ORIGINAL ARTICLE

Normality Distributions Commonly Used in Sport and Health Sciences

Fatma Hilal Yagin[1] | **Burak Yagin**[1] | **Abdulvahap Pinar**[2] *

[1]Biostatistics and Medical Informatics, Faculty of Medicine, Inonu University, Malatya, Türkiye
[2]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye

ABSTRACT

This article provides a comprehensive exploration of univariate and multivariate hypothesis tests commonly employed to assess the normality of data distributions. Normality is a foundational assumption in various statistical analyses, rendering the evaluation of data distribution conformity to the normal distribution paramount. In this article, we discuss the principles behind univariate tests, such as the Kolmogorov-Smirnov test, Anderson-Darling test, and Shapiro-Wilk test, as well as multivariate tests, including the Mahalanobis distance and Mardia's multivariate skewness and kurtosis tests. The article aims to aid researchers and practitioners in selecting the most suitable tests for their specific data analysis requirements.

Keywords: Normality, Multivariate tests, Univariate tests

***Corresponding:** Abdulvahap PINAR; apinar@adiyaman.edu.tr
Journal home page: www.e-jespar.com
Academic Editor: Dr. Mehmet Gülü
https://doi.org/10.5281/zenodo.11544808

ARTICLE HISTORY
Received: 11 May 2024
Accepted: 26 May 2024
Published: 01 July 2024

INTRODUCTION

The normal distribution, often referred to as the Gaussian distribution, stands as one of the cornerstones of statistical theory and practice. Its ubiquity in scientific disciplines and practical applications arises from its mathematical elegance and its alignment with real-world phenomena (Yang, Yu, Xie, & Zhang, 2011). Central to its significance is the assumption that many statistical techniques are anchored in the belief that the data under consideration follows a normal distribution. Consequently, assessing the conformity of data to the normal distribution becomes an exercise of paramount importance, carrying far-reaching implications for statistical analyses and their reliability. This article embarks on a comprehensive exploration of the methodologies employed in the validation of this fundamental assumption, delving into both univariate and multivariate hypothesis tests, which constitute the bedrock of such assessments (Garthwaite, Kadane, & O'Hagan, 2005). Understanding the normal distribution, both theoretically and practically, allows for the application of an extensive range of statistical methods that facilitate hypothesis testing, parameter estimation, and inferential analysis (Boylan & Cho, 2012). When data adheres to a normal distribution, the outcomes of statistical tests and estimations tend to be more accurate and robust. Consequently, many statistical techniques, such as t-tests, analysis of variance (ANOVA), and linear regression, assume that the underlying data follows a normal distribution. In the absence of such conformity, the results of these analyses may be skewed, leading to erroneous conclusions. Therefore, the very premise of numerous statistical methods relies on the assumption that data approximates a normal distribution (Hoekstra, Kiers, & Johnson, 2012).

However, in the real world, data seldom conforms perfectly to the idealized assumptions of statistical techniques. In practice, deviations from the normal distribution are common due to various factors, such as sampling variability, measurement error, or the inherent complexity of the investigation. As a result, it becomes necessary to assess the degree to which data adheres to the normal distribution (Dennis, Ponciano, Lele, Taper, & Staples, 2006). This assessment allows researchers and practitioners to make informed decisions about the suitability of a given statistical analysis. To this end, a suite of statistical tools and tests has been developed to rigorously evaluate normality.

In this article, we focus on these evaluation tools, categorizing them into univariate and multivariate hypothesis tests. The selection of these tests depends on the nature of the data, whether it is univariate (comprised of a single variable) or multivariate (involving multiple variables). By providing a comprehensive overview of these tests, we aim to equip researchers and practitioners with the knowledge and tools necessary to make sound decisions about their data analysis strategies.

UNIVARIATE NORMALITY HYPOTHESIS TESTS

Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test assesses whether a given dataset follows a normal distribution by comparing the empirical cumulative distribution function (ECDF) to the theoretical cumulative distribution function of the normal distribution. It calculates the maximum vertical difference between the two functions (Berger & Zhou, 2014). Small and Medium Samples ($n > 50$) The Kolmogorov-Smirnov test is appropriate and usually gives reliable results.

Anderson-Darling Test

The Anderson-Darling test extends the Kolmogorov-Smirnov test by assigning more weight to the tails of the distribution. Used when there are Large Samples ($n \geq 100$). It provides a more robust evaluation of normality, particularly when focusing on deviations in the distribution's tails (Nelson, 1998).

Shapiro-Wilk Test

The Shapiro-Wilk test assesses normality by considering the correlation between the data and the expected values of a normal distribution. It is particularly effective for small sample sizes, used when there is a small sample ($n < 50$), and is a popular choice for univariate normality testing (Hanusz, Tarasinska, & Zielinski, 2016).

Jarque-Bera Test

A variation of the Lagrange multiplier test is the Jarque-Bera test for normality. Many statistical tests, such as the t-test and the F-test, are based on the assumption of normality. To verify normality, the Jarque-Bera test is usually performed before one of these tests. Alternative normality tests (such as Shapiro-Wilk) are used for sample sizes ($n<2000$) and are usually used for large data sets, as they are unreliable for n values larger than 2000. In particular, the test looks for similarities between the skewness and kurtosis of the data and the normal distribution. The data can be in any of the following formats:

- Data in Time Series.
- Regression model errors
- Information in a Vector.

Kurtosis shows how "crested" the distribution is and how much data is in the tails. A distribution that is properly distributed is perfectly symmetrical about the mean and has zero skewness. The test can be carried out without any knowledge of the data's mean or standard deviation (Thadewald & Büning, 2007).

D'Agostino-Pearson Omnibus Test

The skewness (skewness) and kurtosis (slope) values are computed using this approach to determine the degree to which the data resembles a Gaussian distribution. If the data set is not normally distributed, this test uses a combination of kurtosis and skewness statistics to check. If the data is normally distributed, the test statistic has a chi-square distribution with two degrees of freedom. However, data sets with fewer than 20 elements shouldn't often use this test (Widhiarso & UGM, 2012).

Lilliefors Test

The Kolmogorov-Smirnov test can be performed correctly if the distribution is fully specified, i.e. not only normal but also its mean and standard deviation are known. Without specifying which normal distribution the data come from, i.e. without describing the expected value and variance of the distribution, this test examines the hypothesis whether the data represent a normally distributed population. In other words, the Lilliefors test determines whether the data come from a distribution that is part of the Normal family when the population mean and variance are unknown (the sample provides estimates for these variables). In other words, when extrapolating from a sample, some features of the distribution cannot be used by the Kolmogorov-Smirnov test, or at least important tabular values cannot be used. Sample Lilliefors table data are compared when calculating Kolmogorov-Smirnov test statistics (Lilliefors, 1967). The Lilliefors test is particularly suitable for small ($n < 30$) and medium-sized samples ($n = 30$-100). In such samples, the power and accuracy of the test are usually adequate. In large samples ($n > 100$), the results of the test may be more reliable, but there is a risk that the test becomes hypersensitive.

Cramer Von Mises Test

A goodness-of-fit test of this kind is used to ascertain if a given collection of data is regularly distributed or not. This presumption is frequently used in regression, ANOVA, t-tests, and other statistical tests. The biggest difference between the theoretical distribution function (CDF) and the empirical distribution function (EDF) is computed using this test, which is used on ordinal data (Yildirim & Gökpinar, 2012). Medium-sized samples ($n = 30$-100) enhance the statistical power of the Cramér-von Mises test, yielding more robust and reliable results. Such sample sizes are generally adequate for evaluating conformity to the normal distribution. Additionally, when dealing with large samples ($n > 100$), the reliability of the test outcomes is further increased. However, it is crucial to note that with larger samples, the test's sensitivity is heightened, leading to statistically significant results even for minor deviations from the normal distribution. Consequently, in such scenarios, it is essential to interpret the test results not solely based on statistical significance but also considering the practical magnitude and significance of the deviations (Conover, 1999).

MULTIVARIATE HYPOTHESIS TESTS

Mahalanobis Distance

Multivariate data often require multivariate tests. The Mahalanobis distance measures how far an observation deviates from the multivariate mean in units of the data's covariance structure. A large Mahalanobis distance indicates a departure from multivariate normality. For Mahalanobis distance, medium-sized samples (n = 30-100) and larger samples (n > 100) are more appropriate. These sample sizes increase the likelihood of meeting the assumption of multivariate normality, thus providing more reliable results for Mahalanobis distance calculations (Manly & Manly, 1994).

Mardia's Multivariate Skewness and Kurtosis Tests

Mardia's tests assess multivariate skewness and kurtosis, two essential aspects of multivariate normality. High values of these statistics suggest non-normality in the data distribution. These tests are particularly valuable when investigating the normality of multiple variables simultaneously For small samples (n < 50), the statistical power of Mardia's tests can be diminished, potentially limiting the reliability of the results. Mardia's tests tend to be more robust and valid when applied to medium (n = 50-200) and large samples (n > 200). However, with very large samples (n > 500), the tests may become excessively sensitive, leading to situations where even minor deviations from normality are detected as statistically significant (Izenman, 2008).

Henze-Zirkler Test

The Henze-Zirkler test gauges the separation between two distribution functions using a non-negative functional distance. The test statistic is roughly log-normally distributed if the data are multivariate normal. First, the smoothness parameter, variance, and mean are computed. Next, the p-value is estimated after the variance and mean have been log-normalized (Henze & Zirkler, 1990). The Henze-Zirkler test evaluates the adherence of a dataset to a normal distribution. While there's no strict ceiling on sample size, it's typically advised to have a minimum of 50 to 100 samples. Larger samples enhance the test's reliability in representing the population (Hair, 2009).

Multivariate Q-Q Plot Test

A graphical technique for assessing whether a data set adheres to a multivariate normal distribution is the multivariate Q-Q plot. Using the same mean and covariance matrix, it compares the data's quantiles versus those of a multivariate normal distribution. The plot's points will roughly follow a straight line if the data is multivariate normal (Kutner, Nachtsheim, Neter, & Li, 2005).

To ensure the reliability of the Q-Q plot method for assessing normality, it is recommended that the sample size be at least (n ≥ 20). Larger sample sizes generally enhance the robustness of the results, providing a more accurate reflection of the data's distribution. Ensuring an adequate sample size is crucial for minimizing the effects of sampling variability and for detecting deviations from normality more effectively. This threshold helps to ensure that the graphical assessment yields valid and interpretable results, thereby facilitating more accurate statistical inferences (Alpar, 2014).

Doornink-Hensen Test

For multivariate normality, the Doornik-Hansen test is performed using a multivariate data set that has been adjusted to guarantee independence. This test is the multivariate equivalent of the univariate normalcy test, which Doornik and Hansen created in 1994. The test statistic was developed using skewness and kurtosis coefficients. One can convert the multivariate normal distribution into separate standard normal distributions by using population values. This test uses the data's determinant and moment matrices to assess divergence from the normal distribution. It is possible to compute univariate curvature and steepness coefficients using the translated observation values (Doornik & Hansen, 1994). For the Doornik-Hansen test, it is recommended that the sample size be at least 20 to ensure a minimum level of reliability. However, for more robust and dependable results, a sample size between 50 and 100 is advised. Larger sample sizes reduce the influence of sampling variability, enhance the test's statistical power, and provide a more accurate assessment of multivariate normality. Ensuring an adequate sample size is crucial for minimizing Type I and Type II errors, thereby facilitating more precise and valid statistical inference (Doornik & Hansen, 2008).

Choosing the Right Test

Selecting the appropriate test for normality assessment depends on several factors:

Data Type: Univariate tests are suitable for univariate data, while multivariate tests are designed for multivariate data.

Sample Size: Smaller sample sizes may benefit from more robust univariate tests like the Shapiro-Wilk test.

Data Structure: The choice of test may depend on the correlation structure between variables, with multivariate tests being essential when dealing with correlated data.

Practical Considerations

Data Transformation: In cases where data do not meet normality assumptions, transformations (e.g., logarithmic or Box-Cox) can be applied to approximate a normal distribution.

Robust Techniques: When normality assumptions are not met, using robust statistical methods, which are less sensitive to deviations from normality, can be a valid alternative.

CONCLUSIONS

Assessing the normality of data distributions is a fundamental step in statistical analysis. Univariate and multivariate hypothesis tests provide essential tools for researchers and practitioners to evaluate data conformity to the normal distribution. Careful consideration of the data type, sample size, and correlation structure should guide the choice of the most appropriate test. In cases where normality assumptions are not met, data transformation or the application of robust techniques can ensure accurate statistical analysis. This article serves as a valuable resource for those seeking to make informed decisions regarding the assessment of normality in their data.

Author Contributions

Conceptualization, F.H.Y. and B.Y.; methodology, F.H.Y, A.P.; formal analysis, B.Y.; in vestigation, A.P.; data curation, F.Y.H.; writing—original draft preparation, F.H.Y, B.Y., A.P.; writing—review and editing, F.H.Y, B.Y, A.P.

Informed Consent Statement:

Participants took part in the research voluntarily and the research was conducted in line with the Declaration of Helsinki.

Acknowledgments:

We would like to thank all participants who took part in the research.

Funding:

This research was not funded by any institution or organization.

Conflicts of Interest:

The authors declare that no conflicts interest.

REFERENCES

Alpar, R. (2014). Uygulamalı İstatistik ve Geçerlilik-Güvenirlik: SPSS'de Çözümleme Adımları İle Birlikte, 3. *Baskı, Detay Yayıncılık, Ankara*.

Berger, V. W., & Zhou, Y. (2014). Kolmogorov-smirnov test: Overview. *Wiley statsref: Statistics reference online*. doi: https://doi.org/10.1002/9781118445112.stat06558

Boylan, G. L., & Cho, B. R. (2012). The normal probability plot as a tool for understanding data: A shape analysis from the perspective of skewness, kurtosis, and variability. *Quality and Reliability Engineering International, 28*(3), 249-264. doi:https://doi.org/10.1002/qre.1241

Conover, W. J. (1999). *Practical nonparametric statistics* (Vol. 350): john wiley & sons.

Dennis, B., Ponciano, J. M., Lele, S. R., Taper, M. L., & Staples, D. F. (2006). Estimating density dependence, process noise, and observation error. *Ecological Monographs, 76*(3), 323-341. doi: https://doi.org/10.1890/0012-9615

Doornik, J. A., & Hansen, H. (1994). *A practical test for univariate and multivariate normality*. Retrieved from

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford bulletin of economics and statistics, 70*, 927-939. doi:https://doi.org/10.1111/j.1468-0084.2008.00537.x

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American statistical Association, 100*(470), 680-701. doi:https://doi.org/10.1198/016214505000000105

Hair, J. F. (2009). Multivariate data analysis.

Hanusz, Z., Tarasinska, J., & Zielinski, W. (2016). Shapiro-Wilk test with known mean. *REVSTAT-Statistical Journal, 14*(1), 89-100-189-100. doi:https://doi.org/10.57805/revstat.v14i1.180

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in statistics-Theory and Methods, 19*(10), 3595-3617. doi:https://doi.org/10.1080/03610929008830400

Hoekstra, R., Kiers, H. A., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in psychology, 3*, 137. doi:https://doi.org/10.3389/fpsyg.2012.00137

Izenman, A. J. (2008). *Modern multivariate statistical techniques* (Vol. 1): Springer.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*: McGraw-hill.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association, 62*(318), 399-402. doi:https://doi.org/10.1080/01621459.1967.10482916

Manly, B. F., & Manly, B. (1994). Multivariate statistical methods.

Nelson, L. S. (1998). The Anderson-Darling test for normality. *Journal of Quality Technology, 30*(3), 298-299.

Thadewald, T., & Büning, H. (2007). Jarque–Bera test and its competitors for testing normality–a power comparison. *Journal of applied statistics, 34*(1), 87-105. doi:https://doi.org/10.1080/02664760600994539

Widhiarso, W., & UGM, F. P. (2012). Tanya jawab tentang uji normalitas. *Fakultas Psikologi UGM*, 1-5. doi:https://doi.org/10.1080/00224065.1998.11979858

Yang, J., Yu, X., Xie, Z.-Q., & Zhang, J.-P. (2011). A novel virtual sample generation method based on Gaussian distribution. *Knowledge-Based Systems, 24*(6), 740-748. doi:https://doi.org/10.1016/j.knosys.2010.12.010

Yildirim, N., & Gökpinar, F. (2012). Bazı normallik testlerinin 1. tip hataları ve güçleri bakımından kıyaslanması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 16*(1).