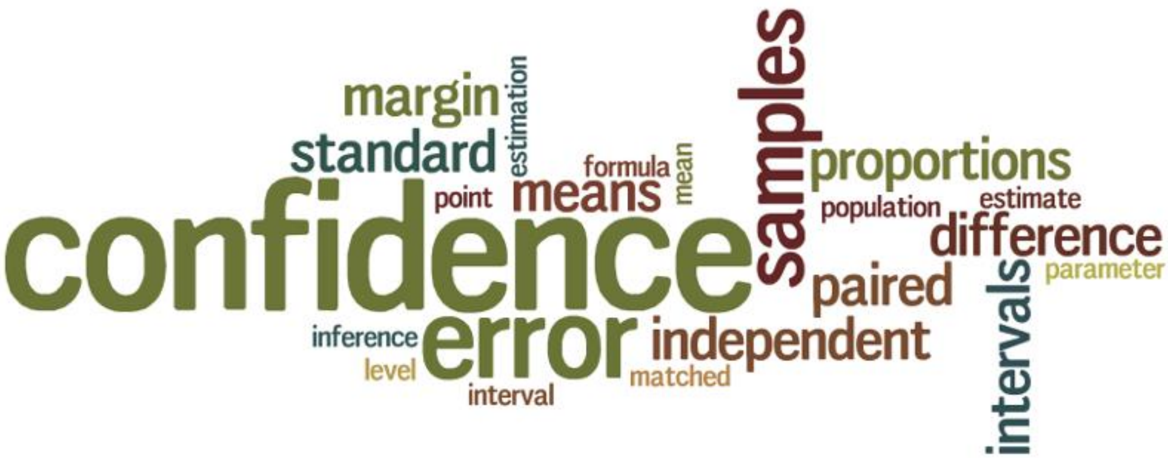
ANATOMY OF A CONFIDENCE INTERVAL



OVERVIEW OF CONFIDENCE INTERVALS

Preface: The term confidence interval is commonly abbreviated (CI). The symbol (±) is read as "plus or minus". The discussion below employs a degree of intentional redundancy for those not versed in the subject matter.

What is a confidence interval?

A confidence interval is an estimation.

A confidence interval estimates a range of values that contains/captures the true population value for a variable of interest. Implicit in this concept, is a level of confidence that this actual/true population value is contained within the estimated range of values.

A confidence interval uses information from a sample (aka sample data) gathered from a larger population. From a purely mathematical viewpoint, ideally this sample has the characteristic of randomness and is known as a *random sample*. However, in practice, true randomness is difficult to achieve. Accordingly, the goal is to have a sample that is representative of the population of interest. Hence, the goal is to obtain a *representative sample*.

The information from this *Representative* sample is used to calculate an estimate and a surrounding range of values for a variable of interest.

CONFIDENCE INTERVAL AND CONFIDENCE LEVEL

While the concepts are related, there is a difference between a confidence interval and a confidence level.

The confidence level is the long-run probability that a series of confidence intervals will contain the true value of a population parameter. As an example, data from a sample are used generate a sample estimate for a population mean. If a different sample is taken, it will likely generate a slightly a different estimate for this true population average. If you repeated this procedure multiple times, by drawing samples (of the same size) from the same population, the confidence level is the percentage of these samples that will capture the actual population parameter.

The confidence level is specified/controlled by those conducting a researcher endeavor. However, the design of the experiment may dictate the magnitude of the level to use.

While there is a level of confidence associated with a CI, a confidence interval is a range of values; it consists of an upper and lower bound. This range is designed to capture the population parameter. Hence, a confidence interval has an associated confidence level and vice versa. This level of confidence determines how the bounds of the confidence interval are calculated. The direct association between the confidence level and the confidence interval means larger confidence levels result in larger/wider confidence intervals.

The following example illustrates the difference between a confidence interval and a confidence level:

Data from a representative sample were used to generate a 95% confidence interval for the proportion of female babies born each year. The result yields an upper bound of 0.56 and a lower bound of 0.48. This 95% confidence interval is reported as *95% CI (0.48, 0.56)*. Associatively, there is a 95% confidence level that this interval will contain the true population proportion.

A confidence interval may be calculated for numerous confidence levels, but the most commonly used value is 95 percent.

BREAKING DOWN CONFIDENCE INTERVALS

Informally, a confidence interval (CI) is a range of values (aka an interval) which is likely to contain a population value (aka a parameter) for "something" that is being estimated. This interval contains/surrounds what is known as an estimate (aka a point estimate). Ideally, a narrow confidence interval (CI) if desirable because it conveys a better estimation of the actual population value.

The "something" being estimated   may include:

- **population proportions**
- **population means (arithmetic averages)**
- **differences between population means or proportions**
- **estimates of variation among groups**

Constructing a CI begins by calculating a point estimate from the sample data. However, a point estimate doesn't provide any information about the variation around this average value. Accordingly, confidence intervals are used to communicate variation around a point estimate. This variation is measured via an interval of values known in statistical jargon as the margin of error.

Remember, both the point estimate and the margin of error are derived from the sample data.

THE POINT ESTIMATE   &   THE MARGIN OF ERROR

THE POINT ESTIMATE

A point estimate is the "average" value calculated from the sample data.

Examples of everyday point estimates:

- **From a representative sample, the average height can be used to estimate the average height of a larger population -- The average height of men in the US is 5 feet 11 inches.**

- **The average proportion voters favoring a ballot proposal from a representative sample can be used to estimate the proportion of a larger population -- 68% of Michigan voters favor proposal A.**

Pulling things together so far…

Informally, a confidence interval (CI) takes the form:

A Calculated Average Value   ±   A Calculated Range of Values

Recall, that this average value is called the point estimate and that this range of values is known as the margin of error.

So, formally we have:

A confidence interval consists of two parts --

A POINT ESTIMATE   ±   A MARGIN OF ERROR

The symbol (±) is read as "plus or minus".

Before delving into what constitutes a margin of error, a couple of examples may help to solidify the discussion thus far.

EXAMPLE #1:

You survey 100 people about their television-viewing habits, and find they view an average of 35 hours of television per week. The survey has a margin of error of 1.4 hours.

Point estimate is 35 hours. Margin of error is 1.4 hours

CI is -- 35 hours ± 1.4 hours

This confidence interval is reported as: 95% CI [33.6, 36.4] *

EXAMPLE #2:

Based on a representative sample, 52% of all births are female. The sample has a margin of error of 4%.

Point estimate is 52% of births. Margin of error is 4% of births.

CI is -- 52% of births ± 4% of births

This confidence interval is reported as: 95% CI [0.48, 0.56] *

***APA Style -- Reporting results may differ and are dependent on various writing style requirements.**

DISSECTING THE MARGIN OF ERROR

THE MARGIN OF ERROR

The margin of error consists of two parts:

1. **A standard error**
2. **A distribution multiplier**

A margin of error is a range of values calculated from the sample data; the point estimate is contained within this interval.

Examples of margin of error estimates:

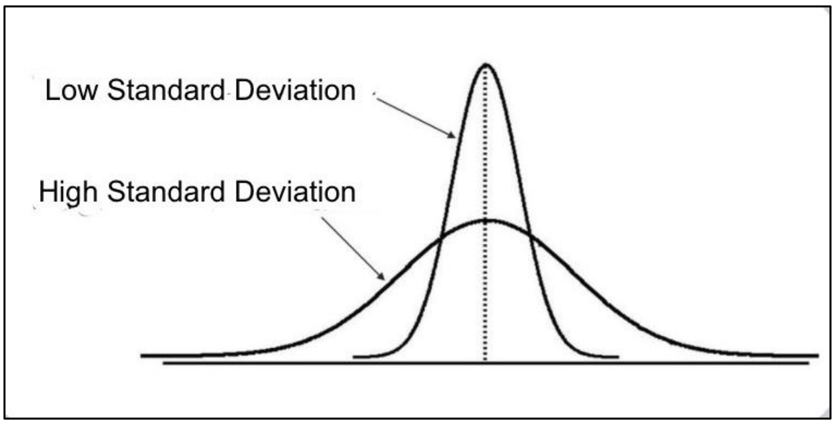
- **From a representative sample, the range of heights can be used to estimate the average range of heights of a larger population -- The average height of men in the US is between 5 feet 8 inches and 6 feet 2 inches. The margin of error is 6 inches (6’ 2” minus 5’ 8”).**

- **The range of the proportion of voters favoring a ballot proposal from a representative sample can be used to estimate the range of the proportions of a larger population -- The percentage of Michigan voters who favor proposal A is between 64% and 72%. The margin of error is 8% (72% minus 64%).**

THE STANDARD ERROR

The standard error (SE) is related to standard deviation of a dataset. The standard error is very similar to standard deviation. Both are measures of spread. The more spread out your data are, the greater the standard deviation

The standard deviation tells you how spread out the data are. It is a measure of how far observed values are from the center (average) of the population distribution. This concept is graphically illustrated below:



NOTE: Smaller standard deviations have values that are concentrated nearer to the center/average value of a dataset

---

When conducting research, you often only collect data for a small sample of the entire population. This sample has an error associated with it. This occurs because each sample will be slightly different and each will produce slightly different estimates.

Consider a study examining the mean of a population. Because your sample is not identical to your population, the sample mean is likely somewhat larger or smaller than the actual population mean. If you were to repeat the sampling process, you would get numerous sample means, each deviating from the true population parameter to some degree. This collection of sample means forms a distribution of its own, known as a sampling distribution. The standard error is calculated using this distribution of estimates that are different form the population mean. In other words, these estimates are called errors because they do not exactly match the true population mean; hence, the terminology standard "error".

The standard error measures how different a sample estimate is likely to be from a population value. It tells you how the estimate would vary if you were to repeat a study using different samples (of the same size) taken from a single population. In other words, it quantifies the difference between a parameter (i.e., a population value) and a statistic (i.e., a sample value).

Depending on the type of statistic being examined (e.g., means, proportions, medians), there are corresponding formulas for calculating the standard error. For example, for a mean the standard error (aka the standard error of the mean or SEM) is calculated as:

$$SEM = \frac{\sigma}{\sqrt{n}}$$

- ***SEM*** **is the Standard Error of the Mean**
- ***σ*** **is the Population Standard Deviation**
- ***n*** **is the Sample Size**

Regardless of the type of statistic being examined, the denominator in the formula used to calculate the standard error contains the size of the sample (n). Hence, as the sample size increases the dominator also increases; this has the effect of decreasing the overall value of the standard error.

The consequence and significance of this relationship is…

The Larger the Sample Size… the Smaller the Standard Error

THE DISTRIBUTION MULTIPLIER

- **most often referred to as just the multiplier**
- **occasionally called a t-multiplier or z-multiplier**
- **sometimes termed a critical value**

The multiplier, is a function of the underlying population and the size of the confidence interval. Typical confidence intervals range from about 50% to 99.9%, with the most common being 95%.

The concept of the multiplier is more easily explained and understood by looking at an example containing specific multiplier values used to generate a margin of error…

The table below lists seven multipliers for normal distributions (aka z distributions) and their corresponding confidence levels. These "z multipliers" are used when an assumption of an underlying normal population distribution is satisfied.

Confidence Level	z multiplier
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291

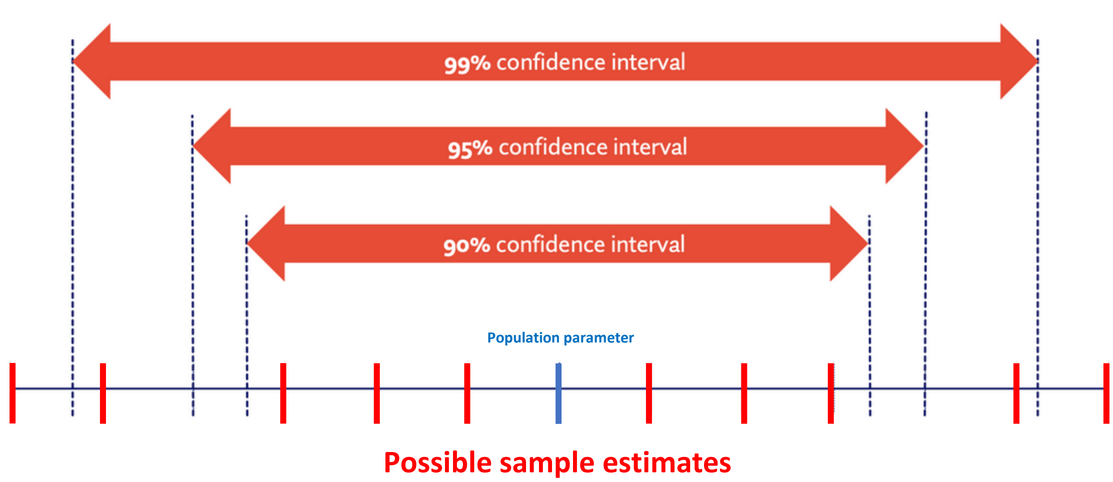
Please note, that as confidence levels increase, so too do the magnitudes of the multipliers. These values are multiplied by the standard error value to generate the margin of error; so, these larger multipliers will increase the magnitude of the margin of error. Recall, that the overall confidence interval is calculated by "adding" and "subtracting" (±) the margin of error to the point estimate.

Consequently, the greater the confidence level, the greater the multiplier, the greater the margin of error, and the greater/wider the associated confidence interval.

Confidence Level ⇧ ⇨ Multiplier ⇧ ⇨ Margin of Error ⇧ ⇨ Confidence Interval range/size ⇧

The size/width of a confidence interval varies depending on the selected level of confidence.

Accordingly, for a given sample, the size/width of a 95% confidence interval is greater than the size/width of a 90% CI.



The fact that a greater degree of confidence generates a wider interval may seem counterintuitive. A higher confidence level implies greater certainty and appears to imply greater precision. However, a larger confidence level produces a wider interval that is less precise. Higher confidence levels, increase the interval width to help ensure that the truth is captured. Conversely, more narrow intervals (associated with lesser confidence levels) are less likely to capture the truth; that's why there is less confidence that the interval contains the population parameter. Hence, a narrow a 50% interval is less likely to contain the truth, but is more precise; compared to a wider 95% interval that has a greater chance to contain the population value, but is less precise.

****

SUMMARIZING THE ANATOMY OF A CONFIDENCE INTERVAL

****

A confidence interval is composed of two parts…

POINT ESTIMATE ± MARGIN OF ERROR

Additionally, the MARGIN OF ERROR consists of two parts:

- **the multiplier**
- **the standard error**

Accordingly, a formal definition of confidence interval is…

POINT ESTIMATE ± (multiplier * standard error)

Working through an example may help to reinforce the procedure used to generate confidence intervals.

Note: The example uses scenarios with the same standard error values and point estimates; this is done for purposes of illustration/comparison. In practice, standard errors and point estimates will almost always differ between samples.

EXAMPLE: **Data from a representative sample were used to determine that the mean number a statisticians it takes to change a light bulb is 3.37 with a standard error of 0.75.**

95% CI Scenario:

Sample data info: POINT ESTIMATE = 3.37 ● z multiplier (95% confidence) = 1.96 ● standard error = 0.75

MARGIN OF ERROR ➔ 1.96 * 0.75 = 1.47

3.37 ± 1.47 ➔ 3.37 - 1.47 = 1.90 and 3.37 + 1.47 = 4.84

95% CI [1.90, 4.84]*

Interpreting these results, we conclude…

There is 95% statistical confidence that the true average (mean) number of statisticians it takes to change a light bulb is between 1.90 and 4.84.

90% CI Scenario:

Sample data info: POINT ESTIMATE = 3.37 ● z multiplier (90% confidence) = 1.645 ● standard error =0.75

MARGIN OF ERROR ➔ 1.645 * 0.75 = 1.23

3.37 ± 1.23 ➔ 3.37 - 1.47 = 2.14 and 3.37 + 1.23 = 4.60

95% CI [2.14, 4.60]*

Interpreting these results, we conclude…

There is 90% statistical confidence that the true average (mean) number of statisticians it takes to change a light bulb is between 2.14 and 4.60.

***APA Style -- Reporting results may differ and are dependent on various writing style requirements.**

CONFIDENCE AND PRACTICAL APPLICATIONS

A confidence level of 0% indicates that you have no faith at all that the CI contains the population value of interest. Conversely, a 100% confidence level means that there is no doubt at all that if the interval captures the true population value. Both examples convey information, but neither is meaningful. In this vein, consider the following statements of results:

A claim professing to be… 100% confident that the true amount of sunshine for a given day is between 0 % and 100% is certain to capture the true level of sunshine. While the

corresponding wide confidence interval will always capture the truth, the information conveyed is of no value.

Whereas, asserting that we are… 95% confident the true amount of sunshine for a given day will be between 20% and 40%, isn't guaranteed to capture the true percentage of sunshine. Although result this has a has a lower level of confidence (and may not always capture the truth), it does convey useful information.

CONFIDENCE INTERVALS AND PROBABILITY

Finally, a common misperception is that a confidence interval is "directly" associated with the probability of the variable being measured. This examination of probability and confidence intervals will be discussed in the context of the following statement -- There is 95% confidence that the true proportion of voters who favor candidate A is between 42% and 48%. The 95% interval associated with this statement is NOT a range of probabilities for the proportion of voters who favor candidate A; rather, the 95% is associated with the confidence in the success rate of the procedure/process used to generate the interval. Hence, it is not correct to say that there is a 95% probability that the true percentage favoring candidate A is between 42% and 48%. Similarly, it is not true that the interval has 95% probability to include the population parameter.

In actuality, the probability that the true proportion (i.e., the population parameter) falls within the confidence interval is either 0 or 1. In other words, a sample is used to generate an interval that either contains the population parameter (i.e., the true proportion) or it does not contain the population parameter; there is no way of being certain, short of sampling everyone. In other words, we cannot know the true population parameter without conducting a census.

If we were to select many different samples (of the same size) and construct the corresponding confidence intervals, 95% of them would actually contain the value of the population proportion and 5% of the time the true population proportion will lie outside the interval. This is a subtle, but important, difference from claiming that there is a 95% probability that the true parameter lies between the two bounds of the interval.

SIDE NOTE: Confidence intervals are associated with the realm of Frequentist Statistics; this is the most common field of statistical philosophy. However, there is a sphere of statistical reasoning known as Bayesian Statistics. In the Bayesian realm, there is a direct connection between probability and the interval of a parameter being estimated. These types of intervals are called "credible intervals".

Q.E.D.