Lesson 37: Is there a shift in men's risk taking behaviour when women are present?

Scenario:

A seatbelt observational study in Massachusetts

- found male drivers wear seatbelts less often than women do
- Men's seatbelt wearing jumped 16% when they had a female passanger along

Seatbelt use was recorded at 161 locations in MA using random sampling
...of 4208 male drivers w/ female passengers, 2777 (66%) were belted
...of 2763 male drivers with male passengers, 1363 (49.3%) were belted
Is there a shift in men's risk taking behavior when women are present? What would we estimate the true size of this gap to be?

Buildup:

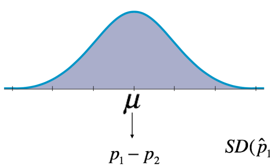
Conditions: Independence, randomization, 10% condition

+

Independent groups assumption: compared groups must be independent of each other

Then

The sampling distribution model for difference between two independent proportions $\hat{p}_1 - \hat{p}_2$ is modeled by



$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_1}{n_1}}$$

A TWO-PROPORTION z-INTERVAL

When the conditions are met, we are ready to find the confidence interval for the difference of two proportions, $p_1 - p_2$. The confidence interval is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \times SE(\hat{p}_1 - \hat{p}_2)$$

where we find the standard error of the difference,

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

from the observed proportions.

The critical value z^* depends on the particular confidence level, C, that we specify.

because we don't know actual proportion, we have only samples

I want to know whether there is a differnce in the seatbelt wearing behavior of men w/ male passengers vs. female passengers.

I am interested in the difference $p_f - p_m$

I will build a 95% confidence interval.

Conditions:

Independence Assumption: random sampling used, behavior independent from car to car

Randomization: NHSTA used random sampling

10% condition: Drivers were less than 10% MA drivers

Independent groups: Random sampling used + no reason to believe that they are not independent

Success/ Failure in both groups:

2777 wore seatbelts	1363 wore seatbelts	≥ 10
1431 did not	1400 didn't	

Two proportion z- interval

$$(\hat{p}_f - \hat{p}_m) \pm z^* \sqrt{\frac{\hat{p}_f \hat{q}_f}{n_1} + \frac{\hat{p}_m \hat{q}_m}{n_2}}$$

$$= (.66 - .49) \pm 1.96 \sqrt{\frac{(.66)(.34)}{4208} + \frac{(.493)(.507)}{2763}}$$

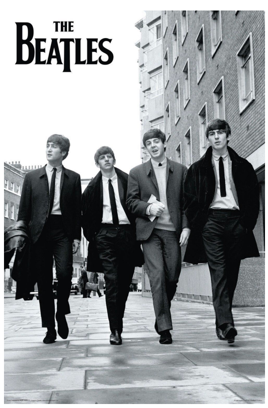
$$= .167 \pm .024$$

$$\Rightarrow 14.35 \text{ to } 19.1\%$$

I am 95% confident that the proportion of male drivers who wear seatbelts when driving next to female drivers is between 14.3% and 19.1% points higher than the proportion who wear seatbelts when driving next to male passengers.

When I get older, losing my hair, many years from now...

Will I snore when I am 64?



Scenario: National Sleep Foundation (in 2001)

Sampled 1010 U.S. Adults about sleep habits

995 Respondents

37% snored a few times a week

26% of the 184 people under age 30 snored

39% of the 811 in older group snored

Is this difference of 13% real or due to natural fluctuations in the sample that we've chosen?

We need to do a hypothesis test!

TWO-PROPORTION z-TEST

The conditions for the two-proportion z-test are the same as for the two-proportion z-interval. We are testing the hypothesis

$$H_0: p_1 - p_2 = 0.$$

Because we hypothesize that the proportions are equal, we pool the groups to find

$$\hat{p}_{\text{pooled}} = \frac{Success_1 + Success_2}{n_1 + n_2}$$

and use that pooled value to estimate the standard error:

$$SE_{\text{pooled}}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_1} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_2}}.$$

Now we find the test statistic,

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_{\text{pooled}}(\hat{p}_1 - \hat{p}_2)}.$$

When the conditions are met and the null hypothesis is true, this statistic follows the standard Normal model, so we can use that model to obtain a P-value.

H_0: There is no difference in snoring rates in the two age groups:

$p_{old} - p_{young} = 0$.

H_A: The rates are different: $p_{old} - p_{young} \neq 0$.

Conditions:

Independence: random sample likely independent

Randomization: respondents were randomly selected

10% condition: respondents in each age group certainly less than 10% of respective populations

Independent groups: random sample -> independent groups

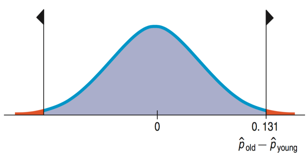
Success/Failure: 48 snored, 136 didn't; 318 snored, 493 didn't ≥ 10

⟶ Two proportion z-test

$$\hat{p}_{pooled} = \frac{y_{old} + y_{young}}{n_{old} + n_{young}} = \frac{318 + 48}{811 + 184} = .3678$$

$$SE(\hat{p}_{old} - \hat{p}_{young}) = \sqrt{\frac{\hat{p}_{pooled}\hat{p}_{pooled}}{811} + \frac{\hat{p}_{pooled}\hat{p}_{pooled}}{184}} \approx .039375$$

$$z = \frac{(\hat{p}_{old} - \hat{p}_{young}) - 0}{SE(\hat{p}_{old} - \hat{p}_{young})} = \frac{.131 - 0}{.039375} = 3.33$$



$$P - value = 2P(z \geq 3.33) = .0008$$

Low p-value. I reject the null hypothesis. I conclude that there is a difference in the rate of snoring between older adults and younger adults. It appears that the older adults are more likely to snore.