

## 8.1: Basics of Confidence Intervals

A point estimator is just the statistic that you have calculated previously. As an example, when you wanted to estimate the population mean,  $\mu$ , the point estimator is the sample mean,  $\bar{x}$ . To estimate the population proportion,  $p$ , you use the sample proportion,  $\hat{p}$ . In general, if you want to estimate any population parameter, we will call it  $\theta$ , you use the sample statistic,  $\hat{\theta}$ .

Point estimators are really easy to find, but they have some drawbacks. First, if you have a large sample size, then the estimate is better. But with a point estimator, you don't know what the sample size is. Also, you don't know how accurate the estimate is. Both of these problems are solved with a confidence interval.

### Definition 8.1.1

**Confidence interval:** This is where you have an interval surrounding your parameter, and the interval has a chance of being a true statement. In general, a confidence interval looks like:  $\hat{\theta} \pm E$ , where  $\hat{\theta}$  is the point estimator and  $E$  is the margin of error term that is added and subtracted from the point estimator. Thus making an interval.

### Interpreting a confidence interval:

The statistical interpretation is that the confidence interval has a probability ( $1 - \alpha$ , where  $\alpha$  is the complement of the confidence level) of containing the population parameter. As an example, if you have a 95% confidence interval of  $0.65 < p < 0.73$ , then you would say, "there is a 95% chance that the interval 0.65 to 0.73 contains the true population proportion." This means that if you have 100 intervals, 95 of them will contain the true proportion, and 5% will not. The wrong interpretation is that there is a 95% chance that the true value of  $p$  will fall between 0.65 and 0.73. The reason that this interpretation is wrong is that the true value is fixed out there somewhere. You are trying to capture it with this interval. So this is the chance is that your interval captures it, and not that the true value falls in the interval.

There is also a real world interpretation that depends on the situation. It is where you are telling people what numbers you found the parameter to lie between. So your real world is where you tell what values your parameter is between. There is no probability attached to this statement. That probability is in the statistical interpretation.

The common probabilities used for confidence intervals are 90%, 95%, and 99%. These are known as the confidence level. The confidence level and the alpha level are related. For a two-tailed test, the confidence level is  $C = 1 - \alpha$ . This is because the  $\alpha$  is both tails and the confidence level is area between the two tails. As an example, for a two-tailed test ( $H_A$  is not equal to) with  $\alpha$  equal to 0.10, the confidence level would be 0.90 or 90%. If you have a one-tailed test, then your  $\alpha$  is only one tail. Because of symmetry the other tail is also  $\alpha$ . So you have  $2\alpha$  with both tails. So the confidence level, which is the area between the two tails, is  $C = 1 - 2\alpha$ .

### Example 8.1.1 stating the statistical and real world interpretations for a confidence interval

- Suppose you have a 95% confidence interval for the mean age a woman gets married in 2013 is  $26 < \mu < 28$ . State the statistical and real world interpretations of this statement.
- Suppose a 99% confidence interval for the proportion of Americans who have tried marijuana as of 2013 is  $0.35 < p < 0.41$ . State the statistical and real world interpretations of this statement

#### Solution

- Statistical Interpretation: There is a 95% chance that the interval  $26 < \mu < 28$  contains the mean age a woman gets married in 2013.  
Real World Interpretation: The mean age that a woman married in 2013 is between 26 and 28 years of age.
- Statistical Interpretation: There is a 99% chance that the interval  $0.35 < p < 0.41$  contains the proportion of Americans who have tried marijuana as of 2013. Real World Interpretation: The proportion of Americans who have tried marijuana as of 2013 is between 0.35 and 0.41.

One last thing to know about confidence is how the sample size and confidence level affect how wide the interval is. The following discussion demonstrates what happens to the width of the interval as you get more confident.

Think about shooting an arrow into the target. Suppose you are really good at that and that you have a 90% chance of hitting the bull's eye. Now the bull's eye is very small. Since you hit the bull's eye approximately 90% of the time, then you probably hit

inside the next ring out 95% of the time. You have a better chance of doing this, but the circle is bigger. You probably have a 99% chance of hitting the target, but that is a much bigger circle to hit. You can see, as your confidence in hitting the target increases, the circle you hit gets bigger. The same is true for confidence intervals. This is demonstrated in *Figure 8.1.1*.

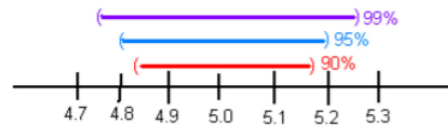


Figure 8.1.1: Affect of Confidence Level on Width

The higher level of confidence makes a wider interval. There's a trade off between width and confidence level. You can be really confident about your answer but your answer will not be very precise. Or you can have a precise answer (small margin of error) but not be very confident about your answer.

Now look at how the sample size affects the size of the interval. Suppose *Figure 8.1.2* represents confidence intervals calculated on a 95% interval. A larger sample size from a representative sample makes the width of the interval narrower. This makes sense. Large samples are closer to the true population so the point estimate is pretty close to the true value.

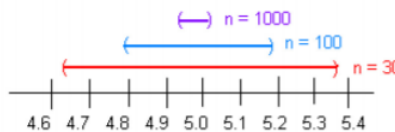


Figure 8.1.2: Affect of Sample Size on Width

Now you know everything you need to know about confidence intervals except for the actual formula. The formula depends on which parameter you are trying to estimate. With different situations you will be given the confidence interval for that parameter.

## Homework

### Exercise 8.1.1

1. Suppose you compute a confidence interval with a sample size of 25. What will happen to the confidence interval if the sample size increases to 50?
2. Suppose you compute a 95% confidence interval. What will happen to the confidence interval if you increase the confidence level to 99%?
3. Suppose you compute a 95% confidence interval. What will happen to the confidence interval if you decrease the confidence level to 90%?
4. Suppose you compute a confidence interval with a sample size of 100. What will happen to the confidence interval if the sample size decreases to 80?
5. A 95% confidence interval is  $6353 \text{ km} < \mu < 6384 \text{ km}$ , where  $\mu$  is the mean diameter of the Earth. State the statistical interpretation.
6. A 95% confidence interval is  $6353 \text{ km} < \mu < 6384 \text{ km}$ , where  $\mu$  is the mean diameter of the Earth. State the real world interpretation.
7. In 2013, Gallup conducted a poll and found a 95% confidence interval of  $0.52 < p < 0.60$ , where  $p$  is the proportion of Americans who believe it is the government's responsibility for health care. Give the real world interpretation.
8. In 2013, Gallup conducted a poll and found a 95% confidence interval of  $0.52 < p < 0.60$ , where  $p$  is the proportion of Americans who believe it is the government's responsibility for health care. Give the statistical interpretation.

### Answer

1. Narrower
3. Narrower
5. See solutions
7. See solutions

This page titled [8.1: Basics of Confidence Intervals](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Kathryn Kozak](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.