

Hypothesis Tests and Confidence Intervals using Normal Distribution!

Stat 120

May 02 2023

How do Malaria parasites impact mosquito behavior?



Experiment Design

- **Mosquitoes are exposed to two groups of mice:**
 - **Malaria-infected mice (experimental group)**
 - **Healthy mice (control group)**
- **Malaria-infected mice go through two stages of infection:**
 - **Stage 1: Not yet infectious (Days 1-8)**
 - **Stage 2: Infectious (Days 9-28)**
- **Response Variable: Whether the mosquito approaches a human.**

Research Questions: Does mosquito behavior differ when exposed to malaria-infected mice compared to healthy mice? Does mosquito behavior differ based on the infection stage of the malaria-infected mice?

Malaria Parasites and Mosquitoes

Malaria parasites would benefit if

- *Mosquitoes approached humans less often after being exposed, but before becoming infectious, because humans are risky*
- *Mosquitoes approached humans more often after becoming infectious, to pass on the infection*



Days 1-8

We'll first look at the mosquitoes before they become infectious (days 1-8).

p_C : **proportion of controls to approach human**

p_E : **proportion of exposed to approach human**

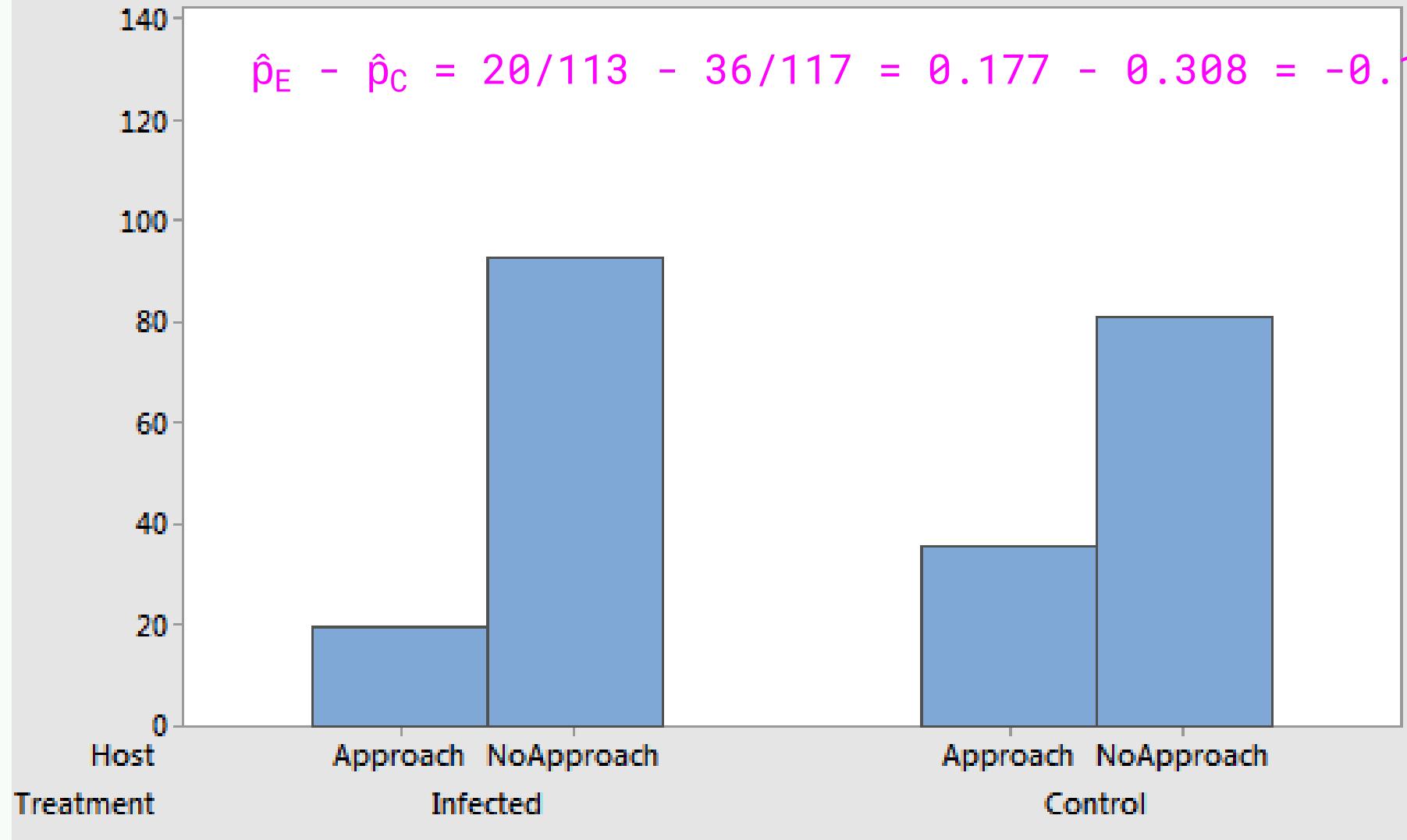
What are the relevant hypotheses?

- A. $H_0 : p_E = p_C, H_a : p_E < p_C$
- B. $H_0 : p_E = p_C, H_a : p_E > p_C$
- C. $H_0 : p_E < p_C, H_a : p_E = p_C$
- D. $H_0 : p_E > p_C, H_a : p_E = p_C$

► Click for answer

Stage = oocyst

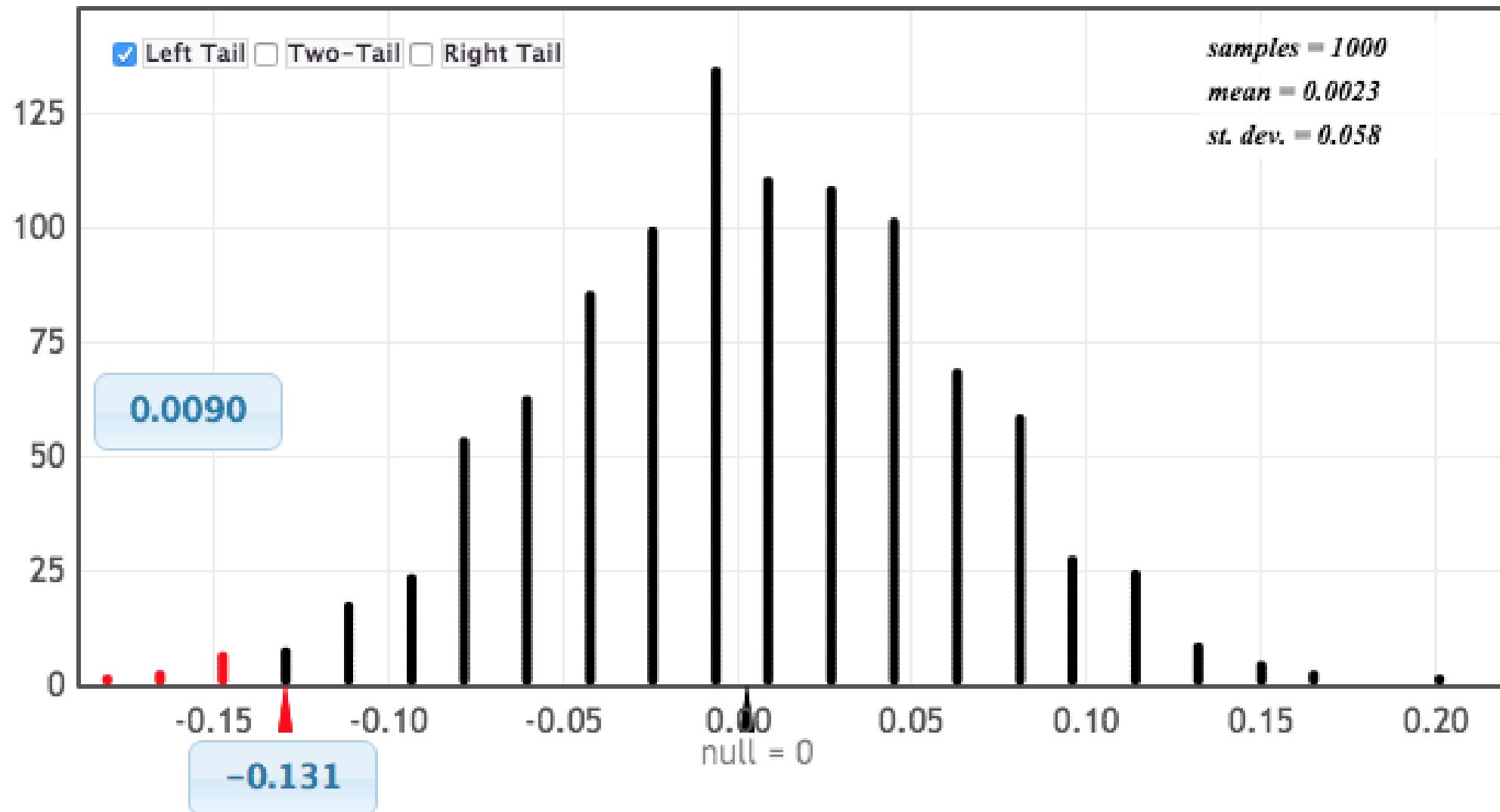
$$\hat{p}_E - \hat{p}_C = 20/113 - 36/117 = 0.177 - 0.308 = -0.131$$

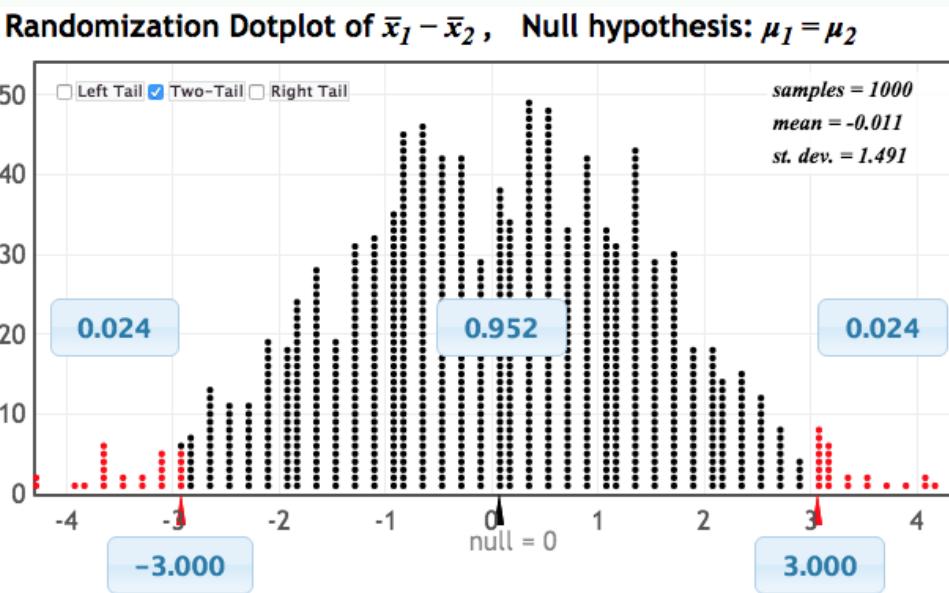
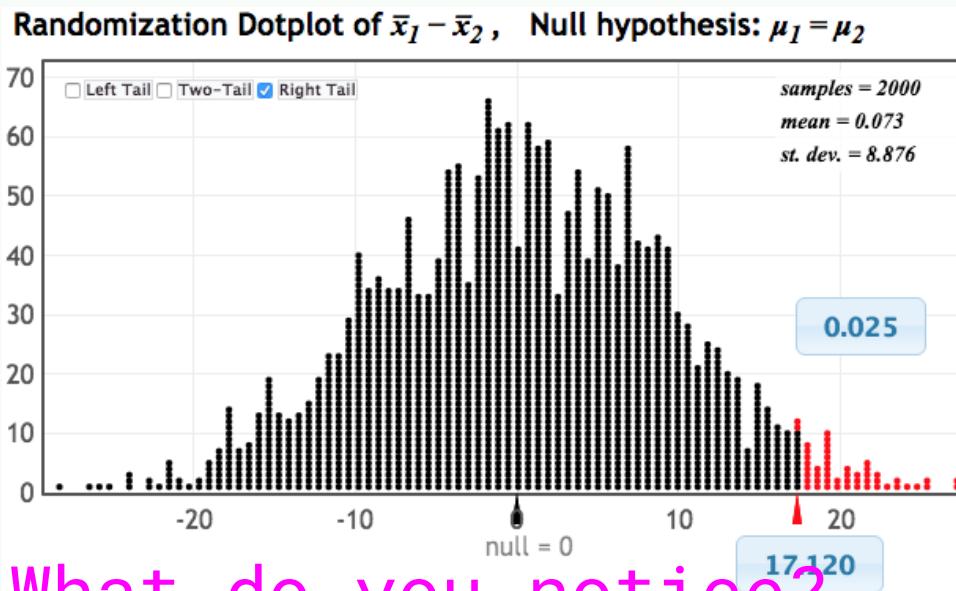
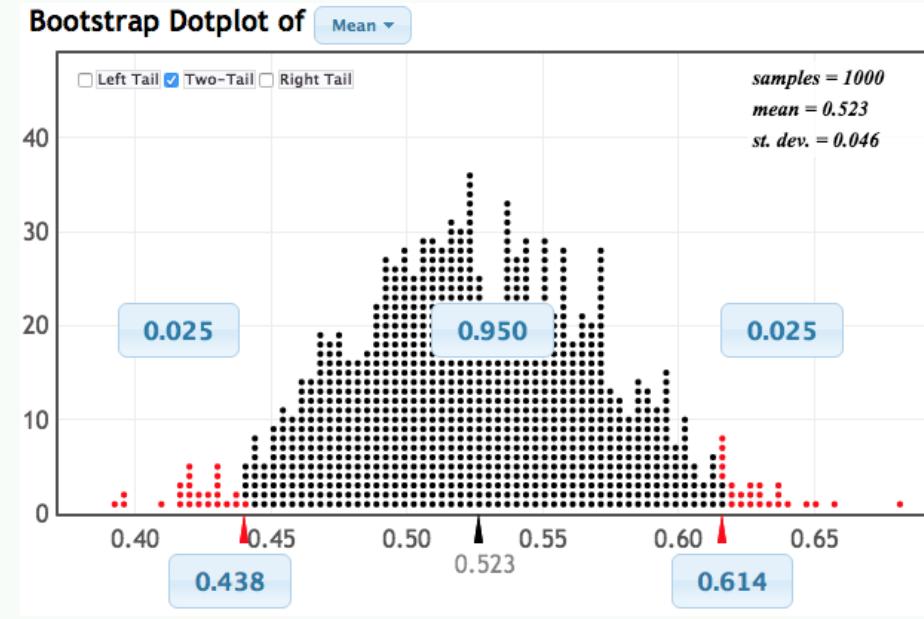
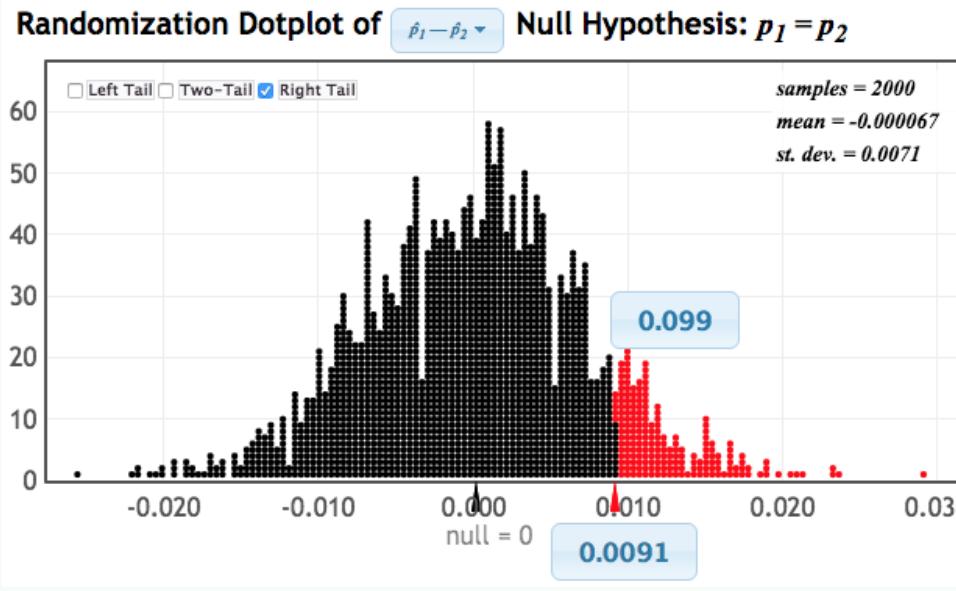


Randomization Dotplot of

$\hat{p}_1 - \hat{p}_2$ ▾

Null Hypothesis: $p_1 = p_2$





What do you notice?

Central Limit Theorem (CLT)

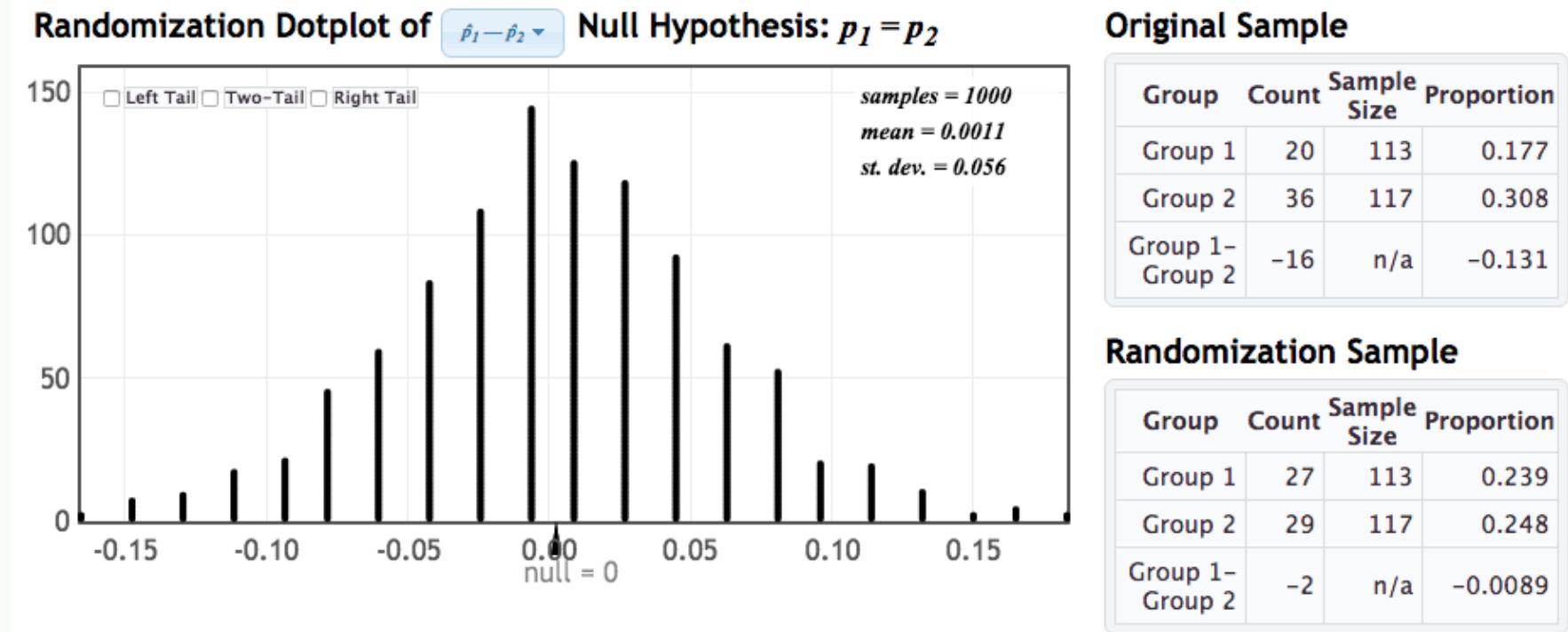
*For random samples with a sufficiently large sample size, the distribution of sample statistics for a **mean** or a **proportion** is normally distributed*

The catch: "sufficiently large sample size"

The **more skewed** the original distribution of data/population is, the larger n has to be for the CLT to work

- For quantitative variables that are not very skewed, $n \geq 30$ is usually sufficient
- For categorical variables, counts of **at least 10** within each category is usually sufficient

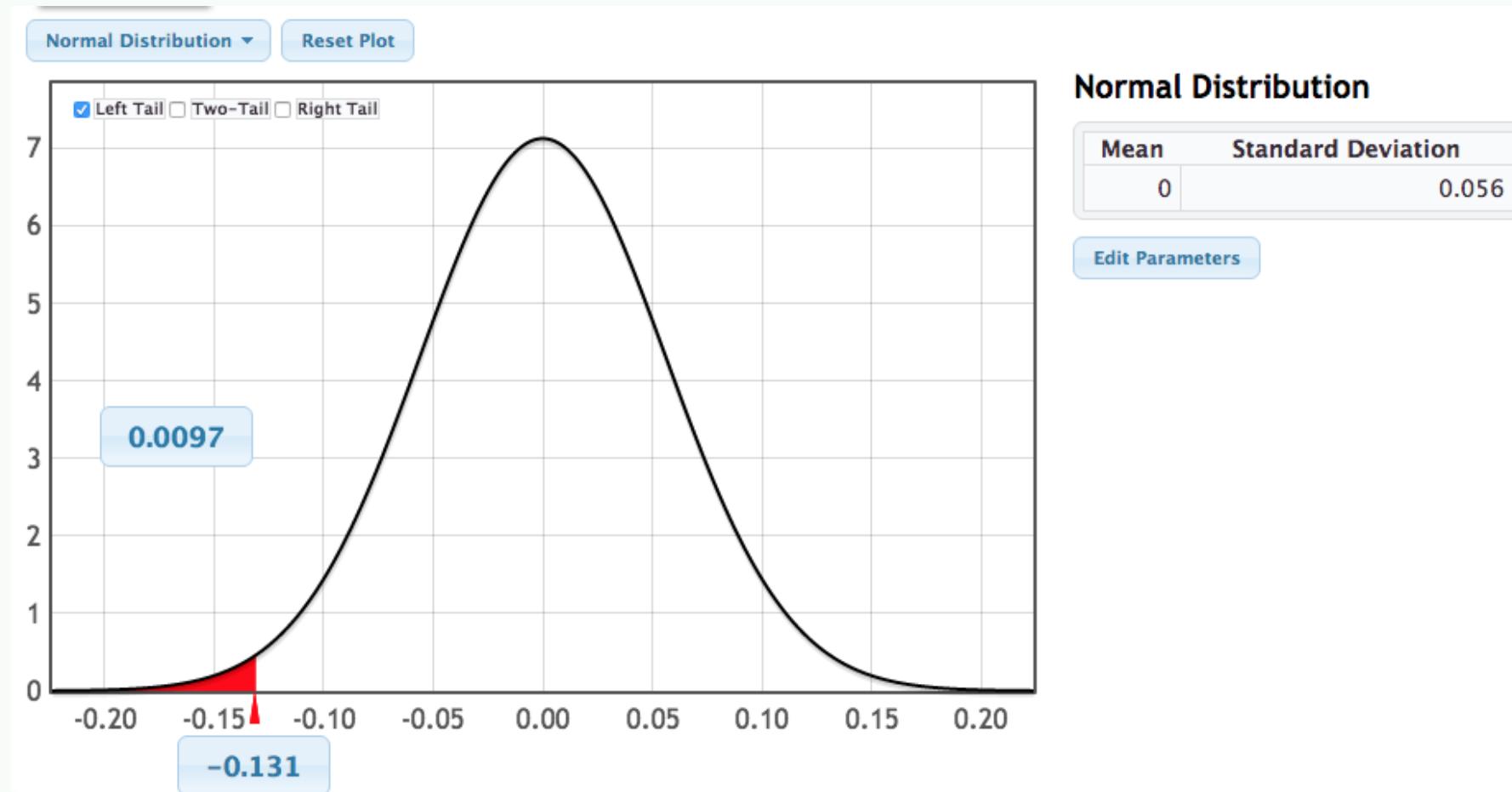
Which normal distribution should we use to approximate this?



- A. $N(0, -0.131)$
- B. $N(0, 0.056)$
- C. $N(-0.131, 0.056)$
- D. $N(0.056, 0)$

► Click for answer

Statkey: p-value from $N(\text{null}, \text{SE})$



Connecting Normal model to hypothesis tests

Suppose: randomization distribution is bell shaped.

- **Center**: hypothesized null parameter value
- **Spread**: the standard error given in the randomization graph (or by formula)
- **P-value**: computed from the normal model the "usual" way - the chance of being as extreme, or more extreme, than the observed statistic.

Standardized Statistic

The standardized test statistic (also known as a z-statistic) is

$$z = \frac{\text{statistic} - \text{null}}{SE}$$

Calculating the number of standard errors a statistic is from the null lets us assess extremity on a common scale.

Malaria and Mosquitos

Does infecting mosquitoes with Malaria actually impact the mosquitoes' behavior to favor the parasite?

- *After the parasite becomes infectious, do infected mosquitoes approach humans more often, so as to pass on the infection?*

Days 9 – 28

For the data after the mosquitoes become infectious (Days (9-28)), what are the relevant hypotheses?

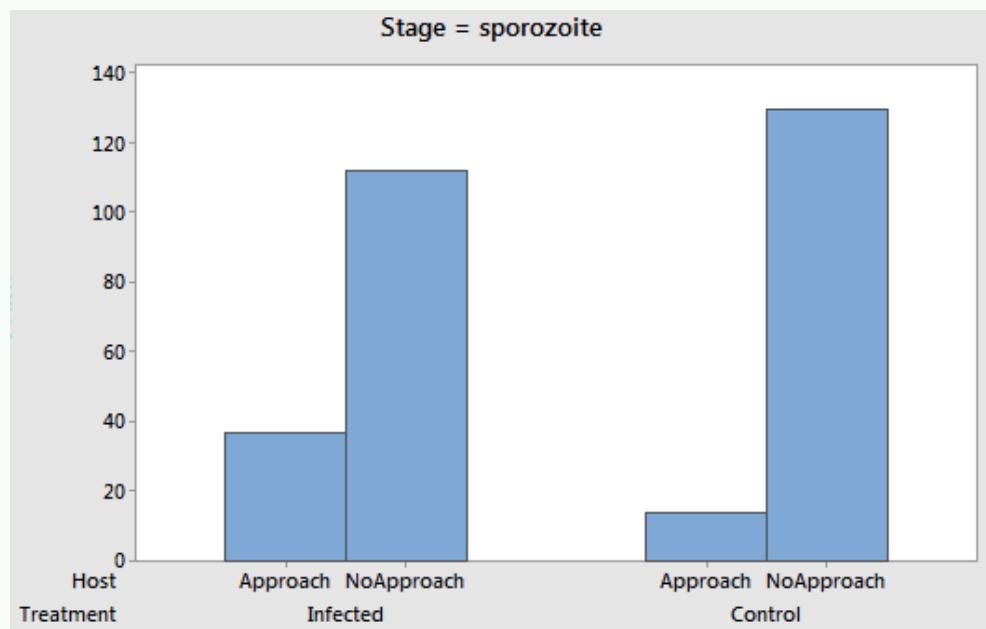
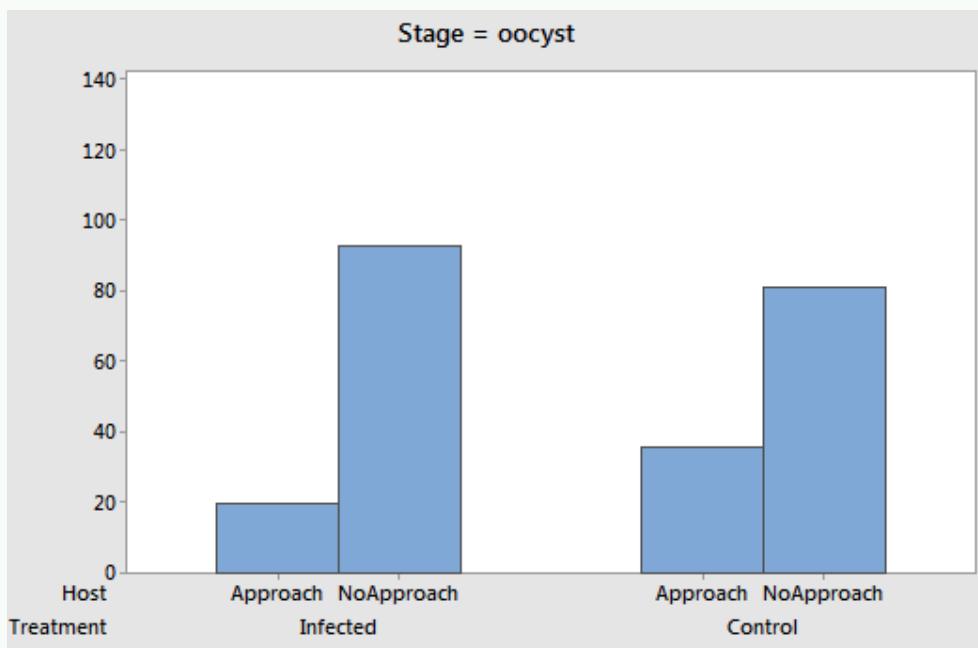
p_C :proportion of controls to approach human

p_E :proportion of exposed to approach human

- A. $H_0 : p_E = p_C, H_a : p_E < p_C$
- B. $H_0 : p_E = p_C, H_a : p_E > p_C$
- C. $H_0 : p_E < p_C, H_a : p_E = p_C$
- D. $H_0 : p_E > p_C, H_a : p_E = p_C$

► Click for answer

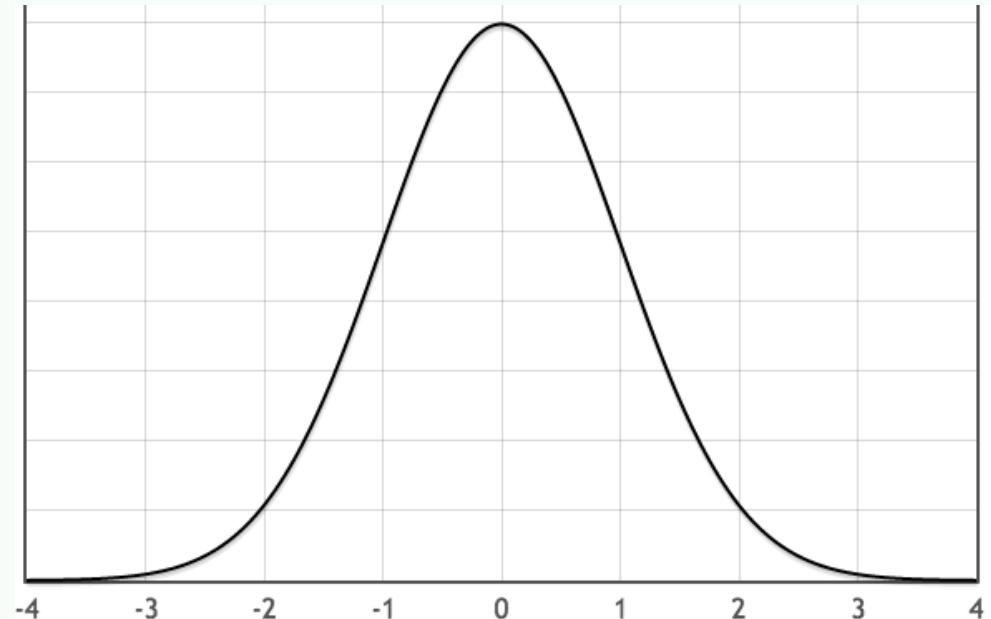
Before and after



Is the difference significant?

The difference in proportions is 0.151 and the standard error is 0.05. Is this significant?

- A. Yes
- B. No



Malaria and Mosquitoes

It appears that mosquitoes infected by malaria parasites do, in fact, behave in ways advantageous to the parasites!

- *Exposed mosquitos are less likely to approach before becoming infectious (so more likely to stay alive)*
- *Exposed mosquitos are more likely to approach humans after becoming infectious (so more likely to pass on disease)*

Formula for p-values Using $N(0,1)$

$$z = \frac{\text{sample statistic} - \text{null value}}{\text{SE}_{\text{From randomization distribution}}}$$

From original data From H_0

Connecting Normal model to Confidence Intervals

Suppose: bootstrap distribution is bell-shaped.

- **Center**: *sample statistic*
- **Spread**: *the standard error given in the bootstrap graph (or by formula)*

Connecting Normal model to Confidence Intervals

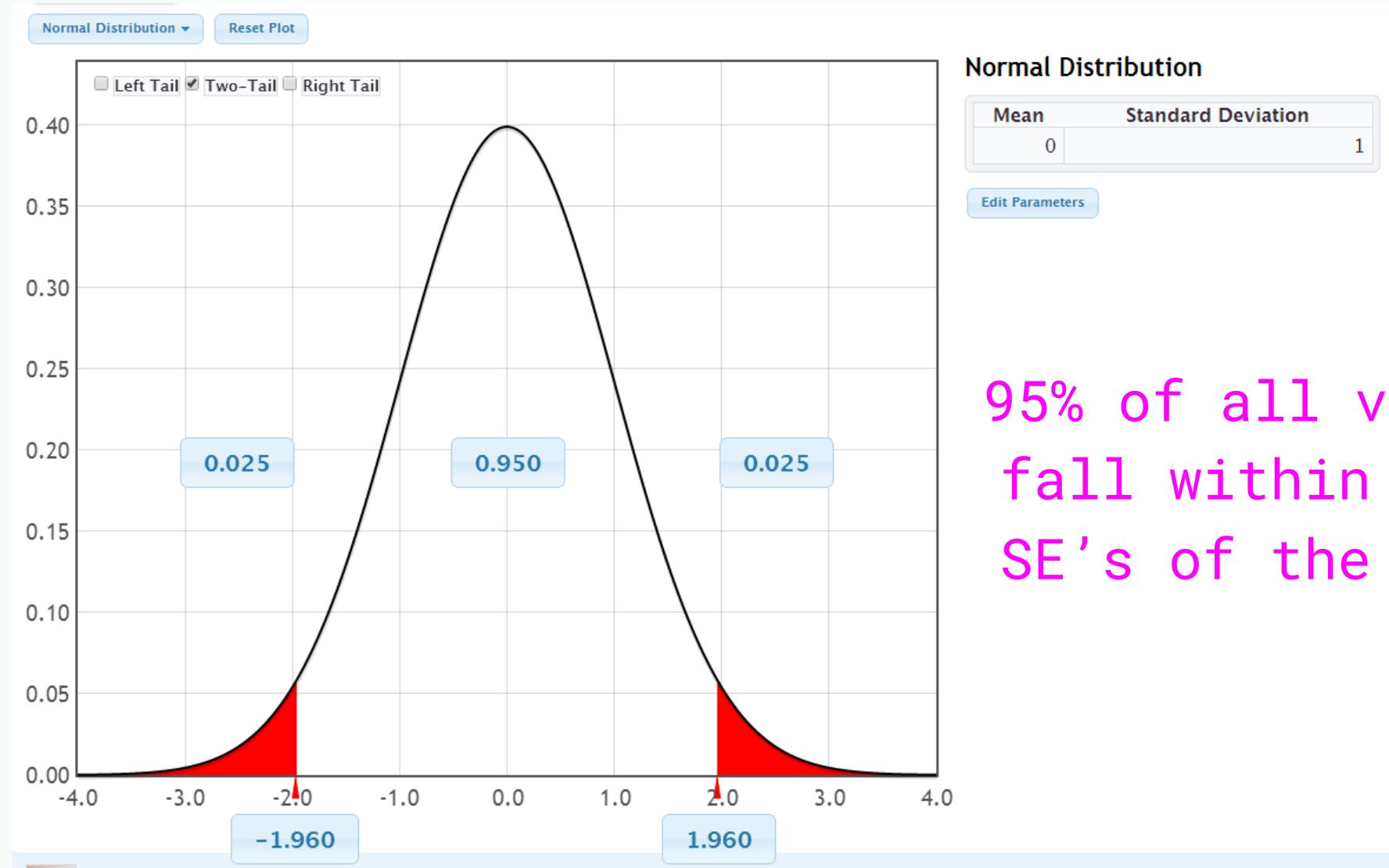
To get a 95% confidence interval we compute:

$$statistic \pm 2(SE)$$

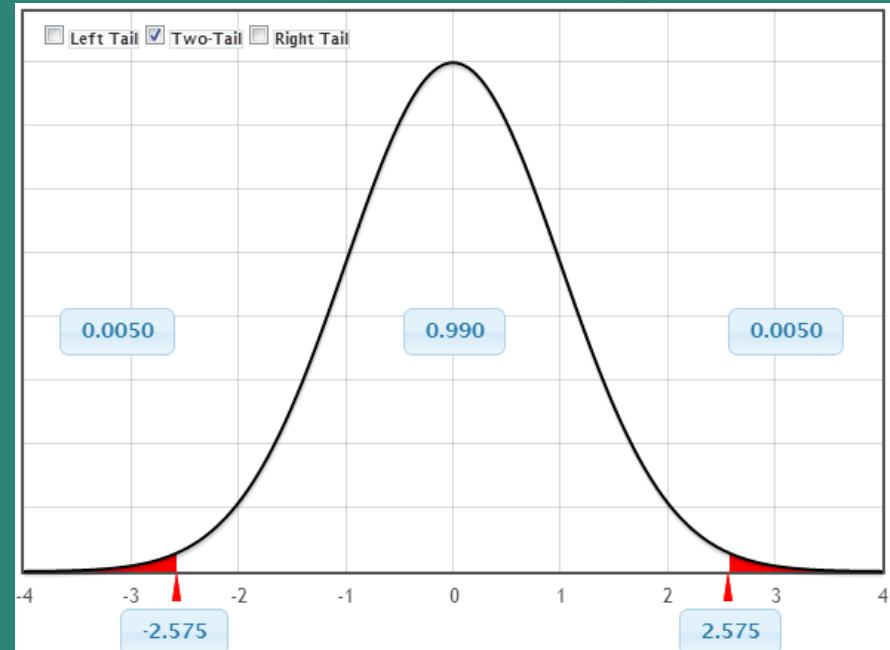
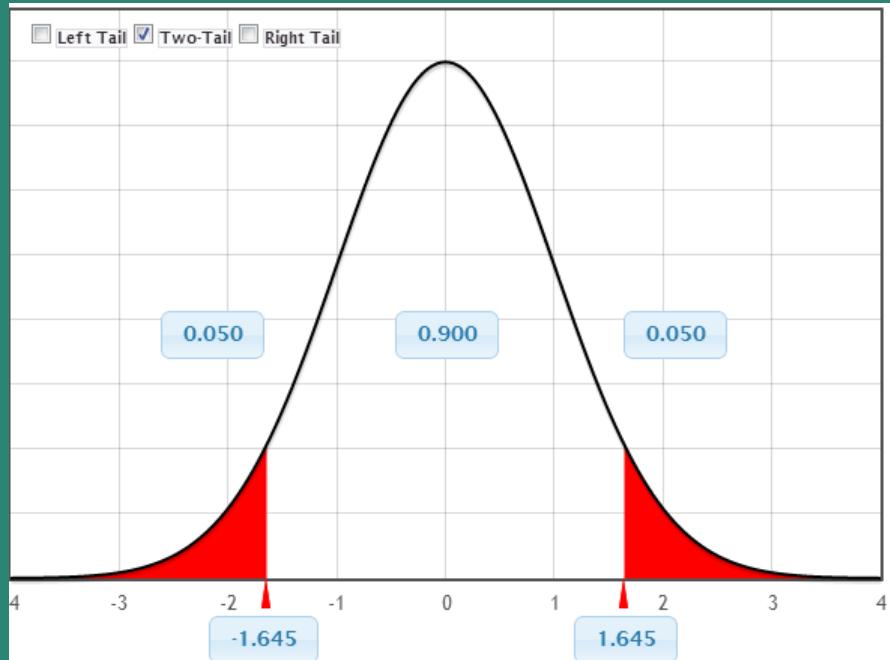
Why 2 SE's?

- 95% **of all sample means fall within 2 SE's of the population mean**
- **The value 2 is a z-score!**
- **Well, actually the precise z-score under a normal model is $z = 1.96$ instead of 2 !**

$N(0,1)$ model



What if we wanted a 90% CI? What z-score should we use to get the margin of error?



```
qnorm(0.95)  
[1] 1.644854
```

90% Confidence: $z^* = 1.645$

```
qnorm(0.995)  
[1] 2.575829
```

99% Confidence: $z^* = 2.576$

Confidence Interval using $N(0,1)$

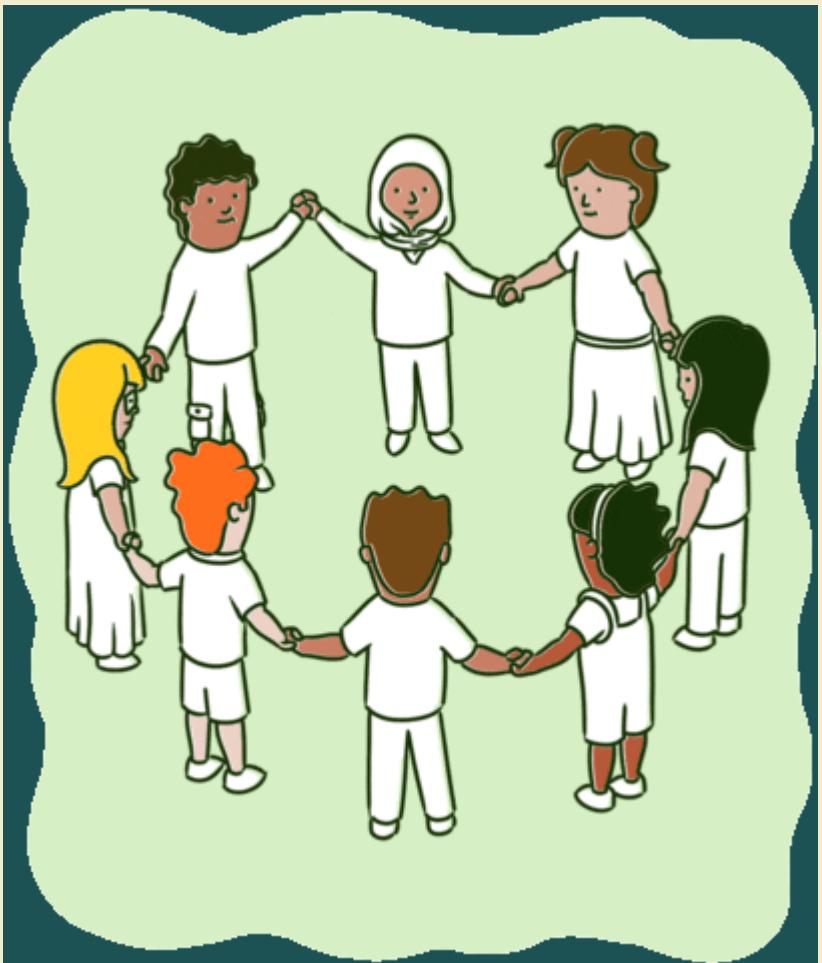
If a statistic is normally distributed, we find a confidence interval for the parameter using

$$\text{statistic} \pm z^* SE$$

where the area between $-z^*$ and $+z^*$ in the standard normal distribution is the desired level of confidence.

YOUR TURN1

05:00



Let's go over to the class activity .Rmd file and complete the tasks for today.