

## Confidence Intervals Introduction

### Prerequisites

[Introduction to Estimators](#), [Characteristics of Estimators](#)

Say you were interested in the mean weight of 10-year old girls living in the United States. Since it would have been impractical to weigh all the 10-year old girls in the United States, you took a sample of 16 and found that the mean weight was 90 pounds. This [sample](#) mean of 90 is a [point estimate](#) of the [population](#) mean. A point estimate by itself is of limited usefulness because it does not reveal the uncertainty associated with the estimate; you do not have a good sense of how far this sample mean may be from the population mean. For example, can you be confident that the population mean is within 5 pounds of 90? You simply do not know.

Confidence intervals provide more information than point estimates. Confidence intervals for means are intervals constructed using a procedure (presented in the [next section](#)) that will contain the population mean a specified proportion of the time, typically either 95% or 99% of the time. These intervals are referred to as 95% and 99% confidence intervals respectively. An example of a 95% confidence interval is shown below:

$$72.85 < \mu < 107.15$$

There is good reason to believe that the population mean lies between these two bounds of 72.85 and 107.15 since 95% of the time confidence intervals contain the true mean.

If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain the population mean. Naturally, 5% of the intervals would not contain the population mean.

It is natural to interpret a 95% confidence interval as an interval with a 0.95 probability of containing the population mean. However, the proper interpretation is not that simple. One problem is that the computation of a confidence interval does not take into account any other information you might have about the value of the population mean. For example, if numerous prior studies had all found sample means above 110, it would not make sense to conclude that there is a 0.95 probability that the population mean is between 72.85 and 107.15. What about situations in which there is no prior information about the value of the population mean? Even here the interpretation is

complex. The problem is that there can be more than one procedure that produces intervals that contain the population parameter 95% of the time. Which procedure produces the "true" 95% confidence interval? Although the various methods are equal from a purely mathematical point of view, the standard method of computing confidence intervals has two desirable properties: each interval is symmetric about the point estimate and each interval is contiguous. Recall from the [introductory section](#) in the chapter on probability that for some purposes, probability is best thought of as subjective. It is reasonable, although not required by the laws of probability, that one adopt a subjective probability of 0.95 that a 95% confidence interval as typically computed contains the parameter in question.

Confidence intervals can be computed for various parameters, not just the mean. For example, later in this chapter you will see how to compute a confidence interval for  $\rho$ , the population value of Pearson's  $r$  based on sample data.