

## Letter to the Editor



Received: Aug 24, 2023

Revised: Dec 12, 2023

Accepted: Jan 16, 2024

Published online: Apr 12, 2024

### Citation

Das S. It is normal to be “not-normal”:  
reporting of correct descriptive statistics  
in dental research. Restor Dent Endod  
2024;49(2):e19.

### \*Correspondence to

**Snigdho Das, MDS**

Department of Dentistry, Ramkrishna Sarada  
Mission Matri Bhawan Hospital, 7A, Sree  
Mohan Lane, Kolkata-700026, WB, India.  
Email: snigdho1991@gmail.com

Copyright © 2024. The Korean Academy of  
Conservative Dentistry

This is an Open Access article distributed  
under the terms of the Creative Commons  
Attribution Non-Commercial License ([https://  
creativecommons.org/licenses/by-nc/4.0/](https://creativecommons.org/licenses/by-nc/4.0/))  
which permits unrestricted non-commercial  
use, distribution, and reproduction in any  
medium, provided the original work is properly  
cited.

### Conflict of Interest

No potential conflict of interest relevant to this  
article was reported.

### ORCID iDs

Snigdho Das <https://orcid.org/0000-0002-9061-748X>

# It is normal to be “not-normal”: reporting of correct descriptive statistics in dental research

**Snigdho Das** \*

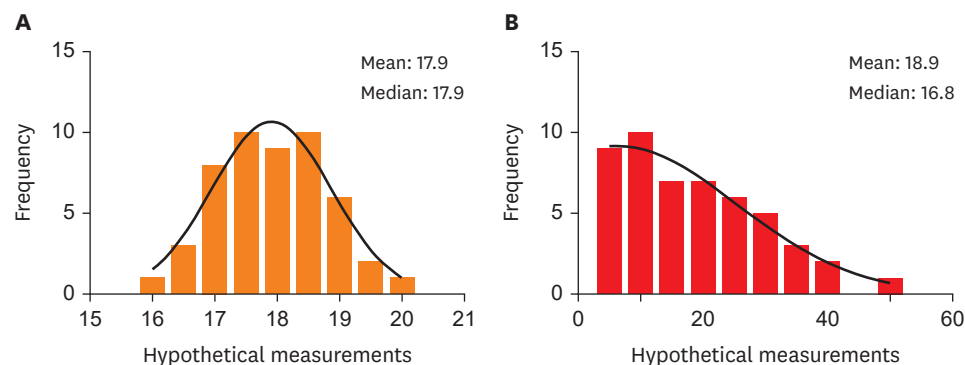
Department of Dentistry, Ramkrishna Sarada Mission Matri Bhawan Hospital, Kolkata, West Bengal, India

Respected Editor-in-Chief

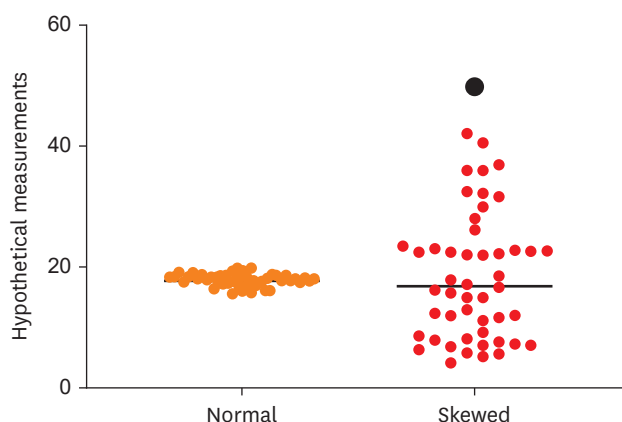
Statistical analysis forms a crucial part of any research and aids in decision-making. It has been observed that data from health research usually do not follow the standard “normal distribution,” and thus are skewed [1]. However, of late there has been a rise in the incorrect reporting of statistical indicators, which ultimately obscures the validity of the research. A normal distribution refers to data where the observations lie symmetrically around an unknown point, forming a bell-shaped curve, whereas a non-normal distribution may lack symmetry, may have extreme values, or may have a steeper or flatter “dome-shape” than a typical bell.

It has been noted that researchers without evaluating the nature of the distribution, *i.e.*, whether the distribution concurs with a normal distribution, report the descriptive statistics using mean and standard deviation as a measure of central tendency and dispersion respectively. Also, they utilize tests meant for analyzing “normally” distributed data. However, these descriptive statistics may not be applicable and often misleading for skewed data, as it is highly influenced by extreme values. **Figure 1** exemplifies two hypothetical data sets plotted in a histogram showing normal distribution and skewed distribution.

In such cases, when observations do not pile up a common point, and evident “outliers” are present, it is desirable to report the measure of central tendency as median as it is not



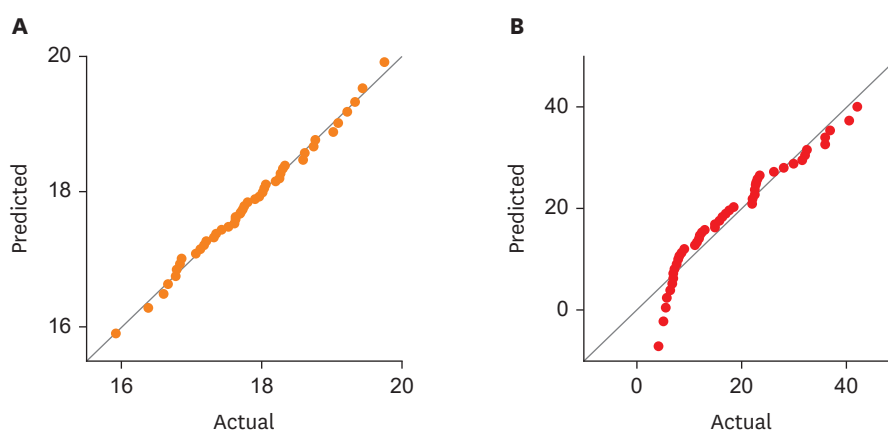
**Figure 1.** Histograms of normally distributed and skewed data (hypothetical data generated digitally with  $n = 50$  each). (A) Normally distributed data and (B) skewed data.



**Figure 2.** Scatter Dot Plot showing the data dispersion in normally distributed data (blue) and skewed data (red) with the black horizontal line at the median. Note the amount of clustering of observations around the median value in the case of normally distributed data and the amount of dispersion in skewed data with an outlier (black dot).

influenced by extreme values [2]. Outliers tend to make the distribution highly dispersive which has been depicted in **Figure 2**. The measure of dispersion can be reported in terms of the interquartile range. It is calculated by dividing the observations into quartiles, where the first quartile denoted by Q1 holds 25% of the values below it and the third quartile (Q3) holds 75% of the data below it, and taking the difference between the first and the third quartile (Q1–Q3) [3].

It is recommended that researchers report the nature of distributions of their variables, which can be either done by eye-balling at the simple histogram or plotting a normal distribution or a quantile-quantile (Q-Q) plot. **Figure 3** illustrates Q-Q plots showing observations from normally distributed data and skewed data. Formal tests to assess normality like the Kolmogorov-Smirnov tests or the Shapiro-Wilk test can also be carried out with the null hypothesis that the variable follows a normal distribution and  $p$  values above the threshold (usually 0.05) suggest normality. However, one must remember that these tests are highly sensitive to extremely large samples and should be used with histograms and normality plots [2].



**Figure 3.** Quantile-Quantile(Q-Q) Plot of two hypothetical data sets. (A) Normally distributed data and (B) skewed data. Note that observations lie on or around the theoretical normal line in the case of normally distributed data, while in skewed data, the observations form a curve or an “S”-shape around the theoretical normal line.

Researchers must acknowledge that summarizing measures applicable for “normally” distributed data may be misleading when pragmatized to non-normal distributions, and reporting the correct indicators may augment the quality of research and significantly contribute to the reduction of the production of low-quality articles.

## REFERENCES

1. Bono R, Blanca MJ, Arnau J, Gómez-Benito J. Non-normal distributions commonly used in health, education, and Social Sciences: A systematic review. *Front Psychol* 2017;8:1602. [PUBMED](#) | [CROSSREF](#)
2. Sainani KL. Dealing with non-normal data. *PM R* 2012;4:1001-1005. [PUBMED](#) | [CROSSREF](#)
3. Madadzadeh F, Ezati Asar M, Hosseini M. Common statistical mistakes in descriptive statistics reports of normal and non-normal variables in biomedical sciences research. *Iran J Public Health* 2015;44:1557-1558. [PUBMED](#)