EDITORIAL Open Access

Normality tests for statistical analysis in dentistry: A brief notes for researcher

Mohamad Shafiq Mohd Ibrahim*

Department of Paediatric Dentistry and Dental Public Health, Kulliyyah of Dentistry, International Islamic University Malaysia (IIUM), 25200, Kuantan, Pahang, Malaysia.

***Corresponding author:**
Address:
Department of Paediatric Dentistry and Dental Public Health, Kulliyyah of Dentistry, International Islamic University Malaysia (IIUM), 25200, Kuantan, Pahang, Malaysia.

Telephone: +6019-9431436

Email address:
shafiq@iium.edu.my

How to cite this article:
Mohd Ibrahim, M. S. (2024). Normality tests for statistical analysis in dentistry: A brief notes for researcher. *IIUM Journal of Orofacial and Health Sciences*, *5*(1), 1–3. https://doi.org/10.31436/ijohs.v5i1.290
Article DOI:
https://doi.org/10.31436/ijohs.v5i1.290

Received:
19 February 2024
Revised:
20 February 2024
Accepted:
20 February 2024
Published Online:
29 February 2024

In the realm of dentistry research, statistical analysis is essential to deriving relevant conclusions from gathered data. The normality of the data distribution is one basic assumption that is frequently evaluated. This is important since it helps choose the right statistical tests and guarantees the validity of the findings. These brief notes provide an understanding of numerous normality tests and how to apply them in statistical analysis, making it an invaluable resource for dental researchers. There are multiple ways to assess if a variable follows a normal distribution, including visual inspections and a variety of statistical tests which are not only applicable to dental research but also other research areas. Usually, the normality of a continuous data distribution can be evaluated using several techniques:

1. Statistical tests of normality

There are mainly two types of statistical tests widely utilized for assessing normality of data distribution: the Kolmogorov-Smirnov (K-S) test and the Shapiro-Wilk test (Orcan, 2020). Even though Shapiro-Wilk test can be used with larger sample size, it is generally recommended for small sample sizes (less than 50 samples) (Barton & Peat, 2014). The Kolmogorov-Smirnov test, on the other hand, is usually applied to samples with $n \geq 50$ (Mishra *et al.*, 2019). A p-value > 0.05 in the Shapiro-Wilk and Kolmogorov-Smirnov tests suggests that the data are normally distributed (Ahmad *et al.*, 2019; Ghasemi & Zahediasl, 2012).

2. Mean and median

A standard normal distribution is typically depicted as a bell-shaped curve, exhibiting symmetrical characteristics wherein the values of the mean and median values are roughly comparable. As a result, the large disparities between the mean and median values can help detect possible skewness and indicate a non-normal distribution. However, the small variations between the mean and median values indicate a normal distribution (Barton & Peat, 2014).

3. Standard deviation

The normal distribution is typically 95% of the data usually fall between -1.96 and +1.96 standard deviations from the two standard deviations above and below the mean (Smeeton, 2016). A practical way to check the normality is by adding and subtracting the variable's double standard deviation from the mean. This produces an approximate range that encompasses 95% of the average values. This predicted range, which covers both the minimum and maximum values, should ideally overlap the actual range of data value. According to rule of thumb, a variable is probably not normally distributed if its standard deviation is greater than half of the mean value (Barton & Peat, 2014).

4. Skewness

A skewness values of zero serves as an indicator of perfect symmetry within the standard normal distribution. The tails are to the right when the skewness value is positive, and to the left when it is negative. The range of skewness values, which range from -1 to +1, represents an approximate normal distribution. Where skewness value <-1 or skewness value > +1 denote a moderate degree of skewness and values <-2 or >+2 denote a severe degree of skewness. A deviation from a normal distribution is strongly suggested by skewness values that are more than or equal to +3 or less than -3 (Mishra *et al.*, 2019; Barton & Peat, 2014; Kim, 2013).

5. Kurtosis

A kurtosis value of zero serves as an indicator of perfect symmetry within the standard normal distribution. A distribution that is more peaked than typical is indicated by a positive kurtosis value, whereas a negative kurtosis value denotes that the distribution shape is flatter than normal. Kurtosis values that lie between -1 and +1 are regarded as being close to a normal distribution. Kurtosis values that deviate from the normal distribution are those that fall below -1 and -3 or exceed +1 and +3. On the other hand, a kurtosis values of -3 or greater than +3 strongly suggests that the data is not normally distributed (Mishra *et al.*, 2019; Barton & Peat, 2014; Kim, 2013).

6. Critical values

Critical values or Z scores can be used to assess normality by dividing the skewness and kurtosis by their corresponding standard errors (Barton & Peat, 2014). In most cases, z-values of ±1.96 offer enough evidence to support the data's normality for small sample sizes (n < 50). On the other hand, an absolute z-value of ±3.29 is required for medium-sized samples (50 ≤ n < 300) in order to determine that the sample distribution is normal. The evaluation of normality for sample sizes larger than 300 is based on histograms, where absolute values of kurtosis greater than seven and skewness greater than two may be used as references for non-normal distributions (Kim, 2013).

7. Histograms and plots

A histogram is the most commonly used graph to visually represents frequency distributions and whether the shape of the continuous data's distribution resembles an approximately normal bell curve (Bluman, 2018). If the graph is exhibiting approximate normal distribution the data is normally distributed. The normal Q-Q plot illustrates by plotting each data point against the expected value for a normal distribution. In an ideal scenario of normal distribution, the plotted points would align precisely along a straight line, and deviations from this line signal varying degrees of non-normality. The variations of the points from the straight line seen in the normal Q-Q plot are shown by the detrended normal Q-Q plots. All points in a normal distribution are uniformly distributed above and below the horizontal line at 0 and randomly cluster about it. In contrast, patterns like a J or an inverted U distribution are displayed by non-normal distributions (Barton & Peat, 2014).

8. Box plot and extreme values

A boxplots or box-whisker diagram is a very helpful tool for summarising a set of data. The boxplot's shape displays distribution of the data by five-number summary (minimum value, first quartile, median, third quartile and maximum value) and any extreme values or outliers. At the center of the box plot is the median and interquartile range as the length of the box which the middle 50% of the observations fall. The whiskers, which represent the minimum and highest values within 1.5 times the interquartile range, are the lines that extend from the top and bottom of the box. Extreme values (asterisks) and outliers (circles) are used to represent values that fall outside of this range. The variable is unlikely to show signs of a normal distribution if there are a lot of extreme values present or if the median is not in the center of the box. Furthermore, the position of the median in the box plot at the bottom of the box indicates positive skewness, while it indicates negative skewness if the median is at the top of the box (Mishra *et al.*, 2019; Barton & Peat, 2014).

In summary, these are the ideal steps to take for consideration whether the data approximates a normal distribution to support the use of parametric tests or a non-normal distribution to support the use of non-parametric tests. However, the researchers tend to choose Shapiro-Wilk test or Kolmogorov-Smirnov (K-S) test to obtain numerical evaluations of normality. Ideally, some of the above steps are still important to be considered to further confirm the normality or non-normality of the data distribution.

Overall, normality test is part of parcel of statistical analysis in research including dentistry. Therefore, it is imperative to understand the ideal steps when performing normality test in order to select the correct statistical tests that suits the available data.

References

Ahmad, W.M.A.W, Ibrahim, M.S.M, Halim, N.A., Mohamad, N.A.A. (2019). *Statistical Analysis Using SPSS Version 24*. Pulau Pinang: Penerbit Universiti Sains Malaysia.

Barton, B., Peat, J. (2014). *Medical Statistics: A Guide to SPSS, Data Analysis and Clinical Appraisal.* 2nd ed. Sydney: Wiley Blackwell, BMJ Books.

Bluman, A.G. (2018). *Elementary Statistics, a Step-by-Step Approach*. 10th ed. New York: McGraw-Hill.

Ghasemi, A., Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489.

Kim, H.Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics*, 38(1), 52-54.

Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C., Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1), 67-72.

Orcan, F. (2020). Parametric or non-parametric: skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education*, 7(2), 255-265.

Smeeton, N.C. (2016). *Dental Statistics Made Easy*. 3rd ed. New York: Chapman, and Hall Book/CRC, Taylor & Francis Group.