JOURNAL SERIES:

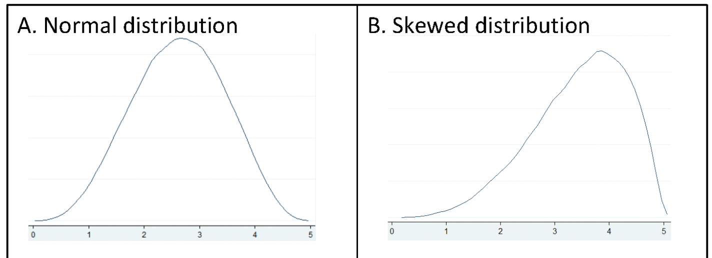
Statistics Review Part 3: Standard Deviations and Confidence Intervals

by Ann Opalka, PharmD, Kevin Look, PharmD, PhD and Amanda Margolis, PharmD, MS, BCACP

This is the third article in a series of articles designed to review the basic, fundamental concepts of biostatistics. This article describes standard deviations, standard error of the mean, and confidence intervals.

Objectives

1. Define standard deviations and standard error of the mean.
2. Define confidence intervals.
3. Describe how to interpret confidence intervals.



A. Characteristics of a normal distribution include a symmetrical bell shaped curve with the mean, median, and mode in the middle of the distribution. The curve is composed of observations that are evenly distributed around the mean. B. Example of a left skewed distribution. Note that the observations are unevenly distributed around the mean, where observations to the right side of the mean are more commonly seen.

FIGURE 1. Comparison of the normal distribution to a skewed distribution

Two common statistical measures that are seen in medical literature are standard deviations and confidence intervals. These are important measures used to describe the variability, or spread, of data. They help clinicians understand the uncertainty of the potential treatment effect, and can be used to determine whether two groups studied differ significantly (i.e., more than by chance alone). Additionally, they can be used to determine if the results of a study are generalizable to a larger population.

Standard Deviations

Standard deviations (SD) determine the approximate distribution of individual data points in regards to a sample mean; therefore, the SD is a measure of the variability in the data. Typically, a study will find a range of individual values or observations, where some values fall either below the mean or above the mean. Thus, individual deviations from the mean can be either positive or negative. Variance is calculated as the mean of the deviations squared, and the SD is calculated as the square root of the variance. The SD provides value to readers when the sample is either normally or nearly normally distributed and is used for continuous variables (Figure 1). If the SD is reported for bimodal or skewed data, it may be falsely larger than expected. Therefore, it would not be an appropriate tool to use to measure the variability for those distributions. A large SD shows that individual data points are further from the mean with a larger spread of the data (more uncertainty) while a small SD shows that they are clustered closer to the mean (less uncertainty).

For example, if the SD for a group's age was 9, the reader can look at the whole sample group and determine how many patients were one SD away from the mean, two SD away from the mean, etc. Essentially this shows how variable the data is around the mean.

Standard Error of the Mean

The SD can also be used to calculate the standard error of the mean (SEM), which is the SD divided by the square root of the sample size (n). Note that a larger sample size will decrease the SEM. The SEM is used to determine how reliable a sample may be when relating it back to a population. It is important to know that the SEM is used in the calculation of the confidence interval, which will be discussed next. Of note, the SEM is often used incorrectly in literature. It should not be used to report variability of a sample (i.e., mean ± SEM); rather, results should be reported as mean ± SD. Using SEM to report variability in the literature provides a false impression of decreased variability. If SEM is provided in a manuscript, the reader can multiply the SEM by the square root of n to obtain the SD in order to help clarify the true variability in the data.

Confidence Intervals

Confidence intervals (CIs) are important as they can be used to determine if the results from a study sample can be applied to the true population of interest and, in some cases, if there is statistical significance. CIs provide additional information on clinical significance when compared to a using a p-value alone. In clinical research a sample of a population is studied and statistics are used to describe this sample (e.g., mean, SD, SEM, etc.). A CI is then calculated such that the true point estimate or mean of the population lies within the range of the CI.

A CI is calculated using the SEM and the mean, where the SEM determines the width of the CI around the mean. CIs are reported in the unit of measure that is being studied (e.g., mg/dl or cm). The CI is applied to continuous data from normal or near-normal distributions, and can be of various widths (e.g., 90%, 95%, 99%, 99.9%). Of note, the 95% CI is most commonly reported. Many practitioners believe a 95% CI means that 95% of the time the true point estimate or value for the population is within the range that was calculated from the sample. However, the technical interpretation of a 95% CI is: within the collection of all 95% CIs that could theoretically be constructed from repeated random samples of the same sample size from the same population, 95% will of them will contain the population value. A CI of 99% (a wider CI) would provide even higher chances that it would include the population value. However, we trade the greater "confidence" of having the population value within our range with lower precision. Consider this: in order to have a 100% CI the range of the confidence interval would be negative infinity to infinity! A larger sample size is more likely to represent the true population value than a smaller sample; therefore, larger sample sizes lead to a smaller SEM and narrowing of the CI which corresponds to increased precision.

CIs can be used to determine statistical significance between groups. If the CI around the difference between two treatment groups includes zero, it can be concluded that there is no statistically significant difference at that level of confidence. However, if it does not include zero and the entire CI is on the same side of zero (either all positive or all negative), it can be concluded that there is a statistically significant effect. As an example, a statistically significant (positive) effect might be reported with a CI of 0.5 to 1.2. In contrast, a CI of -0.6 to 1.2 would not be statistically significant, as the CI includes 0.

CI's can be used with relative risk (RR) or odds ratios (OR) as well. However, in this case, statistical significance is determined based on whether the RR or OR crosses one (instead of zero). For example, if the 95% CI is 0.8 to 1.3 for an RR we can be sure that the "true RR" is between those two numbers. Since the CI crosses one, it is possible that the RR is one, which means that the risk between two groups would be the same and therefore the difference in risk is not statistically significant. When the 95% CI of an RR or OR does not include one then it is considered statistically significant (e.g., 1.3 to 1.7).

For example, suppose the mean age for a group was 40 years of age. The 95% CI for the group may include a range that is 36 to 44 years of age. These results would be reported as a mean of 40 (95% CI: 36-44). That means we would be 95% confident that, based on our sample, the real mean for this population is between 36-44 years of age. If we were to increase the sample size, the 95% range for the CI would get smaller and therefore increase the precision of our estimate.

TABLE 1. Statistical Measures

Statistical measure	*What it is*	*When to use it*
Standard deviation (SD)	SD = √variance Note: variance is the average of the squared deviations from the mean.	• Used to determine variability of a sample around a sample mean • For normally or nearly normally distributed continuous data
Standard Error of the Mean (SEM)	SEM = SD/√n	• Used to calculate confidence intervals • Determines reliability of a sample • Used to describe continuous variables • It is NOT used to describe sample variability
Confidence Intervals (CI)	90%, 95%, 99%	• Used to estimate the range of values that will likely include the value of a population • Used to determine statistical significance at a level of confidence • Used for normal or nearly normally distributed continuous data

Summary

This article reviewed some key statistical concepts including standard deviation, standard error of the mean, and confidence intervals. Standard deviations determine the spread of individual sample points in regards to a sample mean. Confidence intervals can help guide clinicians to interpret data regarding clinical significance and may provide information in regards to generalizing the data from a study sample to the population of interest. The next article in this series will cover probability and statistical inference.

Ann Opalka is a PGY2 Ambulatory Care Pharmacy Resident at the William S. Middleton Memorial Veterans Hospital, Madison, WI. Kevin Look is an Assistant Professor in the Social & Administrative Sciences Division at the University of Wisconsin School of Pharmacy, Madison, WI. Amanda Margolis is a Lecturer at the UW- Madison School of Pharmacy and a Clinical Pharmacist at the William S. Middleton Memorial Veterans Hospital, Madison, WI.

Practice questions

1. Which of the following is the widest confidence interval?
   a. 90%
   b. 95%
   c. 99%
   d. 99.9%

2. The SD is useful for data that is ___________.
   a. Normally distributed only
   b. Normally and nearly normally distributed
   c. Bimodal
   d. It can be used for all data

3. Given that the sample size in a study is 40 subjects and the SD for blood pressure is 15 mmHg, what is the standard error of the mean?
   a. 94.8
   b. 2.37
   c. 0.097
   d. 154.9

Answers:

1. **d** Wider CIs mean that they are more likely to include the true value of the population.

2. **b** SDs can be used for normally distributed and nearly normally distributed data.

3. **b** The SEM is calculated as SD divided by square root of the sample size. Of note, SEM is always smaller than the SD.

References and suggestions for further review:

1. Gaddis GM, Gaddis ML. Introduction to Biostatistics: Part 2, Descriptive Statistics. Annals of Emergency Medicine. 1990;19:309-315.
2. Montori VM, Kleinbart J, Newman TB, et al. Tips for learners of evidence-based medicine: 2. Measures of precision (confidence intervals). Canadian Medical Association or its licensors. 2004;171(6):611-615.
3. Redmond AC, Keenan AM. Understanding statistics: Putting p-values into perspective. Journal of the American Podiatric Medical Association. 2002;92(5):297-305.