Significance level vs. confidence level

Sta102 / BME102

Colin Rundel

March 4, 2015

Significance level vs. confidence level

Agreement of CI and HT

- Confidence intervals and hypothesis tests (almost) always agree, as long as the two methods use equivalent levels of significance / confidence and the SEs are the same.
  - A two sided hypothesis with threshold of $\alpha$ is equivalent to a confidence interval with $CL = 1 - \alpha$.
  - A one sided hypothesis with threshold of $\alpha$ is equivalent to a confidence interval with $CL = 1 - (2 \times \alpha)$.

- If $H_0$ is rejected, a confidence interval that agrees with the result of the hypothesis test should not include the null value.

- If $H_0$ is failed to be rejected, a confidence interval that agrees with the result of the hypothesis test should include the null value.

Example - Experimental Design

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

(a) All 1000 get the drug

(b) 500 get the drug, 500 don’t
Results from the GSS

The GSS asks the same question, below is the distribution of responses from the 2010 survey:

<table>
<thead>
<tr>
<th>All 1000 get the drug</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 get the drug</td>
<td>571</td>
</tr>
<tr>
<td>Total</td>
<td>670</td>
</tr>
</tbody>
</table>

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”.

What are the parameter of interest and the point estimate?

- **Parameter of interest**: Proportion of *all* Americans who have good intuition about experimental design.

\[ p \text{ (a population proportion)} \]

- **Point estimate**: Proportion of *sampled* Americans who have good intuition about experimental design.

\[ \hat{p} \text{ (a sample proportion)} \]

Inference on a proportion

What percent of all Americans have a good intuition about experimental design, i.e. would answer “500 get the drug 500 don’t”?

- We can answer this research question using a confidence interval, which we know is always of the form

\[ \text{point estimate} \pm \text{ME} \]

- And we also know that \( \text{ME} = \text{critical value} \times \text{standard error} \) of the point estimate.

\[ SE_{\hat{p}} = ? \quad CV = ? \]

Proportions and the CLT

What kind of probability model can we use for \( \hat{p} \)?

It may be useful to instead think about \( n\hat{p} \), what distribution will that have?

\[ n\hat{p} \sim \text{Binom}(n, p) \]

\[ n\hat{p} \approx X \sim N(\mu = np, \sigma^2 = np(1 - p)) \]

We can then find the distribution of \( \hat{p} \) by dividing by \( n \),

\[ \hat{p} \approx X/n \sim N(\mu = p, \sigma^2 = p(1 - p)/n) \]
Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population proportion, \( p \), and standard error equal to \( \sqrt{\frac{p(1-p)}{n}} \).

\[ \hat{p} \sim N \left( \mu = p, \sigma^2 = \frac{SE^2}{n} = \frac{p(1-p)}{n} \right) \]

But of course this is true only under certain conditions ... any guesses?

Assumptions/conditions:
1. **Independence:** The sample is random, and \( n < 10\% \) of all Americans, therefore we can assume that one respondent’s response is independent of another.
2. **Normality:** At least 10 successes (\( np \geq 10 \)) and 10 failures (\( n(1-p) \geq 10 \)).

Calculating the Confidence Interval

We are given that \( n = 670, \hat{p} = 0.85 \), we also just learned that the standard error of the sample proportion is \( SE = \sqrt{\frac{p(1-p)}{n}} \). What is the 95% confidence interval for this proportion?

\[ CI = \text{point estimate} \pm \text{margin of error} \]
\[ = \hat{p} \pm z^* \times SE \]
\[ = 0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} \]
\[ = (0.82, 0.88) \]

Choosing a sample size

How many people should you sample in order to reduce the margin of error of a 95% confidence interval down to 1%.

\[ ME = z^* \times SE \]

\[ 0.01 \geq 1.96 \times \sqrt{\frac{p \times (1-p)}{n}} \]

\[ 0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \]

\[ 0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n} \]

\[ n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2} \]

\[ n \geq 4898.04 \rightarrow n \text{ should be at least } 4899 \]
Choosing a sample size when estimating a proportion

What if there isn’t a previous study?

... use \( \hat{p} = 0.5 \). Why?
- if you don’t know any better, 50-50 is a good guess
- \( \hat{p} = 0.5 \) gives the most conservative estimate – largest standard error
  and thus the largest possible sample size.

Hypothesis testing for a proportion

CI vs. HT for proportions

For a test of one proportion our null and alternative hypotheses will be about \( p \), therefore when we assume \( H_0 \) is true we fix \( p = p_0 \). Hence,
- Success-failure condition:
  - CI: At least 10 observed successes and failures, calculated using the
    sample proportion, \( \hat{p} \)
  - HT: At least 10 expected successes and failures, calculated using the
    null value, \( p_0 \)
- Standard error:
  - CI: calculate using observed sample proportion:
    \[
    SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx \sqrt{\hat{p}(1-\hat{p})}
    \]
  - HT: calculate using the null value:
    \[
    SE = \sqrt{\frac{p_0(1-p_0)}{n}}
    \]
Example - Melting ice cap survey

Scientists predict that global warming may have big effects on the polar regions within the next 100 years. One of the possible effects is that the northern ice cap may completely melt. Would this bother you a great deal, some, a little, or not at all if it actually happened?

(a) A great deal
(b) Some
(c) A little
(d) Not at all

Results from the GSS & Duke

The GSS asks this question, below is the distribution of responses from the 2010 survey:

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>Duke</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A great deal</td>
<td>454</td>
<td>56</td>
<td>510</td>
</tr>
<tr>
<td>Not a great deal</td>
<td>226</td>
<td>32</td>
<td>258</td>
</tr>
<tr>
<td>Total</td>
<td>680</td>
<td>88</td>
<td>768</td>
</tr>
</tbody>
</table>

The same question was asked of 88 Duke students, of which 56 said it would bother them a great deal.

We will collapse the data such that we consider only the responses of a great deal or not a great deal.

Collapsed Results

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>Duke</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A great deal</td>
<td>454</td>
<td>56</td>
<td>510</td>
</tr>
<tr>
<td>Not a great deal</td>
<td>226</td>
<td>32</td>
<td>258</td>
</tr>
<tr>
<td>Total</td>
<td>680</td>
<td>88</td>
<td>768</td>
</tr>
</tbody>
</table>

This is an example of a 2 x 2 contingency table.

We are interested in comparing proportion of Duke students who say it would both them a great deal \((P(GD|Duke) = 56/88)\) to the proportion of all Americans who say it would both them a great deal \((P(GD|US) = 454/680)\).

Parameter and point estimate

- **Parameter of interest:** Difference between the proportions of all Duke students and all Americans who would be bothered a great deal by the northern ice cap completely melting.
  \[ P(GD|Duke) - P(GD|US) \]

- **Point estimate:** Difference between the proportions of sampled Duke students and sampled Americans who would be bothered a great deal by the northern ice cap completely melting.
  \[ \hat{P}(GD|Duke) - \hat{P}(GD|US) \]
Inference for comparing proportions

- The details are the same as before...
- CI: point estimate ± margin of error
- HT: Use \( Z = \frac{\text{point estimate} - \text{null value}}{\text{SE}} \) to find appropriate p-value.
- We just need the appropriate standard error of the point estimate (\( SE_{p_{GD|Duke} - p_{GD|US}} \)),

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

CI for difference of proportions

Construct a 95% confidence interval for the difference between the proportions of Duke students and Americans who would be bothered a great deal by the melting of the northern ice cap (\( p_{GD|Duke} - p_{GD|US} \)).

<table>
<thead>
<tr>
<th></th>
<th>Duke</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>A great deal</td>
<td>56</td>
<td>454</td>
</tr>
<tr>
<td>Not a great deal</td>
<td>32</td>
<td>226</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>680</td>
</tr>
</tbody>
</table>

Hypotheses for testing the difference of two proportions

Which of the following is the correct set of hypotheses for testing if the proportion of all Duke students who would be bothered a great deal by the melting of the northern ice cap differs from the proportion of all Americans who do?

- \( H_0 : p_{GD|Duke} = p_{GD|US} \)
- \( H_A : p_{GD|Duke} \neq p_{GD|US} \)
- \( H_0 : p_{GD|Duke} - p_{GD|US} = 0 \)
- \( H_A : p_{GD|Duke} - p_{GD|US} \neq 0 \)
Flashback to working with one proportion

- When constructing a confidence interval for a population proportion, we check if the observed number of successes and failures are at least 10.
  \[ n\hat{p} \geq 10 \quad n(1 - \hat{p}) \geq 10 \]
- When conducting a hypothesis test for a population proportion, we check if the expected number of successes and failures are at least 10.
  \[ np_0 \geq 10 \quad n(1 - p_0) \geq 10 \]

A slight wrinkle ...

- In the case of comparing two proportions where \( H_0 : p_1 = p_2 \), there isn’t a null value we can use to calculated the expected number of successes and failures in each sample or the SE.
- Therefore, we need to first find a common (pooled) proportion for the two groups, and use that in our analysis.
- This involves finding the proportion of total successes among all observations.

\[
\hat{p}_{\text{pooled}} = \frac{\# \text{ of successes in 1} + \# \text{ of successes in 2}}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}
\]

Pooled estimate of a proportion

Calculate the estimated pooled proportion of Duke students and Americans who would be bothered a great deal by the melting of the northern ice cap.

<table>
<thead>
<tr>
<th></th>
<th>Duke</th>
<th>US</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A great deal</td>
<td>56</td>
<td>454</td>
<td>510</td>
</tr>
<tr>
<td>Not a great deal</td>
<td>32</td>
<td>226</td>
<td>258</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>680</td>
<td>788</td>
</tr>
</tbody>
</table>

\[
\hat{p}_{\text{pooled}} = \frac{56 + 454}{88 + 680} = \frac{510}{788} = 0.664
\]

Which sample proportion (\( \hat{p}_{\text{GD|Duke}} \) or \( \hat{p}_{\text{GD|US}} \)) is closer to the pooled estimate? Why?

Implications for the SE

Under the null hypothesis we are stating that \( p_1 = p_2 \) which does not uniquely identify either \( p_1 \) or \( p_2 \). Therefore we are using the pooled proportion (\( \hat{p} \)) as our best guess for \( p_1 \) and \( p_2 \) under the null hypothesis.

For a confidence interval we have seen that

\[
SE = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}} \approx \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

Therefore, for a hypothesis test we will use \( \hat{p} \) as our approximation for \( p_1 \) and \( p_2 \)

\[
SE = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}} \approx \sqrt{\hat{p}(1 - \hat{p}) + \hat{p}(1 - \hat{p})} \\
\approx \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \text{ where } \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}
\]
Do these data suggest that the proportion of all Duke students who would be bothered a great deal by the melting of the northern ice cap differs from the proportion of all Americans who do?

\[ \hat{p}_{\text{pooled}} = 0.664, \quad n_1 = 88, \quad n_2 = 680 \]