

Bootstrap Implementation of Mardia's K^2 Test and Performance Comparison with Royston's H Test

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. (2026) Bootstrap Implementation of Mardia's K^2 Test and Performance Comparison with Royston's H Test. *Journal of Data Analysis and Information Processing*, 14, 83-118.
<https://doi.org/10.4236/jdaip.2026.141006>

Received: January 7, 2026

Accepted: February 2, 2026

Published: February 5, 2026

Copyright © 2026 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

This article has two objectives. The first is to develop an R script that performs Mardia's K^2 test for assessing multivariate normality (MVN), using both its original asymptotic formulation and a bootstrap method. Unlike previous studies that rely on Monte Carlo simulations to obtain critical values, the bootstrap approach focuses on resampling from sample data while preserving its correlational structure. The second objective is to compare both variants of K^2 test with Royston's H test in terms of hit rate and statistical power. A total of 200 samples were generated from MVN distributions and another 200 from multivariate t distributions with five degrees of freedom, varying sample size (7 levels), number of variables (5 levels), and homogeneous correlation among variables (5 levels). The script computes the K^2 and H test statistics, their p -values, and statistical power, and also provides graphical representations of the sampling distribution of the K^2 statistic. Two worked examples, one based on a randomly generated sample from an MVN distribution and another based on an empirical sample without an MVN distribution, demonstrate the script's functionality. The sampling distribution of the bootstrap K^2 statistic is less peaked and exhibits a heavier right tail than the distribution implied by the asymptotic approximation, which is defined by a weighted sum of correlated chi-squared variables. When the null hypothesis was retained (*i.e.*, under multivariate normality), the H test showed slightly better performance in terms of correct retention. In contrast, when the null hypothesis was violated, both variants of Mardia's K^2 test outperformed the H test; in this context, the asymptotic variant generally performed better than the bootstrap variant. In samples drawn from a multivariate t distribution, as well as in analyses aggregated across both distributions, the asymptotic K^2 variant achieved the highest mean power. In multivariate normal samples, however, the bootstrap K^2 variant showed superior performance, followed by the asymptotic variant, with

the H test performing worst. This new variant is a robust option for assessing MVN, and its use and further exploration are recommended.

Keywords

Multivariate Normal Distribution, Statistical Hypothesis Testing, Statistical Power, Multivariate Skewness, Multivariate Kurtosis

1. Introduction

In 1970, Mardia [1] introduced two measures for assessing the shape of multivariate distributions: one for skewness (B_1) and another for kurtosis (B_2). He also developed corresponding significance tests for each measure (Q and Z, respectively). In 1974, he proposed a small-sample correction for the significance test (Q_c) associated with the B_1 skewness measure [2]. Later, in 1980, he introduced an omnibus test ($K^2 = Q + Z^2$), which combines both measures to evaluate multivariate normality [3]. Years afterward, Urzua [4] revisited the omnibus K^2 statistic and computed critical values for sample sizes of 10, 20, 50, 100, 200, and ∞ using Monte Carlo simulation. Koizumi *et al.* [5] also explored this topic through Monte Carlo simulation.

Mardia [1] [2] demonstrated that the asymptotic sampling distribution of the test statistics Q and Q_c follows a chi-square distribution with degrees of freedom given by $k(k+1)(k+2)/6$, while the asymptotic distribution of the Z statistic follows a standard normal distribution. Since these two statistics are asymptotically independent, the asymptotic distribution of the omnibus statistic K^2 is a chi-square distribution with degrees of freedom $1 + k(k+1)(k+2)/6$.

The asymptotic approximation improves with larger sample sizes (e.g., $n \geq 1000$), fewer variables (typically 2 - 4), and low inter-variable correlations ($\rho \leq 0.30$). Conversely, in small samples with many highly correlated variables, the chi-square approximation becomes inadequate, and the distribution of K^2 more closely resembles a generalized chi-square distribution [6]. An approximation to this generalized distribution can be obtained using the Bootstrap method [7].

The departure of K^2 from a standard chi-square distribution arises because the skewness and kurtosis components of Mardia's test are not independent in finite samples, especially when the variables are strongly correlated. In such conditions, the joint variability of Q (skewness) and Z (kurtosis) is affected by their non-negligible covariance, producing a weighted sum of correlated quadratic forms rather than the independent sum assumed asymptotically. This dependence structure inflates the variance of K^2 and produces heavier right tails, making its empirical distribution better described by a generalized chi-square variable. As sample size increases or correlations weaken, the covariance between Q and Z diminishes, and the distribution of K^2 converges toward its nominal chi-square form.

The first objective of this study is to develop an R script [8] for calculating the

omnibus K^2 test using both the asymptotic approach (with a chi-square distribution and $1 + k(k+1)(k+2)/6$ degrees of freedom, where k denotes the number of variables assumed to follow a multivariate normal distribution) and the bootstrap method. This allows for the estimation of critical values, p -values, and statistical power without relying on the limited critical value tables available in the literature [4] [5]. The script also computes asymptotic measures and significance tests for multivariate skewness (B_1 and Q) and multivariate kurtosis (B_2 and Z). The main script is complemented by an auxiliary script that calculates critical values, p -values, and statistical power using the well-known MVN H-test [9]. The implementation of these scripts is illustrated with two examples. The first uses a dataset of 61 observations comprising 3-tuples that represent the D scores from aptitude tests administered to applicants for a master's program in architecture. The second example uses real-world data collected with four 7-point Likert-type scales. The first dataset is based on a randomly generated sample from an MVN distribution, whereas the second relies on an empirical sample without an MVN distribution. R was chosen because it is one of the most comprehensive statistical software environments, offering free and open access. Moreover, it is continuously developed and reviewed by the statistical and mathematical community [10] [11].

The second objective of this statistical-methodological study is to compare the performance of the asymptotic and bootstrap variants of the K^2 test with Royston's H test in terms of hit rate and statistical power. To this end, 200 random samples are generated from multivariate normal distributions (MVN) and 200 from multivariate Student's t-distributions (MVT) with five degrees of freedom, varying the sample size (50, 75, 100, 125, 150, 200, 250, and 500), the number of variables (2, 3, 4, 5, and 6), and the homogeneous correlation among variables (0, 0.3, 0.5, 0.7, and 0.9). The study was limited to two distributions for the sake of simplicity: one in which the null hypothesis should be retained (MVN) and one in which it should be rejected (MVT). The multivariate t-distribution with five degrees of freedom was chosen as the primary alternative to normality because its pronounced leptokurtosis, resulting from heavy tails, represents a common and substantial deviation from the multivariate normal distribution. For simplicity, the study focused on only alternative for multivariate non-normality. The three sample configuration variables (sample size, number of variables, and homogeneous correlation) are known determinants of statistical power in multivariate tests [12]. Royston's H test was selected as the comparison benchmark because it is one of the most widely used and powerful tests for assessing multivariate normality [13] [14].

2. Method

2.1. For the First Objective

The R script consists of two sections, each included in a different appendix, and is available as a Word file (Appendices_with_R_Scripts.docx) in a GitHub repository:

https://github.com/josemoraldelarubia579/R_script-

[for MVN K2 and H tests.git](#).

Appendix 1 presents the script for computing the multivariate skewness coefficient (B_1) and its asymptotic significance test (Q), the multivariate kurtosis coefficient (B_2) and its asymptotic significance test (Z), as well as the K^2 statistic and its significance tests using both asymptotic and bootstrap approaches. The asymptotic significance of the K^2 statistic represents the probability of obtaining a value greater than or equal to the observed one under a chi-square distribution with $1 + k(k + 1)(k + 2)/6$ degrees of freedom.

For the bootstrap approach, both the normative bootstrap distribution and the empirical bootstrap distribution are generated. Let us begin with the normative bootstrap distribution. Given a sample of size n composed of k correlated variables, a normative sample of the same size and dimension is generated to follow a multivariate normal distribution with the same correlational structure. This process begins by creating a square matrix of k independent variables with standard normal distribution, each consisting of n observations. These values are obtained using the probit (quantile) function of the standard normal distribution. First, equally spaced quantile orders are generated, ranging from $0.5/n$ to $1 - 0.5/n$. The quantile orders are then randomly permuted for each variable using a fixed seed (123) to ensure reproducibility [8]. In this way, each of the k variables shares the same set of quantile orders in a different sequence, forming mutually independent vectors.

The resulting matrix of independent Gaussian vectors is multiplied by the lower triangular matrix from the Cholesky decomposition of the original sample's correlation matrix [15], yielding a normative sample with the same correlational structure as the original. From this sample, 1000 samples are randomly drawn with replacement by rows—not by columns—in order to preserve the correlational structure. The K^2 statistic is computed for each resample. These 1000 values constitute the normative bootstrap distribution. In this distribution, which replaces the theoretical chi-square distribution with $1 + k(k + 1)(k + 2)/6$ degrees of freedom, the probability of observing the original sample's K^2 statistic is calculated using the right tail. This yields the bootstrap p -value. The 95th percentile of the distribution serves as the normative bootstrap critical value, which is then used to estimate statistical power.

In contrast, the empirical bootstrap distribution is generated by drawing with replacement 1000 samples directly from the original sample, again using row-wise resampling to preserve the correlational structure. The K^2 statistic is calculated for each resample, and these 1000 values form the empirical bootstrap distribution. Using the normative bootstrap critical value (*i.e.*, the 95th percentile), the proportion of resamples in which the null hypothesis of multivariate normality is rejected (*i.e.*, when the K^2 estimate exceeds the critical value) is calculated. This proportion represents the bootstrap estimate of statistical power at the 5% significance level. If the null hypothesis of multivariate normality holds, power should be below 0.5, preferably under 0.2. Conversely, if the null hypothesis is false, power should ex-

ceed 0.5, preferably above 0.8. Otherwise, the result may be considered ambiguous or questionable.

Finally, the script enables visualization of the sampling distribution of the K^2 statistic through a figure that displays several components. The empirical bootstrap distribution is presented as a histogram with an overlaid smooth density curve, estimated using kernel density estimation via the `density()` function in R's base Stats package. By default, this function employs a Gaussian kernel and determines the bandwidth with the `nrd0` method, an adaptation of Silverman's rule of thumb (Silverman, 1986). The normative bootstrap and chi-square distributions (with $1 + k(k+1)(k+2)/6$ degrees of freedom) are represented by smooth density curves computed in the same way. In addition, the observed value of the K^2 statistic and the critical values from both the normative bootstrap and chi-square distributions are indicated with vertical lines.

Appendix 2 provides the R script for computing Royston's MVN H-test. Using an asymptotic approach, the script calculates both the p -value and statistical power at a 5% significance level. This test is included as a comparative or complementary method for assessing multivariate normality. The aim is to evaluate the consistency of results across the asymptotic and bootstrap variants of the K^2 test and the H test in terms of retaining or rejecting the null hypothesis of multivariate normality.

The operation of the script is illustrated using two examples. The first example is executed with data generated from a multivariate normal distribution; the code used to generate these data is provided in Script 1 (Appendix 1), and the resulting sample is shown in Script 2 (Appendix 2). The second example is executed using the non-normal data provided in Appendix 3.

2.2. For the Second Objective

A total of 200 samples—each consisting of 50, 75, 100, 125, 150, 200, 250, or 500 observations and 2, 3, 4, 5, or 6 variables with homogeneous correlations of 0, 0.3, 0.5, 0.7, or 0.9—were drawn from a multivariate t-distribution with five degrees of freedom (MVTD). An additional set of 200 samples with identical characteristics was drawn from a multivariate normal distribution (MVND). For both sets, two variants of the K^2 test (K^2_a and K^2_b), along with Royston's H test, were computed. The R package `mvtnorm` was used to generate the 400 samples [16]. A single random seed (123) was set at the beginning of each simulation condition to ensure reproducibility [8]. The random number generator state was allowed to advance naturally across successive data generations, so that the 200 samples correspond to independent realizations rather than identical repetitions. The correlation matrices were specified using a compound symmetry (equicorrelation) structure, with unit variances and a common off-diagonal correlation coefficient ρ . Further details can be found in Appendix 4, which is also accessible in the GitHub repository:

https://github.com/josemoraldealarubia579/R_script-for_MVN_K2_and_H_tests.git.

Since the samples were generated using the same seed, the hit rate and statistical power data for the three multivariate normality tests represent repeated measurements. Therefore, statistical tests appropriate for repeated measures were applied.

Confidence intervals for the hit rates were calculated using the Wilson method with continuity correction [17]. Differences in hit rates among the five multivariate normality tests were assessed using Cochran's Q test [18]. Effect size was measured with the eta-squared coefficient [19]. Pairwise comparisons of hit rates were conducted using the paired z-test with an estimated common standard error [20]. Bonferroni correction was applied to control the familywise error rate [21]. Effect size for pairwise comparisons was calculated using a z-based analogue of Cohen's d: $d = z / \sqrt{n}$ [20]. The resulting d values were interpreted according to Cohen's thresholds: [0, 0.2) = trivial, [0.2, 0.5) = small, [0.5, 0.8) = medium, and ≥ 0.8 = large (Cohen, 1988) [22]. All omnibus and pairwise comparisons were performed using SPSS version 27 [23].

The comparison of mean statistical power (based on the 200 samples from the MVT distribution) and its complement (based on the 200 samples from the MVN distribution) was conducted using repeated measures ANOVA. Sample size, number of variables, and homogeneous correlation among variables were included as covariates in the general linear model (GLM), with distribution type (MVT vs. MVN) as a between-subjects factor. The assumption of sphericity—*i.e.*, homogeneity of variances of the differences between all possible pairs of repeated measures—was tested using Mauchly's test [24]. When this assumption was violated, a multivariate approach to within-subject effects was employed using Pillai's trace [25], as recommended for repeated measures ANOVA [26]. Effect sizes for each model component were estimated using partial eta-squared coefficients.

Pairwise comparisons were performed using the paired-samples Student's t-test with Bonferroni correction. Cohen's d [22], with the Hedges–Olkin correction [27], was used to quantify the effect size. The 95% confidence intervals for the mean statistical power of the three MVN tests, as well as the correlations between statistical power and the three cofactors (in the GLM), were computed using the bias-corrected and accelerated (BCa) percentile method based on 1000 bootstrap samples [7]. The reported correlations correspond to Pearson product-moment coefficients. These analyses were executed using SPSS version 27 [23].

In addition to the violation of the sphericity assumption, the assumption of multivariate normality in the repeated measures component (MVN tests), assessed using Royston's H test [9], and the assumption of univariate normality in the distribution of residuals for each repeated measure, assessed using Shapiro–Wilk's W test [28] [29], were not satisfied. Therefore, a nonparametric repeated measures ANOVA [30] was also applied. Pairwise comparisons were conducted using the rank-sum test [31]. Effect size in the omnibus test was estimated using Kendall's W coefficient [32], and in pairwise comparisons using the rank-biserial correlation [33]. The JASP program version 0.19.2 [34] was used to perform these power comparisons.

3. Results

3.1. Example with Multivariate Normality

Three psychometric aptitude scales—assessing mechanical, spatial, and numerical reasoning—were administered to 61 applicants for a Master of Architecture degree. It is assumed that the sample composed of 61 observations of 3-uples follows a multivariate normal distribution. This assumption is tested using both the asymptotic and bootstrap variants of the Mardia's K^2 test, as well as Royston's H test.

The sample was generated using an R script provided in Appendix 1, which also includes the code for running the asymptotic and bootstrap versions of K^2 test. Appendix 2 presents the three scoring vectors and the code for performing H test.

At a 0.05 significance level, the multivariate distribution—based on 61 observations of 3-tuples—shows symmetry and mesokurtosis, as confirmed by Mardia's tests. Both versions of the K^2 test support the assumption of multivariate normality, although the bootstrap variant exhibits lower power ($\phi = 0.016 < 0.2$) compared to the asymptotic one ($\phi = 0.2516 < 0.5$). Royston's H test also supports multivariate normality, with low power ($\phi = 0.1216 < 0.2$). Notably, when the null hypothesis is not rejected, statistical power should fall below 0.5—ideally under 0.2. Otherwise, the result is contradictory or questionable.

Figure 1 shows that the three sampling distributions associated with the K^2 test—namely, the asymptotic (chi-square) distribution in green, the normative bootstrap distribution in red, and the empirical bootstrap distribution in yellow—are similar in overall shape. Among them, the normative bootstrap distribution is the least peaked and exhibits the greatest positive skewness. This divergence indicates that the correlation between multivariate skewness and kurtosis in the empirical data inflates the variability of the K^2 statistic, causing it to behave more like a generalized chi-square variable with substantial tail weight than like the nominal chi-square distribution.

Sample size: $n = 61$.

Number of variables in the original sample: $k = 3$.

Significance level: $\alpha = 0.05$.

Average of the correlation among variables: $m(R_{3 \times 3}) = 0.6846$.

Mardia's multivariate skewness and kurtosis tests

Mardia's multivariate skewness statistic: $B_1 = 4.5598$.

Two-tailed p -value for the null hypothesis of multivariate symmetry: $p = 0.9186$.

Mardia's multivariate kurtosis statistic: $B_2 = -0.5925$.

Two-tailed p -value for the null hypothesis of multivariate mesokurtosis: $p = 0.5535$.

Asymptotic variant of the MVN K-square test

Mardia's multivariate normality statistic: MVN K-square = 4.9108.

Degrees of freedom: $df = 11$

Critical value: $0.95\chi^2 [11] = 19.6751$

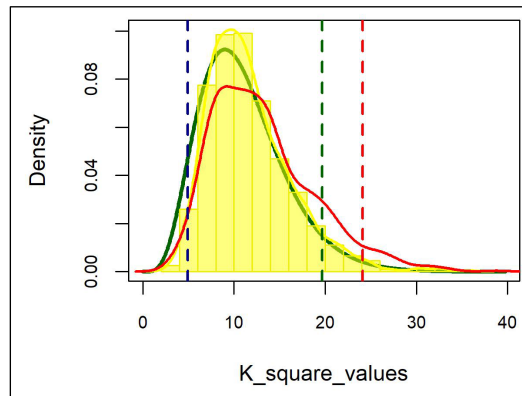


Figure 1. Histogram (yellow) with density curves for the empirical bootstrap (yellow), normative bootstrap (red), and chi-square (green) distributions. The value of the k-square statistic is indicated by a blue vertical line ($K^2 = 4.9108$), the critical value of the normative bootstrap distribution by a red vertical line (0.95 quantile: $0.95q_{boot} = 24.0928$), and the critical value of the chi-square distribution with 11 degrees of freedom by a green vertical line ($0.95\chi^2 [11] = 19.6751$).

Right-tailed p -value under a chi-square distribution with 11 degrees of freedom:

$$p = 0.9354.$$

H_0 (multivariate normality) is maintained at the 0.05 level based on the asymptotic p -value (chi-2 distribution).

Asymptotic statistical power for the K-square statistic: $\phi = 0.2516 < 0.5$.

If the null hypothesis is rejected, statistical power should exceed 0.5—preferably 0.8 or higher. If the null hypothesis is not rejected, statistical power should be below 0.5—preferably under 0.2. Otherwise, the result is contradictory.

Bootstrap variant of the MVN K-square test

Bootstrap critical value (0.95 quantile): $0.95q_{boot} = 24.0928$.

Bootstrap p -value: $p_{boot} = 0.9790$.

H_0 (multivariate normality) is maintained at the 0.05 level based on the bootstrap p -value.

Bootstrap power: $\phi = 0.016 < 0.2$.

Royston's multivariate normality H test

H statistic: $H = 1.7955$.

Degrees of freedom: $df = 2.6596$.

p -value for the null hypothesis of multivariate normality: $p = 0.5498$.

Statistical power of the H test at a significance level of 0.05: $\phi = 0.1844 < 0.2$.

3.2. Example without Multivariate Normality

Next, an illustrative example is presented using real-world data with characteristics typical of psychology studies. A non-probability sampling method was employed, resulting in an incidental sample of 509 participants, including 300 women (59%) and 209 men (41%). The questionnaire was administered digitally and began with a request for informed consent. The Subjective Happiness Scale

[35], adapted for Mexico and validated in the general Mexican population by Quezada *et al.* [36], was used as the measurement instrument. Appendix 3 provides the list of variables used in this example, which is available in the GitHub repository at

[https://github.com/josemoraldearubia579/R_script-](https://github.com/josemoraldearubia579/R_script-for_MVN_K2_and_H_tests.git)

[for MVN K2 and H tests.git](https://github.com/josemoraldearubia579/R_script-for_MVN_K2_and_H_tests.git). It consists of four Likert-type items, that is, ordinal variables with seven ordered categories. The data frame only needs to be modified to include these four variables —`original_data <- data.frame(x1, x2, x3, x4)`— and no further changes are required in the scripts presented in Appendices 1 and 2.

The results clearly reject the null hypothesis of multivariate normality, both for the K^2 test under its asymptotic and bootstrap approaches, and for Royston's H test under its asymptotic or standard approach, which is the only one considered. **Figure 2** displays the histogram with density curves corresponding to the empirical bootstrap, the normative bootstrap, and the chi-square distributions for the K^2 test.

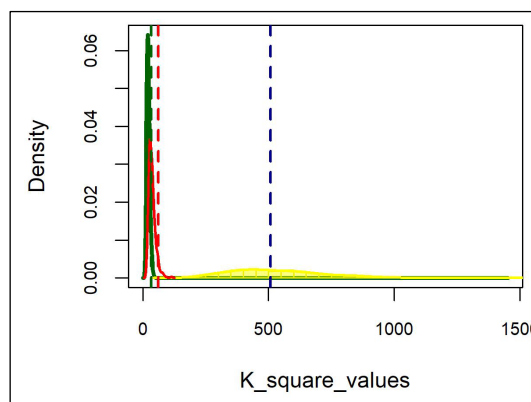


Figure 2. Histogram (yellow) with density curves for the empirical bootstrap (yellow), normative bootstrap (red), and chi-square (green) distributions. The value of the k-square statistic is indicated by a blue vertical line ($k\text{-square} = 507.174$), the critical value of the normative bootstrap distribution by a red vertical line ($0.95q_{\text{boot}} = 62.1757$), and the critical value of the chi-square distribution with 21 degrees of freedom by a green vertical line ($0.95\chi^2[21] = 32.6706$).

Compared to **Figure 1**, which shows the sampling distribution of the K^2 statistic with three variables following a multivariate normal distribution, the empirical distribution is much more leptokurtic and positively skewed. This arises because the four items were negatively asymmetric, and two of them were leptokurtic.

A further aspect worth noting is the marked difference in scale between **Figure 1** and **Figure 2**. In **Figure 1**, the sampling distribution of the K^2 statistic is concentrated within a narrow range, and the empirical, normative-bootstrap, and chi-square curves overlap closely. By contrast, in **Figure 2** the empirical K^2 values extend over a much wider range, with extreme values far exceeding those expected under multivariate normality. This stretching of the horizontal axis reflects the

strong departure from normality, while the clear separation between the three density curves shows that neither the chi-square approximation nor the normative bootstrap can capture the heavy-tailed behaviour of the empirical bootstrap.

Sample size: $n = 509$.

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

Significance level: $\alpha = 0.05$.

Average of the correlation among variables: $m(R) = 0.4281$.

Mardia's multivariate skewness and kurtosis tests

Mardia's multivariate skewness statistic: $B1 = 261.5066$.

Two-tailed p -value for the null hypothesis of multivariate symmetry: $p = 0$.

Mardia's multivariate kurtosis statistic: $B2 = 15.6738$.

Two-tailed p -value for the null hypothesis of multivariate mesokurtosis: $p = 0$.

Asymptotic variant of the MVN K-square test

Mardia's multivariate normality statistic: MVN K-square = 507.174.

Degrees of freedom: $df = 21$.

Critical value: $0.95_chi\text{-sq}[21] = 32.6706$.

Right-tailed p -value under a chi-square distribution with 21 degrees of freedom:

$p_value = 0$.

H_0 (multivariate normality) is rejected at the level 0.05 based on the asymptotic p -value (chi-2 distribution).

Asymptotic statistical power for the K-square statistic: $\phi = 1$.

If the null hypothesis is rejected, the statistical power should exceed 0.5, preferably 0.8 or higher.

If the null hypothesis is not rejected, the statistical power should be below 0.5, preferably under 0.2.

Otherwise, the result is contradictory or questionable.

Bootstrap variant of the MVN K-square test

Bootstrap critical value (0.95 quantile): $q_boot(0.95) = 62.1757$.

Bootstrap p -value: $p_boot = 0$.

H_0 (multivariate normality) is rejected at the level 0.05 based on the bootstrap p -value.

Bootstrap power: $\phi_boot = 1$.

Royston's multivariate normality H test:

H statistic: $H = 281.542$.

Degrees of freedom: $df = 3.9544$.

p -value for the null hypothesis of multivariate normality: $p < 0.0001$.

Statistical power of the H test at a significance level of 0.05: $\phi = 1$.

3.3. Hit Rate Comparisons

In the 200 samples drawn from the multivariate t distribution, the hit rate (HR) of the H test was the lowest, and its confidence interval lay entirely below those of the two variants of the Q test, whose intervals overlapped (**Table 1** and **Figure 3**).

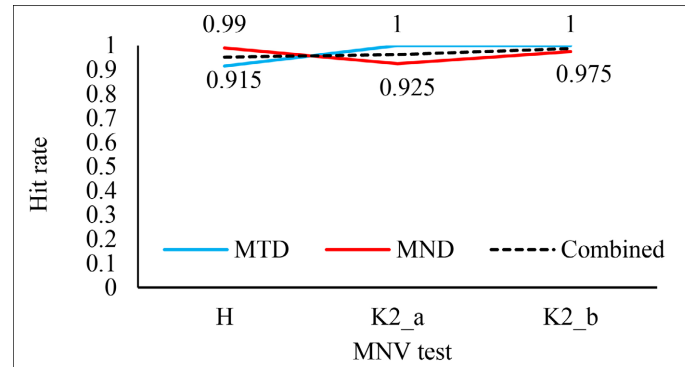


Figure 3. Diagram of the hit rates of the three MVN tests based on 200 samples drawn from multivariate t-distributions, 200 samples drawn from multivariate normal distributions, and a combined sample of 400 observations.

Table 1. Hit rates and 95% continuity-corrected Wilson confidence intervals based on 200 samples from multivariate t distributions, 200 samples from multivariate normal distributions, and a combined sample of 400 observations.

MVN tests	MTD	MND	Combined
	$HR(LL, UL)$	$HR(LL, UL)$	$HR(LL, UL)$
H	0.915 (0.865, 0.948)	0.990 (0.961, 0.998)	0.953 (0.926, 0.970)
K ² _a	1 (0.977, 1)	0.925 (0.877, 0.956)	0.963 (0.938, 0.978)
K ² _b	1 (0.977, 1)	0.975 (0.939, 0.991)	0.988 (0.969, 0.995)

Note. Distribution: MTD = 200 samples drawn from multivariate t distribution, MND = 200 samples drawn from multivariate normal distribution, and combined = 400 samples drawn MTV or MND. MVN test: H = Royston's H test, K²_a = asymptotic variant of the Q test, and K²_b = bootstrap variant of Q test. Statistic: HR = hit rate, LL = lower limit and UL = upper limit of a 95% continuity-corrected Wilson confidence interval.

When comparing HRs among the three MVN tests, a significant difference was found using Cochran's Q test ($X^2[2, N = 200] = 34, p < 0.001$), with a medium effect size ($0.06 < \eta^2 = 0.085 < 0.14$). According to Bonferroni-adjusted paired z tests, the HRs of both Q test variants were significantly higher than that of the H test, although the effect sizes were small (**Table 2**).

In the combined sample, the HRs of the three MVN tests overlapped (**Table 1** and **Figure 3**). However, when comparing HRs among the three MVN tests, a significant difference was found ($X^2[2, N = 400] = 9.176, p < 0.010$), with a small effect size ($0.01 < \eta^2 = 0.011 < 0.06$). After applying Bonferroni's correction, only one significant difference remained: the HR of the bootstrap variant of the Q test was significantly higher than that of the H test, although the effect size was trivial (**Table 2**).

The HR of the H test was significantly higher for samples drawn from multivariate normal distributions than from multivariate t-distributions ($d = -0.075, se = 0.021, z = -3.526, p < 0.001$). In contrast, the HRs of both the asymptotic variant ($d = 0.075, se = 0.019, z = 3.948, p < 0.001$) and the bootstrap variant ($d = 0.025, se = 0.011, z = 2.250, p = 0.024$) of the K² test were significantly higher for

samples drawn from MVT than from MVN distributions.

Table 2. Pairwise comparisons of hit rates using Bonferroni-adjusted paired z-tests on 200 samples from multivariate t-distributions, 200 samples from multivariate normal distributions, and a combined sample of 400 observations.

Distribution	Test1 - Test2	<i>md</i>	<i>se</i>	<i>z</i>	<i>p</i>	<i>p</i> _{Bonf}	<i>d</i>
MTD	K ² a - H	0.085	0.017	5.050	<0.001	<0.001	0.357
	K ² b - H	0.085	0.017	5.050	<0.001	<0.001	0.357
	K ² a - K ² b	0	0.017	0	1	1	0
MND	K ² a - H	-0.065	0.017	-3.862	<0.001	<0.001	-0.273
	K ² b - H	-0.015	0.017	-0.891	0.373	1	-0.063
	K ² a - K ² b	-0.050	0.017	-2.970	0.003	0.009	-0.210
Combined	K ² a - H	0.010	0.012	0.84	0.401	1	0.042
	K ² b - H	0.035	0.012	2.941	0.003	0.010	0.147
	K ² a - K ² b	-0.025	0.012	-2.100	0.036	0.107	-0.105

Note. *md* = mean difference in hit rate between MVN test 1 and MVN test 2, *se* = standard error, *z* = paired z statistic, *p* = two-tailed *p*-value, *p*_{Bonf} = Bonferroni-adjusted two-tailed *p*-value, *d* = Cohen’s d (effect size measure).

3.4. Mean Statistical Power Comparison

3.4.1. In Samples Drawn from MVT Distributions

In the 200 samples drawn from the MVT distributions, the asymptotic variant of the K² test exhibited the highest mean statistical power, while the H test showed the lowest. However, the 95% BCa confidence intervals for the mean power overlapped across the three MVN tests (Table 3 and Figure 4).

Table 3. Mean statistical powers and 95% BCa confidence intervals based on 200 samples from multivariate t distributions, 200 samples from multivariate normal distributions, and a combined sample of 400 observations.

MVN tests	MTD	MND	Combined
H	0.950 (0.906, 1)	0.732 (0.709, 0.755)	0.841 (0.814, 0.872)
K ² _a	0.994 (0.989, 0.998)	0.765 (0.731, 0.798)	0.880 (0.860, 0.898)
K ² _b	0.963 (0.951, 0.974)	0.766 (0.735, 0.799)	0.864 (0.844, 0.885)

Note. Point estimates and 95% BCa confidence intervals for the average statistical power (its complement for 200 samples drawn from multivariate normal distributions) based on 1000 simulations.

The sphericity assumption was strongly violated (Mauchly $W = 0.075$, $X^2[2, N = 200] = 504.182$, $p < 0.001$); therefore, a Huynh-Feldt epsilon correction ($\epsilon = 0.528 < 0.7$) was applied to adjust the degrees of freedom in the analysis of within-subject effects. Additionally, a multivariate test was included, as it is more appropriate when the sphericity assumption is not met.

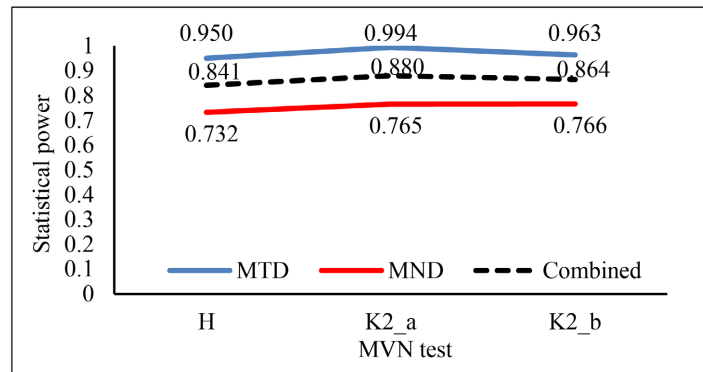


Figure 4. Diagram of the mean statistical power (its complement of the statistical power for 200 samples drawn from multivariate normal distributions) of the three MVN tests.

The effect of the MVN test type was not significant ($F[1.056, 206.928] = 1.591$, $p = 0.209$, $\eta^2 = 0.008$, $\phi = 0.247$), nor were its interactions with sample size ($F[1.056, 206.928] = 1.4$, $p = 0.240$, $\eta^2 = 0.007$, $\phi = 0.223$), the number of variables ($F[1.056, 206.928] = 1.424$, $p = 0.236$, $\eta^2 = 0.007$, $\phi = 0.226$), or the homogeneous correlation among variables ($F[1.056, 206.928] = 1.881$, $p = 0.171$, $\eta^2 < 0.010$, $\phi = 0.283$).

In contrast to these results, the multivariate test of within-subject effects revealed a significant main effect of the MVN test type (Pillai's $\Lambda = 0.158$; $F[2, 195] = 18.316$, $p < 0.001$; $\eta^2 = 0.158$; $\phi = 1$), as were its significant interactions with two of the three cofactors (Pillai's $\Lambda = 0.072$; $F[2, 195] = 7.610$, $p = 0.001$; $\eta^2 = 0.072$; $\phi = 0.944$ with sample size and Pillai's $\Lambda = 0.062$; $F[2, 195] = 6.465$, $p = 0.002$; $\eta^2 = 0.062$; $\phi = 0.902$ with number of variables). The effect size of the MVN test type on power was large, and the effect sizes of the interactions were medium. However, the interaction between the MVN test type and the homogeneous correlation among variables was not significant (Pillai's $\Lambda = 0.010$; $F[2, 195] = 0.964$, $p = 0.383$; $\eta^2 = 0.010$; $\phi = 0.216$).

Based on Pearson's product-moment correlation, with confidence intervals obtained using the BCa method, the statistical power of the three MVN tests was independent of the homogeneous correlation among variables. However, power was positively correlated with sample size. Additionally, the power of both K^2 test variants was positively correlated with the number of variables. The strength of association was low for the H test and the asymptotic variant of the K^2 test, and moderate for the bootstrap variant (see **Table 4**).

In the pairwise comparisons, only one of the three differences was significant both before and after applying the Bonferroni correction. The asymptotic variant of the K^2 test showed significantly higher mean power than its bootstrap variant, with a small effect size (Hedges-Olkin $g = 0.455$, 95% asymptotic *CI*: 0.310, 0.600). See **Table 5**.

Apart from the failure to comply with the assumption of sphericity, the assumption of multivariate normality across the three repeated measures was not met (Royston's test: $H = 342.06$, $df = 3.013$, $p < 0.001$, $\phi = 1$). Additionally, univariate

Table 4. Correlations between the mean statistical powers of the three MVN tests and the three cofactors based on 200 samples drawn from multivariate t distributions.

MVN tests	<i>n</i>	<i>k</i>	<i>r</i>
H	0.109* (0.028, 0.333)	0.114 (-0.060, 0.213)	-0.095 (-0.303, -0.033)
K ² _a	0.196* (0.132, 0.249)	0.252* (0.160, 0.327)	-3.71 × 10 ⁻¹⁷ (-0.128, 0.150)
K ² _b	0.300* (0.238, 0.370)	0.302* (0.225, 0.381)	2.06 × 10 ⁻¹⁷ (-0.135, 0.138)

Note. ϕ = average statistical power (its complement for 200 samples drawn from multivariate normal distributions); *LL* = lower limit, and *UL* = upper limit of the 95% BCa confidence interval based on 1000 simulations.

Table 5. Post hoc pairwise comparisons of mean statistical power between MVN tests (Bonferroni-adjusted) based on 200 samples from multivariate t distributions.

Test1 - Test2	<i>md</i>	<i>se</i>	<i>z</i>	<i>p</i>	<i>p</i> _{Bonf}
H - K ² a	-0.044 (-0.113, 0.024)	0.028	-1.564	0.119	0.358
H - K ² b	-0.013 (-0.080, 0.054)	0.028	-0.474	0.636	1
K ² a - K ² b	0.031 (0.020, 0.042)	0.005	6.835	<0.001	<0.001

Note. *md* = mean difference in statistical power between Test 1 and Test 2 (*LL* = lower limit, and *UL* = upper limit, of the 95% asymptotic confidence interval adjusted using Bonferroni correction to control the Type I error rate for an alpha level of 0.05); *se* = asymptotic standard error; *t*[*df*] = paired t-test statistic [degrees of freedom = 196]; *p* = two-tailed asymptotic *p*-value; *p*_{Bonf} = two-tailed Bonferroni-adjusted asymptotic *p*-value.

normality was violated in the residuals for each dependent variable (Shapiro-Wilk *W* = 0.312, *p* < 0.001 for H; *W* = 0.472, *p* < 0.001 for K²_a; *W* = 0.724, *p* < 0.001 for K²_b). Consequently, the result of the analysis of variance should be interpreted with caution.

The Friedman test revealed a statistically significant difference in the central tendency of statistical power among the three MVN tests ($X^2[2, N = 200] = 174.162, p < 0.001$), with a medium effect size for the MVN test type on statistical power (Kendall *W* = 0.435).

Table 6. Pairwise comparisons of statistical power using the Conover rank-sum test (Bonferroni-adjusted) based on 200 samples from multivariate t distributions.

T1 - T2	<i>w_i</i>	<i>w_j</i>	<i>t</i>	<i>p</i>	<i>p</i> _{Bonf}	<i>r</i> _{tb}
H - K ² a	334	528.5	15.308	<0.001	<0.001	0.964
H - K ² b	334	337.5	0.275	0.783	1	0.011
K ² a - K ² b	528.5	337.5	15.033	<0.001	<0.001	1

Note. *w_i* = sum of ranks for Test 1, *w_j* = sum of ranks for Test 2, |*t*| = absolute value of the t-test statistic (degrees of freedom: *df* = 398), *p* = asymptotic two-tailed *p*-value, *p*_{Bonf} = Bonferroni-adjusted *p*-value for an alpha level of 0.05, |*r*_{tb}| = absolute value of the rank-biserial correlation, used as a measure of effect size.

Pairwise comparisons using the Conover rank sum test indicated that two of

the three differences were statistically significant, even after applying the Bonferroni correction, with a large effect size based on biserial rank correlation. The central tendency of statistical power for the asymptotic variant of the K^2 test was significantly higher than that of both its bootstrap variant and the H test (Table 6).

3.4.2. In Samples Drawn from MVN Distributions

In the 200 samples drawn from the MVN distribution, the two variants of the K^2 test exhibited the lowest mean statistical power ($HR = 0.235$, 95% BCa CI: 0.202, 0.266 for asymptotic variant and $HR = 0.234$, 95% BCa CI: 0.203, 0.265), while H test showed the highest ($HR = 0.268$, 95% BCa CI: 0.244, 0.290). Since the null hypothesis should be retained in these 200 tests, the closer the statistical power is to 0, the better the test performs. However, the 95% BCa confidence intervals for the mean statistical power of the three MVT tests overlapped.

The sphericity assumption was violated (Mauchly $W = 0.869$, $X^2 [2, N = 200] = 27.456$, $p < 0.001$); therefore, a Huynh-Feldt epsilon correction ($\epsilon = 0.884 > 0.7$) was applied to adjust the degrees of freedom in the analysis of within-subject effects. Additionally, the multivariate test was included.

The effect of the MVN type of test was significant ($F[1.768, 346.492] = 34.591$, $p < 0.001$, $\eta^2 = 0.150$, $\phi = 1$), as were its interactions with sample size ($F[1.768, 346.492] = 11.54$, $p < 0.001$, $\eta^2 = 0.056$, $\phi = 0.988$) and the number of variables ($F[1.768, 346.492] = 94.702$, $p < 0.001$, $\eta^2 = 0.326$, $\phi = 1$). The effect size was large for the main effect and the interaction with the number of variables, while it was small for the interaction with sample size. However, the interaction between the MVN type of test and the homogeneous correlation among variables was not significant ($F[1.768, 346.492] = 1.843$, $p = 0.165$, $\eta^2 = 0.009$, $\phi = 0.360$).

The multivariate test for the analysis of within-subject effects yielded very similar results. The effect of the MVN test type was significant (Pillai's $\Lambda = 0.350$; $F[2, 195] = 52.423$, $p < 0.001$; $\eta^2 = 0.350$; $\phi = 1$) as were its interactions with sample size (Pillai's $\Lambda = 0.091$; $F[2, 195] = 9.706$, $p < 0.001$; $\eta^2 = 0.091$; $\phi = 0.981$) and number of variables (Pillai's $\Lambda = 0.588$; $F[2, 195] = 139.404$, $p < 0.001$; $\eta^2 = 0.588$; $\phi = 1$). The effect size was large for both the main effect and the interaction with number of variables, while it was medium for the interaction with sample size. However, the interaction between the MVN test type and the homogeneous correlation among variables was not significant at the 5% level of significance, though it was significant at the 10% level (Λ of Pillai = 0.029; $F[2, 394] = 2.871$, $p = 0.059$; $\eta^2 = 0.029$; $\phi = 0.557$).

Based on Pearson's product-moment correlations, with confidence intervals obtained using the BCa method, the statistical power of the two variants of the Q test was independent of the homogeneous correlation among variables. This cofactor was only negatively correlated with the power of the H test, and the strength of this association was small. Sample size was negatively correlated with the power of both the H test and the bootstrap variant of the K^2 test, and positively correlated with the power of the asymptotic variant. In all three cases, the strength of association was small. The number of variables was negatively correlated with the power

of both K^2 test variants, showing a strong association with the asymptotic variant and a moderate association with the bootstrap variant; in contrast, it was positively correlated with the power of the H test, with a moderate strength of association (Table 7).

Table 7. Correlations between the mean statistical powers of the three MVN tests and the three cofactors based on 200 samples drawn from multivariate normal distributions.

MVN tests	n	k	r
H	-0.188* (-0.289, -0.076)	0.375* (0.257, 0.483)	-0.184* (-0.327, -0.026)
K^2_a	0.146* (-0.028, 0.328)	-0.795* (-0.829, -0.760)	0 (-0.162, 0.146)
K^2_b	-0.221* (-0.309, -0.121)	-0.412* (-0.515, -0.307)	-9.16×10^{-18} (-0.145, 0.148)

Note. n = sample size, k = number of variables, and r = homogeneous correlation among variables. *An asterisk superscript indicates statistical significance at the 0.05 confidence level, as the confidence interval obtained via the BCa method (based on 1000 simulations) includes zero.

In the pairwise comparisons, two significant differences were observed before applying the Bonferroni correction. Both variants of the K^2 test showed lower mean statistical power than the H test. However, the effect sizes were trivial and not statistically significant (Hedges-Olkin $g = 0.098$, 95% asymptotic CI: -0.041, 0.237 for H - K^2_a ; and $g = 0.120$, 95% asymptotic CI: -0.019, 0.259 for H - K^2_b). After applying the Bonferroni correction, neither of these two differences remained statistically significant (Table 8).

Table 8. Post hoc pairwise comparisons of mean statistical power between MVN tests (Bonferroni-adjusted) based on 200 samples from multivariate normal distributions.

T1 - T2	md	se	t[196]	p	p_{Bonf}
H - K^2_a	0.033 (-0.005, 0.072)	0.016	2.096	0.037	0.112
H - K^2_b	0.034 (-0.007, 0.074)	0.017	2.022	0.045	0.134
K^2_a - K^2_b	0.0003 (-0.050, 0.051)	0.021	0.014	0.989	1

Note. md = mean difference in statistical power between Test 1 and Test 2 (LL = lower limit, and UL = upper limit, of the 95% asymptotic confidence interval adjusted using Bonferroni correction to control the Type I error rate for an alpha level of 0.05); se = asymptotic standard error; t[df] = paired t-test statistic [degrees of freedom = 196]; p = two-tailed asymptotic p -value; p_{Bonf} = two-tailed Bonferroni-adjusted asymptotic p -value.

Apart from the violation of the sphericity assumption, the assumption of multivariate normality across the three repeated measures was not met (Royston's test: $H = 132.84$, $df = 2.939$, $p < 0.001$, $\phi = 1$). Additionally, univariate normality was violated in the residuals for each dependent variable (Shapiro-Wilk: $W = 0.966$, $p < 0.001$ for H; $W = 0.939$, $p < 0.001$ for K^2_a ; $W = 0.960$, $p < 0.001$ for K^2_b). Although these deviations are smaller than those observed in the general linear

model based on 200 samples drawn from multivariate t-distributions, the results of the analysis of variance should still be interpreted with caution.

The Friedman test revealed a statistically significant difference in the central tendency of statistical power among the three MVN tests ($X^2[2, N=200] = 275.16$, $p < 0.001$), with a large effect size of test type on power (Kendall $W = 0.688$).

In pairwise comparisons using the Conover rank sum test, all three differences in central tendency were statistically significant, even after applying the Bonferroni correction. Based on the rank-biserial correlation, the effect size was large in both comparisons involving the H test, and small in the comparison between the two K^2 test variants. The H test had the highest sum of ranks (**Table 9**).

Table 9. Pairwise comparisons of statistical power using the Conover rank-sum test (Bonferroni-adjusted) based on 200 samples from multivariate normal distributions.

T1 - T2	w_i	w_j	$ t $	p	p_{Bonf}	$ r_{\text{rb}} $
H - K^2 a	590	326	23.569	<0.001	<0.001	0.995
H - K^2 b	590	284	27.318	<0.001	<0.001	0.996
K^2 a - K^2 b	326	284	3.750	<0.001	<0.001	0.174

Note. w_i = sum of ranks for Test 1, w_j = sum of ranks for Test 2, $|t|$ = absolute value of the t-test statistic (degrees of freedom: $df = 398$), p = asymptotic two-tailed p -value, p_{Bonf} = Bonferroni-adjusted p -value for an alpha level of 0.05, $|r_{\text{rb}}|$ = absolute value of the rank-biserial correlation, used as a measure of effect size.

Table 10. Critical values for the K^2 test at the 10% and 5% significance levels.

Sample		Critical value at 10%			Critical value at 5%		
k	n	Urzua	Boot	Chi-2	Urzua	Boot	Chi-2
2	50	7.83	6.06	9.24	11.57	7.09	11.07
	100	7.98	8.10		11.20	9.91	
	200	7.88	11.14		10.94	13.29	
3	50	11.35	22.02	17.28	16.62	24.14	19.68
	100	11.12	27.16		15.24	30.15	
	200	11.01	22.41		14.37	24.94	
4	50	14.81	35.25	29.62	20.43	38.05	32.67
	100	14.69	45.19		19.16	49.16	
	200	14.00	36.37		18.21	39.46	
5	50	18.06	61.87	47.21	23.95	65.52	51
	100	17.60	71.90		22.91	76.47	
	200	17.00	70.94		21.25	78.03	11.07

Note. k = number of variables; n = sample size. Urzua = critical values for the K^2 test at the 10% and 5% significance levels, based on Urzúa (1997, p. 352); Boot = bootstrap critical value (*i.e.*, the 0.90 or 0.95 quantile of the normative bootstrap distribution); Chi-2 = 0.90 or 0.95 quantile of the chi-squared distribution with $1 + k(k+1)(k+2)/6$ degrees of freedom.

The 10% and 5% critical values reported by Urzua (1997, p. 352) [3], based on Monte Carlo simulations, were lower than the corresponding critical values from both the chi-square distribution with $1 + k(k+1)(k+2)/6$ degrees of freedom and those generated by the script used in this study (from the normative bootstrap distribution). The latter were the highest.

The average absolute difference between Urzua's critical values and those from the chi-square distribution was 12.31, while the difference with the critical values obtained by the script was 10.01 (Table 10).

3.4.3. In the Combined Sample of 400 Observations

In the combined sample, the asymptotic variant of the K^2 test exhibited the highest mean power, while the H test showed the lowest. However, the 95% BCa confidence intervals for the mean statistical power overlapped among the three tests (Table 3 and Figure 4).

The sphericity assumption was violated (Mauchly $W = 0.758$, $X^2[2, N = 400] = 109.11$, $p < 0.001$); therefore, a Greenhouse-Geisser epsilon factor ($\epsilon = 0.805 > 0.7$) was applied to adjust the degrees of freedom in the analysis of within-subject effects and the multivariate test was included.

The effect of the MVN test type was significant ($F[1.61, 636.127] = 6.652$, $p = 0.003$, $\eta^2 = 0.017$, $\phi = 0.864$), as were its interactions with sample size ($F[1.61, 636.127] = 7.062$, $p = 0.002$, $\eta^2 = 0.018$, $\phi = 0.884$) and number of variables ($F[1.61, 636.127] = 22.406$, $p < 0.001$, $\eta^2 = 0.054$, $\phi = 1$). In all three cases, the effect size was small. However, the interaction between MVN test type and homogeneous correlation among variables was not significant ($F[1.61, 636.127] = 0.055$, $p = 0.914$, $\eta^2 < 0.001$, $\phi = 0.058$), nor was the interaction with distribution type ($F[1.61, 636.127] = 0.526$, $p = 0.552$, $\eta^2 = 0.001$, $\phi = 0.128$).

The multivariate test for the analysis of within-subject effects yielded the same result. The effect of the MVN test type was significant (Pillai's $\Lambda = 0.022$; $F[2, 394] = 4.479$, $p = 0.012$; $\eta^2 = 0.022$; $\phi = 0.765$) as were its interactions with sample size (Pillai's $\Lambda = 0.052$; $F[2, 394] = 10.700$, $p < 0.001$; $\eta^2 = 0.052$; $\phi = 0.99$) and number of variables (Pillai's $\Lambda = 0.082$; $F[2, 394] = 17.639$, $p < 0.001$; $\eta^2 = 0.082$; $\phi = 1$). All three significant effects were small. The other two interactions involving the MVN test type were not significant: neither with homogeneous correlation among variables (Pillai's $\Lambda = 0$; $F[2, 394] = 0.037$, $p = 0.963$; $\eta^2 = 0$; $\phi = 0.056$), nor with distribution type (Pillai's $\Lambda = 0.005$; $F[2, 394] = 1.026$, $p = 0.359$; $\eta^2 = 0.005$; $\phi = 0.229$).

Based on Pearson's product-moment correlation, with confidence intervals obtained using the BCa method, the statistical power of the three tests was independent of the homogeneous correlation among variables. The statistical power of the H test correlated positively with sample size, with a small strength of association ($r = 0.114$, 95% BCa CI: 0.034, 0.270). The statistical power of the asymptotic variant of the K^2 test correlated positively with the number of variables, with a moderate strength of association. The statistical power of the bootstrap variant of the K^2 test correlated positively with both sample size and number of variables, with

a small strength of association in each case (Table 11).

Table 11. Correlations between the statistical powers of the three MVN tests and the three cofactors based on a combined sample of 400 observations.

MVN tests	n	k	r
H	0.114* (0.034, 0.270)	-0.022 (-0.189, 0.065)	-0.014 (-0.110, 0.063)
K ² _a	-0.069 (-0.204, 0.045)	0.477* (0.420, 0.539)	4.55×10^{-16} (-0.104, 0.101)
K ² _b	0.187*(0.122, 0.245)	0.296* (0.214, 0.387)	4.65×10^{-18} (-0.093, 0.094)

Note. n = sample size, k = number of variables, and r = homogeneous correlation among variables. * An asterisk superscript indicates statistical significance at the 0.05 confidence level, as the confidence interval obtained via the BCa method (based on 1000 simulations) includes zero.

In the pairwise comparisons, only one of the three differences was statistically significant before applying the Bonferroni correction. The asymptotic approximation of the K² test showed significantly higher mean power than the H test, with a trivial effect size (Hedges-Olkin $g = -0.104$, 95% asymptotic CI: -0.202, -0.006). However, after applying the Bonferroni correction, none of the differences remained statistically significant (see Table 12).

Table 12. Post hoc pairwise comparisons of mean statistical power between MVN tests (Bonferroni-adjusted) based on a combined sample of 400 observations.

T1 - T2	md	se	$t[196]$	p	p_{Bonf}
H - K ² a	-0.039 (-0.081, 0.004)	0.018	-2.182	395	0.030
H - K ² b	-0.023 (-0.064, 0.017)	0.017	-1.388	395	0.166
K ² a - K ² b	0.015 (-0.011, 0.042)	0.011	1.382	395	0.168

Note. md = mean difference in statistical power (its complement for 200 samples drawn from multivariate normal distributions) between Test 1 and Test 2 (LL = lower limit, and UL = upper limit, of the 95% asymptotic confidence interval adjusted using Bonferroni correction to control the Type I error rate for an alpha level of 0.05); se = asymptotic standard error; $t[df]$ = paired t-test statistic [degrees of freedom = 395]; p = two-tailed asymptotic p -value; p_{Bonf} = two-tailed Bonferroni-adjusted asymptotic p -value.

It should be noted that the assumption of homogeneity across the matrices of observed covariances of the dependent variables was not met (Box $M = 987.872$, $F[6, 1147681.811] = 163.3$, $p < 0.001$), nor was the assumption of homogeneity of variance between the residuals of two dependent variables ($F[1, 398] = 30.373$, $p < 0.001$ for K²_a and $F[1, 398] = 109.609$, $p < 0.001$ for K²_b). However, this assumption was met for the H test ($F[1, 398] = 0.044$, $p = 0.833$). Additionally, the assumption of multivariate normality was violated across the three repeated measures (Royston's test: $H = 368.72$, $df = 3.003$, $p < 0.001$, $\phi = 1$), as well as the univariate normality in the residuals for each dependent variable (Shapiro-Wilk $W = 0.377$, $p < 0.001$ for H; $W = 0.905$, $p < 0.001$ for K²_a; $W = 0.947$, $p < 0.001$

for K^2_b). Consequently, the results of the analysis of variance should be interpreted with caution.

The Friedman test revealed a statistically significant difference in the central tendency of statistical power among the three MVN tests ($X^2[2, N=400] = 84.241$, $p < 0.001$), with a small effect size of test type on statistical power (Kendall $W = 0.105$).

In the pairwise comparisons using the Conover rank sum test, all three differences in central tendency were significant, even after applying the Bonferroni correction. The effect size was medium for the comparison between the asymptotic variant of the K^2 test and the H test, and small for the other two comparisons, based on rank-biserial correlation (**Table 13**).

Table 13. Pairwise comparisons of statistical power using the Conover rank-sum test (Bonferroni-adjusted) based on a combined sample of 400 observations.

T1 - T2	w_i	w_j	$ t $	p	p_{Bonf}	$ r_{\text{rb}} $
H - K^2_a	692.5	929.5	9.568	<0.001	<0.001	0.388
H - K^2_b	692.5	778	3.452	<0.001	0.002	0.165
K^2_a - K^2_b	929.5	778	6.116	<0.001	<0.001	0.155

Note. w_i = sum of ranks for Test 1, w_j = sum of ranks for Test 2, $|t|$ = absolute value of the t-test statistic (degrees of freedom: $df = 798$), p = asymptotic two-tailed p -value, p_{Bonf} = Bonferroni-adjusted p -value for an alpha level of 0.05, $|r_{\text{rb}}|$ = absolute value of the rank-biserial correlation, used as a measure of effect size.

4. Discussion

The R program successfully met the study's first objective: automating the computation of Mardia's K^2 test using both the original asymptotic approach [3] and a novel bootstrap method, while also incorporating Royston's H test as a complementary tool for assessing multivariate normality. Running the R script on a dataset of 61 D-scores from three correlated variables generated under an MVN distribution, as well as on an empirical dataset of 509 Likert-type scores from four correlated variables that do not follow an MVN distribution, demonstrated how straightforward it is to interpret the results based on the p -value, the critical value, and the statistical power.

Additionally, the script generates a graph depicting the sampling distribution of the K^2 statistic, estimated using both the asymptotic approach (a chi-square distribution with $1 + k(k+1)(k+2)/6$ degrees of freedom) and resampling methods (empirical and normative bootstrap distributions). These density curves tend to converge when the original sample is drawn from a multivariate normal distribution, especially when inter-variable correlations are low. The greater the deviation from normality, the more pronounced the discrepancies among the three curves.

Given that the hypothesis pertains to the distribution (*i.e.*, multivariate normality), a normative sampling distribution can be generated to replace the asymptotic

approximation (chi-square) for computing the p -value. This also allows for the determination of a critical value to evaluate the rejection rate within the empirical bootstrap distribution, thereby estimating statistical power (*i.e.*, the proportion of estimates exceeding the critical value). The normative bootstrap distribution is less peaked and exhibits a heavier right tail than the chi-square distribution, resembling a generalized chi-square distribution—typical of quadratic forms involving correlated Gaussian variables [6], as illustrated in the example. For the K^2 statistic, this behavior reflects the correlation between multivariate skewness and [1] [37]. To preserve the original interdependence among variables, the lower triangular matrix from the Cholesky decomposition of the correlation matrix is used to generate the normative sample [38], and resampling is performed by rows [39].

Psychological theory holds that intellectual abilities are normally distributed and highly correlated, forming the basis for a general factor [40]. Beyond demonstrating the script's correct execution with an example where all three MVN tests—and the convergence of the sampling distribution curves—supported the multivariate normality of intellectual abilities, the study's second objective was to compare the performance of the two K^2 test variants with that of Royston's H test.

Two hundred samples were simulated under the null hypothesis of multivariate normality, and another 200 under non-normal conditions using a multivariate t-distribution with five degrees of freedom. The number of variables (ranging from 2 to 6), sample size (50, 75, 100, 125, 150, 200, 250, and 500), and the homogeneous correlation among variables were systematically varied. These three factors—number of variables, sample size, and inter-variable correlation—were treated as cofactors, as they are quantitative variables that influence statistical power in hypothesis testing (Jobst *et al.*, 2023). To simplify the simulation, homogeneous correlations were used (0, 0.3, 0.5, 0.7, and 0.9). Combining the two sample sets yielded a dataset of 400 observations, with a “distribution type” variable including two categories (MVT and MVN). As both sample sets were generated using the same random seed (123), they were analyzed using repeated measures statistical methods [31].

To assess the performance of each MVN test, two indicators were considered: hit rate and statistical power [41]. When rejecting the null hypothesis, power should exceed 0.5—ideally reaching 0.8 or 0.9—when 200 samples are drawn from a multivariate t-distribution with five degrees of freedom. Conversely, when the null hypothesis is true (*i.e.*, the 200 samples are from a multivariate normal distribution), power should remain below 0.5, ideally under 0.2 or 0.1 [42]. Because of this inverse directionality, when comparing the power of the three MVN tests in the combined sample, the complement of power (*i.e.*, the type II error rate) was used for the MVN samples.

All three MVN tests demonstrated high hit rates, exceeding 0.9. When the null hypothesis was retained, the H test showed the best performance, whereas when it was rejected, the two K^2 variants outperformed the H test. In the combined sample, the bootstrap variant of the K^2 test demonstrated superior performance com-

pared to the H test.

Mean comparisons were conducted among the MVN tests with respect to statistical power. The distributions of statistical power were non-normal, both univariately and multivariately. Likewise, the residuals from the general linear models violated the assumption of normality. Assumptions of sphericity (*i.e.*, equality of variances in pairwise differences) and homogeneity of observed covariance matrices across dependent variables were also violated. In these models, the powers of the MVN tests were treated as correlated or repeated dependent variables, with sample size, number of variables, and homogeneous correlation as cofactors. Additionally, distribution type served as a between-subjects factor defining two independent groups in the combined model.

As a result, the repeated measures ANOVA results must be interpreted with caution. For this reason, nonparametric tests—specifically the Friedman test and pairwise Conover comparisons—were used. Moreover, Pearson correlations between statistical power and the cofactors were computed, with confidence intervals and significance levels estimated using the bootstrap BCa method. Confidence intervals including zero (*i.e.*, with bounds of opposite signs) were interpreted as indicating non-significant correlations [43].

In the multivariate t-distribution samples, the mean statistical power of the asymptotic K^2 variant was significantly higher than that of both the bootstrap variant and the H test. In the multivariate normal samples, the bootstrap variant showed the best performance, followed by the asymptotic K^2 variant, with the H test performing worst. In the combined sample, the asymptotic variant exhibited the highest mean power, followed by the bootstrap variant, while the H test had the lowest.

Although the asymptotic variant of Mardia's K^2 test consistently achieved higher mean power under non-normal conditions, this advantage must be interpreted in light of its comparatively weaker control of Type I error under multivariate normality. In hypothesis testing, higher power does not necessarily imply superior overall performance if it is accompanied by an inflated rejection rate when the null hypothesis is true [44]. In this sense, the bootstrap variant represents a more balanced compromise between sensitivity to departures from normality and adherence to the nominal significance level [7].

This trade-off reflects a well-known distinction between asymptotic and resampling-based procedures [45]. Asymptotic tests tend to gain power more rapidly with increasing dimensionality and sample size, but may rely on distributional approximations that are less accurate in finite samples. In contrast, bootstrap calibration adapts the critical values to the empirical sampling distribution, improving Type I error control at the cost of a modest reduction in power under certain alternatives.

Accordingly, the choice between the asymptotic and bootstrap variants of the K^2 test should be guided by the primary inferential goal. When sensitivity to non-normality is paramount and moderate deviations from the nominal Type I error

rate are acceptable, the asymptotic variant may be preferred. Conversely, when strict control of the significance level is required—particularly in confirmatory analyses—the bootstrap variant provides a more robust and reliable option.

In terms of hit rate ($HR > 0.90$) and statistical power (high for rejection— $\phi \geq 0.8$ —and low for retention— $\phi \leq 0.2$), Mardia's test slightly outperformed the H test, consistent with Anis *et al.* [13], who emphasized Type II error and power. Specifically, the current power comparison found that the K^2 test outperformed the H test, particularly in its bootstrap variant. However, Moral [46] found that the asymptotic form of the K^2 test performed worse than both the H test and the chi-square approximation of the Q-test in a study involving a broader range of non-normal distributions and greater variability in sample sizes. Notably, this disadvantage was not observed for a multivariate t-distribution with 100 degrees of freedom, and the bootstrap variant of the K^2 test was not included in that study. Therefore, no inconsistency exists between the two findings.

Analyses of within-subject effects in the repeated measures ANOVA and correlation results revealed an interaction between sample size and MVN test type. Larger sample sizes increased statistical power, with the bootstrap K^2 variant benefiting most from the increase—although the strength of association was moderate to low. In contrast, the H test benefited least.

The number of variables also interacted with MVN test type. A larger number of variables increased statistical power, particularly for the asymptotic K^2 variant. The H test showed the smallest benefit, and its improvement was mainly evident in samples drawn from a multivariate normal distribution.

Homogeneous correlation did not interact with MVN test type. The statistical power of both K^2 variants was entirely unaffected by homogeneous correlation among variables. Only in multivariate normal samples did higher correlations slightly enhance the H test's power, with a small effect size.

The invariance of Mardia's tests—for multivariate skewness, kurtosis, and their combined K^2 statistic—across varying levels of homogeneous or heterogeneous correlation is due to their reliance on standardized central moments, which remove the effects of scale and correlation, rendering the K^2 statistic insensitive to correlation magnitude.

When comparing the critical values obtained in this study (asymptotic and bootstrap) with those derived from Urzua's Monte Carlo simulation [4], the latter were consistently lower—even compared to the asymptotic chi-square values (with $1 + k(k + 1)(k + 2)/6$ degrees of freedom). This discrepancy arises from differences in simulation procedures. The present study conditioned simulations on both the empirical sample (original data) and a normative sample (perfectly symmetric and normally distributed), each consisting of correlated variables preserving the original correlational structure. This design preserved not only the number of variables and observations but also their distributional and correlational characteristics, thus more closely approximating the sampling distribution of the statistic.

In contrast, Monte Carlo simulations—where b independent samples of size n are generated from a MVN distribution with zero correlation—condition the sampling distribution on an idealized scenario. While this minimizes error and leads to lower critical values, it sacrifices representativeness of empirical conditions [47].

This study has several limitations related to the scope of a simulation design. Only two generating distributions were considered—a multivariate normal distribution under the null hypothesis and a symmetric but leptokurtic multivariate t distribution with five degrees of freedom—excluding other relevant families such as asymmetric, mesokurtic, or near-normal alternatives.

In addition, the range of sample sizes was restricted, omitting very small and very large samples. The simulations were based on 400 parameter datasets spanning combinations of sample size, dimensionality, and homogeneous correlation, which allowed performance to be assessed on average but did not permit precise estimation of type-I error or power for specific parameter configurations; this limitation could be addressed in future work by generating multiple samples per condition.

Finally, all datasets assumed homogeneous correlations, whereas extending the analysis to heterogeneous dependence structures—using, for example, random positive-definite covariance models—would substantially broaden the applicability of the results and provide a more comprehensive evaluation of the robustness of the bootstrap and asymptotic procedures.

5. Conclusions

The developed R script enables the execution of Mardia's K^2 test using either an asymptotic or bootstrap approach. Unlike previous studies, the bootstrap method focuses on the sample data while preserving its correlational structure. The script allows for the computation of the test statistic, its p -value, and statistical power, as well as the graphical representation of the sampling distribution of the K^2 statistic. Both variants of the Mardia test included in the script are complemented by Royston's H test.

A simulation study comparing two multivariate normality tests—one of which includes two variants—under two types of distributions (one favoring the null hypothesis and the other opposing it) revealed that no single MVN test uniformly dominates across all conditions. The H test performs best for retaining normality, whereas Mardia's K^2 test—particularly its bootstrap variant—shows superior performance for detecting departures from normality. Consequently, the bootstrap K^2 test is recommended when the primary objective is sensitivity to non-normality, while acknowledging that its advantage depends on sample size, dimensionality, and the underlying distribution.

The sample distribution resulting from the bootstrap approach is less peaked and exhibits a heavier right tail compared to the distribution obtained from the asymptotic approximation. This distribution corresponds to a generalized chi-

square, which is typical of sums of squares of correlated normal variables.

6. Suggestions

It is recommended to use the script (available as a Word file in a GitHub repository at

https://github.com/josemoraldearubia579/R_script_for_MVN_K2_and_H_tests.git) when testing for multivariate normality and to further investigate the bootstrap variant of Mardia's K^2 test across various distributions, as well as in comparison with other tests, such as the Henze-Zirkler test [48], which is available in R [49].

Acknowledgements

The author wishes to thank the reviewers and the editor for their valuable comments and corrections, which significantly enhanced the quality of the manuscript.

Conflicts of Interest

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

References

- [1] Mardia, K.V. (1970) Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, **57**, 519-530. <https://doi.org/10.2307/2334770>
- [2] Mardia, K.V. (1974) Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies. *Sankhya: The Indian Journal of Statistics, Series B (1960-2002)*, **36**, 115-128. <https://www.jstor.org/stable/25051892>
- [3] 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)
- [4] Urzua, C.M. (1997) Omnibus Tests for Multivariate Normality Based on a Class of Maximum Entropy Distributions. In: Fomby, T.B. and Hill, R.C., Eds., *Applying Maximum Entropy to Econometric Problems*, Emerald Group Publishing Limited, 341-358. [https://doi.org/10.1108/s0731-9053\(1997\)0000012016](https://doi.org/10.1108/s0731-9053(1997)0000012016)
- [5] Koizumi, K., Okamoto, N. and Seo, T. (2009) On Jarque-Bera Tests for Assessing Multivariate Normality. *Journal of Statistics Advances in Theory and Applications*, **1**, 207-220.
- [6] Das, A. (2025) New Methods to Compute the Generalized Chi-Square Distribution. *Journal of Statistical Computation and Simulation*, **95**, 2608-2642. <https://doi.org/10.1080/00949655.2025.2501401>
- [7] 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>
- [8] R Core Team (2025) R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- [9] 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>

- [10] 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>
- [11] 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>
- [12] 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>
- [13] 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>
- [14] 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>
- [15] 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>
- [16] Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., Maechler, M. and Hothorn, T. (2025) MVTNORM: Multivariate Normal and t Distributions (Version 1.3-3). <https://doi.org/10.32614/CRAN.package.mvtnorm>
- [17] Newcombe, R.G. (1998) Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistics in Medicine*, **17**, 857-872. [https://doi.org/10.1002/\(sici\)1097-0258\(19980430\)17:8<857::aid-sim777>3.0.co;2-e](https://doi.org/10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e)
- [18] Cochran, W.G. (1950) The Comparison of Percentages in Matched Samples. *Biometrika*, **37**, 256-266. <https://doi.org/10.1093/biomet/37.3-4.256>
- [19] Serlin, R.C., Carr, J. and Marascuilo, L.A. (1982) A Measure of Association for Selected Nonparametric Procedures. *Psychological Bulletin*, **92**, 786-790. <https://doi.org/10.1037/0033-2909.92.3.786>
- [20] Sheskin, D.J. (2011) Handbook of Parametric and Non-Parametric Statistical Procedures. 5th Edition. Chapman and Hall. <https://doi.org/10.1201/9780429186196>
- [21] Barnett, M.J., Doroudgar, S., Khosraviani, V. and Ip, E.J. (2022) Multiple Comparisons: To Compare or Not to Compare, That Is the Question. *Research in Social and Administrative Pharmacy*, **18**, 2331-2334. <https://doi.org/10.1016/j.sapharm.2021.07.006>
- [22] Cohen, J. (1988) Statistical Power Analysis for Behavioral Sciences. 2nd Edition, Lawrence Erlbaum Associates.
- [23] IBM Corporation (2022) IBM SPSS Statistics for Windows, Version 27.0. IBM Corporation.
- [24] Mauchly, J.W. (1940) Significance Test for Sphericity of a Normal n -Variate Distribution. *The Annals of Mathematical Statistics*, **11**, 204-209. <https://doi.org/10.1214/aoms/1177731915>
- [25] Pillai, K.C.S. (1955) Some New Test Criteria in Multivariate Analysis. *The Annals of Mathematical Statistics*, **26**, 117-121. <https://doi.org/10.1214/aoms/1177728599>
- [26] Din, I.U. and Hayat, Y. (2021) ANOVA or MANOVA for Correlated Traits in Agricultural Experiments. *Sarhad Journal of Agriculture*, **37**, 1250-1259. <https://doi.org/10.17582/journal.sja/2021/37.4.1250.1259>
- [27] Hedges, L.V. and Olkin, I. (1985) Statistical Methods for Meta-Analysis. Academic

Press.

- [28] 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>
- [29] Royston, P. (1992) Approximating the Shapiro-Wilk W-Test for Non-normality. *Statistics and Computing*, **2**, 117-119. <https://doi.org/10.1007/bf01891203>
- [30] Friedman, M. (1937) The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, **32**, 675-701. <https://doi.org/10.1080/01621459.1937.10503522>
- [31] Conover, W.J. (1999) Practical Nonparametric Statistics. 3rd Edition. John Wiley and Sons.
- [32] Kendall, M.G. and Smith, B.B. (1939) The Problem of m Rankings. *The Annals of Mathematical Statistics*, **10**, 275-287. <https://doi.org/10.1214/aoms/1177732186>
- [33] In, J. and Lee, D.K. (2024) Alternatives to the P Value: Connotations of Significance. *Korean Journal of Anesthesiology*, **77**, 316-325. <https://doi.org/10.4097/kja.23630>
- [34] JASP Team (2024) JASP Version 0.19.2 [Computer Software]. <https://jasp-stats.org/download/>
- [35] Lyubomirsky, S. and Lepper, H.S. (1999) A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation. *Social Indicators Research*, **46**, 137-155. <https://doi.org/10.1023/a:1006824100041>
- [36] Quezada, L., Landero, R. and González, M.T. (2016) A Validity and Reliability Study of the Subjective Happiness Scale in Mexico. *The Journal of Happiness and Well-Being*, **4**, 90-100.
- [37] Jammalamadaka, S.R., Taufer, E. and Terdik, G.H. (2020) On Multivariate Skewness and Kurtosis. *Sankhya A*, **83**, 607-644. <https://doi.org/10.1007/s13171-020-00211-6>
- [38] Burgess, N. (2022) Correlated Monte Carlo Simulation Using Cholesky Decomposition. Social Science Research Network (SSRN). <https://ssrn.com/abstract=4066115>
- [39] Leisch, F. and Kostyshak, S. (2025) Package 'Bootstrap'. <https://cran.r-project.org/web/packages/bootstrap/bootstrap.pdf>
- [40] Korteling, J.E., van de Boer-Visschedijk, G.C., Blankendaal, R.A.M., Boonekamp, R.C. and Eikelboom, A.R. (2021) Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, **4**, Article 622364. <https://doi.org/10.3389/frai.2021.622364>
- [41] Moral De La Rubia, J. (2025) Refined Normality Test Based on the Parametric Seven-Number Summary. *İstatistik ve Uygulamalı Bilimler Dergisi*, **11**, 1-39. <https://doi.org/10.52693/jsas.1670945>
- [42] Quach, N.E., Yang, K., Chen, R., Tu, J., Xu, M., Tu, X.M., et al. (2022) Post-Hoc Power Analysis: A Conceptually Valid Approach for Power Based on Observed Study Data. *General Psychiatry*, **35**, e100764. <https://doi.org/10.1136/gpsych-2022-100764>
- [43] Mokhtar, S.F., Md Yusof, Z. and Sapiri, H. (2023) Confidence Intervals by Bootstrapping Approach: A Significance Review. *Malaysian Journal of Fundamental and Applied Sciences*, **19**, 30-42. <https://doi.org/10.11113/mjfas.v19n1.2660>
- [44] Zhang, Z. and Yuan, K.H. (2018) Practical Statistical Power Analysis Using Web Power and R. ISDSA Press. <https://doi.org/10.35566/power>
- [45] Davidson, R. and MacKinnon, J.G. (2006) The Power of Bootstrap and Asymptotic Tests. *Journal of Econometrics*, **133**, 421-441. <https://doi.org/10.1016/j.jeconom.2005.06.002>
- [46] Moral de la Rubia, J. (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>
- [47] Manly, B.F. and Navarro-Alberto, J.A. (2021) Randomization, Bootstrap and Monte Carlo Methods in Biology. 4th Edition, Chapman and Hall/CRC.
<https://doi.org/10.1201/9780429329203>
- [48] 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>
- [49] Korkmaz, S. (2025) Package ‘MVN’. Multivariate Normality Tests.
<https://cran.r-project.org/web/packages/MVN//MVN.pdf>

Appendix 1

Asymptotic and bootstrap variants of Mardia's MVN K^2 test for samples of at least 20 observations of k-tuples

Define the multivariate random sample

```
library(mvtnorm)
```

Parameters

```
n <- 61
```

```
k <- 3
```

Correlation matrix

```
sigma <- matrix(c(1, 0.8, 0.7,
```

```
0.8, 1, 0.6,
```

```
0.7, 0.6, 1),
```

```
nrow = k, byrow = TRUE)
```

Data simulation

```
set.seed(4567)
```

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

Scaling and rounding

```
x1 <- round(20 * mvn_data[, 1] + 65, 0) # Mechanical reasoning D-scores
```

```
x2 <- round(20 * mvn_data[, 2] + 60, 0) # Spatial reasoning D-scores
```

```
x3 <- round(20 * mvn_data[, 3] + 68, 0) # Numerical reasoning D-scores
```

Database

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

```
# print(original_data)
```

R package required

```
library(MVN)
```

Sample size, number of variables, and average of the correlation among variables

```
n <- length(x1)
```

```
cat("\nSample size: n =", n, ".\n")
```

```
k <- length(original_data)
```

```
cat("Number of variables in the original sample: k =", k, ".\n")
```

```
alpha <- 0.05
```

```
cat("Significance level:  $\alpha$  =", alpha, ".\n")
```

```
COR <- cor(original_data)
```

```
LT_COR <- COR[lower.tri(COR)]
```

```
cat("Average of the correlation among variables: m(R) =", round(mean(LT_COR), 4), ".\n")
```

Mardia's multivariate skewness (B1) and multivariate kurtosis (B2) tests

```
mardia <- mvn(original_data, subset = NULL, mvnTest = "mardia")
```

```

skew <- as.numeric(as.character(mardia$multivariateNormality$Statistic[1]))
kurt <- as.numeric(as.character(mardia$multivariateNormality$Statistic[2]))
p_skew <- as.numeric(as.character(mardia$multivariateNormality$p[1]))
p_kurt <- as.numeric(as.character(mardia$multivariateNormality$p[2]))
cat("\nMardia's multivariate skewness and kurtosis tests\n")
cat("Mardia's multivariate skewness statistic: B1 =", round(skew, 4), ".\n")
cat("Two-tailed p-value for the null hypothesis of multivariate symmetry: p =",
round(p_skew, 4), ".\n")
cat("Mardia's multivariate kurtosis statistic: B2 =", round(kurt, 4), ".\n")
cat("Two-tailed p-value for the null hypothesis of multivariate mesokurtosis: p =",
round(p_kurt, 4), ".\n")

# Mardia's MVN K2 test
# Function to calculate K2 statistic
calculate_k2 <- function(data) {
mardia <- mvn(data, mvnTest = "mardia")
skew <- as.numeric(as.character(mardia$multivariateNormality$Statistic[1]))
kurt <- as.numeric(as.character(mardia$multivariateNormality$Statistic[2]))
p_skew <- as.numeric(as.character(mardia$multivariateNormality$p[1]))
p_kurt <- as.numeric(as.character(mardia$multivariateNormality$p[2]))
k2_stat <- skew + kurt^2
return(k2_stat)
}

# Statistic observed in the original sample
k2_stat <- calculate_k2(original_data)
df_mardia <- 1 + (k * (k + 1) * (k + 2)) / 6
p_val <- pchisq(k2_stat, df = df_mardia, lower.tail = FALSE)
asympt_crit_value <- qchisq(1 - alpha, df = df_mardia, lower.tail = TRUE)
asympt_power <- 1 - pchisq(asympt_crit_value, df = df_mardia, ncp = k2_stat,
lower.tail = TRUE)

# Interpretation from asymptotic approach
cat("\nAsymptotic variant of the MVN K-square test\n")
cat("Mardia's multivariate normality statistic: MVN K-square =", round(k2_stat,
4), ".\n")
cat("Degrees of freedom: df =", df_mardia, ".\n")
cat("Critical value:", 1-alpha, "_chi-sq[" , df_mardia, "] =",
round(asympt_crit_value, 4), ".\n")
cat("Right-tailed p-value under a chi-square distribution with", df_mardia, "de-
grees of freedom: p_value =", round(p_val, 4), ".\n")
if (p_val < alpha) {
cat("H0 (multivariate normality) is rejected at the level", alpha, "based on the

```

```

asymptotic p-value (chi-2 distribution).\n")
} else {cat("H_0 (multivariate normality) is maintained at the level", alpha, " as-
ymptotic p-value (chi-2 distribution).\n")}
cat("Asymptotic statistical power for the K-square statistic: phi =",
round(asymp_power, 4), ".\n")
cat("\nIf 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")

# Normative sample (with the original correlations)
p <- seq(0.5 / n, 1 - 0.5 / n, length.out = n)
independent_data <- matrix(NA, nrow = n, ncol = k)
set.seed(123)
for (j in 1:k) {
independent_data[, j] <- qnorm(sample(p))} # Resampling with replacement by
row
cor_matrix <- cor(original_data)
chol_matrix <- chol(cor_matrix)
normative_data <- independent_data %*% chol_matrix

# Bootstrap normative distribution
B <- 1000
k2_boot_norm_vals <- replicate(B, {
indices <- sample(1:n, size = n, replace = TRUE)
calculate_k2(normative_data[indices, ])}

# Empirical bootstrap distribution. Resampling with replacement by row
n <- nrow(original_data)
set.seed(123)
k2_boot_vals <- replicate(B, {
indices <- sample(1:n, size = n, replace = TRUE)
boot_data <- original_data[indices, ]
calculate_k2(boot_data)})

# Bootstrap critical value, p-value, and power
boot_crit_value <- quantile(k2_boot_norm_vals, probs = 0.95, type = 8)
boot_p_value <- mean(k2_boot_norm_vals >= k2_stat)
boot_power <- mean(k2_boot_vals > boot_crit_value)

# Interpretation from bootstrap approach
cat("\nBootstrap variant of the MVN K-square test\n")

```

```

cat("Bootstrap critical value (0.95 quantile): q_boot(0.95) =",
round(boot_crit_value, 4), ".\n")
cat("Bootstrap p-value: p_boot =", round(boot_p_value, 4), ".\n")
if (boot_p_value < alpha) {
cat("H_0 (multivariate normality) is rejected at the level", alpha, "based on the
bootstrap p-value.\n")
} else {cat("H_0 (multivariate normality) is maintained at the level", alpha, "based
on the bootstrap p-value.\n")}
cat("Bootstrap power: phi_boot =", boot_power, ".\n")

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

# Green density curve of the chi-square distribution as the base graph
x_seq <- seq(0, max(qchisq(0.999, df = df_mardia), max(k2_boot_norm_vals) +
1, k2_stat + 1, max(k2_boot_vals) + 1), length.out = 1000)
y_chi_sq <- dchisq(x_seq, df = df_mardia)
plot(x_seq, y_chi_sq, type = "l", lwd = 3, col = "darkgreen",
main = "", xlab = "K_square_values", ylab = "Density",
xlim = c(0, max(qchisq(0.999, df = df_mardia), max(k2_boot_norm_vals) + 1,
k2_stat + 1, max(k2_boot_vals) + 1)),
ylim = c(0, max(density(k2_boot_vals)$y, y_chi_sq)))
# Yellow histogram of the empirical bootstrap distribution of the K-square test
statistic
hist(k2_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 K-square sta-
tistic
lines(density(k2_boot_vals), col = "yellow", lwd = 2)
# Red density curve of the normative bootstrap distribution of the K-square sta-
tistic
lines(density(k2_boot_norm_vals), col = "red", lwd = 2)
# Blue vertical line for the observed K-square statistic value
abline(v = k2_stat, col = "darkblue", lwd = 2, lty = 2)
# Red vertical line for the normative bootstrap critical value
abline(v = boot_crit_value, col = "red", lwd = 2, lty = 2)
# Green vertical line for the asymptotic critical value
abline(v = qchisq(1-alpha, df = df_mardia), col = "darkgreen", lwd = 2, lty = 2)
# dev.off() # Remove the preceding hashtag symbol to save the figure
cat("\nFigure. Histogram (yellow) with density curves for the empirical bootstrap

```

(yellow), normative bootstrap (red), and chi-square (green) distributions. The value of the k-square statistic is indicated by a blue vertical line (k-square =", round(k2_stat, 4), "), the critical value of the normative bootstrap distribution by a red vertical line ("1-alpha, "q_boot =", round(boot_crit_value, 4), "), and the critical value of the chi-square distribution with", df_mardia, "degrees of freedom by a green vertical line ("1 - alpha, "Chi-square[" df_mardia, "] =", round(asymp_crit_value, 4), ").\n")

Appendix 2

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, 97, 72, 112, 63, 79, 26, 66, 104, 68, 90, 85, 49, 64, 71, 73, 69, 96, 64, 65,
57, 91, 108, 45, 37, 45, 38, 71, 75, 46, 48, 61, 80, 93, 35, 70, 73, 72, 71, 111, 77, 56,
54, 43, 40, 67, 59, 29, 83, 64, 46, 72, 49, 44, 59, 65, 94, 48, 53, 66, 71)
```

```
x2 <- c(39, 97, 66, 106, 64, 71, 22, 61, 94, 61, 70, 79, 52, 51, 50, 68, 51, 102, 55, 58,
59, 83, 121, 41, 33, 58, 22, 70, 65, 57, 33, 65, 82, 75, 35, 49, 86, 44, 89, 88, 67, 36, 62,
55, 35, 73, 58, 50, 97, 39, 35, 85, 58, 35, 44, 40, 87, 43, 52, 78, 73)
```

```
x3 <- c(63, 108, 66, 108, 69, 89, 53, 79, 85, 64, 70, 61, 62, 71, 78, 63, 38, 68, 58, 69,
77, 99, 91, 53, 44, 73, 58, 73, 67, 44, 61, 85, 72, 60, 73, 88, 78, 49, 70, 88, 75, 54, 42,
47, 58, 71, 76, 23, 55, 84, 47, 87, 78, 46, 65, 78, 91, 47, 42, 73, 62)
```

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

```
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 = result$mul-
```

```

tivariateNormality$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 coefficients
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 =", round(df_H, 4), "\n")
cat("p-value for the null hypothesis of multivariate normality: p =", round(re-
sult$multivariateNormality$p, 4), "\n")
cat("Statistical power of the H test at a significance level of", alpha, ": phi =",
round(power_H, 4), "\n")
cat("\nIf 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")

```

Appendix 3. Dataset for Example 2

```

# List of original variables
x1 <- c(4, 3, 4, 5, 5, 4, 3, 6, 5, 5, 5, 5, 7, 6, 6, 6, 6, 5, 6, 6, 7, 7, 6, 7, 6, 6, 6, 6, 7, 6, 6,
7, 7, 7, 6, 7, 5, 7, 7, 7, 6, 7, 6, 7, 7, 3, 1, 4, 6, 4, 4, 5, 3, 3, 3, 4, 4, 3, 5, 4, 4, 4, 4, 5,
5, 4, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 5, 4, 6, 5, 5, 6, 6, 4, 6, 6, 6, 6, 5, 6, 6, 6, 6, 5, 5,
6, 6, 5, 6, 6, 6, 5, 5, 6, 6, 5, 5, 6, 6, 5, 7, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6, 7, 7, 7,
6, 6, 6, 5, 6, 6, 6, 6, 7, 6, 6, 6, 6, 6, 5, 5, 6, 7, 7, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 6, 6,
6, 7, 6, 7, 7, 6, 7, 6, 6, 7, 7, 4, 4, 4, 5, 5, 6, 5, 6, 6, 7, 6, 6, 6, 6, 6, 7, 7, 7, 7, 6, 4, 4,
5, 5, 5, 6, 1, 5, 5, 3, 6, 7, 7, 7, 6, 7, 6, 6, 6, 6, 7, 7, 4, 5, 6, 7, 6, 7, 7, 3, 6, 7, 6, 7, 7, 6,
3, 3, 5, 6, 7, 5, 5, 3, 6, 6, 6, 7, 2, 5, 6, 6, 4, 5, 6, 6, 4, 6, 7, 7, 7, 7, 6, 6, 6, 6, 3, 3, 5, 6,
6, 3, 4, 6, 6, 7, 5, 6, 6, 6, 4, 5, 7, 7, 7, 4, 4, 7, 4, 5, 3, 4, 7, 6, 5, 6, 5, 4, 4, 5, 6, 5, 5, 6, 7,
7, 2, 2, 5, 5, 5, 6, 6, 6, 4, 4, 4, 5, 5, 3, 3, 5, 7, 6, 3, 5, 6, 7, 4, 4, 4, 5, 6, 5, 6, 6, 6, 6, 6,
6, 5, 6, 6, 6, 7, 1, 6, 5, 6, 5, 6, 6, 7, 7, 5, 4, 4, 5, 5, 6, 6, 7, 4, 5, 4, 5, 5, 5, 5, 4, 4, 6, 5,
6, 6, 5, 6, 6, 6, 6, 5, 6, 7, 7, 6, 6, 5, 7, 5, 6, 6, 6, 6, 6, 5, 7, 6, 6, 6, 6, 7, 7, 7, 6, 7, 7,
6, 6, 6, 7, 6, 6, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 7, 6, 7, 6, 4, 4, 5, 5,
5, 7, 5, 6, 6, 6, 7, 6, 5, 6, 6, 5, 7, 7, 6, 6, 5, 6, 7, 5, 7, 6, 5, 5, 4, 6, 4, 6, 5, 6, 5, 6, 5, 7, 6,
5, 6, 7, 6, 5, 4, 6, 7, 6, 2, 5, 3, 5, 5, 6, 5, 3, 5, 3, 6, 4, 6, 6)
x2 <- c(3, 2, 3, 4, 3, 4, 4, 5, 4, 5, 4, 5, 6, 4, 5, 5, 5, 5, 6, 6, 7, 5, 5, 5, 7, 6, 6, 6, 6,
6, 6, 7, 7, 6, 7, 6, 7, 6, 7, 7, 6, 6, 6, 6, 7, 2, 1, 3, 1, 3, 2, 2, 4, 3, 3, 3, 4, 4, 4, 3, 4, 4, 4, 4,
4, 5, 4, 4, 4, 4, 5, 5, 5, 4, 5, 5, 4, 6, 6, 5, 5, 5, 5, 4, 4, 4, 6, 5, 6, 6, 5, 6, 6, 5, 5, 5, 5, 4,
4, 5, 7, 5, 6, 5, 6, 5, 5, 6, 6, 4, 6, 4, 5, 5, 7, 4, 6, 6, 6, 5, 5, 5, 7, 4, 6, 6, 6, 5, 6, 4, 7, 7, 5,
7, 6, 5, 6, 5, 5, 6, 5, 7, 6, 4, 6, 5, 5, 5, 6, 6, 5, 5, 5, 6, 6, 6, 5, 5, 6, 6, 6, 7, 6, 6, 7, 7,
7, 4, 7, 6, 6, 7, 7, 7, 7, 7, 7, 3, 3, 3, 4, 4, 4, 4, 6, 5, 5, 5, 6, 6, 6, 6, 6, 7, 6, 6, 7, 4, 4, 5,
5, 5, 4, 5, 7, 6, 5, 6, 5, 7, 7, 7, 6, 6, 7, 6, 6, 7, 7, 4, 6, 7, 7, 7, 7, 6, 1, 5, 6, 4, 7, 7, 7,

```

```
2, 2, 4, 5, 7, 5, 4, 3, 4, 7, 5, 7, 2, 6, 5, 6, 3, 4, 6, 6, 4, 6, 6, 7, 4, 6, 7, 6, 5, 4, 6, 4, 3, 5, 4,
6, 2, 4, 6, 6, 6, 4, 6, 5, 7, 4, 5, 7, 5, 7, 4, 4, 6, 4, 4, 3, 4, 7, 7, 5, 7, 4, 4, 5, 4, 6, 4, 5, 7, 7,
2, 2, 3, 4, 3, 6, 6, 5, 5, 3, 4, 4, 4, 4, 5, 5, 5, 7, 7, 3, 4, 6, 7, 4, 4, 4, 4, 5, 6, 5, 6, 6, 5, 5, 5,
5, 5, 6, 5, 7, 7, 6, 6, 5, 5, 5, 4, 6, 7, 6, 4, 3, 3, 5, 5, 4, 6, 7, 4, 4, 4, 5, 4, 4, 6, 5, 4, 4, 5, 5,
5, 5, 5, 6, 5, 6, 6, 5, 5, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 4, 4, 6, 7, 6, 6, 4, 5, 6, 7, 7, 6, 6, 6, 7, 6,
7, 7, 7, 7, 6, 6, 7, 7, 6, 5, 7, 7, 6, 6, 6, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 7, 7, 4, 4, 7, 4, 5, 6, 4,
5, 7, 6, 4, 6, 6, 7, 7, 6, 6, 5, 6, 7, 6, 5, 6, 5, 4, 6, 5, 5, 6, 4, 5, 7, 6, 7, 6, 5, 5, 5, 5, 5, 7, 6,
5, 5, 4, 5, 5, 4, 5, 6, 6, 3, 4, 2, 5, 5, 4, 4, 3, 4, 3, 5, 5, 5, 7)
```

```
x3 <- c(4, 2, 3, 4, 4, 2, 4, 2, 4, 4, 3, 4, 4, 4, 6, 5, 6, 3, 5, 5, 5, 5, 5, 6, 4, 6, 6, 7, 6, 5, 6,
6, 5, 6, 6, 6, 7, 6, 6, 6, 7, 7, 6, 7, 6, 7, 6, 1, 1, 2, 3, 2, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 4, 3, 4, 4,
4, 4, 4, 4, 4, 4, 4, 6, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 6, 5, 5, 4, 5, 5, 5, 5, 4, 6, 5, 5, 5, 4, 4, 5, 5,
5, 4, 4, 3, 6, 4, 5, 5, 5, 5, 5, 6, 6, 5, 6, 6, 7, 5, 6, 6, 5, 5, 5, 5, 5, 6, 6, 5, 5, 5, 5, 6, 7, 7, 6,
4, 5, 6, 5, 6, 6, 6, 6, 7, 5, 6, 6, 6, 5, 5, 5, 5, 6, 5, 5, 6, 6, 6, 7, 6, 6, 6, 5, 5, 5, 7, 7, 6, 6, 6,
6, 7, 6, 7, 7, 6, 7, 7, 7, 7, 7, 3, 4, 4, 3, 5, 4, 5, 3, 4, 5, 5, 5, 5, 5, 6, 5, 6, 6, 7, 2, 4, 4,
5, 5, 5, 6, 6, 6, 5, 5, 5, 7, 7, 7, 6, 5, 4, 5, 6, 5, 6, 7, 7, 4, 7, 7, 5, 6, 6, 6, 3, 5, 6, 6, 5, 6, 7,
1, 2, 3, 5, 5, 4, 5, 3, 5, 5, 6, 7, 2, 6, 5, 5, 4, 3, 4, 7, 2, 3, 7, 7, 2, 6, 6, 5, 5, 6, 7, 3, 5, 4, 2,
5, 2, 3, 5, 6, 4, 4, 5, 6, 7, 4, 5, 7, 6, 7, 4, 4, 6, 4, 4, 3, 1, 6, 7, 4, 6, 3, 6, 4, 4, 5, 5, 4, 6, 7,
2, 1, 3, 3, 2, 5, 5, 6, 7, 5, 4, 4, 4, 5, 5, 5, 4, 6, 7, 5, 5, 4, 7, 3, 3, 5, 6, 7, 6, 5, 5, 6, 4, 5, 5,
6, 6, 6, 6, 7, 7, 6, 4, 5, 6, 5, 6, 5, 6, 7, 6, 1, 3, 5, 4, 5, 7, 7, 3, 3, 5, 4, 5, 3, 4, 5, 4, 6, 5, 4,
5, 4, 4, 5, 4, 6, 5, 5, 4, 5, 5, 7, 4, 4, 5, 4, 4, 5, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6, 6, 5, 6, 6, 5, 6, 5, 6,
5, 6, 6, 7, 6, 7, 6, 6, 7, 7, 5, 6, 7, 7, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 4, 5, 7, 2, 5, 5, 5,
5, 6, 5, 7, 5, 5, 6, 5, 7, 6, 5, 6, 6, 5, 5, 4, 4, 6, 5, 6, 6, 5, 3, 5, 2, 5, 5, 4, 6, 5, 1, 5, 7, 6,
5, 5, 6, 5, 5, 1, 6, 6, 6, 3, 5, 3, 4, 4, 6, 5, 3, 6, 7, 5, 6, 5, 6)
```

```
x4 <- c(2, 7, 5, 3, 4, 6, 5, 4, 4, 4, 6, 4, 1, 5, 2, 3, 2, 6, 2, 3, 1, 1, 1, 1, 4, 3, 1, 1, 1, 3, 2,
1, 3, 1, 1, 4, 1, 5, 2, 3, 1, 1, 5, 3, 6, 5, 7, 2, 7, 2, 1, 2, 3, 1, 3, 4, 4, 3, 3, 5, 2, 4, 4, 5, 4, 3,
3, 4, 5, 4, 4, 4, 4, 3, 4, 5, 4, 4, 5, 2, 4, 5, 5, 3, 3, 4, 4, 5, 6, 2, 3, 3, 4, 4, 3, 3, 4, 5, 5, 5, 6,
5, 6, 5, 7, 3, 6, 4, 6, 6, 4, 4, 6, 4, 6, 4, 5, 1, 7, 4, 4, 5, 6, 6, 6, 4, 6, 4, 5, 5, 6, 6, 6, 1, 1, 4,
5, 5, 5, 6, 6, 6, 5, 6, 2, 6, 6, 5, 6, 7, 7, 7, 7, 7, 7, 6, 6, 6, 6, 7, 6, 6, 7, 7, 4, 5, 6, 6, 6,
6, 7, 6, 6, 6, 7, 6, 7, 7, 7, 7, 7, 3, 3, 5, 4, 3, 4, 5, 4, 5, 4, 5, 6, 6, 6, 7, 7, 7, 7, 7, 2, 4, 4,
2, 3, 5, 3, 6, 4, 6, 7, 5, 1, 1, 2, 6, 5, 7, 7, 6, 7, 7, 7, 2, 6, 6, 7, 7, 7, 6, 1, 3, 7, 7, 6, 7, 2,
1, 1, 3, 3, 4, 4, 2, 3, 4, 6, 7, 7, 1, 6, 7, 5, 5, 3, 4, 4, 3, 6, 6, 5, 3, 6, 6, 5, 6, 6, 2, 3, 4, 3, 1,
6, 1, 4, 6, 7, 4, 5, 3, 6, 6, 4, 7, 7, 7, 6, 4, 4, 7, 4, 4, 2, 4, 2, 6, 2, 2, 3, 3, 4, 4, 3, 6, 7, 3, 4,
1, 1, 2, 3, 5, 2, 1, 3, 5, 2, 4, 4, 4, 3, 5, 5, 4, 2, 6, 5, 2, 3, 2, 3, 3, 3, 2, 1, 3, 4, 3, 3, 6, 6, 6,
6, 7, 5, 6, 7, 7, 6, 4, 6, 5, 2, 6, 6, 7, 7, 5, 4, 5, 3, 6, 5, 1, 1, 3, 3, 4, 3, 4, 6, 3, 3, 6, 5, 3, 5,
4, 5, 6, 3, 5, 2, 3, 4, 6, 3, 3, 1, 5, 6, 5, 4, 6, 5, 4, 6, 6, 5, 4, 6, 5, 7, 7, 6, 6, 4, 5, 6, 6, 5, 5,
6, 6, 6, 4, 7, 7, 6, 6, 6, 7, 5, 7, 7, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 5, 3, 5,
4, 1, 5, 4, 6, 6, 6, 4, 3, 6, 6, 1, 1, 7, 3, 6, 7, 4, 7, 7, 6, 6, 4, 5, 6, 3, 6, 5, 4, 3, 6, 2, 4, 1, 6,
5, 3, 4, 4, 3, 4, 4, 4, 3, 2, 3, 6, 4, 6, 6, 5, 4, 2, 7, 6, 4, 6, 6)
```

```
# Database
```

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

Appendix 4. Generation of MVT and MVN samples

```
# Sample of 50 observations in two variables, with a correlation coefficient of 0,
```

drawn from a multivariate t-distribution with five degrees of freedom.

```
library(mvtnorm)
n <- 50 # Number of participants
k <- 2 # Number of variables
df <- 5 # Degrees of freedom for MVT distribution
rho <- 0 # Homogeneous correlation
sigma <- matrix(rho, nrow = k, ncol = k) # Correlation matrix
diag(sigma) <- 1 # Correlation matrix
set.seed(123) # Seed is set for reproducibility
mvt_data_2v <- rmvt(n = n, sigma = sigma, df = df)
x1 <- mvt_data_2v[,1]
x2 <- mvt_data_2v[,2]
original_data <- data.frame(x1, x2) # sample data
print(original_data)
```

Sample of 50 observations in two variables, with a correlation coefficient of 0, drawn from a multivariate standard normal distribution.

```
library(mvtnorm)
n <- 50 # Number of observations
k <- 2 # Number of variables
rho <- 0 # Homogeneous correlation
sigma <- matrix(rho, nrow = k, ncol = k) # Correlation matrix (compound symmetry)
diag(sigma) <- 1 # Correlation matrix
set.seed(123) # Seed is set for reproducibility
mvn_data <- rmvnorm(n = n, sigma = sigma)
x1 <- mvn_data[,1]
x2 <- mvn_data[,2]
original_data <- data.frame(x1, x2)
print(original_data)
```