



Common Statistical Mistakes in Descriptive Statistics Reports of Normal and Non-Normal Variables in Biomedical Sciences Research

Farzan MADADIZADEH^{1,2}, Mohamad EZATI ASAR³, *Mostafa HOSSEINI^{2,4}

1. Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran

2. Dept. of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

3. Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

4. Pediatric Chronic Kidney Disease Research Center, Tehran, Iran

*Corresponding Author: Email: mhossein110@yahoo.com

(Received 14 Sep 2015; accepted 05 Oct 2015)

Dear Editor in Chief

Statistics are the aids of researchers, and the proper use of them requires sufficient knowledge of theoretical and applied concepts of statistics (1).

Today, with the advances in science and the help of computer, statistical analysis tools of data research are easily accessible to any researcher and researchers obtain output and report it in their research without having enough information by selecting a number of options from these black boxes (statistical software). The lack of correct writing of some statistical indicators has been one of the known problems in medical sciences literature in recent years, E.G., standard deviation and the wrong replacement of that with standard error in articles (2). This is so important that for some valid journals the measure of article acceptance is to pinpoint these indicators.

Standard deviation is an indicator of descriptive statistics that describes the distribution of sample data around their mean (3).

x_i : Data of i-th observation

\bar{x} : mean of observations,

n = sample size

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad [1]$$

The standard error is an indicator of inferential statistics to extend the sample results to the population from which the sample was extracted. In other words, this indicator reflects the credibility and reliability of the research and answers the question what the dispersion of results will be if a sample with the same size and by substitution is selected in other times from the same initial population, for example, if the goal is to generalize the results of the sample mean to the population mean, to evaluate the validity of the study after re-sampling for n times and to calculate the means of the sample, their dispersion around the total mean is the standard error (4). But in practice, apart from wasting time, re-sampling technique would be a waste of cost, thus the following simple equation is used to calculate the standard error (5).

$$SE = \frac{SD}{\sqrt{n}} \quad [2]$$

The standard error is always less than the standard deviation and medical researchers sometimes report it mistakenly instead of the standard deviation in descriptive statistics of the study variables, this will distract readers from seeing the dispersion and the research data seems better (3, 5).

The standard deviation alone is meaningful and it is essential to present it in descriptive statistics of normal quantitative variables but the standard error alone is meaningless and it is used only to build confidence interval (6).

Another problem is that some researchers without evaluating the normality of quantitative variable begin to report the descriptive indicators of the variable. It should be noted that standard deviation is only valuable to describe the dispersion in normal quantitative variables and in a case that the variable is not normally distributed another dispersion indicator called interquartile range ($IQR = Q_3 - Q_1$) is used. Based on dividing a data set into quartiles, the first quartile, denoted Q_1 , is the value in the data set that holds 25% of the values below it. The third quartile, denoted Q_3 , is the value in the data set that holds 75% of the values below it (1, 7).

Another problem is the final reporting in the form of $\bar{x} \pm SD$ that creates confusion with the confidence interval for the mean ($\bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot SE(\bar{x})$), it is

recommended to avoid confusion that the results of descriptive statistics are presented for the normal variables as **mean(SD)** and for non-normal variables as **median(IQR)** (8).

In short, respecting the above tips in addition to raising the quality of papers and journals, can significantly contribute to the improvement of health systems research, reduce the production of low-quality articles and thus reduce waste of health care system costs in the field of research.

Accordingly, given that a major concern of the Ministry of Health and its policy makers is to reduce costs and increase productivity and improve

the quality of health systems research, it is recommended that researchers and practitioners of health care system publications pay more attention to these points, through which the high costs of health care system can be controlled and reduced.

Acknowledgements

The authors declare that there is no conflict of interests.

References

1. Marino MJ (2014). The use and misuse of statistical methodologies in pharmacology research. *Biochem Pharmacol*, 87(1):78-92.
2. Herxheimer A(1988). Misuse of the standard error of the mean [letter]. *Br J Clin Pharmacol*, 26(2):7-19.
3. Carter RE (2013). A standard error: distinguishing standard deviation from standard error. *Diabetes*,62(8):e15-e.
4. Biau DJ (2011). In brief: Standard deviation and standard error. *Clin Orthop Relat Res*, 469(9):2661-2664.
5. Lee DK ,In J, Lee S(2015). Standard deviation and standard error of the mean. *Korean J Anesthesiol*, 68(3):220-223.
6. Ko WR, Hung WT, Chang H-C, Lin L-Y(2014). Inappropriate use of standard error of the mean when reporting variability of study samples: A critical evaluation of four selected journals of obstetrics and gynecology. *Taiwan J Obstet Gynecol*, 53(1):26-29.
7. Nagele P(2003). Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br J Anaesth*, 90(4):514-516.
8. Motulsky HJ(2015). Common misconceptions about data analysis and statistics. *Br J Pharmacol*,172(8):2126-2132.