

MVN Q-Test I: A Bootstrap-Based Implementation in R

José Moral de la Rubia

School of Psychology, Universidad Autónoma de Nuevo León, Monterrey, Mexico

Email: jose.morald@uanl.edu.mx

How to cite this paper: Moral de la Rubia, J. (2025) MVN Q-Test I: A Bootstrap-Based Implementation in R. *Journal of Data Analysis and Information Processing*, 13, 389-424. <https://doi.org/10.4236/jdaip.2025.134024>

Received: August 2, 2025

Accepted: September 20, 2025

Published: September 23, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In 2023, a multivariate normality test based on a chi-square approximation was developed. This method assumes independence among Gaussian random variables, and defines the test statistic, denoted by Q , as the sum of squared values. This study aims to develop R scripts that implement the Q -test for multivariate normality using either the Shapiro-Wilk W statistic (QSWa) or the Shapiro-Francia W' statistic (QSFa). A bootstrap version of the Q -test (QSWb and QSFb), which does not assume independence, is also included. Additionally, it incorporates Royston's H -test. The use of the scripts is illustrated with a sample of 50 participants assessed on a variable across four yearly administrations. The sampling distribution generated by the bootstrap method differs from the chi-square distribution and corresponds to a generalized chi-square distribution—namely, the distribution of a sum of squares of correlated variables. This distribution is less peaked and has a heavier right tail than the chi-square distribution. It is concluded that the bootstrap approach is conservative under the null hypothesis of multivariate normality; however, it is theoretically more appropriate than the chi-square approximation. To approximate the distributions of the two versions of the Q -test, it is recommended that the z or z' values set to zero in the calculation of the Q statistic not be subtracted when determining the degrees of freedom in the chi-square approximation. Moreover, a significance level of 10% is suggested for the bootstrap approach, rather than the conventional 5%.

Keywords

Multivariate Normality Tests, Sampling Distribution, Bootstrap Methods, R Program, Computational Statistics

1. Introduction

In 2023, Moral developed the multivariate normality Q -test, which demonstrated

adequate power and efficiency [1] compared with Royston's H-test [2], a well-known test for assessing multivariate normality. In its original formulation, the Q-test's p-value and statistical power were approximated using the chi-square distribution. However, a more flexible bootstrap approach was recommended to relax the test's assumptions [1]. The Q-test comprises two variants: one based on Royston's standardization [3] of the Shapiro-Wilk W statistic [4], and another based on Royston's standardization [5] of the Shapiro-Francia W' statistic [6]. These are referred to as QSWa and QSFa, respectively. In the original simulation study [1], QSWa was found to be slightly more advantageous than QSFa, and its performance closely resembled that of Royston's H-test [2].

The first objective of this article is to develop a script in the R programming language [7] that facilitates the application of the Q-test. The script implements both the chi-square approach of the original proposal (QSWa and QSFa) and the suggested bootstrap approach (QSWb and QSFb). It also includes a test of the serial independence of the sequence of standardized univariate normality statistics for the $2^k - 1$ possible unweighted linear combinations of the original k variables (an assumption of the asymptotic approach), as well as Royston's H-test [2]. Additionally, the script computes statistical power estimates using both the chi-square approximation and the bootstrap method.

The use of the script is illustrated with a dataset comprising fifty 4-tuples. The four variables in this convenience sample are correlated, as they represent total scores on the Toronto Alexithymia Scale [8]—ranging from 20 to 100—collected across four annual administrations. Assessments were conducted online with a cohort of 50 Mexican university students (25 women and 25 men) between 2019 and 2023. The null hypothesis of multivariate normality is expected to be supported by all four Q-test variants and the H-test [9]-[11].

The R programming language was selected because it is one of the most comprehensive statistical software environments, open-access, and continuously developed and reviewed by the mathematical community [12] [13]. The H-test developed by Royston [2] was included as a supplementary test because it is one of the most widely used and powerful tests for assessing multivariate normality [14] [15]. In contrast, Mardia's K^2 test [16] was excluded, as a previous study [1] found that the Q-test outperformed it in terms of hit rate and mean power, particularly when the Q-test's performance closely approximated that of Royston's H-test.

2. The Q-Test for Multivariate Normality and the Development of Its Script

To obtain the test statistic for the MVN Q-test [1], the standardized statistics of the Shapiro-Wilk univariate normality test [4] (Equation (1)) or the Shapiro-Francia test [6] (Equation (2)) are computed, using Royston's (1992, 1993) standardization [3] [5], for all possible unweighted linear combinations of the k variables. The number of such combinations is given by $nc = 2^k - 1$ (Equation (3)). In the first variant, the test statistic is denoted by Q_{SW} as a random variable and q as

a specific observed value. In the second variant, it is denoted by Q_{SF} as a random variable and q' as its observed value.

$$Z(\ln(1-W)) = \frac{W - \mu(\ln(1-W))}{\sigma(\ln(1-W))} = z$$

$$\mu(\ln(1-W)) = 0.0038915 \times \ln^3(n) - 0.083751 \times \ln^2(n) - 0.31082 \times \ln(n) - 1.5861 \tag{1}$$

$$\sigma(\ln(1-W)) = e^{0.0030302 \times \ln^2(n) - 0.082676 \times \ln(n) - 0.4803}$$

$$Z(\ln(1-W')) = \frac{W' - \mu(\ln(1-W'))}{\sigma(\ln(1-W'))} = z'$$

$$u = \ln(\ln(n)) - \ln(n)$$

$$\mu(\ln(1-W')) = 1.0521 \times u - 1.2725 \tag{2}$$

$$v = \ln(\ln(n)) + 2/\ln(n)$$

$$\sigma(\ln(1-W')) = -0.26758 \times v + 1.0308$$

$$nc = \sum_{j=1}^k \binom{k}{j} = \binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k} = 2^k - 1 \tag{3}$$

Negative values of z or z' are truncated to zero to avoid overestimating Q_{SW} or Q_{SF} as negative z or z' values indicate a good fit to normality when calculating the right-tailed p-value under a standard normal distribution. See Equations (4) and (5).

$$Z = \begin{cases} z & Z(\ln(1-W)) > 0 \\ 0 & Z(\ln(1-W)) \leq 0 \end{cases} \tag{4}$$

$$Z' = \begin{cases} z' & Z(\ln(1-W')) > 0 \\ 0 & Z(\ln(1-W')) \leq 0 \end{cases} \tag{5}$$

Once the $2^k - 1$ standardized statistics are obtained and the negative ones have been truncated to zero, the sum of squares (Q statistic) is computed, and the right-tailed p-value is derived from a chi-square distribution. The degrees of freedom (df) correspond to the number of summed values (nc) minus the number of truncated values (a): $df = nc - a$, as per the original proposal. See Equation (6).

$$Q_{SW} = \sum_{i=1}^{nc} z_i^2 = q; Q_{SW} \sim \chi_{nc-a}^2$$

$$Q_{SF} = \sum_{i=1}^{nc} z'_i{}^2 = q'; Q_{SF} \sim \chi_{nc-a}^2 \tag{6}$$

In the script, subtracting truncated values is optional. In the simulation study, this option was not applied; therefore, the default is to set the degrees of freedom equal to the number of combinations (nc). The true distribution of the test statistic is a generalized chi-square distribution, which is less peaked and more positively skewed than a chi-square distribution [17]. For this reason, it was considered the preferable choice.

Since the asymptotic approach based on the chi-square distribution for the Q

statistic assumes not only the normality of the sum variable but also the independence of the variables being summed [18], their serial independence is assessed using the Wald-Wolfowitz runs test [19] via the DescTools package in R [20], and the Ljung-Box test [21] via R's stats package [22]. The latter is included because it allows the specification of a maximum lag (h) over which the analysis is conducted (from 1 to h), whereas the Wald-Wolfowitz test corresponds to a lag of one. For non-seasonal series or those without an expected temporal pattern, it is recommended to examine lags from 1 to the smaller of 10 or one-fifth of the sample size [23]. Schwert's rule [24], which is widely used to determine the maximum lag, is also provided as an option [25]. See Equation (7).

$$h = 12 \times (n/100)^{1/4} \quad (7)$$

Collinearity resulting from possible linear dependencies among k variables is substantially determined by the strength and configuration of their pairwise correlation matrix. As the number of variables increases, so does the probability that certain linear combinations will exhibit high dependency, even if only a few bivariate correlations are statistically significant and of weak or moderate strength. Conversely, the influence of one or two strong correlations may be diluted in the presence of many variables (e.g., 10 or more), as the resulting collinearity will depend more on the overall correlation structure than on isolated relationships.

Nevertheless, serial dependency in test statistic values for all possible linear combinations of k variables, ordered by increasing complexity, is more strongly affected by the number of variables than by the correlations among them. As a result, maintaining the null hypothesis of serial independence becomes increasingly difficult with a large number of variables when using Wald-Wolfowitz or Ljung-Box tests.

Additionally, the script incorporates a bootstrap approach as an alternative for calculating the p-value and statistical power [26] [27], which is particularly useful when the assumption of independence in the sequence of $2^k - 1$ linear combinations is violated. The sampling distribution of the Q statistic is estimated in two ways:

- 1) Using 1000 bootstrap samples drawn from the original dataset of n observations on k correlated variables (empirical bootstrap distribution).
- 2) Using 1000 samples obtained through bootstrap resampling with replacement, drawn from a sample of size n consisting of k normally distributed variables that preserve the correlational structure of the original sample. This sample is referred to as the normative sample, and the resulting distribution as the normative bootstrap distribution.

Consequently, both non-parametric and parametric bootstrap procedures are applied—the latter under the null hypothesis of multivariate normality.

To generate the population or source sample in the parametric bootstrap, an $n \times k$ matrix of independent Gaussian vectors is multiplied by the lower triangular matrix from the Cholesky decomposition of the original variables' correlation matrix [28] [29]. The k Gaussian vectors each consist of n scores drawn from the

standard normal distribution. These scores are generated using equispaced quantile orders ranging from $0.5/n$ to $1 - 0.5/n$, calculated via the probit function. The quantile orders are then permuted using a fixed seed (123) to ensure the stability of the results. In this way, each of the k variables shares the same set of quantile orders, but in a different sequence, thereby forming independent vectors.

From the normative sample, 1000 samples are drawn with replacement and in each sample the test statistic is calculated. These 1000 estimates constitute the normative Bootstrap distribution. This distribution is referred to as normative because it allows for the estimation of the bootstrap critical value (e.g., the 0.95 quantile).

The bootstrap p-value (*i.e.*, the proportion of values greater than or equal to the test statistic q or q') is calculated from the normative bootstrap distribution (a generalized chi-square distribution), serving as a substitute for the chi-square distribution used in the approximate approach. Statistical power (φ) at a given significance level α is estimated as the proportion of rejections in the empirical bootstrap sample, where the rejection region is defined by the $(1 - \alpha)$ th quantile of the normative bootstrap distribution. In both cases, resampling with replacement is performed by rows rather than by columns, in order to preserve the correlations among the variables in the source sample. For stable results, a fixed seed (123) is used when generating all 1000 bootstrap samples [7].

On the other hand, the median of the normative bootstrap distribution is used to calculate the probability associated with this central value in both the left and right tails of the empirical bootstrap sample distribution. Twice the smaller of these two probabilities represents the likelihood that the empirical bootstrap distribution is centered on the median of the normative bootstrap distribution. The closer this probability is to 1, the stronger the evidence of equivalence between the medians of the theoretical and empirical bootstrap distributions; conversely, a value below 0.05 indicates a clear shift of the empirical bootstrap distribution relative to the normative one. A divergence of the empirical bootstrap distribution from the normative distribution indicates a deviation from multivariate normality. It should be noted that the normative q and q' values must be closer to 0 than the number of combinations of the k original variables and far from the median of the normative bootstrap distribution in the case of multivariate normality.

Finally, both distributions are visualized using histograms and density curves to assess whether the sampling distribution of the Q statistic conforms to a chi-square distribution (reference curve) or instead adopts the shape of a generalized chi-square distribution, as is typical for quadratic forms or sums of squares of correlated variables [30]. This generalized distribution exhibits a lower peak and a heavier right tail compared to the standard chi-square distribution. The number of class intervals (k) and their constant width (w) are determined using Rice University's rule (Equation (8)), which is appropriate for a wide range of distributional shapes, including positively skewed and leptokurtic distributions [31].

$$\begin{aligned}
 \text{Number of class intervals : } k &= 2 \times \sqrt[3]{n} = 2 \times \sqrt[3]{1000} = 20 \\
 \text{Interval width : } w &= \begin{cases} \frac{\min(\{q_i\}_{i=1}^n) - \min(\{q_i\}_{i=1}^n)}{20} \\ \frac{\min(\{q'_i\}_{i=1}^n) - \min(\{q'_i\}_{i=1}^n)}{20} \end{cases} \quad (8)
 \end{aligned}$$

where q_i or q'_i ($i = 1, 2, \dots, 1000$) denote the test statistics computed from the 1000 bootstrap samples drawn from the original sample of size n .

It should be noted that, in bootstrap procedures, using 1000 resamples is often considered a computationally efficient choice, as it offers a good balance between statistical accuracy and processing time. Consequently, many applied fields have adopted 1000 as the default setting in software packages such as R, SPSS, SAS, and Stata, as it yields reliable results in typical scenarios without placing excessive demands on computational resources [26] [27]. This number of resamples is also well-suited for the online R platform (Snippets—Run R code online), which allows up to approximately one minute for script execution.

Appendix 1 contains the script for the Q-test calculated using the Shapiro-Wilk univariate normality statistics standardized by Royston's method for samples ranging from 12 to 2000 with k-tuples of measurements [3]. **Appendix 2** presents the Q-test based on the Shapiro-Francia univariate normality statistics [4], standardized according to Royston [5], for sample sizes from 5 to 5000 with k-tuples of measurements [6]. The Q-test is supplemented by Royston's H-test [2], included in **Appendix 3**, for sample sizes from 10 to 2000 with k-tuples of measurements. A significance level (α) of 0.05 is applied to both the H-test and the chi-square approximation of the Q-test, while a level of 0.1 is used for the bootstrap version of the Q-test, which is highlighted in green. To run the scripts, the following R packages must be installed and loaded: `nortest` [32], `MASS` [29], `DescTools` [20], `stats` [22], and `MVN` [33].

Performance benchmarks for the three scripts were obtained on a PC running Windows 11, equipped with a 500 GB solid-state drive, 12 GB of RAM, and a 2.8 GHz processor. In the two variants of the MVN Q test (**Appendix 1**), run times averaged 17.689 seconds and sample standard deviation (sd) 16.409, ranging from 2.030 to 52.950 seconds across the tested configurations. Memory usage averaged 13.317 MB ($sd = 6.702$), with values ranging from 3.554 to 30.212 MB. These benchmarks were obtained for sample sizes ranging from 50 to 500 (in increments of 50) and numbers of variables from 1 to 6 (in increments of 1), with a homogeneous correlation of 0.5. The results indicate a moderate computational cost, with substantial variability in run time depending on data size and dimensionality.

For the two variants of the MVN Q' test (**Appendix 2**), run times averaged 22.014 seconds ($sd = 20.480$), ranging from 2.570 to 66.990 seconds. Memory usage averaged 13.323 MB ($sd = 6.702$), ranging from 3.559 to 30.218 MB. These benchmarks, obtained under the same conditions as the Q test, suggest that Q'

generally requires slightly more processing time, particularly for larger configurations, while its memory footprint remains virtually identical to that of Q.

In the Royston MVN H test (**Appendix 3**), run times averaged 0.030 seconds ($sd = 0.040$), ranging from 0.010 to 0.300 seconds. Memory usage averaged 1.704 MB ($sd = 0.018$), ranging from 1.657 to 1.749 MB. Using the same data configurations, this test was completed in a fraction of a second with minimal memory consumption, making it highly efficient and computationally inexpensive compared to both variants of the MVN Q and Q' tests.

3. Application of the R Script for the Q-Test

The three scripts were applied to four repeated measures in a sample of 50 participants. However, they can be adapted to other datasets by modifying the list of variables highlighted in blue. **Appendix 1** provides the script used to generate a reproducible random sample of 50 observations across four correlated variables following a multivariate normal distribution. The resulting sample is presented in **Appendix 2** and **Appendix 3** under the “List of original variables” section. The scripts can be executed either on a local machine with R installed or online via the Snippets platform, which is available at <https://rdr.io/snippets/>.

In this example, the run time for the two variants of the MVN Q test (**Appendix 1**) was 10.6 seconds, with a memory usage of 23.683 MB. Similarly, the run time for the two variants of the MVN Q' test (**Appendix 2**) was 10.5 seconds, also with a memory usage of 23.683 MB. In contrast, the run time for the Royston MVN H test (**Appendix 3**) was 0.06 seconds, with a memory usage of 1.670 MB. All tests were performed on a PC running Windows 11, equipped with a 500 GB solid-state drive, 12 GB of RAM, and a 2.8 GHz processor.

To save the graphic as a JPG file, uncomment the corresponding lines by removing the hashtag symbols from the `#jpeg()`, `#par()`, and `#dev.off()` commands.

3.1. The Chi-Square Approximation and Bootstrap Version of the Q-Test Based on Shapiro-Wilk W Statistics

Sample size: $n = 50$.

Number of variables in the original sample: $k = 4$.

Arithmetic mean of the Pearson's product-moment correlation coefficients between the variables: $m(R_{4 \times 4}) = 0.5647$.

Level of significance: $\alpha = 0.05$ for the chi-square approximation-based Q-test, and $\alpha = 0.10$ for the bootstrap version of the Q-test, to compensate for the pronounced right-skewness in the sampling distribution of the Q statistic.

Parameters for standardizing the 15 variables created by linear combination of the 4 original variables (**Table 1**):

Expected value of $\ln(1 - W)$: $\mu = -3.850773$.

Standard deviation of $\ln(1 - W)$: $\sigma = 0.4689044$.

According to the Shapiro-Wilk test (Royston, 1992), the null hypothesis of univariate normality is maintained for the 15 linear combinations at a 5% significance

level, as shown in **Table 1**.

Table 1. Shapiro-Wilk’s univariate normality test with Royston’s standardization for the 15 linear combinations of variables.

Comb	<i>W</i>	<i>z</i>	<i>p</i>	Norm	Comb	<i>W</i>	<i>z</i>	<i>p</i>	Norm
c1	0.980	-0.171	0.568	Yes	c9	0.978	0.032	0.487	Yes
c2	0.963	1.181	0.119	Yes	c10	0.975	0.329	0.371	Yes
c3	0.966	1.004	0.158	Yes	c11	0.980	-0.154	0.561	Yes
c4	0.988	-1.131	0.871	Yes	c12	0.986	-0.946	0.828	Yes
c5	0.992	-2.108	0.983	Yes	c13	0.986	-0.906	0.818	Yes
c6	0.987	-0.972	0.834	Yes	c14	0.971	0.671	0.251	Yes
c7	0.990	-1.519	0.936	Yes	c15	0.982	-0.316	0.624	Yes
c8	0.964	1.131	0.129	Yes					

Note. Comb = sum of the corresponding combination without repetition of variables; *W* = Shapiro-Wilk *W* statistic; *z* = standardized value of *W* using Royston’s method; *p* = right-tailed *p*-value under the standard normal distribution; Norm = Univariate normality: Yes = $p \geq \alpha = 0.05$, the null hypothesis of normality is retained; No = $p < 0.05$, the null hypothesis is rejected.

Q-test statistic value in the original sample: $q = 4.2403$.

Number of combinations among the 4 original variables: $nc = 15$.

Number of *z*-values truncated to zero because they are negative: $a = 9$.

Degrees of freedom: $df = 15$. For the chi-square approximation of the QSW test, the values set to zero were not subtracted from the degrees of freedom in order to compensate for the pronounced right-skewness of the sampling distribution—an effect revealed by the bootstrap approach. Subtracting them would reduce the degrees of freedom to 6.

Critical value or 0.95 quantile of the chi-square distribution with 15 degrees of freedom: $_{0.95}\chi^2_{[15]} = 24.9958$. The null hypothesis of multivariate normality is retained at a significance level of 0.05 based on the chi-square approximation critical value.

Right-tailed probability value under the chi-square distribution with 15 degrees of freedom: $p = 0.9968$. The null hypothesis of multivariate normality is retained at a significance level of 0.05 based on the chi-square approximation *p*-value.

Statistical power based on the chi-square approximation at a significance level of 0.05: $\varphi = 0.1889$.

In the sequence of 15 standardized Shapiro-Wilk *W* values [3]—truncated to 0 when negative ($a = 9$)—used to compute the Q-test statistic, no evidence of serial dependence was found according to the Wald-Wolfowitz runs test [19] and the Ljung-Box test [21]. The maximum lag was determined based on a general rule for non-stationary series: $\min(10, nc/5) = \min(10, 15/5) = 3$ [23]. According to Schwert’s rule [25], the maximum lag would be 7. Refer to **Table 2** for further details.

Runs Test for Randomness: runs = 7, $m = 9$, $n = 6$, p-value = 0.5804 > $\alpha = 0.05$.

Table 2. Ljung-Box test for serial independence of z_w values.

Lag	X-square	df	p-value
1	0.0258	1	0.8724
2	0.6219	2	0.7327
3	1.7434	3	0.6273

Note. Lag = step between values forming the autocorrelation pairs; X-square = test statistic; df = degrees of freedom; p-value = right-tailed probability under the chi-square distribution.

Empirical bootstrap distribution (with preserved correlations) derived from the original sample:

Number of bootstrap samples: $B = 1000$.

Right-tailed bootstrap p-value: $p_{boot} = 0.999$. The null hypothesis of multivariate normality is not rejected at the 0.05 significance level based on the bootstrap p-value obtained from the empirical bootstrap distribution.

Normative bootstrap distribution (with preserved correlation):

Number of bootstrap samples: $B = 1000$.

Q-test statistic value in the normative sample: $q_{mvn_boot} = 15.24668$.

Mean of the normative bootstrap distribution: $m(mvn_boot_dist) = 49.14216$.

Median of the normative bootstrap distribution: $mdn(mvn_boot_dist) = 46.84413$.

Two-tailed p-value for testing whether the empirical bootstrap distribution is centered on the median of the normative bootstrap distribution: $p_{2-tailed} = 0.338$. The null hypothesis of equal medians between the normative and empirical bootstrap distributions is not rejected at a significance level of 0.1.

Bootstrap critical value (0.9 quantile of the normative bootstrap distribution): $q_{crit} = 75.4841$. The null hypothesis of multivariate normality is not rejected at a significance level of 0.1 based on the critical value from the normative bootstrap distribution.

Right-tailed p-value in the normative bootstrap distribution: $p_{mvn_boot} = 1$. The null hypothesis of multivariate normality is not rejected at a significance level of 0.1 based on the bootstrap p-value.

Bootstrap statistical power at a significance level of 0.1: $\phi_{boot} = 0.018$.

Figure 1 shows that the value of the Q statistic lies to the left of the critical values of both the chi-square and the normative bootstrap distributions. This indicates that it falls within the acceptance region of the null hypothesis of multivariate normality. The empirical and normative bootstrap distributions are similar and deviate from the chi-square distribution by exhibiting lower peaks, flatter shoulders, and heavier right tails, consistent with a generalized chi-square distribution. However, when the assumption of serial independence is met, the chi-square approximation remains a valid alternative for estimating the p-value and making decisions regarding the null hypothesis.

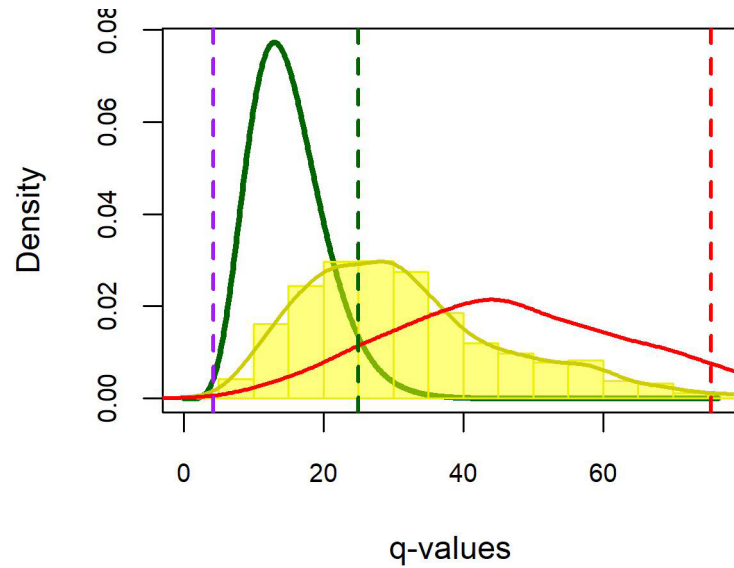


Figure 1. Histogram (yellow) with density curves for the empirical bootstrap (yellow), normative bootstrap (red), and chi-square (green) distributions. The value of the Q-test statistic in original sample is indicated by a purple vertical line ($q = 4.2403$), the critical value of the normative bootstrap distribution by a red vertical line (0.9 quantile: $q_mvn_boot = 75.4841$), and the critical value of the chi-square distribution with 15 degrees of freedom by a green vertical line (0.95 quantile: ${}_{0.95}\chi^2_{[15]} = 24.9958$).

3.2. The Chi-Square Approximation and Bootstrap Version of the Q-Test Based on Shapiro-Francia W' Statistics

The script for running the multivariate normality Q'-test—calculated using the standardized values of the Shapiro-Francia univariate normality W' statistics [5]—is provided in **Appendix 2**. It applies to the same sample of fifty 4-tuples.

Arithmetic mean of the Pearson's product-moment correlation coefficients between the variables: $m(R_{4 \times 4}) = 0.5647$.

Parameters for standardizing the 15 variables created by linear combination of the 4 original variables:

Expected value of $\ln(1 - W')$: $\mu = -3.9532$.

Standard deviation of $\ln(1 - W')$: $\sigma = 0.5290$.

According to the Shapiro-Francia test (Royston, 1993) [5], the null hypothesis of univariate normality is maintained for the 15 linear combinations at a 5% significance level, as shown in **Table 3**.

Table 3. Shapiro-Francia's univariate normality test with Royston's standardization for the 15 linear combinations of variables.

Comb	W	z	p	Norm	Comb	W	z	p	Norm
c1	0.980	0.042	0.483	Yes	c9	0.978	0.222	0.412	Yes
c2	0.963	1.240	0.107	Yes	c10	0.975	0.485	0.314	Yes
c3	0.966	1.083	0.139	Yes	c11	0.980	0.058	0.477	Yes
c4	0.988	-0.808	0.791	Yes	c12	0.986	-0.645	0.740	Yes
c5	0.992	-1.675	0.953	Yes	c13	0.986	-0.609	0.729	Yes
c6	0.987	-0.668	0.748	Yes	c14	0.971	0.789	0.215	Yes

Continued

c7	0.990	-1.153	0.876	Yes	c15	0.982	-0.087	0.535	Yes
c8	0.964	1.196	0.116	Yes					

Note. Comb = sum of the corresponding combination without repetition of variables; W = Shapiro-Francia W' statistic; z = standardized value of W' using Royston's method; p = right-tailed p-value under the standard normal distribution; Norm = Univariate normality: Yes = $p \geq \alpha = 0.05$, the null hypothesis of normality is retained; No = $p < 0.05$, the null hypothesis is rejected.

Q' -test statistic value in the original sample: $q = 1.8161$.

Number of combinations among the 4 original variables: $nc = 15$.

Number of z-values truncated to zero because they are negative: $a = 10$.

Degrees of freedom: $df = 15$.

Critical value or 0.95 quantile of the chi-square distribution with 15 degrees of freedom: $1-\alpha\chi^2_{[df]} = 24.9958$. The null hypothesis of multivariate normality is retained at a significance level of 0.05 based on the chi-square approximation critical value.

Right-tailed probability value under the chi-square distribution with 15 degrees of freedom: $p = 1$. The null hypothesis of multivariate normality is retained at a significance level of 0.05 based on the chi-square approximation p-value.

Statistical power based on the chi-square approximation at a significance level of 0.05: $\varphi = 0.0992$.

In the sequence of 15 standardized Shapiro-Francia W' values—truncated to 0 when negative ($a = 10$)—used to compute the Q' -statistic, no serial dependence was found according to the Wald-Wolfowitz runs test [19] and the Ljung-Box test [21] (Table 4). The maximum lag is determined by a general rule for non-stationary series: $\min(10, nc/5) = \min(10, 15/5) = 3$ [23]. According to Schwert's rule [25], the maximum lag would be 7.

Runs Test for Randomness. data: z_w : runs = 9, $m = 10$, $n = 5$, p-value = 0.5604.

Table 4. Ljung-Box test for serial independence of z_w' values.

Lag	X-square	df	p-value
1	0.1120	1	0.7378
2	0.7789	2	0.6774
3	1.6893	3	0.6393

Note. Lag = step between values forming the autocorrelation pairs; X-square = test statistic; df = degrees of freedom; p-value = right-tailed probability under the chi-square distribution.

Empirical bootstrap distribution (with preserved correlations) derived from the original sample:

Number of bootstrap samples: $B = 1000$.

Right-tailed bootstrap p-value: $p_{boot} = 1$.

The null hypothesis of multivariate normality is not rejected at the 0.05 significance level based on the bootstrap p-value obtained from the empirical bootstrap

distribution.

Normative bootstrap distribution (with preserved correlation):

Number of bootstrap samples: $B = 1000$.

Q²-test statistic value in the normative sample: $q_{m\text{vn_boot}} = 8.0623$.

Mean of the normative bootstrap distribution: $m(\text{mvn_boot_dist}) = 33.31423$.

Median of the normative bootstrap distribution: $mdn(\text{mvn_boot_dist}) = 30.7692$.

Two-tailed p-value for testing whether the empirical bootstrap distribution is centered on the median of the normative bootstrap distribution: $p_{2\text{-tailed}} = 0.49$. The null hypothesis of equal medians between the normative and empirical bootstrap distributions is not rejected at a significance level of 0.1.

Bootstrap critical value (0.9 quantile of the normative bootstrap distribution): $q_{\text{crit}} = 53.4975$. The null hypothesis of multivariate normality is not rejected at a significance level of 0.1 based on the critical value from the normative bootstrap distribution.

Right-tailed p-value in the normative bootstrap distribution: $p_{\text{mvn_boot}} = 1$. The null hypothesis of multivariate normality is not rejected at a significance level of 0.1 based on the bootstrap p-value.

Bootstrap statistical power at a significance level of 0.1: $\text{phi_boot} = 0.037$.

Figure 2 shows that the value of the Q²-statistic lies to the left of both the critical value (95th percentile) of the chi-square distribution with 15 degrees of freedom and the critical value (90th percentile) of the normative bootstrap distribution. This indicates that the null hypothesis of multivariate normality is accepted.

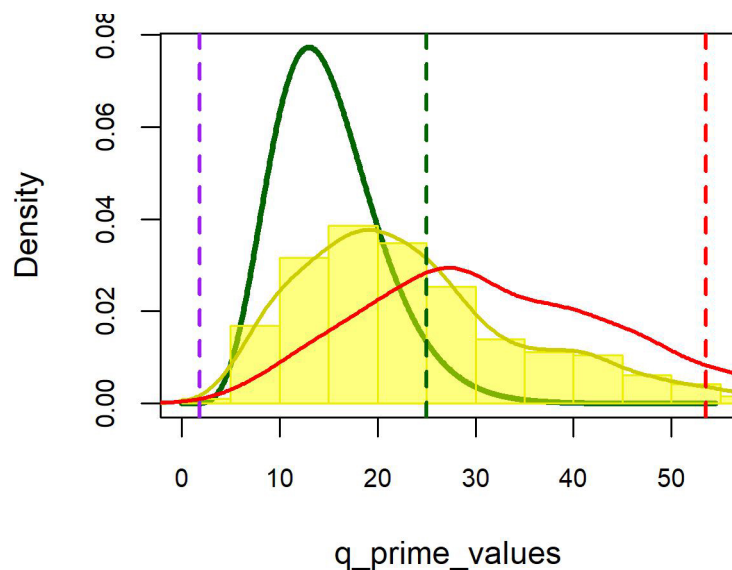


Figure 2. Histogram (yellow) with density curves for the empirical bootstrap (yellow), normative bootstrap (red), and chi-square (green) distributions. The value of the Q²-test statistic in original sample is indicated by a purple vertical line ($q^2 = 1.8160$), the critical value of the normative bootstrap distribution by a red vertical line (0.9 quantile: $q_{\text{mvn_boot}} = 53.4975$), and the critical value of the chi-square distribution with 15 degrees of freedom by a green vertical line (0.95 quantile: ${}_{0.95}\chi^2_{[15]} = 24.9958$).

Furthermore, the distributions of the Q'-statistic generated by nonparametric and parametric bootstrap methods (based on Gaussian vectors with a correlational structure equivalent to that of the original data) are similar. Both deviate from the chi-square distribution with 15 degrees of freedom by exhibiting a lower peak, flatter shoulders, and a heavier right tail, consistent with a generalized chi-square distribution.

To compensate for this heavier right tail, a 90th percentile is used as the critical value for the normative bootstrap distribution. In addition, the correction that subtracts the number of q' values truncated to zero when computing the degrees of freedom of the chi-square distribution was omitted ($df=15$ without correction vs. $df=5$ with correction).

3.3. H-Test

When assessing multivariate normality using Royston's H-test [2], the null hypothesis of normality was not rejected ($H=4.8778$, $df=4.0988$, $p=0.3127 > \alpha=0.05$). Since the null hypothesis was not rejected, the test exhibited statistical power below 0.5 ($\phi=0.3833$) at a 0.05 significance level. **Appendix 3** provides the R script used for these calculations.

4. Discussion

The objective of the study was to develop a script for the R program and demonstrate its use. For practical purposes, three separate scripts were created. The first script applies both the chi-square approximation and the bootstrap version of the Q-test using the Shapiro-Wilk W statistic (QSWa and QSWb). The second script performs the same procedures using the Shapiro-Francia W' statistic (QSFa and QSFb). The third script applies the H-test proposed by Royston (1983). All three scripts include the calculation of the critical value, p-value, and statistical power. As an example, the three scripts were run on a dataset of 50 observations on four correlated variables, representing total scores on a personality scale (alexithymia) collected at four annual time points. The repeatedly measured variable had a discrete range from 20 to 100.

It was expected that the assumption of multivariate normality would be consistently supported by all five MVN tests. As anticipated, the null hypothesis of multivariate normality was retained by all five tests when statistical power was below 0.5. Power was low for the H-test and very low for the four Q-test variants, with the two bootstrap versions showing lower power than the two based on the chi-square approximation.

Since the assumption of serial independence is satisfied, the Q-test version based on the chi-square approximation is a suitable method for calculating the critical value, p-value, and statistical power, as well as for making decisions regarding the null hypothesis. In this case, the statistical power is lower—indicating better test performance when the null hypothesis is true—when using the Shapiro-Francia W' statistic compared to the Shapiro-Wilk W statistic.

The profiles of the sampling distribution generated through repeated sampling—whether using the multivariate standard normal distribution (parametric bootstrap) or the original sample (nonparametric bootstrap)—are similar to each other, but both differ from the contour of the chi-square distribution on which the Q-test's approximate calculation is based. This approximation relies on the assumption that the theoretical sampling distribution of the sum of squares of independent standard normal variables follows a chi-square distribution with degrees of freedom equal to the number of summed variables [1]. Consequently, the bootstrap version is theoretically more robust. It is worth noting that when the Shapiro-Francia W' statistic is used, the three distributional profiles are more similar than when the Shapiro-Wilk W statistic is applied. This is a desirable property, in addition to the lower power when the null hypothesis is retained—as in the present case—or higher power when it is rejected.

The generated profiles are less peaked and exhibit a heavier right tail than the chi-square distribution, as they correspond to a generalized chi-square distribution resulting from the sum of squares of correlated variables [34]. These profiles, characterized by pronounced positive skewness, explain why the number of truncated z' or z -values ($df=15$ uncorrected vs. 6 corrected in QSWa, and $df=15$ vs. 5 in QSFa) is not excluded when calculating the degrees of freedom for the chi-square distribution. As a result, the critical value of the chi-square approximation shifts further to the right. For the normative bootstrap distribution, the quantile is adjusted from 0.95 to 0.90 to account for its substantially heavier right tail, thereby shifting the critical value to the left. These adjustments aim to bring the conclusions derived from both distributions—the chi-square and the generalized chi-square—closer together. In the example shown, closer alignment is more effectively achieved using the variants based on the Shapiro-Francia W' statistic.

The proximity of the two bootstrap distributions—normative and empirical—can be assessed not only visually using the graph provided by the script, but also statistically by testing the equivalence of their medians. In the example shown, this equivalence is positively corroborated, supporting the claim that the two bootstrap distributions are similar.

The normative bootstrap distribution is used in place of the chi-square distribution to calculate the probability value, and it enables the determination of the cut-off point (critical value) for defining the rejection region and estimating statistical power within the empirical bootstrap distribution. It is worth noting that one of the key advantages of the bootstrap method is its ability to approximate the sampling distribution of an estimator or test statistic without requiring parametric assumptions [27].

Accordingly, the script also computes the probability value based on the empirical sampling distribution obtained via non-parametric bootstrap. This approach is particularly common when the null hypothesis pertains to a descriptive or association measure with an unknown sampling distribution. Conversely, when the null hypothesis concerns a known distribution (e.g., multivariate normal), para-

metric bootstrap under that assumption is preferred [27].

The three scripts developed to assess normality assume that the tested distribution is continuous; therefore, the measured variables must be continuous or treated as such, even if their values are recorded as integers—as in the case of intelligence quotient scores in cognitive tests. Nevertheless, several discrete distributions converge toward normality, such as the binomial distribution when n (the number of independent trials) tends toward infinity and p (the probability of success) approaches 0.5. This allows for broader application to discrete values, meaning the data do not necessarily need to include decimal numbers.

One limitation of this study is that it does not assess the performance of the test in terms of hit rate and statistical power under both normal and non-normal multivariate distributions. In this regard, it is important to consider factors such as sample size, number of variables, and inter-variable correlations [35]. Additionally, comparing the four versions of the Q-test with each other and with the H-test—and even with another MVN test, such as the one proposed by Henze and Zirkler [36]—would be valuable. It is recommended that future research address these aspects.

A basic simulation scenario would involve generating samples from multivariate normal distributions as well as from one or two types of non-normal multivariate distributions—specifically, Student's t distributions with 5 degrees of freedom and uniform distributions. The samples would vary in size (ranging from 50 to 1000, in increments of 50), in the number of variables (distribution parameter: from 2 to 6), and in the level of homogeneous correlation among variables (distribution parameter: from 0 to 0.9, in increments of 0.1), resulting in 1000 samples per distribution. Applying the three scripts to these samples and comparing the hit rates and statistical power (or the complement of power in the case of samples drawn from multivariate normal distributions) would yield valuable insights into the performance of the tests.

5. Conclusions

For the Q-test, the chi-square approach is a good option if the assumption of serial independence holds, as in this sample of 50 observations of four-tuples. However, if this assumption is violated, the bootstrap version is theoretically more appropriate. In such cases, it is advisable to compute the Q-test using a 10% significance level. As shown in the example, the bootstrap variant of the Q-test performed best when calculated using the Shapiro-Francia W' statistic. This approach is more conservative regarding the null hypothesis of multivariate normality than the H-test and chi-square approximation-based Q-test variants. Based on the sampling distribution of the Q statistic revealed by the bootstrap method, subtracting the number of z or z' -statistics truncated to zero when calculating the degrees of freedom in chi-square approximation-based Q-test variants is not recommended.

Using these three scripts for the four Q-test variants and the H-test is recommended. Convergence of results between the Q-test and the H-test when running

the scripts reinforces the validity of the conclusions drawn.

Acknowledgements

The author expresses gratitude for the reviewers and editor for their helpful comments.

Conflicts of Interest

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

References

- [1] Rubia, J.M.L. (2023) Proposal and Pilot Study: A Generalization of the W or W' Statistic for Multivariate Normality. *Open Journal of Statistics*, **13**, 119-169. <https://doi.org/10.4236/ojs.2023.131008>
- [2] Royston, J.P. (1983) Some Techniques for Assessing Multivariate Normality Based on the Shapiro-Wilk W. *Applied Statistics*, **32**, 121-133. <https://doi.org/10.2307/2347291>
- [3] Royston, P. (1992) Approximating the Shapiro-Wilk W-Test for Non-Normality. *Statistics and Computing*, **2**, 117-119. <https://doi.org/10.1007/bf01891203>
- [4] Shapiro, S.S. and Wilk, M.B. (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, **52**, 591-611. <https://doi.org/10.2307/2333709>
- [5] Royston, P. (1993) A Toolkit for Testing for Non-Normality in Complete and Censored Samples. *The Statistician*, **42**, 37-43. <https://doi.org/10.2307/2348109>
- [6] Shapiro, S.S. and Francia, R.S. (1972) An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, **67**, 215-216. <https://doi.org/10.1080/01621459.1972.10481232>
- [7] R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- [8] Bagby, R.M., Parker, J.D.A. and Taylor, G.J. (1994) The Twenty-Item Toronto Alexithymia Scale—I. Item Selection and Cross-Validation of the Factor Structure. *Journal of Psychosomatic Research*, **38**, 23-32. [https://doi.org/10.1016/0022-3999\(94\)90005-1](https://doi.org/10.1016/0022-3999(94)90005-1)
- [9] Salminen, J.K., Saarijärvi, S., Äärelä, E., Toikka, T. and Kauhanen, J. (1999) Prevalence of Alexithymia and Its Association with Sociodemographic Variables in the General Population of Finland. *Journal of Psychosomatic Research*, **46**, 75-82. [https://doi.org/10.1016/s0022-3999\(98\)00053-1](https://doi.org/10.1016/s0022-3999(98)00053-1)
- [10] Franz, M., Popp, K., Schaefer, R., Sitte, W., Schneider, C., Hardt, J., *et al.* (2007) Alexithymia in the German General Population. *Social Psychiatry and Psychiatric Epidemiology*, **43**, 54-62. <https://doi.org/10.1007/s00127-007-0265-1>
- [11] Wilczyński, K.M., Cichoń, L., Stasik, A., Kania, K., Rodak, N., Wizner, M., *et al.* (2024) An Analysis of the Time Required for the Diagnosis of ASD and the Factors Influencing Its Duration in a Sample of the Pediatric Population from Poland. *Journal of Clinical Medicine*, **13**, Article 6255. <https://doi.org/10.3390/jcm13206255>
- [12] Braun, W.J. and Murdoch, D.J. (2021) A First Course in Statistical Programming with R. 3rd Edition, Cambridge University Press. <https://doi.org/10.1017/9781108993456>
- [13] Giorgi, F.M., Ceraolo, C. and Mercatelli, D. (2022) The R Language: An Engine for Bioinformatics and Data Science. *Life*, **12**, Article 648. <https://doi.org/10.3390/life12050648>

- [14] Anis, W., Kuntoro, K. and Melaniani, S. (2021) Difference of Power Test and Type II Error (β) on Mardia MVN Test, Henze Zikler's MVN Test, and Royston's MVN Test Using Multivariate Data Analysis. *Jurnal Biometrika dan Kependudukan*, **10**, 153-161. <https://doi.org/10.20473/jbk.v10i2.2021.153-161>
- [15] Khatun, N. (2021) Applications of Normality Test in Statistical Analysis. *Open Journal of Statistics*, **11**, 113-122. <https://doi.org/10.4236/ojs.2021.111006>
- [16] Mardia, K.V. (1980) 9 Tests of Univariate and Multivariate Normality. *Handbook of Statistics*, **1**, 279-320. [https://doi.org/10.1016/s0169-7161\(80\)01011-5](https://doi.org/10.1016/s0169-7161(80)01011-5)
- [17] Rolke, W. and Gongora, C.G. (2020) A Chi-Square Goodness-of-Fit Test for Continuous Distributions against a Known Alternative. *Computational Statistics*, **36**, 1885-1900. <https://doi.org/10.1007/s00180-020-00997-x>
- [18] Chen, T. and Lumley, T. (2019) Numerical Evaluation of Methods Approximating the Distribution of a Large Quadratic Form in Normal Variables. *Computational Statistics & Data Analysis*, **139**, 75-81. <https://doi.org/10.1016/j.csda.2019.05.002>
- [19] Wald, A. and Wolfowitz, J. (1943) An Exact Test for Randomness in the Non-Parametric Case Based on Serial Correlation. *The Annals of Mathematical Statistics*, **14**, 378-388. <https://doi.org/10.1214/aoms/1177731358>
- [20] Signorell, A. (2025) DescTools: Tools for Descriptive Statistics. Version 0.99.6. <https://doi.org/10.32614/CRAN.package.DescTools>
- [21] Ljung, G.M. and Box, G.E.P. (1978) On a Measure of Lack of Fit in Time Series Models. *Biometrika*, **65**, 297-303. <https://doi.org/10.1093/biomet/65.2.297>
- [22] R Core Team (2025) The R Stats Package (Version 4.6.0). <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>
- [23] Chan, K.C.G., Han, J., Kennedy, A.P. and Yam, S.C.P. (2022) Testing Network Autocorrelation without Replicates. *PLOS ONE*, **17**, e0275532. <https://doi.org/10.1371/journal.pone.0275532>
- [24] Schmal, W.B. (2024) Quantitative Tools for Time Series Analysis in Natural Language Processing: A Practitioners Guide. arXiv: 2404.18499. <https://doi.org/10.48550/arXiv.2404.18499>
- [25] Schwert, G.W. (1989) Why Does Stock Market Volatility Change over Time? *The Journal of Finance*, **44**, 1115-1153. <https://doi.org/10.1111/j.1540-6261.1989.tb02647.x>
- [26] Egbert, J. and Plonsky, L. (2020) Bootstrapping Techniques. In: Paquot, M. and Gries, S.T., Eds., *A Practical Handbook of Corpus Linguistics*, Springer, 593-610. https://doi.org/10.1007/978-3-030-46216-1_24
- [27] Rousselet, G., Pernet, C.R. and Wilcox, R.R. (2023) An Introduction to the Bootstrap: A Versatile Method to Make Inferences by Using Data-Driven Simulations. *Meta-Psychology*, **7**, 1-24. <https://doi.org/10.15626/mp.2019.2058>
- [28] Marchev Jr., A. and Marchev, V. (2023) Automated Algorithm for Multi-Variate Data Synthesis with Cholesky Decomposition. *Proceedings of the 7th International Conference on Algorithms, Computing and Systems*, Larissa, 19-21 October 2023, 1-6. <https://doi.org/10.1145/3631908.3631909>
- [29] Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A. and Firth, D. (2025) Support Functions and Datasets for Venables and Ripley's MASS. Version 7.3-65.
- [30] Götze, F. and Tikhomirov, A. (2002) Asymptotic Distribution of Quadratic Forms and Applications. *Journal of Theoretical Probability*, **15**, 423-475. <https://doi.org/10.1023/a:1014867011101>
- [31] Rubia, J.M.D.L. (2024) Rice University Rule to Determine the Number of Bins. *Open*

- Journal of Statistics*, **14**, 119-149. <https://doi.org/10.4236/ojs.2024.141006>
- [32] Gross, J. and Ligges, U. (2022) Package ‘Nortest’, Tests for Normality. Version 1.0-4. <https://cran.r-project.org/web/packages/nortest/nortest.pdf>
- [33] Korkmaz, S. (2025) Package ‘MVN’. Multivariate Normality Tests. <https://cran.r-project.org/web/packages/MVN/MVN.pdf>
- [34] Ferrari, A. (2019) A Note on Sum and Difference of Correlated Chi-Squared Variables. arXiv: 1906.09982. <https://doi.org/10.48550/arXiv.1906.09982>
- [35] Jobst, L.J., Bader, M. and Moshagen, M. (2023) A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models. *Psychological Methods*, **28**, 207-221. <https://doi.org/10.1037/met0000423>
- [36] Henze, N. and Zirkler, B. (1990) A Class of Invariant Consistent Tests for Multivariate Normality. *Communications in Statistics—Theory and Methods*, **19**, 3595-3617. <https://doi.org/10.1080/03610929008830400>

Appendix 1. QSW Test

The chi-square approximation and bootstrap version of the multivariate normality Q-test based on Shapiro–Wilk univariate normality statistics standardized by Royston’s method, for sample sizes ranging from 12 to 2000 with k-tuples of measurements

```
# Load four required R packages
```

```
library(nortest)
```

```
library(MASS)
```

```
library(DescTools)
```

```
library(stats)
```

```
# List of original variables
```

```
# Generate a random sample of 50 observations from a multivariate normal distribution with four correlated variables
```

```
# Load the necessary R package
```

```
library(mvtnorm)
```

```
Parameters
```

```
n <- 50 # Number of observations
```

```
k <- 4 # Number of variables
```

```
rho <- 0.7 # Homogeneous correlation coefficient
```

```
# Create the covariance matrix with rho as off-diagonal values and 1s on the diagonal
```

```
sigma <- matrix(rho, nrow = k, ncol = k)
```

```
diag(sigma) <- 1
```

```
# Set seed for reproducibility
```

```
set.seed(123)
```

```
# Generate multivariate normal data
```

```
mvn_data <- rmvnorm(n = n, sigma = sigma)
```

```
# Scale the values to approximate D-scores
```

```
x1 <- 48 + 10 * round(mvn_data[, 1], 1)
```

```
x2 <- 48 + 10 * round(mvn_data[, 2], 1)
```

```
x3 <- 48 + 10 * round(mvn_data[, 3], 1)
```

```
x4 <- 48 + 10 * round(mvn_data[, 4], 1)
```

```
# Combine into a data frame
```

```
original_data <- data.frame(x1, x2, x3, x4)
```

```

# Display the generated data
print(original_data)

# Sample size, number of variables, and level of significance
n <- length(x1)
cat("\nSample size: n =", n, ".\n")
k <- length(original_data)
cat("Number of variables in the original sample: k =", k, ".\n")
R <- cor(original_data)
m_R <- mean(R[lower.tri(R)])
cat("Arithmetic mean of the Pearson's product-moment correlation coefficients between the variables: m(R) =", round(m_R, 4), ".\n")
alpha <- 0.05

# Generate the sums of all non-repeated combinations of the variables
linear_combinations <- function(df) {
  list_combinations <- list()
  k <- ncol(df)
  # Add the original variables (combinations of size 1)
  for (i in 1:k) {
    list_combinations[[paste0("c", i)]] <- df[, i]
  }
  # Add combinations of size from 2 to k
  counter <- k
  for (i in 2:k) {
    combinations <- combn(k, i)
    for (j in 1:ncol(combinations)) {
      idx <- combinations[, j]
      counter <- counter + 1
      list_combinations[[paste0("c", counter)]] <- rowSums(df[, idx])
    }
  }
  return(list_combinations)
}
list_x <- linear_combinations(original_data)

# Calculation of common parameters for standardizing samples from 12 to 2000
data points
mu <- 0.0038915 * log(n)^3 - 0.083751 * log(n)^2 - 0.31082 * log(n) - 1.5861
sigma <- exp(0.0030302 * log(n)^2 - 0.082676 * log(n) - 0.4803)
cat("\nParameters for standardizing the", 2^k - 1, "variables created by linear
combination of the", k, "original variables:\n")
cat("\nExpected value of ln(1-W):  $\mu$  =", mu, ".\n")
cat("\nStandard deviation of ln(1-W):  $\sigma$  =", sigma, ".\n")

# Apply test and save results to a table
results <- lapply(names(list_x), function(label) {

```

```

data <- list_x[[label]]
result <- shapiro.test(data)
z <- (log(1 - result$statistic) - mu) / sigma
p <- 1 - pnorm(z)
Normality <- if (result$p.value < alpha) "No" else "Yes"
return(c(label, round(result$statistic, 4), round(z, 4), round(p, 4), Normality))
})

# Convert to data frame
table_results <- as.data.frame(do.call(rbind, results), stringsAsFactors =
FALSE)
colnames(table_results) <- c("Combination", "W", "z", "p", "Normality")

# Display the table
cat("\nTabla. Shapiro-Wilk's univariate normality test with Royston's stand-
ardization for the", 2^k-1, "linear combinations of variables\n")
print(table_results, row.names = FALSE)
cat("\nNote. Combination = sum of the corresponding combination without
repetition of variables; W = Shapiro-Wilk W statistic; z = standardized value of
W using Royston's method; p = right-tailed p-value under the standard normal
distribution; Univariate normality: Yes =  $p \geq \alpha$  =", alpha, ", the null hypothesis of
normality is retained; No =  $p < \alpha$  =", alpha, ", the null hypothesis is rejected.\n")

# Function to calculate Q-test statistic (QSW)
calculate_q <- function(list_data) {
z_vals <- sapply(list_data, function(x) {
result_sw <- shapiro.test(x)
z <- (log(1 - result_sw$statistic) - mu) / sigma
return(as.numeric(z))
})
z_values <- ifelse(z_vals < 0, 0, z_vals)
a <- sum(z_vals < 0)
q <- sum(z_values^2)
return(list(q = q, a = a, z_vals = z_vals, z_values = z_values))
}

# Right-tailed p-value and statistical power based on the chi-square distribution
result_q <- calculate_q(list_x)
q_stat <- result_q$q
a <- result_q$a
nc <- length(list_x)
# df_chisq <- nc - a # Correction for truncated z-values
df_chisq <- nc # No correction for truncated z-values

```

```

q_crit <- qchisq(1 - alpha, df = df_chisq)
p_value <- pchisq(q_stat, df = df_chisq, lower.tail = FALSE)
power <- 1 - pchisq(q_crit, df = df_chisq, ncp = q_stat, lower.tail = TRUE)

# Results for QSWa
cat("\nNumber of combinations among the", k, "original variables: nc =", nc,
"\n")
cat("Number of z-values truncated to zero because they are negative: a =", a,
"\n")
cat("Degrees of freedom: df =", df_chisq, "\n")
cat("\nQ-test statistic value in the original sample: q =", round(q_stat, 4), "\n")
cat("\nRight-tailed p-value and critical value based on the chi-square distribu-
tion:\n")
cat("\nCritical value or", 1 - alpha, "quantile of the chi-square distribution with",
df_chisq, "degrees of freedom:  $1-\alpha\chi^2[df]$  =", round(q_crit, 4), "\n")
if (q_stat <= q_crit) {
  cat("The null hypothesis of multivariate normality is retained at a significance
level of", alpha, "based on the chi-square approximation critical value.\n")
} else {
  cat("The null hypothesis of multivariate normality is rejected at a significance
level of", alpha, "based on the chi-square approximation critical value.\n")}
  cat("Right-tailed probability value under the chi-square distribution with",
df_chisq, "degrees of freedom: p =", round(p_value, 4), "\n")
  if (p_value < alpha) {
    cat("The null hypothesis of multivariate normality is rejected at a significance
level of", alpha, "based on the chi-square approximation p-value.\n")
  } else {
    cat("The null hypothesis of multivariate normality is retained at a significance
level of", alpha, " based on the chi-square approximation p-value.\n")}
  cat("Statistical power based on the chi-square approximation at a significance
level of", alpha, ":  $\phi$  =", power, "\n")
  cat("If the null hypothesis is rejected, the statistical power should exceed 0.5,
preferably 0.8 or higher.\n")
  cat("If the null hypothesis is not rejected, the statistical power should be below
0.5, preferably under 0.2.\n")
  cat("Otherwise, the result is contradictory or questionable.\n")

# Test for serial independence in z_w values.
z_w <- result_q$z_values
cat("\nTest of the serial independence of the", nc, "standardized W values (with
negative values truncated to 0) used to compute the Q-test statistic.\n")
RunsTest(z_w, y = NULL, alternative = "two.sided", exact = TRUE, correct =
FALSE)

```

```

lag_max <- min(10, round(nc/5, 0))
# lag_max <- 12 * (n/100)^0.25 # Schwert's rule to determinate the maximum
lag.
ljung_box_test <- lapply(1: lag_max, function(k) Box.test(z_w, lag = k, type =
"Ljung-Box", fitdf = 0))
ljung_box_table <- data.frame(
lag = 1:lag_max,
X_squared = sapply(ljung_box_test, function(x) round(x$statistic, 4)),
df = sapply(ljung_box_test, function(x) round(x$parameter, 4)),
p_value = sapply(ljung_box_test, function(x) round(x$p.value, 4))
)
cat("\nTable: Ljung-Box test for serial independence of z_w values\n")
print(ljung_box_table)
cat("\nNote. Lag = step between values forming the autocorrelation pairs; X-
square = test statistic; df = degrees of freedom; p-value = right-tailed probability
under the chi-square distribution.\n")

# Bootstrap function preserving the correlations of the original sample
resample_preserving_corr <- function(df, B) {
replicate(B, {
df[sample(1:nrow(df), size = nrow(df), replace = TRUE), ]
}, simplify = FALSE)
}

# Function to generate list of combinations
generate_list_boot <- function(df) {
list_boot <- setNames(as.list(df), paste0("c", 1:ncol(df)))
for (i in 2:ncol(df)) {
combinations <- combn(ncol(df), i)
for (j in 1:ncol(combinations)) {
idx <- combinations[, j]
label <- paste0("c", length(list_boot) + 1)
list_boot[[label]] <- rowSums(df[, idx])
}
}
return(list_boot)
}

# Bootstrap simulations preserving correlations
set.seed(123)
B <- 1000
boot_datasets <- resample_preserving_corr(original_data, B)
q_boot_vals <- sapply(boot_datasets, function(df_boot) {

```

```

list_boot <- generate_list_boot(df_boot)
calculate_q(list_boot)$q
})

# Bootstrap p-value (based on the empirical bootstrap distribution)
boot_p_value <- mean(q_boot_vals >= q_stat)
cat("\nEmpirical bootstrap distribution (with preserved correlations) derived
from the original sample:\n")
cat("Number of bootstrap samples: B =", B, "\n")
cat("Right-tailed bootstrap p-value: p_boot =", round(boot_p_value, 4), "\n")
if (boot_p_value < alpha) {
  cat("The null hypothesis of multivariate normality is rejected at the", alpha,
  "significance level based on the bootstrap p-value obtained from the empirical
bootstrap distribution.\n")
} else {
  cat("The null hypothesis of multivariate normality is not rejected at the", alpha,
  "significance level based on the bootstrap p-value obtained from the empirical
bootstrap distribution.\n")
}

# Generate a normative sample using equispaced probabilities and the Cholesky
transformation of the original sample's correlation matrix
p <- seq(0.5 / n, 1 - 0.5 / n, length.out = n)
indep_data <- matrix(NA, nrow = n, ncol = k)
set.seed(123)
for (j in 1:k) {indep_data[, j] <- qnorm(sample(p))}
mat_cor <- cor(original_data)
mat_chol <- chol(mat_cor)
nomative_data <- indep_data %*% mat_chol

# Define the sum-variables of all possible combinations among the original var-
iables
list_x_norm <- linear_combinations(nomative_data)
result_q_norm <- calculate_q(list_x_norm)

# Generate a normative bootstrap distribution using full-row resampling
B <- 1000
q_boot_norm_vals <- replicate(B, {
  indices <- sample(1:n, size = n, replace = TRUE)
  boot_sample <- nomative_data[indices, ]
  list_boot_n <- linear_combinations(boot_sample)
  calculate_q(list_boot_n)$q})

# Bootstrap p-value and critical value (based on the normative bootstrap distri-

```

```

bution)
  boot_norm_p_value <- mean(q_boot_norm_vals >= q_stat)
  # quantile_order <- 1 - alpha # Without adjustment for a heavy right tail
  quantile_order <- 1 - 2 * alpha # With adjustment for a heavy right tail
  boot_norm_critical_value <- quantile(q_boot_norm_vals, probs = quantile_order)
cat("\nNormative bootstrap distribution (with preserved correlation):\n")
cat("Number of bootstrap samples: B =", B, ".\n")
cat("Q-test statistic value in the normative sample: q_mvn_boot =", result_q_norm$q, ".\n")
cat("Mean of the normative bootstrap distribution: m(mvn_boot_dist) =", mean(q_boot_norm_vals), ".\n")
cat("Median of the normative bootstrap distribution: mdn(mvn_boot_dist) =", median(q_boot_norm_vals), ".\n")
p_value_mdn_norm <- 2 * min(mean(q_boot_vals >= median(q_boot_norm_vals)),
  mean(q_boot_vals <= median(q_boot_norm_vals)))
cat("Two-tailed p-value for testing whether the empirical bootstrap distribution is centered on the median of the normative bootstrap distribution: p_2tailed =", round(p_value_mdn_norm, 4), ".\n")
if (p_value_mdn_norm < alpha) {
  cat("The null hypothesis of equal medians between the normative and empirical bootstrap distributions is rejected at a significance level of", 1 - quantile_order, ".\n")
} else {
  cat("The null hypothesis of equal medians between the normative and empirical bootstrap distributions is not rejected at a significance level of", 1 - quantile_order, ".\n")
}
cat("Bootstrap critical value (", quantile_order, " quantile of the normative bootstrap distribution): q_crit =", round(boot_norm_critical_value, 4), ".\n")
if (q_stat <= boot_norm_critical_value) {
  cat("The null hypothesis of multivariate normality is not rejected at a significance level of", 1 - quantile_order, "based on the critical value from the normative bootstrap distribution.\n")
} else {
  cat("The null hypothesis of multivariate normality is rejected at a significance level of", 1 - quantile_order, "based on the critical value from the normative bootstrap distribution.\n")
}
cat("Right-tailed p-value in the normative bootstrap distribution: p_mvn_boot =", round(boot_norm_p_value, 4), ".\n")
if (boot_norm_p_value < 1 - quantile_order) {

```

```

    cat("The null hypothesis of multivariate normality is rejected at a significance
level of", 1 - quantile_order, "based on the bootstrap p-value.\n")
  } else {
    cat("The null hypothesis of multivariate normality is not rejected at a signifi-
cance level of", 1 - quantile_order, "based on the bootstrap p-value.\n")
  }

# Bootstrap power
boot_power <- mean(q_boot_vals > boot_norm_critical_value)
cat("Bootstrap statistical power at a significance level of", 1 - quantile_order, ":",
phi_boot "=", boot_power, ".\n")

# Representation of empirical and normative bootstrap sampling distributions
# Remove the preceding hashtag symbols to save as a JPG or TIFF file
# jpeg("Hist_dens_curves1.jpg", width = 1200, height = 900, units = "px", res =
300)
# tiff("Hist_dens_curves1.tiff", width = 1200, height = 900, units = "px", res =
300)
# par(mar = c(4.5, 4.5, 0.5, 0.5), cex.axis = 0.8)

# Density curve of the chi-square distribution as the base graph
x_seq <- seq(0, max(qchisq(0.999, df = df_chisq), boot_norm_critical_value +
1, q_stat + 1, quantile(q_boot_vals, probs = 0.95, type = 8) + 1), length.out = 1000)
y_chi_sq <- dchisq(x_seq, df = df_chisq)
plot(x_seq, y_chi_sq, type = "l", lwd = 3, col = "darkgreen",
main = "", xlab = "q-values", ylab = "Density",
xlim = c(0, max(qchisq(0.999, df = df_chisq), boot_norm_critical_value + 1,
q_stat + 1, quantile(q_boot_vals, probs = 0.95, type = 8) + 1)),
ylim = c(0, max(density(q_boot_vals)$y, y_chi_sq)))

# Yellow histogram of the empirical bootstrap distribution of the Q-test statistic
hist(q_boot_vals, breaks = 20, freq = FALSE, col = rgb(1, 1, 0, 0.5), border =
"yellow2", add = TRUE)
# Yellow density curve of the empirical bootstrap distribution of the Q-test sta-
tistic
lines(density(q_boot_vals), col = "yellow3", lwd = 2)
# Red density curve of the normative bootstrap distribution of the Q-test statis-
tic
lines(density(q_boot_norm_vals), col = "red", lwd = 2)
# Purple vertical line for the observed Q-test statistic value in the original sam-
ple
abline(v = q_stat, col = "purple", lwd = 2, lty = 2)
# Red vertical line for the normative bootstrap critical value
abline(v = boot_norm_critical_value, col = "red", lwd = 2, lty = 2)

```

```
# Green vertical line for the chi-square approximation critical value
abline(v = qchisq(1 - alpha, df = df_chisq), col = "darkgreen", lwd = 2, lty = 2)
# dev.off() # Remove the preceding hashtag symbol to save the figure
```

```
cat("\nFigure 1. Histogram (yellow) with density curves for the empirical boot-
strap (yellow), normative bootstrap (red), and chi-square (green) distributions.
The value of the Q-test statistic in original sample is indicated by a purple vertical
line (q =", q_stat, "), the critical value of the normative bootstrap distribution by a
red vertical line (" quantile_order, " quantile: q_mvn_boot =", boot_norm_criti-
cal_value, "), and the critical value of the chi-square distribution with", df_chisq,
"degrees of freedom by a green vertical line (" 1 - alpha, " quantile:  $\chi^2$ [" df_chisq,
"] =", q_crit, ").\n")
```

```
cat("\nNote. Yellow histogram (number of bins determined by the Rice rule) =
empirical bootstrap distribution of the Q-test statistic;\n",
"yellow curve = density curve of the empirical bootstrap distribution;\n",
"red curve = density curve of the normative bootstrap distribution (generated
from normally distributed variables with the same correlational structure as the
original data);\n",
"green curve = chi-square distribution with", df_chisq, "degrees of freedom;\n",
"purple vertical line = observed Q-test statistic value in original sample;\n",
"red vertical line = critical value or", quantile_order, " quantile of the normative
bootstrap distribution;\n",
"green vertical line = critical value or", 1 - alpha, " quantile of the chi-square
distribution with", df_chisq, "degrees of freedom.\n")
```

Appendix 2. QSF Test

The chi-square approximation and bootstrap version of the multivariate normality Q-test based on Shapiro–Francia univariate normality statistics standardized by Royston’s method, for sample sizes ranging from 5 to 5000 with k-tuples of measurements

```
# Load four required R packages
library(nortest)
library(MASS)
library(DescTools)
library(stats)

# List of original variables
x1 <- c(47, 52, 46, 55, 47, 33, 41, 43, 63, 50, 44, 52, 55, 56, 38, 46, 41, 54, 55, 42,
50, 52, 53, 50, 67, 41, 32, 48, 37, 41, 44, 63, 45, 44, 39, 44, 30, 67, 47, 53, 60, 46, 53,
49, 49, 39, 49, 53, 48, 44)
x2 <- c(49, 61, 47, 54, 33, 38, 35, 56, 63, 47, 47, 39, 50, 63, 49, 41, 49, 60, 46, 37,
```

```

52, 55, 61, 46, 64, 46, 37, 55, 45, 37, 38, 49, 50, 42, 51, 39, 36, 48, 39, 48, 48, 50, 52,
60, 51, 52, 49, 46, 43, 34)
x3 <- c(59, 54, 56, 50, 48, 33, 49, 51, 62, 46, 41, 43, 52, 55, 47, 42, 50, 46, 46, 45,
48, 59, 61, 57, 54, 43, 33, 47, 48, 36, 41, 54, 59, 33, 58, 31, 31, 60, 44, 56, 47, 52, 49,
45, 51, 43, 56, 50, 40, 37)
x4 <- c(51, 44, 51, 63, 42, 35, 45, 47, 62, 45, 60, 43, 51, 64, 47, 38, 47, 36, 56, 43,
54, 55, 58, 46, 50, 43, 28, 53, 47, 35, 42, 53, 53, 50, 39, 32, 43, 59, 43, 48, 72, 45, 51,
43, 46, 40, 51, 47, 59, 34)
original_data <- data.frame(x1, x2, x3, x4)

# Sample size, number of variables, and level of significance
n <- length(x1)
cat("\nSample size: n =", n, ".\n")
k <- length(original_data)
k <- length(original_data)
cat("Number of variables in the original sample: k =", k, ".\n")
R <- cor(original_data)
m_R <- mean(R[lower.tri(R)])
cat("Arithmetic mean of the Pearson's product-moment correlation coeffi-
cients between the variables: m(R) =", round(m_R, 4), ".\n")
alpha <- 0.05

# Generate the sums of all non-repeated combinations of the variables
linear_combinations <- function(df) {
list_combinations <- list()
k <- ncol(df)
# Add the original variables (combinations of size 1)
for (i in 1:k) {
list_combinations[[paste0("c", i)]] <- df[, i]}
# Add combinations of size from 2 to k
counter <- k
for (i in 2:k) {
combinations <- combn(k, i)
for (j in 1:ncol(combinations)) {
idx <- combinations[, j]
counter <- counter + 1
list_combinations[[paste0("c", counter)]] <- rowSums(df[, idx])}}
return(list_combinations)}
list_x <- linear_combinations(original_data)

# Calculation of common parameters for standardizing samples from 12 to 2000
data points
u <- log(log(n)) - log(n)

```

```

mu <- 1.0521 * u - 1.2725
v <- log(log(n)) + 2 / log(n)
sigma <- -0.26758 * v + 1.0308
cat("\nParameters for standardizing the", 2^k - 1, "variables created by linear
combination of the", k, "original variables:\n")
cat("\nExpected value of ln(1-W'):  $\mu =$ ", mu, ".\n")
cat("\nStandard deviation of ln(1-W'):  $\sigma =$ ", sigma, ".\n")

# Apply test and save results to a table
results <- lapply(names(list_x), function(label) {
  data <- list_x[[label]]
  result <- shapiro.test(data)
  z <- (log(1 - result$statistic) - mu) / sigma
  p <- 1 - pnorm(z)
  Normality <- if (result$p.value < alpha) "No" else "Yes"
  return(c(label, round(result$statistic, 4), round(z, 4), round(p, 4), Normality))
})

# Convert to data frame
table_results <- as.data.frame(do.call(rbind, results), stringsAsFactors =
FALSE)
colnames(table_results) <- c("Combination", "W'", "z", "p", "Normality")

# Display the table
cat("\nTabla. Shapiro-Francia's univariate normality test with Royston's stand-
ardization for the", 2^k-1, "linear combinations of variables\n")
print(table_results, row.names = FALSE)
cat("\nNote. Combination = sum of the corresponding combination without
repetition of variables; W = Shapiro-Francia W' statistic; z = standardized value
of W' using Royston's method; p = right-tailed p-value under the standard normal
distribution; Univariate normality: Yes =  $p \geq \alpha =$ , alpha, ", the null hypothesis of
normality is retained; No =  $p <$ , alpha, ", the null hypothesis is rejected.\n")

# Function to calculate Q2-test statistic (QSF)
calculate_q <- function(list_data) {
  z_vals <- sapply(list_data, function(x) {
    result_sf <- sf.test(x)
    z <- (log(1 - result_sf$statistic) - mu) / sigma
    return(as.numeric(z))
  })
  z_values <- ifelse(z_vals < 0, 0, z_vals)
  a <- sum(z_vals < 0)
  q <- sum(z_values^2)
}

```

```

return(list(q = q, a = a, z_vals = z_vals, z_values = z_values))
}

# Right-tailed p-value and statistical power based on the chi-square distribution
result_q <- calculate_q(list_x)
q_stat <- result_q$q
a <- result_q$a
nc <- length(list_x)
# df_chisq <- nc - a # Correction for truncated z-values
df_chisq <- nc # No correction for truncated z-values
q_crit <- qchisq(1 - alpha, df = df_chisq)
p_value <- pchisq(q_stat, df = df_chisq, lower.tail = FALSE)
power <- 1 - pchisq(q_crit, df = df_chisq, ncp = q_stat, lower.tail = TRUE)

# Results for QSFa
cat("\nNumber of combinations among the", k, "original variables: nc =", nc,
"\n")
cat("Number of z-values truncated to zero because they are negative: a =", a,
"\n")
cat("Degrees of freedom: df =", df_chisq, "\n")
cat("\nQ-test statistic value in the original sample: q =", round(q_stat, 4), "\n")
cat("\nRight-tailed p-value and critical value based on the chi-square distribu-
tion:\n")
cat("\nCritical value or", 1 - alpha, "quantile of the chi-square distribution with",
df_chisq, "degrees of freedom:  $1-\alpha\chi^2[df] =$ ", round(q_crit, 4), "\n")
if (q_stat <= q_crit) {
cat("The null hypothesis of multivariate normality is retained at a significance
level of", alpha, "based on the chi-square approximation critical value.\n")
} else {
cat("The null hypothesis of multivariate normality is rejected at a significance
level of", alpha, "based on the chi-square approximation critical value.\n")}
cat("Right-tailed probability value under the chi-square distribution with",
df_chisq, "degrees of freedom: p =", round(p_value, 4), "\n")
if (p_value < alpha) {
cat("The null hypothesis of multivariate normality is rejected at a significance
level of", alpha, "based on the chi-square approximation p-value.\n")
} else {
cat("The null hypothesis of multivariate normality is retained at a significance
level of", alpha, " based on the chi-square approximation p-value.\n")}
cat("Statistical power based on the chi-square approximation at a significance
level of", alpha, ":  $\varphi =$ ", power, "\n")
cat("If the null hypothesis is rejected, the statistical power should exceed 0.5,
preferably 0.8 or higher.\n")

```

```

cat("If the null hypothesis is not rejected, the statistical power should be below
0.5, preferably under 0.2.\n")
cat("Otherwise, the result is contradictory or questionable.\n")

# Test for serial independence in z_w' values.
z_w <- result_q$z_values
cat("\nTest for the serial independence of the", nc, "standardized W' values
(truncated to 0 when negative) used to compute the Q'-test statistic.\n")
RunsTest(z_w, y = NULL, alternative = "two.sided", exact = TRUE, correct =
FALSE)
lag_max <- min(10, round(nc/5, 0))
# lag_max <- 12 * (n/100)^0.25 # Schwert's rule to determinate the maximum
lag.
ljung_box_test <- lapply(1: lag_max, function(k) Box.test(z_w, lag = k, type =
"Ljung-Box", fitdf = 0))
ljung_box_table <- data.frame(
Lag = 1:lag_max,
X_squared = sapply(ljung_box_test, function(x) round(x$statistic, 4)),
df = sapply(ljung_box_test, function(x) round(x$parameter, 4)),
p_value = sapply(ljung_box_test, function(x) round(x$p.value, 4))
)
cat("\nTable: Ljung-Box test for serial independence of z_w' values\n")
print(ljung_box_table)
cat("\nNote. Lag = step between values forming the autocorrelation pairs; X-
square = test statistic; df = degrees of freedom; p-value = right-tailed probability
under the chi-square distribution.\n")

# Bootstrap function preserving the correlations of the original sample
resample_preserving_corr <- function(df, B) {
replicate(B, {
df[sample(1:nrow(df), size = nrow(df), replace = TRUE), ]
}, simplify = FALSE)
}

# Function to generate list of combinations
generate_list_boot <- function(df) {
list_boot <- setNames(as.list(df), paste0("c", 1:ncol(df)))
for (i in 2:ncol(df)) {
combinations <- combn(ncol(df), i)
for (j in 1:ncol(combinations)) {
idx <- combinations[, j]
label <- paste0("c", length(list_boot) + 1)
list_boot[[label]] <- rowSums(df[, idx])
}
}
}

```

```

}
}
return(list_boot)
}

# Bootstrap simulations preserving correlations
set.seed(123)
B <- 1000
boot_datasets <- resample_preserving_corr(original_data, B)

q_boot_vals <- sapply(boot_datasets, function(df_boot) {
list_boot <- generate_list_boot(df_boot)
calculate_q(list_boot)$q
})

# Bootstrap p-value (based on the empirical bootstrap distribution)
boot_p_value <- mean(q_boot_vals >= q_stat)
cat("\nEmpirical bootstrap distribution (with preserved correlations) derived
from the original sample:\n")
cat("Number of bootstrap samples: B =", B, "\n")
cat("Right-tailed bootstrap p-value: p_boot =", round(boot_p_value, 4), "\n")
if (boot_p_value < alpha) {
cat("The null hypothesis of multivariate normality is rejected at the", alpha,
"significance level based on the bootstrap p-value obtained from the empirical
bootstrap distribution.\n")
} else {
cat("The null hypothesis of multivariate normality is not rejected at the", alpha,
"significance level based on the bootstrap p-value obtained from the empirical
bootstrap distribution.\n")
}

# Generate a normative sample using equispaced probabilities and the Cholesky
transformation of the original sample's correlation matrix
p <- seq(0.5 / n, 1 - 0.5 / n, length.out = n)
indep_data <- matrix(NA, nrow = n, ncol = k)
set.seed(123)
for (j in 1:k) {indep_data[, j] <- qnorm(sample(p))}
mat_cor <- cor(original_data)
mat_chol <- chol(mat_cor)
nomative_data <- indep_data %*% mat_chol

# Define the sum-variables of all possible combinations among the original var-
iables

```

```

list_x_norm <- linear_combinations(nomative_data)
result_q_norm <- calculate_q(list_x_norm)

# Generate a normative bootstrap distribution using full-row resampling
B <- 1000
q_boot_norm_vals <- replicate(B, {
  indices <- sample(1:n, size = n, replace = TRUE)
  boot_sample <- nomative_data[indices, ]
  list_boot_n <- linear_combinations(boot_sample)
  calculate_q(list_boot_n)$q
})

# Bootstrap p-value and critical value (based on the normative bootstrap distribution)
boot_norm_p_value <- mean(q_boot_norm_vals >= q_stat)
# quantile_order <- 1 - alpha # Without adjustment for a heavy right tail
quantile_order <- 1 - 2 * alpha # With adjustment for a heavy right tail
boot_norm_critical_value <- quantile(q_boot_norm_vals, probs = quantile_order)

cat("\nNormative bootstrap distribution (with preserved correlation):\n")
cat("Number of bootstrap samples: B =", B, "\n")
cat("Q2-test statistic value in the normative sample: q_mvn_boot =", result_q_norm$q, "\n")
cat("Mean of the normative bootstrap distribution: m(mvn_boot_dist) =", mean(q_boot_norm_vals), "\n")
cat("Median of the normative bootstrap distribution: mdn(mvn_boot_dist) =", median(q_boot_norm_vals), "\n")
p_value_mdn_norm <- 2 * min(mean(q_boot_vals >= median(q_boot_norm_vals)),
  mean(q_boot_vals <= median(q_boot_norm_vals)))
cat("Two-tailed p-value for testing whether the empirical bootstrap distribution is centered on the median of the normative bootstrap distribution: p_2tailed =", round(p_value_mdn_norm, 4), "\n")
if (p_value_mdn_norm < alpha) {
  cat("The null hypothesis of equal medians between the normative and empirical bootstrap distributions is rejected at a significance level of", 1 - quantile_order, "\n")
} else {
  cat("The null hypothesis of equal medians between the normative and empirical bootstrap distributions is not rejected at a significance level of", 1 - quantile_order, "\n")
}
cat("Bootstrap critical value (", quantile_order, " quantile of the normative bootstrap distribution): q_crit =", round(boot_norm_critical_value, 4), "\n")

```

```

    if (q_stat <= boot_norm_critical_value) {
      cat("The null hypothesis of multivariate normality is not rejected at a significance level of", 1 - quantile_order, "based on the critical value from the normative bootstrap distribution.\n")
    } else {
      cat("The null hypothesis of multivariate normality is rejected at a significance level of", 1 - quantile_order, "based on the critical value from the normative bootstrap distribution.\n")
    }
    cat("Right-tailed p-value in the normative bootstrap distribution: p_mvn_boot =", round(boot_norm_p_value, 4), ".\n")
    if (boot_norm_p_value < 1 - quantile_order) {
      cat("The null hypothesis of multivariate normality is rejected at a significance level of", 1 - quantile_order, "based on the bootstrap p-value.\n")
    } else {
      cat("The null hypothesis of multivariate normality is not rejected at a significance level of", 1 - quantile_order, "based on the bootstrap p-value.\n")
    }
  }

  # Bootstrap power
  boot_power <- mean(q_boot_vals > boot_norm_critical_value)
  cat("Bootstrap statistical power at a significance level of", 1 - quantile_order, ": phi_boot =", boot_power, ".\n")

  # Representation of empirical and normative bootstrap sampling distributions
  # Remove the preceding hashtag symbols to save as a JPG or TIFF file
  # jpeg("Hist_dens_curves2.jpg", width = 1200, height = 900, units = "px", res = 300)
  # tiff("Hist_dens_curves2.tiff", width = 1200, height = 900, units = "px", res = 300)
  # par(mar = c(4.5, 4.5, 0.5, 0.5), cex.axis = 0.8)

  # Density curve of the chi-square distribution as the base graph
  x_seq <- seq(0, max(qchisq(0.999, df = df_chisq), boot_norm_critical_value + 1, q_stat + 1, quantile(q_boot_vals, probs = 0.95, type = 8) + 1), length.out = 1000)
  y_chi_sq <- dchisq(x_seq, df = df_chisq)
  plot(x_seq, y_chi_sq, type = "l", lwd = 3, col = "darkgreen",
       main = "", xlab = "q_prime_values", ylab = "Density",
       xlim = c(0, max(qchisq(0.999, df = df_chisq), boot_norm_critical_value + 1, q_stat + 1, quantile(q_boot_vals, probs = 0.95, type = 8) + 1)),
       ylim = c(0, max(density(q_boot_vals)$y, y_chi_sq)))
  # Yellow histogram of the empirical bootstrap distribution of the Q'-test statistic

```

```

hist(q_boot_vals, breaks = 20, freq = FALSE, col = rgb(1, 1, 0, 0.5), border =
"yellow2", add = TRUE)
# Yellow density curve of the empirical bootstrap distribution of the Q'-test sta-
tistic
lines(density(q_boot_vals), col = "yellow3", lwd = 2)
# Red density curve of the normative bootstrap distribution of the Q'-test sta-
tistic
lines(density(q_boot_norm_vals), col = "red", lwd = 2)
# Purple vertical line for the observed Q'-test statistic value in the original sam-
ple
abline(v = q_stat, col = "purple", lwd = 2, lty = 2)
# Red vertical line for the normative bootstrap critical value
abline(v = boot_norm_critical_value, col = "red", lwd = 2, lty = 2)
# Green vertical line for the chi-square approximation critical value
abline(v = qchisq(1-alpha, df = df_chisq), col = "darkgreen", lwd = 2, lty = 2)
# dev.off() # Remove the preceding hashtag symbol to save the figure

```

cat("\nFigure 2. Histogram (yellow) with density curves for the empirical boot-strap (yellow), normative bootstrap (red), and chi-square (green) distributions. The value of the Q-test statistic in original sample is indicated by a purple vertical line (q = ", q_stat, "), the critical value of the normative bootstrap distribution by a red vertical line (" ", quantile_order, " quantile: q_mvn_boot =", boot_norm_criti- cal_value, "), and the critical value of the chi-square distribution with", df_chisq, "degrees of freedom by a green vertical line (" ", 1 - alpha, " quantile: χ^2 [" ", df_chisq, "] =", q_crit, ").\n")

Appendix 3. Royston's H-test

Royston's multivariate normality H-test for samples size from 10 to 2000 with k-tuples of measurements

```

# Load required R package
library(MVN)
# List of original variables
x1 <- c(47, 52, 46, 55, 47, 33, 41, 43, 63, 50, 44, 52, 55, 56, 38, 46, 41, 54, 55, 42,
50, 52, 53, 50, 67, 41, 32, 48, 37, 41, 44, 63, 45, 44, 39, 44, 30, 67, 47, 53, 60, 46, 53,
49, 49, 39, 49, 53, 48, 44)
x2 <- c(49, 61, 47, 54, 33, 38, 35, 56, 63, 47, 47, 39, 50, 63, 49, 41, 49, 60, 46, 37,
52, 55, 61, 46, 64, 46, 37, 55, 45, 37, 38, 49, 50, 42, 51, 39, 36, 48, 39, 48, 48, 50, 52,
60, 51, 52, 49, 46, 43, 34)
x3 <- c(59, 54, 56, 50, 48, 33, 49, 51, 62, 46, 41, 43, 52, 55, 47, 42, 50, 46, 46, 45,
48, 59, 61, 57, 54, 43, 33, 47, 48, 36, 41, 54, 59, 33, 58, 31, 31, 60, 44, 56, 47, 52, 49,
45, 51, 43, 56, 50, 40, 37)
x4 <- c(51, 44, 51, 63, 42, 35, 45, 47, 62, 45, 60, 43, 51, 64, 47, 38, 47, 36, 56, 43,

```

```

54, 55, 58, 46, 50, 43, 28, 53, 47, 35, 42, 53, 53, 50, 39, 32, 43, 59, 43, 48, 72, 45, 51,
43, 46, 40, 51, 47, 59, 34)
original_data <- data.frame(x1, x2, x3, x4)

n <- length(x1)
k <- length(original_data)
alpha <- 0.05
result <- mvn(original_data, subset = NULL, mvnTest = "royston")

# Statistical power for H-test
R <- cor(original_data)
lambda <- 5
mu <- 0.715
ln_n <- log(n)
v_n <- 0.21364 + 0.015124 * ln_n^2 - 0.0018034 * ln_n^3
r_ij <- R[lower.tri(R)]
c_ij <- r_ij^lambda * (1 - (mu / v_n) * (1 - r_ij))^mu
c_bar <- mean(c_ij)
df_H <- k / (1 + (k - 1) * c_bar)
power_H <- 1 - pchisq(qchisq(1-alpha, df = df_H), df = df_H, ncp = re-
sult$multivariateNormality$H, lower.tail = TRUE)

# Display results
cat("\nSample size: n =", n, ".\n")
cat("Number of variables in the original sample: k =", k, ".\n")
cat("Arithmetic mean of the Pearson's product-moment correlation coeffi-
cients between the variables: m(R) =", round(mean(r_ij), 4), ".\n")
cat("\nRoyston's multivariate normality H-test\n")
cat("H statistic: H =", round(result$multivariateNormality$H, 4), ".\n")
cat("Degrees of freedom: df =", df_H, ".\n")
cat("p-value for the null hypothesis of multivariate normality: p =", result$mul-
tivariateNormality$p, ".\n")
cat("Statistical power of the H-test at a significance level of", alpha, ":  $\phi$  =",
power_H, ".\n")
cat("If the null hypothesis is rejected, the statistical power should exceed 0.5,
preferably 0.8 or higher.\n")
cat("If the null hypothesis is not rejected, the statistical power should be below
0.5, preferably under 0.2.\n")
cat("Otherwise, the result is contradictory or questionable.\n")

```