Hypothesis Testing with Two Samples

MAT 152 – STATISTICAL METHODS I

LECTURE 3

INSTRUCTOR: DUSTIN ROTEN

FALL 2020

Previous Example

June believes that 50% of first-time brides in the United States are younger than their grooms. She performs a hypothesis test to determine if the percentage is the same or different from 50%. June samples 100 first-time brides and 53 reply that they are younger than their grooms. For the hypothesis test, she uses a 1% level of significance.

$H_0: p = 0.50$

$H_a: p \neq 0.50$ (Can be above or below, so two-tailed test)

$\alpha = 0.01$

Random variable: P' = percent of first-time brides who are younger than their grooms.

Previous Example (Cont.)

This is a proportion problem so a binomial distribution must be used.

From June’s assumptions, we can build an approximate normal distribution:

$$P' \sim N\left(p, \sqrt{\frac{p \cdot q}{n}}\right) = N\left(0.5, \sqrt{\frac{0.5 \cdot 0.5}{100}}\right) = N(0.5, 0.05)$$

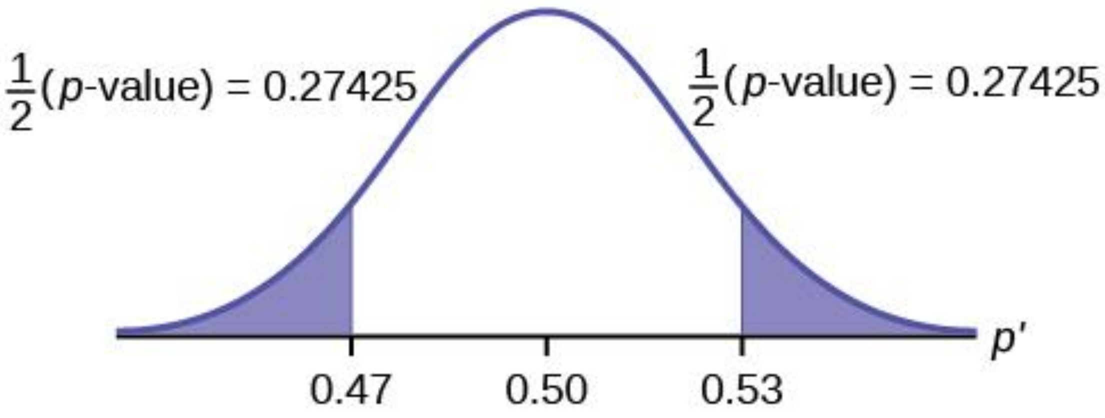
We need to examine the likelihood of obtaining her sample proportion

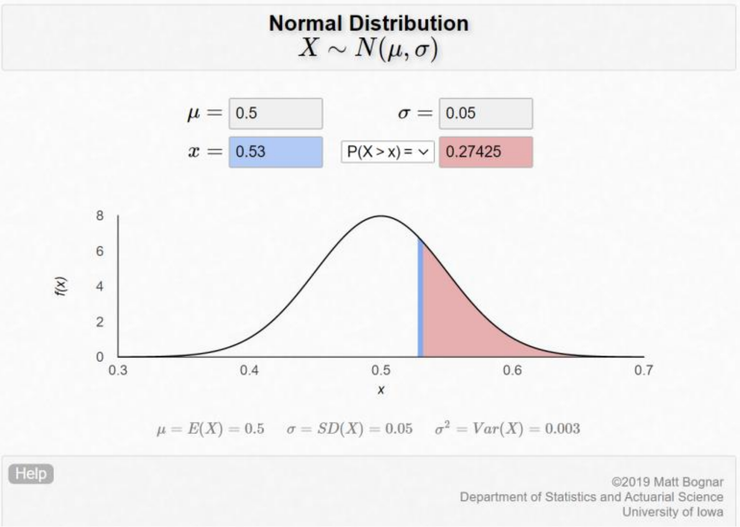
$$p' = \frac{53}{100} = 0.53$$

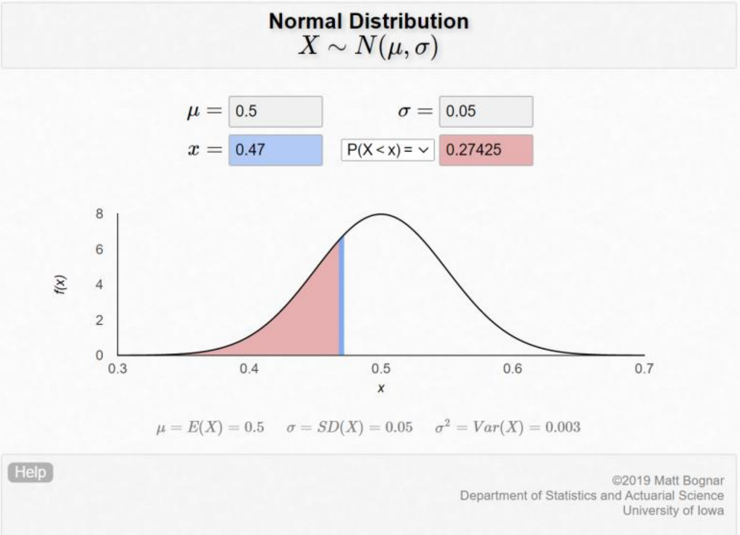
Previous Example (Cont.)

This is a two-tailed test so the probability must be investigated on BOTH sides of the distribution.

$$P(p' < 0.47 \mid p' > 0.53) = P(p' < 0.47) + P(p' > 0.53) = \; 0.5485$$







Previous Example (Cont.)

Interpretation:

If the null hypothesis is true, then there is 0.5485 (54.85%) probability that the sample proportion p' is 0.53 or more OR 0.47 or less.

Decision:

Since $\alpha < p$ (0.01 < 0.5485), we cannot reject the null hypothesis.

Conclusion:

At the 1% level of significance, the sample data do not show sufficient evidence that the percentage of first-time brides who are younger than their grooms is different from 50%.

So, June may be wrong be we don’t have sufficient evidence to prove it.

Comparing Two Independent Population Proportions

Characteristics of a hypothesis test:

1. Two independent simple random samples
2. At least five successes and five failures
3. The population must be 10 to 20 times the size of the sample

The proportions are "pooled":

$$p'_c = \frac{x_1 + x_2}{n_1 + n_2} \text{ and } q'_c = 1 - p'_c$$

Variances are combined to calculate the standard error:

$$SE = \sqrt{(\mathrm{Var}_1 + \mathrm{Var}_2)} = \sqrt{\frac{p'_1 \cdot q'_1}{n_1} + \frac{p'_2 \cdot q'_2}{n_2}} \approx \sqrt{p'_c q'_c \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Example

Two types of medication for hives are being tested to determine if these is a **difference in the proportions** of adult patient reactions. Researchers want to know if Medication A or Medication B is more effective at treating hives within 30 minutes.

- 20 patients out of a random sample of 200 adults were given Medication A and still had hives after 30 minutes.
- 12 patients out of a random sample of 200 adults were given Medication B and still had hives after 30 minutes.

Compare at the 1% significance level.

Random Variable:

$P'_A - P'_B =$ difference in the proportions of adult patients who did not react to the medication after 30 minutes.

Example (Cont.)

Two types of medication for hives are being tested to determine if these is a **difference in the proportions** of adult patient reactions. Researchers want to know if Medication A or Medication B is more effective at treating hives within 30 minutes. Test at $\alpha = 0.01$.

- 20 patients out of a random sample of 200 adults were given Medication A and still had hives after 30 minutes.
- 12 patients out of a random sample of 200 adults were given Medication B and still had hives after 30 minutes.

$H_0: p_A = p_B$ $\qquad$ $H_0: p_A - p_B = 0$ $\qquad$ $H_0: \Delta p = 0$

$H_a: p_A \neq p_B$ $\qquad$ $H_a: p_A - p_B \neq 0$ $\qquad$ $H_a: \Delta p \neq 0$

(The word “difference” implies a two-tailed test)

Since this is a test of two binomial population proportions (success/failure; yes/no), the distribution can be approximated as normal.

Example (Cont.)

Now, all required components are calculated.

Pool the proportions

$$p_c' = \frac{x_A+x_B}{n_A+n_B} = \frac{20+12}{200+200} = 0.08 \text{ and } q_c' = 1 - p_c' = 1 - 0.08 = 0.92$$

Assuming that the null hypothesis is true, what is the approximate normal distribution?

$$\Delta P' \sim N\left[\Delta p', \sqrt{p_c' q_c' \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}\right] = N\left[0, \sqrt{(0.08 \cdot 0.92)\left(\frac{1}{200} + \frac{1}{200}\right)}\right] \approx N[0, 0.02713]$$

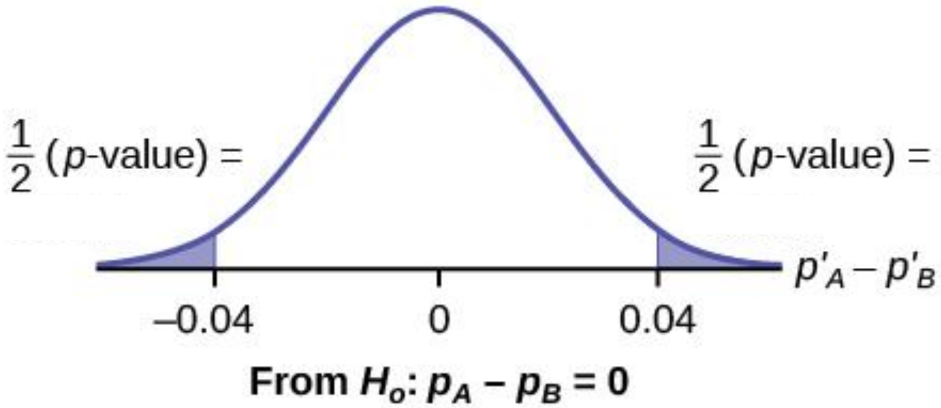
What is the experimental difference between the proportions?

$$\Delta p' = p_A' - p_B' = 0.1 - 0.06 = 0.04$$

Example (Cont.)

Since this is a two-sided test, we want to test the probability (p-value) of the sample difference ($\Delta p'$) being above OR below the assumed difference ($\Delta p = 0$).

It is important to remember that our calculated p-value will be SPLIT between the tails!



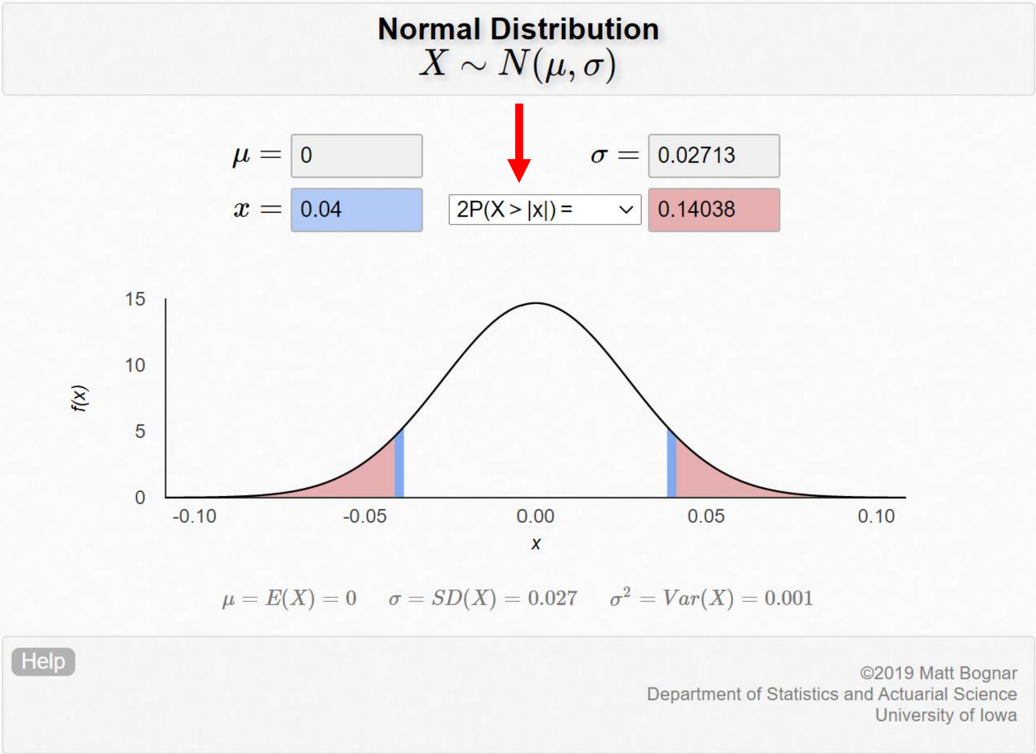
Example (Cont.)

The provided normal distribution calculator will calculate both tails at once IF you select the proper option!

~0.0702 in each tail

Note that $\alpha < \mathrm{p-value}$ where 0.01< 0.14038

Since the p-value isn't smaller than our significance level, our observed difference isn't an extreme event. (We CANNOT reject H_0.)



Example (Cont.)

Two types of medication for hives are being tested to determine if these is a **difference in the proportions** of adult patient reactions. Researchers want to know if Medication A or Medication B is more effective at treating hives within 30 minutes.

- 20 patients out of a random sample of 200 adults were given Medication A and still had hives after 30 minutes.
- 12 patients out of a random sample of 200 adults were given Medication B and still had hives after 30 minutes.

Conclusion:

At a 1% level of significance, from the sample data, there is NOT sufficient evidence to conclude that there is a difference in the proportions of adult patients who did not react after 30 minutes to Medication A and Medication B.

A Quick Review

When testing two-sample population proportions, the process can be summarized as follows.

1. Determine the null and alternative hypotheses.
2. Determine what type of test to use (left-, right-, two-tailed test).
3. Pool the proportions (p_c').
4. Assuming the null hypothesis to be true, construct the normal distribution ($\Delta p = 0$).
5. Calculate the probability of the observed difference in proportions.
6. Draw a conclusion.