

# OpenIntro Inference Guide

## Means and Proportions, 1-2 Samples

When creating a confidence interval (CI) or running a test for a mean, proportion, or difference in means or proportions, we first identify an appropriate *point estimate*, which we calculate using a sample. Next, we calculate the *standard error (SE)* of the point estimate, which is a measure of the point estimate's uncertainty. The general form for a confidence interval is

$$\text{point estimate} \pm (z^* \text{ or } t_{df}^*) * SE_{\text{estimate}}$$

The value  $z^*$  or  $t_{df}^*$  is found from the appropriate table (normal or t table) and is chosen based on the confidence level. The general form for a hypothesis test statistic is

$$\text{test statistic} = \frac{\text{point estimate} - \text{null value}}{SE_{\text{estimate}}}$$

where the *null value* is the value under question in  $H_0$ . For instance, if  $H_0 : \mu_1 - \mu_2 = 7.3$ , then the null value is 7.3. Alternatively, if  $H_0 : p = 0.4$ , then the null value is 0.4.

To identify the point estimate and standard error:

- Determine the number of samples: 1 or 2.
- If individual outcomes are categorical, use proportions. If numerical, use means.
- Determine whether you would like to create a confidence interval or run a hypothesis test.

Identify the proper CI or test in the table below. The mechanics for inference will be the same in each case. Additional instructions and special circumstances for using the table below:

- Plot the data, and check conditions/assumptions! Use OpenIntro Statistics as a reference.
- If the inference is for proportions or the standard deviation is known, use the normal distribution. If it is for means and the standard deviation is unknown, use  $t$ , i.e. the  $t$  distribution.
- For the 2-proportion test, use the pooled test when  $H_0$  is  $p_1 - p_2 = 0$  (or  $p_1 = p_2$ ). For the pooled test,  $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$ .
- When the data are numerical, there are two samples, and the data are also paired, compute the difference of each pair and analyze these differences. Note:  $n_{\text{diff}} = \#$  of differences, i.e. 2 paired samples each of size 10 implies there are  $n_{\text{diff}} = 10$  differences.
- Reminder:  $s$ ,  $s_1$ ,  $s_2$ , and  $s_{\text{diff}}$  are standard deviations of the samples.

Circumstance	parameter	estimate	SE <sub>estimate</sub>
1-prop CI	$p$	$\hat{p}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
1-prop test, $p_0 = \text{expected}$	$p$	$\hat{p}$	$\sqrt{\frac{p_0(1-p_0)}{n}}$
2-prop, unpooled, test or CI	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
2-prop, pooled, test only	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$
1-samp t-test or CI	$\mu$	$\bar{x}$	$s/\sqrt{n}$
2-samp, unpaired, t-test or CI	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
2-samp, paired, t-test or CI	$\mu_{\text{diff}} = \mu_1 - \mu_2$	$\bar{x}_{\text{diff}}$	$s_{\text{diff}}/\sqrt{n_{\text{diff}}}$