Lecture 13 - Tests of Proportions

Sta102 / BME102

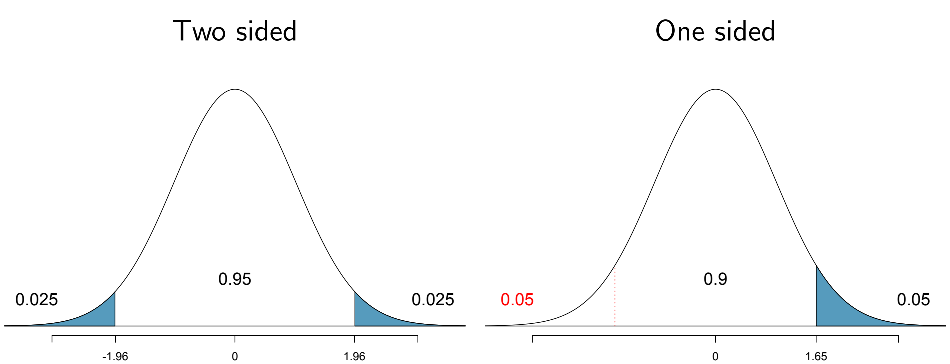
Colin Rundel

October 15, 2014

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 α is equivalent to a confidence interval with $CL = 1 - \alpha$.
  - A one sided hypothesis with threshold of α 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.

Significance level vs. confidence level



Example - Waiting Times

A 95% confidence interval for the average waiting time at an emergency room is (128 minutes, 147 minutes).

Determine if the following statements are *true* or *false*,

(a) a hypothesis test of $H_A : \mu \neq 120$ min at $\alpha = 0.05$ is equivalent to this CI.

(b) a hypothesis test of $H_A : \mu > 120$ min at $\alpha = 0.025$ is equivalent to this CI.

(c) This interval does not support the claim that the average wait time is 120 minutes.

(d) The claim that the average wait time is 120 minutes would not be rejected using a 90% confidence interval.

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:

All 1000 get the drug	99
500 get the drug 500 don't	571
Total	670

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

  $$\textit{point estimate} \pm \textit{ME}$$

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

  $$SE_{\hat{p}} = ? \qquad CV = ?$$

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?

$$n\hat{p} \sim Binom(n,\, p)$$

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

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 $\sqrt{\frac{p\,(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\,(1-p)}{n}}\right)$$

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

Assumptions/conditions:

1. *Independence*:
   - *Random sample*
   - *10% condition*: If sampling without replacement, $n <$ 10% of the population.
2. *Normality*: At least 10 successes ($np \geq 10$) and 10 failures ($n(1-p) \geq 10$).

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have a good intuition about experimental design?

Given: $n = 670$, $\hat{p} = \frac{571}{670} = 0.85$.

Are CLT conditions met?

1. *Independence*: The sample is random, and $670 <$ 10% of all Americans, therefore we can assume that one respondent's response is independent of another.
2. *Success-failure*: 571 people answered correctly (successes) and 99 answered incorrectly (failures), both are greater than 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?

$$\begin{aligned} CI &= \text{point estimate} \pm \text{margin of error} \\ &= \text{point estimate} \pm \text{critical value} \times SE \\ &= \hat{p} \pm z^* \times SE \\ &= 0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} = (0.82, 0.88) \end{aligned}$$

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%.

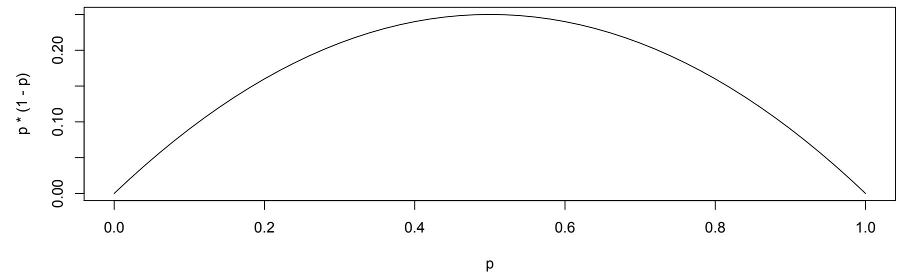
$$ME = z^\star \times SE$$

$$\begin{aligned}
0.01 &\geq 1.96 \times \sqrt{\frac{p \times (1-p)}{n}} \\
0.01 &\geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \textit{Using } \hat{p} \textit{ from previous study} \\
0.01^2 &\geq 1.96^2 \times \frac{0.85 \times 0.15}{n} \\
n &\geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2} \\
n &\geq 4898.04 \rightarrow n \textit{ should be at least 4,899}
\end{aligned}$$

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.



CI vs. HT for proportions

- Success-failure condition:
  - CI: At least 10 *observed* successes and failures
  - HT: At least 10 *expected* successes and failures, calculated using the null value, p_0
- Standard error:
  - CI: calculate using observed sample proportion: $SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
  - HT: calculate using the null value: $SE = \sqrt{\frac{p_0(1-p_0)}{n}}$

With means SE only depended on s or σ and n, so it was not determined in any way by H_0.

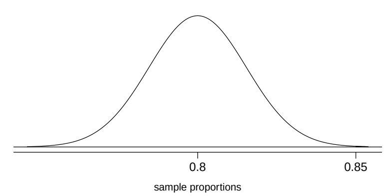
With proportions the mean and SE both depend on p, therefore H_0 affects the SE.

Back to the GSS

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

$$\begin{aligned}
SE &= \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154 \\
Z &= \frac{0.85 - 0.80}{0.0154} = 3.25 \\
p-value &= 1 - 0.9994 = 0.0006
\end{aligned}$$



Since p-value is small we reject H_0. The data provide convincing evidence that more than 80% of Americans have a good intuition on experimental design.

Common Misinterpretations

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey a is $\pm 3\%$. A news piece on this study's findings states: "More than 10% of all Americans have objections on religious grounds to celebrating Halloween."

Is this statement justified?

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:

A great deal	454
Some	124
A little	52
Not at all	50
Total	680

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

	US	Duke	Total
A great deal	454	56	510
Not a great deal	226	32	258
Total	680	88	768

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

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

What does it mean if

$$P(GD|Duke) = P(GD|US)?$$

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.

$$p_{Duke} - p_{US}$$

- *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.

$$\hat{p}_{Duke} - \hat{p}_{US}$$

Inference for comparing proportions

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

$$SE_{(p_1 - p_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Conditions for inference on the difference of proportions

1. *Independence within groups:*
   - The US group is sampled randomly and we're assuming that the Duke group represents a random sample as well.
   - $n_{Duke} < 10\%$ of all Duke students and $680 < 10\%$ of all Americans.

   We can assume that the attitudes of Duke students in the sample are independent of each other, and attitudes of US residents in the sample are independent of each other as well.
2. *Independence between groups:* The sampled Duke students and the US residents are independent of each other.
3. *Success-failure:*
   At least 10 observed successes and 10 observed failures in both groups.

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 ($p_{Duke} - p_{US}$).

	Duke	US
A great deal	56	454
Not a great deal	32	226
Total	88	680

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?

$$H_0 : p_{Duke} = p_{US} \qquad H_0 : p_{Duke} - p_{US} = 0$$
$$H_A : p_{Duke} \neq p_{US} \qquad H_A : p_{Duke} - p_{US} \neq 0$$

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.

$$n\hat{p} \geq 10 \qquad n(1 - \hat{p}) \geq 10$$

- When conducting a hypothesis test for a population proportion, we check if the *expected* number of successes and failures are at least 10.

$$np_0 \geq 10 \qquad n(1 - p_0) \geq 10$$

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.

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

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.

	Duke	US
A great deal	56	454
Not a great deal	32	226
Total	88	680

$$\hat{p}_{pooled} = \frac{56 + 454}{88 + 680} = 0.664$$

Which sample proportion ($\hat{p}_{Duke}$ or $\hat{p}_{US}$) 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

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$
$$\approx \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \text{ where } \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

HT for comparing proportions

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?

$$\hat{p}_{pooled} = 0.664, \quad n_1 = 88, \quad n_2 = 680$$

Recap - inference for one proportion

- Population parameter: p, point estimate: $\hat{p}$
- Conditions:
  - independence
    - random sample and 10% condition
  - at least 10 successes and failures
    - *observed* for CI
    - *expected* for HT
- Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$
  - for CI: use $\hat{p}$
  - for HT: use p_0

Recap - comparing two proportions

- Population parameter: $(p_1 - p_2)$, point estimate: $(\hat{p}_1 - \hat{p}_2)$
- Conditions:
  - independence within groups
    - random sample and 10% condition met for both groups
  - independence between groups
  - at least 10 successes and failures in each group
    - *observed* for CI
    - *expected* for HT
- $SE_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
  - for CI: use $\hat{p}_1$ and $\hat{p}_2$
  - for HT:
    - when $H_0 : p_1 = p_2$: use $\hat{p}_{pool} = \frac{\#suc_1 + \#suc_2}{n_1 + n_2}$
    - when $H_0 : p_1 - p_2 =$ *(some value other than 0)*: use $\hat{p}_1$ and $\hat{p}_2$
      - this is pretty rare

Reference - standard error calculations

	one sample	two samples
mean	$SE = \frac{\sigma}{\sqrt{n}}$	$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
proportion	$SE = \sqrt{\frac{p(1-p)}{n}}$	$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

- When working with means, it's very rare that σ is known, so we usually use s as an approximation.
- When working with proportions, we will not know p therefore
  - if doing a hypothesis test, p comes from the null hypothesis
  - if constructing a confidence interval, use $\hat{p}$ instead