MAT111 REVIEW

Ch. 7-8

STATISTICS VS PARAMETERS

Statistics

- Sample Mean: $\bar{x}$
- Sample Standard Deviation: s
- Sample Proportion: $\hat{p}$

- Based on observed data

Parameters

- Population Mean: μ
- Population Standard Deviation: σ
- Population Proportion: p

- Typically unknown

SAMPLING DISTRIBUTION

- Sampling Distribution is the probability distribution of $\hat{p}$
- The standard deviation of a sampling distribution is the Standard Error
  - $SE = \sqrt{\frac{p(1-p)}{n}}$ and $SE_{est} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

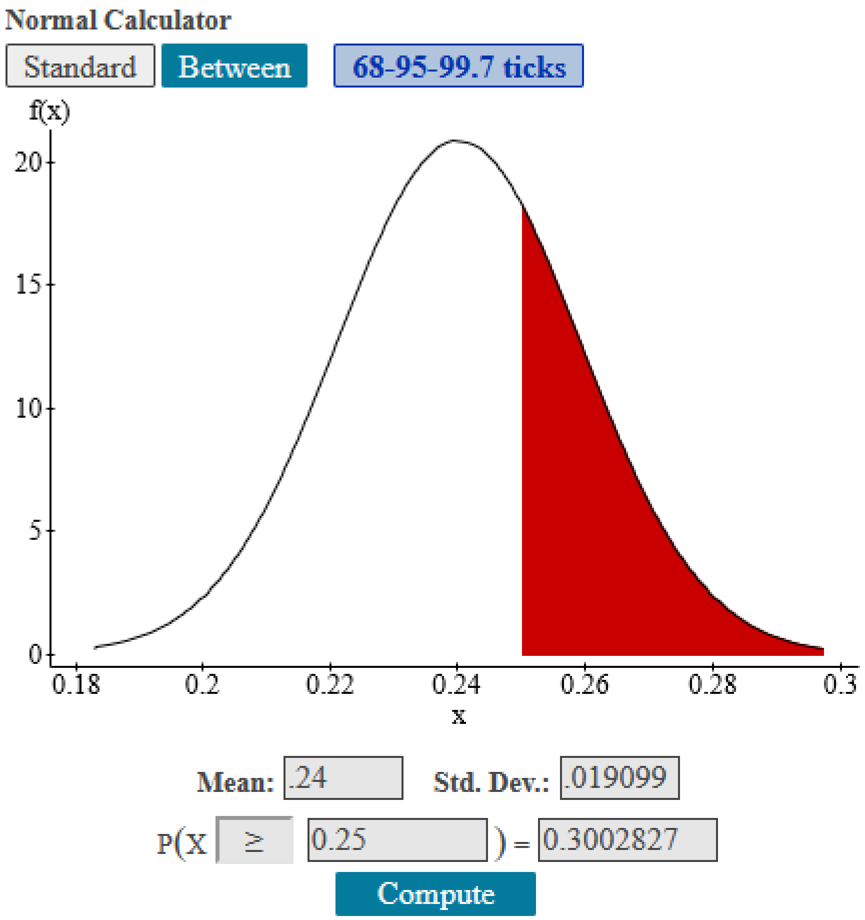
EXAMPLE:

- According to a candy company, packages of a certain candy contain 24% orange candies. Find the approximate probability that the random sample of 500 candies will contain 25% or more orange candies.

Using a normal approximation, what is the probability that at least 25% of 500 randomly sampled candies will be orange?

Find the Standard Error: $SE = \sqrt{\frac{0.24(1-0.24)}{500}} = 0.019099738$

StatCrunch – Stat – Calculators – Normal
Use Standard Error for Standard Deviation

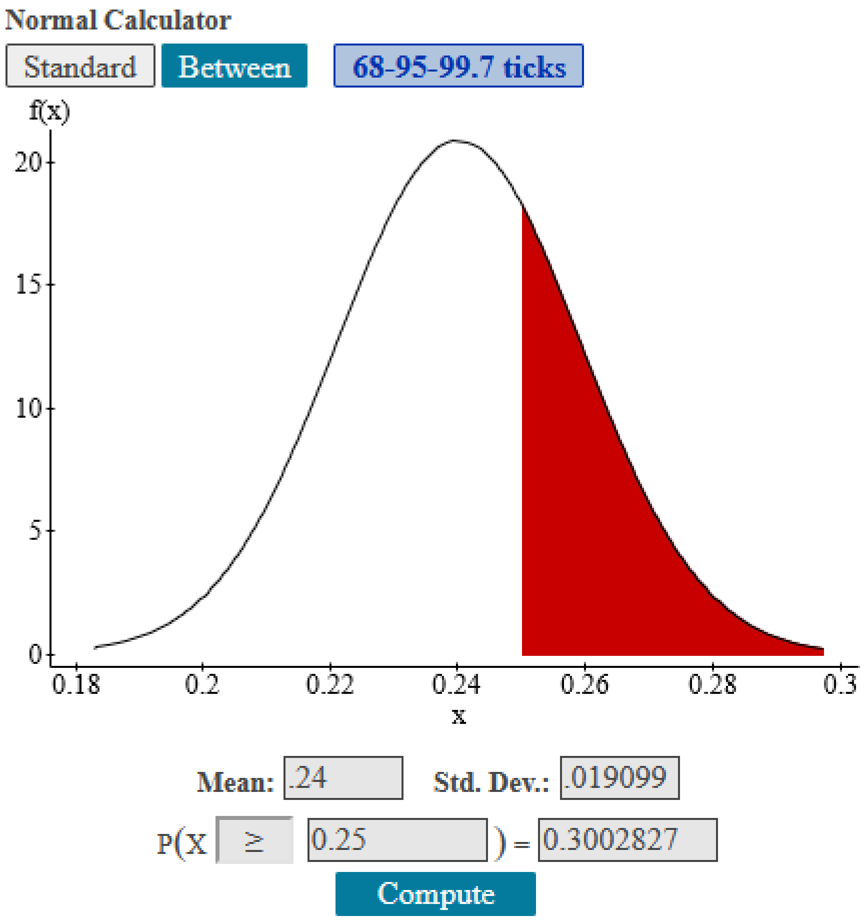


EXAMPLE:

- According to a candy company, packages of a certain candy contain 24% orange candies. Find the approximate probability that the random sample of 500 candies will contain 25% or more orange candies.

Using a normal approximation, what is the probability that at least 25% of 500 randomly sampled candies will be orange?

$$P(\hat{p} \geq 0.25) = 0.300$$



CENTRAL LIMIT THEOREM

CLT assures us that no matter what the shape of the population distribution, if a sample is selected so that the following conditions are met, then the sampling distribution follows an approximately Normal distribution

- Condition 1: Random Sample and Independence
  - Each observation is collected randomly from the population and observations are independent of each other.
- Condition 2: Large sample
  - The sample size, n, is large enough that the sample expects at least 10 successes and 10 failures.
- Condition 3: Big Population
  - The population must be at least 10 times larger than the sample size.

EXAMPLE:

- According to a regional Bar Association, approximately 62% of the people who take the bar exam to practice law in the region pass the exam. Find the approximate probability that at least 66% of 200 randomly sampled people taking the bar exam will pass. Answer the questions below.

The value of the population proportion is ________

EXAMPLE:

- According to a regional Bar Association, approximately 62% of the people who take the bar exam to practice law in the region pass the exam. Find the approximate probability that at least 66% of 200 randomly sampled people taking the bar exam will pass. Answer the questions below.

The value of the population proportion is $\boldsymbol{p = 0.62}$

EXAMPLE:

- According to a regional Bar Association, approximately 62% of the people who take the bar exam to practice law in the region pass the exam. Find the approximate probability that at least 66% of 200 randomly sampled people taking the bar exam will pass. Answer the questions below.

Check the conditions for Central Limit Theorem.

Condition 1: Random and Independent ✓

Condition 2: Large Sample ✓

Condition 3: Big Population ✓

EXAMPLE:

- According to a regional Bar Association, approximately 62% of the people who take the bar exam to practice law in the region pass the exam. Find the approximate probability that at least 66% of 200 randomly sampled people taking the bar exam will pass. Answer the questions below.

Find the Standard Error.

$$SE = \sqrt{\frac{0.62(1-0.62)}{200}} = 0.034322005$$

The Standard Error is 0.034

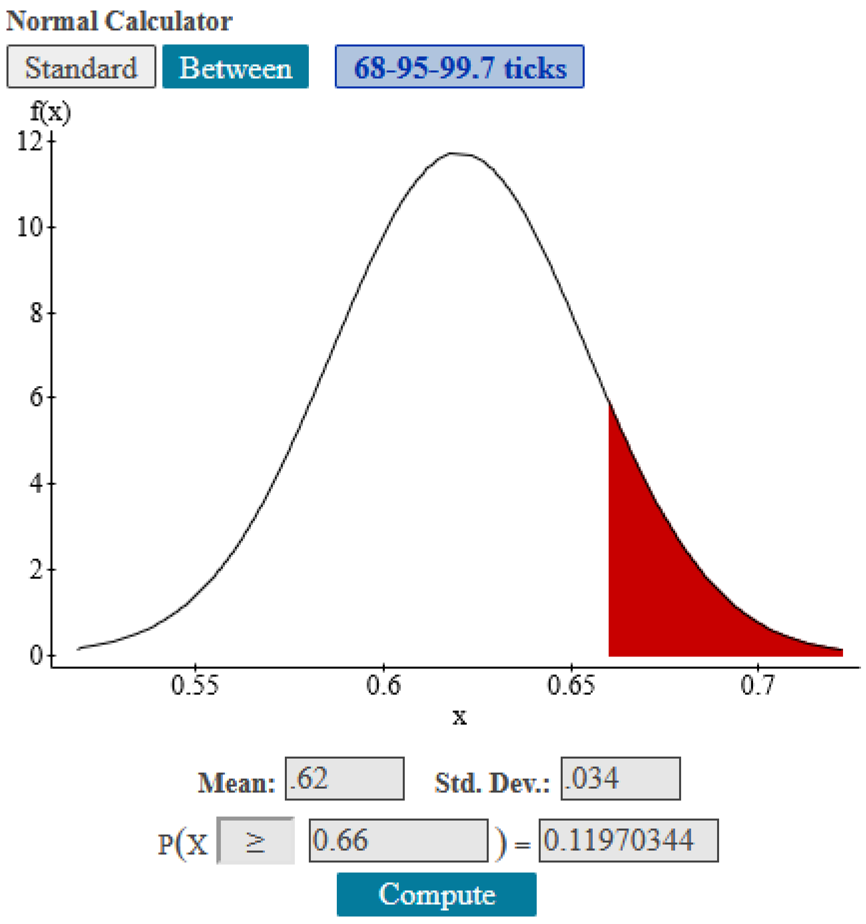
EXAMPLE:

- According to a regional Bar Association, approximately 62% of the people who take the bar exam to practice law in the region pass the exam. Find the approximate probability that at least 66% of 200 randomly sampled people taking the bar exam will pass. Answer the questions below.

Use a normal distribution to find the approximate probability that at least 0.66 pass the exam.

StatCrunch – Stat – Calculator – Normal
Use Standard Error for Standard Deviation

$$P(\hat{p} \geq 0.66) = 0.120$$



CONFIDENCE INTERVALS

- Confidence intervals are found by:

$$estimator \pm \ margin\ of\ error$$

Note: The higher the level of confidence, the wider the confidence interval.

StatCrunch – Stat – Proportion Stats – One or Two Sample

EXAMPLE:

In a simple random sample of 1200 people age 20 and over in a certain country, the population with a certain disease was found to be 0.085 (or 8.5%).

- What is the standard error of the estimate of the proportion of all people in the country age 20 and over with the disease?

$$SE_{est} = \sqrt{\frac{0.085(1-0.085)}{1200}} = 0.008050621$$

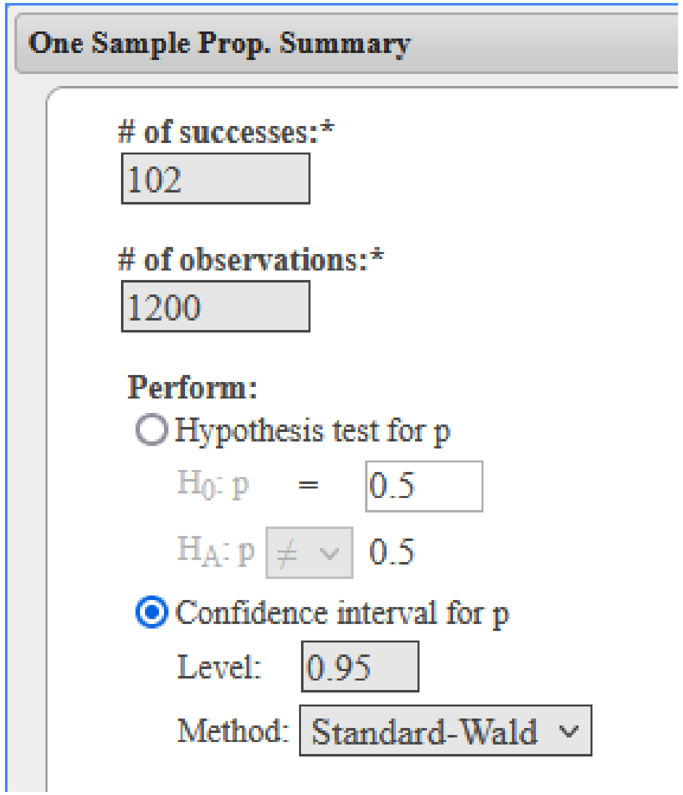
The Standard Error is 0.0081

EXAMPLE:

In a simple random sample of 1200 people age 20 and over in a certain country, the population with a certain disease was found to be 0.085 (or 8.5%).

- Report the 95% confidence interval for the proportion of all people in the country age 20 and over with the disease.

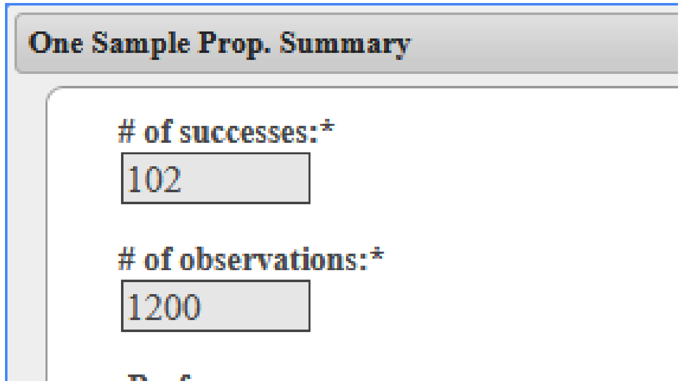
StatCrunch – Stat – Proportion Stats – One Sample
Note: $Successes = np$



EXAMPLE:

In a simple random sample of 1200 people age 20 and over in a certain country, the population with a certain disease was found to be 0.085 (or 8.5%).

- Report the 95% confidence interval for the proportion of all people in the country age 20 and over with the disease.



StatCrunch – Stat – Proportion Stats – One Sample
Note: $Successes = np$

One sample proportion summary confidence interval:
p : Proportion of successes
Method: Standard-Wald

95% confidence interval results:

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	102	1200	0.085	0.0080506211	0.069221073	0.10077893

(0.069, 0.101)

EXAMPLE:

In a simple random sample of 1200 people age 20 and over in a certain country, the population with a certain disease was found to be 0.085 (or 8.5%).

- Find the margin of error, using a 95% confidence level, for estimating this proportion.

Note the Confidence Interval is $(0.069, 0.101)$

To find the margin of error, take ½ of the width of the confidence interval, or use the formula $m = z^{*}SE$

So, margin of error is: $\mathbf{m = 0.016}$

EXAMPLE:

In a simple random sample of 1200 people age 20 and over in a certain country, the population with a certain disease was found to be 0.085 (or 8.5%).

- According to a government agency, nationally, 9.3% of all people in the country age 20 or over have the disease. Does the 95% confidence interval you found support or refute this claim? Explain.

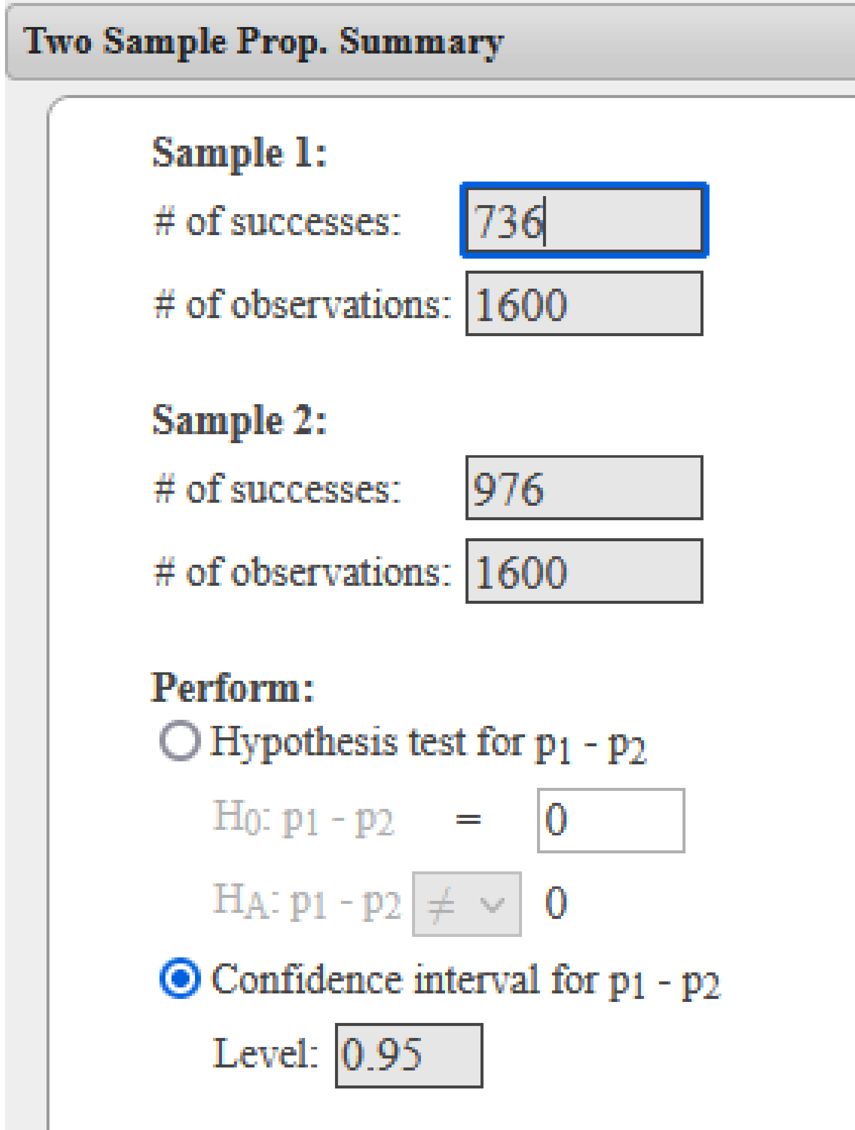
The confidence interval (0.069, 0.101) **supports** this claim, since the value **0.093 is contained** within the interval for the proportion.

EXAMPLE:

In 2002 a Pew Poll based on a random sample of 1600 people suggested that 46% of Americans approved of stem cell research. In 2009 a new poll of a different sample of 1600 people found that 61% approved.

- Find a 95% confidence interval for the difference in proportions approved in the two groups.

StatCrunch – Stat – Proportion Stats – Two Sample
Note: $Successes = np$

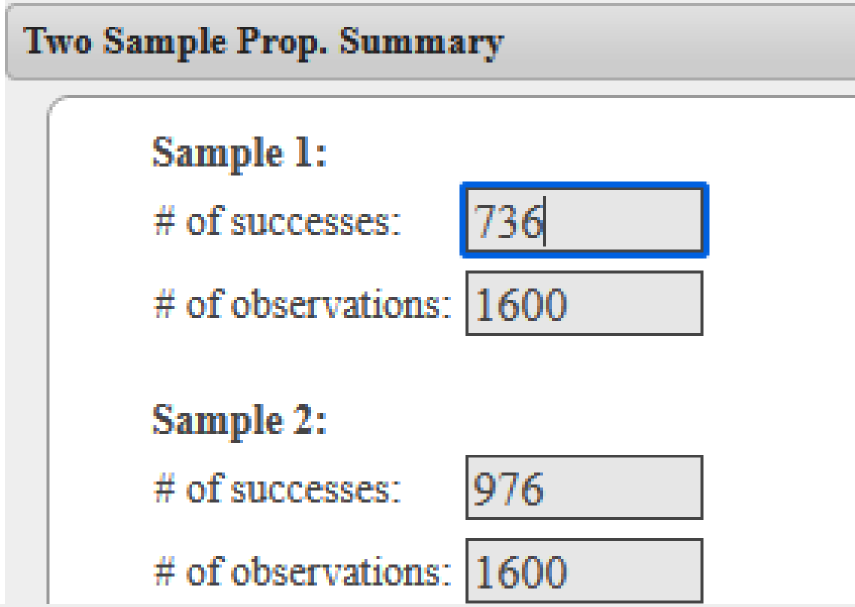


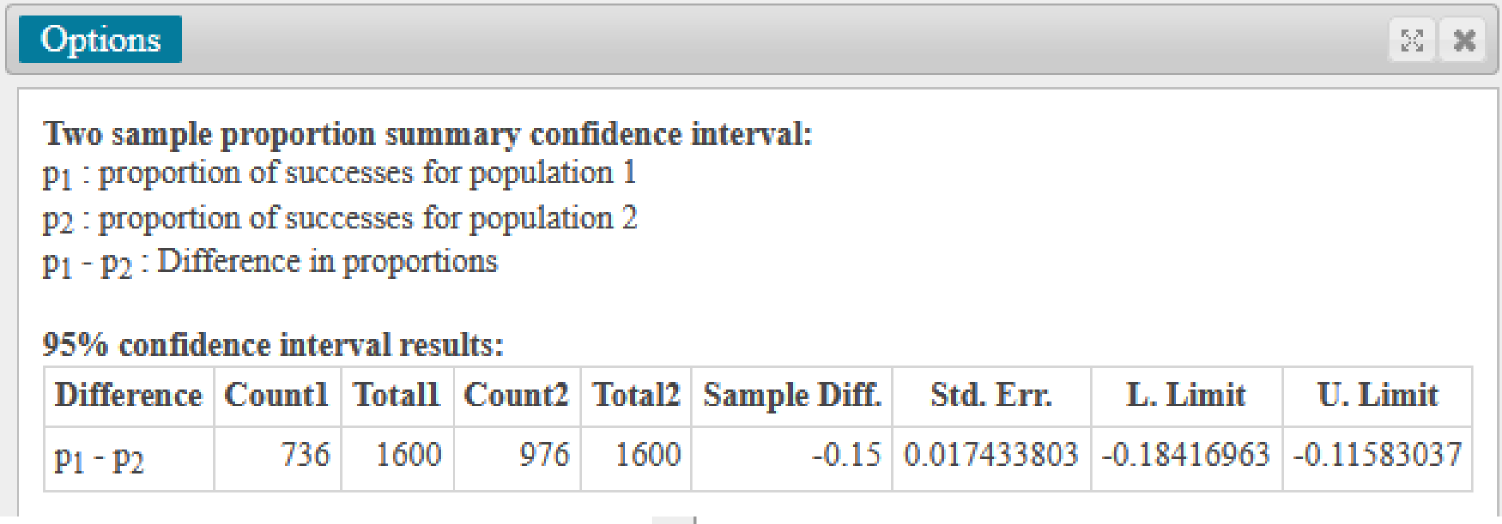
EXAMPLE:

In 2002 a Pew Poll based on a random sample of 1600 people suggested that 46% of Americans approved of stem cell research. In 2009 a new poll of a different sample of 1600 people found that 61% approved.

- Find a 95% confidence interval for the difference in proportions approved in the two groups.

StatCrunch – Stat – Proportion Stats – Two Sample
Note: $Successes = np$




Options

Two sample proportion summary confidence interval:
p_1 : proportion of successes for population 1
p_2 : proportion of successes for population 2
$p_1 - p_2$: Difference in proportions

95% confidence interval results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	L. Limit	U. Limit
$p_1 - p_2$	736	1600	976	1600	-0.15	0.017433803	-0.18416963	-0.11583037

(-0.1842, -0.1158)

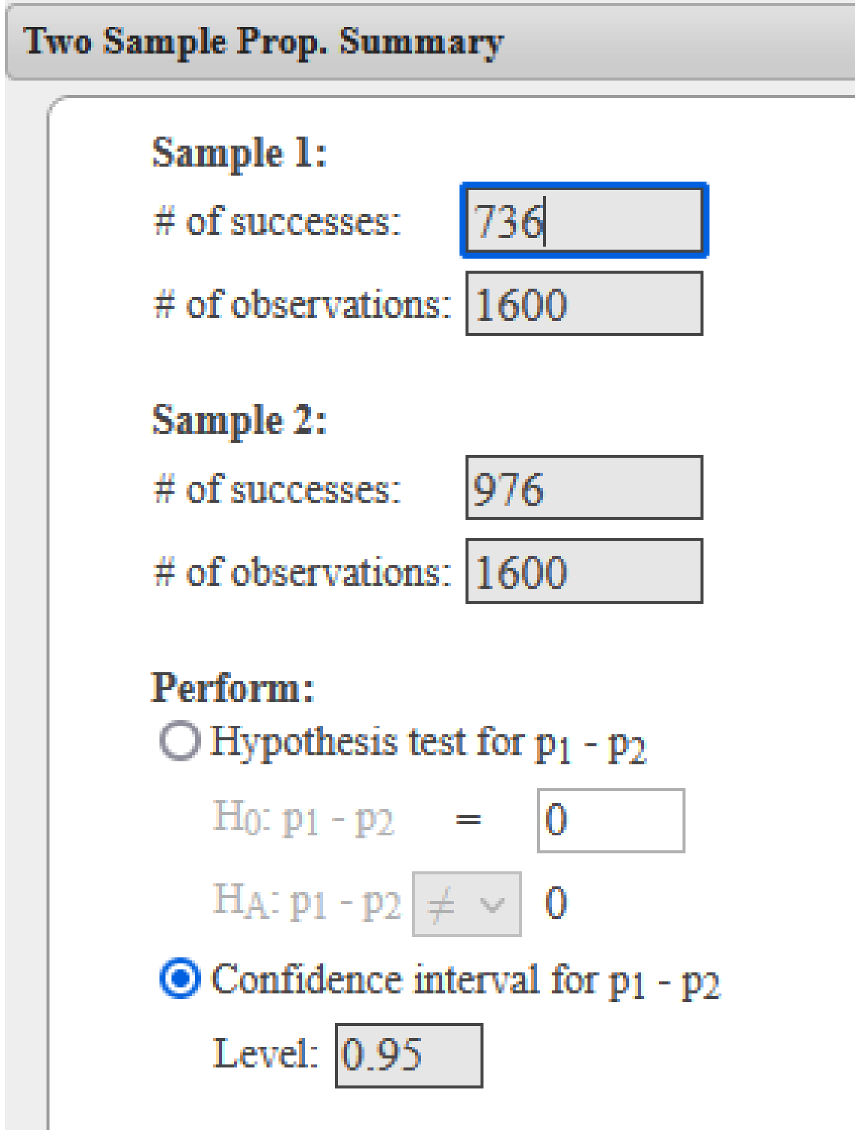
EXAMPLE:

In 2002 a Pew Poll based on a random sample of 1600 people suggested that 46% of Americans approved of stem cell research. In 2009 a new poll of a different sample of 1600 people found that 61% approved.

- Explain whether this interval captures 0 and what this means.

*The interval **does not capture** 0, suggesting that **is not** plausible that **the proportions are the same.***

*Since the boundary values are **both negative**, we can conclude that the proportion of the **second population** (2009 poll) **is larger.***



HYPOTHESIS TESTING

- Step 1: Hypothesize
  - State your hypothesis about the population parameter
- Step 2: Prepare
  - Choose a significance level and test statistic, check conditions and assumptions
- Step 3: Compare and Compute
  - Compute a test statistic and p-value
- Step 4: Interpret
  - Do you reject the null hypothesis or not? What does this mean?
  - If $p < \alpha$, then we reject the Null Hypothesis.

StatCrunch – Stat – Proportion Stats –

Important Formulas:

$$z = \frac{\hat{p} - p_0}{SE}$$

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

EXAMPLE: HYPOTHESES

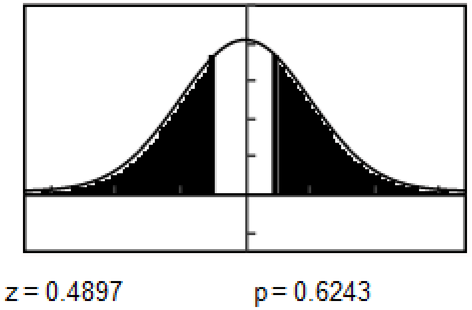
Hypotheses are always statements about which of the following?

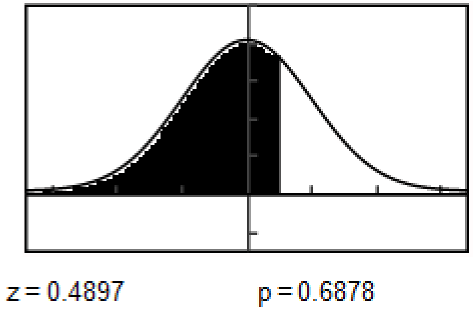
a. Sample size
b. Population parameters
c. Estimators
d. Sample statistics

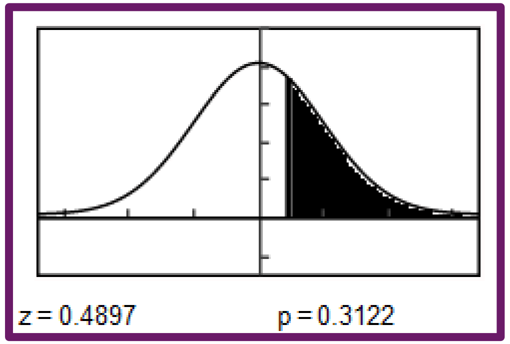
EXAMPLE:

When asked whether marriage is becoming obsolete, 795 out of 2011 randomly selected adults answering a particular survey said yes. We are testing the hypothesis that the population proportion that believes marriage is becoming obsolete is more than 39% using a significance level of 0.05.

- One of the following figures is correct. Indicate which graph matches the alternative hypothesis, $p>0.39$. Report and interpret the correct p-value.







Because the p-value is greater than the significance level, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the population proportion is greater than 0.39

EXAMPLE: HYPOTHESIS TESTING

A vaccine to prevent a severe virus was given to children within the first year of life as part of a drug study. The study reported that of the 3223 children randomly assigned the vaccine, 46 got the virus. Of the 1694 children randomly assigned the placebo, 41 got the virus.

- Find the sample percentage of children who caught the virus in each group. Is the sample percentage lower for the vaccine group, as investigators hoped?

Percentage of children who caught the virus in the vaccine group: 1.43%

Percentage of children who caught the virus in the placebo group: 2.42%

Yes! The sample percentage is lower for the vaccine group.

EXAMPLE: HYPOTHESIS TESTING

A vaccine to prevent a severe virus was given to children within the first year of life as part of a drug study. The study reported that of the 3223 children randomly assigned the vaccine, 46 got the virus. Of the 1694 children randomly assigned the placebo, 41 got the virus.

- Determine whether the vaccine is effective in reducing the chance of catching the virus, using a significance level of 0.10.
  - Step 1: State the null and alternate hypotheses
    Note: We are testing if Sample 1(Vaccine) < Sample 2(Placebo)

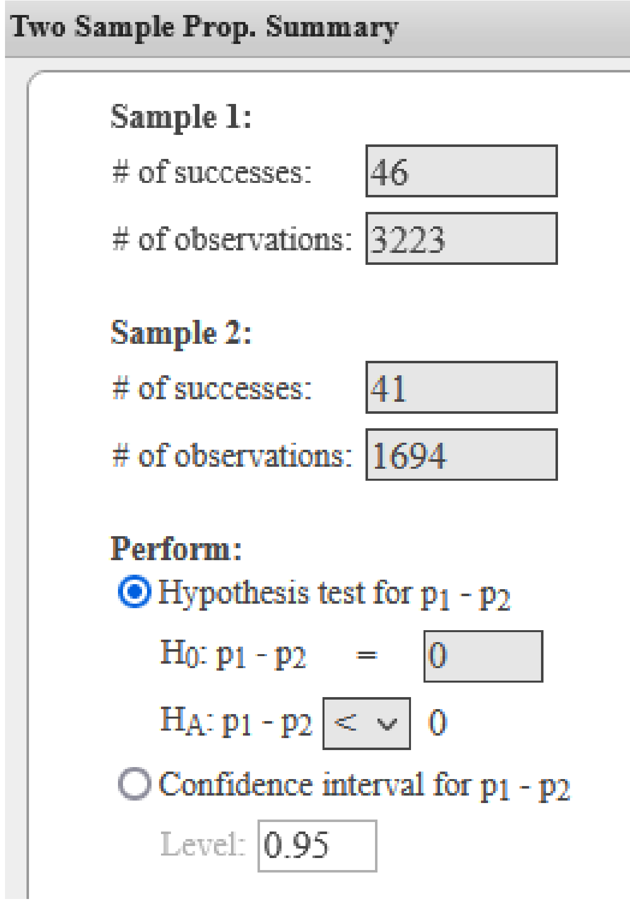
$$\boldsymbol{H_0: p_1 - p_2 = 0}$$
$$\boldsymbol{H_a: p_1 - p_2 < 0}$$

EXAMPLE: HYPOTHESIS TESTING

A vaccine to prevent a severe virus was given to children within the first year of life as part of a drug study. The study reported that of the 3223 children randomly assigned the vaccine, 46 got the virus. Of the 1694 children randomly assigned the placebo, 41 got the virus.

- Determine whether the vaccine is effective in reducing the chance of catching the virus, using a significance level of 0.10.
  - Step 2: Find the test statistic and the p-value.

StatCrunch – Stats – Proportion Stats – Two Sample



EXAMPLE: HYPOTHESIS TESTING

A vaccine to prevent a severe virus was given to children within the first year of life as part of a drug study. The study reported that of the 3223 children randomly assigned the vaccine, 46 got the virus. Of the 1694 children randomly assigned the placebo, 41 got the virus.

- Determine whether the vaccine is effective in reducing the chance of catching the virus, using a significance level of 0.10.
  - Step 2: Find the test statistic and the p-value.

Two sample proportion summary hypothesis test:

p_1 : proportion of successes for population 1

p_2 : proportion of successes for population 2

$p_1 - p_2$: Difference in proportions

$H_0 : p_1 - p_2 = 0$

$H_A : p_1 - p_2 < 0$

Hypothesis test results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	46	3223	41	1694	-0.0099306527	0.0039563639	-2.5100453	0.006

$$t = -2.51\,, \qquad p = 0.006$$

EXAMPLE: HYPOTHESIS TESTING

A vaccine to prevent a severe virus was given to children within the first year of life as part of a drug study. The study reported that of the 3223 children randomly assigned the vaccine, 46 got the virus. Of the 1694 children randomly assigned the placebo, 41 got the virus.

- What is the conclusion for this test?

$0.006 < 0.10$ so we reject H_0.
There is sufficient evidence to conclude that the vaccine is effective in reducing the chance of catching the virus at the significance level of 0.10.

EXAMPLE: HYPOTHESIS TESTING

Historically, the percentage of residents of a certain country who support stricter gun control laws has been 53%. A recent poll of 932 people showed 505 in favor of stricter gun control laws. Assume the poll was given to a random sample of people. Test the claim that the proportion of those favoring stricter gun control has changed. Perform a hypothesis test, using a significance level of 0.05.

- State the null and alternate hypotheses.

$$\boldsymbol{H_0: p = 0.53}$$
$$\boldsymbol{H_a: p \neq 0.53}$$

EXAMPLE: HYPOTHESIS TESTING

Historically, the percentage of residents of a certain country who support stricter gun control laws has been 53%. A recent poll of 932 people showed 505 in favor of stricter gun control laws. Assume the poll was given to a random sample of people. Test the claim that the proportion of those favoring stricter gun control has changed. Perform a hypothesis test, using a significance level of 0.05.

- Compute the sample proportion $\hat{p}$.

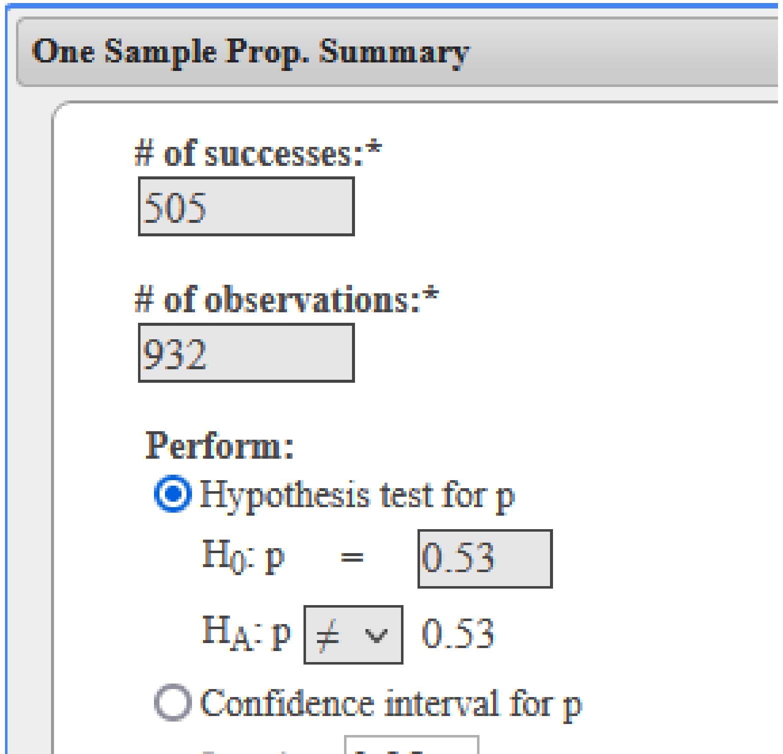
Note: The sample proportion is what we observed. (505 out of the 932)

$$\boldsymbol{\hat{p} = 0.5418}$$

EXAMPLE: HYPOTHESIS TESTING

Historically, the percentage of residents of a certain country who support stricter gun control laws has been 53%. A recent poll of 932 people showed 505 in favor of stricter gun control laws. Assume the poll was given to a random sample of people. Test the claim that the proportion of those favoring stricter gun control has changed. Perform a hypothesis test, using a significance level of 0.05.

- Compute the test statistic and p-value.



EXAMPLE: HYPOTHESIS TESTING

Historically, the percentage of residents of a certain country who support stricter gun control laws has been 53%. A recent poll of 932 people showed 505 in favor of stricter gun control laws. Assume the poll was given to a random sample of people. Test the claim that the proportion of those favoring stricter gun control has changed. Perform a hypothesis test, using a significance level of 0.05.

- Compute the test statistic and p-value.

$$z = 0.72$$
$$p - value = 0.469$$

One Sample Prop. Summary

Options

One sample proportion summary hypothesis test:
p : Proportion of successes
$H_0 : p = 0.53$
$H_A : p \neq 0.53$

Hypothesis test results:

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	505	932	0.54184549	0.016348537	0.72455983	0.4687

EXAMPLE: HYPOTHESIS TESTING

Historically, the percentage of residents of a certain country who support stricter gun control laws has been 53%. A recent poll of 932 people showed 505 in favor of stricter gun control laws. Assume the poll was given to a random sample of people. Test the claim that the proportion of those favoring stricter gun control has changed. Perform a hypothesis test, using a significance level of 0.05.

- What is the conclusion for this test?

$0.469 \nless 0.05$ Do not reject H_0. The percentage is not significantly different from 53%.

CENTER FOR ACADEMIC SUPPORT

Hearnes 213
816-271-4524