Perform Multivariate Normality Tests in Python

Authored by

Mohammed looti

November 7, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Perform Multivariate Normality Tests in Python*. PSYCHOLOGICAL STATISTICS. Retrieved from https://statistics.arabpsychology.com/?p=12112

The Foundational Role of Distributional Assumptions

In the expansive discipline of statistical modeling and inference, the integrity of many widely used parametric tests, such as the ubiquitous t-tests and Analysis of Variance (ANOVA), rests upon a critical, often unspoken, prerequisite: that the underlying data adheres to a normal distribution. This assumption of normality is not merely a formality; it is the mathematical bedrock that validates the formulas used to calculate standard errors, confidence intervals, and p-values.

When researchers proceed with sophisticated statistical procedures without first verifying this fundamental assumption, they risk drawing conclusions that are statistically unsound. Violating the normality assumption, especially in small sample sizes, can severely inflate Type I or Type II error rates, leading to inaccurate modeling and potentially misleading research outcomes. Therefore, the preliminary step of data diagnostics--specifically testing for normality--is an essential safeguard in any rigorous quantitative study.

While a normal distribution is often visualized as a perfectly symmetric, bell-shaped curve defined by its mean and standard deviation, real-world empirical data rarely achieves this theoretical perfection. Consequently, data analysts must deploy a systematic approach involving both exploratory data visualization and formal hypothesis testing to determine if the deviations from ideal normality are statistically significant enough to invalidate the intended parametric model. This rigorous assessment process ensures that the chosen statistical tools are appropriate for the data structure, thereby maintaining the validity and reliability of the research findings.

Distinguishing Univariate from Multivariate Normality

When initial data screening involves assessing a single variable in isolation, the procedure is termed univariate analysis of normality. This process is relatively straightforward and employs several established techniques to scrutinize the shape of the data distribution.

For exploratory visual assessment, the Q-Q plot (Quantile-Quantile plot) remains the gold standard. This graphical tool compares the quantiles of the observed data against the theoretical quantiles of a standardized normal distribution. If the data is truly normally distributed, the plotted points should fall tightly along a 45-degree straight line. Departures from this line, such as S-shapes or heavy tails, indicate issues like skewness or kurtosis, suggesting non-normality.

Beyond visual inspection, univariate normality is formally tested using statistical procedures designed to quantify the difference between the observed distribution and a theoretical normal distribution. Prominent examples include the Shapiro-Wilk test (often preferred for smaller samples), the Kolmogorov-Smirnov test (though less sensitive), and the Jarque-Bera test. These tests yield a p-value that helps determine the likelihood of the data originating from a normal population.

However, the landscape shifts dramatically when we move from single variables to multiple, interdependent variables that must be considered simultaneously. This introduces the concept of **multivariate normality**. Crucially, multivariate normality is a far stricter condition than simply requiring that each variable independently satisfy the univariate normality assumption. A dataset can exhibit perfect univariate normality for every variable yet still fail the multivariate normality test if the variables are related in a non-linear or non-normal fashion.

Multivariate normality implies that not only is the marginal distribution of every component variable normal, but also that any linear combination of these variables is normally distributed, and that the relationships between variables are completely characterized by the mean vector and the covariance matrix. When this collective assumption is violated, the core inferences derived from advanced statistical models become unreliable.

The Critical Need for Multivariate Normality Tests

The requirement for **multivariate normality** is not academic; it is essential for the valid application of several highly influential statistical models routinely used in data science, psychometrics, and experimental research. Models like Multivariate Analysis of Variance (MANOVA), structural equation modeling (SEM), factor analysis, and multivariate regression often assume that the residuals, or sometimes the independent variables themselves, jointly follow a multivariate normal distribution.

When this collective normality is assumed but not verified, the consequences can be severe. In MANOVA, for instance, non-normality can lead to the instability of the covariance matrices, affecting the reliability of test statistics like Wilk's Lambda. In factor analysis, violation of this assumption can complicate model estimation and lead to inaccurate loading interpretations, particularly when using methods based on maximum likelihood estimation.

Detecting deviations from multivariate normality is significantly more complex than detecting univariate non-normality because it must account for the intricate structure of the covariance. The tests must assess the joint distribution across multiple dimensions simultaneously. This necessity mandates the use of specialized **multivariate normality tests** designed specifically to probe the structure of the multivariate space, rather than relying on a collection of inadequate univariate checks.

Furthermore, in the context of identifying outliers in multivariate data, a related metric often employed is the Mahalanobis distance. This measure quantifies the distance of a data point from the center of the distribution, adjusting for the variance and covariance structure. High Mahalanobis distance values flag potential outliers that deviate significantly from the assumed ellipsoidal shape of a multivariate normal distribution, reinforcing its link to the overall normality assessment.

Introducing the Robust Henze-Zirkler (HZ) Test

Among the tools available for assessing multivariate normality, the **Henze-Zirkler Multivariate Normality Test** (often abbreviated as the HZ test) is highly regarded by statisticians for its power and consistency, particularly when dealing with datasets that have a large number of variables (higher dimensions). It is a nonparametric test that provides a robust measure against various types of non-normality, including skewness and kurtosis.

The Henze-Zirkler test operates by comparing the empirical characteristic function of the observed data to the theoretical characteristic function expected under the assumption of multivariate normality. The characteristic function is a unique, complex-valued function that mathematically defines a probability distribution. By measuring the integrated squared distance between the two characteristic functions, the HZ test statistic quantifies the overall departure of the sample distribution from the theoretical multivariate normal distribution.

Like all formal statistical hypothesis tests, the HZ test operates under a defined set of hypotheses. Understanding these is crucial for correct interpretation:

H0 (Null Hypothesis): The set of variables collectively follows a multivariate normal distribution.
Ha (Alternative Hypothesis): The set of variables does *not* follow a multivariate normal distribution.

The outcome of the test hinges on the calculated p-value. A small p-value indicates that the observed data is unlikely if the null hypothesis were true, thus leading to the rejection of H0 and the conclusion that the data is not multivariate normal. Due to its foundation in characteristic functions, the Henze-Zirkler test is known for its computational efficiency and sensitivity across a wide range of dimensionality and sample sizes, making it an excellent choice for modern data analysis.

Preparing the Python Environment and Data Generation

To implement the robust Henze-Zirkler test efficiently in a computational environment, we turn to Python, leveraging the specialized capabilities of the Pingouin statistical package. Pingouin is celebrated for its user-friendly syntax and its clean integration with Pandas DataFrames, offering immediate access to complex statistical procedures, including the HZ test, via the straightforward function `multivariate_normality()`.

Before executing any code, it is essential to prepare your environment. If you are new to this package, the installation is handled quickly using the standard Python package installer:

pip install pingouin

Beyond Pingouin, we require the standard scientific computing libraries: Pandas for structuring and manipulating our data into the necessary DataFrame format, and NumPy for efficient numerical operations, particularly for generating synthetic data for demonstration purposes. Using synthetic data generated from a known distribution allows us to validate the test's performance; if we intentionally create data that *is* multivariate normal, we expect the test to correctly confirm that assumption.

For our demonstration, we will construct a dataset comprising three variables (x1, x2, and x3). We will ensure **multivariate normality** is met by drawing 50 independent observations for each variable directly from NumPy's standard normal distribution function (`np.random.normal`). This synthetic control allows us to anticipate the outcome: the HZ test should conclude that the null hypothesis of normality cannot be rejected.

Executing and Interpreting the Henze-Zirkler Results

With the environment prepared and the synthetic data generated, we proceed to execute the `multivariate_normality()` function, passing the Pandas DataFrame as the primary argument. We must also explicitly set the significance level (alpha), which conventionally defaults to 0.05, representing the threshold for statistical significance.

The following Python code performs the necessary imports, generates the controlled dataset, and applies the Henze-Zirkler test:

 # Import necessary packages for statistical testing and data handling
from pingouin import multivariate_normality
import pandas as pd
import numpy as np

# Create a synthetic dataset ensuring multivariate normality is met
df = pd.DataFrame({'x1':np.random.normal(size=50),
'x2': np.random.normal(size=50),
'x3': np.random.normal(size=50)})

# Perform the Henze-Zirkler Multivariate Normality Test with alpha = 0.05
multivariate_normality(df, alpha=.05)

HZResults(hz=0.5956866563391165, pval=0.6461804077893423, normal=True)

The output, formatted as a named tuple `HZResults`, delivers the metrics required for a statistically sound conclusion. To interpret these findings correctly, we focus on three key elements:

HZ Test Statistic (hz): The calculated value of the Henze-Zirkler test (0.59569). This value quantifies the degree of deviation from the theoretical normal distribution.
p-value (pval): The probability (0.64618) of observing the current data, or data more extreme, assuming the null hypothesis (H0) of normality is true.
Normal (boolean): A direct logical interpretation provided by the Pingouin package based on the comparison of the p-value and alpha (True).

The final decision in statistical hypothesis testing hinges on comparing the calculated p-value against the predefined alpha level (0.05):

If the p-value is less than the alpha threshold ($p < 0.05$), we possess sufficient statistical evidence to reject the null hypothesis (H0), concluding that the data significantly deviates from a multivariate normal distribution.
If the p-value is greater than or equal to alpha ($p \geq 0.05$), we lack sufficient evidence to reject the null hypothesis, meaning we accept the possibility that the data follows a multivariate normal distribution.

In our specific example, the calculated p-value (0.64618) is substantially larger than the **significance level** (0.05). Therefore, we **fail to reject the null hypothesis**. This result confirms our expectation, as the data was deliberately generated from a known normal process. We can confidently proceed with subsequent multivariate statistical analyses, knowing that the foundational assumption of normality has been met.

Understanding and applying the Henze-Zirkler test is an indispensable skill for any analyst working with complex, interdependent datasets, ensuring that the foundation for subsequent statistical modeling is robust and reliable.

Related: *For an academic demonstration of the practical application of the Henze-Zirkler test in medical and biological contexts, consult this research paper, which explores its utility in comparative studies.*