Hypothesis Tests and Confidence Intervals using Normal Distribution!

Stat 120

February 09 2023

How do Malaria parasites impact mosquito behavior?

✨This experiment looks at the behavior of mosquito when exposed to malaria infected mice and healthy mice.

✨The parasites go through two stages: not yet infectious (Days 1-8) and infectious (Days 9-28).

Malaria Parasites and Mosquitoes

> ✨*The response variable is whether the mosquito approached a human in a cage with them.*
>
> ✨*The experiment looks to see if this behavior differs by exposed vs control, and if it differs by infection stage.*

Cator LJ, George J, Blanford S, Murdock CC, Baker TC, Read AF, Thomas MB. (2013). 'Manipulation' without the parasite: altered feeding behaviour of mosquitoes is not dependent on infection with malaria parasites. Proc R Soc B 280: 20130711.

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?

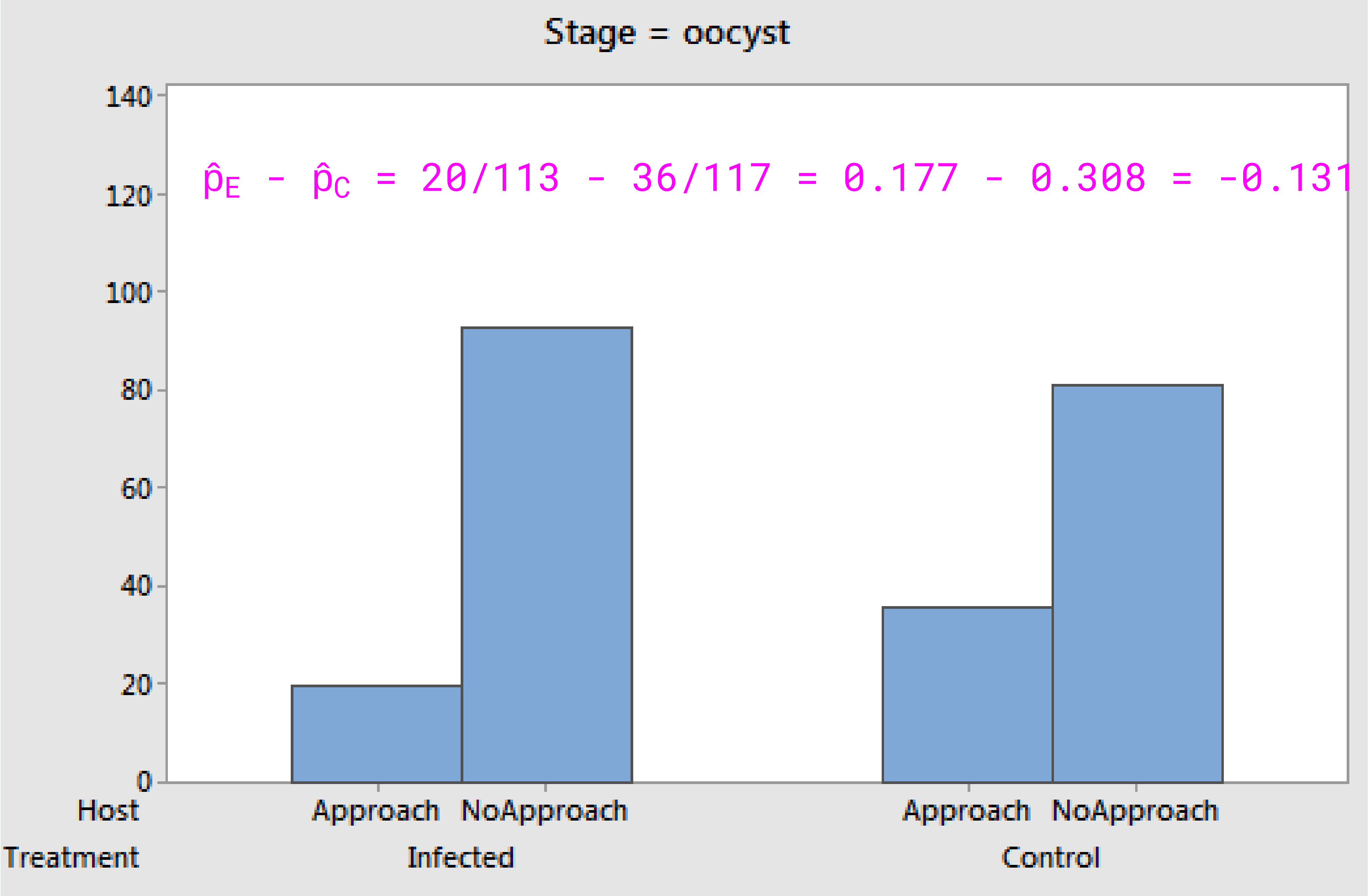
A. $\mathrm{H}_0 : p_\mathrm{E} = p_\mathrm{C}, \mathrm{H}_\mathrm{a} : p_\mathrm{E} < p_\mathrm{C}$

B. $\mathrm{H}_0 : p_\mathrm{E} = p_\mathrm{C}, \mathrm{H}_\mathrm{a} : p_\mathrm{E} > p_\mathrm{C}$

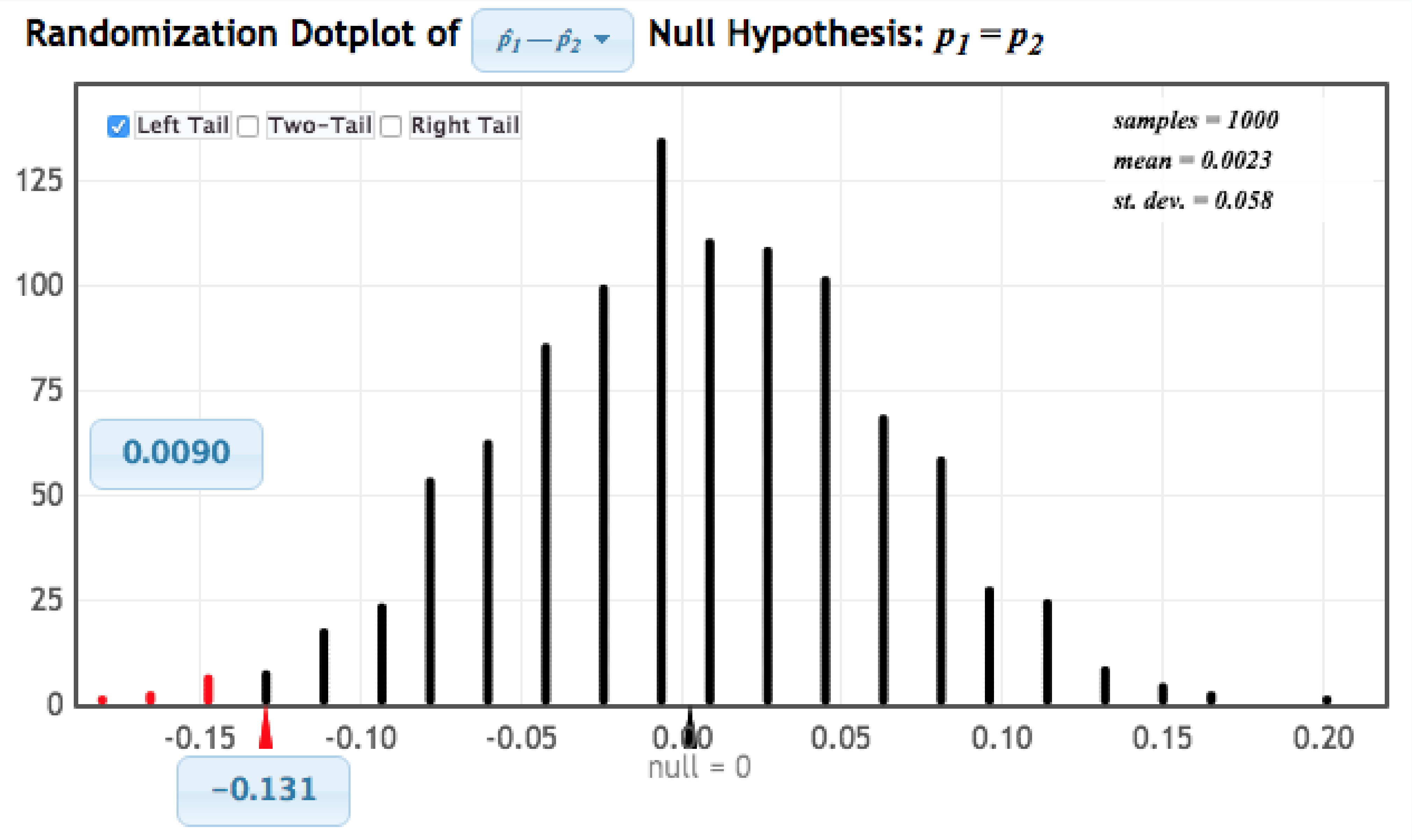
C. $\mathrm{H}_0 : p_\mathrm{E} < p_\mathrm{C}, \mathrm{H}_\mathrm{a} : p_\mathrm{E} = p_\mathrm{C}$

D. $H_0 : p_\mathrm{E} > p_\mathrm{C}, \mathrm{H}_\mathrm{a} : p_\mathrm{E} = p_\mathrm{C}$

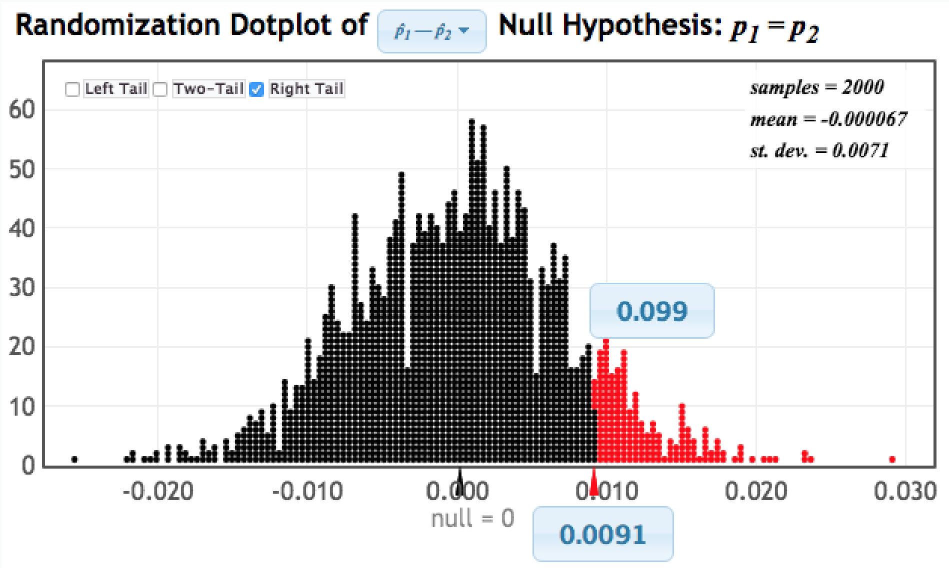
▶ Click for answer

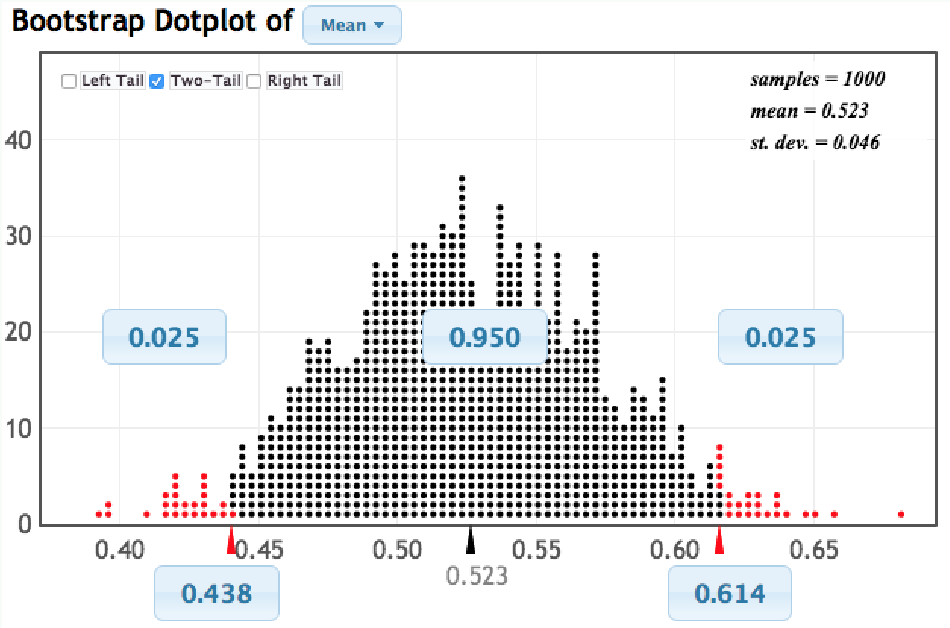


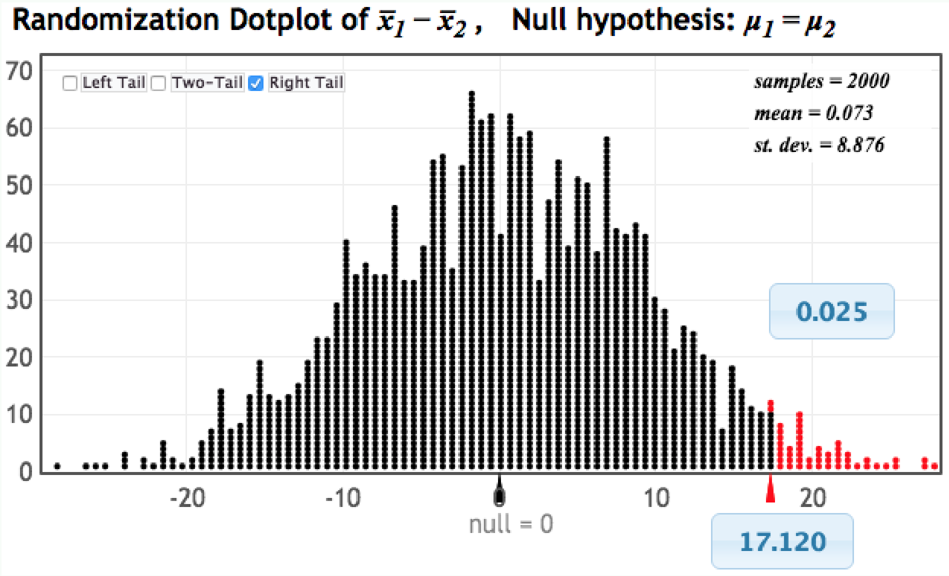
Stage = oocyst
$\hat{p}_E - \hat{p}_C = 20/113 - 36/117 = 0.177 - 0.308 = -0.131$
140
120
100
80
60
40
20
0
Host
Treatment
Approach
NoApproach
Infected
Approach
NoApproach
Control

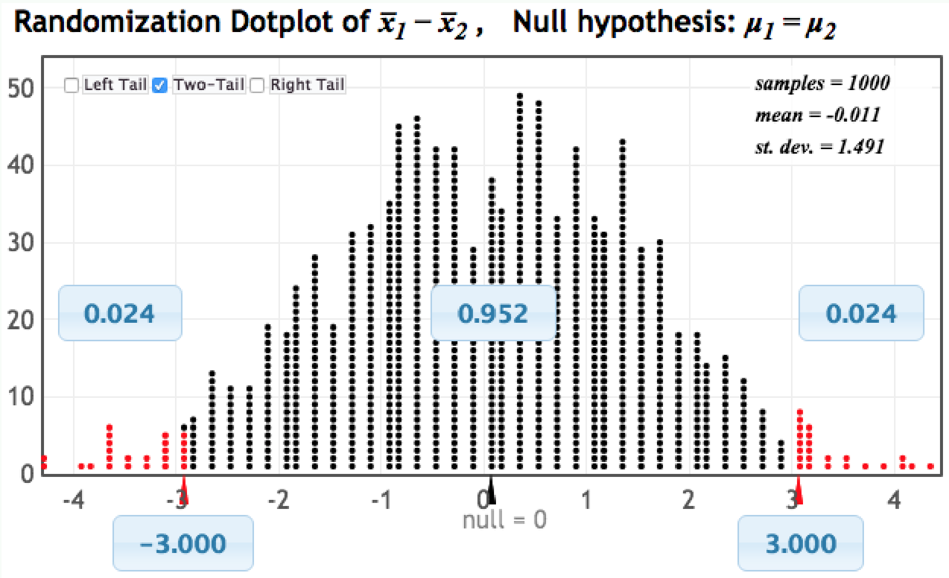


Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$
Left Tail
Two-Tail
Right Tail
samples = 1000
mean = 0.0023
st. dev. = 0.058
125
100
75
50
25
0
0.0090
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
null = 0
-0.131









What do you notice?

Central Limit Theorem

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

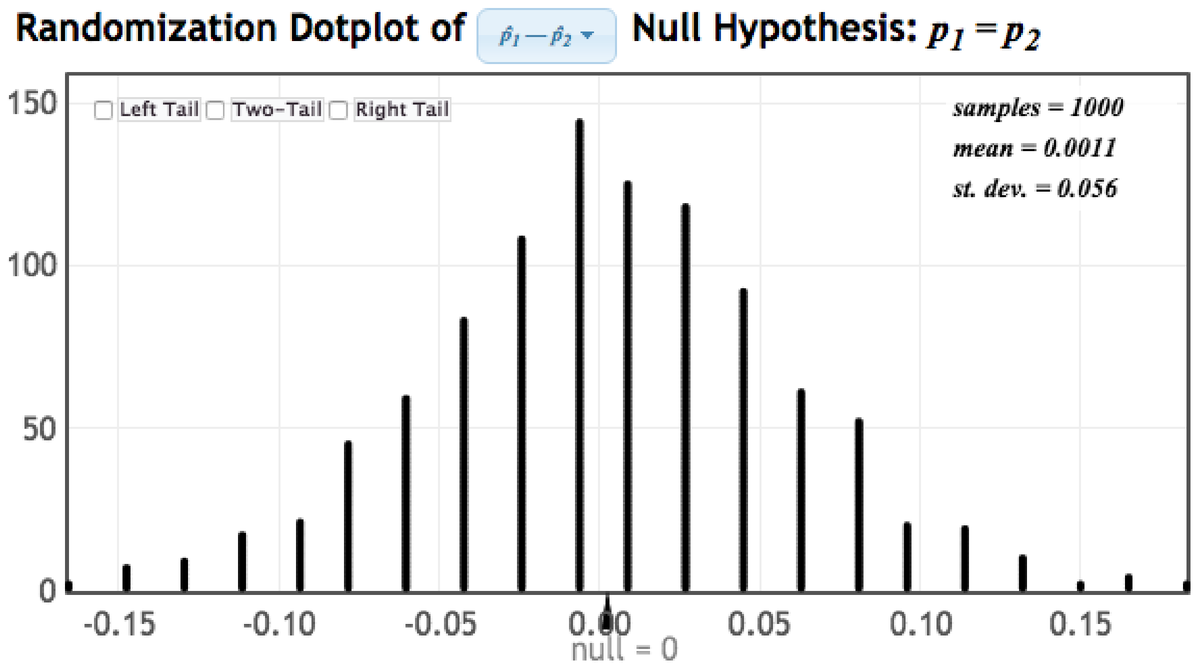
The catch: "sufficiently large sample size"

✨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, $\boldsymbol{n \geq 30}$ 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?



Original Sample

Group	Count	Sample Size	Proportion
Group 1	20	113	0.177
Group 2	36	117	0.308
Group 1−Group 2	-16	n/a	-0.131

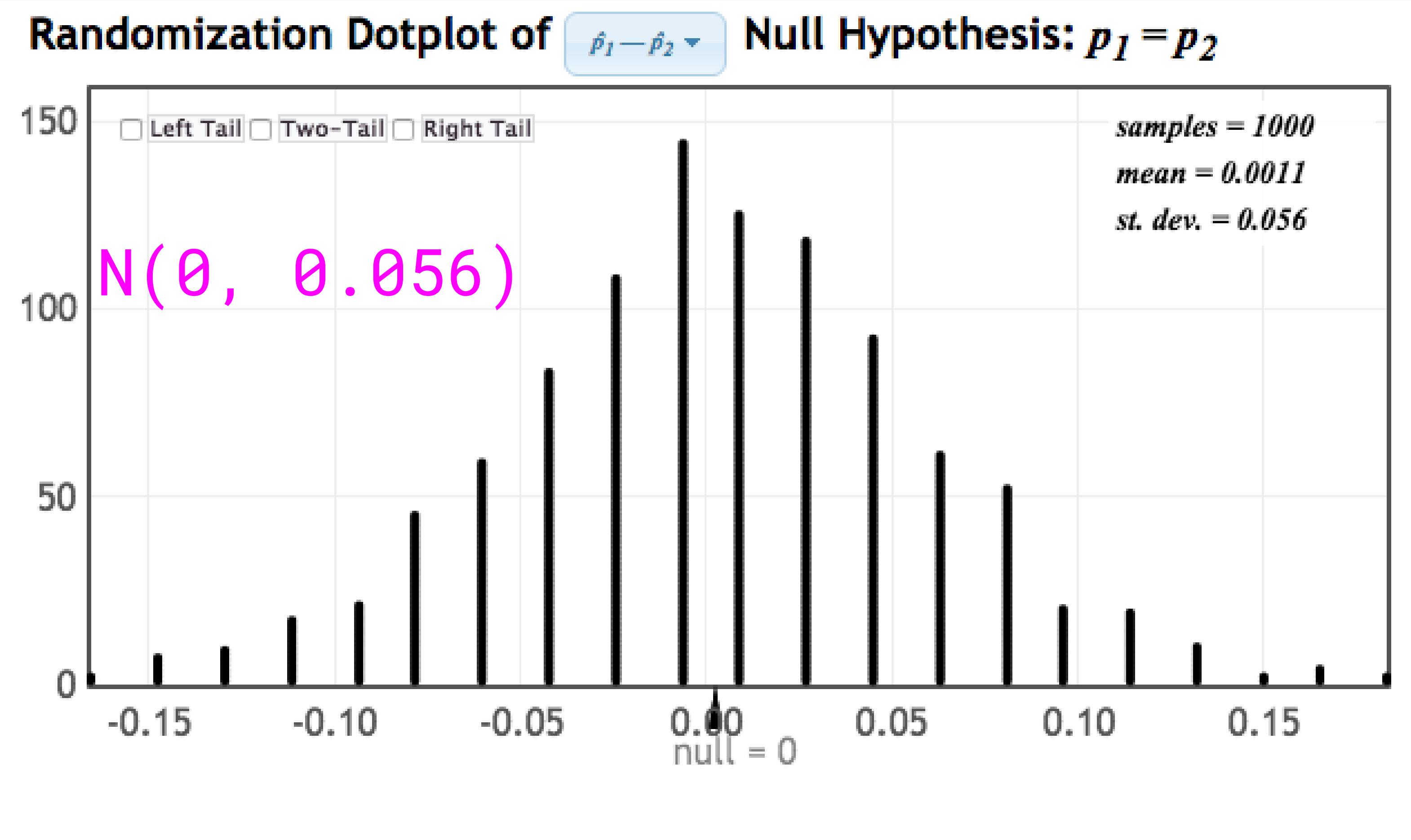
Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	27	113	0.239
Group 2	29	117	0.248
Group 1−Group 2	-2	n/a	-0.0089

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

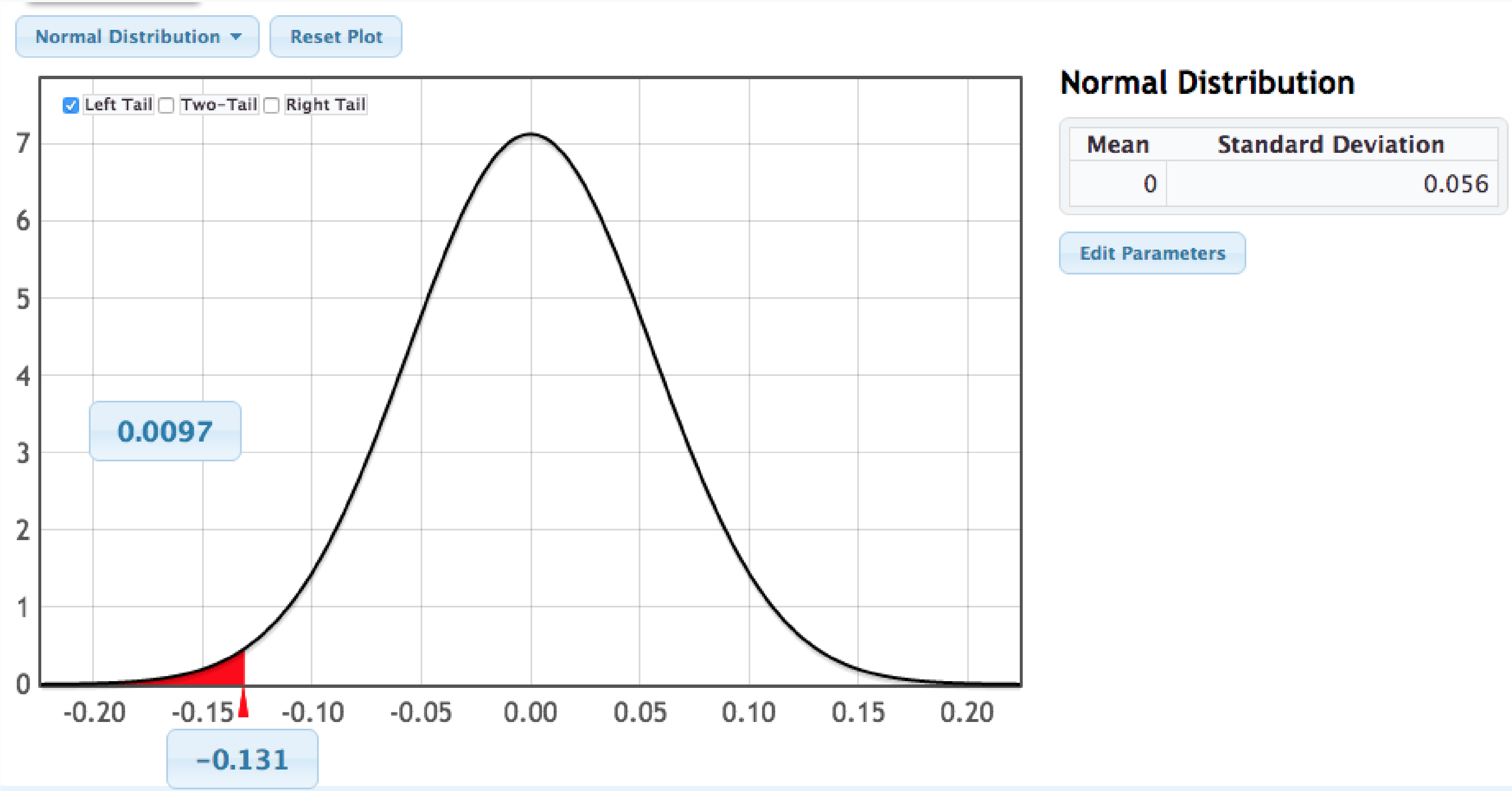
▶ Click for answer

Normal Distribution



We can compare the original statistic to this Normal distribution to find the p-value!

Statkey: p-value from N(null, 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

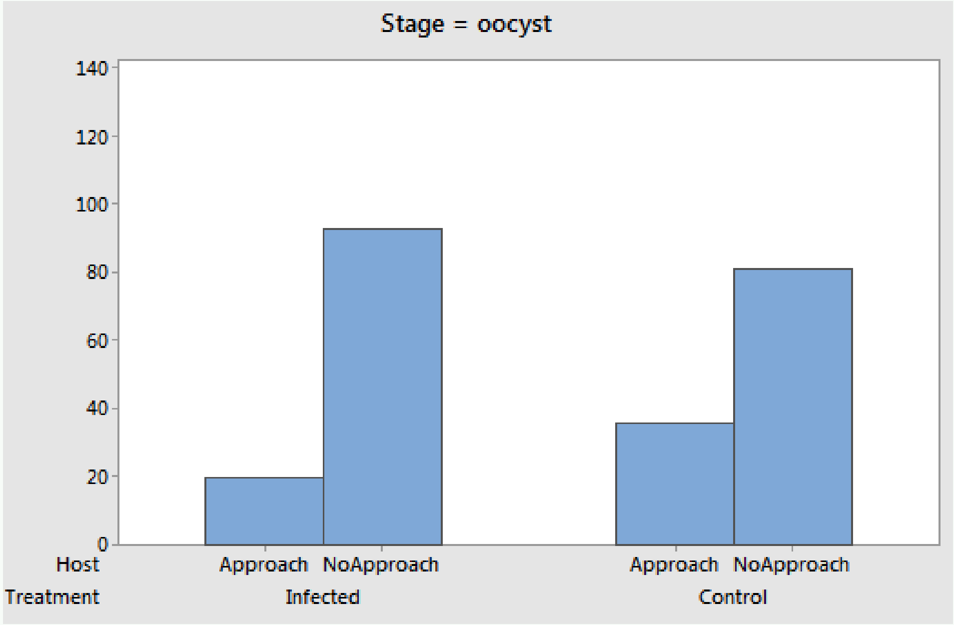
For the data after the mosquitoes become infectious (Days $9 - 28$), what are the relevant hypotheses?

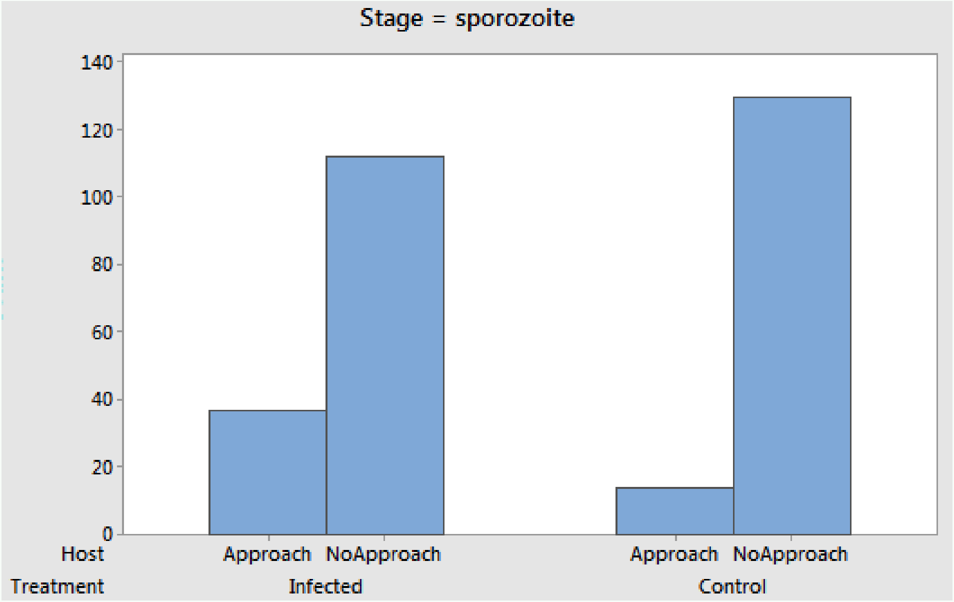
$$\boldsymbol{p}_C \text{ :proportion of controls to approach human}$$
$$\boldsymbol{p}_E \text{ :proportion of exposed to approach human}$$

$$
\begin{array}{ll}
A. & \mathrm{H}_0 : p_{\mathrm{E}} = p_{\mathrm{C}}, \mathrm{H}_{\mathrm{a}} : p_{\mathrm{E}} < p_{\mathrm{C}} \\
B. & \mathrm{H}_0 : p_{\mathrm{E}} = p_{\mathrm{C}}, \mathrm{H}_{\mathrm{a}} : p_{\mathrm{E}} > p_{\mathrm{C}} \\
C. & \mathrm{H}_0 : p_{\mathrm{E}} < p_{\mathrm{C}}, \mathrm{H}_{\mathrm{a}} : p_{\mathrm{E}} = p_{\mathrm{C}} \\
D. & \mathrm{H}_0 : p_{\mathrm{E}} > p_{\mathrm{C}}, \mathrm{H}_{\mathrm{a}} : p_{\mathrm{E}} = p_{\mathrm{C}}
\end{array}
$$

▶ Click for answer

Before and after



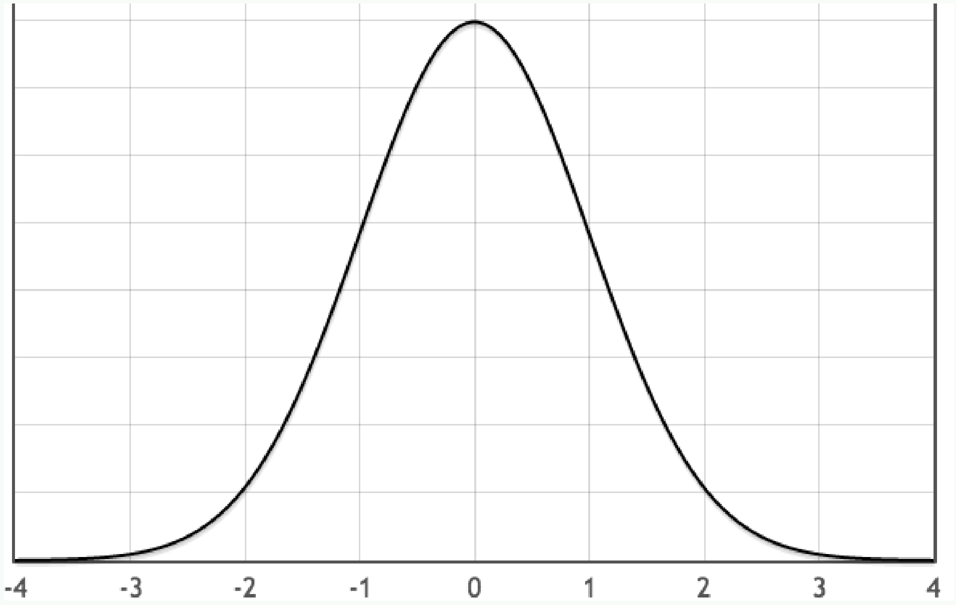


Is the difference significant?

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)

From original data

From H_0

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

From randomization distribution0

Connecting Normal model to Confidence Intervals

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

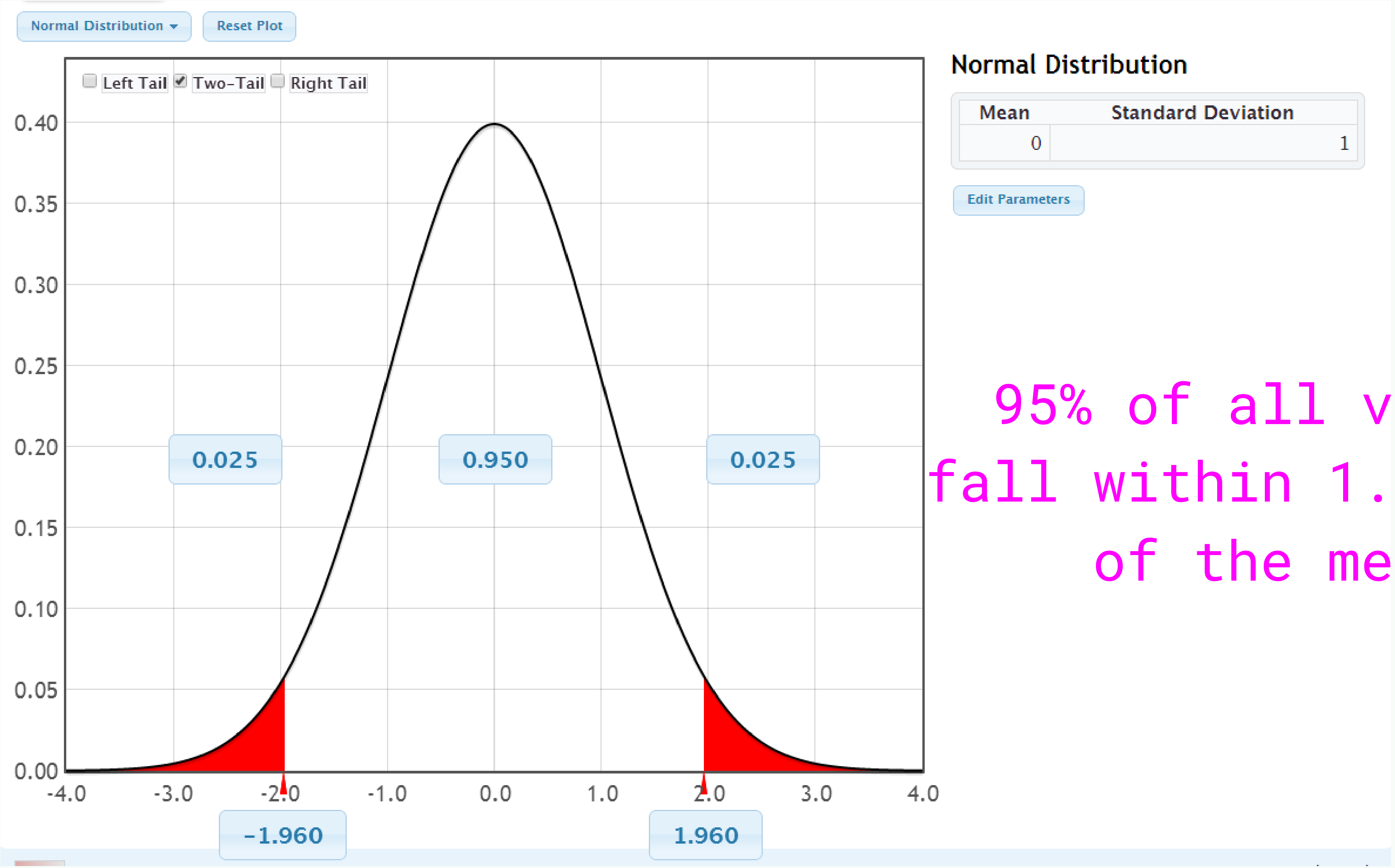
To get a 95% confidence interval we compute:

$$statistic \pm 2(SE)$$

Why 2 SE's?

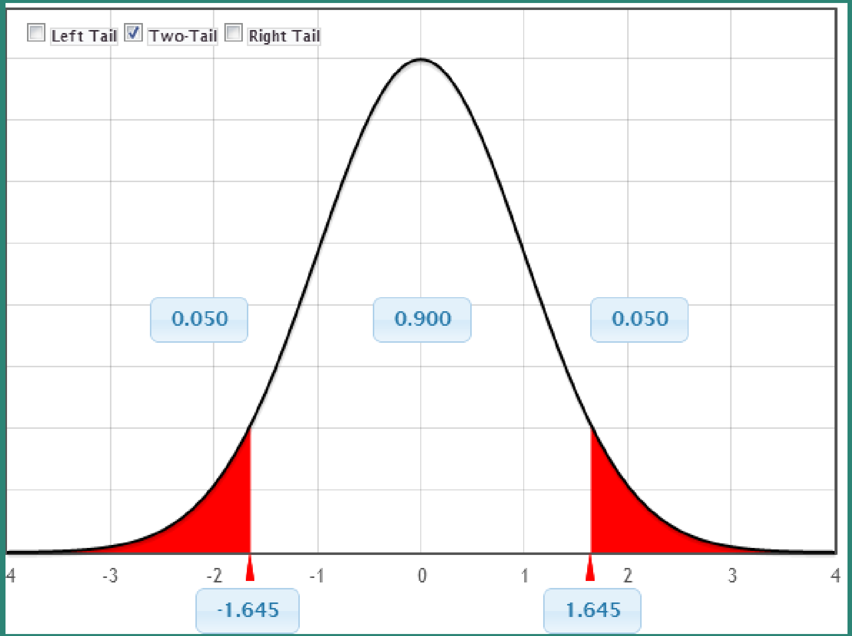
- ✨95% ***of all sample means fall within 2 SE's of the population mean****
- ✨***The value 2 is a z-score!***
- ✨***Well, actually the precise z-score under a normal model is $z = 1.96$ instead of 2 !***

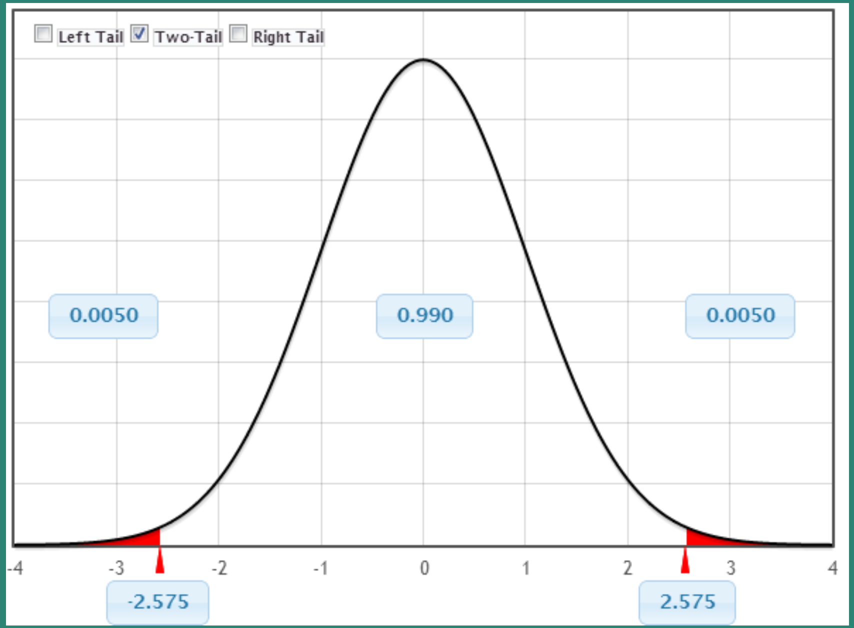
N(0,1) model



95% of all values fall within 1.96 SE's of the mean

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
```

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

90% Confidence: $z^* = 1.645$

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

$$statistic \pm z^* SE$$

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

Global Warming

What percentage of Americans believe in global warming?

A survey on 2,251 randomly selected individuals conducted in October 2010 found that 1328 answered "Yes" to the question

"Is there solid evidence of global warming?"

Give and interpret a $95\% \mathrm{CI}$ for the proportion of Americans who believe there is solid evidence of global warming.

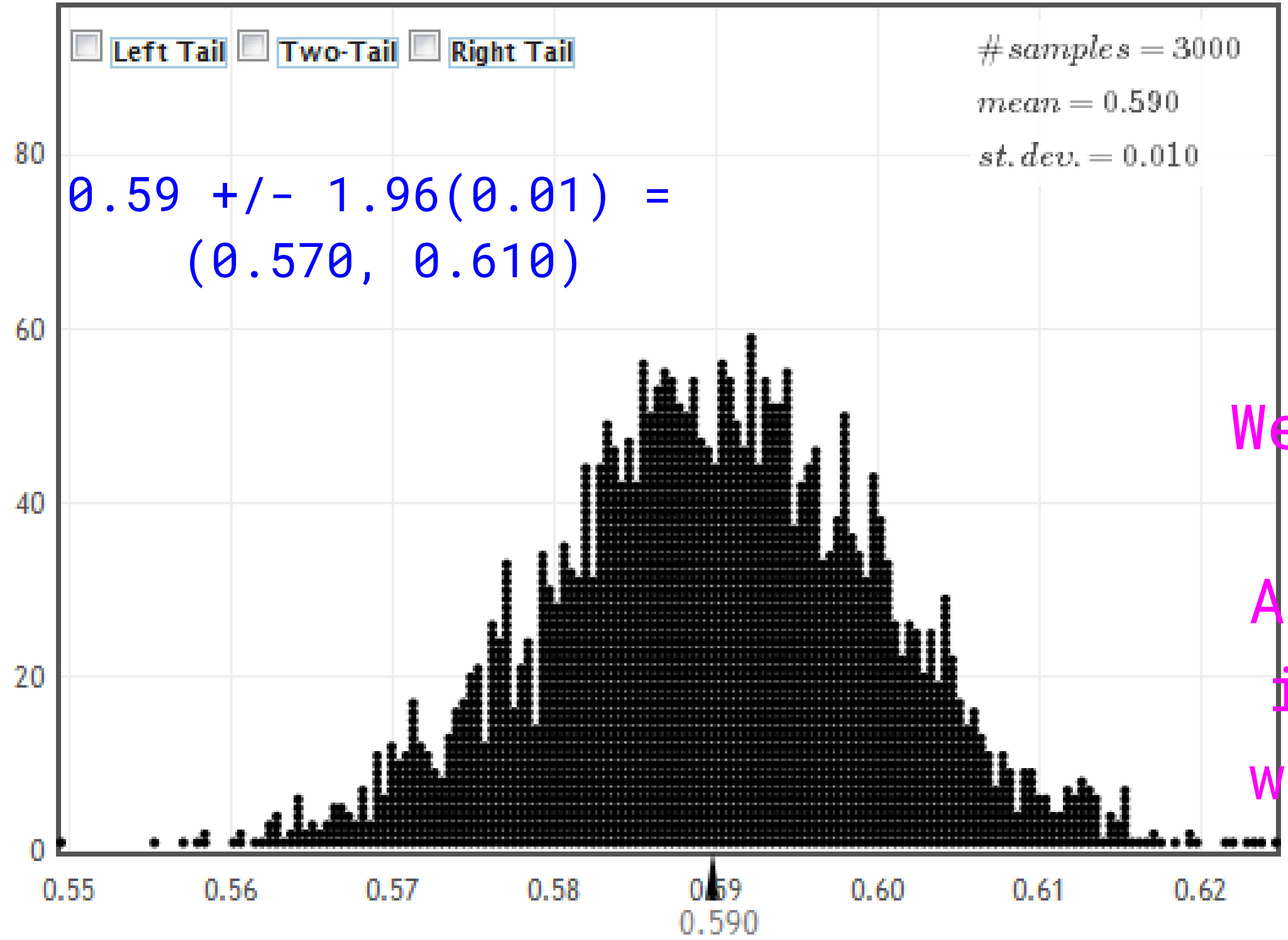
Source: “Wide Partisan Divide Over Global Warming”, Pew Research Center, 10/27/10. s

Bootstrap For One Categorical Variable [Return to StatKey Index]

Custom Data | Edit Data

Generate 1 Samples | Generate 10 Samples | Generate 100 Samples | Generate 1000 Samples | Reset Plot

Bootstrap Dotplot of Proportion



0.59 +/- 1.96(0.01) =
(0.570, 0.610)

Original Sample

Count	n	Proportion
1328	2251	0.590

Bootstrap Sample

Count	n	Proportion
1304	2251	0.579

We are 95% confident that the true percentage of all Americans that believe there is solid evidence of global warming is between 57.0% and 61.0%

Global Warming

What is a 90% confidence interval for the proportion of US adults who believe in global warming?

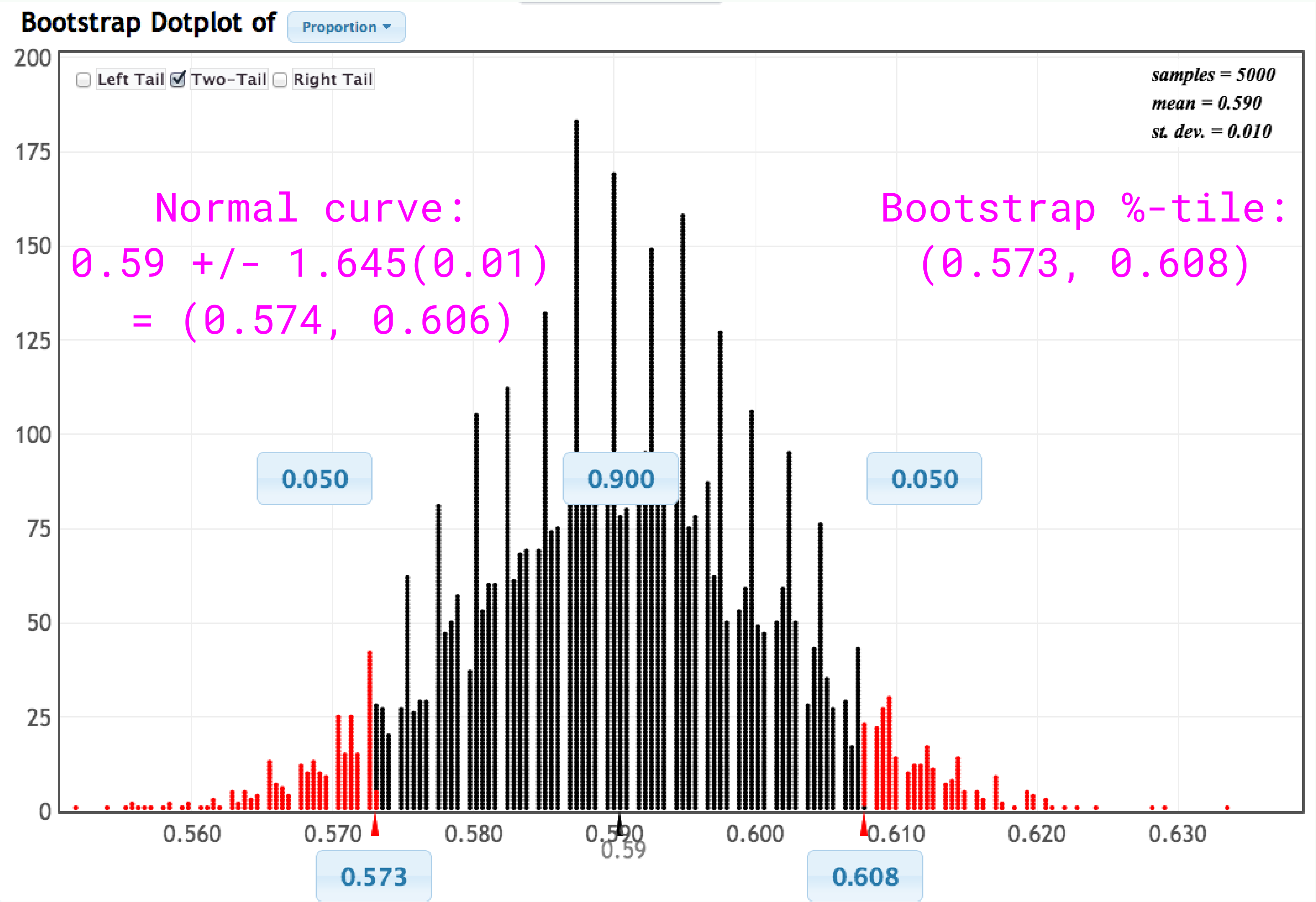
$$0.59 \pm 1.645(0.01) = (0.574, 0.606)$$

What is a 99% confidence interval?

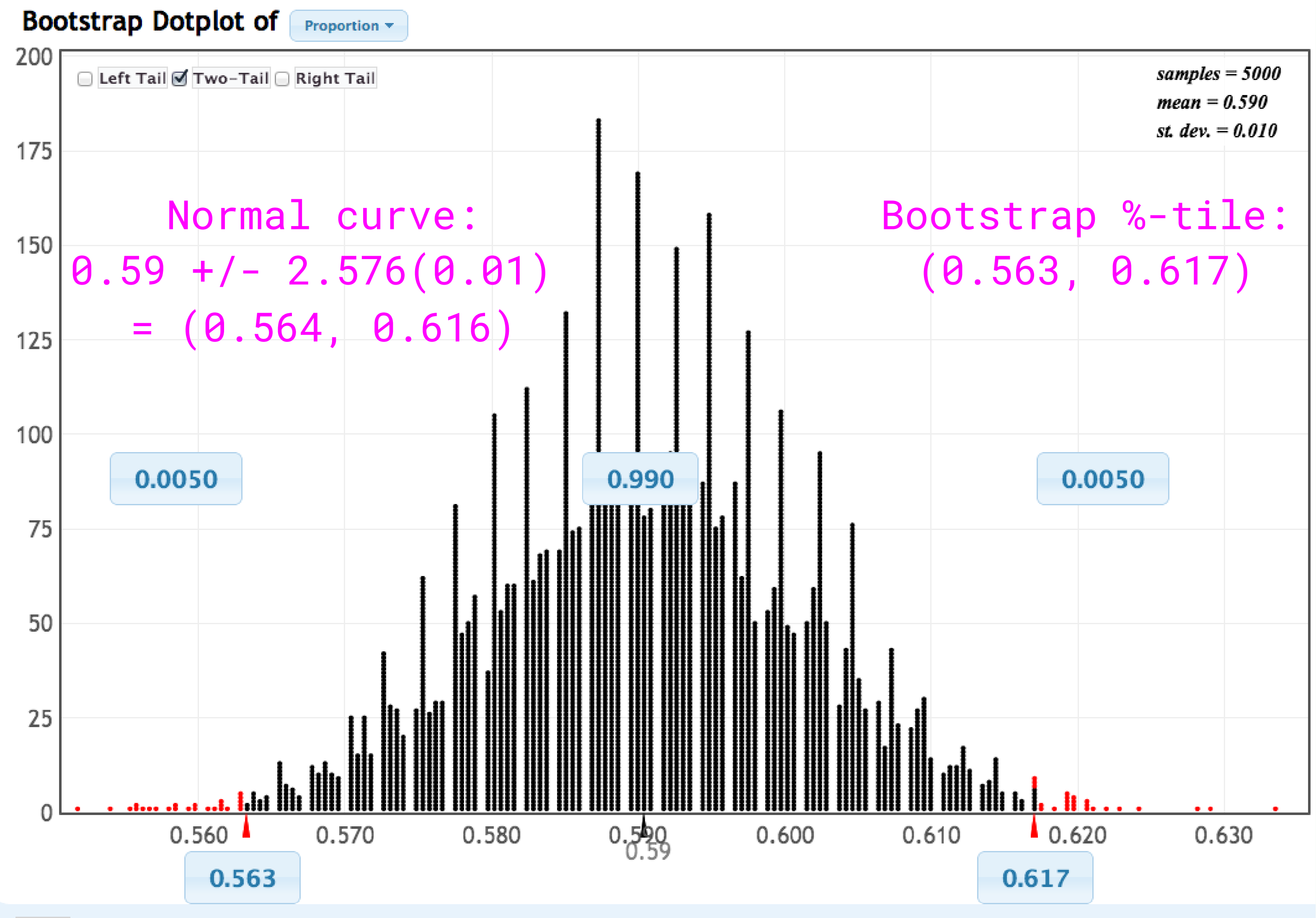
$$0.59 \pm 2.576(0.01) = (0.564, 0.616)$$

Remember, more confidence = wider interval. So how do these compare to the bootstrap CI ?

Global Warming: 90% CI

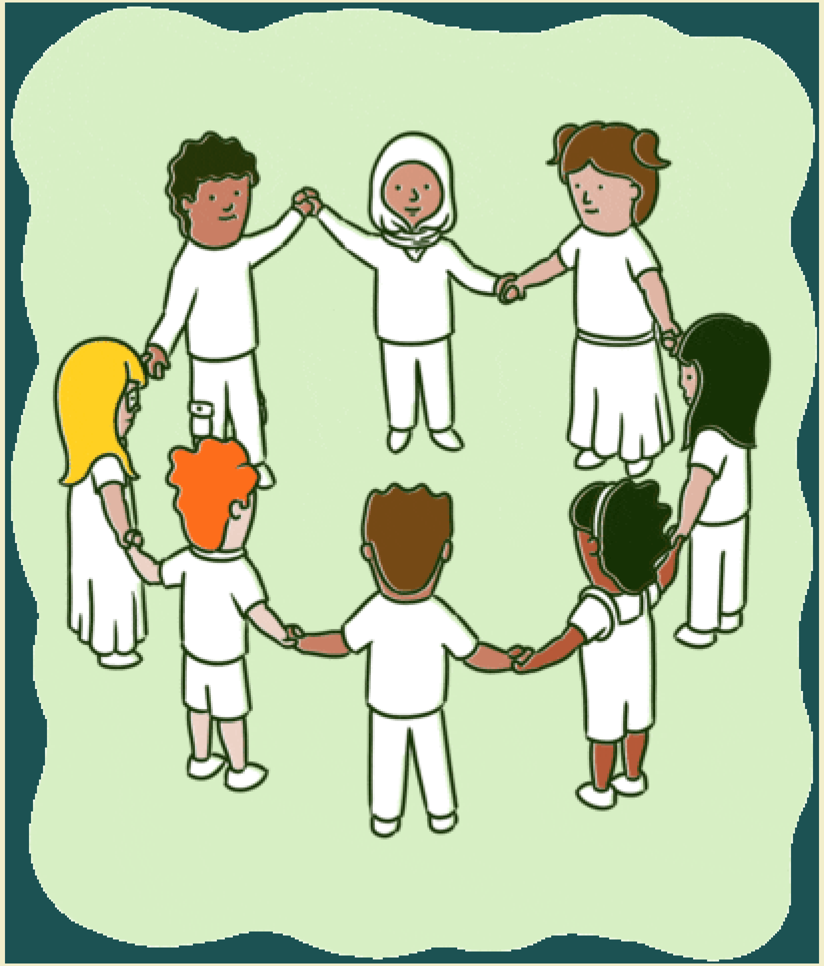


Global Warming: 99% CI



Your Turn 1

10:00



Please go over the class activity 16