# Confidence Intervals (Page 351-363, Chapter 14)

**TODAY YOU WILL BE ABLE TO…**

* Define statistical inference
* Describe statistical estimation
* Describe the parts of a confidence interval
* Interpret a confidence level
* Construct and interpret a confidence interval for the mean of a Normal population
* Describe how confidence intervals behave

**RECALL**

* The mean, , of a sampling distribution is equal to the population mean, .
* A sample mean from a single sample, , is only an **estimate** of the population mean.
* The mean, either of a sampling distribution or a single sample, is useless without a measure of spread.
  + The standard deviation of a sampling distribution is given by .
  + The standard deviation of a single sample is given by

**STATISTICAL INFERENCE**

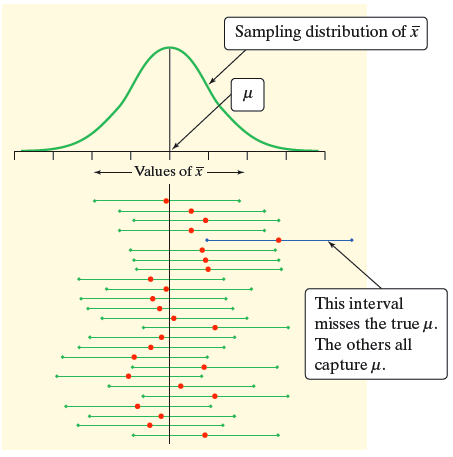
Statistical inference provides methods for drawing conclusions about a population from sample data. Statistical inference uses the language of probability to say how trustworthy our conclusions are.

Knowing that the sampling distribution of a statistic, such as the mean, is Normally distributed when the sample size is large allows us to draw conclusions about a population even when we cannot take many samples.

**STATISTICAL ESTIMATION**

The sample mean and sample standard deviation are **point estimates** and, by themselves, are of limited use because they do not contain any measure of uncertainty.

A **confidence interval** includes both the point estimate and a measure of uncertainty. This interval estimate is one of the most common forms of statistical inference.



**PARTS OF A CONFIDENCE INTERVAL**

The sampling distribution of tells us how close to µ the sample mean is likely to be.

A **confidence interval** for a parameter has two parts:

1. An interval calculated from the data, which has the general form

**estimate ± margin of error**

The **margin of error** is themeasure of uncertainty and is calculated as the product of the standard deviation of the statistic and a critical value† based on the chosen level of confidence such as 90%, 95%, or 99%.

1. A **confidence level**, which gives the probability that the interval will capture the true parameter value in repeated samples. That is, the confidence level is the success rate for the method.

† When the population is Normal and the population standard deviation is known, the critical value is a z-score, denoted, chosen at (1-α)/2 where α is the probability of making an error.

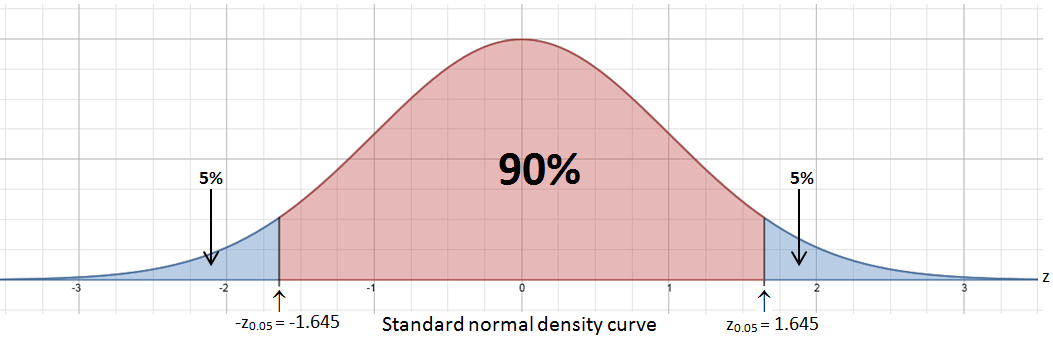
**CONFIDENCE LEVEL**

The **confidence level**, C%, is the overall capture rate if the method is used many times. The sample mean will vary from sample to sample, but when we use the method *estimate ± margin of error* to get an interval based on each sample, C% of these intervals capture the unknown population mean *µ*.

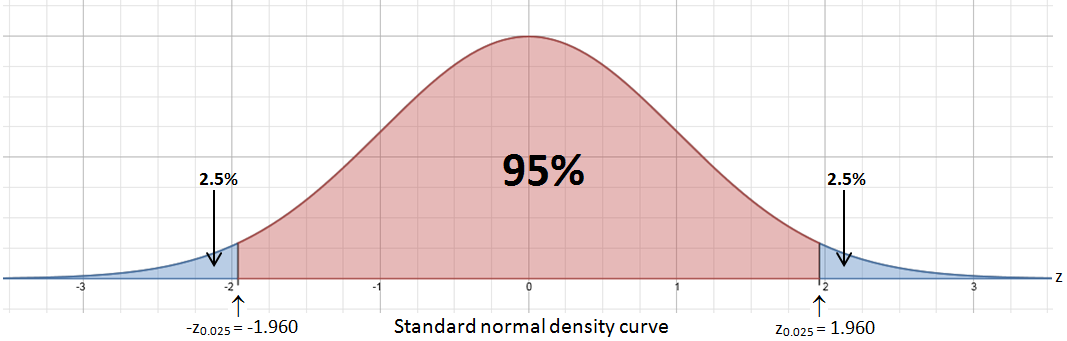
We usually choose a confidence level of 90% or higher because we want to be quite sure of our conclusions. The most common confidence level is 95%.

To say that we are 95% *confident* is shorthand for“95*% of* all possible samples of a given size from this population will result in an interval that captures the unknown parameter.”

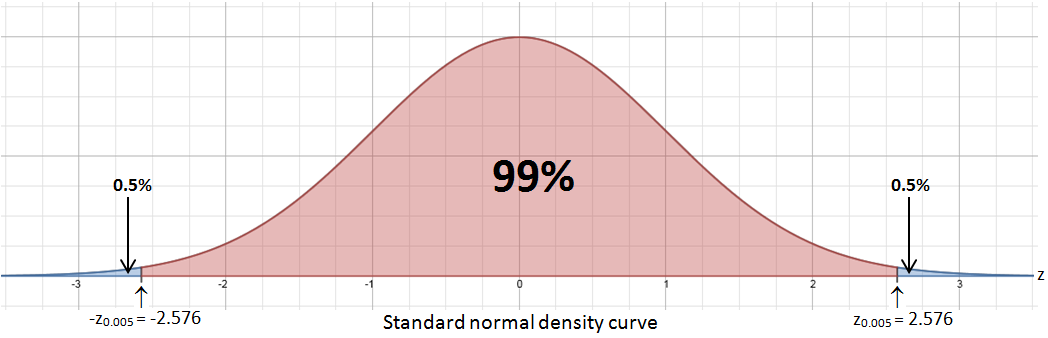
A 90% confidence interval is an interval that “captures” the middle 90% of the sampling distribution of the statistic. Notice that the middle 90% of the distribution is obtained by considering the numbers that are 1.645 standard deviations to the left and to the right of the mean. The number 1.645 is denoted by z0.05 since it “cuts off” the upper and lower 5% of any normal distribution.



Likewise, a 95% confidence interval captures the middle 95% of the sampling distribution of the statistic. The middle 90% of the distribution is obtained by considering the numbers that are 1.960 standard deviations to the left and to the right of the mean. The number 1.960 is denoted by z0.025 since it “cuts off” the upper and lower 2.5% of any normal distribution.



Finally, a 99% confidence interval captures the middle 99% of the sampling distribution of the statistic. The middle 99% of the distribution is obtained by considering the numbers that are 2.576 standard deviations to the left and to the right of the mean. The number 2.576 is denoted by z0.005 since it “cuts off” the upper and lower 0.5% of any normal distribution.



These (and similarly determined) numbers are important in constructing a confidence interval.

**CONDITIONS FOR ESTIMATION**

There will be different formulas for the **margin of error** depending on the population and sample conditions, but the basic idea of confidence intervals is the same no matter what the conditions are. Therefore, we will introduce confidence intervals with the assumption that the conditions outlined below are true.

**Simple Conditions for Inference About a Mean**

*The conditions that we have a perfect SRS, that the population is exactly Normal, and that we know the population standard deviation are all unrealistic.*

1. We have an SRS from the population of interest. There is no nonresponse or other practical difficulty.
2. The variable we measure has an exactly Normal distribution *N*(*μ*,*σ*) in the population.
3. We don’t know the population mean μ, but we do know the population standard deviation *σ*.

**THE IDEA OF STATISTICAL ESTIMATION**

Consider a sample of size 16 and with mean = 240.79. This sample is from a population that has a Normal distribution. We know the population standard deviation, σ = 20, but not the population mean, μ = ?. We could guess that *µ* is “somewhere” around 240.79. **But how close to 240.79 is *µ* likely to be?**

To answer this question, we must ask another:

**How would the sample mean vary if we took many SRSs of size 16 from the population?**

|  |  |  |
| --- | --- | --- |
| Picture 1.png | Picture 1.png | Picture 1.png |

* In repeated samples, the values of the sample mean will follow a Normal distribution with mean *µ* and standard deviation 5.

= = 5

* The 68-95-99.7 Rule tells us that in 95% of all samples of size 16, the sample mean will be within about two standard deviations of *µ*, specifically within 1.96 standard deviations: 1.96\*5 or 9.8.
* If the sample mean is within 9.8 points of *µ*, then *µ* is within 9.8 points of the sample mean.
* Therefore, the interval from 9.8 points below to 9.8 points above the sample mean will “capture” *µ* in about 95% of all samples of size 16.
* If we estimate that the interval 230.79 to 250.79 contains *µ*, we’d be calculating an interval using a method that captures the true *µ* in about 95% of all possible samples of this size.

**Conclusion:** We are 95% confident that the interval, 230.79 to 250.79, captures the true population mean *µ.*

**CONFIDENCE INTERVAL ABOUT THE MEAN**

From the general form, **estimate ± × (standard deviation of the statistic)**, we derive the following formula for the confidence interval about the mean when the **Simple Conditions for Inference About a Mean** on page 4 are true: Normal population with known standard deviation.

± ×

**HOW CONFIDENCE INTERVALS BEHAVE**

The confidence interval for the mean of a Normal population illustrates several important properties that are shared by all confidence intervals in common use.

* The user chooses the confidence level and the margin of error follows.
* We would like high confidence and a small margin of error.
* High confidence suggests our method almost always gives correct answers.

A small margin of error suggests we have pinned down the parameter precisely.

The margin of error for the confidence interval is:

×

The margin of error gets smaller when

* gets smaller (the same as a lower confidence level *C%*)
* *σ* is smaller. It is easier to pin down *µ* when *σ* is smaller.
* *n* gets larger. Since *n* is under the square root sign, we must take four times as many observations to cut the margin of error in half.

**FOUR-STEP PROCESS**

**State:** What is the practical question that requires estimating a parameter?

**Plan:** Identify the parameter, choose a level of confidence, and select the type of confidence interval that fits your situation.

**Solve:** Carry out the work in two phases:

1. **Check the conditions** for the interval that you plan to use.

2. Calculate the **confidence interval**.

**Conclude:** Return to the practical question to describe your results in this setting.