sciendo

Short Communication

DOI:10.2478/rrlm-2022-0030

Normality assessment, few paradigms and use cases

Călin Avram*, Marius Mărușteri

Department of Medical Informatics and Biostatistics, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania

Abstract

Background: *The importance of applying the normality tests is underlined by the way of continuing the statistical protocol for numerical data within inferential statistics, respectively by the parametric or non-parametric tests that we will apply further on.* ***Methods***: *To check the calculation mode, we used sets of random values and we performed the normality assessment using statistical calculation programs. We took non-Gaussian data (n = 30, n = 50, n = 100, n = 500) and Gaussian data (n = 30, n = 50, n = 100, n = 500) for which we checked the normality of the data. Data chosen for this study were most representative for each batch (n).* ***Results***: *The application of normality tests to the data under study confirms that the data are non-Gaussian for the first data set. For the Gaussian data sample, the verification of normality is confirmed by the results.* ***Conclusion***: *For data up to 50 subjects, it is recommended to apply the Shapiro-Wilk test, but also to apply graphical methods to confirm the accuracy of the result. If the data samples have more than 50 values, the D'Agostino & Pearson omnibus normality test should be applied and if the statistical program does not contain this test, the Shapiro-Wilk test can be applied (in the case of SPSS). Graphical methods, although they require some experience, are useful for identifying the normality of distributions with a small number of data.*

Keywords: *normality test, Gaussian distribution, Q-Q plot diagrams*

Received: 26th May 2022; Accepted: 16th June 2022; Published: 30th June 2022

Introduction

The processing of numerical data with the help of statistical tests must often - depending on the statistical protocol - go through an important stage that leads to a correct interpretation of the results. The stage that must be performed is Normality assessment - the verification of the normality of the data allows the correct identification of the tests that are to be performed further in a statistical protocol (parametric tests or non-parametric tests). Normality assessment can be performed both by graphical methods and by mathematical methods. Graphical methods are useful for more experienced users, while mathematical methods involving statistical tests allow to find results that are easy to interpret by any user. (1).

* **Corresponding author**: Călin Avram, Department Medical Informatics and Biostatistics, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania. E-mail: calin.avram@umfst.ro

Since 1900, there have been a lot of normality assessment tests (2) that allow the identification of Gaussian distributions. To name a few: Shapiro – Wilk test (3); Shapiro – France test; Kolomogorov – Smirnov test (4,5); Anderson – Darling test; Jarque – Bera test; chi-squared test (6); skewness – kurtosis test, and the list could go on. Undoubtedly, some of these tests have shown various corrections over time (7).

Based on the result of the normality test that indicates or not the existence of normal/ Gaussian distributions, various analysis methods can be performed including regression, correlation, comparison of central tendency ,and analysis of variance (8), following - as appropriate - parametric or nonparametric paths of statistical analysis.

A standard Gaussian distribution is one of the most important distributions and has a bell-shaped curve described by its mean and SD, while the extreme values do not have a significant impact on the mean value, namely 68.2%, 95.4%, and 99.7% of the data are contained between mean and ± 1SD, mean and ± 2SD, mean and ± 3SD (1). Various specialists (8,9) consider that, for a sample with over 100 measurements, the violation of normality is not a major problem, however, the performance and identification of data normality should be done regardless of its size (8,9). If the data do not correspond to a Gaussian distribution, then the mean is no longer a representative element for our data set and the representative element will become the median (1). A wrong selection of the representative element of the data will implicitly lead to a misinterpretation. This can be avoided by performing a normality assessment and depending on the test result we choose the representative value of the data. If the data have passed the normality test, then we will apply parametric tests, otherwise the medians are used to compare the groups and therefore we will apply nonparametric tests.

Normality assessment tests

For the data in the paper, we recommend the statistical analysis programs MedCalc (https://www.medcalc.org/), SPSS (https://www.ibm.com/analytics/spss-statistics-software), and GraphPad Prism (https://www.graphpad.com/scientific-software/prism/), where one can find the normality tests that will be presented.

1. The Kolmogorov-Smirnov test is based on the maximum difference between the observed distribution and the expected cumulative-normal distribution. Because it uses the sample mean and standard deviation to calculate the expected normal distribution, the Lilliefors adjustment is used. The smaller the maximum difference, the more likely the distribution is to be normal. This test proved to be less potent than the other tests in most cases (7,10).
2. The Shapiro-Wilk test proved to be the most powerful test in most situations (11,12). It is the ratio of two estimates of the variance of a normal distribution based on a random sample of *n* observations (7,13).
3. The D'Agostino-Pearson test first analyzes the data and determines the asymmetry (to quantify the asymmetry of the distribution) and the vault (to quantify the shape of the distribution). Then, it calculates how much each of these values differ from the expected value with a normal distribution and calculate a single P value from the sum of the squares of these discrepancies (14). This statistical test combines two tests (the kurtosis test and the skewness test) and gives it a chance to detect deviations from normality (15,16).
4. The graphic methods of type Q-Q plot or cumulative frequency (P-P) plots we have represented the two data sets (observed and expected) and are recommended specially to experienced users (1). If the data come from a Gaussian distribution, we should see the

points forming a line that is approximately straight (17).

Material and method

Data collection

To check the calculation mode, we took sets of random values and we performed the normality test using statistical calculation programs (GraphPad Prism vers. 9, SPSS vers. 22.0 and MedCalc vers. 19). The graphical method was performed using the MedCal vers. 19 and SPSS vers. 22.0

For each data set we took a different number of values n=30, n = 50, n = 100, n = 500.

For non-Gaussian data (n = 30, n = 50, n = 100, n = 500) we applied the data generator that is based on the principle of Vale and Maurelli (18), the data represents the value of creatinine from urine at 24 hours. The Gaussian data sets (n = 30, n = 50, n = 100, n = 500) were generated using the MedCalc program and met the following conditions: the mean of the values was chosen 110 and SD = 25 (the data correspond to the measurement of Creatinine in the urine for 24 hours which may have normal values in the 0-200 mg range). Data chosen for this study were most representative for each batch (n).

Applied statistical test

We applied the normality tests from the SPSS statistical program (Kolmogorov –Smirnov test, Lilliefors Sig. Correlation, Shapiro-Wilk test), from the MedCalc statistical program (Shapiro-Wilk test for Normal distribution, Shapiro-Francia test for Normal distribution, D'Agostino-Pearson test for Normal distribution, Kolmogorov-Smirnov test for Normal distribution (Lilliefors significance correction)), and from the GraphPad Prism statistical program (Shapiro-Wilk test, D'Agostino-Person omibus test, Kolmogorov-Smirnov whit test Dallal-Wilkinson -Lilliefor P value). Graphical methods for identifying normality were made using the MedCalc and SPSS.

Throughout the paper we will use the following calculation formulas:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

$$CV\% = \frac{S}{\bar{x}} * 100$$

where x equals the mean, S standard deviation, CV the variation coefficient

Working hypotheses, test elaboration

The following were established in the normality test: null hypothesis (H0) - the data do not differ significantly from a normal distribution. The alternative hypothesis (H1) - data set differs significantly from a normal distribution. The statistical test measures the discrepancy between the data set and the reference one, and, thus, we will obtain the p value (19). If the value *p* (where is the significance level chosen for our test, usually 0.05), then the normality hypothesis is rejected.

Results

For non-Gaussian data on creatinine in urine at 24 hours we have a coefficient of variation between 4% and 8%. The values for the normality check are found in Table 1. The results for verifying the normality of the data for the sets of values (N = 30, N = 50, N = 100 and N = 500) confirm that these data have a non-Gaussian distribution. All tests indicated the correct answer for samples with more than 30 measurements.

Fig. 1 also shows the evaluation of the normality of the data made by graphical methods - Q-Q plot diagrams for the data that do not have a Gaussian distribution. The data were analyzed using the

MedCalc program. It can be seen that the results found in the diagrams below (Fig. 1) confirm the test results in Table 1, the data are non-Gaussian. For the second data set for creatinine in urine for 24-hour, it was chosen so that the data respect a Gaussian distribution, we chose the mean values equal to 110 and the standard deviation (SD) of 25. In this case the data scattering (the coefficient of variation) is 23% with relatively homogeneous data. The verification of normality

Table 1. Values of the normality test for data that do not have a Gaussian distribution

Normality test	N=30 (p)	N=50 (p)	N=100 (p)	N=500 (p)
Normality test with GraphPad Prism program				
Shapiro-Wilk normality test	<0.0001*	<0.0001*	<0.0001*	<0.0001*
D'Agostino & Pearson omnibus normality test	0.0005*	<0.0001*	<0.0001*	<0.0001*
KS normality test with Dallal-Wilkinson-Lilliefor P value	<0.0001*	0.0017*	<0.0001*	<0.0001*
Normality test with SPSS program				
Shapiro-Wilk	0.000*	0.000*	0.000*	<0.000*
Kolmogorov-Smirnov Lilliefors Significance Correction	0.000*	0.002*	0.000*	<0.000*
Normality test with MedCalc program				
Shapiro-Wilk test for Normal distribution	<0.0001*	<0.0001*	<0.0001*	<0.0001*
Shapiro-Francia test for Normal distribution	0.0001*	0.0001*	<0.0001*	< 0.0001*
D'Agostino-Pearson test for Normal distribution	0.0005*	0.0001*	<0.0001*	< 0.0001*
Kolmogorov-Smirnov test for Normal distribution (Lilliefors significance correction)	<0.0001*	0.0017*	<0.0001*	< 0.0001*

*reject Normality

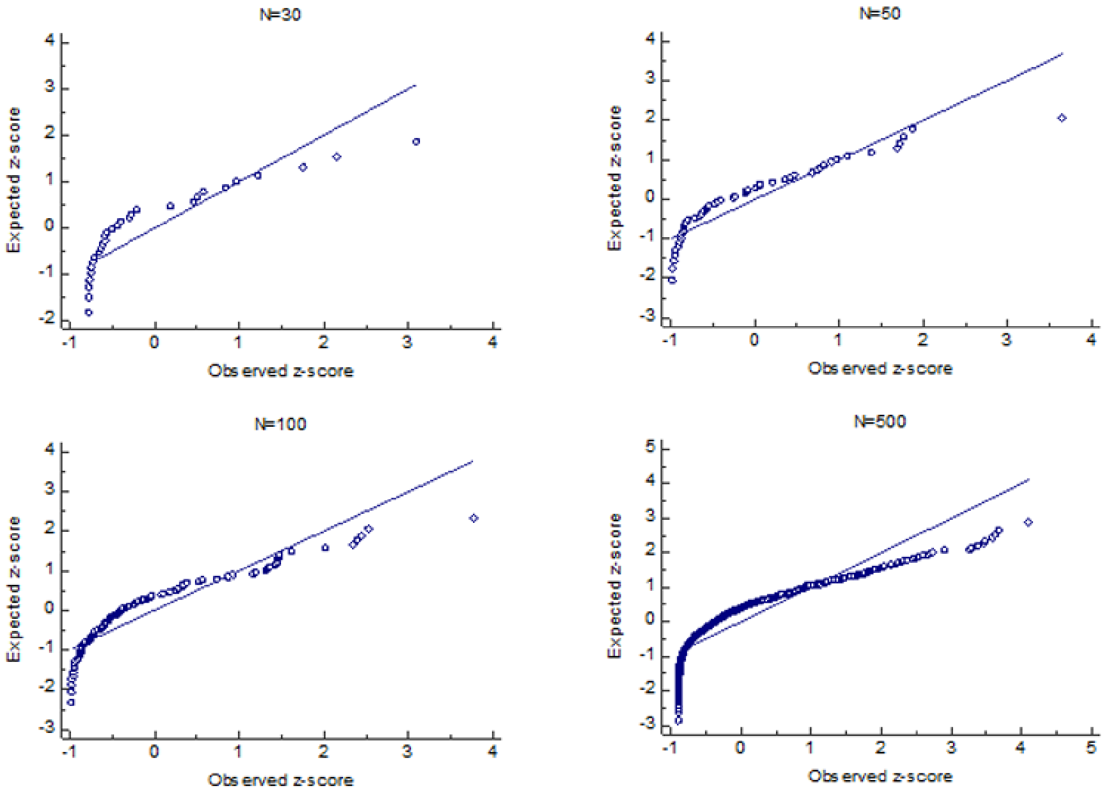

Fig. 1. Q-Q plot diagrams for data with non-Gaussian distribution

was performed with the help of various statistical calculation programs and we obtained the following values, which are found in table 2.
Figure 2 shows the Q-Q plot to verify the normality of the data, this being a graphical method. These data were analyzed using the SPSS statistical program. The results in Table 2 are also confirmed by the Q-Q plot which are in Fig. 2.

Table 2. Values of the normality test for data with Gaussian distribution

Normality test	**N=30 (p)**	**N=50 (p)**	**N=100 (p)**	**N=500 (p)**
Normality test with GraphPad Prism program				
Shapiro-Wilk normality test	0.3711	0.2944	0.6735	0.7365
D'Agostino & Pearson omnibus normality test	0.2256	0.0642	0.4622	0.9699
KS normality test with Dallal-Wilkinson-Lilliefor P value	0.2000	0.2000	0.200	0.200
Normality test with SPSS program				
Shapiro-Wilk	0.371	0.294	0.673	0.736
Kolmogorov-Smirnov Lilliefors Significance Correction	0.200	0.200	0.200	0.200
Normality test with MedCalc program				
Shapiro-Wilk test for Normal distribution	0.3711	0.2944	0.6735	0.7365
Shapiro-Francia test for Normal distribution	0.6037	0.1266	0.7363	0.6327
D'Agostino-Pearson test for Normal distribution	0.2256	0.0642	0.4622	0.9699
Kolmogorov-Smirnov test for Normal distribution (Lilliefors significance correction)	>0.10	>0.10	>0.10	>0.10

*reject Normality

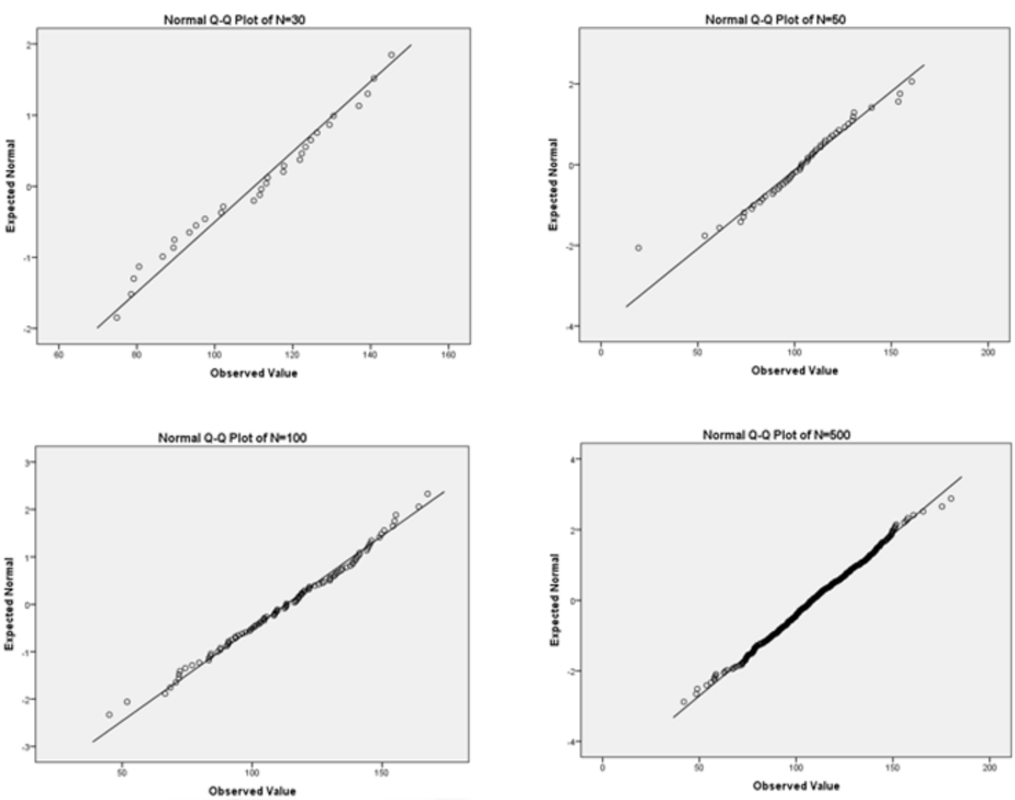

Fig. 2. Q-Q plot diagrams for data with Gaussian distribution

Discussions

The importance of applying the normality assessment is underlined by the way of continuing the statistical protocol for numerical data within the inferential statistics, respectively by the parametric or non-parametric tests that we will apply further on. Depending on the result of the normality assessment, we will apply differentiated statistical methods: parametric statistical tests if the data have a Gaussian distribution or non-parametric statistical tests if we do not have a Gaussian distribution (1,20).

The normality of the data can be verified by various methods, both numerical and graphical, but each method in turn has both advantages and disadvantages. The various methods for verifying the normality of the data involve an important element, namely the number of measurements that are performed (the study sample) (12), and the precision of the result display for the normality test applied.

Statistical tests have the advantage of making an objective assessment of the normality of the data, but they also have the disadvantage that they are not sensitive enough to small data samples or are overly sensitive to large data samples (1).

Depending on the sample size, some specialists consider that for a small sample of data, for example those with fewer than 30 measurements, no normality test should be applied and non-parametric tests should be applied automatically (21,22).

According to the central limit theorem in which we find the words "large numbers" and "approximately" for the normality of the data, no number is identified related to the sample size (23). Because a limit number for the sample is not specified in this theory, we must apply normality tests regardless of the sample size. Although there are various opinions about such a limit, which is 30, for which one should not apply any test of normality and automatically apply non-parametric tests (21,22). This limit was established by various Monte Carlo simulations (from rather old paper), but these simulations were performed on the basis of perfect theoretical samples and in ideal conditions (23), which is unlikely to be met in the case of data collected by various methods. William Gossett (23) applied the t student test for samples smaller than 30 where this condition was imposed, meaning that if the limit of a sample is greater than 30 measurements, then we will never apply this test for small data samples and which meet the conditions of application (they have Gaussian distributions). Likewise, D'Agostino and Person checked the computational power of the normality tests for samples with n = 20 (24). Of course, for some studies a sample of 10 or 20 it is enough, while for other studies a large sample may be insufficient (as in the case of pandemic data) (25). It is very clear that this central limit theorem should be applied on a case-by-case basis (26).

When we have a large data sample, the central limit theorem states that the violation of normality is not a major problem, although, to obtain some significant conclusions it is advisable to assume the normality of the data regardless of sample size (19,27).

For example, to verify the normality of the data, the Shapiro-Wilk test is recommended for small data samples ($n < 50$), although it can be applied just as well for large sample sizes (12,27). Unfortunately, this test, like the others studied, can misidentify the distributions for data sets less than to 30, this test should be supplemented with graphical methods to have a confirmation of the test result (Figure 3).

The Kolmogorov-Smirnov test with Lilliefors significance correction is based on the largest discrepancy between the cumulative sample distribution and the cumulative normal distribution, it is used for samples with $n \geq 50$ (1); this test proved to be less powerful than the other tests in most situations and is included in statistical

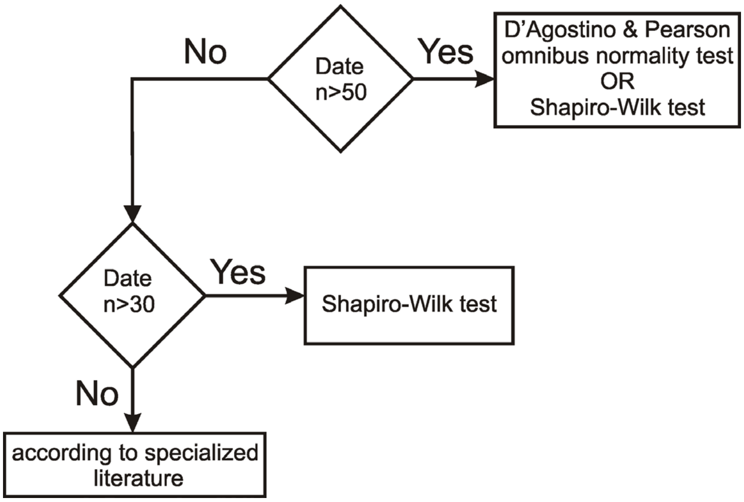

Fig. 3. Logic diagram for applying the normality assessment

programs more due to its popularity and for historical reasons (7).
Another check of data normality can be done with tests that take into account both data skewness and kurtosis, these tests are strong and easy to apply, such a test is D'Agostino & Pearson omnibus normality test (11).
In our opinion, it is recommended that the analytical values be supplemented with a verification of the graphical representation of the data - especially for small data sets (fewer than 30 measurements) - which allows the identification of abnormal values that may influence the test result (27). We must not forget that the graphical and semigraphic methods (CDF, QQ plot), although present in most statistical programs, require a certain degree of user experience in using these methods (1). The simplest of them, the QQ plot method which checks if the points are found on a line, then the normality of the data is accepted, a removal of the points from this right leads to the rejection of normality for the respective data (1,15). These methods allow to visualize the degree of discrepancy of the data in comparison with the theoretical distribution together with the specific values that are distanced, but do not offer an objective measurement of this discrepancy (1). There are other methods that allow the identification of normal data such as Bootstrap Diagnostics (28).
Regardless of the test applied, it is advisable to specify which of these tests have been applied especially in scientific research. In many of the articles, only brief statistical data are specified and people who either review the article or read these papers find it very difficult to identify the normality assessment applied (8).
Normality assessment of samples with 30 values or fewer was not performed in this study and a new research will be made in which we will perform Monte Carlo simulations to obtain evidence-based conclusions.

Conclusions

For numerical data, the part of the normality assessment does not have to be something optional, it is something compulsory and allows the correct identification of the distribution of the data from the analyzed samples, for each sample. For a sample of data of fewer than 30 subjects, the existing international literature recommends that no normality assessment be applied and that the data be considered data that do not have a Gaussian distribution. In our opinion, however, for data up to 50 subjects, it is recommended to apply the Shapiro-Wilk test, but also to apply graphical methods to confirm the correctness of the result. If the data samples have more than 50 values, the D'Agostino & Pearson omnibus normality test should be applied and if the statistical program does not contain this test, the Shapiro-Wilk test can be applied (in the case of SPSS). Graphical methods, although they require some experience, are useful for identifying the normality of distributions with a small number of data.

Authors' contribution

Conceptualization: CA, MM; Methodology CA, MM; Formal Analysis: CA; Writing original draft: CA; Review & Editing, CA, MM.

Conflicts of interest.
There are no conflicts of interests.

References

1. Gupta A, Mishra P, Pandey CM, Singh U, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anaesthesia. 2019;22:67-72. DOI: 10.4103/aca.ACA_157_18
2. Thode HC. Testing for normality. 2002; New York: Marcel Dekker. DOI: 10.1201/9780203910894
3. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52:591-611. DOI: 10.1093/biomet/52.3-4.591
4. Kolmogorov AN. Sulla determinazione empirica di une Legge di distribuzione. Giornale dell'Istituto Italiano Degli Attuari. 1933;4:83-91.
5. Smirnov NV. Sui la distribution de w2 (Criterium de M.R.v. Mises). Comptes Rendus (Paris). 1936;202:449-52.
6. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1900;50:157-75. DOI: 10.1080/14786440009463897
7. NCSS 2020 Data Analysis & Graphics. Normality Tests. 2020; https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Normality_Tests.pdf. 2020 accessed 15 Jan 2021
8. Altman DG, Bland JM. Statistics notes: the normal distribution. BMJ. 1995;310(6975):298. DOI: 10.1136/bmj.310.6975.298
9. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 2012 Spring;10(2):486-9. DOI: 10.5812/ijem.3505
10. SPSS Tutorials: Official Site. SPSS Kolmogorov-Smirnov Test for Normality. 2021 https://www.spss-tutorials.com/spss-kolmogorov-smirnov-test-for-normality/
11. Yap BW, Sim CH. Comparisons of various types of normality tests. J Stat Comput Simul. 2011;81:2141-55. DOI: 10.1080/00949655.2010.520163
12. Noughabi HA, Arghami NR. Monte Carlo comparison of seven normality tests. J Stat Comput Simul. 2011;81(8):965-72. DOI: 10.1080/00949650903580047
13. SPSS Tutorials: Official Site. SPSS Shapiro-Wilk Test for Normality. 2021 https://www.spss-tutorials.com/spss-shapiro-wilk-test-for-normality/
14. Öztuna D, Elhan AH, Tüccar E. Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. Turk J Med Sci. 2006;36(3):171-6.
15. D'Agostino R, Belanger A. A Suggestion for Using Powerful and Informative Tests of Normality. Am Stat. 1990;44(4):316-21. DOI: 10.1080/00031305.1990.10475751
16. MedCalc Manual. Tests for Normal distribution. 2021 https://www.medcalc.org/manual/testsfornormaldistribution.php
17. GraphPad Prism: Tutorials Official Site. QQ plot. 2021 https://www.graphpad.com/guides/prism/latest/statistics/stat_qq-plot.htm
18. Vale CD, Maurelli VA. Simulating multivariate nonormal distributions. Psychometrika. 1983;48(3):465-71. DOI: 10.1007/BF02293687
19. Motulsky H. Intuitive biostatistics: a nonmathematical guide to statistical thinking. 4th Edition ed. New York: Oxford University Press; 2018.
20. Ranganathan P, Gogtay NJ. An Introduction to Statistics - Data Types, Distributions and Summarizing Data. Indian J Crit Care Med. 2019;23(2 Suppl):S169-S170. DOI: 10.5005/jp-journals-10071-23198
21. Simundic AM. Practical recommendations for statistical analysis and data presentation in Biochemia Medica journal. Biochemia Medica. 2012;22:15-23. DOI: 10.11613/BM.2012.003
22. Dawson-Saunders B, Trapp RG. Basic Clinical biostatistics. 4th ed. New York: McGraw-Hill; 2005.
23. Chakrapani C. Statistical Reasoning vs. Magical Thinking. Shamanism as Statistical Knowledge: Is a Sample Size of 30 All You Need? 2011; http://www.chuck-

chakrapani.com/articles/PDF/0411Chakrapani.pdf. accessed 26 May 2021.
24. D'Agostino R, Pearson ES. Tests for Departure from Normality. Empirical Results for the Distributions of b 2 and √b 1. Biometrika. 1973;60(3):613. DOI: 10.2307/2335012
25. Sitanshu K, Ramalingam A. Is 30 the magic number? issues in sample size estimation. Natl J Commun Med. 2013;4:175-9.
26. Fischer H. A History of the Central Limit Theorem: From Classical to Modern Probability Theory. Springer New York: New York, NY, USA. 2011. DOI: 10.1007/978-0-387-87857-7_8
27. Sánchez-Espigares JA, Grima P, Marco-Almagro L. Mosaic normality test. Communications in Statistics - Theory and Methods. 2021;50(23):5561-73. DOI: 10.1080/03610926.2020.1734828
28. Hesterberg T. It's Time To Retire the "n >= 30" rule. Proceedings of the Joint Statistical Meetings, American Statistical Association, Alexandria VA. 2008.