

## 9.3: Two Proportion Z-Test and Confidence Interval

This section will look at how to analyze a difference in the proportions for two independent samples. As with all other hypothesis tests and confidence intervals, the process of testing is the same, though the formulas and assumptions are different.

There are three types of hypothesis tests for comparing the difference in 2 population proportions  $p_1 - p_2$ , see Figure 9-7.




Two-tailed Test	Right-tailed Test	Left-tailed Test
$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$	$H_0: p_1 = p_2$ $H_1: p_1 > p_2$	$H_0: p_1 = p_2$ $H_1: p_1 < p_2$
		
$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 \neq 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 > 0$	$H_0: p_1 - p_2 = 0$ $H_1: p_1 - p_2 < 0$

Figure 9-7

Note that for our purposes,  $p_1 - p_2 = 0$ . We could also use a variant of this model to test for a magnitude difference for when  $p_1 - p_2 \neq 0$ , but we will not cover that scenario.

The z-test is a statistical test for comparing the proportions from two populations. It can be used when the samples are independent,  $n_1\hat{p}_1 \geq 10$ ,  $n_1\hat{q}_1 \geq 10$ ,  $n_2\hat{p}_2 \geq 10$ , and  $n_2\hat{q}_2 \geq 10$ .

The formula for the z-test statistic is:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p} \cdot \hat{q} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$\text{Where } \hat{p} = \frac{(x_1 + x_2)}{(n_1 + n_2)} = \frac{(\hat{p}_1 \cdot n_1 + \hat{p}_2 \cdot n_2)}{(n_1 + n_2)}, \quad \hat{q} = 1 - \hat{p}, \quad \hat{p}_1 = \frac{x_1}{n_1}, \hat{p}_2 = \frac{x_2}{n_2}.$$

The pooled proportion  $\hat{p}$  is a weighted mean of the proportions and  $\hat{q}$  is the complement of  $\hat{p}$ . Some texts or software may use different notation for the pooled proportion, note that  $\hat{p} = \bar{p}$ .

A vice principal wants to see if there is a difference between the number of students who are late to class for the first class of the day compared to the student's class right after lunch. To test their claim to see if there is a difference in the proportion of late students between first and after lunch classes, the vice-principal randomly selects 200 students from first class and records if they are late, then randomly selects 200 students in their class after lunch and records if they are late. At the 0.05 level of significance, can a difference be concluded?

**First Class After Lunch Class** Sample Size 200 200 Number of late students 13 16

### Solution

Assumptions: We are comparing the proportion of late students' first and after lunch classes. The number of "successes" and "failures" from each population must be greater than 10 ( $= 13 \geq 10$ ,  $= 187 \geq 10$ ,  $= 16 \geq 10$ , and  $= 184 \geq 10$ ). We must assume that the samples were independent.

Using the Traditional Method The claim is that there is a difference between the proportion of late students. Let population 1 be the first class, and population 2 be the class after lunch. Our claim would then be  $p_1 \neq p_2$ .

The correct hypotheses are:  $H_0: p_1 = p_2$

$$H_1: p_1 \neq p_2.$$

Compute the  $z_{\alpha/2}$  critical values. Draw and label the sampling distribution.

Use the inverse normal function  $\text{invNorm}(0.025, 0, 1)$  to get  $z_{\alpha/2} = \pm 1.96$ . See Figure 9-8.

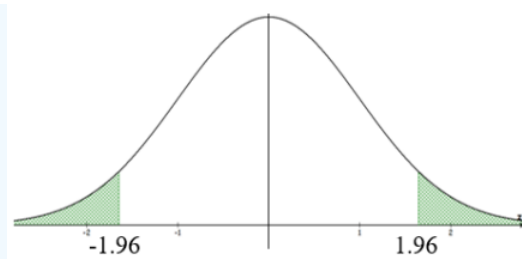


Figure 9-8

In order to compute the test statistic, we must first compute the following proportions:

$$\hat{p} = \frac{(x_1 + x_2)}{(n_1 + n_2)} = \frac{(13 + 16)}{(200 + 200)} = 0.0725 \quad \hat{q} = 1 - \hat{p} = 1 - 0.0725 = 0.9275$$

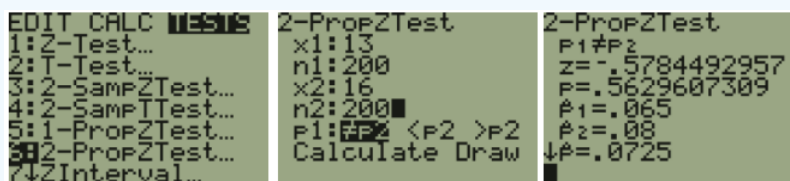
$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{13}{200} = 0.065 \quad \hat{p}_2 = \frac{x_2}{n_2} = \frac{16}{200} = 0.08$$

The test statistic is,  $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p} \cdot \hat{q} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(0.065 - 0.08)}{\sqrt{(0.0725 \cdot 0.9275) \left( \frac{1}{200} + \frac{1}{200} \right)}} = -0.5784$ .

Decision: Because the test statistic is between the critical values, we do not reject  $H_0$ .

Summary: There is not enough evidence to support any difference in the proportion of students that are late for their first class compared to the class after lunch.

**TI-84:** Press the [STAT] key, arrow over to the [TESTS] menu, arrow down to the option [6:2-PropZTest] and press the [ENTER] key. Type in the  $x_1$ ,  $n_1$ ,  $x_2$ , and  $n_2$  arrow over to the  $\neq$ ,  $<$ ,  $>$  sign that is the same in the problem's alternative hypothesis statement, then press the [ENTER] key, arrow down to [Calculate] and press the [ENTER] key. The calculator returns the z-test statistic and the p-value.



**TI-89:** Go to the [Apps] Stat/List Editor, then press [2<sup>nd</sup>] then F6 [Tests], then select **6: 2-PropZTest**. Type in the  $x_1$ ,  $n_1$ ,  $x_2$ , and  $n_2$  arrow over to the  $\neq$ ,  $<$ ,  $>$  and select the sign that is the same in the problem's alternative hypothesis statement. Press the [ENTER] key to calculate. The calculator returns the z-test statistic, sample proportions, pooled proportion, and the p-value.



## Two Proportions Z-Interval

A  $100(1 - \alpha)\%$  confidence interval for the difference between two population proportions  $p_1 - p_2$ :

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \sqrt{\left( \frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2} \right)} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \sqrt{\left( \frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2} \right)}$$

Or more compactly as  $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\left( \frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2} \right)}$

The requirements are identical to the 2-proportion hypothesis test. Note that the standard error does not rely on a hypothesized proportion so do not use a confidence interval to make decisions based on a hypothesis statement.

Find the 95% confidence interval for the difference in the proportion of late students in their first class and the proportion who are late to their class after lunch.

**First Class** **After Lunch Class** Sample Size 200 200 Number of late students 13 16

### Solution

First, compute the following:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{13}{200} = 0.065 \quad \hat{q}_1 = 1 - \hat{p}_1 = 1 - 0.065 = 0.935$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{16}{200} = 0.08 \quad \hat{q}_2 = 1 - \hat{p}_2 = 1 - 0.08 = 0.92$$

Find the  $z_{\alpha/2}$  critical value. Use the inverse normal to get  $z_{\alpha/2} = 1.96$ .

Now substitute the numbers into the interval estimate:  $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\left(\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}\right)}$

$$\Rightarrow (0.065 - 0.08) \pm 1.96 \sqrt{\left(\frac{0.065 \cdot 0.935}{200} + \frac{0.08 \cdot 0.92}{200}\right)}$$

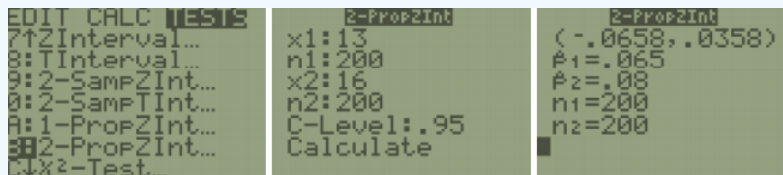
$$\Rightarrow -0.015 \pm 0.0508$$

$$\Rightarrow (-0.0508, 0.0358).$$

Use interval notation  $(-0.0508, 0.0358)$  or standard notation  $-0.0508 < p_1 - p_2 < 0.0358$ . Note that we can have negative numbers here since we are taking the difference of two proportions.

Since  $p_1 - p_2 = 0$  is in the interval, we are 95% confident that there is no difference in the proportion of late students between their first class or those who are late for their class after lunch.

**TI-84:** Press the [STAT] key, arrow over to the [TESTS] menu, arrow down to the option [2-PropZInt] and press the [ENTER] key. Type in the  $x_1$ ,  $n_1$ ,  $x_2$ ,  $n_2$ , the confidence level, then press the [ENTER] key, arrow down to [Calculate] and press the [ENTER] key. The calculator returns the confidence interval.



**TI-89:** Go to the [Apps] **Stat/List Editor**, then press [2<sup>nd</sup>] then F7 [Ints], then select **6: 2-PropZInt**. Type in the  $x_1$ ,  $n_1$ ,  $x_2$ ,  $n_2$ , the confidence level, then press the [ENTER] key to calculate. The calculator returns the confidence interval.