

# Confidence Intervals

February 5, 2025

## Contents

<b>1</b>	<b>Definitions and Variables</b>	<b>1</b>
<b>2</b>	<b>Sampling Distributions</b>	<b>2</b>
2.1	Sampling Distributions - Mean . . . . .	3
<b>3</b>	<b>Confidence Intervals</b>	<b>3</b>
3.1	Confidence Interval - Proportions . . . . .	3
3.2	Confidence Interval - One Proportion . . . . .	4
3.3	Confidence Interval - Two Proportions . . . . .	4
3.4	Confidence Intervals - Mean . . . . .	4
3.5	Confidence Intervals - Two Means . . . . .	5

## 1 Definitions and Variables

**Definition 1.1** (Confidence Interval). An interval computed from sample data that provides a range of plausible values for a population parameter.

**Definition 1.2** (Confidence Level). A number that tells how "confident" we are in the interval.

**Definition 1.3** ( $z^*$ ). Critical Z value or critical value is the number of standard deviations a data point needs to be from the mean to be considered statistically significant in the population.

There are two methods to finding  $z^*$ :

1. using the formula

$$z^* = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (1)$$

2. using the calculator:

- 2nd + vars
- 3. or invNorm()

- tail as center and use the given area, the positive number will be your  $z^*$
- likewise, tail as left, take 1 - area and divide by 2

**Definition 1.4** ( $t^*$ ). The critical t value is a "cut off point" which helps you decide whether to support or reject a null hypothesis; If the test lays within our interval then we can support otherwise we can reject.

There are two methods to finding  $t^*$ :

1. using the formula

$$t^* = \frac{\bar{x} - \mu}{\frac{S_x}{\sqrt{n}}} \quad (2)$$

2. using the calculator:

- 2nd + vars
- 4. or invT()
- area = 1 - given percent divided by 2
- DF as sample size - 1

**Definition 1.5** ( $\hat{p}$ ). Point estimate, the midpoint of the confidence interval and the value of a statistic that estimates the parameter of interest.

**Definition 1.6** (Level of Confidence). The expected proportion of intervals that will contain the parameter if many samples are taken.

*Example 1.1.* A 95% level of confidence means that we expect 95% of all confidence intervals to contain the true population parameter.

**Definition 1.7** (Margin of Error). Also denoted as ME, is the distance from the point estimate and the upper bound given as the second number in the range of the confidence interval.

**Definition 1.8** (Standard Error). Also denoted as SE, is the standard deviation of a sampling distribution or an estimate of that standard deviation.

## 2 Sampling Distributions

It is important that we briefly discuss sampling distributions for both means and proportions. A sample is a small subset of the population.

From the sample we can get the sample statistics:  $\bar{x}$ ,  $S_x$ ,  $\hat{p}$ .

Likewise, from the population, we can get population parameters:  $\mu$ ,  $\sigma$ ,  $p$ .

Samples drawn at random from a given population are likely to differ. This is known as sampling variability or sampling error. Sampling error can be reduced by increasing the size of the sample. The larger the sample the closer we get

to the population parameter. To get closer to the true population measure, we can examine a sampling distribution instead of just one sample. A sampling distribution takes all possible samples of a given size and combines their sample statistics to make an overall data set.

## 2.1 Sampling Distributions - Mean

It is important to quickly mention that the mean of the sampling distribution will equal the population parameter or mean. The standard deviation will also be different if the population is AT LEAST 10 times the sample size.

$$\begin{aligned} N &\geq 10n \\ \Rightarrow \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} \end{aligned} \tag{3}$$

## 3 Confidence Intervals

Confidence intervals differ slightly based on whether you are given the mean or the proportion of a distribution.

We will start with focusing on proportion.

### 3.1 Confidence Interval - Proportions

To start off, we must establish the conditions for a sampling distribution for proportions.

- Random -  $\hat{p}$  must be from simple random samples
- Normal (large enough) - number of successes and failures must both be greater than 10

$$N\hat{p} \geq 10 \wedge N(1 - \hat{p}) \geq 10 \tag{4}$$

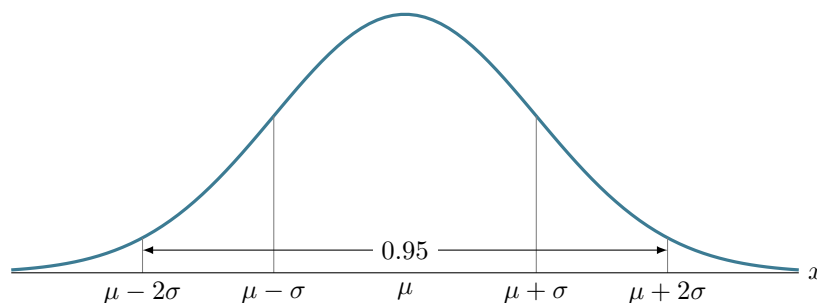
where N is the number of trials and  $\hat{p}$  is the proportion

- Independent (NOT large enough) - If we sample less than 10% of the population, we can treat observations as independent

$$\begin{aligned} n &\leq \frac{N}{10} \\ N &\geq 10n \end{aligned} \tag{5}$$

Next, we must understand the correlation between the critical value and the number of standard deviations from the parameter.

The above graph demonstrates that 95% confidence would correlate to approximately  $2\sigma$  away from the parameter  $\mu$ . We can also observe that the tail ends are of equal size and respectively would be 2.5% of the area in this example. To find the area of the tail end, you may use either the formula or the calculator method described in section 1.3.



### 3.2 Confidence Interval - One Proportion

If you are only given one sample then the 1-PropZInt() function in the calculator will be used.

The calculator solves the following equation:

$$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (6)$$

### 3.3 Confidence Interval - Two Proportions

Here instead we are trying to find the difference between two proportions. Both samples must be

- independent of each other,
- both must pass the 10% condition.

The center will be equal to the first proportion minus the second

$$\bar{x} = \hat{p}_1 - \hat{p}_2 \quad (7)$$

The standard error, defined in section 1.8, has the following equation for two proportions:

$$SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \quad (8)$$

These problems can be trivialized using the 2-SampZInt() function.

### 3.4 Confidence Intervals - Mean

Means are slightly different in that we can either know or not know the population standard deviation,  $\sigma$ . For both one sample and two samples if the population standard deviation is known then  $z^*$  can be found through either the formula or calculator method described in section 1.3. However, it is more

likely that  $\sigma$  is not known and the rest of the next two sections will be under this assumption. Here the confidence interval is defined as

$$CI = \bar{x} \pm t^* \frac{s}{\sqrt{n}} \quad (9)$$

For the mean, most conditions are the same as in proportions with the exception of the large enough condition.

- Normal (large enough) - number of samples must exceed 30

$$n > 30 \quad (10)$$

If this condition is not met, then the population must be approximately normal. Additionally, if we are given a data table, the condition can be checked with a box plot. If there are no outliers and it is mostly symmetrical, then we can assume the population is normally distributed.

### 3.5 Confidence Intervals - Two Means

The same assumptions can be made as the ones above except they have to apply for both samples. Here the equation simply subtracts both means before being added and subtracted to the margin of error:

$$CI = \bar{x}_1 - \bar{x}_2 \pm t^* \frac{s}{\sqrt{n}} \quad (11)$$

the standard error will also be as follows:

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (12)$$

finally, we will derive sample size from margin of error,  $z^*$ , and  $\sigma$ .

$$\begin{aligned} ME &= z^* \times SE \\ SE &= \frac{\sigma}{\sqrt{n}} \\ ME &= z^* \frac{\sigma}{\sqrt{n}} \\ ME^2 &= (z^*)^2 \frac{\sigma^2}{n} \\ n &= \lceil (z^*)^2 \frac{\sigma^2}{ME^2} \rceil \end{aligned} \quad (13)$$

We round up at the end as we cannot have a fraction of a sample while also not going below the calculated value.