Journal of BUON **15:** 182-187, 2010

SPECIAL ARTICLE

Statistical presentation of data in biomedical publications

R. Igić[1,2], S. Stoisavljevic-Šatara[2]

[1]Department of Anesthesiology and Pain Management, Stroger Hospital of Cook County, Chicago, IL, USA; [2]Department of Pharmacology and Toxicology, Medical Faculty, University of Banja Luka, Banja Luka, Republic of Srpska, Bosnia and Herzegovina

Summary

Scientists rely upon statistics to report their findings accurately and to interpret the published findings of others. The proper application of this discipline provides confidence in statements based upon experimental results. Professional statisticians may help at some phases of planning experiments even before results are obtained, but not all scientists avail themselves of such services. As a result, errors can occur in presentation of variability of data, and more often than not, such errors are ultimately published in biomedical journals. This paper examines the most common descriptive statistics for quantitative and categorical data. Standard deviation (SD), standard error of the mean (SEM), confidence intervals (CI), and various technical details, including how to present data in publications, and when to use particular statistical tools, are discussed as well.

Key words: confidence intervals, data display, data variability, descriptive statistics, standard deviation, types of data

Introduction

Most biomedical and clinical studies involve the selection of a small sample from a larger population (humans, animals, microorganisms, etc), from which the findings are then extrapolated to the larger population with varying degrees of confidence. One obstacle common to some researchers is how to report the data. Statisticians suggest that if the purpose is to describe findings from data that are "normally" distributed (Gaussian distribution), then the SD should be used. However, if the purpose is to describe study outcomes, the SEM or a CI should be applied. It is common mistake to use the SEM to describe data [1-3]. Thus, it is important to understand what each of these statistical tools can do and how to use them properly. Also, with measures of central tendency (mean, median, mode), CIs, and data transformation one should be familiar with.

Types of data

The data may be grouped into several types (Table 1).

Table 1. Types of data

*Type**	*Examples*
Quantitative	
Continuous	Age, weight, height
Discrete	Number of arrhythmic attacks per month
Categorical	
Ordered categories	Grade of prostate cancer; better, same, worse
Nominal	Male/female; blood group 0, A, B, AB

*This typifying may be of help to an investigator to display and analyze the data [2]

Because it is easier to summarise categorical variables, quantitative variables are often converted to categorical for descriptive purposes [2]. Thus, arterial blood pressure may be converted into a nominal variable by defining "prehypertension" as a diastolic blood pressure greater than 85 mm Hg, "hypertension" as blood pressure larger than 90 mm Hg, or "normotension" as blood pressure less than or equal to 85 mm Hg. Similar conversion can be done if age is converted into "young", "middle aged" or "old". But, with such categorization of con-

Correspondence to: Rajko Igić, MD, PhD. John Stroger hospital of Cook County, Anesthesiology Research, Chicago, IL 60612, USA.
Fax: +1 312 864 9641, E-mail: r.igic@excite.com

Received 04-05-2009; Accepted 03-06-2009

tinuous variable, the amount of information available is reduced. Also, statistical analysis of the nominal type of data is less powerful.

Variability

There are several and multiple reasons for variability among scientific data. Even if every member of a population is assumed to be identical, it does not mean that the series of samples randomly drawn from a given population will be identical. Variability can arise from the sampling technique. Or, variability can result from innate differences in the samples themselves. Finally, individual observations of identical samples differ because the population from which they are drawn is distributed over a range of possible values. Variability of the data can be described in several ways, including the measures of central tendency and degree of dispersion.

One way to begin is to present the central tendency (mean, median, mode). The mean is the average value; we add up the observed values and divide by the number of them. The weakness of the mean is that extreme values or outliers affect it. For such skewed data, the median is far better measure of central tendency. To identify the median (mid point), we have to find the point that divides the data in two halves, one greater than it, and the other less than it. If the data have a bimodal distribution, the mode is best use because it presents the most common value.

Degree of dispersion

For normal distribution, the degree of dispersion can be described by the mean, range, variance, SD, and SEM. However, for the skewed (non-normal) distributions these measures of dispersions may be misleading. For that reason, it is better to divide the data into four, and get quartiles. The median is in the second quartile. Such data may be summarized in the interquartile range. Such type of data is possible to display as a dot plot by showing all data together with the median for each set. If the data sets are large, plotting individual points is not practical, and a box-whisker plot may be used instead [2]. The box is marked as the first and third quartile, and the whiskers show the range (Figure 1).

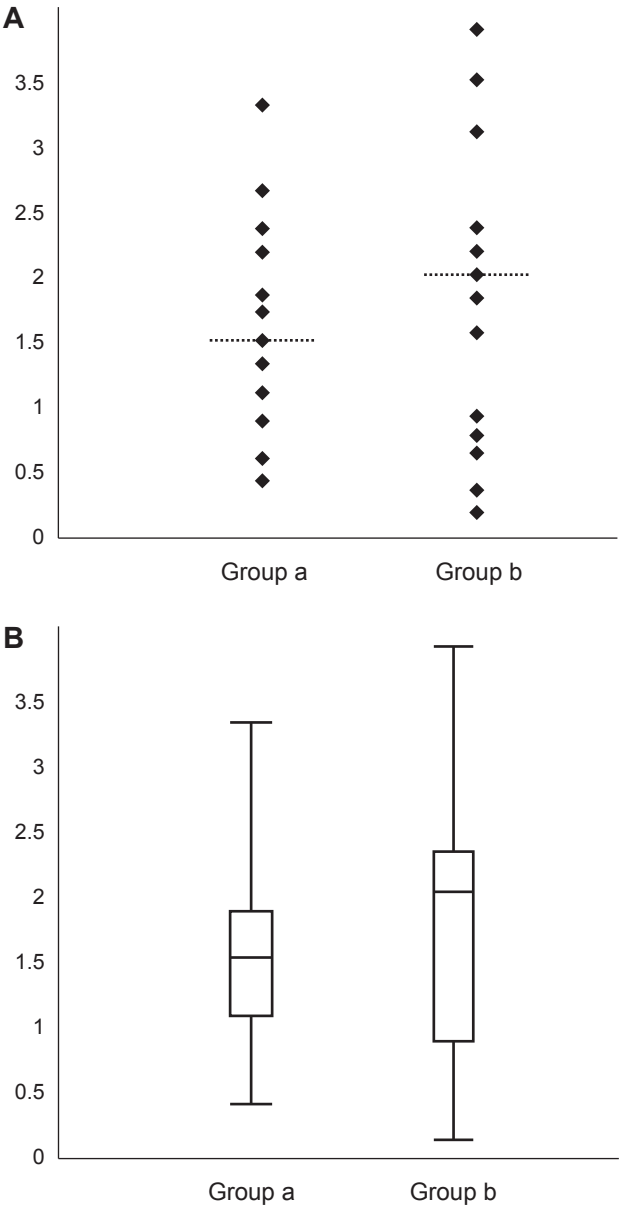

Figure 1. Data display as a dot plot **(A)** and box-whisker plot **(B)**. The median for each set of the data is marked both at the dot plots and box-whisker plots. The box is marked by the first and third quartile. The whiskers show the range.

Standard deviation

The range of data points about the mean is an important statistical parameter, but it does not accurately indicate the spread of observations about the mean. One way to express variability of the sample or the population under study is to use the respective variances. However, the units for variance for a sample are different from those for the population as a whole. Thus, it is better to obtain the square root of the variance, or the SD.

SD is a statistic that describes the degree of variation among the individual observations in the sample. If all individual observations had the same value, the SD would be zero; the bigger spread of these observa-

tions from their mean, the larger SD. In other words, a large SD indicates that the individual points are scattered far from the mean, and a small SD indicates that these points are clustered closely around the mean. To use the SD properly, one has to be sure that the data have an approximately normal distribution.

If a random sample has small SD, then sample means are more likely to be close to the population mean than when SD is large. The SD for the entire population (σ) and the SD for the sample (*s*) are necessarily different, as mentioned earlier. For data that fall within a normal distribution, approximately 68% of the individual observations will have values within one SD on either side of the mean; the other 32% are equally scattered above and below these limits. Ninety-five percent of all observations will fall within 1.96 SD of the mean, and the remaining 5% are scattered equally above and below these limits. The range covered by three SD from the mean includes approximately 99.7% of all observations. Thus, if we know the mean and SD of a particular set of observations, we can estimate the range that would include a certain percentage of the data.

Sometimes it is appropriate to transform the data to fit a normal distribution. For example, using $\log_e$ transformation or the square root transformation will reduce larger values more than smaller values. Such transformations are indicated when there are extreme outliers. Alternatively, one could use the median and the inter-quartile range. But, the inter-quartile range is not applicable for small data sets.

Coefficient of variation

The ratio of SD to the sample mean is known as the coefficient of variation (CV). CV is obtained when the intra-subject SD is divided by the mean (*m*), expressed as a percentage. The CV% indicates the measurement error.

$$CV\% = SD/m$$

This estimation is generally used as a measure of repeatability for biochemical assays, such as when the assay is performed on separate occasions on the same sample or when repeated measurements are made on an individual specimen. The CV% should not be used as an indication of between-subject variability.

Standard error of the mean

The sample mean is only an estimate of the mean of the population. For this measurement to be useful, we need to know how closely the collected data approximate the true mean for the population. When the SD is divided by the square root of the number of observations in a sample, the result is an estimate of the standard error of the mean, or SEM (SEM is frequently abbreviated as SE):

$$SEM=SD/\sqrt{n}$$

The term "standard error" is a misnomer, because it is neither standard nor an error. The term originated during the Industrial Revolution when reproducibility of measurements was important; repeating the measurement multiple times and determining the average measurement gave the best result. SEM, as a measure of precision for an estimated characteristic or treatment effect, thus indicates the "error" with which the measurement was made. Based upon random sampling from a given population, SEM depends upon the variation of the population and the size of the sample. Since we do not know the variation in the population, we use variation in the sample to estimate it. As sample size increases, the sample SD (*s*) approaches the entire population SD (σ), and SEM decreases because it depends on sample size. If the sample size expands to include the entire population, then $s = \sigma$ and SEM = 0.

It is also possible to calculate the SEM associated with a percentage or a proportion of a study population. Here, the sample size will influence the size of SEM, but the amount of variation is determined by the percentage or proportion in the population itself, so we do not need to estimate the SD. If *p* represents one percent, 100 - *p* represents the remainder. The SEM of each of these percentages is obtained by the following formula:

$$SEM\% = \sqrt{p\,(100-p)/n}$$

For studies of clinical conditions, we can rarely obtain a random sample. Instead, we must use the patients or subjects that are available. The conditions of clinical research necessarily impose certain biases, such as the fact that hospital patients are not the same as those in the general community, volunteers are not typical of non-volunteers, and the patients who return questionnaires are a particular set, different from those who do not. For these reasons, the *Methods* section of clinical reports frequently contain details related to the selection process as well as the various parameters of age, gender, social status, response rate, etc.

Statistical methodology has advanced considerably, and statistical software has become available to enable researchers to carry out complex data analysis and data presentation. Unfortunately, much of this statistical methodology is used rather uncritically[4]. Pop-

ular software packages are Minitab, SPSS, SAS, BMDP, InStat, and R which is free software for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

Difference between standard deviation and standard error of the mean

SD reflects the dispersion of individual sample observations about the sample mean and thus shows the variability of those observations. In contrast, SEM reflects the theoretical dispersion of sample means about some overall population mean; it characterizes the degree of uncertainty about the true value of that population mean. Because SEM depends upon sample size, it is an inappropriate estimate of variability among observations. However, the standard error is useful in calculation of a CI.

Reference ranges and confidence intervals

When a set of observations has a normal distribution, SD indicates the limits of the scatter of the observations. For example, 1.96 SD both above and below the mean indicate the range within which 95% of the observations lie. The 95% limits are often referred to as a *reference range*, and for many biological variables they define what is regarded as the normal (standard or typical) range. For example, if we measure the mean systolic blood pressure in 100 school children and find it to be 120 mm Hg, with a SD = 6 mm Hg, we can see that the mean plus or minus 1.96 its SD results in 120 + (1.96 × 6) = 131.8 mm Hg and 120 – (1.96 × 6) = 108.2 mm Hg. From these calculations we can conclude that only 1 in 20 (or 5%) of school children in this population would be expected to have systolic blood pressure below 108.2 or above 131.8 mm Hg.

Confidence intervals refer to estimates. If a series of samples are drawn from the population and the mean of each set is calculated, 95% of those means will fall within two SEM above and two below the mean of all sets. This common mean is expected to be very close to the mean of the population. This is the confidence level. The 95% CI mark is obtained similarly to the reference range, but using SEM instead of SD. For example, if in the previous sample of 100 school children, SEM was 0.6 mm Hg, this sample mean plus or minus 1.96 times its standard error gives the following:

$$120 + (1.96 \times 0.6) = 121.2 \text{ mm Hg and}$$
$$120 - (1.96 \times 0.6) = 118.8 \text{ mm Hg}$$

These figures indicate the 95% confidence interval. We can say that there is only a 5% chance that the mean of the population falls outside the range 118.8 to 121.8 mm Hg.

When we take the mean plus or minus three times its SEM (this is done only infrequently), the range would be: 120 + (3 × 0.6)=121.8 mm Hg and 120 – (3 × 0.6)=118.2 mm Hg, indicating 99.73% CI. The chance of the population mean falling outside this range is now only 1 in 370. If different investigators randomly sample the same population (assuming that there will be some sample variability) they may report different 95% confidence intervals, yet for 95 of every 100 investigators the confidence interval will include the overall population mean.

We can see that reference ranges refer to individuals and confidence intervals refer to estimates about the parameters of the population. CI quantifies the particular degree of confidence that the interval drawn in this way will actually include the population mean. The relationship between a reference range and a confidence interval is the same as between SD and SEM.

Standard deviation, standard error of the mean and confidence intervals in publications

To present the results of an investigation, the authors should always indicate the number of subjects (n), the range of results, the central tendency (mean, SD), and the spread (CI for the mean). SD (and SEM when needed) can be presented either graphically or numerically. Graphical representation may be easier for a reader to see, but the numerical values more accurately present the sizes.

Numerical measures of dispersion (SD and SEM) can be shown in two ways. Usually, one records the numerical mean and the most appropriate measure of dispersion, either SD or SEM. Presentation of data collected for measurement of systolic blood pressure in the sample of the school children would look like this: 120 mm Hg (SD = 6, n =100), or 12 ± 6 mm Hg (SD, n =100). Alternatively, using SEM, it would be: 120 mm Hg (SEM = 0.6, n = 100), or 120 ± 0.6 mm Hg (SEM, n =100).

Many journals still report SD with a ± symbol. However, a SD is a single positive number, and many journals have eliminated this symbol in accordance with the recommendation of the *Scientific Style and Format* [5].

For clarity and ease of reading, data should be presented either in tables or figures and sometimes in the text. However, to help reader understand a table or figure, the text should include a brief explanation and

expand upon major findings. The International Committee's statistical guidelines states: "When possible, quantify and present them with appropriate indicators of measurement error or uncertainty (such as CIs)". These guidelines also offer another important suggestion: "Give numbers of observations. Report losses to observation (such as dropouts from a clinical trial)" [6].

When to use standard deviation and standard error of the mean

SD indicates the variability of single observations. SD is useful, for example, when we want to establish normal ranges for a particular diagnostic result. If we know the diastolic blood pressure and the SD of those measurements in healthy people, we can calculate a 95% CI. Then we can tell when an individual's diastolic pressure lies outside this "normal range".

In other circumstances, such as reporting differences between treated and control groups, we might compare means of the different groups and SEM to make conclusions of the study. CIs even better describe the precision of such an estimate. The weakness of SEM is that the mean ± SEM gives only a 68% CI, but the mean ± 1.96 × SEM approximates a 95% CI. When we know *n*, we can convert SD to SEM or *vice versa*, but this mathematical manipulation is not so easy, particularly if we want to indicate differences between two sample means. Statisticians recommend that investigators report the SD or the SEM via graph or text, depending upon the information to be conveyed. The general rule is, if an investigator wants to compare population means, he will report the SEM, and if he wants to emphasize variations among a group of individuals, he will report the SD.

When to report confidence intervals

CIs help the investigator to decide if the experimental result is sufficiently distinct to be relevant. CI characterizes the uncertainty about the true value of a population parameter. This measure of dispersion has become generally used in clinical research, as well as in the broader biomedical field. If we present a CI for a population mean, we give bounds to the expected discrepancy between the sample and the population mean. CI is a strong tool for inference because it provides the same statistical information as the P value [2,4]. In publications, this type of uncertainty is presented graphically or numerically.

CIs are frequently used for the results of clinical trials, including the number needed to treat (NNT) in trials that have a binary outcome. In a trial comparing two treatments, the NNT is the estimated number of patients who need to be treated with the new rather than the standard treatment to find one additional patient of benefit [7,8].

Common errors in reporting the standard error of the mean

Because the reported mean and SEM can refer either to a sample or to a population, it is important for the investigator to define how they are used. Generally, the mean and SD are used to describe data, while the mean, SEM and 95% CI are preferred for reporting an estimate and its measure of precision to the main outcome of a study [9].

Some medical journals provide guidelines for reporting statistics, and a few have an editor especially trained to evaluate statistics. In summarizing results from experiments, an investigator should report variability among the actual individual measurements, as indicated by the SD. In fact, guideline No. 5 for reporting statistics in journals published by the American Physiological Society states "Report variability using a standard deviation" [10]. However, if an investigator repeated the same experiment many times, and each time calculated a sample mean, the SD of these sample means will be the SEM. When SEM is simply a statistic calculated from SD, it is a single sample mean. For that reason, the authors of these guidelines feel that it is wrong to report SEM, a theoretical estimate of the variability of possible values of a sample mean about a population mean.

Statisticians have documented the fact that the scientific literature commonly publishes statistical errors. Roughly 50% of published articles have at least one error [8, 11-13], which clearly undermines the process of scientific discovery and dispersion of information. In 2005 alone, 25-38% of articles published in leading scientific journals, such as *Nature, British Medical Journal (BMJ),* and *Nature Medicine*, contained some disparity between reported statistics (*t*-tests, *F*-tests, etc.) and their corresponding *P* values [14].

The errors in journals comprise several different misuses of statistics. Inappropriate application of SEM as a descriptive statistic for measure of variation occurs all too frequently in published papers, possibly because SEM is always smaller than SD. Investigators wish to show data as more precise than they really are. Even worse, some authors use the mean +/–, a number to describe variability without identifying the +/– number as

either SD or SEM. Finally, some authors use SEM as an inferential statistic instead of the 95% CI. To confirm such findings, the authors of this report examined 10 issues of the *Cancer Research and Treatment* from 2008 and 2009, and found just one report in which investigators used the SEM inappropriately.

Conclusion

When data are normally distributed, one should use the mean and SD to describe the measure of central tendency and variability of the observation. If data are not normally distributed, SD is not a good indicator of variability. In that case, mathematical transformation can be used to normalize the data. Alternatively, the median, range, or inter-quartile range may be used to describe the center and variability. For normally distributed data, the CI should be reported when a range for some characteristic of a population is important. This helps us to determine whether the experimental effect is sufficiently different to be relevant. The mean and SEM define a particular characteristic (the mean) of a population and a measure of its precision (SEM). SEM should not be taken as a measure of the variability of observations but rather a means to describe the outcome of a study. In presenting the results of a study of sample observations, one should give the number of subjects (n), the range of results, the central tendency (mean, SD), and the spread (CI for the mean). Proper use of these statistical tools will strengthen confidence in any conclusions drawn by the investigators.

References

1. Igic R, Muacevic G. Introduction to statistical planning and evaluation in biological, pharmacological and clinical research. Tuzla, Medicinski fakultet, 1985 (in Serbo-Croatian).
2. Swinscow TDV (Ed). Statistics at square one. BMJ, London, 1996, 9th Edn.
3. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel. Adv Physiol Educ 2007; 31: 295-298.
4. Campbell MJ (Ed). Statistics at square two: Understanding modern statistical applications in medicine. BMJ, London, 2001.
5. Council of Science Editors SMC. Scientific style and format: The CSE manual for authors, editors, and publishers. Reston, The Council, 2006.
6. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biochemical journals. Med Educ 1999; 33: 66-78.
7. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995; 10: 452-454.
8. Altman DG. Confidence intervals for the number needed to treat. BMJ 1998; 317: 1309-1312.
9. Lang T. Twenty statistical errors even you can found in biomedical research articles. Croat Med J 2004; 45: 361-370.
10. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. Adv Physiol Edu 2004; 28: 85-87.
11. Brown GW. Standard deviation, standard error. Which standard should we use? Am J Dis Child 1982; 136: 937-941.
12. Altman DG, Bland JM. Improving doctors' understanding of statistics. J R Stat Soc Ser A 1991; 154: 223-267.
13. Altman DG. Statistics in medical journal: some recent trends. Stat Med 2000; 19: 3275-3289.
14. Clayton MK. How should we achieve high-quality reporting of statistics in scientific journals? A commentary on "Guidelines for reporting statistics in journals published by the American Physiological Society". Adv Physiol Edu 2007; 31: 302-304.