STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 195–208, https://doi.org/10.59139/stattrans-2025-034
Received – 10.12.2024; accepted – 24.04.2025

Normality tests for transformed large measured data: a comprehensive analysis

Abu Feyo Bantu[1], **Andrzej Kozyra**[2], **Józef Wiora**[3]

Abstract

In statistical analysis, evaluating the normality of large datasets is crucial for validating parametric tests, particularly in areas such as Global Navigation Satellite System (GNSS) measurements, where data often exhibit non-normal characteristics resulting from their variability and errors. This research aims to transform the measured GNSS data and to assess the effectiveness of transformation methods in achieving normality. Techniques like logarithmic, quantile and rank-based Inverse Normal Transformation (INT) were evaluated using visual methods (histograms, Q-Q plots), descriptive statistics (skewness, kurtosis) and statistical tests, including Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Lilliefors (LF), D'Agostino K-squared (DA), Shapiro-Wilk (SW), Jarque-Bera (JB), Cramér-von Mises (CM), and Pearson Chi-square (Chi2) tests. The sensitivity of these tests to deviations from normality was assessed through the Receiver Operating Characteristic (ROC) analysis and the Area Under the Curve (AUC) values at a significance level of 0.1, using Monte Carlo (MC) simulations across the varying sample sizes. The results showed that untransformed latitude data consistently failed normality tests, while transformed data displayed normal characteristics. The rank-based INT showed superior effectiveness, influenced by the original distribution and characteristics of the dataset. The findings underscore the importance of tailored transformations in large-scale data applications, enhancing the accuracy and applicability of parametric statistical methods in geospatial and other industrial domains.

Key words: GNSS, ROC, normality test, statistical analysis.

1. Introduction

In statistical analysis, assessing the normality of datasets is essential, as many parametric tests rely on the assumption of normally distributed data as described in (Tabachnick et al., 2019). In highly competitive manufacturing industries, pilot runs typically involve very small sample sizes to accelerate the launch of new products. While these small datasets may approximate normal distributions due to their limited variability, large datasets, such as those collected from Global Navigation Satellite Systems (GNSS), often deviate significantly from normality due to various sources of error (Li et al., 2016). Yap and Sim (2011) highlighted that these non-normal characteristics, influenced by factors such as atmospheric

[1]Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland. E-mail: abu.feyo.bantu@polsl.pl. ORCID: https://orcid.org/0000-0001-6463-9864.
[2]Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland. E-mail: andrzej.kozyra@polsl.pl. ORCID:https://orcid.org/0000-0003-2645-3537.
[3]Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland. E-mail: jozef.wiora@polsl.pl. ORCID: https://orcid.org/0000-0002-8450-8623.

disturbances, multipath effects, and satellite geometry, pose challenges to the application of traditional parametric methods. Fault detection in GNSS systems is essential for ensuring reliability; however, traditional methods often assume Gaussian noise, which limits their effectiveness in real-world scenarios where measurement noise deviates from normality. Yan (2024) proposed a jackknife-based test statistic for fault detection in linearized pseudorange positioning systems, which does not rely on specific noise distribution assumptions. This method, combined with hypothesis testing using the Bonferroni correction, enhances robustness against non-Gaussian noise, making it particularly suitable for GNSS applications.

While such robust techniques improve fault detection and identify sources of errors in GNSS applications, the assumption of normality in large datasets remains a debated topic. Wilcox (2010) argues that, based on the central limit theorem, data naturally tend toward a normal distribution as sample size increases. This suggests that normality assumptions hold in large datasets, regardless of the method used to assess normality. However, Demir (2022) challenges this notion; it is not true: since the number of data is large, the data will not always have a normal distribution, particularly in the presence of measurement errors and non-Gaussian noise.

To address this limitation, data transformation techniques are widely used to modify data distributions and achieve a closer approximation to normality. According to Osborne (2010), these methods are crucial for fulfilling statistical assumptions, enhancing effect sizes, and preparing datasets for comprehensive analysis. Huang et al. (2023) emphasizes that data transformation is a commonly employed technique for normalizing data to enhance statistical modeling. The study also outlines various transformation methods, such as logarithmic, log-sinh, Box-Cox, Yeo-Johnson, and square root approaches, each designed to address different types of non-normality. For instance, the Box-Cox transformation is a flexible family of power transformations that adjusts data to stabilize variance and reduce skewness. Peterson (2021) further elaborates on the implementation of the Box-Cox transformation, emphasizing its ability to select an optimal parameter (λ) that maximizes normality. Similarly, the Yeo-Johnson transformation extends the Box-Cox method to accommodate negative values, making it more versatile for diverse datasets (Cai and Xu, 2024).

Despite the availability of these transformation techniques, selecting the most appropriate method for large, real-world datasets remains challenging due to their varying characteristics and sources of variability (Khatun, 2021). Normality tests such as Shapiro-Wilk, Anderson-Darling, and D'Agostino's K-squared provide insights into dataset distributions but offer limited guidance on effectively transforming data to satisfy normality assumptions (Razali and Wah, 2011). Moreover, studies like Ogaja (2022) highlight the importance of preprocessing steps in GNSS data analysis, where raw observables undergo rigorous processing to estimate geodetic parameters. Barba et al. (2021) demonstrate the use of adapted R packages to analyze GNSS time series, focusing on displacement velocities, noise levels, and temporal forecasts. Their work underscores the necessity of identifying and addressing outliers, gross errors, and noise to ensure reliable interpretation of GNSS data.

This research aims to transform the measured data and evaluate the efficiency of transformation methods in the normality of large GNSS datasets. It compares log, quantile, and rank-based INT transformations to provide insights for improving the reliability of parametric statistical methods in measurement-driven fields. We evaluated data normality using

graphical, descriptive, and statistical methods, including p-value tests, to examine the effects of transformations. The transformed data were then compared with untransformed data to assess improvements in normality.

The remainder of this paper is as follows: Section 2 reviews the relevant data transformation and normality tests. Section 3 introduces our methods and data collection techniques. Section 4 presents the results and discussion through graphical representations, descriptive and statistical analysis, and evaluation of normality test performance. Finally, Section 5 concludes the study.

2. Theory

2.1. Data transformation techniques

Raymaekers and Rousseeuw (2024) describe data transformation as a technique for handling non-Gaussian data by preprocessing variables to approximate normality. This process ensures that the transformed data are nearly normal at their core, while a few outliers may still deviate. Here is a breakdown of common techniques, including their formula as shown in Table 1.

Table 1: Data transformation techniques

Transformation techniques	Formula	When to use
Log Transform	$y = \log(x)$	When data are positively skewed.
Quantile Transform	$y_i = \text{Quantile}(x_i)$	When you want to map data to a target distribution (e.g. normal).
Rank-Based INT	$y_i = \Phi^{-1}\left(\frac{r_i}{N+1}\right)$	When data are highly non-normal or non-monotonic.

where y_i is the transformed data point after mapping, x_i is the original data point, Φ^{-1} is the inverse CDF of the normal distribution, N is the total number of data points, and r_i is the rank of the i-th data point in the dataset.

Log transformation is a widely used statistical method primarily applied to reduce data skewness and mitigate variability caused by outliers. It is commonly believed to enhance normality by making the data distribution more closely resemble a normal distribution (Sun and Xia, 2024). As described by Ghasemi and Zahediasl (2012), it involves taking the logarithm of each data value, which compresses the range of large values while preserving the relationships between data points.

According to Pham (2021), quantile transformation is a statistical technique that converts data from its original distribution to a target distribution. This approach effectively normalizes non-normal data and mitigates issues caused by outliers and heavy tails.

As explained by McCaw et al. (2020), the rank-based inverse normal transformation (INT) is commonly applied to highly skewed data and non-normally distributed traits. INT is a non-parametric mapping that replaces sample quantiles by quantiles from the standard normal distribution. After INT, the marginal distribution of any continuous outcome is asymptotically normal. INT has the effect of symmetrizing and concentrating the residual

distribution around zero. After applying data transformations to the dataset, normality was evaluated using various testing techniques to assess how well the data conform to a normal distribution.

2.2. Normality tests

A normality test is a statistical procedure used to determine whether a dataset follows a normal distribution. Since many statistical methods rely on the assumption of normality in population data, it is crucial to verify whether this assumption holds before applying such methods (Kwak and Park, 2019). There are three major methods to check the normality: graphical, descriptive, and statistical (Zygmonta, 2023).

Graphs allow the easy assessment of major data departures from normality (Obilor and Amadi, 2018). The normal Quantile-Quantile (Q-Q) plot is a widely used and successful method for determining data normality. As explained by Stine (2017), it validates the assumption of normality in a sample dataset. It compares the two datasets to see if they have the same distribution. A histogram is a bar graph representing the frequency distribution of a dataset. It divides data into bins and counts the number of observations in each bin. For a normally distributed dataset, it exhibits a bell-shaped curve, symmetric around the mean. If the dataset is non-normal, the histogram deviates from the bell-shaped curve.

The second method is descriptive analysis, which uses skewness (S) and kurtosis (K) to evaluate the shape of data distributions. This is one of the most commonly employed techniques for this purpose. Skewness S evaluates the asymmetry of a probability distribution, indicating how much the data deviate from a symmetrical distribution. A normal distribution, which is symmetrical and bell-shaped, has zero skewness. A positively skewed distribution ($S > 0$) has a long right tail, while a negatively skewed one ($S < 0$) has a long left tail with most values concentrated on the right (Tabachnick et al., 2019).

Kurtosis K is a statistical measure that describes the shape of a probability distribution by analysing its tails and peak relative to a normal distribution. According to Tabachnick et al. (2019), positive kurtosis indicates a distribution with heavier tails and a sharper peak, whereas negative kurtosis reflects lighter tails and a flatter peak. Excess kurtosis, calculated as $K-3$, compares the kurtosis of a given distribution to that of a normal distribution. There are three main types of kurtosis: mesokurtic $(K = 3, \text{excess kurtosis} = 0)$, representing a normal distribution; leptokurtic $(K > 3, \text{excess kurtosis} > 0)$, indicating a distribution with heavier tails and a sharper peak; and platykurtic $(K < 3, \text{excess kurtosis} < 0)$, which corresponds to a distribution with lighter tails and a flatter peak than a normal distribution.

The third technique is statistical, and p-values of tests such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Lilliefors (LF), D'Agostino's (DA), Shapiro-Wilk (SW), Jarque-Bera (JB), Cramér-von Mises (CM), and Pearson's Chi-square (Chi2).

The KS test compares the empirical Cumulative Distribution Function (CDF) of a sample to the CDF of a reference distribution (e.g. normal). The test statistic, as described in Kumbhar et al. (2024), is determined by the maximum absolute difference between the two CDFs. The Anderson-Darling test assesses whether a sample comes from a specified distribution. It leverages the principle that, under the assumption that the data originate from the hypothesized distribution, their CDF should follow a uniform distribution (Kwak and Park,

2019). The AD test assesses normality by comparing the test statistic to a critical value (cv) at a given significance level. It is a variation of the KS test, which applies additional weighting to emphasize differences in the tails of the distribution. The LF test is a variation of the KS test; it corrects for the fact that the parameters (mean and variance) of the normal distribution are estimated from the sample, which is calculated as in (Uyanto, 2022). The test statistic is the maximal absolute difference between empirical and hypothetical CDF. The DA test checks the S and K of the data and compares it to a normal distribution. It is used to assess if a dataset deviates from normality in terms of its symmetry (skewness) and peakness (kurtosis), with the detailed formula for its calculation provided in (D'Agostino, 2017).

The SW test evaluates how well the sample data fits a normal distribution by comparing the sample to the expected values under normality, with its statistical formula provided in (Kwak and Park, 2019).

The JB test is employed to verify the normality of a data set before applying standard statistical tests such as the t-test, z-test, or F-test. For a detailed explanation of the formula and parameters, refer to (Aslam et al., 2021). It evaluates S and K in the data, comparing them to their expected values under a normal distribution.

The CVM test measures the difference between the empirical distribution function of a sample and the CDF of the normal distribution. It is particularly useful when analyzing the tails of the probability density function (PDF). For a detailed description of its equation (Von Mises, 2014). The Chi2 test compares the observed frequencies in bins with the expected frequencies for a normal distribution. The original test by Pearson was designed to see whether an observed set of frequency O was in agreement with a multinomial distribution with parameters $m, p_1, \ldots, p_k$. This is done by calculating the expected frequency $E_i = mp_i$ and the test statistic $X = \sum(O-E)^2/E$ (Rolke and Gongora, 2021).

Different normality tests often produce different results; some tests reject the null hypothesis of normality while others fail to reject it (Demir, 2022). Therefore, it is better to crosscheck the normality of the data using different techniques. The statistical value and median p-value are the essential elements of the comparison tests (Khatun, 2021). The statistical value measures the difference between the observed and normal distributions. The median p-value for a normality test is defined as:

$$p_{\text{median}} = \begin{cases} p\left(\frac{n+1}{2}\right) & \text{if } n \text{ is odd,} \\ \dfrac{p\left(\frac{n}{2}\right) + p\left(\frac{n}{2}+1\right)}{2} & \text{if } n \text{ is even.} \end{cases} \tag{1}$$

It is the middle value of the p-values obtained from repeated tests on different samples $p_1, p_2, \ldots, p_n$, where p_i represents the p-value from the i-th normality test. It is a non-parametric measure of the central tendency of the p-values (Tabachnick et al., 2019). It helps to assess the normality by considering its skewness and kurtosis. By analyzing these p-values at 5% and 10% significance levels and varying sample sizes, we can gauge how well the data conform to a normal distribution. A p-value below 0.05 suggests rejecting the null hypothesis (H0), indicating non-normality, while a high p-value suggests insufficient evidence to reject H0, indicating possible normality.

After conducting normality tests, the performance of these tests was evaluated using the ROC curve with Area Under the Curve (AUC) values, offering insight into their ability to detect normality in both the original and transformed data. ROC is a graphical tool used to assess the diagnostic ability of a binary classifier, such as normality tests in statistics (Patrício et al., 2017). It is a plot of the True Positive Rate (TPR) versus the False Positive Rate (FPR) at various threshold settings. The TPR is:

$$\text{TPR} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \tag{2}$$

and FPR is:

$$\text{FPR} = \frac{\text{False Positive (FP)}}{\text{False Positive (FP)} + \text{True Negative (TN)}} \tag{3}$$

so to plot the ROC curve, we need to compute TPR and FPR for multiple threshold values. The AUC summarizes the entire location of the ROC curve rather than depending on a specific operating point. It is an effective and combined measure of sensitivity and specificity that describes the inherent validity of diagnostic tests (Nahm, 2022). The AUC value ranges from 0 to 1:

- AUC = 1: Perfect classification,
- AUC = 0.5: No discrimination (i.e. random guessing),
- AUC < 0.5: Worse than random guessing.

A higher AUC indicates better performance of the normality test, as it means the test can better distinguish between normal and non-normal data. It can be approximated by the trapezoidal rule as follows:

$$\text{AUC} = \sum_{i=1}^{n-1} 2\left(\text{FPR}_i - \text{FPR}_{i-1}\right) \cdot \left(\text{TPR}_i + \text{TPR}_{i-1}\right) \tag{4}$$

where FPR_i and TPR_i are the values at each threshold.

3. Methods

To collect GNSS sample data, we deployed an L76X GPS module in a fixed outdoor location. This module, known for its compact design and high positioning accuracy, was connected to a Raspberry Pi, which served as the control system. The Raspberry Pi handled module configuration, data logging, and real-time monitoring, ensuring continuous and reliable data collection throughout the experiment.

Over a continuous period of 96 hours, the GPS module recorded a substantial amount of GNSS data, generating a total of over two million lines formatted in the National Marine Electronics Association (NMEA) standard. This widely used format in GNSS systems organizes navigation and positional data into standardized sentences, making it ideal for subsequent analysis. From the entire recorded dataset, `$GPGGA` sentences, with a specific

type of NMEA message were extracted. These sentences contain key positional and signal quality parameters, including fix quality, satellite count, horizontal dilution of precision (HDOP), and latitude / longitude, all of which are essential for evaluating the GNSS performance. A total of 270 791 rows of `$GPGGA` sentences were isolated from the raw dataset for detailed analysis. Python scripts were utilized to efficiently filter, parse, and structure the extracted data into a log file.

4. Results and discussion

4.1. Graphical data analysis

Figure 1 highlights the analysis of untransformed latitude data. The histogram on the left shows a positively skewed distribution, deviating from the normal curve, indicating non-normality. The Q-Q plot on the right further confirms this by showing significant deviations of data points from the reference line, particularly in the tails. These results emphasize the challenges of applying parametric statistical methods to non-normal latitude data.

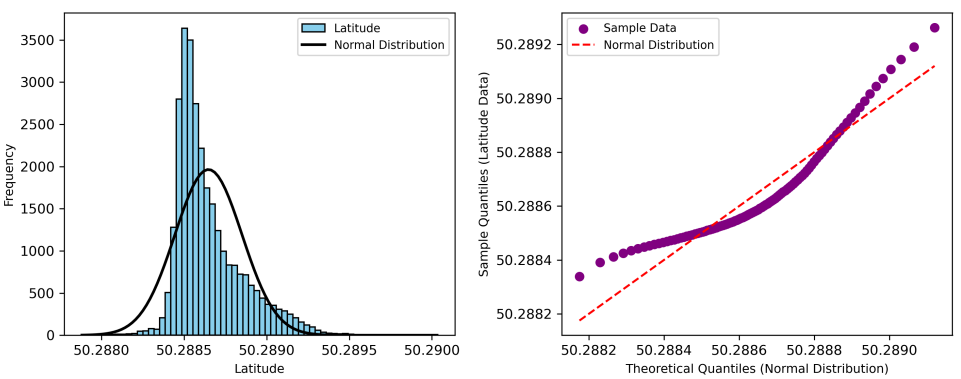

Figure 1: Distribution of untransformed GNSS latitude data histogram plot (left) and Q-Q plot (right)

Figure 2 demonstrates the success of a rank-based inverse normal transformation (INT) in normalizing latitude data. The histogram shows a symmetric, bell-shaped distribution, while the Q-Q plot confirms that the transformed data points closely align with those of a standard normal distribution. These results indicate the INT effectiveness in making the data suitable for parametric statistical methods.

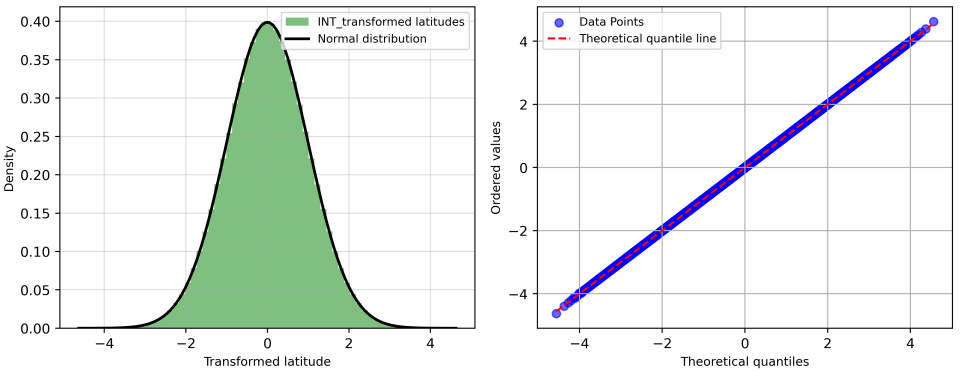

Figure 2: Distribution of transformed GNSS latitude data histogram plot (left) and Q-Q plot (right)

4.2. Descriptive data analysis (skewness and kurtosis)

Table 2 provides a statistical comparison of skewness and kurtosis values between transformed and untransformed datasets, applying rank-based INT, quantile transformation, and log transformation. This comparison underscores the effectiveness of these methods in addressing distributional issues such as skewness and kurtosis, which are pivotal for accurate statistical analysis and enhancing model performance.

Table 2: Comparison of transformed and untransformed descriptive data statistics

Test Type	Transformed data			Untransformed data
	Rank-based INT	Quantile Transform	Log Transform	
Skewness	0.000001	0.019956	1.241654	1.241669
Kurtosis	-0.000215	0.116283	1.961968	1.962022

The rank-based INT transformation achieves a skewness of 0.000 001, effectively neutralizing asymmetry in the data. This indicates an almost perfect symmetric distribution. The quantile transformation reduces skewness to 0.019 956, which is close to zero, reflecting minimal asymmetry. Both the log transform and untransformed data retain significant skewness: 1.241 654 and 1.241 669, respectively. This suggests the data are positively skewed and remain far from a symmetric distribution.

In the case of kurtosis, rank-based INT yields a kurtosis value of -0.000 215, achieving a near-normal distribution by minimizing the impact of extreme values. The quantile transformation also reduces kurtosis to 0.116 283, demonstrating a flattened distribution compared to the untransformed data. However, the log transform shows kurtosis values of 1.961 968, nearly similar to the untransformed data 1.962 022, indicating a heavy-tailed distribution persists.

Overall, the rank-based INT transformation is the most effective method for addressing distributional issues in the dataset, achieving near-normal distribution properties with minimal skewness and kurtosis. Quantile Transformation also performs well but is less ef-

fective than INT. Log transformation retains significant skewness and kurtosis, making it less suitable for analyses requiring normality.

4.3. Statistical and p-value of tests

Table 3 compares statistical test results for transformed and untransformed data to assess normality at a significance level of $\alpha = 0.05$. The untransformed data exhibit significant deviation from a normal distribution, as indicated by high statistical values and consistently low p-values (0.000) across most tests.

Table 3: Comparison of transformed and untransformed statistics and p-value

Test Type	Transformed data		Untransformed data	
	Statistic value	p-value	Statistic value	p-value
Kolmogorov-Smirnov (KS)	0.001	0.964	0.133	0.000
Anderson-Darling (AD)	0.020	0.787(cv)	$9.58\text{x}10^{3}$	0.787(cv)
Lilliefors (LF)	0.000	0.990	0.132	0.001
D'Agostino's K-squared (DA)	0.000	1.000	$5.60\text{x}10^{4}$	0.000
Shapiro-Wilk (SW)	1.000	1.000	0.897	0.000
Jarque-Bera (JB)	0.001	1.000	$1.13\text{x}10^{5}$	0.000
Cramér-von Mises (CM)	0.004	1.000	$9.03\text{x}10^{4}$	0.000
Pearson's Chi-square (Chi2)	$4.12\text{x}10^{5}$	0.000	$5.77\text{x}10^{5}$	0.000

The transformed data consistently produce statistical values near 0.000 and p-values greater than $\alpha = 0.05$, indicating a failure to reject the H0. However, the Chi2 test reports a high statistical value and a p-value below $\alpha = 0.05$, suggesting the H0 is rejected. At a significance level of $\alpha = 0.05$, the AD test's critical value (0.787) exceeds the calculated statistical value (0.020), indicating a failure to reject the H0. All tests effectively validated the normality of the transformed data, confirming no significant deviations from normality, except for the Chi2 test. The p-values of all tests are 0.000, except for the AD test, where the critical value (0.787) is lower than the corresponding statistical value, indicating a rejection of H0 and significant deviations from normality in the case of untransformed data.

4.4. Performance evaluation of normality tests

4.4.1 Comparison of ROC for untransformed latitudes

From Figure 3, the ROC curves are clustered around the diagonal line, with AUC values close to 0.5 for most tests. This suggests that at smaller sample sizes (45, 55, 65, 75, 85, 95, 105), the tests struggle to differentiate between normality and deviations from it, exhibiting limited sensitivity and specificity.

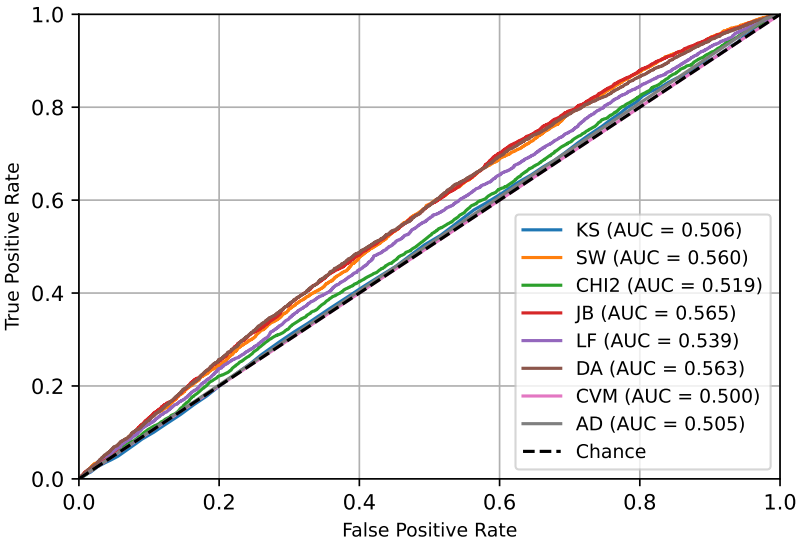

Figure 3: ROC analysis for untransformed GNSS latitude data with a small sample size

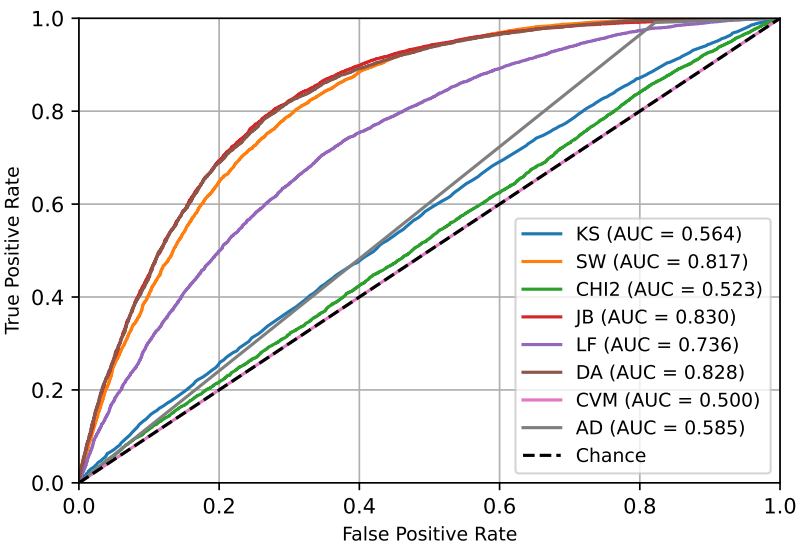

Figure 4: ROC analysis for untransformed GNSS latitude data with a large sample size

Figure 4 describes ROC curve analysis of tests at 0.1% significance level. At larger sample sizes (250, 350, 450, 550, 650, 750, 850, 950, 1050), tests like SW, JB, and DA demonstrate enhanced discriminatory power, becoming more effective at detecting deviations from normality. This improvement is likely due to the increased statistical power associated with larger datasets.

4.4.2 Comparison of ROC for transformed latitudes

Figure 5 and Figure 6 demonstrate the ROC curves and corresponding AUC metrics for various normality tests applied to transformed latitude data for small and large sample sizes at the significance level of 0.1%. From Figure 5, all AUC values are approximately between 0.500 and 0.511, which indicates that the performance of the normality tests is nearly equivalent to random classification. This suggests that the transformations applied to the latitude data were successful in normalizing the data, making it indistinguishable from a truly normal distribution at smaller sample sizes.

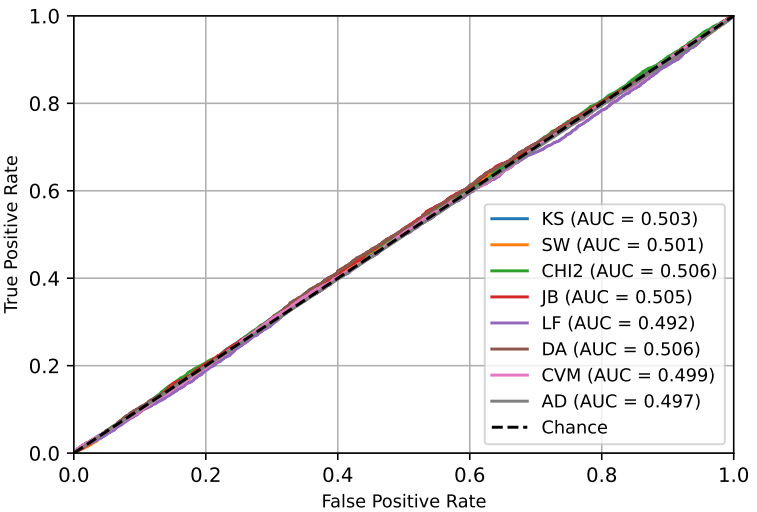

Figure 5: ROC analysis for transformed GNSS latitude data with a small sample size

From Figure 6, all tests have AUC values near 0.5, suggesting that the tests are performing no better than random classification in distinguishing normality from non-normality in the transformed data. This indicates no significant separation between the true positive rate (sensitivity) and the false positive rate. For larger sample sizes, the transformations applied to the latitude data effectively normalize the dataset. If the transformed data closely resemble the synthetic normal data, the tests may struggle to identify significant differences, leading to a high number of false positives and false negatives. This ultimately results in a lower AUC score. An AUC near 0.5 suggests that the test is essentially making random guesses due to the indistinguishability of the data.

A moderate AUC (0.6 - 0.7) implies that the transformation of latitude data into a normal distribution is not entirely successful. Conversely, a high AUC > (0.8) indicates that the transformation has not completely normalized the data, allowing for detectable differences.

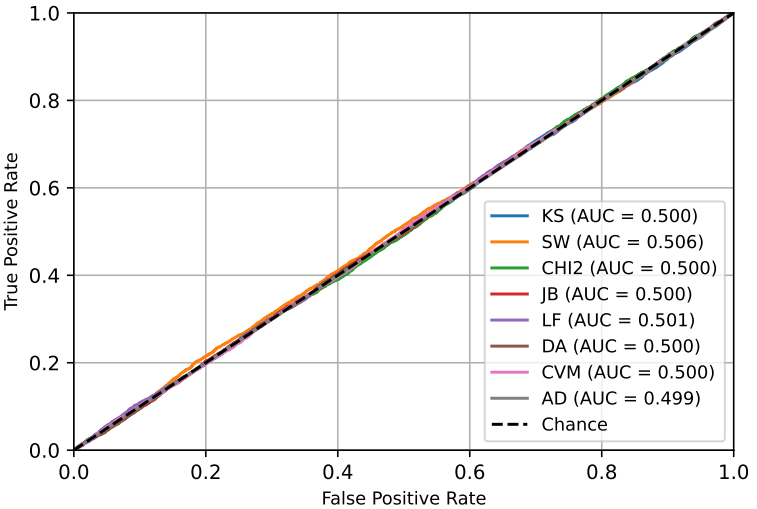

Figure 6: ROC analysis for transformed GNSS latitude data with a large sample size

5. Conclussion

The study emphasizes the importance of assessing and transforming large datasets, such as GNSS measurements, to ensure normality for the validation of parametric statistical tests. The untransformed GNSS latitude data were identified as non-normal using various visual and statistical tests, including histograms, Q-Q plots, skewness, kurtosis, and statistical tests: KS, AD, DA, SW, JB, CVM, Chi2, and LF. Among the transformation techniques, the rank-based Inverse Normal Transformation (INT) demonstrated high effectiveness in enhancing data normality, as validated by various testing methods. The efficiency of statistical tests' was assessed using ROC and AUC analysis, which successfully categorized untransformed data as non-normal and transformed data as normal. These findings underscore the necessity of using tailored transformation methods in large-scale data applications, particularly in geospatial and industrial fields, to enhance the reliability and applicability of parametric statistical methods.

Acknowledgements

This research was funded by the Polish Ministry of Science and Higher Education, grant number 02/050/BKM-24/0042 (AF) and 02/050/BK-24/0032 (JW, AK).

References

Aslam, M., Sherwani, R. A. K. and Saleem, M., (2021). Vague data analysis using neutrosophic Jarque-Bera test. *Plos one*, 16(12), e0260689.

Barba, P., Rosado, B., Ramírez-Zelaya, J. and Berrocoso, M., (2021). Comparative anal-

ysis of statistical and analytical techniques for the study of GNSS geodetic time series. *Engineering Proceedings*, 5(1), p. 21.

Cai, J. and Xu, X., (2024). Bayesian analysis of mixture models with yeo-johnson transformation. *Communications in Statistics-Theory and Methods*, 53(18), pp. 6600–6613.

D'Agostino, R. B., (2017). Tests for the normal distribution. *in* Goodness-of-fit-techniques, Routledge, pp. 367–420.

Demir, S., (2022). Comparison of normality tests in terms of sample sizes under different skewness and kurtosis coefficients. *International Journal of Assessment Tools in Education*, 9(2), pp. 397–409.

Ghasemi, A., and Zahediasl, S., (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10, pp. 486–489.

Huang, Z., Zhao, T., Lai, R., Tian, Y. and Yang, F., (2023). A comprehensive implementation of the log, Box-Cox and log-sinh transformations for skewed and censored precipitation data. *Journal of Hydrology*, 620, pp. 129347.

Khatun, N., (2021). Applications of normality test in statistical analysis. *Open Journal of Statistics*, 11(1), pp. 113–122.

Kumbhar, D. D., Kumar, S., Dubey, M., Kumar, A., Dongale, T. D., Pawar, S. D. and Mukherjee, S., (2024). Exploring statistical approaches for accessing the reliability of y2o3-based memristive devices. *Microelectronic Engineering*, 288, p. 112166.

Kwak, S. G. and Park, S. H., (2019). Normality test in clinical research. *Journal of Rheumatic Diseases*, 26(1), pp. 5–11.

Li, D.-C., Wen, I.-H. and Chen, W.-C., (2016). A novel data transformation model for small data-set learning. *International Journal of Production Research*, 54(24), pp. 7453–7463.

McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. and Lin, X., (2020). Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics*, 76(4), pp. 1262–1272.

Nahm, F. S., (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, 75(1), pp. 25–36.

Obilor, E. I. and Amadi, E. C., (2018). Test for significance of Pearson's correlation coefficient. *International Journal of Innovative Mathematics, Statistics & Energy Policies*, 6(1), pp. 11–23.

Ogaja, C. A., (2022), GNSS data processing. *in* Introduction to GNSS Geodesy: Foundations of Precise Positioning Using Global Navigation Satellite Systems, Springer, pp. 119–134.

Osborne, J., (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, 15(1).

Patrício, M., Ferreira, F., Oliveiros, B. and Caramelo, F., (2017). Comparing the performance of normality tests with roc analysis and confidence intervals. *Communications in Statistics: Simulation and Computation*, 46, pp. 7535–7551.

Peterson, R. A., (2021). Finding optimal normalizing transformations via best normalize. *R Journal*, 13(1).

Pham, L., (2021). Frequency connectedness and cross-quantile dependence between green bond and green equity markets. *Energy Economics*, 98, p. 105257.

Raymaekers, J. and Rousseeuw, P. J., (2024). Transforming variables to central normality. *Machine Learning*, 113(8), pp. 4953–4975.

Razali, N. M. and Wah, Y. B., (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2, pp. 13–14.

Rolke, W. and Gongora, C. G., (2021), A chi-square goodness-of-fit test for continuous distributions against a known alternative. *Computational Statistics*, 36(3), pp. 1885–1900.

Stine, R. A., (2017). Explaining normal quantile-quantile plots through animation: the water-filling analogy. *The American Statistician*, 71(2), pp. 145–147.

Sun, J. and Xia, Y. (2024). Pretreating and normalizing metabolomics data for statistical analysis. *Genes & Diseases*, 11(3), p. 100979.

Tabachnick, B. G., Fidell, L. S. and Ullman, J. B., (2019). *Using Multivariate Statistics*, 7th ed., Pearson.

Uyanto, S. S., (2022). An extensive comparisons of 50 univariate goodness-of-fit tests for normality. *Austrian Journal of Statistics*, 51(3), pp. 45–97.

Von Mises, R., (2014). *Mathematical theory of probability and statistics*, Academic press.

Wilcox, R. R., (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*, Vol. 249, Springer.

Yan, P., (2024). Jackknife test for faulty GNSS measurements detection under non-gaussian noises. *in* 'Proceedings of the 37th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024)', pp. 1619–1641.

Yap, B. W. and Sim, C. H., (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81, pp. 2141–2155.

Zygmonta, C. S., (2023). Managing the assumption of normality within the general linear model with small samples: Guidelines for researchers regarding if, when and how. *The Quantitative Methods for Psychology*.