Inference for Proportions

Inference for a Single Proportion

PBS Chapter 8.1
Objectives (PBS Chapter 8.1)

Inference for a single proportion

- Conditions for inference on $p$
- Large-sample confidence interval for $p$
- Plus four confidence interval for $p$
- Significance test for a proportion
- Choosing a sample size
Introduction

- Many studies collect data on categorical variables, such as race or occupation of a person, the make of a car, etc.

- The parameters of interest in these settings are population proportions.

- The statistic used to estimate a population proportion is the sample proportion.
Sampling distribution of a sample proportion

The sampling distribution of a sample proportion \( \hat{p} \) is approximately normal (normal approximation of a binomial distribution) when the sample size is large enough.
Conditions for inference on $p$

Assumptions:

1. The data used for the estimate are an SRS from the population studied.

2. The population is at least 10 times as large as the sample used for inference. This ensures that the standard deviation of $\hat{p}$ is close to $\sqrt{p(1-p)/n}$

3. The sample size $n$ is large enough that the sampling distribution can be approximated with a normal distribution. How large a sample size is required depends in part on the value of $p$ and the test conducted. Otherwise, rely on the binomial distribution.
Large-sample confidence interval for $p$

Confidence intervals contain the population proportion $p$ in $C\%$ of samples. For an SRS of size $n$ drawn from a large population and with sample proportion $\hat{p}$ calculated from the data, an \textbf{approximate level $C$ confidence interval} for $p$ is:

$$\hat{p} \pm m, \ m \text{ is the margin of error}$$

$$m = z^* \ SE = z^* \sqrt{\hat{p}(1 - \hat{p})/n}$$

Use this method when the number of successes and the number of failures are both at least 15.

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

Arthritis is a painful, chronic inflammation of the joints. An experiment on the side effects of pain relievers examined arthritis patients to find the proportion of patients who suffer side effects. The experiment found that 23 of a sample of 440 arthritis patients suffered some “adverse symptoms.”
Let’s calculate a 90% confidence interval for the population proportion of arthritis patients who suffer some “adverse symptoms.”

What is the sample proportion $\hat{p}$?

$$\hat{p} = \frac{23}{440} \approx 0.052$$

What is the sampling distribution for the proportion of arthritis patients with adverse symptoms for samples of 440?

For a 90% confidence level, $z^* = 1.645$.

Using the large sample method, we calculate a margin of error $m$:

$$m = z^* \sqrt{\hat{p}(1-\hat{p})/n}$$

$$m = 1.645 \times \sqrt{0.052(1-0.052)/440} \approx 0.023$$

$\Rightarrow$ With 90% confidence level, between 2.9% and 7.5% of arthritis patients taking this pain medication experience some adverse symptoms.
Because we have to use an estimate of $p$ to compute the margin of error, confidence intervals for a population proportion are not very accurate.

$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Specifically, we tend to be incorrect more often than the confidence level would indicate. But there is no systematic amount (because it depends on $p$).

Use with caution!
A simple adjustment produces more accurate confidence intervals. We act as if we had four additional observations, two being successes and two being failures. Thus, the new sample size is \( n + 4 \) and the count of successes is \( X + 2 \).

The “plus four” estimate of \( p \) is:

\[
\tilde{p} = \frac{\text{counts of successes} + 2}{\text{count of all observations} + 4}
\]

And an approximate level \( C \) confidence interval is:

\[
CI : \quad \tilde{p} \pm m \quad \text{with}
\]

\[
m = z \times SE = z \times \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{(n + 4)}}
\]

Use this method when \( C \) is at least 90% and sample size is at least 10.
We now use the “plus four” method to calculate the 90% confidence interval for the population proportion of arthritis patients who suffer some “adverse symptoms.”

What is the value of the “plus four” estimate of $p$?

\[
\hat{p} = \frac{23 + 2}{440 + 4} = \frac{25}{444} \approx 0.056
\]

An approximate 90% confidence interval for $p$ using the “plus four” method is:

\[
m = z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{(n + 4)}}
\]

\[
m = 1.645 \sqrt{0.056(1 - 0.056)} / 444
\]

\[
m = 1.645 \times 0.011 \approx 0.018
\]

90% CI for $p$: $\hat{p} \pm m$

or $0.056 \pm 0.018$

➤ With 90% confidence level, between 3.8% and 7.4% of arthritis patients taking this pain medication experience some adverse symptoms.
Significance test for $p$

The sampling distribution for $\hat{p}$ is approximately normal for large sample sizes and its shape depends on $p$ and $n$.

Thus, we can easily test the null hypothesis:

$H_0: p = p_0$ (a given value we are testing).

If $H_0$ is true, the sampling distribution is known →

The likelihood of our sample proportion given the null hypothesis depends on how far from $p_0$ our $\hat{p}$ is in units of standard deviation.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

This is valid when both expected counts—expected successes $np_0$ and expected failures $n(1 - p_0)$—are each 10 or larger.
P-values and one or two sided hypotheses—reminder

\[ H_a: p > p_0 \text{ is } P(Z \geq z) \]

\[ H_a: p < p_0 \text{ is } P(Z \leq z) \]

\[ H_a: p \neq p_0 \text{ is } 2P(Z \geq |z|) \]

If the P-value is as small or smaller than the chosen significance level \( \alpha \), then the difference is statistically significant and we reject \( H_0 \).
A national survey by the National Institute for Occupational Safety and Health on restaurant employees found that 75% said that work stress had a negative impact on their personal lives.

You investigate a restaurant chain to see if the proportion of all their employees negatively affected by work stress differs from the national proportion $p_0 = 0.75$.

$$H_0: p = p_0 = 0.75 \text{ vs. } H_a: p \neq 0.75 \text{ (2 sided alternative)}$$

In your SRS of 100 employees, you find that 68 answered “Yes” when asked, “Does work stress have a negative impact on your personal life?”

The expected counts are $100 \times 0.75 = 75$ and 25. Both are greater than 10, so we can use the z-test.

The test statistic is:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\"}$$

$$= \frac{0.68 - 0.75}{\sqrt{(0.75)(0.25) \frac{1}{100}}} = 1.62$$
From Table A we find the area to the left of $z = 1.62$ is 0.9474. Thus $P(Z \geq 1.62) = 1 - 0.9474$, or 0.0526. Since the alternative hypothesis is two-sided, the $P$-value is the area in both tails, and $P = 2 \times 0.0526 = 0.1052$.

The chain restaurant data are not significantly different from the national survey results ($\hat{p} = 0.68$, $z = 1.62$, $P = 0.11$).
Software gives you summary data (sample size and proportion) as well as the actual p-value.

### Minitab

**Test and Confidence Interval for One Proportion**

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95.0 % CI</th>
<th>Z-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>100</td>
<td>0.680000</td>
<td>(0.588572, 0.771428)</td>
<td>-1.62</td>
<td>0.106</td>
</tr>
</tbody>
</table>

### Crunch It!

**Hypothesis test results:**

- *p* = proportion of successes for population
- Parameter: *p*
- H0: Parameter = 0.75
- HA: Parameter not = 0.75

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Count</th>
<th>Total</th>
<th>Sample Prop.</th>
<th>Std. Err.</th>
<th>Z-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>68</td>
<td>100</td>
<td>0.68</td>
<td>0.04330127</td>
<td>-1.6165807</td>
<td>0.106</td>
</tr>
</tbody>
</table>
Interpretation: magnitude vs. reliability of effects

The **reliability** of an interpretation is related to the strength of the evidence. The smaller the **p-value**, the stronger the evidence against the null hypothesis and the more confident you can be about your interpretation.

The **magnitude** or **size** of an effect relates to the real-life relevance of the phenomenon uncovered. The P-value does NOT assess the relevance of the effect, nor its magnitude.

A **confidence interval** will assess the magnitude of the effect. However, magnitude is not necessarily equivalent to how theoretically or practically relevant an effect is.
Sample size for a desired margin of error

You may need to choose a sample size large enough to achieve a specified margin of error. However, because the sampling distribution of $\hat{p}$ is a function of the population proportion $p$, this process requires that you guess a likely value for $p$: $p^*$. 

$$p \sim N \left( p, \sqrt{p(1-p)/n} \right) \Rightarrow n = \left( \frac{z^*}{m} \right)^2 p^*(1-p^*)$$

The margin of error will be less than or equal to $m$ if $p^*$ is chosen to be 0.5.

Remember, though, that sample size is not always stretchable at will. There are typically costs and constraints associated with large samples.
What sample size would we need in order to achieve a margin of error no more than 0.01 (1%) for a 90% confidence interval for the population proportion of arthritis patients who suffer some “adverse symptoms.”

We could use 0.5 for our guessed \( p^* \). However, since the drug has been approved for sale over the counter, we can safely assume that no more than 10% of patients should suffer “adverse symptoms” (a better guess than 50%).

For a 90% confidence level, \( z^* = 1.645 \).

\[
n = \left( \frac{z^*}{m} \right)^2 p^* (1 - p^*) = \left( \frac{1.645}{0.01} \right)^2 (0.1)(0.9) \approx 2434.4
\]

\( \Rightarrow \) To obtain a margin of error no more than 1%, we would need a sample size \( n \) of at least 2435 arthritis patients.
Inference for Proportions
Comparing Two Proportions

PBS Chapter 8.2
Objectives (PBS Chapter 8.2)

Comparing two proportions

- Comparing two independent samples
- Large-sample confidence interval for two proportions
- Plus four confidence interval for two proportions
- Significance tests
- Relative risk
Comparing two independent samples

We often need to compare two treatments used on independent samples. We can compute the difference between the two sample proportions and compare it to the corresponding, approximately normal sampling distribution for \((\hat{p}_1 - \hat{p}_2)\):

\[
\text{Standard deviation} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}
\]
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

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

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

$C$ 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 both samples.
“No Sweat” Garment Labels

“No Sweat” labels on apparel indicate proper working conditions. Is there a gender difference in the proportion of label users? We want to calculate a 95% confidence interval for the difference in the proportions of label users.

Standard error of the difference $p_1 - p_2$:

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

$$SE = \sqrt{\frac{0.213(0.787)}{63} + \frac{0.108(0.892)}{27}} = 0.0308$$

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>n</th>
<th>$\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>63</td>
<td>296</td>
<td>0.213</td>
</tr>
<tr>
<td>Men</td>
<td>27</td>
<td>251</td>
<td>0.108</td>
</tr>
</tbody>
</table>

The confidence interval is $(\hat{p}_1 - \hat{p}_2) \pm z \times SE$

So the 95% CI is $0.105 \pm 1.96 \times 0.0308 = 0.105 \pm 0.060$

We are 95% confident that the difference in the proportions is between 0.04 and 0.16.
**Plus four CI for two proportions**

The plus four method again produces more accurate confidence intervals. We act as if we had four additional observations: one success and one failure in each of the two samples. The new combined sample size is $n_1 + n_2 + 4$ and the proportions of successes are:

$$\tilde{p}_1 = \frac{X_1 + 1}{n_1 + 2} \quad \text{and} \quad \tilde{p}_2 = \frac{X_2 + 1}{n_2 + 2}$$

An approximate level $C$ confidence interval is:

$$CI : \ (\tilde{p}_1 - \tilde{p}_2) \pm z^* \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2 + 2}}$$

Use this when $C$ is at least 90% and both sample sizes are at least 5.
“No Sweat” Garment Labels – plus four CI

“No Sweat” labels on apparel indicate proper working conditions. Is there a gender difference in the proportion of label users? We want to calculate a 95% confidence interval for the difference in the proportions of label users.

For the plus four procedure, we would use $X_1 = 64$, $n_1 = 298$, $X_2 = 28$, and $n_2 = 253$.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>n</th>
<th>$\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women (1)</td>
<td>63</td>
<td>296</td>
<td>0.213</td>
</tr>
<tr>
<td>Men (2)</td>
<td>27</td>
<td>251</td>
<td>0.108</td>
</tr>
</tbody>
</table>
Test of significance

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.

$H_0 : p_1 = p_2 = p$

Our best estimate of $p$ is $\hat{p}$, the pooled sample proportion

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

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

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).
"No Sweat" Garment Labels

"No Sweat" labels on apparel indicate proper working conditions. Is there a gender difference in the proportion of label users? Are men and women equally likely to be label users?

Here is the data summary:

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>n</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Women (1)</td>
<td>63</td>
<td>296</td>
<td>0.213</td>
</tr>
<tr>
<td>Men (2)</td>
<td>27</td>
<td>251</td>
<td>0.108</td>
</tr>
</tbody>
</table>

\begin{align*}
H_0 & : p_1 = p_2 \\
H_a & : p_1 \neq p_2 \\
\hat{p}_{\text{pooled}} & = \frac{63 + 27}{296 + 251} = 0.1645 \\
\end{align*}

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.213 - 0.108}{\sqrt{0.1645*0.8355\left(\frac{1}{296} + \frac{1}{251}\right)}} = \frac{0.105}{0.03181} = 3.30
\]

The P-value is \(2P(Z > 3.3) = 2(0.0005) = 0.001\). The difference is statistically significant.
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.

The age at which a woman gets her first child may be an important 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 30\(^{th}\) birthday or after.

<table>
<thead>
<tr>
<th></th>
<th>Birth age 30+</th>
<th>Sample size</th>
<th>(\hat{p})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>683</td>
<td>3220</td>
<td>21.2%</td>
</tr>
<tr>
<td>No</td>
<td>1498</td>
<td>10,245</td>
<td>14.6%</td>
</tr>
</tbody>
</table>

\[ RR = \frac{.212}{.146} \approx 1.45 \]

Women with a late first child have 1.45 times the risk of developing breast cancer.