Unit 3: Inference for Categorical and Numerical Data

2. Inference for the difference of two proportions

10/31/2016

Quiz 3 - Confidence Intervals

Recap from last time

1. We can use the CLT to make inferences about proportions
2. Confidence intervals can be used to make inferences about a population proportion
3. Confidence intervals can be used to do Hypothesis Tests

Key ideas

1. You can use the Normal approximation for the difference of two proportions
2. The margin of error is not just the sum of the margin of errors for each proportion
3. If you think two proportions come from the same population, you can use a pooled estimate

Melting ice caps

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.

How much would it bother you if this actually happened?

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

Results from the NSF SEEI2012

National Science Board
SCIENCE & ENGINEERING INDICATORS 2016

The National Science Foundation asked this question as part of a survey on general scientific literacy in 2010. Here are the results:

	SEEI2012	Previous Class	PSYC 20100
A great deal	454	69	39
Some	124	30	2
A little	52	4	0
None at all	50	2	0
Total	680	105	41

Estimating the population difference

Parameter of interest: Difference between the proportions of previous students and all Americans who would be bothered a great deal by the northern ice cap completely melting.

$p_{class} - p_{US}$

Point estimate: Difference between the proportions of sampled students and sampled Americans who would be bothered a great deal by the northern ice cap completely melting.

$\hat{p}_{class} - \hat{p}_{US}$: a sample proportion

Inference for comparing proportions

Details almost the same as before...

CI: point estimate ± margin of error

HT: Use $Z = \frac{point\ estimate - null\ vaue}{Standard\ Error}$

We just need the appropriate standard error for the point estimate ($SE_{class\text{-}US}$)

$$SE_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

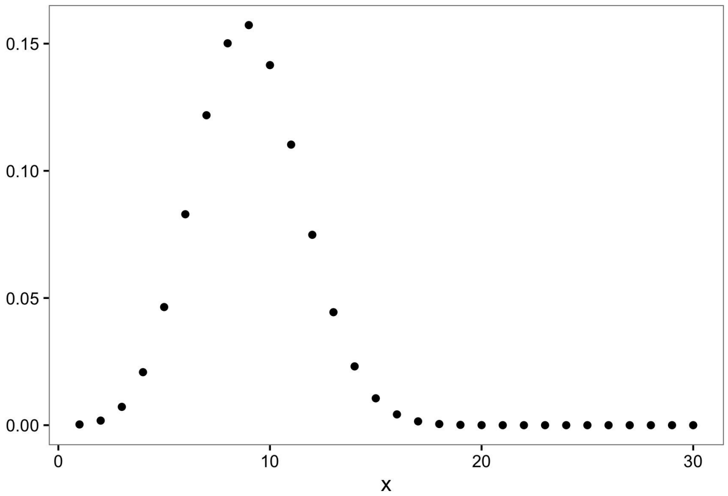
Practice Question 1: Why the new SE estimate?

Naïve intuition: Find the SE for the class data, find the SE for the US data.
Add them up

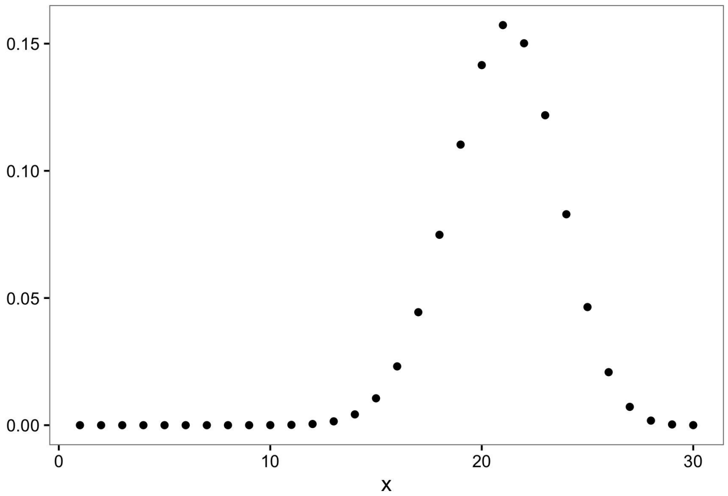
$$SE_{\hat{p}_1} = \sqrt{\frac{p_1(1-p_1)}{n_1}} \qquad SE_{\hat{p}_2} = \sqrt{\frac{p_2(1-p_2)}{n_2}}$$

$$SE_{(\hat{p}_1-\hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Why is the correct SE estimate smaller?



p=.3



p=.7

Conditions for CI for difference of proportions

Independence within groups

The people in the US group are sampled independently of each-other.

The people in the class group are sampled independently of each-other.

Independence between groups

The sampled students and US residents are independent of each-other

Success-failure

At least 10 observed successes and 10 observed failures in each group.

Proportion of differences are also nearly-normally distributed

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

$$(\hat{p}_1 - \hat{p}_2) \pm Z^{\star} \times \sqrt{\frac{\hat{p}_{class}(1-\hat{p}_{class})}{n_{class}} + \frac{\hat{p}_{US}(1-\hat{p}_{US})}{n_{US}}}$$

$$\begin{aligned}
&= (0.657 - 0.668) \pm 1.96 \times \sqrt{\frac{0.657 \times 0.343}{105} + \frac{0.668 \times 0.332}{680}} \\
&= -0.011 \pm 0.097 \\
&= (-0.108, 0.086)
\end{aligned}$$

Practice Question 2

Which of the following is the correct set of hypotheses for testing if the proportion of 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?

(a) $H_0: p_{class} = p_{US}$
$H_A: p_{class} \neq p_{US}$

(b) $H_0: \hat{p}_{class} - \hat{p}_{US} = 0$
$H_A: \hat{p}_{class} - \hat{p}_{US} \neq 0$

(c) $H_0: p_{class} - p_{US} = 0$
$H_A: p_{class} - p_{US} \neq 0$

(d) $H_0: p_{class} = p_{US}$
$H_A: p_{class} < p_{US}$

Practice Question 2

Which of the following is the correct set of hypotheses for testing if the proportion of 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?

(a) $H_0: p_{class} = p_{US}$
$H_A: p_{class} \neq p_{US}$

(b) $H_0: \hat{p}_{class} - \hat{p}_{US} = 0$
$H_A: \hat{p}_{class} - \hat{p}_{US} \neq 0$

(c) $H_0: p_{class} - p_{US} = 0$
$H_A: p_{class} - p_{US} \neq 0$

(d) $H_0: p_{class} = p_{US}$
$H_A: p_{class} < p_{US}$

A pooled estimate of the population proportion

If you think that two samples come from the same population (*p*). Or you want to test whether they do, you used a *pooled estimate* of *p̂*.

$$\hat{p} = \frac{\#\ of\ successes_1 + \#\ of\ successes_2}{n_1 + n_2}$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(\hat{p}_{pool}, \sqrt{\frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_1} + \frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_2}}\right)$$

Key ideas

1. You can use the Normal approximation for the difference of two proportions
2. The margin of error is not just the sum of the margin of errors for each proportion
3. If you think two proportions come from the same population, you can use a pooled estimate