# Independent Samples–Comparing Proportions

Lecture 40
Section 11.5

Robb T. Koether

Hampden-Sydney College

Fri, Apr 6, 2012

# Outline

# Outline

# Introduction

- Every week, pollster Scott Rasmussen surveys 3,500 likely voters and asks them whether the country is heading in the "right direction."

# Introduction

- Every week, pollster Scott Rasmussen surveys 3,500 likely voters and asks them whether the country is heading in the "right direction."
- In the week of Feb 6 - 12, the percentage who said "yes" reached its highest value (34%) in over a year.

# Introduction

- Every week, pollster Scott Rasmussen surveys 3,500 likely voters and asks them whether the country is heading in the "right direction."
- In the week of Feb 6 - 12, the percentage who said "yes" reached its highest value (34%) in over a year.
- This past week, the value was 29%

# Introduction

- Every week, pollster Scott Rasmussen surveys 3,500 likely voters and asks them whether the country is heading in the "right direction."
- In the week of Feb 6 - 12, the percentage who said "yes" reached its highest value (34%) in over a year.
- This past week, the value was 29%
- Does the drop of 5 percentage points indicate that fewer people think the country is headed in the right direction now than did two months ago?

# Introduction

- Every week, pollster Scott Rasmussen surveys 3,500 likely voters and asks them whether the country is heading in the "right direction."
- In the week of Feb 6 - 12, the percentage who said "yes" reached its highest value (34%) in over a year.
- This past week, the value was 29%
- Does the drop of 5 percentage points indicate that fewer people think the country is headed in the right direction now than did two months ago?
- Or can the change be attributed to sampling error?

# Introduction

- In this situation, we have two populations:

# Introduction

- In this situation, we have two populations:
  - The population of likely voters on Feb 6.

# Introduction

- In this situation, we have two populations:
  - The population of likely voters on Feb 6.
  - The population of likely voters on Mar 26.

# Introduction

- In this situation, we have two populations:
  - The population of likely voters on Feb 6.
  - The population of likely voters on Mar 26.
- Let $p_1$ be the (true) proportion who believed on Feb 6 that the country is headed in the right direction.

# Introduction

- In this situation, we have two populations:
  - The population of likely voters on Feb 6.
  - The population of likely voters on Mar 26.
- Let $p_1$ be the (true) proportion who believed on Feb 6 that the country is headed in the right direction.
- Let $p_2$ be the (true) proportion who believed on Mar 26 that the country is headed in the right direction.

# Introduction

- In this situation, we have two populations:
  - The population of likely voters on Feb 6.
  - The population of likely voters on Mar 26.
- Let $p_1$ be the (true) proportion who believed on Feb 6 that the country is headed in the right direction.
- Let $p_2$ be the (true) proportion who believed on Mar 26 that the country is headed in the right direction.
- We *know* that $\hat{p}_1 > \hat{p}_2$.

# Introduction

- In this situation, we have two populations:
  - The population of likely voters on Feb 6.
  - The population of likely voters on Mar 26.
- Let $p_1$ be the (true) proportion who believed on Feb 6 that the country is headed in the right direction.
- Let $p_2$ be the (true) proportion who believed on Mar 26 that the country is headed in the right direction.
- We *know* that $\hat{p}_1 > \hat{p}_2$.
- The question is, is $p_1 > p_2$?

# Outline

# Hypothesis Testing Procedure

**Example (Testing hypotheses concerning $p_1 - p_2$)**

- Test the hypothesis, at the 5% level, that the proportion who believe that the country is headed in the right direction has decreased from Feb 6 to Mar 26.

# Hypothesis Testing Procedure

**Example (Testing hypotheses concerning $p_1 - p_2$)**

(1) Let $p_1$ be the proportion who on Feb 6 believed the country is headed in the right direction.
Let $p_2$ be the proportion who on Mar 26 believed the country is headed in the right direction.
The hypotheses are

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 > p_2$$

# Hypothesis Testing Procedure

**Example (Testing hypotheses concerning $p_1 - p_2$)**

(1) Let $p_1$ be the proportion who on Feb 6 believed the country is headed in the right direction.
Let $p_2$ be the proportion who on Mar 26 believed the country is headed in the right direction.
The hypotheses are

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 > p_2$$

(2) The significance level is $\alpha = 0.05$.

# Outline

# Hypothesis Testing Procedure

- What is the test statistic?
- It depends on the sampling distribution of $\hat{p}_1 - \hat{p}_2$.

# Hypothesis Testing Procedure

- What is the test statistic?
- It depends on the sampling distribution of $\hat{p}_1 - \hat{p}_2$.
- Here we go again...

# The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

- If the sample sizes are large enough, then $\hat{p}_1$ is

$$N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right)$$

and $\hat{p}_2$ is

$$N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right).$$

CENTRAL LIMIT THEOREM

# The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

- If the sample sizes are large enough, then $\hat{p}_1$ is

$$N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right)$$

and $\hat{p}_2$ is

$$N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right).$$

CENTRAL LIMIT THEOREM

# The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

- The sample sizes will be large enough if

$$n_1 p_1 \geq 5, \text{ and } n_1(1 - p_1) \geq 5,$$

and

$$n_2 p_2 \geq 5, \text{ and } n_2(1 - p_2) \geq 5.$$

# The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

- This is equivalent to saying that we had at least 5 yes's and at least 5 no's in each sample.

# The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

- Therefore, $\hat{p}_1 - \hat{p}_2$ is normal with mean

$$\begin{aligned}\mu_{\hat{p}_1 - \hat{p}_2} &= \mu_{\hat{p}_1} - \mu_{\hat{p}_2} \\ &= p_1 - p_2\end{aligned}$$

  and standard deviation

$$\begin{aligned}\sigma_{\hat{p}_1 - \hat{p}_2} &= \sqrt{\sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_1}} \\ &= \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}.\end{aligned}$$

# The Test Statistic

- Therefore, the test statistic *would* be

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

  except that...

# The Test Statistic

- ...we do not know the values of $p_1$ and $p_2$.
- We will use $\hat{p}_1$ and $\hat{p}_2$ to approximate $p_1$ and $p_2$.
- Therefore, the test statistic *would* be

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

except that...

# Pooled Estimate of $p$

- ...we can do better by pooling the data.
- The null hypothesis assumes that $p_1 = p_2$.
- Therefore, $\hat{p}_1$ and $\hat{p}_2$ are both estimators of a common value, which we will call $p$.

# The Test Statistic

- By pooling the data, we have a total count of $x_1 + x_2$ out of a combined sample of $n_1 + n_2$, for a pooled proportion of

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

- This leads to a better estimator of the standard deviation of $\hat{p}_1 - \hat{p}_2$.

$$\sigma_{\hat{p}_1 - \hat{p}_2} \approx \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

## Example (Testing hypotheses concerning $p_1 - p_2$)

(3) The test statistic is

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

# Outline

# Calculating the Test Statistic

**Example (Testing hypotheses concerning $p_1 - p_2$)**

(4)
- The sample sizes were both 3500.
- 34% of 3500 is 1190.
- 29% of 3500 is 1015.
- The pooled estimate for $p$ is

$$\hat{p} = \frac{1190 + 1015}{3500 + 3500} = \frac{2205}{7000} = 0.315.$$

- Now we may compute the value of the test statistic.

# The Value of the Test Statistic

## Example (Testing hypotheses concerning $p_1 - p_2$)

(4) First compute

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{(0.315)(0.685)\left(\frac{1}{3500}+\frac{1}{3500}\right)} = 0.00753.$$

# The Value of the Test Statistic

## Example (Testing hypotheses concerning $p_1 - p_2$)

(4) First compute

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{(0.315)(0.685)\left(\frac{1}{3500} + \frac{1}{3500}\right)} = 0.00753.$$

(5) Then compute $z$:

$$z = \frac{0.34 - 0.29}{0.00753} = \frac{0.05}{0.00753} = 6.640.$$

# The Value of the Test Statistic

**Example (Testing hypotheses concerning $p_1 - p_2$)**

(5) Compute the *p*-value:

$$\begin{aligned} p\text{-value} &= \texttt{normalcdf(6.640, E99)} \\ &= 0.1.577 \times 10^{-11}. \end{aligned}$$

(6) Reject $H_0$.

(7) The proportion of likely voters who believe that the country is headed in the right direction decreased from Feb 6 to Mar 26.

# Outline

# TI-83 - Testing Hypotheses Concerning $p_1 - p_2$

## TI-83 Testing Hypotheses Concerning $p_1 - p_2$

- Press `STAT > TESTS > 2-PropZTest...`
- Enter $x_1$
- Enter $n_1$
- Enter $x_2$
- Enter $n_2$
- Choose the correct alternative hypothesis.
- Select `Calculate` and press `ENTER`.

# TI-83 - Testing Hypotheses Concerning $p_1 - p_2$

## TI-83 Testing Hypotheses Concerning $p_1 - p_2$

- A window appears with the following information.
  - The title.
  - The alternative hypothesis.
  - The value of the test statistic $z$.
  - The $p$-value.
  - $\hat{p}_1$.
  - $\hat{p}_2$.
  - The pooled estimate $\hat{p}$.
  - $n_1$.
  - $n_2$.

# Practice

## Practice

- Work the previous example on the TI-83.

# Outline

# Assignment

## Homework

- Read Section 11.5, pages 718 - 724.
- Let's Do It! 11.8, 11.9.
- Exercises 34(omit d,e), 35, page 725.
- Chapter Review 45(e), 46, 48(omit d), 50, 51(omit f), 52, 54 - 56, page 728.