2-Proportion Confidence Interval and Significance Test

Is yawning contagious? Results from Mythbusters experiment:

exp.

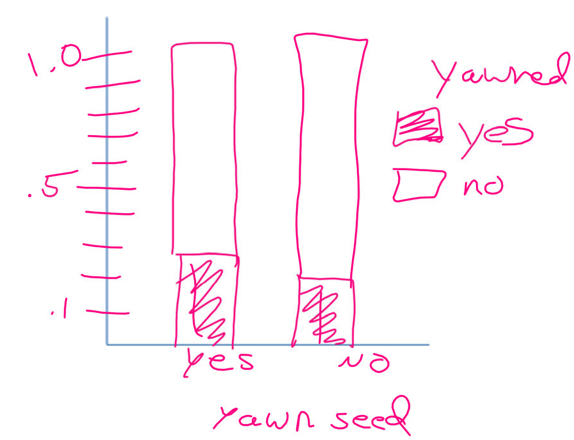
resp.

Did subject yawn?	Yawn seed (1)	No yawn seed (2)	Total
Yes	10	4	14
No	24	12	36
Total	34	16	50

What proportion of those who had the yawn seed yawned? $\hat{p}_1 =$ 10/34 = .294

What proportion those who didn't have the yawn seed yawned? $\hat{p}_2 =$ 4/16 = .25

Draw a segmented bar graph:



Difference in the sample proportions: $\hat{p}_1 - \hat{p}_2 =$ .294 − .25 = .044

Relative Risk: .294/.25 = 1.176

Those with the yawn seed were 1.176 times more likely to yawn than those without.

Are the variables whether or not there was a yawn seed and whether or not the subject yawned independent?

yes → bars ≈ same

If the variables are independent (yawn seed had no effect on whether or not subject yawned): combined

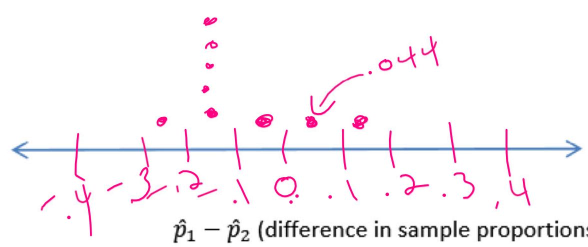
What proportion yawned? $\hat{p}_c =$ 14/50 = .28

How many would you expect to yawn in each group?

① .28(34) = 9.52 ② (16)(.28) = 4.48

What would you expect the difference in the proportions between the two groups to be?

$p_1 - p_2 =$ 0



$\hat{p}_1 - \hat{p}_2$ (difference in sample proportions that yawned)

What proportion of times did we get the observed difference $\hat{p}_1 - \hat{p}_2 =$ .044? 257/509 ≈ .5

What does this number represent (interpret)?

p-value - prob. randomly get a diff. of .044 or greater, if there's actually no diff.

Does this provide evidence that more people yawn with the yawn seed than without?

NO .5 not sign.

2-Proportion Confidence Interval and Significance Test

Repeat the simulation if $\hat{p}_1 = \frac{12}{34} = .353$ and $\hat{p}_2 = \frac{2}{16} = .125$

What proportion of times did we get the observed difference $\hat{p}_1 - \hat{p}_2 = .228$? 45/509 ≈ .088

Does this provide evidence that more people yawn with the yawn seed than without?

yes, some .088 is sign. @ .10

Statistical significance: the likelihood of an observed result by asking how often such an extreme result would occur by chance alone. If the sample result is unlikely to occur by chance alone, it is said to be statistically significant. The probability of obtaining a result at least as extreme as the sample by chance alone is known as the p-value

The CLT says the sampling distribution of $\hat{p}_1 - \hat{p}_2$ (the difference in the sample proportions) will be:

- Approx. normal if

  $n_1\hat{p}_1 \geq 10 \quad n_1(1-\hat{p}_1) \geq 10 \qquad n_2\hat{p}_2 \geq 10 \quad n_2(1-\hat{p}_2) \geq 10$

- Have a mean of $\boxed{p_1 - p_2}$ (the difference in the actual proportions) (usually 0)
- Have a standard deviation of

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

2-Proportion Confidence Interval and Significance Test

21-2

Infant sleeping position	1992	1996	Total
Stomach	700	240	940
Back	300	760	1060
Total	1000	1000	2000

Did they Reduce prop. of babies ~~of babies~~ sleeping on Stomach from '92 to '96?

$\hat{p}_1 = .7$ $\hat{p}_2 = .24$ $\hat{p}_c = \frac{940}{2000} = .47$

Define Parameter/Write Hypotheses:

p_{92} = actual prop. of babies sleep stomach in 1992.

p_{96} = " " " in 1996

$H_0: p_{92} = p_{96}$

$H_a: p_{92} > p_{96}$

Check Conditions:

Large counts:

$1000(.7) = 700$

$1000(1-.7) = 300$

$1000(.24) = 240$

$1000(1-.24) = 760$

≥ 10

* can use 5 instead of 10

Two independent random sample the two populations (obs. Study)

Or random assignment to two groups (experiment)

Test statistic/p-value:

$$Z = \frac{.7 - .24}{\sqrt{\frac{.47(1-.47)}{1000} + \frac{.47(1-.47)}{1000}}}$$

$Z = 20.61$

p-value ≈ 0

Conclusion:

p-value ≈ 0

Sign. at any reasonable level

Reject H_0

Evid. prop. sleep on stomach dec. from '92 to '96.

2-Proportion Confidence Interval and Significance Test

Confidence Interval:

(conditions are the same as significance test)

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

$$(.7 - .24) \pm 1.96 \sqrt{\frac{.7(1-.7)}{1000} + \frac{.24(1-.24)}{1000}}$$

$$.46 \pm .03883$$

$$(.42, .499)$$

95% conf. the diff in prop. of babies sleep on stomach is in this interval.

Confidence interval estimate the *difference* in the population proportions, not the actual values.

What does it tell you if the values are both positive?

$$\hat{p}_1 > \hat{p}_2$$

Both negative?

$(-.499, -.42)$ $\hat{p}_1 < \hat{p}_2$

What if 0 is included in the interval?

ex. $(-.21, .08)$

0 is in interval, there might be no diff.