Understanding Confidence Intervals

Prepared by: Edward Waltz, Ph.D.
University at Albany School of Public Health; ecwaltz@albany.edu

Graphic 1: At least two measures are needed to describe the distribution of any statistical variable.

(a) **Measure of Central Tendency—** i.e., where is the 'center' of the distribution ?

(b) **Measure of Dispersion—** i.e., how much variation exists in the distribution ?

Statisticians have devised a number of different ways of characterizing the central tendency and dispersion of a distribution. Some of the more common of these are summarized in Graphic 1. Most people have a commonsense understanding of the **average** (or **mean**); the **median** and **mode** are somewhat less widely known.

Measures of dispersion can be as simple as the **range**. Most measures of variation that we use are derived in some way. from the statistical **variance**. The **standard deviation**, for example, is simply the square root of the variance. Ultimately in this training, we are interested in understanding **95% confidence intervals**.

Graphic 2: In this graphic we see that we can sample a large population multiple times. Each of our samples has an average and a variance associated with it. If we aggregated all of the averages from these samples, we can create a sampling distribution of the mean for the population. The standard deviation of this distribution is more properly known as the **standard error of the mean** (SEM or, simply, the **standard error**, SE).

Graphic 3: When our sample is based on at least 100 observations, the distribution of means resembles a normal distribution. This allows us to use standard errors to calculate the 95% confidence intervals. (See the manual for the technique used with smaller samples.) In a normal distribution, the area under the curve is a function of the number of standard deviations from the mean. For example, the range from one standard deviation below the mean to one standard deviation above contains about 68% of the total area. Phrased another way, this means that about 68% of the observations are within one standard deviation of the mean. To calculate 95% confidence intervals we need to know that 95% of the observations in a normal distribution can be found within 1.96 standard deviations either side of the mean.

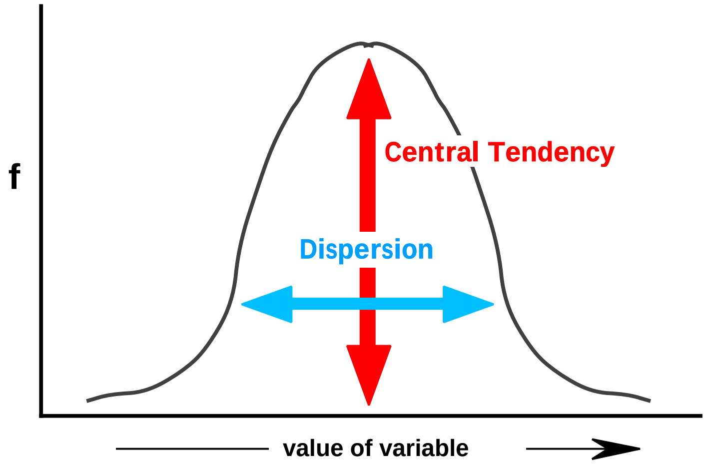
This property of normal distributions allows us to calculate 95% confidence intervals by taking our sample average and adding (and, later, subtracting) the product of 1.96 times the standard error of the mean. The average ± 1.96 * SEM represent the 95% confidence limits of our population. The true mean of our population will be somewhere between these values (or within this **confidence interval**) in 95/100 samples.

Graphic 4: One use of confidence intervals is a "quick-and-dirty" test for differences between the averages measured in two populations. If we calculate the confidence intervals for two populations and find that they do not overlap, then we know that a difference as large as this will occur by chance less than 5% of the time. (i.e., the difference is statistically significant at the $p<0.05$ level) If the confidence intervals do overlap, we conclude that the populations do not differ at $p< 0.05$. Note that this way of testing for significant differences should only be used when at least one sample has fewer than 100 observations. (See the manual for methods for larger samples.)

Note also that this is a fairly conservative way to test for significant differences. In some cases where the overlap is small, other statistical tests (such as t-tests) will detect a significant difference for the same two populations. For comparisons with important consequences, it is advisable to have the results analyzed by someone with more statistical expertise, especially when the overlap in confidence intervals is small.

1. Describing the Distribution of a Variable

Two important characteristics of any distribution are the 'center' and the variability.



Some Measures of Central Tendency

Mean (Average): M= sum (∑) of the values divided by N

Median: 50^{th} percentile of the distribution

Mode: most commonly observed value

Some Measures of Dispersion

Range: Maximum value minus Minimum value

Interquartile range: 25^{th} to 75^{th} percentile

Variance:

s^2= sum of squared deviations from the Mean divided by N

Derivatives of the variance:

Standard deviation: $s= \sqrt{s^2}$

Standard error of the mean:† S.E.M.= $s / \sqrt{N}$

Coefficient of variation: C.V. = (s / M) * 100%

95% Confidence limits (which define the confidence interval):

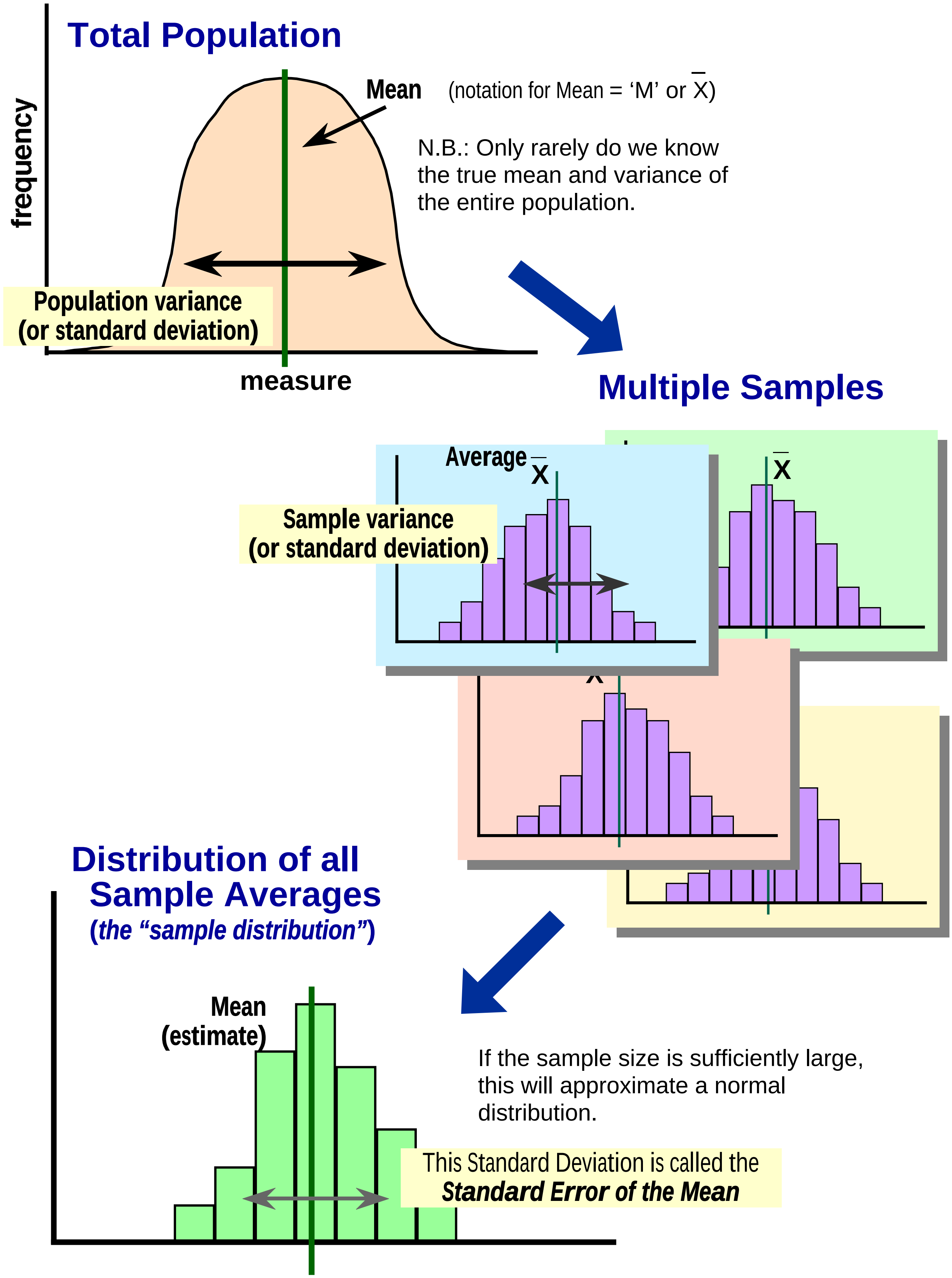
$$M \pm (S.E.M. * C_{95}); \text{ where } C_{95} = 1.96$$

$$(99\%, C_{99}= 2.576 \quad ; 90\%, C_{90}= 1.645)$$

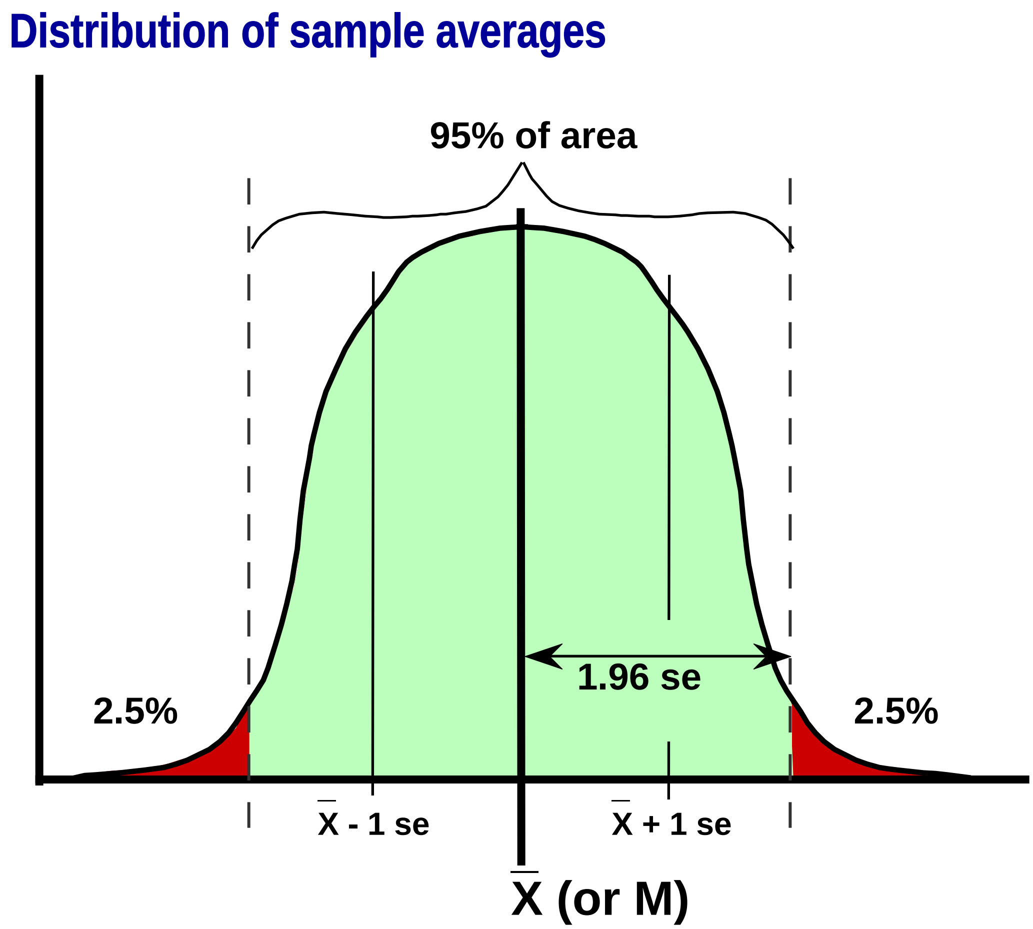
---

† Sometimes simplified as "Standard Error"

2. Visualizing Multiple Distributions



3. Determining Confidence Intervals



In this distribution, 95% of the area under the curve is within the area bounded by

$$\bar{X} - 1.96\ \text{se} \longleftarrow \bar{X} \longrightarrow \bar{X} + 1.96\ \text{se}$$

These 95% confidence limits define the boundaries of the 95% confidence interval.

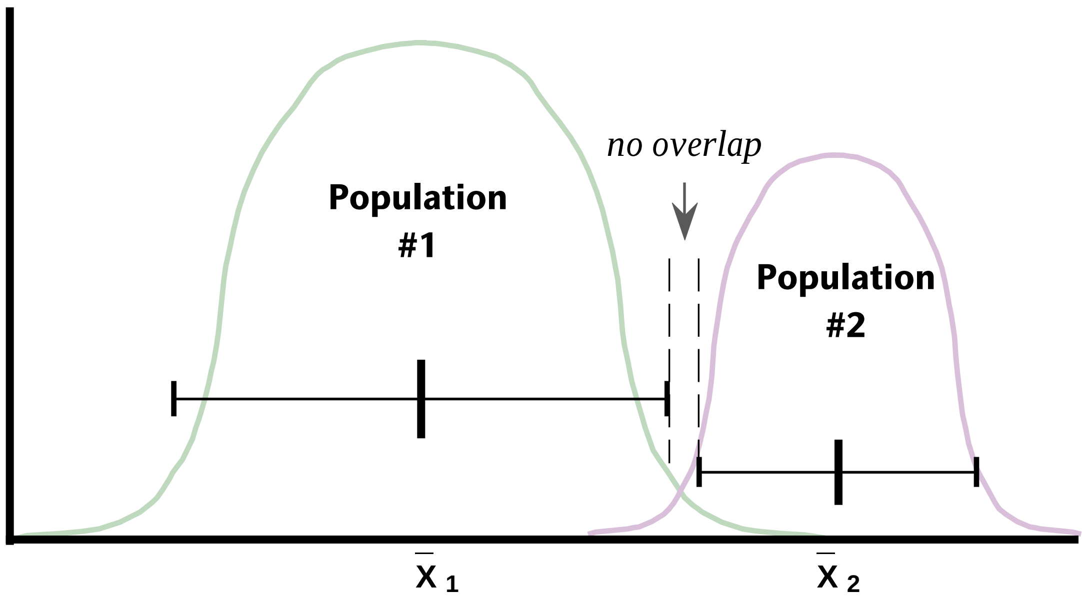
If the sample size is sufficiently large ($n \geq 100$) for any sample average that we measure, then the true mean of the distribution will be within that confidence interval 95 out of 100 times. Smaller sample sizes require alternative methods of calculating confidence intervals.

Note also that 90% of the area under the curve is within 1.645 standard errors of the mean; and 99% of the area is within 2.576 standard errors of the mean. These figures would be used to calculate 90% and 99% confidence limits, respectively.

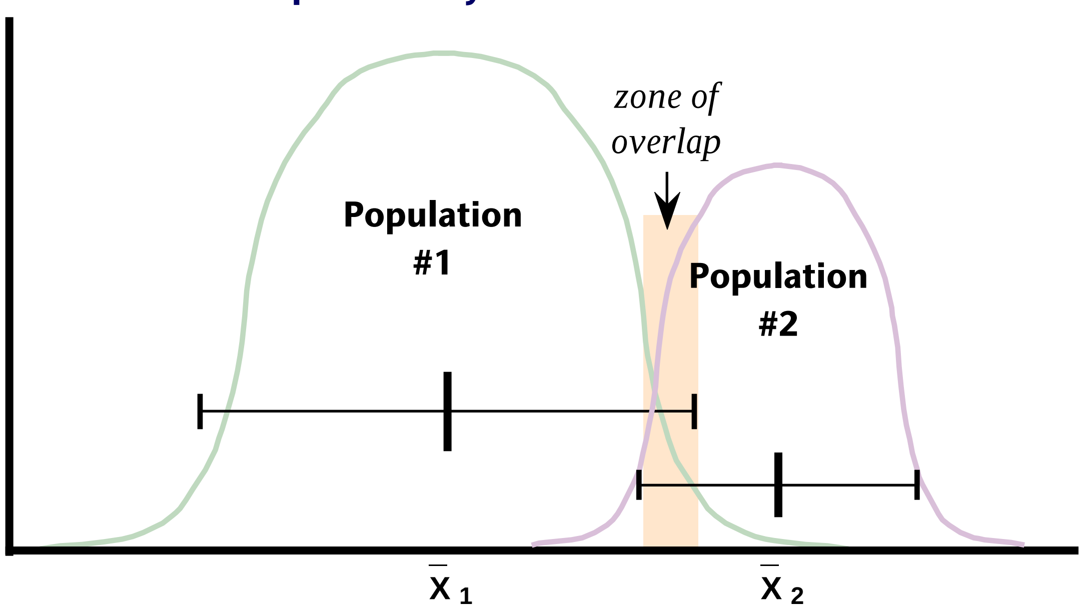
4. *Using Confidence Intervals to Test Hypotheses*

When either of two populations being compared has fewer than 100 observations, their confidence intervals may be used to test for differences between the two averages.

When the confidence intervals do not overlap, the averages are different with a probablility < 0.05



When the confidence intervals do overlap, the averages are not different at probablility < 0.05



Note that this is a conservative method of testing. Other statistical tests may reveal statistical differences between populations which do not differ statistically with this method.