In most clinical and experimental studies, the standard deviation (SD) and the estimated standard error of the mean (SEM) are used to present the characteristics of sample data and to explain statistical analysis results. However, some authors occasionally muddle the distinctive usage between the SD and SEM in medical literature. Because the process of calculating the SD and SEM includes different statistical inferences, each of them has its own meaning. SD is the dispersion of data in a normal distribution. In other words, SD indicates how accurately the mean represents sample data. However the meaning of SEM includes statistical inference based on the sampling distribution. SEM is the SD of the theoretical distribution of the sample means (the sampling distribution). While either SD or SEM can be applied to describe data and statistical results, one should be aware of reasonable methods with which to use SD and SEM. We aim to elucidate the distinctions between SD and SEM and to provide proper usage guidelines for both, which summarize data and describe statistical results.

**Key Words:** Standard deviation, Standard error of the mean

A data is said to follow a normal distribution when the values of the data are dispersed evenly around one representative value. A normal distribution is a prerequisite for a parametric statistical analysis [1]. The mean in a normally distributed data represents the central tendency of the values of the data. However, the mean alone is not sufficient when attempting to explain the shape of the distribution; therefore, many medical literatures employ the standard deviation (SD) and the standard error of the mean (SEM) along with the mean to report statistical analysis results [2].

The objective of this article is to state the differences with regard to the use of the SD and SEM, which are used in descriptive and statistical analysis of normally distributed data, and to propose a standard against which statistical analysis results in medical literatures can be evaluated.

Medical studies begin by establishing a hypothesis about a population and extracting a sample from the population to test the hypothesis. The extracted sample will take a normal distribution if the sampling process was conducted via an appropriate randomization method with a sufficient sample size. As with all normally distributed data, the characteristics of the sample are represented by the mean, variance or SD. The variance or SD includes the differences of the observed values from the mean (Fig. 1); thus, these values represent the variation of the data [1-3]. For instance, if the observed values are scattered closely around the mean value, the variance — as well as the SD — are reduced. However, the variance can confuse the interpretation of the data because it is computed by squaring the units of the observed values. Hence, the SD, which uses the same units used with the mean, is more appropriate [3] (Equations 1 and 2).

**Sample:** \(x_1, x_2, x_3, \ldots, x_n\) (sample size = \(n\))

\[
\text{Mean} (\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{Equation 1}
\]

\[
\text{Variance} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}, \quad \text{Standard Deviation (SD)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \quad \text{Equation 2}
\]
As mentioned previously, using the SD concurrently with the \( \text{mean} \) can more accurately estimate the variation in a normally distributed data. In other words, a normally distributed statistical model can be achieved by examining the \( \text{mean} \) and the SD of the data \([1]\) (Fig. 1, Equations 1 and 2). In such models, approximately 68.7% of the observed values are placed within one SD from the \( \text{mean} \), approximately 95.4% of the observed values are arranged within two SDs from the \( \text{mean} \), and about 99.7% of the observed values are positioned within three SDs from the \( \text{mean} \) \([1,4]\). For this reason, most medical literatures report their samples in the form of the \( \text{mean} \) and SD \([5]\).

The sample as referred to in medical literature is a set of observed values from a population. An experiment must be conducted on the entire population to acquire a more accurate confirmation of a hypothesis, but it is essentially impossible to survey an entire population. As a result, an appropriate sampling — a process of extracting a sample that represents the characteristics of a population — is essential to acquire reliable results. For this purpose, an appropriate sample size is determined during the research planning stage and the sampling is done via a randomization method. Nevertheless, the extracted sample is still a part of the population; thus, the sample mean is an estimated value of the population mean. When the samples of the same sample size are repeatedly and randomly taken from the same population, they are different each other because of sampling variation as well as sample means (Fig. 2, Level B).
distribution of different sample means, which is achieved via repeated sampling processes, is referred to as the sampling distribution and it takes a normal distribution pattern (Fig. 2, Level C) [1,6,7]. Therefore, the SD of the sampling distribution can be computed; this value is referred to as the SEM [1,6,7]. The SEM is dependent on the variation in the population and the number of the extracted samples. A large variation in the population causes a large difference in the sample means, ultimately resulting in a larger SEM. However, as more samples are extracted from the population, the sample means move closer to the population mean, which results in a smaller SEM. In short, the SEM is an indicator of how close the sample mean is to the population mean [7]. In reality, however, only one sample is extracted from the population. Therefore, the SEM is estimated using the SD and a sample size (Estimated SEM). The SEM computed by a statistical program is an estimated value calculated via this process [5] (Equation 3).

Estimated Standard Error of the Mean (SEM) = \( \frac{SD}{\sqrt{n}} \)  
95% Confidence Interval : \( \overline{X} \pm (1.96 \times SEM) \)

A confidence interval is set to illustrate the population mean intuitively. A 95% confidence interval is the most common [3,7]. The SEM of a sampling distribution is estimated from one sample, and a confidence interval is determined from the SEM (Fig. 2, Level C). In the strict sense, the 95% confidence interval provides the information about a range within which the 95% sample means will fall, it is not a range for the population mean with 95% confidence. For example, a 95% confidence interval signifies that when 100 sample means are calculated from 100 samples from a population, 95 of them are included within the said confidence interval and 5 are placed outside of the confidence interval. In other words, it does not mean that there is a 95% probability that the population mean lies within the 95% confidence interval.

When statistically comparing data sets, researchers estimate the population of each sample and examine whether they are identical. The SEM — not the SD, which represents the variation of each sample data set — is set to illustrate the population mean [7]. In such cases, it would be easy to misinterpret the population mean within the error range specified by the pre-set significance level [4,6,8].

The SEM is smaller than the SD, as the SEM is estimated usually the SD divided by the square root of the sample size (Equations 2 and 3). For this reason, researchers are tempted to use the SEM when describing their samples. It is acceptable to use either the SEM or SD to compare two different groups if the sample sizes of the two groups are equal; however, the sample size must be stated in order to deliver accurate information. For example, when a population has a large amount of variation, the SD of an extracted sample from this population must be large. However, the SEM will be small if the sample size is deliberately increased. In such cases, it would be easy to misinterpret the population from using the SEM in descriptive statistics. Such cases are common in medical research, because the variables in medical research impose many possible biases originated from inter- and intra-individual variations originated from underlying general conditions of the patients and so on. When interpreting the SD...
and SEM, however, the exact meaning and purposes of the SD and SEM should be considered to deliver correct information. [3,4,6,7,10].

We examined 36 clinical or experimental studies published in Volume 6, Numbers 1 through 6 of the Korean Journal of Anesthesiology and found that a few of the studies inappropriately used the SD and SEM. First, examining the descriptive statistics, we found that all of the studies used the mean and SD or the observed number and percentage. One study suggested a 95% confidence interval; this particular study appropriately stated the sample size along the confidence interval, offering a clearer understanding of the data suggested in the study [11]. Among the 36 studies examined, only one study described the results of a normality test [12]. Second, all 36 studies used the SD, the observed number or the percentage to describe their statistical results. One study did not specify what the values in the graphs and tables represent (i.e., the mean, SD or interquartile range). There was also a study that used the mean in the text but showed an interquartile range in the graphs. Sixteen studies used either the observed number or the percentage, and most of them reported their results without a confidence interval. Only two studies stated confidence intervals, but only one of those two studies appropriately used the confidence interval [13]. As shown above, we found that some of the studies have inappropriately used the SD, SEM and confidence intervals in reporting their statistical results. Such instances of the inappropriate use of statistics must be meticulously screened during manuscript reviews and evaluations because they may hamper an accurate comprehension of a study’s data.

In conclusion, the SD reflects the variation in a normally distributed data, and the SEM represents the variation in the sample means of a sampling distribution. With this in mind, it is pertinent to use the SD (paired with a normality test) to describe the characteristics of a sample; however, the SEM or confidence interval can be used for the same purpose if the sample size is specified. The SEM, paired with the sample size, is more useful when reporting statistical results because it allows an intuitive comparison between the estimated populations via graphs or tables.

ORCID

Junyong In, http://orcid.org/0000-0001-7403-4287
Sangseok Lee, http://orcid.org/0000-0001-7023-3668

References

1. Curran-Everett D, Taylor S, Kafadar K. Fundamental concepts in statistics: elucidation and illustration. J Appl Physiol (1985) 1998; 85: 775-86.
2. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel. Adv Physiol Educ 2007; 31: 295-8.
3. Altman DG, Bland JM. Standard deviations and standard errors. BMJ 2005; 331: 903.
4. Carlin JB, Doyle LW. Statistics for clinicians: 4: Basic concepts of statistical reasoning: hypothesis tests and the t-test. J Paediatr Child Health 2001; 37: 72-7.
5. Livingston EH. The mean and standard deviation: what does it all mean? J Surg Res 2004; 119: 117-23.
6. Rosenbaum SH. Statistical methods in anesthesia. In: Miller's Anesthesia. 8th ed. Edited by Miller RD, Cohen NH, Eriksson LI, Fleisher LA, Wiener-Kronish JP. Young WL. Philadelphia, Elsevier Inc. 2015, pp 3247-50.
7. Curran-Everett D. Explorations in statistics: standard deviations and standard errors. Adv Physiol Educ 2008; 32: 203-8.
8. Carley S, Lecky F. Statistical consideration for research. Emerg Med J 2003; 20: 258-62.
9. Mahler DL. Elementary statistics for the anesthesiologist. Anesthesiology 1967; 28: 749-59.
10. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anesthesiology journals. Br J Anaesth 2003; 90: 514-6.
11. Koh MJ, Park SY, Park EJ, Park SH, Jeon HR, Kim MG, et al. The effect of education on decreasing the prevalence and severity of neck and shoulder pain: a longitudinal study in Korean male adolescents. Korean J Anesthesiol 2014; 67: 198-204.
12. Kim HS, Lee DC, Lee MG, Son WR, Kim YB. Effect of pneumoperitoneum on the recovery from intense neuromuscular blockade by rocuronium in healthy patients undergoing laparoscopic surgery. Korean J Anesthesiol 2014; 67: 20-5.
13. Lee H, Shon YJ, Kim H, Paik H, Park HP. Validation of the APACHE IV model and its comparison with the APACHE II, SAPS 3, and Korean SAPS 3 models for the prediction of hospital mortality in a Korean surgical intensive care unit. Korean J Anesthesiol 2014; 67: 115-22.