
400	-40.8138988	...	49.94166	...	99.69524
450	-5.7418691	...	49.28067	...	94.29004
500	-9.2878793	...	49.86201	...	121.60666
550	-26.2334393	...	49.62307	...	102.91311
600	0.7197058	...	50.27604	...	106.89202
650	-8.4086652	...	50.04799	...	99.44288
700	-35.1873205	...	49.93920	...	124.71132
750	-4.3566404	...	49.76694	...	106.39675
800	-27.0054577	...	50.10379	...	135.88705
850	-6.2540337	...	49.47890	...	160.39084
900	-23.8530300	...	49.94832	...	112.82285
950	-21.2402997	...	49.16473	...	109.56630
1000	-10.3551434	...	50.22167	...	121.36216
1050	-10.9874128	...	49.84465	...	126.42730
1100	-17.6727651	...	49.13120	...	112.89938
1150	-19.3386696	...	49.82527	...	114.30884
1200	-34.2327092	...	50.41611	...	112.19349
1250	-11.8510114	...	50.06876	...	105.43839
1300	1.8400462	...	49.71477	...	119.94513
1350	-24.1708992	...	49.83466	...	129.00185
1400	-17.6022695	...	50.03964	...	108.28709
1450	-34.8161317	...	50.20337	...	106.84666
1500	-26.7905853	...	49.88449	...	111.58504
2000	-27.9725620	...	50.11953	...	117.70299

5.2. Statistical Analysis

Two criteria were considered to evaluate the performance of the normality tests. One criterion was the proportion of hits, defined as the ratio between the number of correct or successful identifications and the total number of original samples drawn from each distribution (31 samples). A hit occurs when the null hypothesis of normality is retained at the 5% significance level for a sample drawn from a normal distribution with a mean of 50 and a standard deviation of 10. Conversely, a hit also occurs when the hypothesis is rejected at the 5% level for a sample drawn from a non-normal distribution. Another criterion was statistical power at the 5% significance level. For samples drawn from a non-normal distribution, power was defined as the probability of correctly rejecting the null hypothesis of normality, that is, detecting a deviation from normality when it truly exists. In the case of normal samples, its complement, the Type II error (i.e., the probability of failing to reject the null hypothesis when it is false), was assessed. A value close to 1 in this context would indicate a high probability of retaining the null hypothesis when the data are actually normal.

Since a single seed was used (123 for non-normal distributions and 12 for the normal distribution) to generate the 31-size samples across 12 distributions, the four normality tests are treated as a repeated-measures factor. Accordingly, hit rates were compared using Cochran's Q-test, with post-hoc comparisons performed using Dunn's test with Bonferroni correction [42]. The effect size in the omnibus test was measured using the eta-squared coefficient [43]. Confidence intervals for the hit rates were calculated using Wilson's procedure [44].

Statistical power means were compared using a repeated measures analysis of variance. The normality test served as the within-subjects factor with four levels. Distribution was included as a between-subjects factor with twelve independent groups, and sample size was treated as a covariate. The model included the main effect of the test as well as its interactions with sample size and distribution type.

The assumption of homogeneity of covariance matrices was assessed using Box's M test, and sphericity was evaluated using Mauchly's test. Due to a violation of the sphericity assumption, the multivariate approach based on Wilks' lambda was applied [41]. Effect size was estimated using partial eta squared (η^2_p) [43].

Normality in the distributions of mean differences was assessed using the Shapiro-Wilk W test [12-13]. In case of violation of this assumption, pairwise comparisons between the 4 levels of the repeated measures factor were performed using Student's t-test, with the p-value calculated using the bootstrap procedure (by drawing 1000 samples with replacement from the original sample), and confidence intervals were computed using the Bias-Corrected and Accelerated (BCa) percentile method. The interactions between the test and the distribution were examined by checking whether the 95% bootstrap BCa confidence intervals for the power means overlapped (not significant) or did not overlap (significant), in addition to analyzing crossovers in the power mean plot with lines separated by normality test (4 lines) [46]. The interactions between the test and the sample size were assessed using the power mean plot with lines separated by normality test, also evaluating the crossovers.

Statistical analyses were executed with R version 4.4 [41], Excel version 365 [47] and SPSS version 2021 [48].

6. Results

6.1. Comparison of Hit Ratio or Test Accuracy

When comparing the number of hits among the four normality tests using Cochran's *Q* test for the equality of binary proportions in related samples, the difference was significant ($Q_{[df = 3,\, N = 372]} = 158{,}980$, asymptotic two-tailed p-value < 0.001), with a large effect size calculated through the eta-squared coefficient ($\eta^2 = 0.142 > 0.14$). Comparisons performed using Dunn's post-hoc test revealed that five out of six differences were statistically significant, even after applying Bonferroni correction. The Shapiro-Wilk *W* test showed the highest hit rate (0.976, 95% Wilson-type *CI* [0.955, 0.987]), followed by the *Q* test based on bootstrap procedures (0.952, 95% Wilson-type *CI* [0.925, 0.969]), with their hit rates being statistically equivalent ($t = 1.137$, $p = 0.256$, $p_{BC} = 1$). Third place was for the Q test based on the chi-square distribution with seven degrees of freedom (0.849, 95% Wilson-type CI [0.810, 0.882]). The Q test based on a quadratic form using a generalized chi-square distribution for p-value calculation yielded the lowest hit rate (0.766, 95% Wilson-type *CI* [0.721, 0.806]). Refer to Table 5 and Figure 1 for further details. In the representation of the hit-and-miss proportion for each normality test in Figure 1, yellow is used for hits and red for misses.

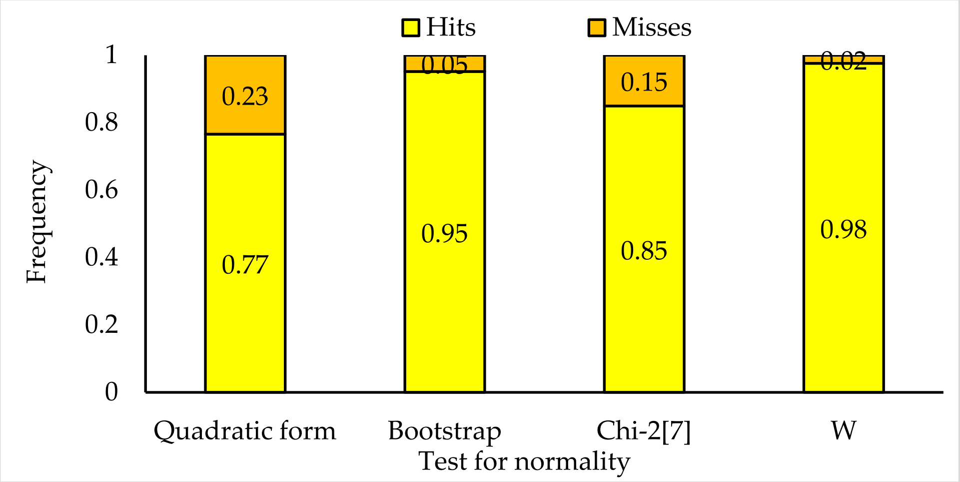

Figure 1. Stacked bar plot illustrates the hit-and-miss proportions for each of the four normality tests. These proportions reflect the retention of the null hypothesis of normality for samples drawn from a normal distribution, or its rejection for samples drawn from non-normal distributions, using a significance level of 0.05. Normality tests: Quadratic Form = *Q* test based on the quadratic form and using the generalized chi-square distribution to calculate the probability value; Bootstrap = *Q* test employing the bootstrap approach for this calculation; Chi-2[7] = *Q* test with approximate probability values calculated using the chi-square distribution

with seven degrees of freedom, assuming independence among random variables; and W = Shapiro-Wilk *W* test standardized using Royston's method [12-13].

Table 5. Pairwise comparisons of the proportions of hits using Dunn's test

Test 1 - Test 2	*MSD*	*se*	*t*	*p*	p_c
Hits_Chi2 - Hits_Quadratic	0.083	0.019	4.406	< 0.001	< 0.001
Hits_Bootstrap - Hits_Quadratic	0.188	0.019	9.949	< 0.001	< 0.001
Hits_W - Hits_Quadratic	0.210	0.019	11.086	< 0.001	< 0.001
Hits_Bootstrap - Hits_Chi2 -	-0.105	0.019	-5.543	< 0.001	< 0.001
Hits_W - Hits_Chi2	0.126	0.019	6.680	< 0.001	< 0.001
Hits_W - Hits_Bootstrap	0.022	0.019	1.137	0.256	1

Note. *MSD* = Minimum Significant Difference. Each row tests the null hypothesis that the proportions of hits in Test 1 and Test 2 are equal. *se* = asymptotic standard error, *t* = test statistic value, *p* = asymptotic two-tailed probability value, and p_c = adjusted probability value using Bonferroni correction at a significance level of 0.05. Hits corresponding to each normality test: Quadratic = *Q* test based on the quadratic form, using the generalized chi-square distribution to compute the probability value; Bootstrap = *Q* test employing the bootstrap approach for this computation; Chi2 = *Q* test using an approximate probability calculation based on the chi-square distribution with seven degrees of freedom (which assumes independence between random variables); and *W* = Shapiro-Wilk *W* test standardized using Royston's method.

When calculating the confidence intervals for the hit rates per distribution using Wilson's formula [44], the Q_T test showed significantly lower proportions compared to the other three tests (Q_B, Q, and W) for the chi-square and uniform distributions. For the Rayleigh and logistic distributions, the hit rates of the Q_T and Q tests were significantly lower than those of the W and Q_B tests, whose confidence intervals overlapped. Across the full set of 12 distributions, the hit rates for the Q_T and Q tests were significantly lower than those of the W and Q_B tests, which were statistically equivalent (Table 6).

Table 6. Proportions of hits and their 95% Wilson-type confidence intervals for each normality test across different distributions

Distribution	Q_T	Q_B	Q	W
Laplace	0.871 [0.711, 0.949]	0.935 [0.793, 0.982]	0.903 [0.751, 0.967]	0.968 [0.838, 0.994]
Normal	1 [0.890, 1]	1 [0.890, 1]	1 [0.890, 1]	1 [0.890, 1]
Student	0.968 [0.838, 0.994]	0.935 [0.793, 0.982]	0.968 [0.838, 0.994]	1 [0.890, 1]
Chi-2	**0.484** **[0.320, 0.652]**	1 [0.890, 1]	0.968 [0.838, 0.994]	1 [0.890, 1]
Arcsine	0.935 [0.793, 0.982]	1 [0.890, 1]	0.968 [0.838, 0.994]	1 [0.890, 1]
Uniform	**0.774** **[0.602, 0.886]**	0.968 [0.838, 0.994]	0.871 [0.711, 0.949]	1 [0.890, 1]
Triangular	0.839 [0.674, 0.929]	0.968 [0.838, 0.994]	0.871 [0.711, 0.949]	1 [0.890, 1]
Semicircle	0.839 [0.674, 0.929]	0.903 [0.751, 0.967]	0.871 [0.711, 0.949]	1 [0.890, 1]
Exponential	1 [0.890, 1]	1 [0.890, 1]	1 [0.890, 1]	1 [0.890, 1]
HSD	0.871 [0.711, 0.949]	0.935 [0.793, 0.982]	0.774 [0.602, 0.886]	0.935 [0.793, 0.982]
Rayleigh	**0.645** **[0.469, 0.789]**	0.968 [0.838, 0.994]	**0.645** **[0.469, 0.789]**	0.968 [0.838, 0.994]
Logist	**0.355**	0.839	**0.516**	0.871

	[0.211, 0.531]	[0.674, 0.929]	**[0.348, 0.680]**	[0.711, 0.949]
Total	**0.766**	0.954	**0.849**	0.976
	[0.721, 0.806]	[0.928, 0.971]	**[0.810, 0.882]**	[0.955, 0.987]

Note. Normality tests: Q_T = Q test based on the quadratic form using the generalized chi-square distribution to calculate the probability value; B = Q test employing the bootstrap approach for this calculation; $\chi^2[7]$ = Q test with approximate probability values derived from the chi-square distribution with seven degrees of freedom, assuming independence among random variables; and W = Shapiro-Wilk W test standardized using Royston's method [12–13].

6.2. Comparison of Mean Statistical Power

A repeated-measures model was defined to compare the mean statistical power of the four normality tests. The type of distribution was included as a fixed-effects factor for independent measures, while sample size was treated as a covariate. The assumption of sphericity was not met, as indicated by the Mauchly test ($W = 0.900$, $\chi^2[5] = 37.615$, $p < 0.001$), and the null hypothesis of equal covariance matrices was rejected by the Box test (M of Box = 3005.680, $F[90, 107675.456] = 28.462$, $p < 0.001$). Consequently, the multivariate omnibus test of Wilks' lambda was used.

The effect of normality tests (an intragroup factor with four levels) was significant, with a large effect size. Additionally, the interaction between the normality test and sample size (a covariate with 31 values) was significant, showing a medium effect size, as was the interaction between the normality test and the type of distribution (a fixed-effect factor with 12 levels), which had a large effect size. The statistical power for these differences was unitary. For more details, refer to Table 7.

Table 7. Multivariate test: Wilks lambda

Effect	**Λ**	***F***	***df_1***	***df_2***	***p***	**η_p^2**	***NCP***	**ϕ**
Test	0.710	48,700	3	357	< 0.001	0.290	146.101	1
Test*N	0.840	22,679	3	357	< 0.001	0.160	68.037	0.999
Test*Dist	0.407	11.375	33	1052.492	< 0.001	0.259	367.578	1

Note. Intra-subject design: normality test (with four levels). Interaction of the test with N (sample size) and Distr. (type of distribution). The sample size includes 31 values (treated as a covariate), and the type of distribution encompasses 12 categories (a fixed-effects factor for independent groups). λ = Wilks Lambda statistic, where the asterisk denotes its exact calculation; df = degrees of freedom; p = right-tailed probability in the Snedecor-Fisher F distribution; η_p^2 = partial eta-squared coefficient; NCP = non-centrality parameter; and φ = statistical power calculated at a significance level of 0.05.

The assumption of a normal distribution in the six power differences among the four normality tests is satisfied only with the logistic distribution in three out of six comparisons and with the normal distribution in two out of six comparisons. For the other distributions, as well as in the combined sample, the null hypothesis of normality is rejected using the Shapiro-Wilk test [12-13] at a significance level of 5% (Table 8).

Table 8. Testing normality for the difference in power between the four normality tests

Distribution	Diff.	*n*	*W*	*z*	*p*	Diff.	*n*	*W*	*z*	*p*
Laplace	Q_T-B	31	0.651	5.036	<0.001	B-χ^2	31	0.445	5.998	<0.001
Normal		31	0.919	2.010	0.022		31	0.928	1.766	0.039
Student		31	0.514	5.723	<0.001		31	0.472	5.894	<0.001
Chi-2		31	0.789	3.994	<0.001		31	0.326	6.400	<0.001
Arcsine		31	0.269	6.568	<0.001		31	0.269	6.568	<0.001
Uniform		31	0.451	5.975	<0.001		31	0.533	5.640	<0.001
Triangular		31	0.616	5.234	<0.001		31	0.489	5.826	<0.001
Semicircle		31	0.54	5.609	<0.001		31	0.787	4.013	<0.001
Exponential		31	0.176	6.816	<0.001		31	0.192	6.776	<0.001
Hyperbolic		31	0.69	4.791	<0.001		31	0.643	5.083	<0.001
Rayleigh		31	0.522	5.688	<0.001		31	0.723	4.558	<0.001
Logistic		31	0.965	0.272	0.389		31	0.945	1.208	0.116

Total		372	0.683	10.444	<0.001		372	0.704	10.281	<0.001
Laplace	Q_T-χ^2	31	0.676	4.882	<0.001	B-W	31	0.4	6.159	<0.001
Normal		31	0.97	-0.048	0.513		31	0.945	1.208	0.116
Student		31	0.334	6.375	<0.001		31	0.44	6.016	<0.001
Chi-2		31	0.829	3.558	<0.001		31	0.263	6.585	<0.001
Arcsine		31	0.269	6.568	<0.001		31	0.202	6.750	<0.001
Uniform		31	0.566	5.488	<0.001		31	0.326	6.400	<0.001
Triangular		31	0.652	5.030	<0.001		31	0.414	6.110	<0.001
Semicircle		31	0.694	4.764	<0.001		31	0.375	6.244	<0.001
Exponential		31	0.282	6.531	<0.001		31	0.179	6.809	<0.001
Hyperbolic		31	0.65	5.042	<0.001		31	0.644	5.078	<0.001
Rayleigh		31	0.717	4.602	<0.001		31	0.54	5.609	<0.001
Logistic		31	0.875	2.909	0.002		31	0.829	3.558	<0.001
Total		372	0.667	10.56	<0.001		372	0.549	11.280	<0.001
Laplace	Q_T-W	31	0.639	5.106	<0.001	χ^2-W	31	0.435	6.035	<0.001
Normal		31	0.738	4.442	<0.001		31	0.755	4.303	<0.001
Student		31	0.356	6.306	<0.001		31	0.313	6.440	<0.001
Chi-2		31	0.797	3.914	<0.001		31	0.332	6.382	<0.001
Arcsine		31	0.282	6.531	<0.001		31	0.276	6.548	<0.001
Uniform		31	0.456	5.956	<0.001		31	0.527	5.666	<0.001
Triangular		31	0.616	5.234	<0.001		31	0.487	5.835	<0.001
Semicircle		31	0.559	5.521	<0.001		31	0.721	4.573	<0.001
Exponential		31	0.331	6.385	<0.001		31	0.277	6.545	<0.001
Hyperbolic		31	0.694	4.764	<0.001		31	0.686	4.817	<0.001
Rayleigh		31	0.468	5.910	<0.001		31	0.702	4.709	<0.001
Logistic		31	0.927	1.795	0.035		31	0.956	0.746	0.232
Total		372	0.691	10.383	<0.001		372	0.651	10.672	<0.001

Note. Diff = random variable of the power difference between tests of normality: Q_T = *Q* test from quadratic form using the generalized chi-square distribution for the calculation of the probability value, B = *Q* test from its bootstrap approach and $\chi^2_{[7]}$ = *Q* test from the chi-square approximation with seven degrees of freedom, and *W* = Shapiro-Wilk test with Royston's standardization. Test to check the normality of power differences: *W* = Shapiro-Wilk test statistic, *n* = sample size, *z* = standardized value of the logarithmic transformation of *W* (Royston's standardization), and *p* = probability to the right tail in a standard normal distribution.

When pairwise comparisons were conducted using the paired-samples Student's t-test with a bootstrap procedure for calculating p-values and obtaining BCa percentile confidence intervals (1,000 replications), the mean power of the W test was significantly higher than that of the other three tests. The statistical power of the bootstrap version of the Q test was significantly higher than that of the other two versions, which did not differ significantly from each other. See Table 9, which presents the group means, the mean differences with their corresponding 95% BCa confidence intervals, and the two-tailed bootstrap p-values.

Table 9. Comparison of mean statistical power using Student's t-test for two matched samples, with the BCa method applied for calculating the bootstrap confidence interval.

Tets	m_1	m_2	*md*	*p_boot*
Q_T-B	0.902 [0.882, 0.922]	0.935 [0.920, 0.950]^b	-0.033 [-0.049, -0.018]^b	0.001
Q_T-$\chi^2_{[7]}$		0.902 [0.881, 0.920]^b	0 [-0.014, 0.014]^b	0.983
Q_T-W		0.961 [0.946, 0.975]^b	-0.059 [-0.075, -0.046]^b	0.001
B-$\chi^2_{[7]}$	0.935	0.902	0.034	0.001

	[0.920, 0.950][b]	[0.881, 0.920][b]	[0.020, 0.049][b]	
B-W		0.961 [0.946, 0.975][b]	-0.040 [-0.055, -0.026][b]	0.001
χ^2-W	0.902 [0.881, 0.920][b]	0.961 [0.946, 0.975][b]	-0.060 [-0.074, -0.046][b]	0.001

Note. Statistical tests of normality: Q_T = *Q* test based on the quadratic form, using the generalized chi-square distribution to compute the p-value; B = *Q* test employing the bootstrap approach; χ^2[7] = *Q* test using the chi-square approximation with seven degrees of freedom; and *W* = Shapiro-Wilk test standardized using Royston's method. m_1 = mean power of the test serving as the minuend, m_2 = mean power of the test serving as the subtrahend, *md* = difference between the two means (mean difference) with a 95% bootstrap confidence interval computed using the BCa method, and p_boot = bootstrap probability for the null hypothesis of mean equality.

Figure 2 displays the average power values for the non-normal distributions and the average complement of power (beta error) for the normal distribution for the four normality tests. The curves are represented in different colors: the Shapiro–Wilk test is shown in yellow, the Q test (quadratic version) in blue, the bootstrap version of the Q test in red, and the version of the Q test that uses an approximate p-value based on the chi-square distribution with seven degrees of freedom in black. Statistical power increased with sample size. The W test showed the fastest improvement as sample size increased, exhibiting higher average power than the other tests. The bootstrap version of the Q test outperformed the other two Q test versions from the smallest sample size. At a sample size of 1,000, all four tests reached unit power.

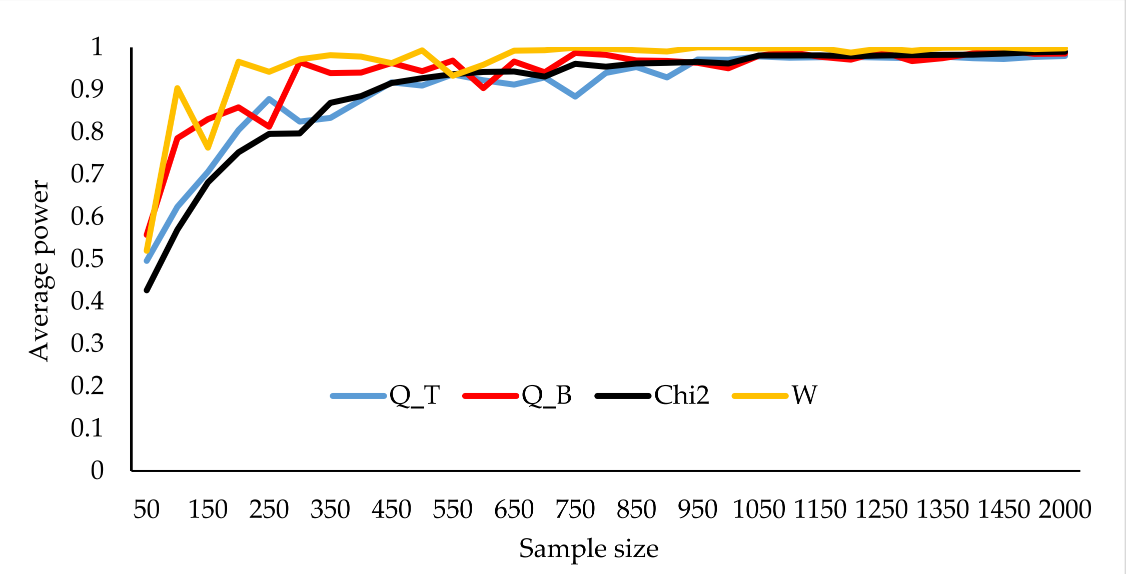

Figure 2. Diagram of the marginal means of power by normality test in relation to sample size. Tests: Q_T = *Q* test based on the quadratic form, using the generalized chi-square distribution to calculate the probability value; B = *Q* test employing the bootstrap approach; Chi2 = *Q* test with approximate probability values derived from the chi-square distribution with seven degrees of freedom (assuming independence among random variables); and W = Shapiro-Wilk test standardized using Royston's method [12-13].

The interaction effect between the distribution and the normality test was assessed using the mean diagram (Figure 3) and the overlap of the 95% bootstrap confidence intervals obtained via the BCa percentile method. If no overlap was observed, the difference was considered significant in a two-tailed test at a 5% significance level (Table 10).

Table 10. Mean power in the interaction between the distribution and the test, with 95% bootstrap confidence intervals obtained via the BCa method

Distr.	Q_T	**Bootstrap**	χ^2**[7]**	*W*
Laplace	0.909 [0.834, 0.969][b]	0.950 [0.871, 0.998][b]	0.935 [0.863, 0.991][b]	0.967 [0.895, 1][b]
Normal	0.830 [0.812, 0.845][b]	**0.724 [0.667, 0.776]**[b]	0.853 [0.836, 0.870][b]	0.947 [0.910, 0.973][b]

Student	0.979 [0.927, 1][b]	0.946 [0.891, 0.993][b]	0.985 [0.942, 1][b]	0.989 [0.963, 1][b]
Chi-2	**0.806 [0.708, 0.890][b]**	0.997 [0.988, 1][b]	0.986 [0.963, 1][b]	0.995 [0.983, 1][b]
Arcsine	0.992 [0.976, 1][b]	0.999 [0.995, 1][b]	0.981 [0.943, 1][b]	0.991 [0.991, 1][b]
Uniforme	0.964 [0.915, 0.996][b]	0.982 [0.938, 1][b]	0.941 [0.874, 0.989][b]	0.955 [0.956, 1][b]
Triangular	0.873 [0.778, 0.956][b]	0.988 [0.956, 1][b]	0.944 [0.882, 0.993][b]	0.985 [0.959, 1][b]
Semicircular	0.908 [0.821, 0.974][b]	0.935 [0.857, 0.993][b]	0.830 [0.728, 0.916][b]	0.954 [0.882, 0.998][b]
Exponential	1 [0.999, 1][b]	1 [0.999, 1][b]	1 [0.999, 1][b]	1 [0.999, 1][b]
Secant	0.909 [0.830, 0.971][b]	0.924 [0.846, 0.982][b]	0.866 [0.768, 0.949][b]	0.934 [0.830, 0.996][b]
Rayleigh	0.959 [0.906, 0.992][b]	0.951 [0.889, 0.996][b]	0.795 [0.678, 0.898][b]	0.948 [0.859, 0.999][b]
Logistic	0.693 [0.611, 0.774][b]	0.830 [0.739, 0.909][b]	0.705 [0.608, 0.794][b]	0.836 [0.719, 0.927][b]

Note. Normality test: Q_T = *Q* test from the quadratic form using the generalized chi-square distribution for the calculation of the probability value, Bootstrap = *Q* test from the bootstrap approach for this calculation, $\chi^2[7]$ = *Q* test with the approximate probability calculation from the chi-square distribution with seven degrees of freedom (which implies assuming independence between random variables), and *W* = Shapiro-Wilk *W* test with Royston's standardization.

The mean diagram in Figure 3 shows that the highest average power is achieved by the Shapiro–Wilk W test across all distributions, except for the Rayleigh distribution, where the maximum occurs with the quadratic form of the Q test (Q_T). For the normal distribution, the average power of the bootstrap version of the Q test (Q_B) was significantly lower than that of the other three tests (Table 10). Additionally, for the chi-square distribution, the average power of the quadratic form of the Q test was significantly lower (Q_T) than that of the other three tests (Table 10). For the remaining distributions, no significant differences in average power were found among the four normality tests (Table 10).

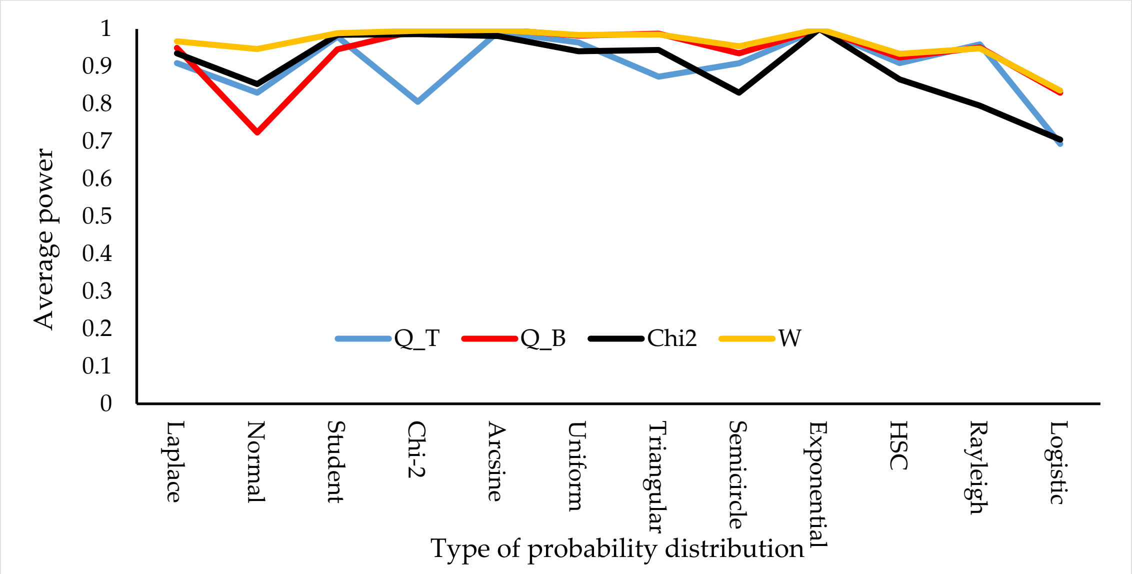

Figure 3. Diagram of the marginal mean power of the normality tests in relation to the type of distribution. Normality test: Q_T = Q test based on the quadratic form, using the generalized chi-square distribution to compute the probability value; B = Q test employing the bootstrap approach for this computation; Chi2 = Q test using an approximate probability calculation based on the chi-square distribution with seven degrees of freedom (which assumes independence between random variables); and W = Shapiro-Wilk W test standardized using Royston's method.

6.3. Application of R scripts to real-world data

Alexithymia was measured using the 20-item Toronto Alexithymia Scale (TAS-20) [18], adapted for the Mexican population by Moral [19]. The scale consists of 20 Likert-type items with six response options (0 to 5), yielding a total score ranging from 0 to 100. The sample consisted of 385 participants (85 men and 300 women) who took the entrance examination for a psychology program at a public university in northeastern Mexico. Below is the random vector of TAS-20 total scores.

```
x <- c(34, 14, 13, 17, 44, 4, 6, 33, 22, 28, 14, 23, 13, 12, 32, 12, 7, 48, 55, 45, 24, 13, 61, 23, 55, 35, 11, 13, 29, 30,
10, 29, 29, 26, 14, 41, 11, 17, 21, 26, 10, 38, 24, 20, 50, 26, 27, 27, 25, 17, 16, 27, 20, 28, 35, 28, 27, 40, 10, 12,
21, 19, 23, 40, 22, 4, 8, 26, 14, 35, 47, 24, 21, 31, 33, 23, 0, 33, 36, 31, 24, 25, 25, 4, 43, 21, 41, 26, 27, 21, 21, 32,
12, 22, 18, 30, 36, 26, 9, 38, 18, 35, 31, 41, 26, 15, 17, 31, 24, 44, 56, 16, 52, 30, 24, 25, 52, 44, 4, 19, 34, 22, 28,
31, 3, 12, 0, 34, 10, 44, 18, 17, 7, 48, 20, 17, 17, 42, 9, 22, 18, 58, 49, 32, 24, 16, 22, 9, 25, 28, 27, 46, 30, 10, 15,
29, 22, 18, 16, 13, 34, 13, 12, 14, 25, 33, 52, 5, 35, 20, 16, 28, 17, 20, 5, 25, 35, 9, 35, 45, 16, 26, 32, 27, 20, 27,
10, 23, 24, 20, 43, 28, 15, 21, 26, 32, 22, 51, 11, 27, 39, 21, 26, 18, 20, 19, 26, 14, 29, 25, 29, 23, 6, 19, 31, 5, 20,
33, 12, 32, 14, 24, 45, 39, 21, 20, 33, 18, 23, 15, 6, 25, 8, 54, 36, 16, 12, 28, 25, 17, 50, 7, 31, 28, 20, 20, 20, 27,
29, 32, 34, 41, 29, 31, 18, 4, 19, 30, 31, 5, 12, 35, 34, 17, 39, 14, 23, 18, 21, 22, 31, 36, 19, 50, 23, 10, 31, 36, 29,
21, 13, 28, 43, 17, 24, 43, 20, 17, 60, 39, 22, 30, 19, 37, 25, 21, 30, 9, 39, 21, 10, 17, 45, 33, 32, 33, 14, 31, 33, 16,
24, 19, 8, 4, 27, 21, 58, 38, 40, 14, 21, 14, 2, 29, 18, 42, 25, 28, 49, 37, 24, 34, 33, 30, 23, 10, 30, 46, 12, 38, 11,
32, 27, 37, 7, 25, 24, 12, 22, 10, 5, 24, 24, 20, 19, 16, 29, 28, 18, 19, 28, 17, 13, 47, 27, 15, 13, 25, 35, 6, 19, 45,
36, 17, 32, 25, 5, 30, 37, 41, 46, 15, 14, 13, 32)
```

The script in Appendix A yields the following results. First, it produces a density histogram of the empirical distribution with the empirical and normal density curves overlaid (Figure 4). The histogram was constructed following the Freedman-Diaconis rule [49], and the density was estimated using Epanechnikov's kernel [50] and the Sheather-Jones bandwidth [51]. The histogram profile (12 yellow-filled bins) and its corresponding density curve (dark blue line) resemble the normal curve (red line), although a slight positive (right-tailed) skewness is noticeable (Figure 4). Second the script provides the sample quantiles corresponding to the PSNS (Table 11), along with the statistics needed to standardize these quantiles (sample size, mean, and standard deviation), and a table showing the calculations used to calculate the Q-statistic, which is obtained by summing the values in the last column of this table and equals 13.7043 (Table 12).

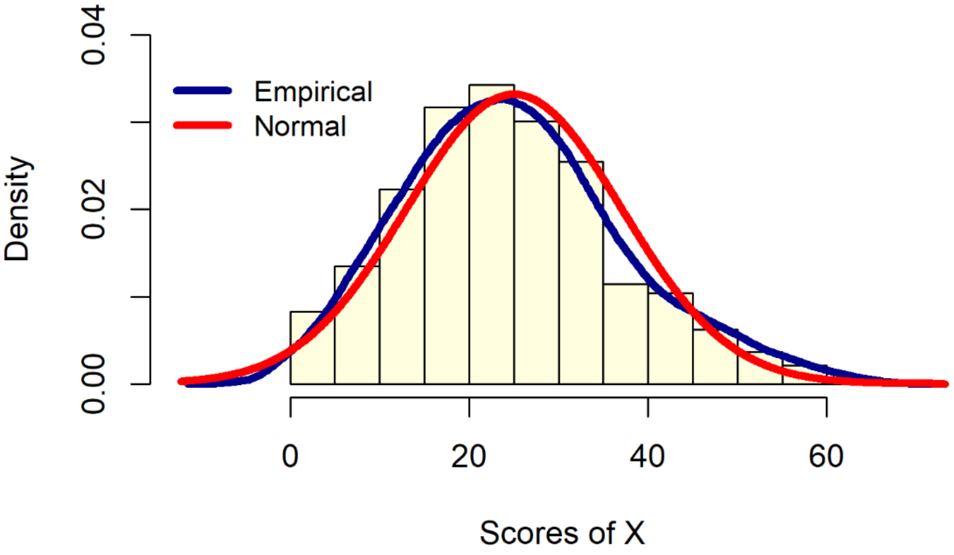

Figure 4. Density histogram of the empirical distribution (using the Freedman-Diaconis rule) with the empirical density curve (based on Epanechnikov’s kernel and the Sheather–Jones bandwidth) and the normal density curve overlaid.

Table 11. Sample quantiles corresponding to Parametric Seven-Number Summary (PSNS)

2.275013%	9.121122%	25.24925%	50%	74.75075%	90.87888%	97.72499%
4	10	17	24	32	43	52

Statistics needed to standardize sample quantiles corresponding to the PSNS:
Sample size: 385.
Sample mean: 25.01818.
Sample standard deviation: 12.01441.

Table 12. Testing normality using the test based on the Parametric Seven-Number Summary

e_zp	p	d_zp	ee_zp	x_p	z_xp	z	z_sq
-2	0.0228	0.054	0.1407	4	-1.7494	1.7804	3.1698
-1.3333	0.0912	0.164	0.0895	10	-1.25	0.9313	0.8673
-0.6667	0.2525	0.3194	0.0693	17	-0.6674	-0.0103	1e-04
0	0.5	0.3989	0.0639	24	-0.0847	-1.3268	1.7603
0.6667	0.7475	0.3194	0.0693	32	0.5811	-1.2342	1.5233
1.3333	0.9088	0.164	0.0895	43	1.4967	1.8259	3.3339
2	0.9772	0.054	0.1407	52	2.2458	1.7463	3.0495
Sum							13.7043

Note: e_zp = expected value for the p-th quantile in a N(0, 1) distribution, p = quantile order, d_zp = density of the p-th quantile in a N(0, 1) distribution, ee_zp = standard error of the p-th quantile for n normally distributed data points, x_p = p-th sample quantile, z_xp = standardized quantile (calculated using the sample mean and standard deviation), z = (z_xp - e_zp) / ee_zp = standardization of the z_xp statistic under the assumption of normality, z^2 = squared z_xp statistic.

Value of test statistic in original sample 13.70427.

The script in Appendix B applies bootstrap procedures to compute the critical value for a 5% significance level. This value is obtained from the normative bootstrap distribution (Figure 5) as its 0.95 quantile, using R's Type 8 quantile estimator (Q_critical = 6.43475). The script also calculates the p-value, defined as the proportion of values in the normative bootstrap distribution that are greater than the observed test statistic, based on the 2000 Q estimates that constitute the distribution (Figure 6). In addition, it computes the effect size as the standardized distance between the test statistic and its expected value (in the normative bootstrap distribution), using its standard deviation for scaling. The null hypothesis of normality is rejected (Q statistic= 13.70427 > Q_critical value = 6.43475; p-value = 0.0005 < α = 0.05), with a large effect size: Z = (13.70427 - 2.961112) / 1.902137 = 5.6479 > 3.

Bootstrap sampling distribution of the Q statistic generated from the original sample (Figure 5).

Number of extractions with replacement = 2000.

Bootstrap estimation of the Q = 17.59577.

Bootstrap standard error of the Q = 7.13478.

Bootstrap bias of the Q = 3.891502.

Normative bootstrap sampling distribution of the Q statistic (Figure 6).

This is generated from a normal distribution with mean 25.01818 and standard deviation 12.01441

Based on 2000 extractions of size 385.

Bootstrap expected value of Q under normal = 2.961112.

Bootstrap standard error of Q under normal = 1.902137.

Bootstrap critical value for the Q statistic with a significance level of 0.05 = 6.43475.

Bootstrap p-value = 0.0005.

The null hypothesis that the data follow a normal distribution is rejected with a significance level of 0.05 based on the bootstrap p-value.

Effect size as standardized distance: Z_Q = 5.647943.

Interpretation of effect size
< 1.7: Trivial
[1.7, 2.5): Small
[2.5, 3): Medium
$\geqslant$ 3: Large

Based on the Z statistic = 5.6479, the size effect of the deviation from the normal distribution model is classified as large.

The normative bootstrap distribution of the Q statistic (derived from normally distributed samples) is represented in Figure 5 by a histogram and a density curve, along with the chi-square distribution with 7 degrees of freedom (red curve). The vertical dotted lines indicate the test statistic (black) and the bootstrap critical value (purple). The normative bootstrap distribution appears more peaked and has a longer right tail than the chi-square distribution with 7 degrees of freedom.

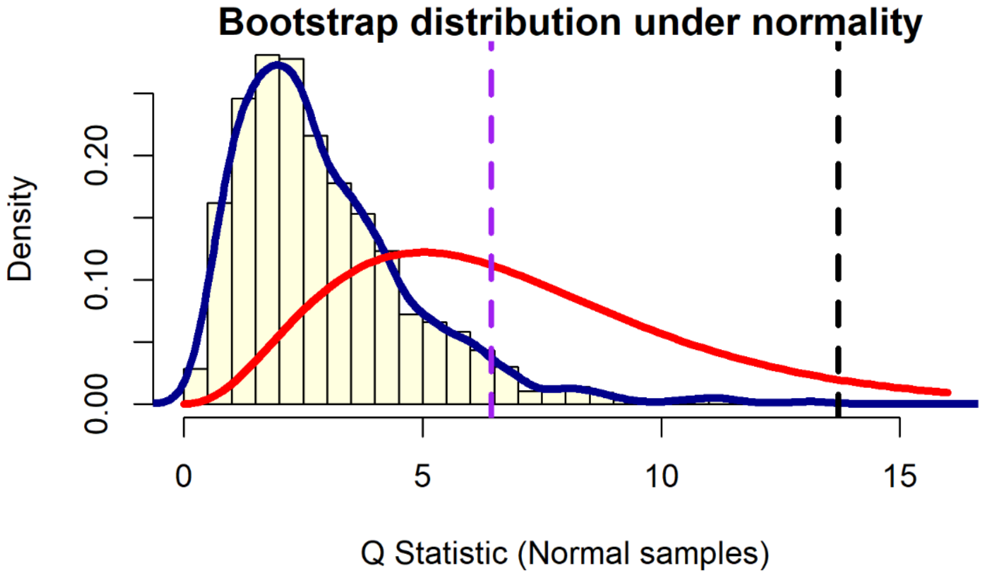

Figure 5. Histogram of normative bootstrap distribution (under null hypothesis of normality) of Q-Statistic with empirical and chi-2[7] density curves overlaid. The critical value (from the normative bootstrap distribution of the Q-statistic) is represented by a purple dotted line, while the test statistic value is shown with a black dotted line.

The empirical bootstrap distribution of the Q statistic (derived from the original sample) is shown in Figure 6 as a histogram and a density curve, together with the chi-square distribution with 7 degrees of freedom (red curve). The vertical dotted lines represent the test statistic (black) and the bootstrap critical value (purple). In contrast to what is observed in Figure 5, the chi-square distribution appears more peaked and exhibits a longer right tail than the empirical bootstrap distribution.

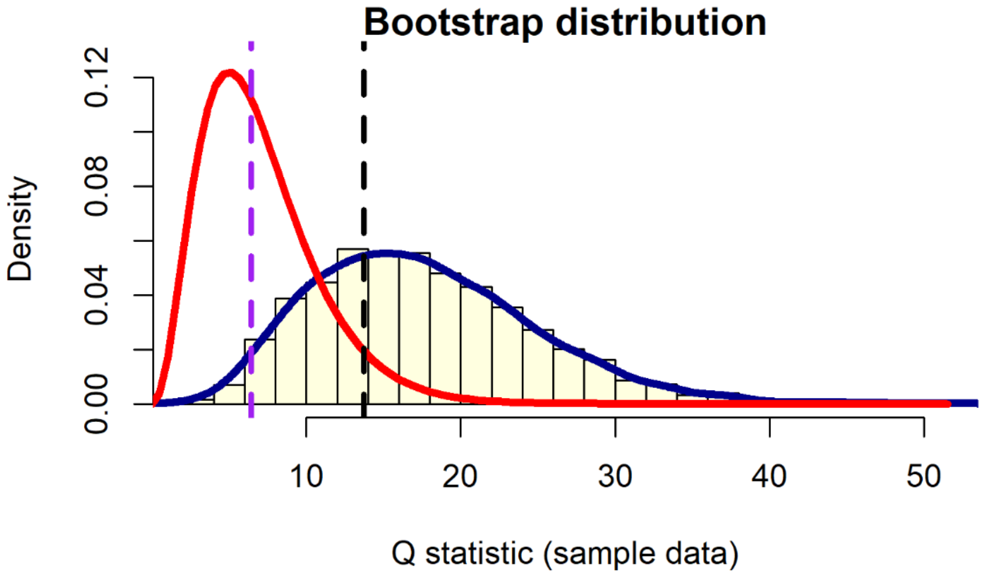

Figure 6. Histogram of bootstrap distribution (from original sample) of Q-Statistic with empirical and chi-2[7] density curves overlaid. The critical value (from the normative bootstrap distribution of the Q-statistic) is represented by a purple dotted line, while the test statistic value is shown with a black dotted line.

Appendix C script using 10,000 bootstrap samples (generated from original sample) computes statistical power or proportion of rejections across simulations. When you run the script, you get a very high power value ($\phi = 1 > 0.90$).

Number of bootstrap samples = 10000.

Number of data per bootstrap sample = 385.

Bootstrap power at a significance level of 0.05 = 1.

The script in Appendix D yields the quadratic form of the Q statistic. First, it presents the Gaussian vector with quantiles standardized under the null hypothesis of normality. Second, it displays the matrix of correlations between quantiles under the null hypothesis (Table 13). Third, it computes the scalar resulting from the quadratic form—namely, the row Gaussian vector multiplied by the square correlation matrix, and then by the column Gaussian vector—which gives the value of the test statistic (Q_T = 13.08073). The p-value is calculated using a generalized chi-square distribution and is greater than the significance level, so the null hypothesis of normality is retained (p-value = 0.9103499 > α = 0.05). Finally, the effect size is computed as the standardized distance between the test statistic and its expected value, which turns out to be trivial: Z_Q = (13.08073 − 7) / 4.829811 = 1.259 < 1.7.

Gaussian vector z:

z <- (1.780379, 0.9313009, -0.01029847, -1.326763, -1.23424, 1.825896, 1.746296)

Table 13. Correlation matrix of PSNS quantiles (under the assumption of normality)

Quantile	**[,1]**	**[,2]**	**[,3]**	**[,4]**	**[,5]**	**[,6]**	**[,7]**
[1,]	1	0.4816	0.2625	0.1526	0.0887	0.0483	0.0233
[2,]	0.4816	1	0.5451	0.3168	0.1841	0.1004	0.0483
[3,]	0.2625	0.5451	1	0.5812	0.3378	0.1841	0.0887
[4,]	0.1526	0.3168	0.5812	1	0.5812	0.3168	0.1526
[5,]	0.0887	0.1841	0.3378	0.5812	1	0.5451	0.2625
[6,]	0.0483	0.1004	0.1841	0.3168	0.5451	1	0.4816
[7,]	0.0233	0.0483	0.0887	0.1526	0.2625	0.4816	1

The Q statistic value in the quadratic form = 13.08073.

Eigenvalues: 2.747683, 1.535561, 0.9552459, 0.6382596, 0.4665808, 0.3647546, and 0.2919151.

Number of eigenvalues: 7.

Sum of eigenvalues: 7.

p-value for the Q_T statistic = 0.1019042.

Statistical power: 0.9103499.

Expected value of the Q_T statistic = 7.

Standard deviation of the Q_T statistic = 4.829811.

Effect size as standardized distance: Z_Q = 1.259.

Interpretation of effect size.

< 1.7: Trivial

[1.7, 2.5): Small

[2.5, 3): Medium

⩾ 3: Large

Based on Z statistic = 1.259, the effect size of the deviation from the normal distribution model is classified as trivial.

When you run the Appendix E script to calculate the p-value, statistical power, and effect size of the Q statistic using the chi-square approximation with 7 degrees of freedom, the result resembles the quadratic form of the Q statistic. This aligns with the hypothesized distribution of the TAS-20 total score but differs from the result obtained using the bootstrap version of the Q statistic, which does not support this hypothesis.

Test Statistic: Q = 13.70427, critical value with a significance level of 0.05 = 14.06714 , asymptotic p-value = 0.05669822.

The null hypothesis that the data follow a normal distribution is not rejected with a significance level of 0.05 by the Q test.

The statistical power to the right tail for the alternative hypothesis of non-normality for the Q test: ϕ = 0.2222.

If the null hypothesis of normality is maintained, the power must be less than 0.5; while, if rejected, it must be greater than 0.5. In the latter case, it is classified as good with a value of 0.8 and very good with a value of 0.9.

Effect size via Cramer's V-coefficient: V = 0.0713.

Interpretation of effect size based on Cohen (1988) [35]:

< 0.1: Trivial

[0.1, 0.3): Small

[0.3, 0.5): Medium

⩾ 0.5: Large

Based on the V statistic = 0.0713, the size of the effect of the deviation from the normal distribution model is classified as trivial.

The Shapiro–Wilk test, using Royston's procedure (Appendix H), confirms the result obtained with the bootstrap version of the Q statistic and the histogram of deviations from normality. In contrast, the quadratic form and chi-square versions of the Q statistic support the null hypothesis of normality. Consequently, this example validates the results of the analysis with simulated data, indicating that the bootstrap version is the most appropriate option for the PSNS Q test.

Shapiro-Wilk W-statistic: w = 0.9819425.

Logarithmic transformation of W-statistic: ln(1 - w) = -4.014193.

Expected value of ln(1 - w) = -5.583649.

Standard deviation of ln(1 - w) = 0.4210108.

Royston´s standardized W-statistic: z_w = 3.727829.

Right-tailed p-value = 9.656814e-05.

The null hypothesis that the data follow a normal distribution is rejected with a significance level of 0.05 by the Shapiro-Wilk W-test using Royston's procedure.

Statistical power: ϕ = 0.9813733.

It should be noted that both the Kolmogorov-Smirnov-Lilliefors test [52] (D = 0.051253, p-value = 0.01674) and the Anderson-Darling test [53] (A^2 = 1.4911, p-value = 0.000751) reject the null hypothesis of normality at the 5% significance level. The Anscombe-Glynn kurtosis test [54] indicates mesokurtosis (kurt = 2.99577, z = 0.16173, p-value = 0.8715), whereas the D'Agostino asymmetry test [55] reveals right-tail asymmetry (skew = 0.46965, z = 3.65637, p-value = 0.0002558). However, no outliers are detected (Grubbs test [56] for one outlier: G = 2.99489, U = 0.97658, p-value = 0.5018). See Appendix H for the implementation of these tests.

7. Discussion

The initial proposal of the *Q* test for normality, based on the PSNS, defined a sum of squares of seven standardized quantiles without interaction terms [5]. Although it does not assume independence between quantiles, their correlations are ignored to allow the use of the chi-square distribution with seven degrees of freedom as an approximate sampling distribution. This approximation greatly simplifies the calculation of the probability value, Type II error, statistical power, and effect size, as is also done in the K^2 test [55, 57-60]. The test requires a large random sample of a continuous variable.

Based on a simulation study and the approximate calculation of the probability value and statistical power using the chi-square distribution with seven degrees of freedom, Moral [5] suggested a minimum sample size of 150 participants for the *Q* test to achieve high accuracy (proportion of correct decisions in retaining or rejecting the null hypothesis of normality) and high power (for non-normal distributions) or complement of power (for the normal distribution). However, for the normal distribution, the test exhibited an accuracy of 1 in detecting or retaining the null hypothesis of normality with a sample size as small as 20, while maintaining a power below 0.2.

The lower power of the *Q* test in rejecting the null hypothesis for non-normal distributions was particularly evident in symmetric distributions with slight leptokurtosis (logistic) or platykurtosis (uniform, semicircular, and arcsine). This reduced power is explained by the high correlations observed among the four central numbers of the seven-number summary.

The present study does not ignore the correlation between quantiles but addresses it in two ways: by generating the sampling distribution of the *Q* statistic using the bootstrap method (Q_B) or by introducing a new expression of the statistic that incorporates interaction terms between pairs of random variables, weighted by their corresponding correlations. This new approach takes the form of a quadratic expression, where the vector of standardized quantiles remains the same as in the previous proposal, but the symmetric square matrix is replaced by the correlation matrix of the quantiles under the null hypothesis of normality, instead of an identity matrix. Since the standardization of the quantiles and the correlation matrix is based on their expected values under the null hypothesis of normality, the subscript T (theoretical distribution) is added to *Q* (quadratic form) to denote the test statistic.

A sample simulation was used as in the previous study. Thus, simulations were performed for 31 sample sizes across 12 distributions (one normal and 11 non-normal) using a fixed seed, resulting in 372 random samples, each of which was subjected to four normality tests. On the one hand, normality was assessed using three versions of the *Q*-test: one ignoring the correlations between quantiles, one generating the sampling distribution using a bootstrap procedure, and one specifying the quadratic form with the correlation matrix of the quantiles. On the other hand, normality was also tested using the Shapiro-Wilk *W* test [12-13]. The effectiveness (hit rate) and power of the tests were evaluated, and both metrics were compared across the four normality tests using two repeated measures tests: Cochran's *Q* test for proportion of correct classifications and Friedman's test for mean power.

As in the previous study [5], the *W* test exhibited the highest accuracy and statistical power. Its power was significantly higher than those of the three versions of the *Q* test and its hit ratio was significantly higher than those of the approach Q test based on chi-square distribution and generalized chi-square distribution, but not than bootstrap approach. The latter showed significant differences in accuracy (hit ratio) y statistical power than other two versions of Q test.

The Q_B test demonstrated perfect accuracy, or hit ratio, for normal distribution that is a mesokurtic symmetrical distribution, as well as for distributions very distant from normality (exponential, arcsine, and chi-square), except for these last two distributions are results shared by the four normality tests. Its accuracy was very high for platykurtic symmetrical distributions (uniform, arcsine, and semicircular) with an average of 0.957, which was the weakness of the original version of Q test. It also was very high for leptokurtic symmetrical distribution (Laplace, Student, and hyperbolic secant) with an average of 0.935, even for a distribution with slight positive asymmetry ($\sqrt{\beta_1} \approx 0.631$) and slight pointing ($\beta_2 - 3 \approx 0.256$) as is the Rayleigh distribution with a value of 0.968. For the slightly leptokurtic symmetrical distribution included in the study, namely the logistic distribution, its accuracy was high of 0.806. In fact, its accuracy was equivalent to that of the *W* test across all 12 distributions. Meanwhile, the Q_T test showed lower accuracy than the *W* test for the chi-square, uniform, logistic, and Rayleigh distributions, as well as lower accuracy than the original version of the *Q* test for the logistic and Rayleigh distributions. Consequently, the bootstrap approach is the best choice among the three versions of the *Q* test.

Is it worth questioning why the quadratic form is not the best option for the Q test—or, in fact, why it performs the worst? The quadratic form is the most theoretically grounded expression, clearly based on the assumption of normality, which serves as the null hypothesis of the distributional model. Accordingly, it is denoted with the subscript T (for theoretical) and has a specific implementation in R. The issue may lie in the second-degree polynomial involving squared standardized values and their cross-products, weighted by the corresponding correlation coefficients. This arithmetic structure makes the test overly conservative with respect to the null hypothesis, ultimately rendering it inadequate and lacking in statistical power. In contrast, the bootstrap procedure generates the sampling distribution in a way that incorporates the influence of quantile correlations, but without the excessive conservatism of the explicit quadratic form. It also avoids the theoretical limitation of assuming independence, as in the original test form, which proved to be both appropriate and powerful.

While the bootstrap approach is as accurate as the other two versions of the Q test in maintaining the null hypothesis under normally distributed data, it exhibits a lower type II error rate (i.e., higher statistical power). Thus, it represents the least conservative formulation with respect to the null hypothesis of normality in the Q test.

Although the *W* test was a better option than the Q_B test in its bootstrap approach, the latter remains an important alternative when using the seven-number summary to assess normality. Both tests showed similar accuracy (0.952, 95% Wilson-type *CI* [0.925, 0.969] vs. 0.976, 95% Wilson-type *CI* [0.955, 0.987]) and average power (0.935, 95% BCa *CI* [0.919, 0.950] vs. 0.961, 95% BCa *CI* [0.946, 0.975]). Notably, the *W* test does not clearly outperform the *Q* test with samples of 50, and a previous study had already reported this finding with smaller samples of 20 and 30 data points [5]. Similarly, Souza, Toebe, Mello, and Bittencourt [61] found that the Shapiro-Wilk test performed poorly with small samples.

The example with real data demonstrates that the bootstrap version of the Q test leads to the same conclusion as traditional normality tests—such as the Shapiro-Wilk test—namely, the rejection of the null hypothesis. In contrast, the other two versions of the Q test assume normality. Although this assumption is supported by previous studies conducted with psychology students [62], high school students [63], and both clinical [20] and general [64] populations, the current sample consists of students specifically aspiring to enter a psychology program during the selection process. Therefore, the deviation from normality, characterized by positive skewness, may be attributed to impression management (i.e., social desirability). Since alexithymia entails low emotional intelligence, it is not a desirable trait for candidates seeking to become psychologists. Consequently, participants provide lower scores on the alexithymia scale.

Regarding the limitations of the study, it should be noted that the simulation was restricted to a single sample per condition (sample size × distribution type) due to computational processing constraints. However, a seed was used to ensure the reproducibility of results and to achieve greater statistical power in comparisons among the four normality tests when applying repeated measures statistical tests (Cochran's *Q* and Friedman's test). Small sample sizes (20, 30, 40) were not included, as a previous study on the *Q* test indicated the necessity of larger samples, including *W* test [5]. Additionally, only 12 continuous distributions were considered, meaning the behavior of the *Q* test with discrete distributions, such as uniform, binomial, negative binomial, hypergeometric, negative hypergeometric, or Poisson, was not examined [65]. The comparison was limited to the Shapiro-Wilk test [12-13], which is currently regarded as the most powerful test for sample sizes ranging from 3 to 2000 [66]. However, other alternatives exist, such as the Shapiro-Francia test [16-17], which can be used for samples of up to 5,000 data points and has a statistical power similar to that of the Shapiro-Wilk test [15].

8. Conclusions

It is concluded that the bootstrap approach in the *Q* test, based on the parametric seven-number summary (PSNS), represents a substantial improvement over the original version, which relied on the chi-square distribution with seven degrees of freedom and did not account for correlations among the seven quantiles. The Q_B variant significantly improves both the test's accuracy and statistical power. Furthermore, its performance is comparable to that of the Shapiro–Wilk W test, which is currently regarded as the most powerful normality test for samples of up to 2,000 observations.

9. Suggestions

Further investigation of the Q_B test with discrete distributions that converge to normality is recommended, considering both parameterizations that deviate from normality and parameter values that ensure proximity to it. For instance, the binomial distribution B(10, 0.1) can be considered far from normal, warranting rejection of the null hypothesis of normality in this case. In contrast, the binomial distribution B(100, 0.5) has parameter values that closely approximate the normal distribution, supporting the retention of the null hypothesis.

Similarly, in the case of the Poisson distribution, a rate parameter of 3 results in a distribution that deviates significantly from normality, justifying rejection of the null hypothesis. Conversely, a rate parameter of 100 provides an acceptable approximation to normality, reinforcing the retention of the null hypothesis.

It is also suggested to include larger sample sizes (ranging from 1,000 to 10,000 in increments of 500) and to compare the Q_B test with the Shapiro-France normality test [16-17] and the quantile-based test by Avdović and Jevremović [6]. The expectation is that the Q_T test will exhibit similar or slightly lower accuracy and power values. If additional tests such as the Kolmogorov-Smirnov test [67] or the G-test [68] are included, the Q_B test is expected to demonstrate superior performance.

The application of the Q_B test is recommended either as a standalone procedure—accompanied by graphical representations such as bar and box plots, histograms with overlaid density and normal curves, and normal quantile-quantile plots [69-70]—particularly when the seven-number summary is reported [4], or as a complementary test alongside the Shapiro–Wilk test [12-13]. The R scripts developed in this study are available for this purpose.

The Freedman–Diaconis rule is used to determine the bin width and number of bins for the histogram [49] representing the sample data. Epanechnikov's kernel function [50] and the Sheather–Jones bandwidth selector [51] are employed to estimate the density of the overlaid empirical curve. These options were selected due to their flexibility regarding the normality assumption and their closer alignment with empirical data [71]. Additionally, a normal curve is included as a theoretical reference in the plot, which can be saved in Tagged Image File Format (TIFF). This format, developed by Aldus Corporation, is known for its high quality and support for both lossless and lossy compression, making it suitable for various purposes, including archiving and printing [72].

Funding: This research received no external funding.

Acknowledgments: The author expresses his gratitude to the reviewers and editors for the suggestions and corrections received for the improvement of the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

References

[1] A. L. Bowley, Elements of Statistics. P. S. King, London, 1901.

[2] A. L. Bowley, Elementary Manual of Statistics. P. S. King, London, 1910.

[3] P. Muthudoss, S. Kumar, E. Y. C. Ann, K. J. Young, R. L. R. Chi, R. Allada, B. Jayagopal, A. Dubala, I. B. Babla, S. Das, S. Mhetre, I. Saraf, and A. Paudel, "Topologically directed confocal Raman imaging (TD-CRI): Advanced Raman imaging towards compositional and micromeritic profiling of a commercial tablet components," J. Pharm. Biomed. Anal., vol 210, article 114581, February 2022. https://doi.org/10.1016/j.jpba.2022.114581

[4] J. W. Tukey, Exploratory Data Analysis. Addison and Wesley, Reading, MA, 1977.

[5] J. Moral, "Testing for normality from the parametric seven number summary," Open J. Stat., vol. 12, no. 1, pp. 118-154, February 2022. https://doi.org/10.4236/ojs.2022.121009

[6] A. Avdović, and V. Jevremović, "Quantile-zone based approach to normality testing," Mathematics, vol. 10, no. 11, article 1828, May 2022. https://doi.org/10.3390/math10111828

[7] G. Dikta, and M. Scheer, Bootstrap Methods: with Applications in R. Springer Nature, Cham, Switzerland, 2021. https://doi.org/10.1007/978-3-030-73480-0

[8] B. Efron, and B. Narasimhan, "The automatic construction of bootstrap confidence intervals," J Comput Graph Stat., vol. 29, no. 3, pp. 608-619, Mach 2020. https://doi.org/10.1080/10618600.2020.1714633

[9] G. Rousselet, C. R. Pernet, and R. R. Wilcox, "An introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations", Meta-Psychology, vol. 7, article 2058. https://doi.org/10.15626/MP.2019.2058

[10] A. Stuart, and J. K. Ord, Kendall's Advanced Theory of Statistics. Volume 1. Distribution Theory, 6th ed., Edward Arnold, London, 1994.

[11] P. Lafaye de Micheaux, Package CompQuadForm: Distribution function of quadratic forms in normal variables, 2025. https://cran.r-project.org/web/packages/CompQuadForm/CompQuadForm.pdf

[12] S. S. Shapiro, and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol. 52, no. 3-4, pp. 591-611, December 1965. https://doi.org/10.1093/biomet/52.3-4.591

[13] J. P. Royston, "Approximating the Shapiro-Wilk W-test for non-normality," Statistics and Computing, vol. 2, no. 3, pp. 117-119, September 1992. https://doi.org/10.1007/BF01891203

[14] J. Arnastauskaitė, T. Ruzgas, and M. Bražėnas, "An exhaustive power comparison of normality tests," Mathematics, vol. 9, no. 7, article 788, April 2021. https://doi.org/10.3390/math9070788

[15] N. Khatun, "Applications of normality test in statistical analysis," Open J. Stat., vol. 11, no. 1, pp. 113-122, January 2021. https://doi.org/10.4236/ojs.2021.111006

[16] S. S. Shapiro, and R. S. Francia, "An approximate analysis of variance test for normality," J. Am. Stat. Assoc., vol. 67, no 337, pp. 215-216, June 1972. https://doi.org/10.1080/01621459.1972.10481232

[17] J. P. Royston, "A toolkit of testing for non-normality in complete and censored samples," J. Roy. Stat. Soc. Ser. D-The Statistician, vol. 42, no. 1, pp. 37-43, March 1993. https://doi.org/10.2307/2348109

[18] R. M. Bagby, J. D. Parker, and G. J. Taylor, "The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure," J. Psychosom. Res., vol. 38, no. 1, pp. 23-32, Janaury 1994. https://doi.org/10.1016/0022-3999(94)90005-1

[19] J. Moral, "Propiedades psicométricas de la Escala de Alexitimia de Toronto de 20 reactivos en México," Revista Electrónica de Psicología Clínica de Iztacala, vol. 11, no. 2, pp. 97-114, July 2008. Available at: https://www.iztacala.unam.mx/carreras/psicologia/psiclin/

[20] G. Pedersen, E. Normann‐Eide, I. U. M. Eikenæs, E. H. Kvarstein, and T. Wilberg, "Psychometric evaluation of the Norwegian Toronto Alexithymia Scale (TAS‐20) in a multisite clinical sample of patients with personality disorders and personality problems," J. Clin. Psychol., June 2022, vol. 78, no. 6, pp. 1118-1136, 2022. https://doi.org/10.1002/jclp.23270

[21] H. Zhang, J. Shen, and Z. Wu, "A fast and accurate approximation to the distributions of quadratic forms of Gaussian variables," J Comput. Graph. Stat., vol. 31, no. 1, pp. 304-311, January 2022. https://doi.org/10.1080/10618600.2021.2000423

[22] M. Estaki, L. Jiang, N. A. Bokulich, D. McDonald, A. González, T. Kosciolek, C. Martino, Q. Zhu, A. Birmingham, Y. Vázquez-Baeza, M. R. Dillon, E. Bolyen, J. Caporaso, G., and R. Knight, "QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data," Curr. Protoc. Bioinformatics, vol. 70, article e100, April 2020. https://doi.org/10.1002/cpbi.100

[23] J. M. O. Machado, Outlier detection in accounting, master thesis, University of Porto. Institutional repository of the University of Porto, 2018. https://sigarra.up.pt/fep/en/pub_geral.show_file?pi_doc_id=423163

[24] S. Cai, J. Zhou, and J. Pan, "Estimating the sample mean and standard deviation from order statistics and sample size in meta-analysis," Stat. Methods Med. Res., 30, vol. 12, pp. 2701-2719, October 2021. https://doi.org/10.1177/09622802211047348

[25] R. J. Hyndman, and Y. Fan, "Sample quantiles in statistical packages," Am. Stat., vol. 50, no. 4, pp. 361-365, November 1996. https://doi.org/10.1080/00031305.1996.10473566

[26] L. Makkonen, and M. Pajari, "Research article defining sample quantiles by the true rank probability," J. Probab. Stat., vol. 2014, article 326579, December 2014. http://dx.doi.org/10.1155/2014/326579

[27] D. M. Hawkins, "Quantile-quantile methodology-detailed results," arXiv, article 2303.03215, October 2023. https://doi.org/10.48550/arXiv.2303.03215

[28] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," Math. Proc. Camb. Philos. Soc., vol. 30, no. 2, pp. 178-191, April 1934. https://doi.org/10.1017/S0305004100016595

[29] P. McCullagh, "Gaussian distributions," in Ten Projects in Applied Statistics, Springer Series in Statistics. Springer, Cham, pp. 251-277. https://doi.org/10.1007/978-3-031-14275-8_15

[30] G. Casella, and R. Berger, Statistical Inference. CRC Press, Hoboken, NJ, 2024. https://doi.org/10.1201/9781003456285

[31] W. Y. Chen, G. W. Peters, R. H. Gerlach, and S. A. Sisson, "Dynamic quantile function models," Quantitative Finance, vol. 22, no. 9, pp. 1665-1691, April 2022. https://doi.org/10.1080/14697688.2022.2053193

[32] A. Canty, B. Ripley, and A. R. Brazzale, Package ‘boot’. Bootstrap functions, 2024 https://cran.r-project.org/web/packages/boot/boot.pdf

[33] J. Cohen, "A power primer," in A. E. Kazdin, Ed., Methodological Issues and Strategies in Clinical Research, 4th ed. American Psychological Association, Washington, DC, 2016, pp. 279-284. https://doi.org/10.1037/14805-018

[34] R. B. Davies, "Algorithm AS 155: The distribution of a linear combination of chi-2 random variables." J. R. Stat. Soc. Ser. C Appl., vol. 29, no. 3, pp. 323-333, December 1980. https://doi.org/10.2307/2346911

[35] J. Cohen, Statistical Power Analysis for Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

[36] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification," Ann. Math. Stat., vol. 25, no. 3, pp. 484-498, September 1954. https://doi.org/10.1214/aoms/1177728717

[37] A. Das, "New methods to compute the generalized chi-square distribution," arXiv, article 2404.05062, February 2024. https://doi.org/10.48550/arXiv.2404.05062

[38] H. Cramer, Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ, 1946.

[39] S. J. Peterson, and S. Foley, "Clinician's guide to understanding effect size, alpha level, power, and sample size," Nutr. Clin. Pract., vol. 36, no. 3, pp. 598-605, June 2021. https://doi.org/10.1002/ncp.10674

[40] C. Baumgarten, and T. Patel, "Automatic random variate generation in Python," Proceedings of the 21st Python in Science Conference, July 11, 2022., pp. 46-51. https://doi.org/10.25080/majora-212e5952-007

[41] R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

[42] N. Smeeton, N. Spencer, and P. Sprent, Applied Nonparametric Statistical Methods. CRC Press, Boca Raton, FL, 2025. https://doi.org/10.1201/9780429326172

[43] C. F. Fey, T. Hu, and A. Delios, "The measurement and communication of effect sizes in management research," Manag. Organ. Rev., vol. 19, no. 1, pp. 176-197, February 2023. https://doi.org/10.1017/mor.2022.2

[44] L. A. O. Orawo, "Confidence intervals for the binomial proportion: a comparison of four methods," Open J. Stat., vol. 11, no. 5, pp. 806-816, October 2021. https://doi.org/10.4236/ojs.2021.115047

[45] H. Pham, Springer Handbook of Engineering Statistics, 2023. https://doi.org/10.1007/978-1-4471-7503-2

[46] M. J. Blanca-Mena, R. Alarcón, J. Arnau, J. García-Castro and R. Bono, "How to proceed when normality and sphericity are violated in the repeated measures ANOVA," Ann. Psychol., vol. 40, no. 3, pp. 466-480, July 2024. https://doi.org/10.6018/analesps.594291

[47] Microsoft Corporation. Microsoft Excel (version 2021) [Software], Microsoft Corp., 2021.

[48] IBM Corporation, IBM SPSS Statistics (version 27) [Software], IBM Corp., 2020.

[49] D. Freedman, and P. Diaconis, "On the histogram as a density estimator: L2 theory," Probab. Theory Relat. Fields, vol. 57, no. 4, pp. 453-476, December 1981. https://doi.org/10.1007/BF01025868

[50] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," Theory Probab. Appl., vol. 14, no. 1, pp. 153-158, January 1969. https://doi.org/10.1137/1114019

[51] S. J. Sheather, and Jones, M. C. "A reliable data-based bandwidth selection method for kernel density estimation," J. R. Stat. Soc. Ser. B Stat. Method., vol. 53, no. 3, pp. 683-690, December 1991. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x

[52] H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," J. Am. Stat. Assoc., vol. 62, no. 318, pp. 399-402, June 1967. https://doi.org/10.1080/01621459.1967.10482916

[53] T. W. Anderson, and D. A. Darling, "Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes," Ann. Math. Stat., vol. 23, no. 2, pp. 193-212, June, 1952. https://doi.org/10.1214/aoms/1177729437

[54] F. J. Anscombe and W. J. Glynn, "Distribution of the kurtosis statistic b2 for normal samples," Biometrika, vol. 70, no. 1, pp. 227-234, April 1983 https://doi.org/10.1093/biomet/70.1.227R.

[55] R. B. D’Agostino, "Transformation to normality of the null distribution of g1," Biometrika, vol. 57, no. 3, pp. 679-681, December 1970. https://doi.org/10.1093/biomet/57.3.679

[56] F. E. Grubbs, "Sample criteria for testing outlying observations," Ann. Math. Statist. Vol. 21, no. 1, pp. 27-58, March, 1950. https://doi.org/10.1214/aoms/1177729885

[57] D’Agostino, and E. S. "Pearson, Tests for departure from normality: empirical results for the distributions of b2 and √b1," Biometrika, vol. 60, no. 3, pp. 613-622, 01 December 1973. https://doi.org/10.1093/biomet/60.3.613

[58] R. B. D’Agostino, A. Belanger, and R. B. Jr. D’Agostino, "A suggestion for using powerful and informative tests of normality," Am. Stat., vol. 44, no. 4, pp. 316-321, December 1990. https://doi.org/10.1080/00031305.1990.10475751

[59] C. M. Jarque, and A. Bera, "A test for normality of observations and regression residuals," Int. Stat. Rev., vol. 55, no. 2, pp. 163-172, August 1987. https://doi.org/10.2307/1403192

[60] C. Urzua, "On the correct use of omnibus tests for normality," Econ. Lett., vol. 53, no. 3, pp. 247-251, December 1996. https://doi.org/10.1016/S0165-1765(96)00923-8

[61] R. R. De Souza, M. Toebe, A. C. Mello, and K. C. Bittencourt, "Sample size and Shapiro-Wilk test: An analysis for soybean grain yield," Eur. J. Agron., vol. 142, no. 1, article 126666, January 2023. https://doi.org/10.1016/j.eja.2022.126666

[62] Y. Şener, and Y. Günaydın, "Attitudes Toward Dating Violence, Social Impact, and Alexithymia in University Students: A Structural Equation Modeling," Psychol Reports, 0(0), July, 2024. https://doi.org/10.1177/00332941241265618

[63] M. A. Terzioğlu, and A. Büber, "Alexithymia, internet addiction, and cyber-victimisation among high school students in Turkey: an exploratory study," Behav. Inform. Technol., vol. 44, no. 7, pp. 1350-1361, May 2024. https://doi.org/10.1080/0144929X.2024.2353273

[64] M. Miniati, C. Poidomani, C. Conversano, G. Orrù, R. Ciacchini, D. Marazziti, G. Perugi, A. Gemignani, and L. Palagini, "Alexithymia and interoceptive confusion in Covid-19 pandemic distress," Clinical Neuropsychiatry: Journal of Treatment Evaluation, vol. 20, no. 4, pp. 264-270, January 2023. https://doi.org/10.36131/cnfioritieditore20230405

[65] J. L. Devore, K. N. Berk, and M. A. Carlton, Modern Mathematical Statistics with Applications. Springer Nature, Berlin, 2021. https://doi.org/10.1007/978-3-030-55156-8

[66] S. Demir, "Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients," IJATE, vol. 9, no. 2, pp. 397-409, June 2022. https://doi.org/10.21449/ijate.1101295

[67] J. Vrbik, "Deriving CDF of Kolmogorov-Smirnov test statistic," Appl. Math., vol. 11, pp. 227-246, March 2020. https://doi.org/10.4236/am.2020.113018

[68] S. S. Mangiafico, "G-test of goodness-of-fit," in An R Companion for the Handbook of Biological Statistics, Version 1.3.9, 2023. https://rcompanion.org/rcompanion/b_04.html

[69] H. Wickham, W. Chang, L. Henry, K. Takahashi, C. Wilke, K. Woo, H. Yutani, D. Dunnington, T. van den Brand, and P. L. Pedersen, ggplot2: Create elegant data visualisations using the grammar of graphics, 2024. https://doi.org/10.32614/CRAN.package.ggplot2

[70] S. Phuyal, M. Rashid, and J. Sarkar, GIplot: An R package for visualizing the summary statistics of a quantitative variable, 2021. https://cran.r-project.org/web/packages/GIplot/index.html

[71] L. Gonzales-Fuentes, L. Barbé, K., Barford, L. Lauwers, and L. Philips, "A qualitative study of probability density visualization techniques in measurements," Measurement, vol. 65, no. 1, pp. 94-111, April 2015. https://doi.org/10.1016/j.measurement.2014.12.022

[72] A. J. Qasim, and F. Q. A. Alyousuf, "History of image digital formats using in information technology," Qalaai Zanist Scientific Journal, vol. 6, no. 2, pp. 1098-1112, June 2021. https://doi.org/10.25212/lfu.qzj.6.2.41

Appendices

Appendix A. Q-statistic

```
# Shared script for the QB and Q tests.
# It serves as the introductory section of the scripts found in Appendices C and E.

# Define data vector.
x <- c()

# Sample representation using a histogram (based on the Freedman-Diaconis rule) with density curves
(Epanechnikov’s kernel and Sheather-Jones bandwidth) and normal curves overlaid.
# Remove the hash symbol from tiff() and dev.off to save the graphic as a TIFF file.
tiff("histogram.tiff", width = 1600, height = 900, units = "px", res = 300)
par(mar = c(4, 4, 1, 1) + 0.1) # Set the plot margins.

# Define a wider range for x to ensure complete visualization
x_range <- range(x)
x_buffer <- 0.2 * diff(x_range)
xlim_adjusted <- c(x_range[1] - x_buffer, x_range[2] + x_buffer)
# Compute density for ylim
density_x <- density(x, kernel = "epanechnikov", bw = "SJ")
y_normal <- dnorm(x, mean = mean(x), sd = sd(x))
# Histogram based on the Freedman-Diaconis rule.
hist(x, breaks = "FD", freq = FALSE, col = "lightyellow", border = "black",
```

```
main = "", xlab = "Scores of X", xlim = xlim_adjusted, ylab = "Density", ylim = c(0, max(density_x$y,
y_normal) + 0.01))
# Overlay a density curve (Epanechnikov’s kernel and Sheather-Jones bandwidth)
lines(density_x, col = "darkblue", lwd = 4)
# Overlay the expected density curve if data follow a normal distribution.
x_seq <- seq(xlim_adjusted[1], xlim_adjusted[2], length.out = 1000)
y_normal <- dnorm(x_seq, mean = mean(x), sd = sd(x))
lines(x_seq, y_normal, col = "red", lwd = 4)
# Add legend.
legend("topleft", legend = c("Empirical", "Normal"), col = c("darkblue", "red"),
lwd = 4, bty = "n", title = "", cex = 0.8)
dev.off()

# Calculation of sample quantiles at the orders corresponding to the parametric seven-number summary
for a normal distribution, using the R type-8 rule.
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
q <- quantile(x, probs = p, type = 8)
cat("\nSample quantiles corresponding to Parametric Seven-Number Summary (PSNS)\n")
print(round(q, 4))

cat("\nStatistics needed to standardize sample quantiles corresponding to the PSNS:\n")
n <- length(x) # sample size
m <- mean(x) # sample mean
s <- sd(x) # sample standard deviation
cat("\nSample size:", n, ".\n")
cat("\nSample mean:", m, ".\n")
cat("\nSample standard deviation:", s, ".\n")

# Computation of the statistical value for the normality test, derived from the parametric seven-number
summary.
e_zp <- qnorm(p) # Expected value under normal.
d_zp <- dnorm(qnorm(p)) # Expected density under normal.
ee_zp <- sqrt((p * (1 - p)) / (n * dnorm(qnorm(p))^2)) # Standard error under normal.
zx <- (q - m) / s # Standardized quantile.
z <- (zx - e_zp) / ee_zp # Standardization of zx under assumption of normality.
z_sq <- z^2 # squared of the z_q statistic.
Q_stat <- sum(z_sq) # test statistic for the normality Q test.

# Creating a table from a data frame with the previous calculations.
tabla <- data.frame(e_zp = round(e_zp, 4), p = round(p, 4), d_zp = round(d_zp, 4), ee_zp = round(ee_zp,
4), x_p = round(q, 4), z_xp = round(zx, 4), z = round(z, 4), z_sq = round(z_sq, 4))
fila_suma <- data.frame(e_zp = "Sum", p = "", d_zp = "", ee_zp = "", x_p = "", z_xp = "",
z = "", z_sq = format(Q_stat, digits = 6))
tabla <- rbind(tabla, fila_suma)
# Displaying the table.
cat("\nTable: Testing normality using the test based on the Parametric Seven-Number Summary\n")
print(tabla, row.names = FALSE)
cat("\nNote: e_zp = expected value for the p-th quantile in a N(0, 1) distribution,
p = quantile order,
d_zp = density of the p-th quantile in a N(0, 1) distribution,
ee_zp = standard error of the p-th quantile for n normally distributed data points,
```

```
x_p = p-th sample quantile,
z_xp = standardized quantile (calculated using the sample mean and standard deviation),
z = (z_xp - e_zp) / ee_zp = standardization of the z_xp statistic under the assumption of normality,
z^2 = squared z_xp statistic.\n")

# Display value of test statistic in original sample.
cat("\nValue of test statistic in original sample: Q-statistic =", Q_stat, ".\n")
```

Appendix B. Bootstrap version of PSNS Q test

```
# Script for the Q_B test: Bootstrap probability and effect size.
# This script is designed to be used in conjunction with the script from Appendix B, but it can also be executed independently.

# Function to compute the Q statistic from a sample vector using R’s type-8 quantiles.
compute_q_statistic <- function(sample, type = 8) {
# Define the seven standard normal quantiles used in the PSNS:
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
# Obtain the sample quantiles at those probabilities
q <- quantile(sample, probs = p, type = type)
# Standardize the sample quantiles using the mean and standard deviation of the bootstrap sample.
zx <- (q - mean(sample)) / sd(sample)
# Compute the squared deviations between the standardized sample quantiles and the corresponding theoretical quantiles, scaled by its standard error (under null hypothesis of normality)
z_sq <- ((zx - qnorm(p)) / sqrt((p * (1 - p)) / (length(sample) * dnorm(qnorm(p))^2)))^2
sum(z_sq)
}

# Value of test statistic in original sample.
Q_stat <- compute_q_statistic(x)

# Bootstrap sampling distribution for the Q statistic in the normality test based on the Parametric Seven-Number Summary (PSNS).
set.seed(123) # Seed for reproducibility of results.
Q_bootstrap <- numeric(2000)
for (i in 1:2000) {
x_boot <- sample(x, length(x), replace = TRUE)
Q_bootstrap[i] <- compute_q_statistic(x_boot)
}
# print(round(Q_bootstrap, 4)) # Remove the hash symbol if you want to display the empirical bootstrap distribution.

# Bootstrap critical value for the null hypothesis of normality using R's type-9 quantiles in the function that computes the Q statistic.
set.seed(123) # Seed for reproducibility of results,
Q_null <- numeric(2000)
for (i in 1:2000) {
x_norm_boot <- rnorm(length(x), mean = mean(x), sd = sd(x))
Q_null[i] <- compute_q_statistic(x_norm_boot, type = 9)
}
# print(round(Q_null, 4)) # Remove the hash symbol if you want to display the normative bootstrap distribution.
```

```
# Critical value and bootstrap p-value
alpha <- 0.05
Q_critical <- quantile(Q_null, 1 - alpha, type = 8)

# Print results.
cat("\nValue of test statistic in original sample", Q_stat, ".\n")
cat("\nBootstrap sampling distribution of the Q statistic generated from the original sample\n")
cat("\nNumber of extractions with replacement =", length(Q_bootstrap), ".\n")
cat("\nBootstrap estimation of the Q =", mean(Q_bootstrap), ".\n")
cat("\nBootstrap standard error of the Q =", sd(Q_bootstrap), ".\n")
cat("\nBootstrap bias of the Q =", mean(Q_bootstrap) - Q_stat, ".\n")
cat("\nNormative bootstrap sampling distribution of the Q statistic\n")
cat("\nThis is generated from a normal distribution with mean", mean(x), "and standard deviation",
sd(x), "\n")
cat("\nBased on", length(Q_null), "extractions of size", length(x), ".\n")
cat("\nBootstrap expected value of Q under normal =", mean(Q_null), ".\n")
cat("\nBootstrap standard error of Q under normal =", sd(Q_null), ".\n")
p_bootstrap <- mean(Q_null > Q_stat)
cat("\nBootstrap critical value for the Q statistic with a significance level of", alpha, "=", Q_critical,
".\n")
cat("\nBootstrap p-value =", format(p_bootstrap, scientific = FALSE, digits = 5), ".\n")
if (p_bootstrap < alpha) {
cat("\nThe null hypothesis that the data follow a normal distribution is rejected \nwith a significance
level of", alpha, "based on the bootstrap p-value.\n")
} else {
cat("\nThe null hypothesis that the data follow a normal distribution is not rejected \nwith a
significance level of", alpha, "based on the bootstrap p-value.\n")
}

# Effect size.
Z_Q <- abs(Q_stat - mean(Q_null)) / sd(Q_null)
cat("\nEffect size as standardized distance: Z_Q =", Z_Q, ".\n")
cat("\nInterpretation of effect size\n")
cat("< 1.7: Trivial\n[1.7, 2.5): Small\n[2.5, 3): Medium\n ≥ 3: Large\n")
effect_size <- ifelse(Z_Q < 1.7, "trivial",
ifelse(Z_Q < 2.5, "small",
ifelse(Z_Q < 3, "medium", "large")))
cat("\nBased on the Z statistic = ", round(Z_Q, 4), ", the size effect of the deviation from the normal
distribution model is classified as", effect_size, ".\n")

# Histogram of the bootstrap sampling distribution of the Q statistic from the sample data,
# with an overlaid kernel density estimate and a chi-square density curve with seven degrees of
freedom.
# Remove the hash symbol from tiff() and dev.off to save the graphic as a Tagged Image File Format
(TIFF) file.
# tiff("Empiric_hist.tiff", width = 1600, height = 900, units = "px", res = 300)
par(mar = c(4, 4, 1, 1) + 0.1) # Set the plot margins
# Define x-axis limits
x_min <- min(Q_bootstrap)
x_max <- max(Q_bootstrap)
```

```
# Compute kernel density and chi-square density
dens_kernel <- density(Q_bootstrap)
dens_chisq <- dchisq(dens_kernel$x, df = 7)
# Determine max y value for ylim
y_max <- max(dens_kernel$y, dens_chisq)
# Plot histogram with adjusted ylim
hist(Q_bootstrap, breaks = "fd", freq = FALSE, col = "lightyellow", border = "black",
main = "Bootstrap distribution", xlab = "Q statistic (sample data)",
ylab = "Density", xlim = c(x_min, x_max), ylim = c(0, y_max * 1.05))
# Overlay kernel density estimate (Gaussian kernel)
lines(dens_kernel, col = "darkblue", lwd = 4)
# Overlay chi-square density curve with 7 degrees of freedom
curve(dchisq(x, df = 7), from = 0, to = x_max, col = "red", lwd = 4, add = TRUE)
# Add vertical line for Q_critical
abline(v = Q_critical, col = "purple", lwd = 3, lty = 2)
# Add vertical line for Q_statistic
abline(v = Q_stat, col = "black", lwd = 3, lty = 2)
# dev.off()

# Histogram of the bootstrap sampling distribution of the Q statistic under normality,
# with an overlaid kernel density estimate and a chi-square density curve with seven degrees of
freedom.
# Remove the hash symbol from tiff() and dev.off to save the graphic as a TIFF file.
# tiff("Normative_hist.tiff", width = 1600, height = 900, units = "px", res = 300)
par(mar = c(4, 4, 1, 1) + 0.1) # Set the plot margins.
# Density histogram using Freedman-Diaconis rule.
# Check if the upper limit of 16 for the ordinate axis is appropriate.
hist(Q_null, breaks = "fd", freq = FALSE, col = "lightyellow", border = "black",
main = " Bootstrap distribution under normality", xlab = "Q Statistic (Normal samples)", ylab =
"Density", xlim = c(0, 16))
# Overlay kernel density estimate (Gaussian kernel)
lines(density(Q_null), col = "darkblue", lwd = 4)
# Overlay chi-square density curve with 7 degrees of freedom
# Check if the upper limit of 16 for the curve is appropriate.
curve(dchisq(x, df = 7), from = 0, to = 16, col = "red", lwd = 4, add = TRUE)
# Add vertical line for Q_critical
abline(v = Q_critical, col = "purple", lwd = 3, lty = 2)
# Add vertical line for Q_statistic
abline(v = Q_stat, col = "black", lwd = 3, lty = 2)
# dev.off()
```

Appendix C. Statistical power of bootstrap version of PSNS Q test

```
# Script for computing the bootstrap power of the normality test using the QB statistic.

x <- c() # Define data vector.
Q_critical <- 6.43475 # Bootstrap critical value. It is taken from the result of the previous script.
alpha <- 0.05 # Level of significance or complement of the critical value order (Q_critical).

# Bootstrap power calculation.
set.seed(456) # Seed for reproducibility of results.
# Create a vector to store results of 10000 simulated tests
```

```
for (i in 1:10000) {
x_boot <- sample(x, length(x), replace = TRUE)
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
q <- quantile(x_boot, probs = p, type = 8)
zx <- (q - mean(x_boot)) / sd(x_boot)
z_sq <- ((zx - qnorm(p)) / sqrt((p * (1 - p)) / (length(x_boot) * dnorm(qnorm(p))^2)))^2
Q_sim <- sum(z_sq)
}
# Compute empirical power as the proportion of rejections across simulations.
power <- mean(Q_sim > Q_critical)

# Display results
cat("\nNumber of bootstrap samples =", length(Q_sim), ".\n")
cat("\nNumber of data per bootstrap sample =", length(x_boot), ".\n")
cat("\nBootstrap power at a significance level of", alpha, "=", format(power, scientific = FALSE, digits = 5), ".\n")
```

Appendix D. Quadratic form version of PSNS Q test

```
# Script for calculating the Q_T statistic (quadratic form), its probability value, statistical power, and effect size.

# Replace the example with your own data vector.
x <- c(-0.23, -1.39, 0.38, 0.52, -0.49, 0.28, -0.04, 0.11, 1.03, -0.33, -0.33, 0.06, 0.16, 0.29, -0.16, -1.06, 0.54, 0.88, -1.64, -0.31)

# Gaussian vector z.
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
q <- quantile(x, probs = p, type = 8) # Parametric seven-number summary
n <- length(x) # Sample size
m <- mean(x) # Sample Mean
s <- sd(x) # Standard Sample Deviation
e_zp <- qnorm(p) # Expected value under normal
d_zp <- dnorm(qnorm(p)) # Expected density under normal
ee_zp <- sqrt((p * (1 - p)) / (n * dnorm(qnorm(p))^2)) # Standard error under normal
zx <- (q - m) / s # Standardized quantile
z <- (zx - e_zp) / ee_zp # Standardization of zx statistic under assumption of normality
cat("\nGaussian vector z:\n")
cat(z, "\n")

# Matrix of correlations R under the assumption of normality.
p <- pnorm(c(-2, -4/3, -2/3, 0, 2/3, 4/3, 2)) # Cumulative probabilities corresponding to PSNS z-scores
d <- dnorm(c(-2, -4/3, -2/3, 0, 2/3, 4/3, 2)) # Density values corresponding to PSNS z-scores
n <- 1000 # Number of simulations
# Initialize a square matrix of zeros to store the theoretical covariance values.
cov_matrix <- matrix(0, nrow = length(p), ncol = length(p))
# Compute the diagonal elements (variances) of the covariance matrix.
# These are based on the variance of the sample quantiles under normality.
for (i in 1:length(p)) {
cov_matrix[i, i] <- (p[i] * (1 - p[i])) / (n * d[i]^2)}
# Compute the off-diagonal elements (covariances) of the covariance matrix.
# These reflect the expected covariances between different quantiles under normality.
```

```
for (i in 1:(length(p) - 1)) {
for (j in (i + 1):length(p)) {
cov_matrix[i, j] <- (p[i] * (1 - p[j])) / (n * d[i] * d[j])
cov_matrix[j, i] <- cov_matrix[i, j]}}
# Convert the covariance matrix to a correlation matrix.
# This is done by standardizing each covariance by the square roots of the corresponding variances.
corr_matrix <- cov_matrix / sqrt(outer(diag(cov_matrix), diag(cov_matrix)))
cat("\nCorrelation matrix under the assumption of normality\n")
print(round(corr_matrix, 4))

# Quadratic form.
Q_T <- t(z) %*% corr_matrix %*% z
cat("\nThe Q statistic value in the quadratic form =", Q_T, ".\n")

# Probability value with the generalized chi-square distribution (Davies method).
library(CompQuadForm)
eigenvalues <- eigen(corr_matrix, symmetric = TRUE, only.values = TRUE)$values
p_value <- davies(Q_T, lambda = eigenvalues)$Qq
cat("\nEigenvalues:", eigenvalues, ".\n")
cat("\nNumber of eigenvalues:", length(eigenvalues), "\n")
cat("\nSum of eigenvalues:", sum(eigenvalues), ".\n")
cat("\np-value for the Q_T statistic =", p_value, ".\n")
# Exact calculation of power using the generalized chi-square distribution.
alpha <- 0.05
Q_alpha <- qchisq(1 - alpha, df = length(eigenvalues)) # Initial approach
delta <- eigenvalues * z^2
power <- davies(Q_alpha, lambda = eigenvalues, delta = delta)$Qq
cat("\nStatistical power:", power, ".\n")

# Effect size.
ev_Q <- sum(diag(corr_matrix)) # Expected value of the Q_T statistic.
sd_Q <- sqrt(2 * sum(eigenvalues ^2)) # Standard deviation of the Q_T statistic.
Z_Q <- abs(Q_T - ev_Q) / sd_Q # Standarized value of the Q_T statistic.
cat("\nExpected value of the Q_T statistic =", ev_Q, ".\n")
cat("\nStandard deviation of the Q_T statistic =", sd_Q, ".\n")
cat("\nEffect size as standardized distance: Z_Q =", Z_Q, ".\n")
cat("\nInterpretation of effect size.\n")
cat("< 1.7: Trivial\n[1.7, 2.5): Small\n[2.5, 3): Medium\n ≥ 3: Large\n")
effect_size <- ifelse(Z_Q < 1.7, "trivial",
ifelse(Z_Q < 2.5, "small",
ifelse(Z_Q < 3, "medium", "large")))
cat("\nBased on Z statistic =", round(Z_Q, 4), ", the effect size of the deviation from the normal
distribution model is classified as", effect_size, ".\n")
```

Appendix E. Chi-square approximation to PSNS Q test

```
# Script for calculating the probability value, statistical power, and effect size based on the chi-square
approximation with 7 degrees of freedom.
# To be used following script in Appendix B.

cat("\nCalculation of probability value, statistical power, and effect size using the chi-square
approximation with 7 degrees of freedom.\n")
```

```
# Calculation of the probability value from the chi-square approximation with 7 degrees of freedom.
alpha <- 0.05
p_asint <- pchisq(Q_stat, df = 7, lower.tail = FALSE)
cat("\nTest Statistic: Q =", Q_stat, ",", "critical value with a significance level of", alpha,"=",
qchisq(alpha, df = 7, lower.tail = FALSE), ",", "asymptotic p-value =", p_asint, ".\n")
if (p_asint < alpha) {
cat("\nThe null hypothesis that the data follow a normal distribution is rejected \nwith a significance
level of", alpha, "by the Q test.\n")
} else {cat("\nThe null hypothesis that the data follow a normal distribution is not rejected \nwith a
significance level of", alpha, "by the Q test.\n")}

# Statistical power of the Q test using the chi-square distribution approximation with 7 degrees of
freedom.
power <- 1 - pchisq(qchisq(alpha, df = 7, lower.tail = FALSE), df = 7, ncp = Q_stat, lower.tail = FALSE,
log.p = FALSE)
cat("\nThe statistical power to the right tail for the alternative hypothesis of non-normality for the Q
test: ϕ =", round(power, 4), ".\n")
cat("\nIf the null hypothesis of normality is maintained, the power must be less than 0.5; \nwhile, if
rejected, it must be greater than 0.5. In the latter case, it is classified as \ngood with a value of 0.8 and
very good with a value of 0.9.\n")

# Effect size calculation using the chi-square distribution approximation with 7 degrees of freedom.
v <- sqrt(Q_stat / (n * 7))
cat("\nEffect size via Cramer's V-coefficient: V =", round(v, 4), ".\n")

# Interpretation based on Cohen (1988) [29].
cat("\nInterpretation of effect size based on Cohen (1988):\n")
cat("< 0.1: Trivial\n[0.1, 0.3): Small\n[0.3, 0.5): Medium\n≥ 0.5: Large\n")
effect_size <- ifelse(v < 0.1, "trivial",
ifelse(v < 0.3, "small",
ifelse(v < 0.5, "medium", "large")))
cat("\nBased on the V statistic =", round(v, 4), ", the size of the effect of the deviation from the normal
distribution model is classified as", effect_size, ".\n")
```

Appendix F. Correlation matrix

```
# Population correlation matrix of PSNS quantiles in a normal distribution
set.seed(123) # Seed for reproducibility of results.
n <- 1000 # Original sample size and each bootstrap sample.
B <- 10000 # Number of bootstrap samples.
p <- pnorm(c(-2, -4/3, -2/3, 0, 2/3, 4/3, 2)) # Quantile orders.
# Function for extracting sample quantiles by R type-9 rule.
get_quantiles <- function(sample) {quantile(sample, probs = p, type = 9)}
# Generating bootstrap distributions.
original_sample <- rnorm(n, 0, 1)
bootstrap_samples <- replicate(B, sample(original_sample, n, replace = TRUE))
# Calculation of quantiles for each bootstrap sample.
quantiles_matrix <- apply(bootstrap_samples, 2, get_quantiles)
# Calculation and printing of the correlation matrix.
correlation_matrix <- cor(t(quantiles_matrix))
print(correlation_matrix)
```

Appendix G. Generation of original samples

```
# Generation of original samples of different sizes, drawn from various distributions

# Required libraries.
library(VGAM)
library(EnvStats)
library(extraDistr)

# Define an empty dataframe to store the results.
results <- data.frame(Sample_Size = integer(), Min = numeric(), Gap1 = character(), Mdn = numeric(),
Gap2 = character(), Max = numeric())

# Define sample sizes.
n_values <- c(seq(50, 1500, by = 50), 2000) # De 50 a 1500 en pasos de 50, más 2000

# Set a seed for reproducibility.
set.seed(123)
# set.sed (12) # For normal distribution
# Iterate over sample sizes
for (sample_size in n_values) {

# Generate data from a distribution.
# To select a distribution, remove the hash symbol (#) from the corresponding sample generator. In this
run, the hash was removed from the generator for the Laplace distribution.
x <- rlaplace(sample_size, mu = 50, sigma = 10) # Laplace’s distribution.
# x <- rnorm(sample_size, mean = 50, sd = 10) # Normal distribution.
# x <- rt(sample_size, df = 4) * 10 + 50 # Student's t-distribution with 4 degrees of freedom.
# x <- rchisq(sample_size, df = 5) # Chi-square distribution with 4 degrees of freedom.
# x <- rbeta(sample_size, shape1 = 0.5, shape2 = 0.5) # Arcsine distribution.
# x <- runif(sample_size, min = 0, max = 100) # Continuous uniform distribution.
# x <- rtri(sample_size, min = 0, max = 3, mode = 2.99) # Triangular distribution.
# x <- 3 * (2 * rbeta(sample_size, shape1 = 1.5, shape2 = 1.5) - 1) # Wigner’s semicircle distribution.
# x <- rexp(sample_size, rate = 2) # Exponential distribution.
# x <- 50 + 10 * 2/pi * log(tan(pi/2 * runif(sample_size, min = 0, max = 1))) # Hyperbolic secant distribution.
# x <- rlogis(sample_size, location = 50, scale = 10) # Logistic distribution.
# x <- rrayleigh(sample_size, sigma = 12) # Rayleigh’s distribution.

# Calculate the minimum, median and maximum.
results <- rbind(results, data.frame(Sample_Size = sample_size, Min = min(x), Gap1 = "...", Mdn =
median(x), Gap2 = "...", Max = max(x)))
}
# Print the final table.
# Remove the hash symbol from the corresponding distribution.
cat("\nOriginal sample drawn from a Laplace's distribution\n")
# cat("\nOriginal sample drawn from a normal distribution\n")
# cat("\nOriginal sample drawn from a Student's t-distribution with 4 degrees of freedom\n")
# cat("\nOriginal sample drawn from a chi-square distribution with 4 degrees of freedom\n")
# cat("\nOriginal sample drawn from an arcsine distribution\n")
# cat("\nOriginal sample drawn from a continuous uniform distribution\n")
# cat("\nOriginal sample drawn from a triangular distribution\n")
```

```
# cat("\nOriginal sample drawn from a Wigner's semicircle distribution\n")
# cat("\nOriginal sample drawn from an exponential distribution\n")
# cat("\nOriginal sample drawn from a hyperbolic secant distribution\n")
# cat("\nOriginal sample drawn from a logistic distribution\n")
# cat("\nOriginal sample drawn from a Rayleigh's distribution\n")
colnames(results) <- c("Sample size", "Minimum", "...", "Median", "...", "Maximum")
print(results, row.names = FALSE)
```

Appendix H. Shapiro-Wilk W test and other tests of normality

```
# Shapiro-Wilk W test using Royston's procedure for samples of 12 to 2000 data points and statistical power estimation.

# Define a vector of data points
x <- c()

# Load required library
library(nortest)

# Define significance level
alpha <- 0.05

# Compute statistics
sw <- shapiro.test(x)
w <- sw$statistic
p_value <- sw$p.value
sample_size <- length(x)
m <- 0.0038915 * log(sample_size)^3 - 0.083751 * log(sample_size)^2 - 0.31082 * log(sample_size) - 1.5861
sd <- exp(0.0030302 * log(sample_size)^2 - 0.082676 * log(sample_size) - 0.4803)
z_w <- (log(1 - w) - m) / sd
zc <- qnorm(0.95, mean = m, sd = sd)
power <- 1 - pnorm((zc - log(1 - w)) / sd)

# Display results
cat("Shapiro–Wilk W statistic: w =", w, "\n")
cat("Logarithmic transformation of W statistic: ln(1 - w) =", log(1 - w), "\n")
cat("Expected value of ln(1 - w):", m, "\n")
cat("Standard deviation of ln(1 - w):", sd, "\n")
cat("Royston's standardized W statistic: z_w =", z_w, "\n")
cat("Right-tailed p-value:", p_value, "\n")

if (p_value < alpha) {
cat("\nThe null hypothesis that the data follow a normal distribution is rejected\nat the", alpha,
"significance level by the Shapiro–Wilk W test using Royston's procedure.\n")
} else {
cat("\nThe null hypothesis that the data follow a normal distribution is not rejected\nat the", alpha,
"significance level by the Shapiro–Wilk W test using Royston's procedure.\n")
}

cat("\nStatistical power: ϕ =", power, "\n")

# Lilliefors (Kolmogorov-Smirnov) normality test
```

```
lillie.test(x)
# Anderson-Darling normality test
ad.test(x)
# Outliers, symmetry, and kurtosis tests.
library(outliers)
grubbs.test(x, type = 10)
library(moments)
anscombe.test(x)
agostino.test(x)
```