Normality test: Is it really necessary?

Michail Tsagris[a] and Nikolaos Pandis[b]
Rethymnon, Greece, and Bern, Switzerland

Statistical tests used to analyze data require assumptions for the results to be valid. For instance, when performing an independent *t* test for the equality of 2 means, the normality assumption mandates that the populations from which the 2 samples were drawn follow the normal distribution. In linear regression analysis, a basic assumption is that the residuals are normally distributed.

This short article aims to discuss this normality assumption and to answer the following 3 questions: (1) what does the normality assumption mean?; (2) how can this assumption be tested?; and (3) is the testing of the normality assumption important?

When referring to the normality assumption, we literally state that the sample was drawn from a population that was normally distributed. When constructing a histogram of the sample data, we expect to see the bell-shaped curve illustrated in the Figure. The assumption of normality is powerful and implies some nice theoretical properties. For example, certain percentages of the sample observations are distributed symmetrically about the mean. More specifically, 68% and 95% of the data were located 1 and 2 standard deviations above and below the mean, respectively.

Evidently, in a study with a finite sample size, we will only observe the approximations of the Figure. Then, the question of interest is how much our histogram will deviate from this nice theoretical histogram and, more importantly, what constitutes an acceptable deviation from the ideal. To answer this question, researchers usually select a normality test from the large pool of available tests and apply it to their sample. Frequently used tests to assess normality are the Kolmogorov-Smirnov and the Shapiro-Wilk tests.[1] However, the choice of the testing procedure and the procedure itself subsequently can raise more questions regarding the validity

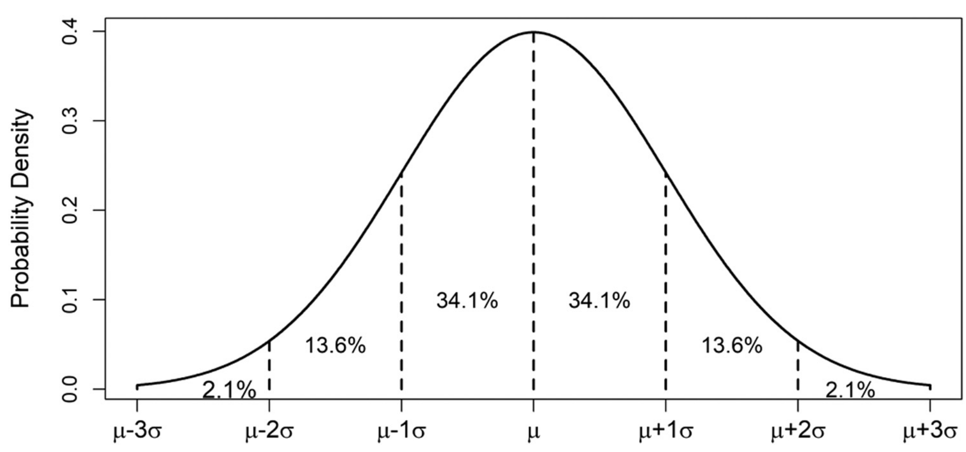

Fig. Histogram of the normal distribution.
Note. The *t* test and the analysis of variance are special cases of a linear regression model in which the dependent variable is categorical.

[a]University of Crete, Rethymnon, Greece.
[b]University of Bern, Bern, Switzerland.
Address correspondence to: Nikolaos Pandis, Department of Orthodontics and Dentofacial Orthopedics, University of Bern, Freiburgstrasse 7, CH-3010 Bern, Switzerland; e-mail, npandis@yahoo.com.
Am J Orthod Dentofacial Orthop 2021;159:548-9
0889-5406/$36.00

https://doi.org/10.1016/j.ajodo.2021.01.003

of the results: (1) with low sample sizes, tests will tend to not reject the normality assumption, even in cases they should, because of low power; (2) with large sample sizes, the tests will reject normality with a high probability even in the presence of small and acceptable deviations from normality; and (3) the results of the normality assumption can differ depending on the test used.

Therefore, accepting or rejecting normality may depend on parameters other than normality itself. Therefore, it is not recommended to rely on such tests for the normality assumption but rather to look at the actual distribution of the data at hand. This approach is likely to be much more informative than statistical testing.

Montgomery[2] stated that the analysis of variance (and related procedures, such as multiple comparisons) is robust to the normality assumption. There is evidence that when the number of observations per variable is greater than 10, violations of the normality assumption often do not noticeably affect results.[1] Tsagris et al[3] investigated the effect of normality on 2 independent samples *t* test via large-scale simulation studies and concluded that even large deviations from the normality assumption did not influence the validity of the *t* test corroborating the previous results.[2]

As stated earlier, our suggestion is that we should ignore testing and focus more on visual assessment. Certain outcomes may inherently follow a particular type of distribution. For example, the number of bracket failures over the treatment period may follow a Poisson distribution. Time to treatment completion may follow a type of distribution called Gamma. If we know that a particular outcome follows a specific distribution, it is probably better to assume this distribution despite what the test to verify the distribution of the data is suggesting.

REFERENCES

1. Schmidt AF, Finan C. Linear regression and the normality assumption. J Clin Epidemiol 2018;98:146-51.
2. Montgomery D. Design and analysis of experiments. 5th ed. New York: John Wiley & Sons; 2001.
3. Tsagris M, Alenazi A, Verrou KM, Pandis N. Hypothesis testing for two population means: parametric or non-parametric test? J Stat Comput Simul 2020;90:252-70.