Objectives

8.2 Comparing two proportions

- Examples
- Significance test for a difference in proportions
- Large-sample CI for a difference in proportions
- Relative risk
In this Chapter we compare proportions between two populations.

The applications traverse several areas from the social sciences to medicine.

The main assumptions that we will be making is that the samples are completely independent and as usual so as not to have a bias we assume that the samples are SRS (simple random sample from the populations of interest).

We start by looking at several motivating examples.
Example 1: Differences in opinions?

- Some suggest that people in the South have a **different** attitude towards renewable energy than those elsewhere in the country. A survey was conducted. 400 people were surveyed and asked should be government focus federal funding on renewable or non-renewable energy resources.

- The results of the survey are given below.

<table>
<thead>
<tr>
<th></th>
<th>South</th>
<th>North</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renewable</td>
<td>130</td>
<td>145</td>
</tr>
<tr>
<td>Non-renewable</td>
<td>70</td>
<td>55</td>
</tr>
</tbody>
</table>

\[ \hat{p}_S = 0.65 \quad \hat{p}_N = 0.725 \]

From the data there is a difference in the proportions:

\[ \hat{p}_S = 0.65 \text{ and } \hat{p}_N = 0.725 \]

But is this difference statistically significant (can we obtain such a difference when the proportions in the population are the same)?
As we are only looking for a difference we are testing $H_0: p_S - p_N = 0$ against $H_A: p_S - p_N \neq 0$. The results are given below.

**Hypothesis test results:**
- $p_1$: proportion of successes for population 1
- $p_2$: proportion of successes for population 2
- $p_1 - p_2$: difference in proportions
- $H_0: p_1 - p_2 = 0$
- $H_A: p_1 - p_2 \neq 0$

<table>
<thead>
<tr>
<th>Difference</th>
<th>Count1</th>
<th>Total1</th>
<th>Count2</th>
<th>Total2</th>
<th>Sample Diff.</th>
<th>Std. Err.</th>
<th>Z−Stat</th>
<th>P−value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1 - p_2$</td>
<td>145</td>
<td>200</td>
<td>130</td>
<td>200</td>
<td>0.075</td>
<td>0.04635124</td>
<td>1.6180797</td>
<td>0.1056</td>
</tr>
</tbody>
</table>

As the p-value is 10.56% (which is larger than the usual 5% significance level). Based on this there is a 10.56% chance of observing the difference by random chance, as this chance is quite large there isn’t anything in the data that suggests there is a difference in the opinions of the South and elsewhere.

**Remark** This is a two-sided test so the p-value is double the smallest area. The area to the RIGHT of 1.618 is 5.23%

In Statcrunch go to Stat -> Proportion Stat -> Two Sample -> With Summary
Example 2: Dogs or Cats?

- Popular belief is that males tend to be dog people and females tend to prefer cats. Is there any evidence that females really prefer cats?
- We want to test whether the proportion of females who like cats is greater than the proportion of males who like cats: $H_0: p_F - p_M \leq 0$ against $H_A: p_F - p_M > 0$.
- A simple random sample of 50 male and 50 female students were taken. The results of the survey are summarized below

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cats</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Dogs</td>
<td>25</td>
<td>20</td>
</tr>
</tbody>
</table>

$\hat{p}_M = 0.5$ $\hat{p}_F = 0.6$

From this data, our estimate of the proportion of males who like cats = 0.5 and the proportion of females who like cats is 0.6. There is a difference in preference in this sample, but is it statistically significant?
Hypothesis test results:

\[ p_1 \] : proportion of successes for population 1
\[ p_2 \] : proportion of successes for population 2
\[ p_1 - p_2 \] : difference in proportions
\[ H_0 \] : \( p_1 - p_2 = 0 \)
\[ H_A \] : \( p_1 - p_2 \neq 0 \)

<table>
<thead>
<tr>
<th>Difference</th>
<th>Count1</th>
<th>Total1</th>
<th>Count2</th>
<th>Total2</th>
<th>Sample Diff.</th>
<th>Std. Err.</th>
<th>Z–Stat</th>
<th>P–value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 - p_2 )</td>
<td>30</td>
<td>50</td>
<td>25</td>
<td>50</td>
<td>0.1</td>
<td>0.09949874</td>
<td>1.005</td>
<td>0.3149</td>
</tr>
</tbody>
</table>

The above gives the output for a two-sided test, but we can deduce the p-value for our one-sided test. The above output is the **double area to the right of 1.005**. Thus the area to the right of 1.005 = 31.49/2 = 15.74%.

We see that the z-transform for our test is:

\[
z = \frac{0.6 - 0.5}{0.099} = 1.005
\]

- The alternative is pointing to the RIGHT, thus the p-value is the area to the RIGHT of 1.005.
- Either looking up the tables or halving the above p-value to gives 15.74%.

There is a 15.5% chance of observing the differences we are seeing in the sample when there is no difference in male female preference. Since 15.74% is large (and larger than 5%), there is no evidence to in the data to suggest there is a difference.
Below we construct a 95% confidence interval for the difference in preference.

95% confidence interval results:
\( p_1 \) : proportion of successes for population 1
\( p_2 \) : proportion of successes for population 2
\( p_1 - p_2 \) : difference in proportions

<table>
<thead>
<tr>
<th>Difference</th>
<th>Count1</th>
<th>Total1</th>
<th>Count2</th>
<th>Total2</th>
<th>Sample Diff.</th>
<th>Std. Err.</th>
<th>L. Limit</th>
<th>U. Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 - p_2 )</td>
<td>30</td>
<td>50</td>
<td>25</td>
<td>50</td>
<td>0.1</td>
<td>0.09899495</td>
<td>−0.094026536</td>
<td>0.294026</td>
</tr>
</tbody>
</table>

We see from the output that the 95% confidence interval for the difference between true proportions of males and females who prefer cats lies somewhere between \([-0.09,0.29]\).
Example 3: Thai HIV vaccine trials

- In 2006 drug trials were done for a vaccine against HIV.
- 16,000 volunteers in Thailand were put on the trial, 8000 given the vaccine and 8000 a placebo: http://en.wikipedia.org/wiki/RV_144

<table>
<thead>
<tr>
<th></th>
<th>vaccine</th>
<th>placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contracted HIV</td>
<td>51</td>
<td>72</td>
</tr>
<tr>
<td>Did not contract</td>
<td>7949</td>
<td>7968</td>
</tr>
</tbody>
</table>

$\hat{p}_V = 0.0065$ and $\hat{p}_P = 0.009$

- Clearly the vaccine is not effective, in the sense that some of the people given the drug went on to develop HIV, but in this sample the proportion of people who were given the vaccine and went on to develop HIV is slightly less than those who were not given the vaccine:

$$\hat{p}_V = 0.0065 \text{ and } \hat{p}_P = 0.009$$

Is this difference statistically significant?
Hypothesis test results:

\[ p_1 : \text{proportion of successes for population 1} \]
\[ p_2 : \text{proportion of successes for population 2} \]
\[ p_1 - p_2 : \text{difference in proportions} \]
\[ H_0 : p_1 - p_2 = 0 \]
\[ H_A : p_1 - p_2 > 0 \]

| Difference | Count1 | Total1 | Count2 | Total2 | Sample Diff. | Std. Err. | Z–Stat     | P–value 
|------------|--------|--------|--------|--------|--------------|-----------|------------|--------
| \( p_1 - p_2 \) | 74     | 8000   | 51     | 8000   | 0.002875     | 0.0013920726 | 2.065266 | 0.0194 |

- Here we are testing \( H_0: p_P - p_V \leq 0 \) against \( H_A: p_P - p_V > 0 \).

- This matches with the test above and the p-value is 1.94%. This suggests that though the difference between the two groups is small it is still significant (but not strongly so), and it is plausible that the vaccine does have some protective effect.

- There has been a lot of criticism in how this trial was done, in particular that it should have targeted vulnerable groups such as sex workers, as their chance of contracting HIV tends to larger. This allows to really test the efficacy of the vaccine.
Example 4: SA HIV vaccine trials

- In 2005 drug trials were done for a vaccine against HIV.
- About 1800 men in South Africa were put on the trial.

<table>
<thead>
<tr>
<th></th>
<th>vaccine</th>
<th>placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contracted HIV</td>
<td>49</td>
<td>33</td>
</tr>
<tr>
<td>Did not contract HIV</td>
<td>865</td>
<td>889</td>
</tr>
<tr>
<td>Total</td>
<td>914</td>
<td>922</td>
</tr>
</tbody>
</table>

\[
\hat{p}_V = \frac{49}{914} = 0.053 \quad \hat{p}_P = \frac{33}{922} = 0.035
\]

- Look at the proportions,

\[
\hat{p}_V = 0.053 \text{ and } \hat{p}_P = 0.035
\]

We want to test: \(H_0: p_V - p_P \geq 0\) against \(H_A: p_V - p_P < 0\). But we see that for this sample the proportion of people who contracted HIV after taking the vaccine was GREATER than the proportion who contracted HIV using the placebo.
We test $H_0: p_V - p_P \geq 0$ against $H_A: p_V - p_P < 0$.

- It is immediately clear that there is no evidence in the data to support the null (the p-value is huge).

- In 2007 the trial was abandoned.

- It has been suggested that the vaccine may have had a negative effect, is there any evidence of this? What sided test would you do?
Example 5: Gastric Freezing

Gastric freezing was once a treatment for ulcers. Patients would swallow a deflated balloon with tubes, and a cold liquid would be pumped for an hour to cool the stomach and reduce acid production, thus relieving ulcer pain. The treatment was shown to be safe, significantly reducing ulcer pain, and was widely used for years. However, these studies were done without a control group.

A randomized comparative experiment later compared the outcome of gastric freezing with that of a placebo: 28 of the 82 patients subjected to gastric freezing improved, while 30 of the 78 in the control group improved.

<table>
<thead>
<tr>
<th></th>
<th>Gastric freezing</th>
<th>placebo</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>28</td>
<td>30</td>
<td>58</td>
</tr>
<tr>
<td>No improvement</td>
<td>54</td>
<td>48</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>78</td>
<td>160</td>
</tr>
</tbody>
</table>

\[ \hat{p}_G = \frac{28}{82} = 0.3415 \quad \hat{p}_P = \frac{30}{78} = 0.3846 \quad \hat{p} = \frac{58}{160} = 0.3625 \]
It is clear there is no evidence to suggest that Gastric freezing is any better than the placebo.

After this randomized trial Gastric freezing was abandoned. Indeed if the placebo and gastric freezing have the same effect then the proportion of people who would see an improvement regardless of whether they used a placebo or freezing is the combined ratio = 0.3625.

Conclusion: The gastric freezing was no better than a placebo. Consequently, the treatment was abandoned. Moral: ALWAYS USE A CONTROL GROUP!
A drug firm wants to test the efficacy of a hair cream (Minodoxil) on increasing hair. They did a randomized trial, one group was given the drug and the other was given a placebo. They are interested in testing:

- Observe the sample sizes do not have to be the same.
- There is a difference in sample proportions, can this difference be explained by just sampling difference or is it statistically significant?

<table>
<thead>
<tr>
<th></th>
<th>Minidoxil</th>
<th>placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>99</td>
<td>60</td>
</tr>
<tr>
<td>No improvement</td>
<td>211</td>
<td>242</td>
</tr>
<tr>
<td>Total</td>
<td>310</td>
<td>302</td>
</tr>
</tbody>
</table>

\[
\hat{p}_M = \frac{99}{310} = 0.32 \quad \hat{p}_P = \frac{60}{302} = 0.2
\]
Hypothesis test results:

- $p_1$: proportion of successes for population 1
- $p_2$: proportion of successes for population 2
- $p_1 - p_2$: difference in proportions
- $H_0: p_1 - p_2 = 0$
- $H_A: p_1 - p_2 > 0$

<table>
<thead>
<tr>
<th>Difference</th>
<th>Count1</th>
<th>Total1</th>
<th>Count2</th>
<th>Total2</th>
<th>Sample Diff.</th>
<th>Std. Err.</th>
<th>Z-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1 - p_2$</td>
<td>99</td>
<td>310</td>
<td>60</td>
<td>302</td>
<td>0.12067934</td>
<td>0.035455827</td>
<td>3.4036531</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Clearly the firm wants to see whether their remedy is better than the placebo so they are testing $H_0: p_M - p_P \leq 0$ against $H_A: p_M - p_P > 0$.

The above is a one-sided test that corresponds to the above test (just make sure you enter the data in correctly).

- The p-value is 0.03%, which is less than the standard significance levels, thus there is evidence to suggest that Minodixil works.
The theory: Comparing two independent samples

The main assumption made when doing the tests and making confidence intervals in the previous example is that the difference between the two sample proportion \( \hat{p}_1 - \hat{p}_2 \) is approximately normally distributed.

This means the sample sizes have to be relatively large. How large depends on how close \( p_1 \) and \( p_2 \) are from either zero or one, and the sample sizes.

The standard error is

\[
\sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}
\]
Test of significance for $H_0: p_1 = p_2$

If the null hypothesis is true, then we can rely on the properties of the sampling distribution to estimate the probability of drawing 2 samples with proportions $\hat{p}_1$ and $\hat{p}_2$ at random. The null hypothesis says $p_1 = p_2 = p$, with $p$ denoting the common value.

Our best estimate of $p$ is the **pooled sample proportion**:

$$\hat{p} = \frac{\text{total successes}}{\text{total observations}} = \frac{\text{count}_1 + \text{count}_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}.$$ 

The test statistic is then

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

This test is appropriate when the populations are at least 10 times as large as the samples and all counts are at least 5 (number of successes and number of failures in each sample).
Example 1: What does the null and pooling mean?

We test the hypothesis $H_0: p_M - p_P \leq 0$ against $H_A: p_M - p_P > 0$.

- The null hypothesis basically says, there is no difference between Minodixil and the placebo. Simply rubbing ones head simply increases your chance of hair growth. What is this chance? Under the null being true, we POOL the data together (this means combining the data) to estimate the proportion who benefit from a head massage:

<table>
<thead>
<tr>
<th></th>
<th>Minidoxil</th>
<th>placebo</th>
<th>Combined Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>99</td>
<td>60</td>
<td>159</td>
</tr>
<tr>
<td>No improvement</td>
<td>211</td>
<td>242</td>
<td>463</td>
</tr>
<tr>
<td>Total</td>
<td>310</td>
<td>302</td>
<td>612</td>
</tr>
</tbody>
</table>

$\hat{p}_M = \frac{99}{310} = 0.32$  $\hat{p}_P = \frac{60}{302} = 0.2$  $\hat{p} = \frac{159}{612} = 0.259$

If the null is true and there is no difference between placebo and Minodixil then our best estimate of those who gain hair growth from a head massage is $\hat{p} = 0.259$

$$s.e = \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n} + \frac{1}{m} \right)} = \sqrt{0.259 \times (1 - 0.259) \left( \frac{1}{310} + \frac{1}{302} \right)} = 0.035$$
Similarly the best estimator of the standard error (under the null there is no difference is) the pooled standard error:

\[
s.e = \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n} + \frac{1}{m} \right)} = \sqrt{0.259 \times (1 - 0.259) \left( \frac{1}{310} + \frac{1}{302} \right)} = 0.035
\]

We use this quantity when testing (since we always do the test under the null being true) BUT not when constructing confidence intervals.
What does “best” estimate mean?

- Let us return to the Minidoxil example:

<table>
<thead>
<tr>
<th></th>
<th>Minidoxil</th>
<th>placebo</th>
<th>Combined Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>99</td>
<td>60</td>
<td>159</td>
</tr>
<tr>
<td>No improvement</td>
<td>211</td>
<td>242</td>
<td>463</td>
</tr>
<tr>
<td>Total</td>
<td>310</td>
<td>302</td>
<td>612</td>
</tr>
</tbody>
</table>

|               | \( \hat{p}_M = \frac{99}{310} = 0.32 \) | \( \hat{p}_P = \frac{60}{302} = 0.2 \) | \( \hat{p} = \frac{159}{612} = 0.259 \) |

- Suppose there Minidoxil has the same effect as a placebo (basically a head massage; this is the null being true). How to estimate the proportion of people who gain more hair after several weeks of massaging ones head:
  - One estimate is 60/302 = 0.2.
  - Another estimate is 99/310 = 0.32 (since minidoxil has the same effect as a head massage).
  - The best estimate uses **the largest sample size** (since this minimizes the standard error). Since minidoxil and the placebo have the same effect we pool the data (combine the data) to increase the sample size. This leads to the estimate 159/612 = 0.259. This estimate is based on a sample size of 612 which is larger than the above two. Therefore it is the best estimate.
Example 2: Pooling data (dogs and cats)

- Let us return to the dog/cat and gender data set:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cats</td>
<td>25</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>Dogs</td>
<td>25</td>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

\[ \hat{p}_M = 0.5 \quad \hat{p}_F = 0.6 \quad \hat{p} = \frac{55}{100} = 0.55 \]

Suppose gender has no influence on one's preference for a cat or dog. What is the best estimate for the proportion of people who prefer cats to dogs?

**Answer:** If gender has no influence on preference, we should use the sample with the largest sample size. This means pooling/combining the male and female data. The total number of people who preferred cats to dogs is 55 out of 100. Therefore the best estimate for those who prefer cats to dogs is \( \frac{55}{100} = 0.55 \).
Large-sample CI for two proportions

For two independent SRSs of sizes $n_1$ and $n_2$ with sample proportion of successes $\hat{p}_1$ and $\hat{p}_2$ respectively, an approximate level $C$ confidence interval for $p_1 - p_2$ is

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm m, \quad m \text{ is the margin of error}$$

$$m = z^* \times SE_{\text{diff}} = z^* \times \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

$C$ is the area under the standard normal curve between $-z^*$ and $z^*$.

Use this method only when the populations are at least 10 times larger than the samples and the number of successes and the number of failures are each at least 10 in each sample.
Example 1: confidence intervals and Minidoxil

Now we want to see what the proportion gain would be when using Minidoxil – thus our aim is to construct a confidence interval for the difference in the proportions.

<table>
<thead>
<tr>
<th></th>
<th>Minidoxil</th>
<th>placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>99</td>
<td>60</td>
</tr>
<tr>
<td>No improvement</td>
<td>211</td>
<td>242</td>
</tr>
<tr>
<td>Total</td>
<td>310</td>
<td>302</td>
</tr>
</tbody>
</table>

\[
\hat{p}_V = \frac{99}{310} = 0.32 \quad \hat{p}_P = \frac{60}{302} = 0.2
\]

From the output we see that the 95% confidence interval is calculated as:

\[
[(0.32 - 0.2) - 1.96 \times 0.035, (0.32 - 0.2) + 1.96 \times 0.035] = [0.0514, 0.1886]
\]
The 95% confidence interval tells us that the gain in using Minodixil over the Placebo, that is with 95% confidence between 5% to 18.9% of the population would notice a difference over using a placebo.

Note that the standard error is slightly different than in the testing case, this is because we did not POOL the data. When constructing a confidence interval we do not make any assumptions about the relationship between the two proportions:

\[
s.e. = \sqrt{\frac{(99/310)(211/310)}{310} + \frac{(60/302)(142/302)}{302}} = 0.035
\]
Example 2: Cholesterol and heart attacks

How much does the cholesterol-lowering drug Gemfibrozil help reduce the risk of heart attack? We compare the incidence of heart attack over a 5-year period for two random samples of middle-aged men taking either the drug or a placebo.

So a 90% CI is $(0.0414-0.0273) \pm 1.645 \times 0.0057 = 0.0141 \pm 0.0094$, or $(0.0235, 0.0047)$.

We estimate with 90% confidence that the percentage of middle-aged men who suffer a heart attack is between 0.47% to 2.35% lower when taking the cholesterol-lowering drug over a placebo.
Brief introduction to relative risk

Another way to compare two proportions is to study the ratio of the two proportions, which is often called the relative risk (RR). A relative risk of 1 means that the two proportions are equal.

The procedure for calculating confidence intervals for relative risk is more complicated (use software) but still based on the same principles that we have studied.

Let us return to the previous example. We want to compare the effect that Gemfibrozil has on heart attacks compared to the Placebo. We see that 2.73% of the group on Gemfibrozil had a heart attack compared with 4.14% in the Placebo group.

<table>
<thead>
<tr>
<th></th>
<th>Heart attack</th>
<th>Sample size</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gemfibrozil</td>
<td>56</td>
<td>2051</td>
<td>2.73%</td>
</tr>
<tr>
<td>Placebo</td>
<td>84</td>
<td>2030</td>
<td>4.14%</td>
</tr>
</tbody>
</table>

The Relative Risk (RR) is $RR = \frac{4.14}{2.73} = 1.51$

Thus the placebo increases the chance of getting a heart attack by 50% compared with the drug.

Warning: Relative Risk is relative! In real terms the chance of a heart attack in either group is quite small. These quantities are estimates!
The age at which a woman gives birth to her first child may be a factor in the risk of later developing breast cancer. An international study selected women with at least one birth and recorded if they had breast cancer or not and whether they had their first child before their 30th birthday or after.

<table>
<thead>
<tr>
<th>Age at birth of 1st child</th>
<th>Number with cancer</th>
<th>Sample size</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>30+</td>
<td>683</td>
<td>3220</td>
<td>21.2%</td>
</tr>
<tr>
<td>&lt;30</td>
<td>1498</td>
<td>10,245</td>
<td>14.6%</td>
</tr>
</tbody>
</table>

$RR = \frac{.212}{.146} = 1.45.$

Women with a late first child have 1.45 times the risk of developing breast cancer. However, we need to take care, as these numbers are simply based on data. What are require is a confidence interval built around this estimate.
Summary of inference for $p_1 - p_2$

- The estimate for $p_1 - p_2$ is the difference in sample proportions
  \[ \hat{p}_1 - \hat{p}_2. \]

- The confidence interval for a population proportion $p_1 - p_2$ is
  \[ \hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}. \]
  $z^*$ is obtained from the normal distribution for confidence level $C$.

- A significance test of the hypothesis $H_0: p_1 = p_2$ uses test statistic
  \[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}, \text{ with } \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}. \]

The $P$-value is computed from the normal distribution according to the alternative hypothesis.

- Note the distinction between the SE for a confidence interval and the SE for a hypothesis test.

- The investigator chooses the confidence level $C$ and/or the significance level $\alpha$ prior to viewing the data.
Summary: Different standard errors for CI and tests

- From the output given in the previous examples, you may have noticed that the standard errors for the CI and the test are different (similar to the one-proportion case but different to the mean examples that we encountered previously).
- The `true' standard in the two sample case is

\[
\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}
\]

- But of course the proportions are unknown and need to be replaced by the estimates based on the samples.
- In the case of constructing confidence intervals we estimate the proportion for each sample and use:

\[
\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}
\]
However, in the case of hypothesis testing \( H_0: p_1 - p_2 \leq 0 \) against \( H_A: p_1 - p_2 > 0 \). We construct the standard error under the null. Under the null the proportions for both populations are assumed to be the same, in this case the standard error becomes

\[
\sqrt{p(1 - p) \left( \frac{1}{m} + \frac{1}{n} \right)}
\]

Of course the \( p \) proportion is unknown even under the null. But in this case we `pool' the two samples to obtain an estimator of \( p \):

\[
\hat{p} = \frac{\text{no. of successes in sample 1 plus number of successes in sample 2}}{\text{no. in sample 1 plus no. in sample two}}
\]
Accompanying problems associated with this Chapter

- Quiz 16
- Homework 8 (Questions 4(b))
- Homework 9 (Question 1)