hypothesis tests for comparing two proportions

Dr. Mine Çetinkaya-Rundel
Duke University

Duke
UNIVERSITY

A SurveyUSA poll asked respondents whether any of their children have ever been the victim of bullying. Also recorded on this survey was the gender of the respondent (the parent). Below is the distribution of responses by gender of the respondent.

	Male	Female
Yes	34	61
No	52	61
Not sure	4	0
Total	90	122
$\hat{p}$	0.38	0.50

34 / 90   61 / 122

$H_0: p_{male} - p_{female} = 0$

$H_A: p_{male} - p_{female} \neq 0$

✔ check conditions

✔ calculate test statistic & p-value

Link to poll: http://www.surveyusa.com/client/PollReport.aspx?g=1823ef50-44c7-4d2a-9efc-ead711b4ad9c

flashback to working with one proportion: $\hat{p}$ vs. p

	observed confidence interval	*expected* hypothesis test
success-failure condition	$n\hat{p} \geq 10$ $n(1-\hat{p}) \geq 10$	$np \geq 10$ $n(1-p) \geq 10$
standard error	$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$SE = \sqrt{\frac{p(1-p)}{n}}$

working with two proportions: $\hat{p}$ vs. p

	observed confidence interval	expected hypothesis test
success-failure condition	$n_1\hat{p}_1 \geq 10$ $\quad n_2\hat{p}_2 \geq 10$ $n_1(1-\hat{p}_1) \geq 10$ $\quad n_2(1-\hat{p}_2) \geq 10$	$H_0 : p_1 = p_2$
standard error	$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	

pooled proportion

$H_0 : p_1 = p_2$ = ?

Pooled proportion:

$$\hat{p}_{pool} = \frac{total\ successes}{total\ n}$$
$$= \frac{\#\ of\ successes_1 + \#\ of\ successes_2}{n_1 + n_2}$$

practice

Calculate the estimated pooled proportion of males and females who said that at least one of their children has been a victim of bullying.

$$\hat{p}_{pool} = \frac{34 + 61}{90 + 122}$$

$$\approx 0.45$$

	Male	Female
Yes	34	61
No	52	61
Not sure	4	0
Total	90	122
$\hat{p}$	0.38	0.50

revisit: working with two proportions: $\hat{p}$ vs. p

	observed confidence interval	*expected* hypothesis test
success-failure condition	$n_1\hat{p}_1 \geq 10$ $n_1(1-\hat{p}_1) \geq 10$ $n_2\hat{p}_2 \geq 10$ $n_2(1-\hat{p}_2) \geq 10$	$n_1\hat{p}_{pool} \geq 10$ $n_1(1-\hat{p}_{pool}) \geq 10$ $n_2\hat{p}_{pool} \geq 10$ $n_2(1-\hat{p}_{pool}) \geq 10$
standard error	$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$SE = \sqrt{\frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_1} + \frac{\hat{p}_{pool}(1-\hat{p}_{pool})}{n_2}}$

what about means?

parameter of interest: **μ**

$$H_0 : \mu = null \;\; value$$

$$SE = \frac{s}{\sqrt{n}}$$

μ doesn't appear in SE

parameter of interest: ***p***

$$H_0 : p = null \;\; value$$

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

p appears in SE

practice

Are conditions for inference met for conducting a hypothesis test to compare the two proportions?

	Male	Female
Total	90	122
$\hat{p}$	0.38	0.50
$\hat{p}_{pool}$	0.45	

1. independence:

✓ within groups: random sample & 10% condition

Sampled males independent of each other, sampled females are as well.

✓ between groups:

No reason to expect sampled males and females to be dependent.

2. sample size / skew: ✓ Males: $90 \times 0.45 = 40.5$ and $90 \times 0.55 = 49.5$

✓ Females: $122 \times 0.45 = 54.9$ and $122 \times 0.55 = 67.1$

We can assume that the sampling distribution of the difference between two proportions is nearly normal.

practice

Conduct a hypothesis test, at 5% significance level, evaluating if males and females are equally likely to answer "Yes" to the question about whether any of their children have ever been the victim of bullying.

	Male	Female
Total	90	122
$\hat{p}$	0.38	0.50
$\hat{p}_{pool}$	0.45	

$H_0: p_{male} - p_{female} = 0$ $\quad H_A: p_{male} - p_{female} \neq 0$

$$(\hat{p}_{male} - \hat{p}_{female}) \sim N\left(mean = 0, SE = \sqrt{\frac{0.45 \times 0.55}{90} + \frac{0.45 \times 0.55}{122}} \approx 0.0691\right)$$

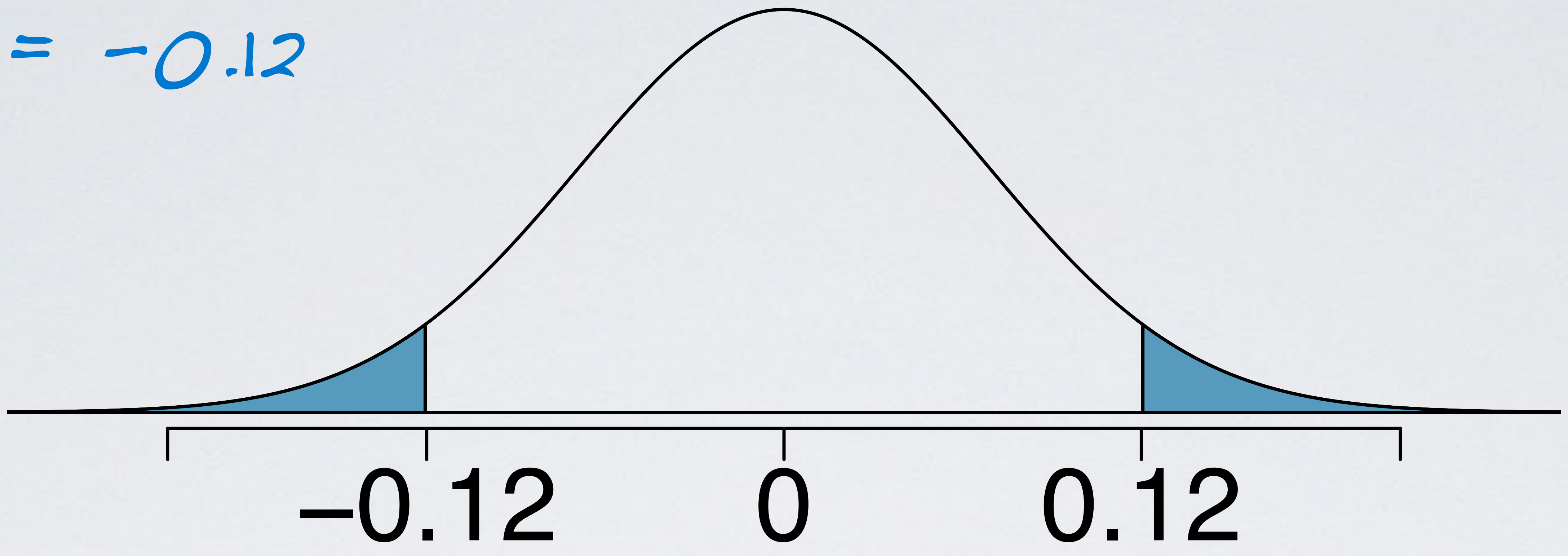
$$\text{point estimate} = \hat{p}_{male} - \hat{p}_{female} = 0.38 - 0.50 = -0.12$$

practice

point estimate = -0.12

null value = 0

SE = 0.0691



	Male	Female
Total	90	122
$\hat{p}$	0.38	0.50
$\hat{p}_{pool}$	0.45	

$$Z = \frac{-0.12 - 0}{0.0691} \approx -1.74$$

$$\text{p-value} = P(|Z| > 1.74) \approx 0.08$$