

# Comparing Two Proportions

## Sections 23.1, 23.2, 23.3

### Lecture 43

Robb T. Koether

Hampden-Sydney College

Thu, Apr 7, 2016

# Outline

- 1 Two Proportions
- 2 Confidence Intervals
- 3 Assignment

# Outline

1 Two Proportions

2 Confidence Intervals

3 Assignment

# Two Proportions

- We will now consider situations where we have two populations and we wish to compare the proportion found in one to the proportion found in the other.

# Two Proportions

- We will now consider situations where we have two populations and we wish to compare the proportion found in one to the proportion found in the other.
- For example, we might compare the voting habits of men vs. women.

# Two Proportions

- We will now consider situations where we have two populations and we wish to compare the proportion found in one to the proportion found in the other.
- For example, we might compare the voting habits of men vs. women.
- Or the engagement of freshmen vs. upperclassmen in campus life.

# Two Proportions

- We will now consider situations where we have two populations and we wish to compare the proportion found in one to the proportion found in the other.
- For example, we might compare the voting habits of men vs. women.
- Or the engagement of freshmen vs. upperclassmen in campus life.
- Or social media use between two different age groups.

# Two Proportions

- Let  $p_1$  be the population proportion in Population #1.
- We take a sample of size  $n_1$  from Population #1 and find a sample proportion  $\hat{p}_1$ .
- Define  $p_2$ ,  $n_2$ , and  $\hat{p}_2$  similarly for Population #2.



- To compare  $p_1$  to  $p_2$ , we will use  $\hat{p}_1 - \hat{p}_2$  to estimate their difference.
- Thus, the question, what is the sampling distribution of the statistic  $\hat{p}_1 - \hat{p}_2$ ?

- To compare  $p_1$  to  $p_2$ , we will use  $\hat{p}_1 - \hat{p}_2$  to estimate their difference.
- Thus, the question, what is the sampling distribution of the statistic  $\hat{p}_1 - \hat{p}_2$ ?
- That is,

- To compare  $p_1$  to  $p_2$ , we will use  $\hat{p}_1 - \hat{p}_2$  to estimate their difference.
- Thus, the question, what is the sampling distribution of the statistic  $\hat{p}_1 - \hat{p}_2$ ?
- That is,
  - What is the **mean** of  $\hat{p}_1 - \hat{p}_2$ ?

- To compare  $p_1$  to  $p_2$ , we will use  $\hat{p}_1 - \hat{p}_2$  to estimate their difference.
- Thus, the question, what is the sampling distribution of the statistic  $\hat{p}_1 - \hat{p}_2$ ?
- That is,
  - What is the **mean** of  $\hat{p}_1 - \hat{p}_2$ ?
  - What is the **standard deviation** of  $\hat{p}_1 - \hat{p}_2$ ?

- To compare  $p_1$  to  $p_2$ , we will use  $\hat{p}_1 - \hat{p}_2$  to estimate their difference.
- Thus, the question, what is the sampling distribution of the statistic  $\hat{p}_1 - \hat{p}_2$ ?
- That is,
  - What is the **mean** of  $\hat{p}_1 - \hat{p}_2$ ?
  - What is the **standard deviation** of  $\hat{p}_1 - \hat{p}_2$ ?
  - What is the **shape** of the distribution?

## Fact

- The mean of  $\hat{p}_1 - \hat{p}_2$  is  $p_1 - p_2$ .
- The standard deviation of  $\hat{p}_1 - \hat{p}_2$  is

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}.$$

- The distribution is approximately normal, provided  $n_1$  and  $n_2$  are large.

- As usual, we cannot compute

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

because we do not know the values of  $p_1$  and  $p_2$ .

- So we substitute  $\hat{p}_1$  and  $\hat{p}_2$  and get the **standard error**:

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

# Outline

1 Two Proportions

**2 Confidence Intervals**

3 Assignment



- The formula for a confidence interval for  $p_1 - p_2$  is

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

# Example

## Example

- A survey conducted by Rasmussen reports that 80% of Republicans believe that illegal immigrants should not be allowed to vote, while 52% of Democrats believe that they should not be allowed to vote.
- Assume that the survey included 350 Republicans and 400 Democrats.
- Find a 95% confidence interval for the difference in proportions between all Republicans and all Democrats.

# Example

## Example

- One year ago, 42% of Democrats in a survey opposed allowing illegal immigrants to vote. Today, in a survey, 52% oppose it.
- Assume that both surveys included 400 Democrats.
- Find a 95% confidence interval for the difference in proportions between Democrats today and Democrats one year ago.

# Example

## Example

- The same survey showed that 52% of Democrats oppose allowing illegal immigrants to vote and 39% support it.
- Why should we not calculate a confidence interval for this difference?

# Outline

1 Two Proportions

2 Confidence Intervals

**3 Assignment**

# Assignment

## Assignment

- Read Section 23.1, 23.2, 23.3.
- Apply Your Knowledge: 1, 2, 3.
- Check Your Skills: 9, 10, 13.
- Exercises 17, 18, 19, 28.