

Standard deviation and standard error of the mean

Dong Kyu Lee¹, Junyong In², and Sangseok Lee³

Department of Anesthesiology and Pain Medicine, ¹Korea University Guro Hospital, Seoul, ²Dongguk University Ilsan Medical Center, Goyang, ³Sanggye Paik Hospital, Inje University College of Medicine, Seoul, Korea

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

자료가 하나의 대푯값을 중심으로 좌우 대칭인 산포(dispersion)를 하는 경우 정규분포(normal distribution)를 따른다고 하며, 정규분포는 대부분의 모수적 통계분석(parametric statistical analysis)의 전제조건이 된다[1]. 정규분포를 전제로 하는 자료에서의 평균은 자료의 중심경향성을 나타내는 대푯값으로 사용된다. 그러나, 평균만으로 자료의 분포모양을 설명하기는 부족하기 때문에 연구논문에서는 표준편차(standard deviation; SD)나 표준오차(standard error of the mean; SEM)를 함께 사용하여 자료의 분포를 설명하거나 통계분석 결과를 보고한다[2].

이 글은 정규분포를 하는 자료의 특성을 설명하거나 통계분석

결과에 사용되는 SD와 SEM의 차이와 이용방법에 대해서 기술하고, 연구논문의 통계분석 결과를 평가하고 판단하는 기준을 제시하고자 한다.

의학연구는 관심이 되는 모집단(population)에 대하여 가설을 설정하고 이것을 입증하기 위한 표본(sample)의 추출로부터 시작된다. 이 때, 추출 과정에서 적절한 무작위법(randomization method)과 충분한 표본수(sample size)를 확보한다면, 표본은 정규분포를 따른다. 그러므로, 이 표본의 특성은 평균과 분산(variance) 혹은 SD로 나타낸다. 분산 혹은 SD는 모든 관측값(observed value)에서 평균까지의 차이를 포함하므로(Fig. 1), 자료의 변동(variation of data)을 나타내는 값이다[1-3]. 예를 들어, 관측값이 평균 주변에 모여 있다면, 분산은 작아지며, SD 또한 작아진다. 하지만 분산은 계산과정 중 관측값의 단위를 제공하게 되어 자료를 해석하는데 혼란을 야기할 수 있으므로, 자료의 평균과 같은 단위를 사용하는 SD를 이용하는 것이 더 적절하다[3](Equations 1 and 2).

Sample: $x_1, x_2, x_3, \dots, x_n$ (sample size = n)

$$\text{Mean } (\bar{x}) = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots \text{Equation 1}$$

$$\text{Variance} = \frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n-1}, \text{ Standard Deviation (SD)} = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n-1}} \dots\dots\dots \text{Equation 2}$$

Corresponding author: Dong Kyu Lee, M.D., Ph.D.
Department of Anesthesiology and Pain Medicine, Korea University Guro Hospital, 148, Gurodong-ro, Guro-gu, Seoul 152-703, Korea
Tel: 82-2-2626-3237, Fax: 82-2-2626-1437
E-mail: entopic@naver.com
ORCID: <http://orcid.org/0000-0002-4068-2363>

Received: April 2, 2015.
Revised: 1st, April 21, 2015; 2nd, April 24, 2015; 3rd, May 6, 2015.
Accepted: May 7, 2015.

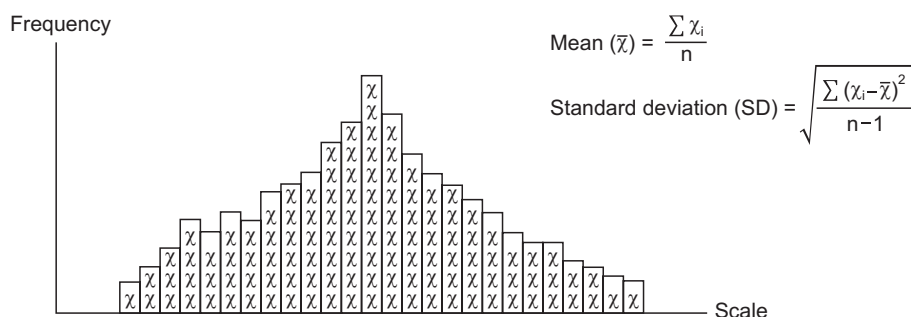
Korean J Anesthesiol 2015 June 68(3): 220-223
<http://dx.doi.org/10.4097/kjae.2015.68.3.220>

A

	A	B	C	D	E	F	G
1	χ	χ	χ	χ	χ	...	χ
2	χ	χ	χ	χ	χ	...	χ
3	χ	χ	χ	χ	χ	...	χ
4	χ
5	χ	χ	χ	χ	χ	χ	χ



B



C

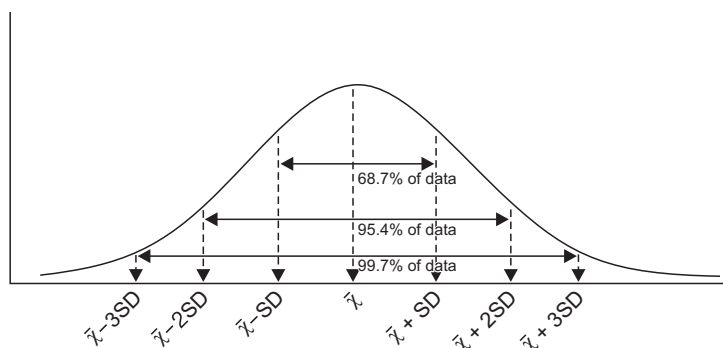


Fig. 1. Process of data description. First, we gather raw data from the population by means of randomization (A). We then arrange the each value according to the scale (frequency distribution); we can presume the shape of the distribution (probability distribution) and can calculate the mean and standard deviation (B). Using these mean and standard deviation, we produce a model of the normal distribution (C). This distribution represents the characteristics of the data we gathered and is the normal distribution, with which statistical inferences can be made (\bar{x} : mean, SD: standard deviation, x_i : observation value, n : sample size).

이와 같이 평균과 SD를 이용하면 정규분포를 전제하는 표본의 변동 정도를 추측할 수 있다. 다시 말하면, 표본으로부터 계산된 평균과 SD를 이용하여 하나의 정규분포를 전제하는 통계적 모형을 구하는 방법이며[1](Fig. 1)(Equations 1 and 2), 정규분포의 경우 평균에서 1배의 SD만큼 떨어진 범위 안에 약 68.7%의 자료가 위치하며, 2배의 SD 범위에는 약 95.4%가, 3배의 범위에서는 약 99.7%의 자료가 포함된다[1,4]. 이런 이유로 대부분의 연구논문에서 표본을 평균과 SD로 보고한다[5].

의학연구에서 표본은 연구대상인 모집단에서 관측값을 추출한 것이다. 정확한 연구결과를 위해서는 모집단 전체를 대상으로 가설을 적용하여 실험해야 하지만, 현실에서 모집단 전체를 조사하는 것은 불가능하다. 그렇기 때문에 모집단의 특성을 유지하는 표본을 추출하는 과정이 신뢰할 수 있는 결과를 얻기 위해 필요

하다. 이런 목적으로 연구계획 단계에서 필요한 표본수를 결정하고 표본을 추출할 때는 무작위법(randomization method)을 사용한다. 하지만 이 표본도 모집단의 일부이므로, 표본평균은 모평균(population mean)의 추정값이다. 만약 같은 모집단에서 무작위법으로 같은 표본수를 가지는 표본을 다시 추출한다면 이 과정에서 발생하는 변이(sampling variation)에 의하여 처음과 다른 표본이 추출되며, 표본평균도 달라진다(Fig. 2, Level B). 이런 방법으로 반복적으로 표본을 추출하여 얻어지는 표본평균들로 만든 분포를 표집분포(sampling distribution)라 하며, 이 분포는 정규분포를 한다(Fig. 2, Level C)[1,6,7]. 그러므로, 표집분포의 SD를 계산할 수 있으며, 이 값이 SEM이다[1,6,7]. SEM은 모집단의 변동과 추출된 표본의 개수에 따라 좌우된다. 모집단의 변동이 크다면, 추출된 표본의 평균들의 차이도 커져서, SEM은 커진다. 하지만 모

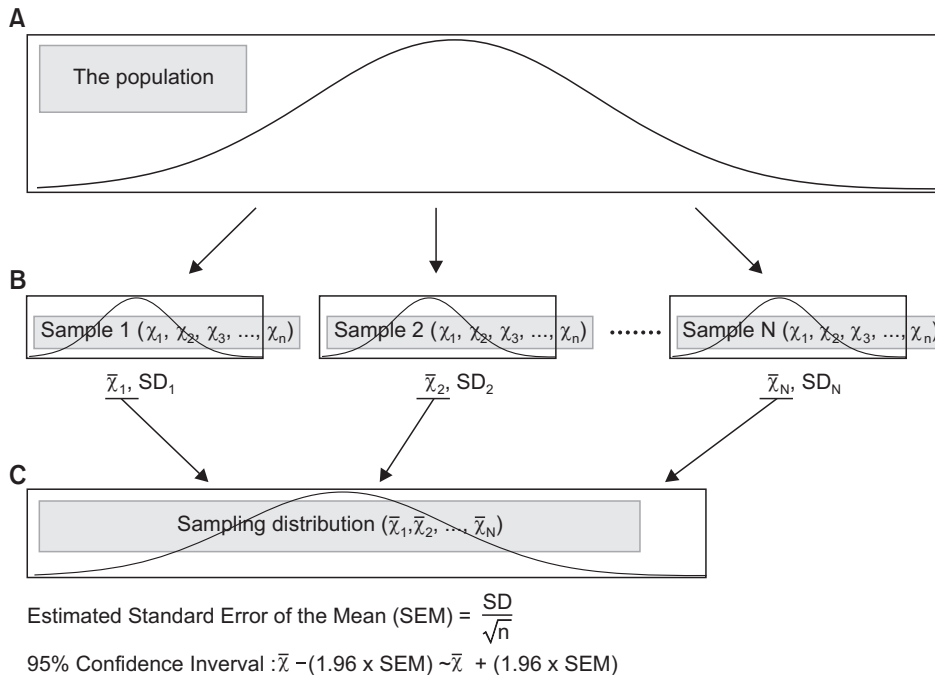


Fig. 2. Process of statistical inference. Level A indicates the population. In most experiments, we only obtain one set of sample data from the population using randomization (Level B); the mean and standard deviation are calculated from sample data we have. For statistical inference purposes, we assume that there are several sample data sets from the population (Level B); the means of each sample data set produce the sampling distribution (Level C). Using this sampling distribution, statistical analysis can be conducted. In this situation, the estimated standard error of the mean or the 95% confidence interval has an important role during the statistical analysis process (\bar{x} : mean, SD: standard deviation, n: sample size, N: number of sample data sets extracted from population).

집단으로부터 표본을 많이 추출할수록, 표집분포를 구성하는 평균들이 모평균에 가까이 모이게 되어 SEM은 작아진다. 한마디로 SEM은 표본의 평균이 얼마나 모평균에 가까운지 나타내는 지표이다[7]. 그러나, 실제로는 하나의 표본만 추출하므로 이 표본의 SD와 표본수를 이용하여 SEM을 추정한다. 통계 프로그램에서 출력되는 SEM은 이 과정에 의하여 계산된 추정값이다[5](Equation 3).

$$\text{Estimated Standard Error of the Mean (SEM)} = \frac{SD}{\sqrt{n}} \cdots \text{Equation 3}$$

모집단의 평균을 직관적으로 표현하기 위하여 신뢰구간(confidence interval)을 사용하며, 95% 신뢰구간을 사용하는 것이 일반적이다[3,7]. 신뢰구간은 하나의 표본에서 표집분포의 SEM을 추정하여, 이 SEM에 의해 결정된다(Fig. 2, Level C). 따라서, 신뢰구간은 모평균의 존재 가능성에 대한 구간이 아니라, 표집분포를 이루는 표본평균들의 존재 가능성에 대한 구간이다. 예를 들면 95% 신뢰구간은 모집단에서 100개의 표본을 추출하여 100개의 표본평균을 계산하였을 때, 95개는 이 신뢰구간에 포함되며, 5개는 이 신뢰구간 밖에 존재한다는 의미이다. 다시 말하면, 모평균이 이 신뢰구간에 95%의 확률로 존재한다는 의미가 아니다.

통계를 사용하여 자료들을 비교할 때에는 각 표본의 모집단을 추정하여 이들이 서로 같은지에 대하여 검정한다. 따라서 이 과정에는 자료의 변동을 대변하는 SD가 아니라 SEM을 이용하여 모평균을 추정한다(Fig. 2)[4,8,9]. 이로써 미리 상정한 유의수준(significant level) 안에서 발생할 수 있는 오차의 범위에서, 연구에 사용된 표본이 모집단의 특성을 대변한다고 결론 내릴 수 있다[4,6,8].

SEM은 SD를 표본수의 제곱근으로 나누어 계산되므로 SD보

다 작은 숫자를 가진다(Equations 2 and 3). 이런 이유로 연구자들은 표본을 기술할 때 SEM을 이용하려는 유혹에 빠진다. 서로 다른 두 집단을 비교할 때, 두 집단의 표본수가 같다면 각 표본의 SEM나 SD중 어떤 값을 비교하여도 무관하지만, 표본수가 다른 경우, 이를 명시하여야 정확한 정보를 전달할 수 있다. 예를 들어, 모집단의 변동이 커서 표본의 SD가 크게 계산 되었을 때 의도적으로 표본수를 늘리면 SEM은 작아지게 된다. 이 경우 기술통계(descriptive statistics)에 SEM을 이용하면 모집단의 모양을 잘못 이해할 수 있다. 의학연구에서 변수들은 기저질환 등에 의한 개인간의 차이, 개인의 상태 변화에서 기인하는 편의(偏倚, bias)를 내재하므로, 이런 상황이 흔하게 발생된다. 따라서, SD와 SEM을 해석하고 사용할 때는 각각의 사용 목적을 고려하여 정확한 정보가 전달되도록 하여야 한다[3,4,6,7,10].

Korean Journal of Anesthesiology (KJA) 67권 1호에서 6호에 게재된 36편의 임상 혹은 실험연구를 분석한 결과, 일부 연구에서 SD와 SEM이 부적절하게 사용된 것을 확인하였다. 먼저, 기술통계를 살펴보면, 모든 연구에서 평균과 SD 또는 관측수(observed number)와 백분위수(%)를 사용하였으며, 95% 신뢰구간을 제시한 경우가 1편 있었다. 95% 신뢰구간을 사용한 연구에서는 신뢰구간과 함께 표본수를 적절하게 기술하여 자료의 특성을 이해하기 쉽게 하였다[11]. 그러나, 전체 36편의 연구들 중 정규성검정(normality test)의 결과를 기술한 연구는 1편 뿐이었다[12]. 두 번째로, 통계분석 결과는 기술통계에서와 마찬가지로 모든 연구에서 SD와 관측수 혹은 백분위수가 사용되었다. 1편의 연구에서는 그래프나 표에 제시된 값이 SD나 SEM, 사분위범위(interquartile range) 등 무엇인지 기술되지 않았으며, 본문에는 평균이, 그래프에는 사분위범위가 기술된 경우도 있었다. 결과를 제시할 때 관측수 혹은 백분위수가 사용된 연구는 16편이었으며, 대부분 신뢰구

간 없이 백분위수로만 제시하였다. 신뢰구간이 함께 기술한 경우는 2편이었으나, 그 중 1편만 신뢰구간의 의미에 맞게 기술되었다 [13]. 이와 같이 일부 논문에서 SD와 SEM, 신뢰구간이 정확하지 않게 기술되었다. 이는 독자들에게 정보전달의 혼란을 초래하므로, 논문을 작성하는 과정이나 심사과정에서 주목되어야 할 부분들이다.

결론적으로, 정규분포를 전제하는 표본에서 SD는 표본을 구성하는 자료의 변동을 반영하며, SEM는 표집분포를 구성하는 평균들의 변동을 반영한다. 그러므로 표본의 특성을 기술할 때는 정규성검정이 선행된 상태에서 SD를 사용하는 것이 바람직하나, 자료

에 따라 표본수를 제시한다면 SEM 혹은 신뢰구간을 통한 표현도 가능하다. 통계분석 결과를 제시할 때는 SD보다는 표본수와 함께 SEM을 사용하면 추정된 모집단들의 직관적인 비교가 표나 그래프를 통해 가능하여 결과해석이 용이해진다.

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 anaesthesia 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.