estimating the difference between two proportions

Dr. Mine Çetinkaya-Rundel
Duke University

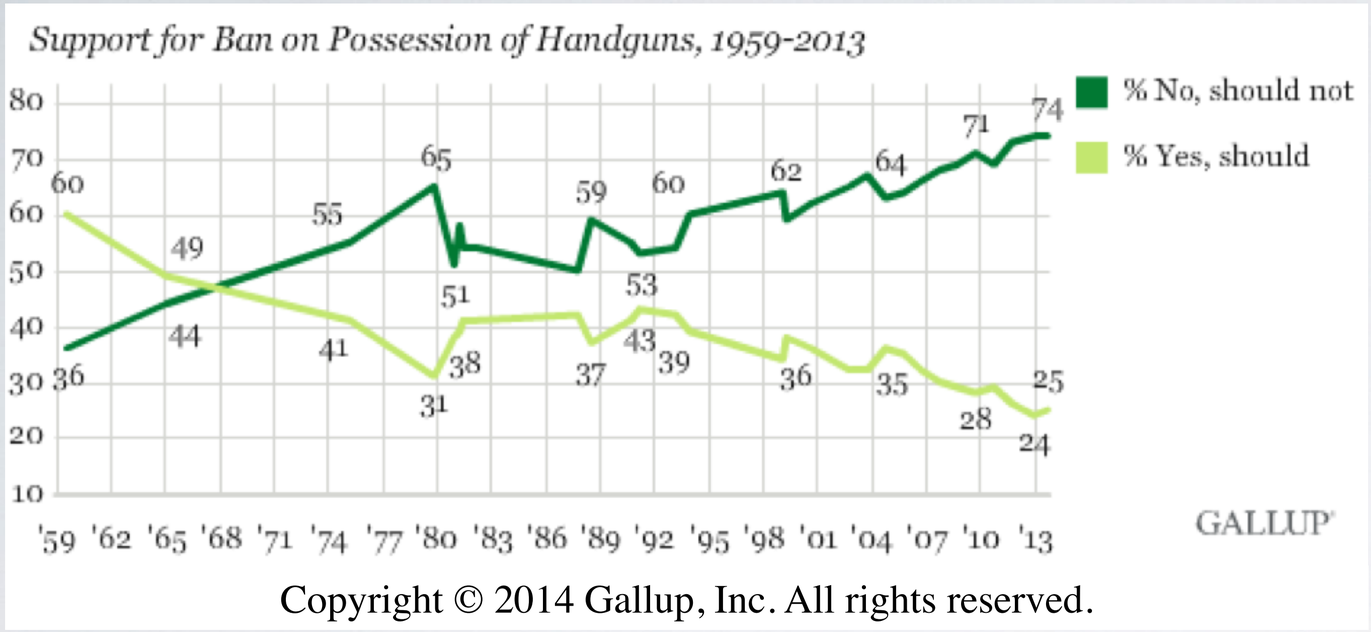
In early October 2013, a Gallup poll asked "Do you think there should or should not be a law that would ban the possession of handguns, except by the police and other authorized persons?"

(a) No, there should not be such a law
(b) Yes, there should be such a law
(c) No opinion

Link to poll: http://www.gallup.com/poll/165563/remains-divided-passing-stricter-gun-laws.aspx

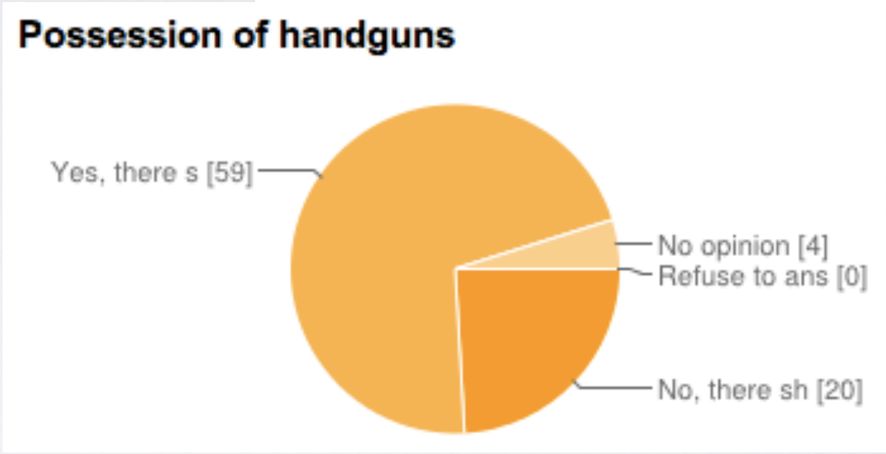
Image source: http://openclipart.org/detail/14213/not-for-sale-by-cibo00-14213

US



Coursera

No, there should not be such a law	**20**	24%
Yes, there should be such a law	**59**	71%
No opinion	**4**	5%



	$suc.$	n	$\hat{p}$
US	257	1028	0.25
Coursera	59	83	0.71

How do Coursera students and the American public at large compare with respect to their views on laws banning possession of handguns?

parameter of interest

Difference between the proportions of **all** Coursera students and **all** Americans who believe there should be a ban on possession of handguns.

$$p_{Coursera} - p_{US}$$

point estimate

Difference between the proportions of **sampled** Coursera students and **sampled** Americans who believe there should be a ban on possession of handguns.

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

estimating the difference between two proportions

point estimate ± margin of error

$$(\hat{p}_1 - \hat{p}_2) \pm z^\star SE_{(\hat{p}_1 - \hat{p}_2)}$$

Standard error for difference between two proportions, for calculating a confidence interval:

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

Conditions for inference for comparing two independent proportions:

1. ***Independence:***
   - ✓ **within groups:** sampled observations must be independent within each group
     - random sample/assignment
     - if sampling without replacement, n < 10% of population
   - ✓ **between groups:** the two groups must be independent of each other (non-paired)
2. ***Sample size/skew:*** Each sample should meet the success-failure condition:
   - ✓ $n_1p_1 \geq 10$ and $n_1(1-p_1) \geq 10$
   - ✓ $n_2p_2 \geq 10$ and $n_2(1-p_2) \geq 10$

practice

Using a 95% confidence interval, estimate how Coursera students and the American public at large compare with respect to their views on laws banning possession of handguns.

	$suc.$	n	$\hat{p}$
US	257	1028	0.25
Coursera	59	83	0.71

1. independence: ✓ random sample: yes for US, no for Coursera
   ✓ 10% condition: met for both

Sampled Americans independent of each other, sampled Courserians may not be.

2. sample size / skew: ✓ US: 257 successes, 1028 − 257 = 771 failures
   ✓ Coursera: 59 successes, 83 − 59 = 24 failures

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

practice

	$suc.$	n	$\hat{p}$
US	257	1028	0.25
Coursera	59	83	0.71

$$(\hat{p}_{Coursera} - \hat{p}_{US}) \pm z^* \, SE =$$

$$= (0.71 - 0.25) \pm 1.96 \sqrt{\frac{0.71 \times 0.29}{83} + \frac{0.25 \times 0.75}{1028}}$$

$$= 0.46 \pm 1.96 \times 0.0516$$

$$= 0.46 \pm 0.10$$

$$= (0.36, 0.56)$$

does the order matter?

remember $(\hat{p}_1 - \hat{p}_2) \pm z^\star \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

can be - or +

always +

$$(p_{Coursera} - p_{US}) = (p_{US} - p_{Coursera}) =$$

$$= (0.71 - 0.25) \pm 0.10 \qquad = (0.25 - 0.71) \pm 0.10$$

$$= 0.46 \pm 0.10 \qquad = -0.46 \pm 0.10$$

$$= (0.36, 0.56) \qquad = (-0.56, -0.36)$$

practice

Based on the confidence interval we calculated, should we expect to find a significant difference (at the equivalent significance level) between the population proportions of Coursera students and the American public at large who believe there should be a law banning the possession of handguns?

$$(p_{Coursera} - p_{US}) = (0.36, 0.56)$$

$H_0: p_{Coursera} - p_{US} = 0$

reject H_0