This is the second article of a series on fundamental concepts in biostatistics and research. In this article, the author reviews the manner in which researchers characterize data. Normality, standard deviation, mean, P value, and other concepts related to parametric statistics are discussed in common language, with a minimum of jargon and mathematics, and with clinical examples. Emphasis is given to conceptual understanding.

**Keywords:** mean; median; mode; average; normal distribution; standard deviation; p-value; statistics

Received: 25 June 2015; Accepted: 3 August 2015; Published: 19 October 2015

In part one of this article (1), we discussed the fact that groups of measurements for a given value can be characterized by the center of the cluster of measurements (frequently represented by the mean) and the dispersion of measurements around the center. Dispersion is characterized in most simple terms by ‘range’, the distance between the least and the greatest value. In this article, we will discuss ‘standard deviation’ (SD) and how it can be used to simplify discussions of variability. Central to this discussion is an understanding of ‘normal distribution’.

**Normal distribution**

Many biological measurements have a typical pattern of dispersion: normal distribution. A group of measurements that has a normal distribution may be described as ‘normal’ or as having the characteristic of ‘normality’. A normal distribution (sometimes also described as a Gaussian distribution or bell curve) is useful to researchers. Normal distributions are predictable, and scientists can calculate what percent of a group of values lies between various boundary values.

In a normal distribution, the most frequent value (mode) is the average value (the mean). Both values are equivalent and lay dead center in the middle of the grouping. To say this in another way, 50% of the measurements will lie above the mean (and mode) and 50% will lie below. As one moves away from the middle value (the most frequent value) the frequency decreases continuously at a variable but predictable rate. Furthermore, the distribution is symmetrical. For example, the frequency of RBCs 5 fL less in volume than the mean volume will be equivalent to the frequency of RBCs 5 fL greater than the mean volume. Such symmetry is the same for any given value, any distance from the midpoint (2).

Researchers are always careful to assess normality. Various tests have been devised to assess whether or not a group of values is normally distributed. The assurance of normality is important because if the values are not normally distributed then the above described characteristics may not be present, and the use of calculations based on normality may lead to incorrect conclusions. Tests and calculations are referred to as being ‘parametric’ if their validity is based on the assumption that values are normally distributed. Common parametric tests are the T test and analysis of variance.

**SD as a measure of variability in normal distributions**

When describing normal distributions, researchers use the SD to describe variability rather than the simple measurement of ‘range’. Researchers prefer the SD for at least two reasons. First, its standardized unit can be used to describe diverse groups with different underlying units of measure (e.g., centimeters vs. inches). Second, the use of SD allows researchers to predict the frequency of measurements above, below, and within a set of boundary values. For example, it is a commonly used approximation that 95% of measurements will fall within 2 SD of the center (the mean).

The SD may be considered to be a type of average distance (deviation) from the center for all measures. The following example will allow you to understand the
The differences between the mean and each of these values are respectively (in fL):

\[3, 2, 1, 0, 1, 2, 3\]

The sum of the differences from the mean is:

\[3 + 2 + 1 + 0 + 1 + 2 + 3 = 12\]

The average difference (deviation) from the mean is:

\[\frac{12}{7} = 1.7 \text{ fL}\]

Consider how this value (1.7 fL) could be used if it were the actual SD. Remember that a useful ‘rule of thumb’ approximation is that 95% of measures fall within 2 SD of the mean. Using this rule and the mean and SD, we can predict that 95% of RBCs in the given specimen have volumes between 2 SD (2 \times 1.7 = 3.4 fL) above and below the mean corpuscular volume (MCV) of 90 fL. Thus 95% of RBCs in the original specimen have volumes falling between 86.6 and 93.4 fL. The remaining 5% of RBCs (if there were a much larger group of measurements) would have volumes less than 86.6 fL or greater than 93.4 fL. Because the normal curve is symmetric, the remaining 5% is split between those RBCs that are less and those that are greater than the stated boundaries. We can, therefore, deduce that 2.5% of the RBCs are less than 86.6 fL in volume, and 2.5% of the RBCs are greater than 93.4 fL in volume (Fig. 1).

### P value

Ninety-five percent of all measures (by definition) occur within 2 SD of the mean (3). Measures beyond 2 SD from the mean will occur only 5% of the time. By convention, a value so extreme that it occurs only 5% of the time, or less, is considered in statistical jargon as ‘unlikely’.

\(^1\)Note that the distance from the mean is considered to be positive, regardless as to whether or not the measure lays above or below the center. If we were to assign negative directional values to those measures below the mean, the positive and negative differences would cancel each other.

\[(-3) + (-2) + (-1) + 1 + 2 + 3 = 0\]

\(^2\)The actual SD formula involves squaring the deviations, summing them, and ultimately taking the square root (2). These actions weight deviations based on their extremeness and resolve the problem of positive and negative differences cancelling each other. Those requiring precise knowledge of statistical calculations are referred to a textbook of statistics.

---

**Fig. 1.** Standard deviation and tails. Illustrated is a frequency distribution of 100,000 RBCs normally distributed with a mean MCV of 90 fL and a SD of 1.5 fL. Approximately 95% of all values lay within 2 SD of the mean in any normal distribution. In this case, 95,000 RBCs have MCVs between 86 and 93 fL which is 2 SD (=1.5 fL \times 2 = 3 fL) above and below the mean of 90 fL. The areas in green are the two tails of the distribution and represent 5% of the total RBCs, which is 5,000 RBCs; 2.5% of RBCs lie in each tail. Values that lie in the green area occur with a frequency of less than 5% and are considered unlikely.

The **P** in ‘P value’ can be translated as ‘probability’. A ‘P value’ is the probability that a value as extreme as that stated will occur. For example, using the example above, RBCs with a volume greater than 3.4 fL above or below the mean of 90 fL are considered to have a P value of 0.05 (5%). Note that this formulation results in half the values (2.5%) being 2 SD above the mean and the other half (2.5%) being 2 SD below the mean. This is the conventional way of interpreting P values. The term ‘2 Tailed P value’ refers to this convention and the fact that the unlikely values lie split between the two tails of the frequency distribution (Fig. 1).

### Sample and inference

To this point, we have looked at particular sets of measures and how they are characterized in terms of the center and dispersion. For example, a researcher may evaluate and characterize the MCV of 100 patients with iron deficiency anemia. He could define the mean MCV and the SD for those 100 patients. However, a researcher may be interested in using the results of a study to make predictions about subjects other than those in the study. For example, the researcher might use the information gathered from 100 patients with iron deficiency anemia and make a prediction about the MCV of all patients with iron deficiency anemia. The term ‘inferential statistics’ is used to describe that area of study that systematically explores the prediction of the characteristics of large populations based on the results of a sample.

A set of measures in a study is considered a ‘sample’. A person in the study is considered a ‘subject’. A second use for the term ‘sample’ is to refer to the group of subjects from whom the measures are derived. When the researcher generalizes the findings to a broader group than the
study group, that broader group is considered to be the 'population' (Fig. 2). As an example, a population might be considered to be all patients with iron deficiency anemia or all of the measures of MCV in those persons. The mean and SD are generally measured from a sample and are estimates of those values in the whole population. The true and unknown mean and SD for the entire population are considered to be the 'parameters' of that population. They cannot be absolutely verified unless every member of that population can be identified and evaluated. In many circumstances, this is considered to be impossible.

Samples are very important because based on the study of the sample the researcher will often make predictions about a large population. It is critical that samples be representative of the populations they represent. Samples must be precisely defined in terms of inclusion and exclusion characteristics. Additionally, because selection of samples by the researcher may lead to unsuspected biases, good research often incorporates random selection of the sample.

**P values for comparisons**

In medical research, groups are often compared. One commonly encountered scenario is the comparison of two groups of patients receiving different therapies. In this situation, the mean response is calculated in each group and the mathematical characteristics of normal distributions are used to estimate the likelihood that the difference in the means could be due to chance alone (if there were indeed no true difference in response). In this context, the \( P \) value represents this chance difference. A \( P \) value of less than 0.05 is interpreted as meaning that if therapies were equal it would be unlikely that such a chance difference would occur (Fig. 3). Hence, a real difference in response likely exists. This touches on the topic of 'Standard Error of the Mean'. This will be addressed in a future article in this series.

**Fig. 2.** Sample and population. This figure illustrates that the sample is a group of subjects selected from the population. If the population were considered to be all persons in the world with iron deficiency anemia, the population would be very large indeed. The researcher determines the sample size.

**Fig. 3.** \( P \) value: what is the likelihood that the difference in outcomes could occur by chance if there were actually no difference in the treatments? In this fictitious illustration, two groups receive different treatments for iron deficiency anemia. Both groups respond to therapy with an increase in the mean MCV. The difference in the outcomes between Groups 1 and 2 is 5 fL (10 fL in group 1 and 15 fL in group 2). The \( P \) value of 0.05 suggests that the likelihood of this outcome is only 5% if there is no actual difference in the effectiveness of therapies. By convention outcomes with a \( P \) value of 0.05 or less are considered 'unlikely'. This low \( P \) value supports the conclusion that the treatment in Group 2 is truly superior to the treatment in Group 1.

**Main points**

- Normal distributions are symmetric and can be defined in terms of center (mean) and variability (SD).
- Within normal distributions, the highest frequency value (mode) is at the center (mean), that is, the mean equals the mode.
- In a normal distribution approximately 95% of values fall within 2 SD of the mean.
- The \( P \) value describes the probability of finding a value more extreme (in distance from the mean) than the stated value. An observation with a \( P \) value of less than 5% (0.05) is by convention considered to be 'unlikely'.
- When the mean and SD are calculated based on a sample they remain estimates. The true values (parameters) for the larger population remain unknown.

**Acknowledgements**

The author gratefully acknowledges the contributions of the following individuals whose thoughtful reviews and constructive criticisms contributed to the completion of this document: Alan T. Kaell, M.D.; Neha Naik, M.D.; and Mohit Sharma, M.D.

**Conflict of interest and funding**

The author has not received any funding or benefits from industry or elsewhere to conduct this study.

**References**

1. Cardinal LJ. Central tendency and variability in biological systems. J Community Hosp Intern Med Perspect 2015; 5(3): 1–4.
2. Indrayan A. Medical biostatistics. 3rd ed. Boca Raton, FL: CRC Press; 2012.
3. Glantz SA. Primer of biostatistics. 7th ed. New York: McGraw-Hill; 2011.