STAT 200

Elementary Statistics

9.3 - Comparing Two Independent Proportions

Printer-friendly version (https://onlinecourses.science.psu.edu/stat200/print/book/export/html/61)

1. Check any necessary assumptions and write null and alternative hypotheses.

For a right- or left-tailed test, a minimum of 10 successes and 10 failures in each group are necessary (i.e., $np \geq 10$ and $n(1-p) \geq 10$). Two-tailed tests are more robust and require only a minimum of 5 successes and 5 failures in each group.

The two groups that are being compared must be unpaired and unrelated (i.e., independent).

Below are the possible null and alternative hypothesis pairs:

Research Question	**Are the proportions of group 1 and group 2 different?**	**Is the proportion of group 1 greater than the proportion of group 2?**	**Is the proportion of group 1 less than the proportion of group 2?**
Null Hypothesis, H_0	$p_1 - p_2 = 0$	$p_1 - p_2 \leq 0$	$p_1 - p_2 \geq 0$
Alternative Hypothesis, H_a	$p_1 - p_2 \neq 0$	$p_1 - p_2 > 0$	$p_1 - p_2 < 0$
Type of Hypothesis Test	Two-tailed, non-directional	Right-tailed, directional	Left-tailed, directional

2. Calculate an appropriate test statistic.

The null hypothesis is that there is not a difference between the two proportions (i.e., $p_1 = p_2$). If the null hypothesis is true then the population proportions are equal. When computing the standard error for the difference between the two proportions a pooled proportion is used as opposed to the two proportions separately (i.e., unpooled). This pooled estimate will be symbolized by $\hat{p}$. This is similar to a weighted mean, but with two proportions.

Pooled Estimate of p

$$\hat{p} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$

The standard error for the difference between two proportions is symbolized by SE_0. The subscript 0 tells us that this standard error is computed under the null hypothesis ($H_0 : p_1 - p_2 = 0$).

Standard Error of $\hat{p}$

$$SE_0 = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Note that the default in many statistical programs, including Minitab Express, is to estimate the two proportions separately (i.e., unpooled). In order to obtain results using the pooled estimate of the proportion you will need to change the test method.

Also note that this standard error is different from the one that you will use later when constructing a confidence interval for $p_1 - p_2$. While the hypothesis testing procedure is based on the null hypothesis that $p_1 - p_2 = 0$, the confidence interval approach is not based on this premise. The hypothesis testing approach uses the pooled estimate of p while the confidence interval approach will use an unpooled method.

Test Statistic for Two Independent Proportions

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_0}$$

3. Determine a *p* value associated with the test statistic.

The *z* test statistic found in Step 2 is used to determine the *p* value.

4. Decide between the null and alternative hypotheses.

If $p \leq \alpha$ reject the null hypothesis. If $p > \alpha$ fail to reject the null hypothesis.

5. State a "real world" conclusion.

Based on your decision in Step 4, write a conclusion in terms of the original research question.

Example: Ice Cream Cones

Hand calculations are shown as well as Minitab Express procedures.

The Creamery wants to compare men and women in terms of preference for eating their ice cream out of a cone. They take a sample of 500 customers (240 men and 260 women) and ask if they prefer cones over bowls. They found that 124 men preferred cones and 90 women preferred cones.

Example: Same Sex Marriage

A survey was given to a sample of college students. They were asked whether they think same sex marriage should be legal. We'll compare the proportion of males and females who responded "yes." Of the 251 females in the sample, 185 said "yes." Of the 199 males in the sample, 107 said "yes."

1. Check any necessary assumptions and write null and alternative hypotheses.

For females, there were 185 who said "yes" and 66 who said "no." For males, there were 107 who said "yes" and 92 who said "no." There are at least 5 successes and failures in each group. This assumption has been met.

$$\hat{p}_f = \frac{185}{251}$$
$$\hat{p}_m = \frac{107}{199}$$

This is a two-tailed test because we are looking for a difference between males and females, we were not given a specific direction.

$$H_0 : p_f - p_m = 0$$
$$H_a : p_f - p_m \neq 0$$

2. Calculate an appropriate test statistic.

$$\hat{p} = \frac{185+107}{251+199} = \frac{292}{450} = .6489$$

$$SE_0 = \sqrt{\frac{292}{450}\left(1-\frac{292}{450}\right)\left(\frac{1}{251}+\frac{1}{199}\right)} = .0453$$

$$z = \frac{\frac{185}{251}-\frac{107}{199}}{.0453} = 4.400$$

Our test statistic is $z = 4.400$

3. Determine a *p* value associated with the test statistic.

$P(z > 4.400) = .0000054$, this is a two-tailed test, so this value must be multiplied by two: $.0000054 \times 2 = .0000108$

$p < .0001$

4. Decide between the null and alternative hypotheses.

$p \leq .05$, therefore we reject the null hypothesis.

5. State a "real world" conclusion.

There is evidence that there is a difference between the proportion of females and males who think that same sex marriage should be legal.

Above is one example of conducting a hypothesis test for two proportions including calculations by hand. Conducting these tests using Minitab Express will provide you will all of the information that you need to make a decision:

Test

Null hypothesis	$H_0: p_1 - p_2 = 0$
Alternative hypothesis	$H_1: p_1 - p_2 \neq 0$

Method	Z-Value	P-Value
Fisher's exact		<0.0001
Normal approximation	4.40	<0.0001

The pooled estimate of the proportion (0.648889) is used for the tests.

‹ 9.2 - Review of the Five Step Hypothesis Testing Procedure (/stat200/node/59)

up (/stat200/node/57)

9.4 - Comparing Two Independent Means › (/stat200/node/60)