

## EDITORIAL

# Biostatistics Series – 2: Descriptive Statistics

Gangaboraiah Bilagumba

<sup>1</sup>Professor Emeritus and Research Head,  
Dayananda Sagar College of Dental Sciences,  
Karnataka, India

## Correspondence

Gangaboraiah Bilagumba, Dayananda Sagar  
College of Dental Sciences, Bangalore.  
Email: gangaboriah@dscds.edu.in

## Abstract

This editorial provides a comprehensive overview of descriptive statistics, focusing on how to effectively summarize quantitative variables. It details the primary measures of central tendency—mean, median, and mode—and emphasizes that choosing the correct measure depends heavily on data normality and the presence of skewness. Because central tendency alone provides an incomplete picture, the text also explores essential measures of dispersion, including standard deviation, variance, range, and interquartile range (IQR). It outlines practical methods for evaluating data distribution, testing for normality using non-parametric tests, and detecting outliers via Tukey's fences and Box-and-Whisker plots. Additionally, the document defines the fundamental scales of measurement (nominal, ordinal, interval, and ratio), alongside specialized scoring systems like the Likert scale for evaluating knowledge, attitude, and practice. Finally, it summarizes the key mathematical properties, strengths, and limitations of these statistical measures.

## KEYWORDS

Descriptive Statistics, Measures of Central Tendency, Measures of Dispersion, Data Normality, Skewness, Outlier Detection, Scales of Measurement

## 1 | INTRODUCTION TO DESCRIPTIVE STATISTICS

Data given in Table 1 and Table 2 relates to the BPI of subjects with and without Alzheimer disease (AD). From a Statistical point of view, is a numerical measurement of BPI among subjects with and without AD. So, it is based on quantitative variable and hence it needs to be described statistically using summary statistics like mean, median, range, standard deviation, inter quartile range etc.

## 2 | MEASURES OF CENTRAL TENDENCY

The computation of mean, median, and mode which are called as Measures of Central Tendency describes the central location of the data. Particularly, median is called as positional average because it finds the exact central position in the data.

Now, the question is which of the three appropriate measures viz., mean, median, and mode to be chosen to describe the data. This will be explained a little while after defining other measures like standard deviation. (Figure 1)

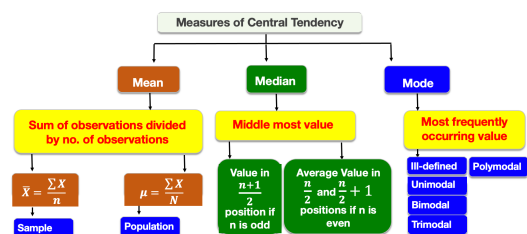


FIGURE 1 Measures of Central Tendency

## 3 | MEASURES OF DISPERSION

For the above data, the mean, median, and mode for both the groups is 41.00. Does this show, the distribution of data in both groups are same? The answer is no, because all the three measures viz., mean, median, and mode indicate only central location, but the spread in the data is not explained by them. This shows the description of data is incomplete without the other measure of dispersion like for example standard deviation. The following are the measures of dispersion. (Figure 2)

For the data on DHA, the standard deviation (SD) is respectively 9.29, 7.42 and 2.01. Now what do we understand by these values of mean of 41.00 and SDs of 9.29, 7.42 and 2.01. As mentioned before mean is pointing out the central position, may not be exactly center but in

**TABLE 1** BPI of subjects with and without Alzheimer disease (AD)

|          |       |       |       |       |       |      |      |      |
|----------|-------|-------|-------|-------|-------|------|------|------|
| SI No    | 1     | 2     | 3     | 4     | 5     | 6    | 7    | 8    |
| BPI - G1 | 100.0 | 100.0 | 100.0 | 21.8  | 100.0 | 17.8 | 64.2 | 43.7 |
| BPI - G2 | 60.7  | 41.1  | 35.0  | 37.0  | 26.3  | 40.0 | 12.5 | 19.4 |
| SI No    | 9     | 10    | 11    | 12    | 13    | 14   | 15   | 16   |
| BPI - G1 | 100.0 | 21.8  | 100.0 | 100.0 | 100.0 | 75.0 | 90.0 | 91.6 |
| BPI - G2 | 15.6  | 25.0  | 31.2  | 60.7  | 27.7  | 39.2 | 50.0 | 53.5 |
| SI No    | 17    | 18    | 19    | 20    | 21    | 22   | 23   | 24   |
| BPI - G1 | 76.7  | 95.0  | 100.0 | 93.7  | 100.0 | 28.6 | 45.0 | 42.0 |
| BPI - G2 | 25.0  | 18.8  | 18.7  | 21.8  | 21.8  | 31.2 | 35.5 |      |

**TABLE 2** Baseline and Follow-up DHA measurements

|                 |       |       |       |       |       |       |       |       |       |       |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Baseline DHA    | 37.79 | 29.99 | 28.78 | 47.09 | 41.00 | 48.58 | 41.00 | 29.01 | 52.55 | 54.21 |
| DHA at 1 month  | 33.26 | 30.76 | 54.69 | 40.85 | 41.00 | 41.43 | 41.00 | 34.07 | 42.27 | 50.67 |
| DHA at 3 months | 41.00 | 38.50 | 41.88 | 39.33 | 42.80 | 41.00 | 38.16 | 40.00 | 43.85 | 43.48 |

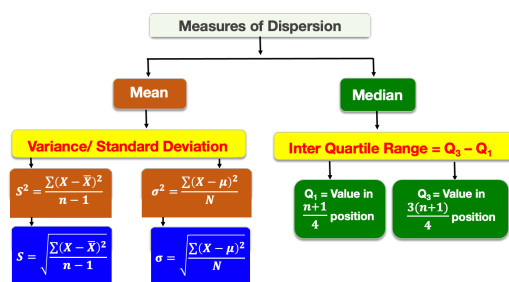


**FIGURE 2** Measures of Dispersion

and around it, whereas the SD explains the average spread around the mean. Lower the SD implies more similarities in the data and larger SD, the wide spread. In the above example, the DHA at baseline has an SD of 9.29 showing wider spread in the data, followed by DHA at 1 month with an SD of 7.42 and the least SD for DHA at 3 months.

## 4 | NORMALITY AND SKEWNESS

Now that, the mean, median and mode are same, which one to be used in the description of data. For this, first we need to understand the skewness. Skewness is defined as deviation from normal or asymmetric. Here normal refers to bell shaped curve. If the data follows a bell shape or near bell shape will be called as normal data. (More about normal, let us lean later under normal distribution). Negative skewness: A long tail to the left with fewer values and majority of the distributed to the right. In this case Mean < Median < Mode. (Figure 3,4)

Positive skewness: A long tail to the right with fewer values and majority of the distributed to the left. In this case Mean > Median > Mode

To decide to choose between mean and median, the data should be checked for normality using a non-parametric statistical (let us learn this in more detail later) tests like, Kolmogorov-Smirnov test, Shapiro-Wilk or D'Agostini-Pearson omnibus normality test. In the on DHA, incidentally,

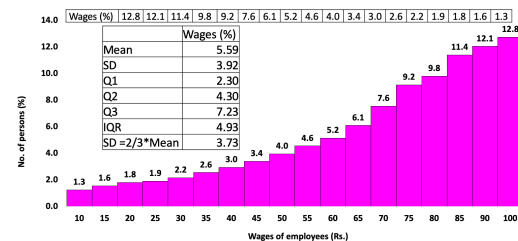


**FIGURE 3** Negative skewness showing Mean < Median < Mode

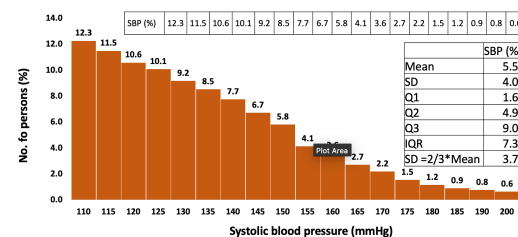


**FIGURE 4** Positive skewness showing Mean > Median > Mode

all of them are found to be normal ( $P > 0.05$ ). However, in the AD study, the data of Group 1 is non-normal ( $P = 0.0001$ ) and of Group 2 is normal ( $P = 0.14.13$ ).

## 5 | OUTLIER DETECTION

It is important to note here always that, if the data is normal, then both mean and SD must be used together for the complete description (not mean alone or SD alone) and when it is non-normal (usually  $SD > 0.667 * \text{mean}$ ), the median and inter quartile range (IQR) should be used to describe the data. Any possible outliers can be detected by Box-and-Whisker plot. (Figure 5)

Source: Unknown. For the detection of outliers, the fences are computed using the Tuckey's method given by: Lower Inner Fence (LIF):  $Q1$

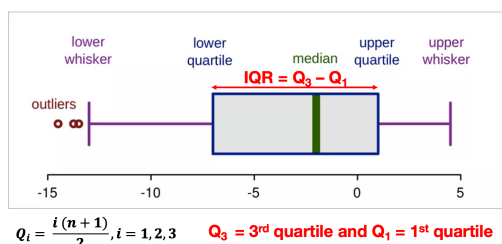


**FIGURE 5** Box-and-Whisker plot for outlier detection

-  $(1.5 \times IQR)$  – values below this are mild lower outliers; Upper Inner Fence (UIF):  $Q3 + (1.5 \times IQR)$  – values above this are mild upper outliers; Lower Outer Fence (LOF):  $Q1 - (3 \times IQR)$  – values below this are extreme lower outliers; Upper Outer Fence (UOF):  $Q3 + (3 \times IQR)$  – values above this are extreme upper outliers. Besides SD and IQR, Range, Variance and Coefficient of Variation are appropriately also used to measure the spread of data.

## 6 | SCALES OF MEASUREMENT

For measuring the statistical data, the following scales are to be made use.

| Scale    | Criteria |       |       |              |       |                 |                  |
|----------|----------|-------|-------|--------------|-------|-----------------|------------------|
|          | Identn.  | Order | Dist. | Start. Point | Ratio | Stat. measure   | Exam.            |
| Nominal  | Yes      | X     | X     | X            | X     | Proportion Mode | Oral hygiene     |
| Ordinal  | Yes      | Yes   | X     | X            | X     | Median          | Physic. activity |
| Interval | Yes      | Yes   | Yes   | Arbitrary    | X     | All             | Temperature      |
| Ratio    | Yes      | Yes   | Yes   | Fixed        | Yes   | All             | DMFT             |

**FIGURE 6** Scales of measurement: Nominal, Ordinal, Interval, and Ratio

In addition to the above four scales, Knowledge is measured using (0 – 1) scale, wherein for every wrong answer assign a score of '0' and for correct answer a score of '1'. The Attitude will be measured by 5-point Likert scale (Strongly disagree – 1, Disagree – 2, Neutral – 3, Agree – 4 and Strongly agree – 5 for positive attitude and for negative attitude the scores will be reversed). The practice is also measured by (0 – 1) scale with wrong practice assigning '0' and correct practice '1'. (Figure 6)

## 7 | PROPERTIES OF STATISTICAL MEASURES

Some properties on mean, median, mode and SD. For mean: It measures stability. Mean is the most stable among other measures of central tendency because every score contributes to the value of the mean. It is

rigidly defined and therefore suitable for further mathematical analysis. It may easily be affected by the extreme scores (outliers). The sum of each score's distance from the mean is zero  $(\sum (X - \text{Mean}) = 0)$ . It can be applied to interval and ratio level of measurement. It may not be an actual score in the distribution. It is very easy to compute. For median: It may not be an actual observation in the data set. Not affected by extreme values because median is a positional measure. Can be applied in ordinal level. For Mode: Not affected by extreme values. There may be an ill-defined Mode (all values repeat once). There may also be several modes. Used for either Numerical or Categorical data. For SD: Non-Negativity: Standard deviation is always greater than or equal to zero ( $\sigma \geq 0$  or  $s \geq 0$ ). It can never be negative because it is derived from the square root of squared distances. Zero Variance: The standard deviation is exactly zero if and only if all data values are identical. This indicates that there is no spread or variation in the data. Sensitivity to Outliers: Because the formula squares the deviations from the mean, extreme values (outliers) heavily impact the standard deviation, inflating the result even if the rest of the data is tightly grouped. Same Units: Unlike variance (which is measured in squared units), standard deviation is expressed in the exact same units as the original data (e.g., meters, dollars, kilograms), making it highly interpretable.

## ACKNOWLEDGMENTS

Data source courtesy of Dr. Savita A M, Professor and HoD of Periodontology, DSCDS, Bangalore.

### How to cite this article:

Bilagumba G. Biostatistics Series – 2: Descriptive Statistics. World J Dent Excell. 2026;2(1):1.