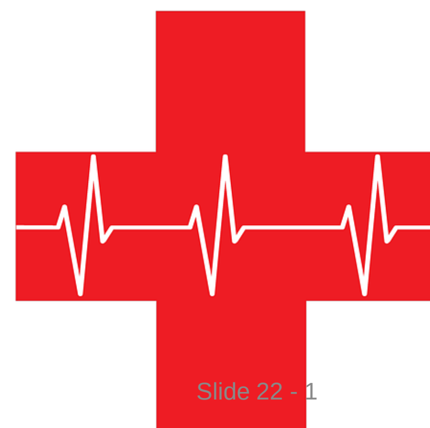
Tuesday, February 26, 2019

- Warm-up

  A recent medical study focused on the relationship between estrogen use and premature death rates among post-menopausal women. Researchers searched the medical records of a large health care maintenance organization for women born between 1900 and 1915 who had taken estrogen supplements for at least one year starting in 1969. There were **232** such women and **53** of them had died prematurely (not defined) from all causes. The researchers also selected a sample of records of women born between 1900 and 1915 who had not taken estrogen supplements at all. There were **222** women in this sample and **87** of them had died prematurely from all causes. Assuming that these samples of women are representative samples from the populations of women born between 1900 and 1915 who did and did not take estrogen supplements, do these data provide evidence that the premature death rates for these two populations are different?

- Check Homework
- 2 proportion confidence intervals



Objectives

- Content Objective: I will make connections between past learnings and the new chapter ideas.
- Social Objective: I will listen well and participate in class.
- Language Objective: I will listen well and take good notes so the reading assignment goes well.

A recent medical study focused on the relationship between estrogen use and premature death rates among post-menopausal women. Researchers searched the medical records of a large health care maintenance organization for women born between 1900 and 1915 who had taken estrogen supplements for at least one year starting in 1969. There were **232** such women and **53** of them had died prematurely (not defined) from all causes. The researchers also selected a sample of records of women born between 1900 and 1915 who had not taken estrogen supplements at all. There were **222** women in this sample and **87** of them had died prematurely from all causes. Assuming that these samples of women are representative samples from the populations of women born between 1900 and 1915 who did and did not take estrogen supplements, do these data provide evidence that the premature death rates for these two populations are different?

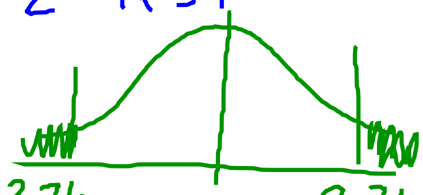
two proport. z-test

$H_0: p_e = p_n \quad (p_e - p_n = 0)$

$H_A: p_e \neq p_n \quad (p_e - p_n \neq 0)$

z → -3.76

p-value → 0.000164



estrogen | not
represent = stated → ''
10% 232 < 10% of women on estrogen | 222 < 10% of all women not on estrogen
assume independent groups
53 ≥ 10
232-53 ≥ 10 → success/fail ← 87 ≥ 10
222-87 ≥ 10

Due to a very low p-value of 0.000164, we reject the null. These data provide evidence that the premature death rates for these two populations are different.

Proportions observed in independent random samples *are* independent. Thus, we can add their variances. So…

The standard deviation of the difference between two sample proportions is

$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

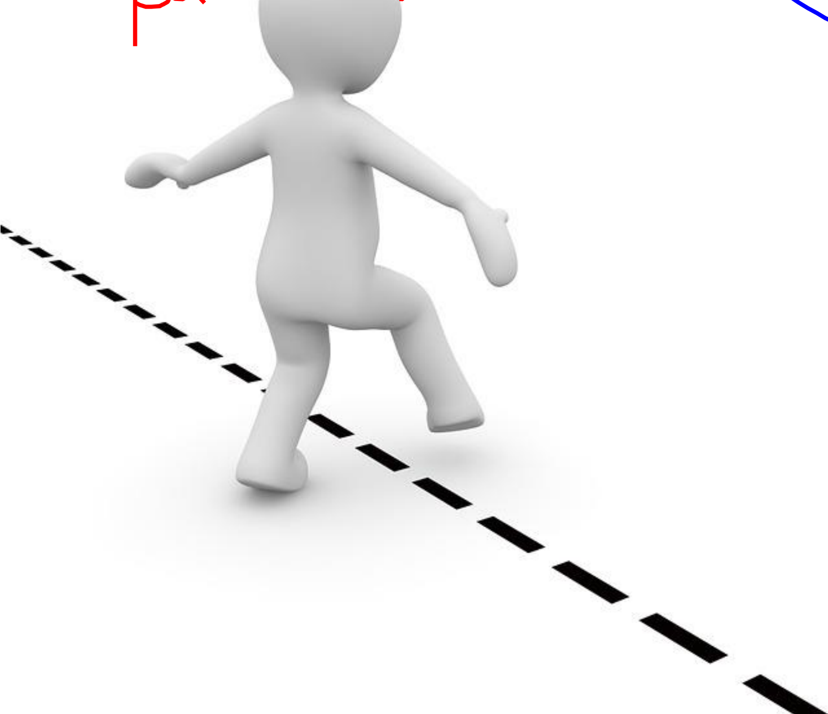
population → parameter

Thus, the standard error is

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

Sample data

Statistic



Two-Proportion *z*-Interval

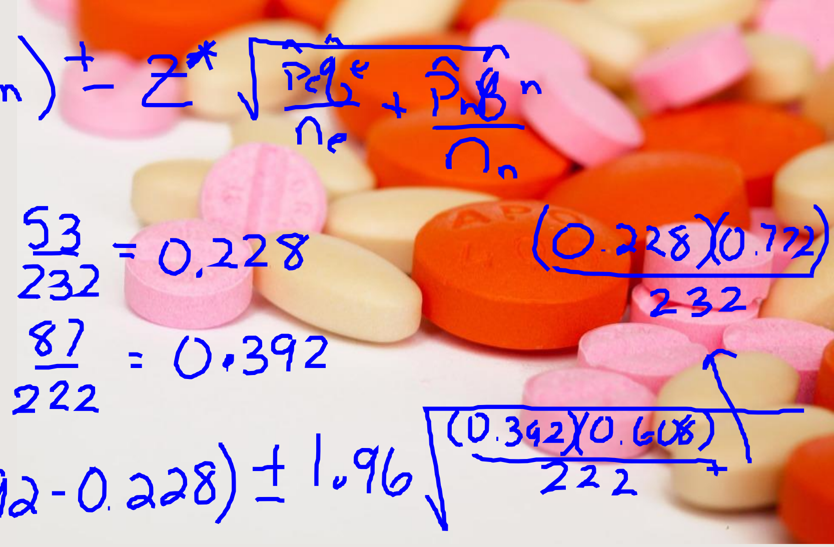
- When the conditions are met, we are ready to find the confidence interval for the difference of two proportions: $\hat{p} \pm z^* SE(\hat{p})$
- The confidence interval is

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

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

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

A recent medical study focused on the relationship between estrogen use and premature death rates among post-menopausal women. Researchers searched the medical records of a large health care maintenance organization for women born between 1900 and 1915 who had taken estrogen supplements for at least one year starting in 1969. There were **232** such women and **53** of them had died prematurely (not defined) from all causes. The researchers also selected a sample of records of women born between 1900 and 1915 who had not taken estrogen supplements at all. There were **222** women in this sample and **87** of them had died prematurely from all causes. Assuming that these samples of women are representative samples from the populations of women born between 1900 and 1915 who did and did not take estrogen supplements, do these data provide evidence that the premature death rates for these two populations are different? How big might the difference in the two population death rates be?



$$(\hat{p}_e - \hat{p}_n) \pm z^* \sqrt{\frac{\hat{p}_e\hat{q}_e}{n_e} + \frac{\hat{p}_n\hat{q}_n}{n_n}}$$

$$\hat{p}_e = \frac{53}{232} = 0.228$$

$$\hat{p}_n = \frac{87}{222} = 0.392$$

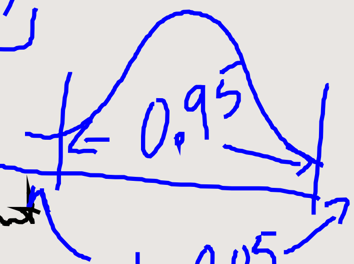
$$(0.392 - 0.228) \pm 1.96\sqrt{\frac{(0.392)(0.608)}{222} + \frac{(0.228)(0.772)}{232}}$$

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

$\alpha = 0.05$

$0.164 \pm 1.96 \cdot 0.0427$

0.164 ± 0.083



I am 95% confident that the true proportion difference between population death rates is 16.4% ± 8.3%

Another Example

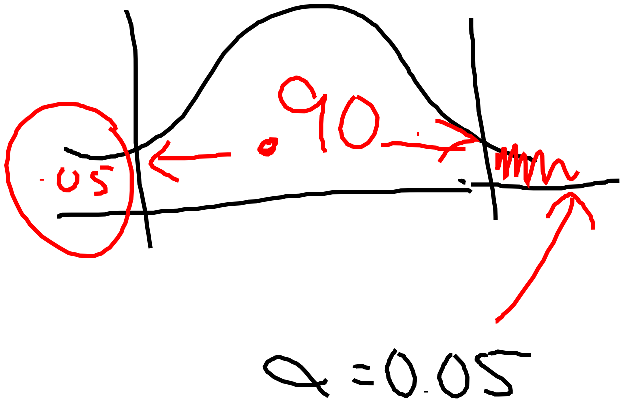
Women tend to fear crime more than men, even though they are less likely to be victims of crime. One study recruited separate random samples of 56 black women and 63 black men over the age of 65 from Atlantic City, New Jersey. Of the women, 27 said they “felt vulnerable” to crime and 46 of the men said this. Create a confidence interval to verify this claim at α = 0.05.

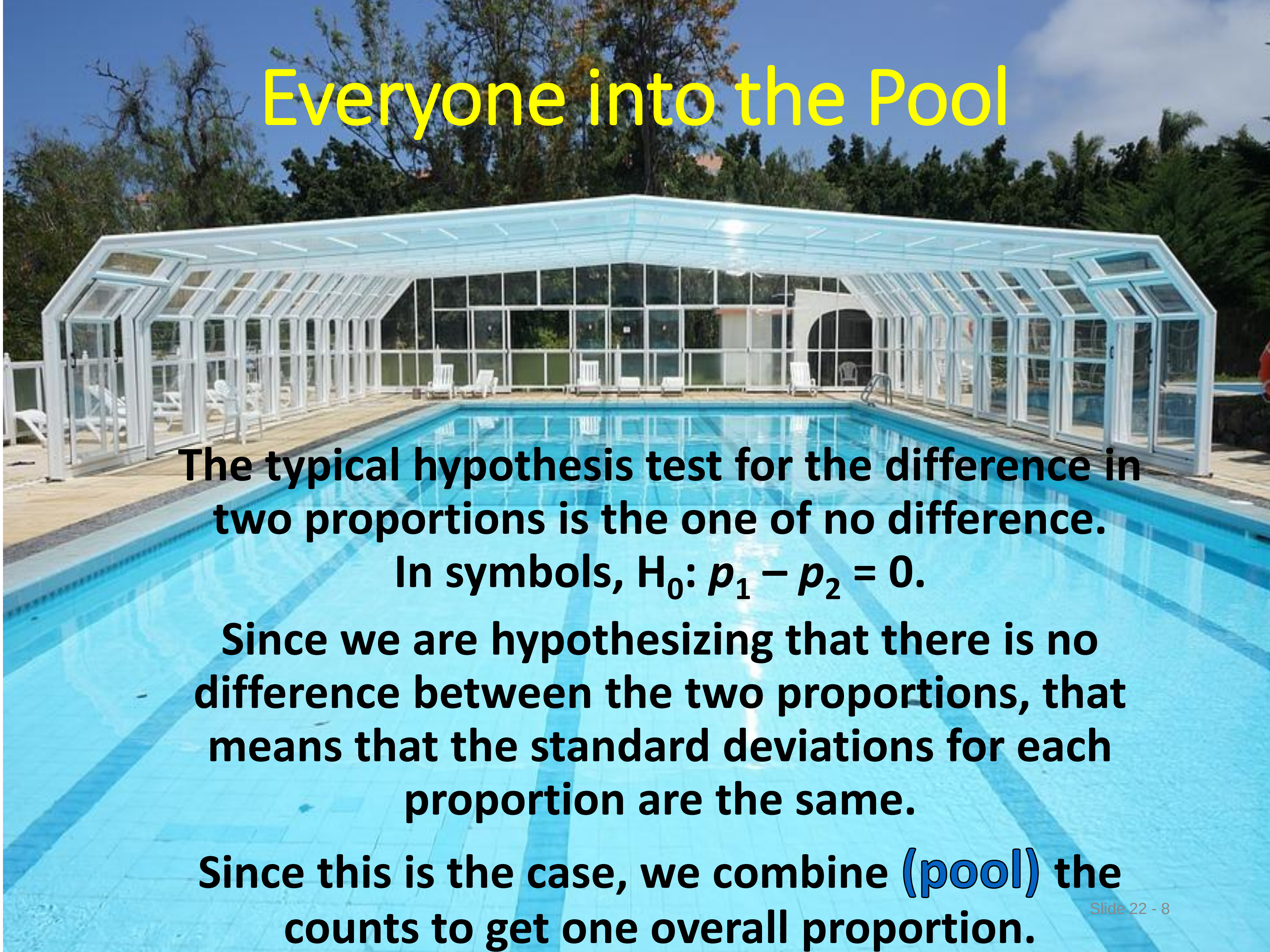
2 prop confidence interval

-0.24 ± 0.14

X_1 = women

2 = men





Everyone into the Pool

The typical hypothesis test for the difference in two proportions is the one of no difference.
In symbols, H_0: $p_1 - p_2 = 0$.

Since we are hypothesizing that there is no difference between the two proportions, that means that the standard deviations for each proportion are the same.

Since this is the case, we combine (pool) the counts to get one overall proportion.

What Can Go Wrong?

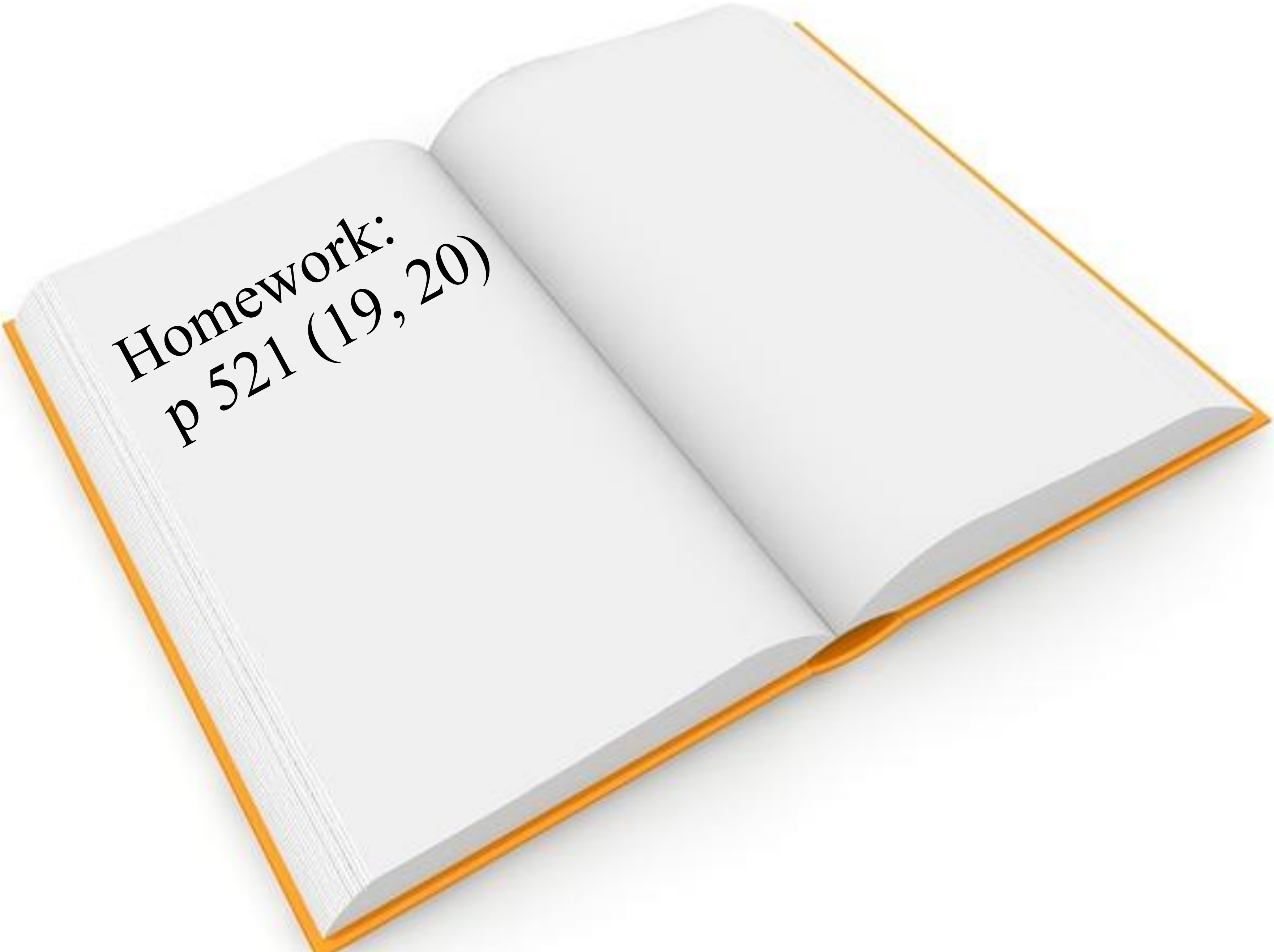
- Don't use two-sample proportion methods when the samples aren't independent.
  - These methods give wrong answers when the independence assumption is violated.
- Don't apply inference methods when there was no randomization.
  - Our data must come from representative random samples or from a properly randomized experiment.
- Don't interpret a significant difference in proportions causally.
  - Be careful not to jump to conclusions about causality.

What have we learned?

Perhaps the most important thing to remember is that the concepts and interpretations are essentially the same—only the mechanics have changed slightly.

Hypothesis tests and confidence intervals for the difference in two proportions are based on Normal models.

- Both require us to find the standard error of the difference in two proportions.
  - We do that by adding the variances of the two sample proportions, assuming our two groups are independent.


Homework:
p 521 (19, 20)