

# Inference for multiple proportions

Stat 120

May 11 2022

## Tests for Categorical Variable(s)

### Goodness-of-fit test:

- Test a claim about the distribution of one categorical variable
- E.g. are 6 M&M colors equally likely?
- E.g. is Biden's approval rating 50%?

## Tests for Categorical Variable(s)

### Chi-square test for association:

- Determine if a relationship between two categorical variables is statistically significant
- E.g. Does M&M color distribution depend on type (chocolate vs. peanut)?

## Test for one categorical variable

Seen single proportion tests before

- Example : Test if the proportion of Reese's Pieces that are orange is different from 1/3.

$$H_0 : p = 1/3$$

$$H_a : p \neq 1/3$$

## Test for one categorical variable (different proportions)

What if we want to test proportions for several categories at once?

- Example: Are the three colors (orange, yellow, brown) of Reese's Pieces equally likely?
- $H_0$  specifies a proportion,  $p_i$ , for each category.

## Test for one categorical variable (different proportions)

The proportions do not have to be the same. e.g: Grade distribution

$$H_0 : p_A = 0.2 \quad p_B = 0.3 \quad p_C = 0.3 \quad p_D = 0.1 \quad p_F = 0.1$$

$$H_a : \text{at least one proportion different}$$

# Rock-Paper-Scissors

ROCK	PAPER	SCISSORS	TOTAL
36	12	37	85

How would we test whether all of these categories are equally likely?

Conduct a hypothesis test

- State Hypothesis
- Calculate a test statistic, based on your sample data
- Create a distribution of this statistic, as it would be observed if the null hypothesis were true
- Measure how extreme your test statistic is, as compared to the distribution generated under null

## Test Statistic

Why can't we use the familiar formula to get the test statistic?

$$\frac{\text{sample statistic} - \text{null value}}{\text{SE}}$$

- More than one sample statistic
- More than one null value

We need something a bit more complicated ...



# Observed Counts

- The observed counts are the actual counts observed in the study

ROCK	PAPER	SCISSORS	TOTAL
36	12	37	85

- The expected counts are the expected counts if the null hypothesis were true
- For each cell, the expected count is the sample size ( $n$ ) times the null proportion,  $p_i$

	ROCK	PAPER	SCISSORS	TOTAL
Observed	36	12	37	85
Expected	28.33	28.33	28.33	85

## Chi-Square Statistic

- A test statistic is one number, computed from the data, which we can use to assess the null hypothesis
- The chi-square statistic is a test statistic for categorical variables:

$$\chi^2 = \sum \frac{(\textit{observed} - \textit{expected})^2}{\textit{expected}}$$

## Rock-Paper-Scissors

	ROCK	PAPER	SCISSORS	TOTAL
Observed	36	12	37	85
Expected	28.33	28.33	28.33	85

$$\chi^2 = \dots$$
$$\approx \dots$$

# What next?

We have a test statistic. What else do we need to perform the hypothesis test?

- A distribution of the test statistic assuming  $H_0$  is true

How do we get this? Two options:

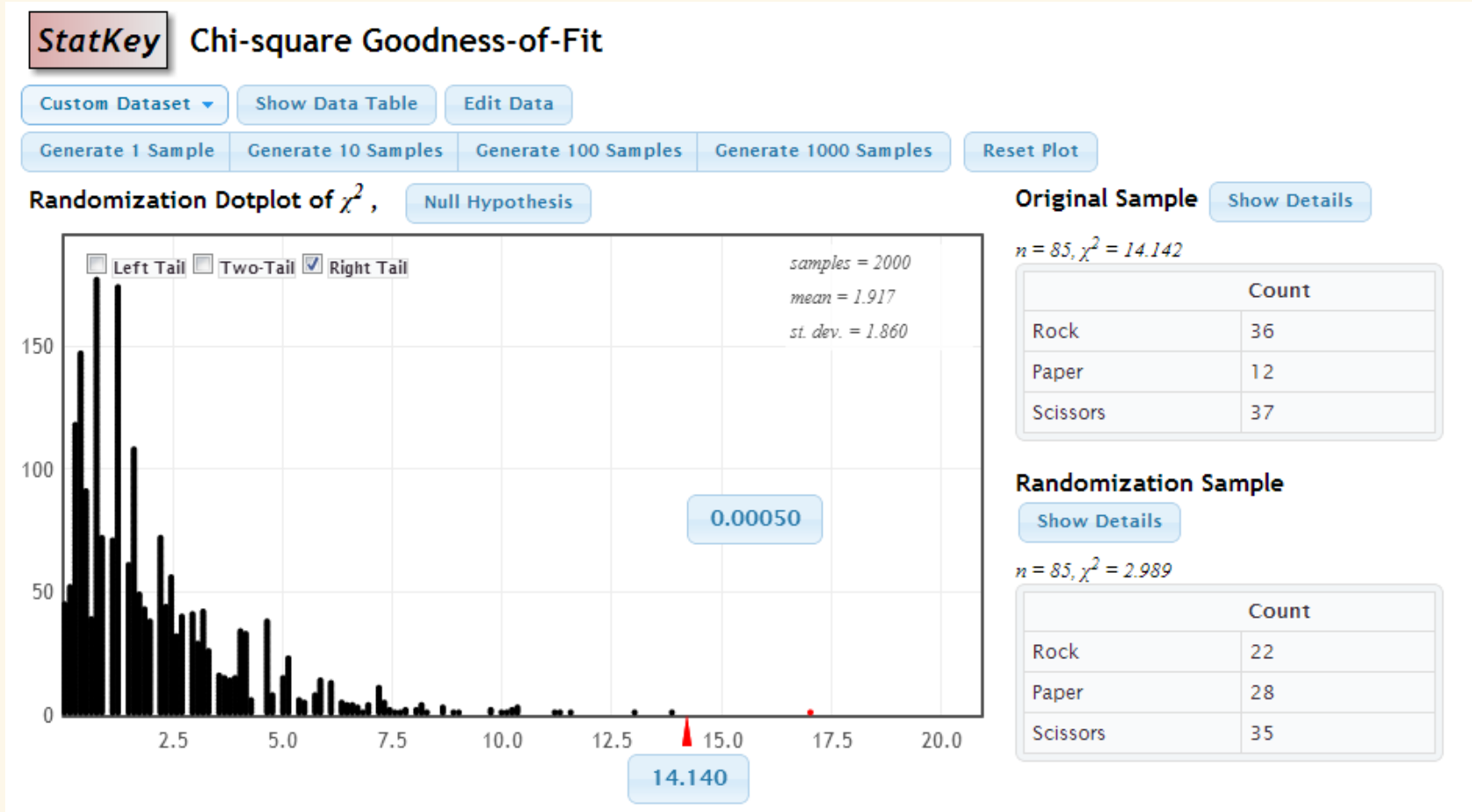
1. Simulation
2. Theoretical Distribution

# Simulation

1. Take 3 scraps of paper and label them Rock, Paper, Scissors. Fold or crumple them so they are indistinguishable. Choose one at random and record the result.
2. Repeat a number of times to match the original sample size and get a table of observed counts.
3. Calculate the  $\chi^2$ -statistic.
4. Repeat this many times to get a randomization distribution of many  $\chi^2$ -statistics.

How extreme is the actual test statistic in this randomization distribution?

# Chi-Square Distribution



# Chi-Square Distribution

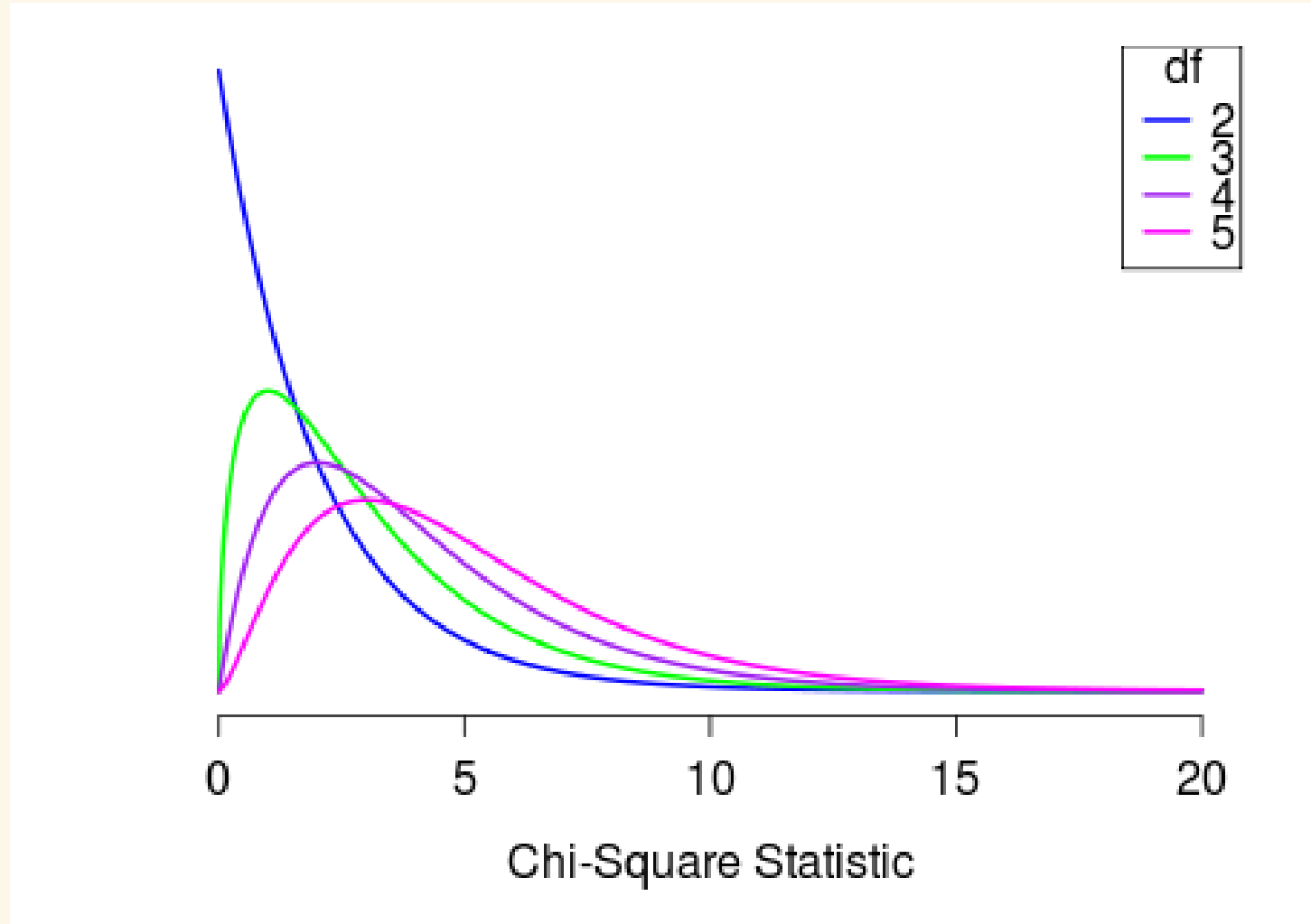
- If each of the expected counts are at least 5, AND if the null hypothesis is true, then the  $\chi^2$  statistic follows a  $\chi^2$  distribution, with degrees of freedom equal to

$$df = \text{number of categories} - 1$$

- Rock-Paper-Scissors:

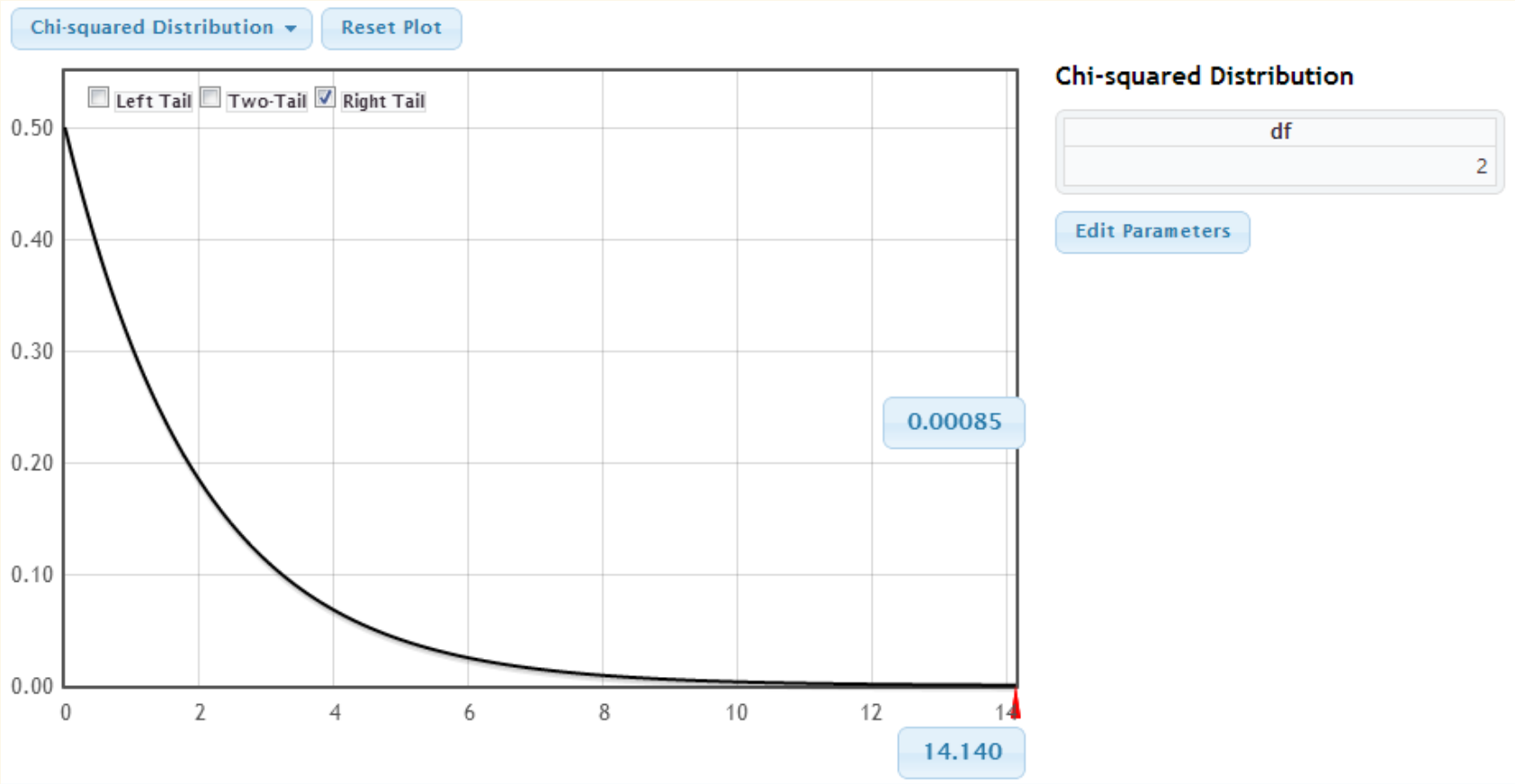
$$df = 3 - 1 = 2$$

# Chi-Square Distribution





# p-value using Chi-square distribution



## Goodness of Fit

- A  $\chi^2$  test for goodness of fit test determines whether the distribution of a categorical variable is the same as some null hypothesized distribution
- The null hypothesized proportions for each category do not have to be the same

# Chi-Square Test for Goodness of Fit

- State null hypothesized proportions for each category,  $p_i$ . Alternative is that at least one of the proportions is different than specified in the null.
- Calculate the expected counts for each cell as  $n \cdot p_i$ . Make sure they are all greater than 5 to proceed.
- Calculate the  $\chi^2$  statistic:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

- Compute the p-value as the area in the tail above the  $\chi^2$  statistic, for a  $\chi^2$  distribution with  $df = (\text{number of categories} - 1)$ .
- Interpret the p-value in context.

## Mendel's Pea Experiment

Between 1856 and 1863 Gregor Mendel cultivated and tested roughly 29,000 pea plants. This study showed that one in four pea plants was purebred recessive, two out of four were hybrid and one out of four were purebred dominant.

















Mendel's work was rejected at first and was not widely accepted until after he died. He is now known as the "father of modern genetics"

# Mendel's Pea Experiment



$Y$  = yellow seed  
 $y$  = green seed  
 $S$  = round shape  
 $s$  = wrinkly shape

↓  
 $F_2$

 $SSYY$	 $SSYy$	 $SsYY$	 $SsYy$
 $SSyY$	 $SSyy$	 $SsyY$	 $Ssyy$
 $sSYy$	 $sSYy$	 $ssYY$	 $ssYy$
 $sSyY$	 $sSyy$	 $ssyY$	 $ssyy$

$S$ ,  $Y$ : Dominant and  $s$ ,  $y$ : Recessive

Mate SSYY with ssyy → 1st Generation:  
all  $Ss Yy$

Mate 1st Generation: → 2nd Generation

Phenotype	Theoretical Proportion
Round, Yellow	9/16
Round, Green	3/16
Wrinkled, Yellow	3/16
Wrinkled, Green	1/16

# Mendel's Pea Experiment

Phenotype	Theoretical Proportion	Observed Counts
Round, Yellow	9/16	315
Round, Green	3/16	101
Wrinkled, Yellow	3/16	108
Wrinkled, Green	1/16	32

Let's test this data against the null hypothesis of each  $p_i$  equal to the theoretical value, based on genetics

$$H_0 : p_1 = 9/16, p_2 = 3/16, p_3 = 3/16, p_4 = 1/16$$

$$H_a : \text{At least one } p_i \text{ is not as specified in } H_0$$

## Mendel's Pea Experiment

What is the expected count for round green peas?

- A. 0.182
- B. 108
- C. 104.25
- D. 139

Phenotype	Theoretical Proportion	Observed Counts
Round, Yellow	9/16	315
Round, Green	3/16	101
Wrinkled, Yellow	3/16	108
Wrinkled, Green	1/16	32
Total	1	556

# Mendel's Pea Experiment

Phenotype	Theoretical Proportion	Observed Counts
Round, Yellow	9/16	315
Round, Green	3/16	101
Wrinkled, Yellow	3/16	108
Wrinkled, Green	1/16	32
Total	1	556

The  $\chi^2$  statistic is the sum of 4 components (because there are 4 categories). What is the contribution to the  $\chi^2$  statistic from the “Round, Yellow” category?

1. 0.016
2. 1.05
3. 5.21
4. 107.2

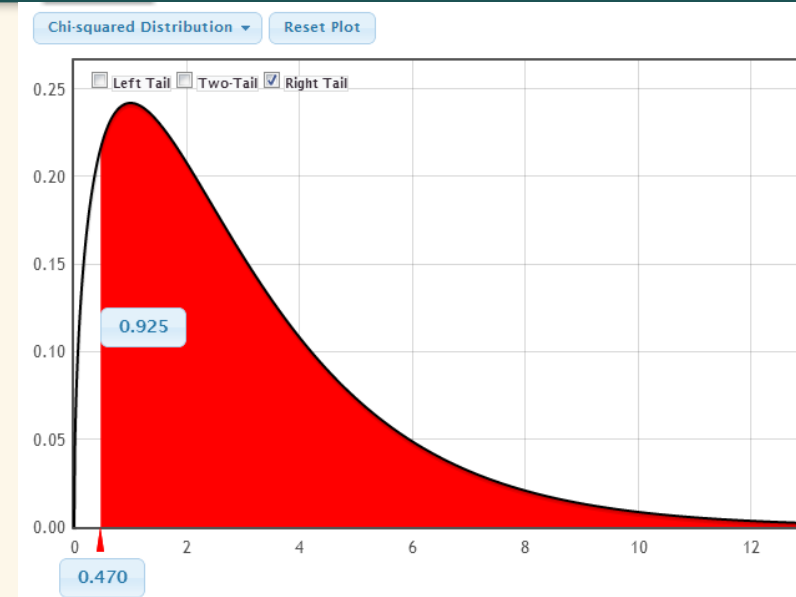
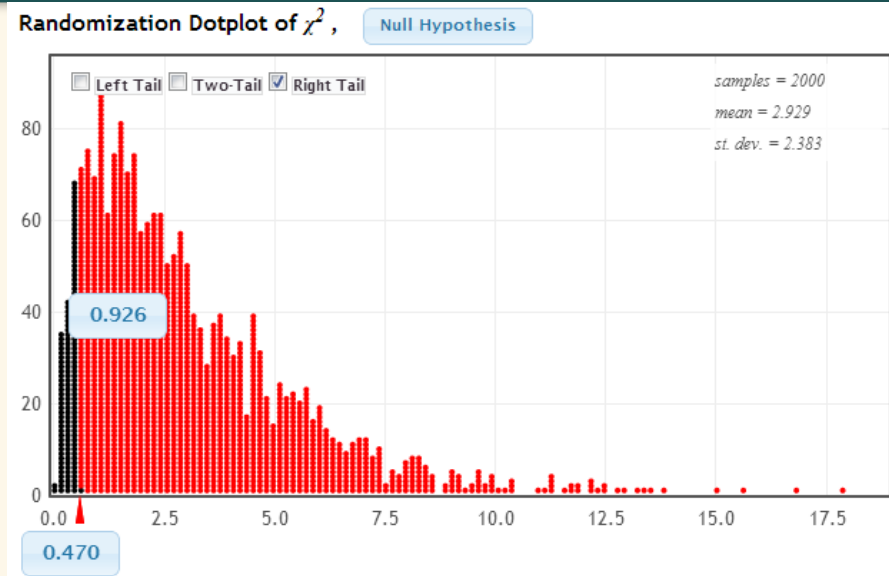


# Mendel's Pea Experiment

$$\chi^2 = 0.47$$

Two options:

- Simulate a randomization distribution
- Compare to a  $\chi^2$  distribution with 4-1



## Mendel's Pea Experiment

Does this prove Mendel's theory of genetics? Or at least prove that his theoretical proportions for pea phenotypes were correct? (p-value = 0.925 )

1. Yes
2. No

# Your Turn 1

05:00



- Go over to the in class activity file
- Complete the activity in your group