Randomization Distributions and P-values

Stat 120

April 21 2023

Statistical Hypothesis

Null Hypothesis (H_0) : ***Claim that there is no effect or difference.***

Alternative Hypothesis (H_a) : ***Claim for which we seek evidence.***

Always claims about population parameters.

Statistical Significance

Hypothesis testing is similar to how our justice system works (or is suppose to work).

$$H_0 : \text{Defendant is innocent vs. } H_a \text{: Defendant is guilty}$$

Assumption: Defendant is innocent (H_0)

Verdict:

- **Guilty:** evidence (data) "beyond a reasonable doubt" points to guilt (Statistically significant)
- **Not Guilty:** evidence (data) not beyond a reasonable doubt, but we don't know if they are truly innocent (H_0)

Extrasensory Perception (ESP)

p= Proportion of correct guesses

$$H_0 : \boldsymbol{p} = \mathbf{1}/5$$
$$H_a : \boldsymbol{p} > \mathbf{1/5}$$

Extrasensory Perception (ESP)

p= Proportion of correct guesses

$$\mathrm{H}_0 : \boldsymbol{p} = \mathbf{1}/5$$
$$\mathrm{H}_\mathrm{a} : \boldsymbol{p} > \mathbf{1/5}$$

- ***Suppose we try this $n = 10$ times and get 3 correct guesses.***
- ***What kinds of statistics (sample proportions) would we observe just by chance, if the null were true and ESP does not exist?***
- ***How can we generate this distribution?***

Extrasensory Perception (ESP)

p= Proportion of correct guesses

$$\mathrm{H}_0 : \boldsymbol{p} = \mathbf{1}/5$$
$$\mathrm{H}_a : \boldsymbol{p} > \mathbf{1}/\mathbf{5}$$

- ***Suppose we try this $n = 10$ times and get 3 correct guesses.***
- ***What kinds of statistics (sample proportions) would we observe just by chance, if the null were true and ESP does not exist?***
- ***How can we generate this distribution?***

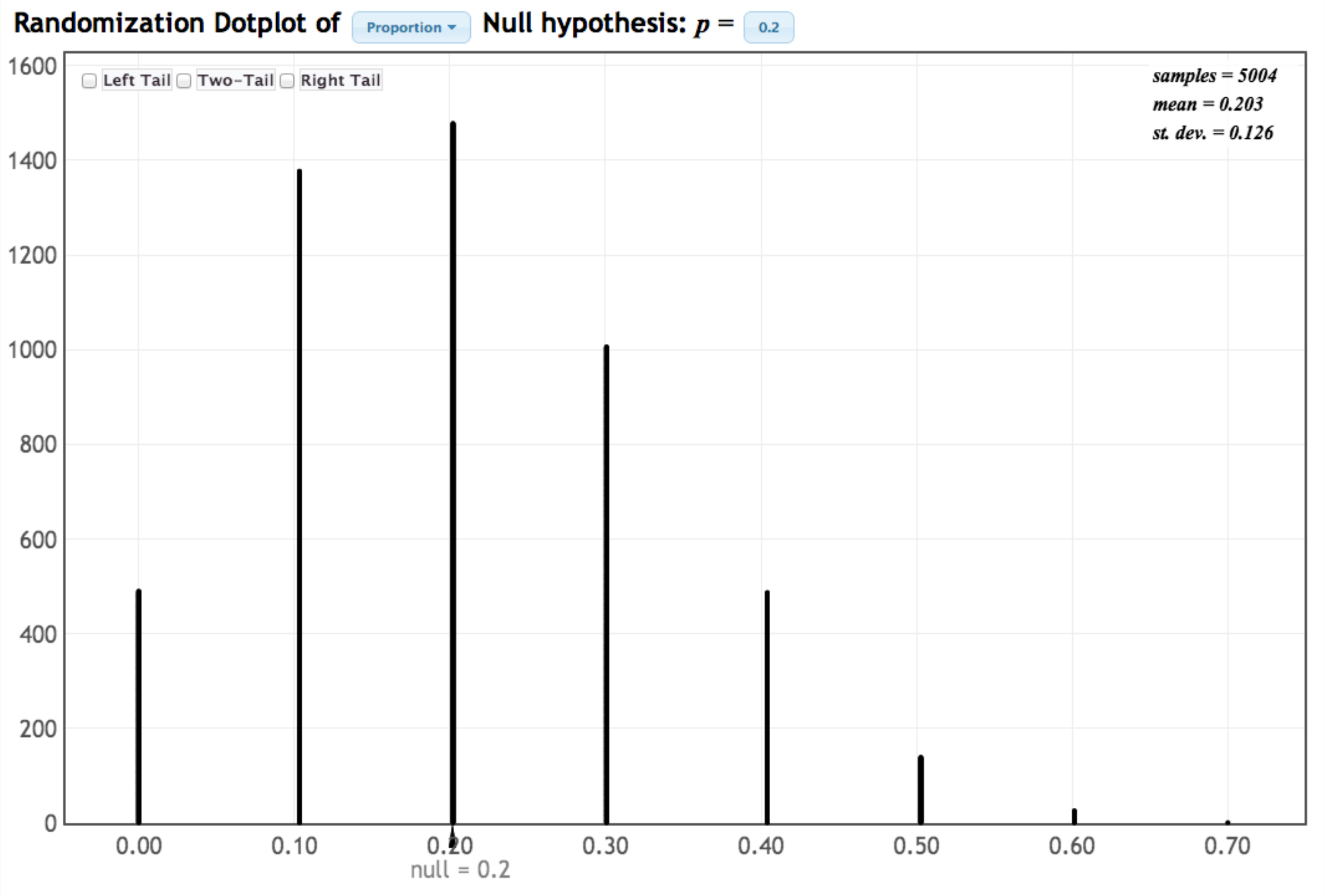
Simulate many samples of size $n = 10$ with $p = 0.2$ and look at the distribution of sample proportions.

Randomization Distribution

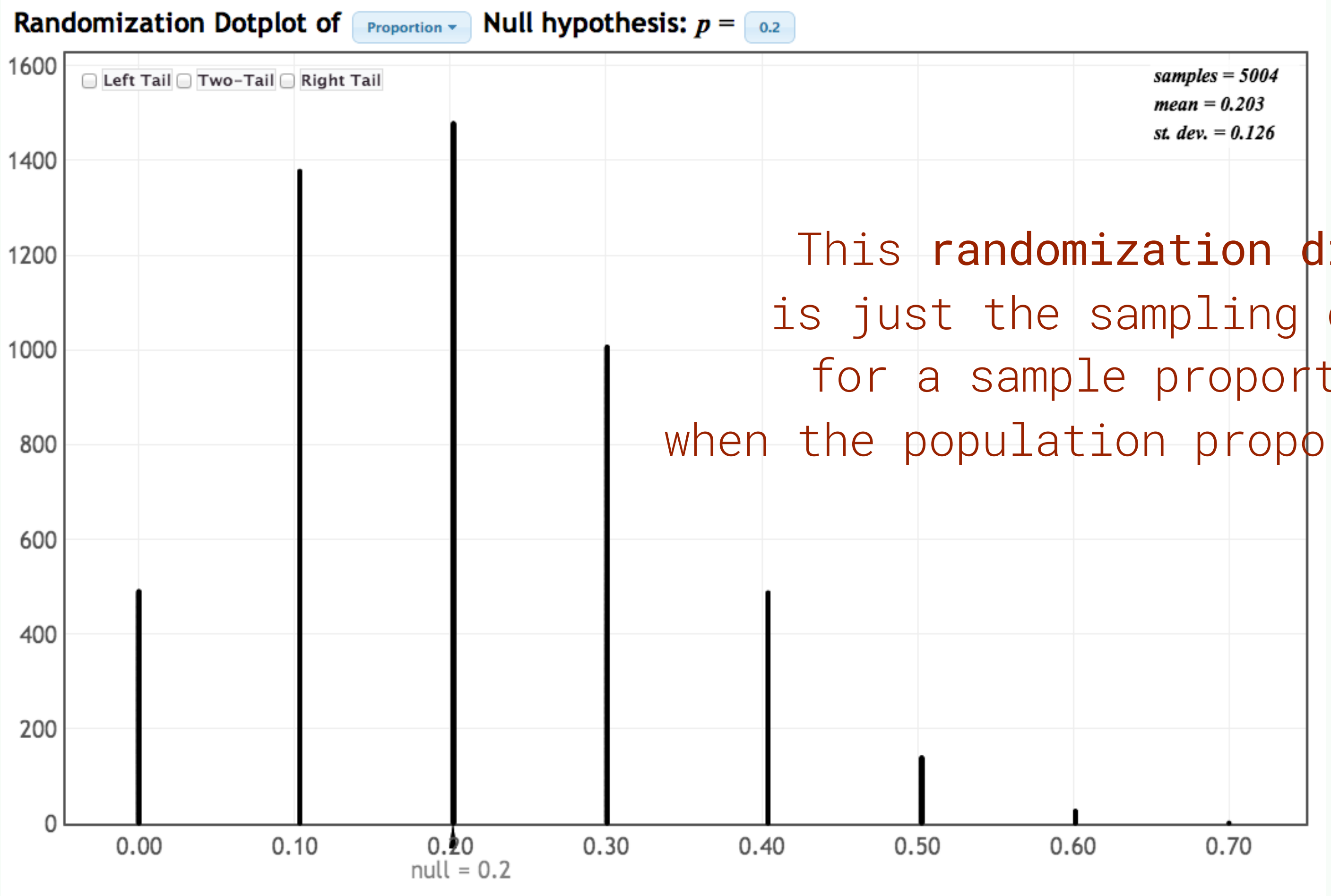
A randomization distribution is a collection of statistics from samples simulated assuming the null hypothesis is true

- Also known as a **permutation distribution**.
- A randomization distribution is centered at the value of the parameter given in the **null hypothesis**.

Randomization Distribution for ESP



Randomization Distribution for ESP



This **randomization distribution** is just the sampling distribution for a sample proportion (n=10) when the population proportion is **p=0.20.**

P-value

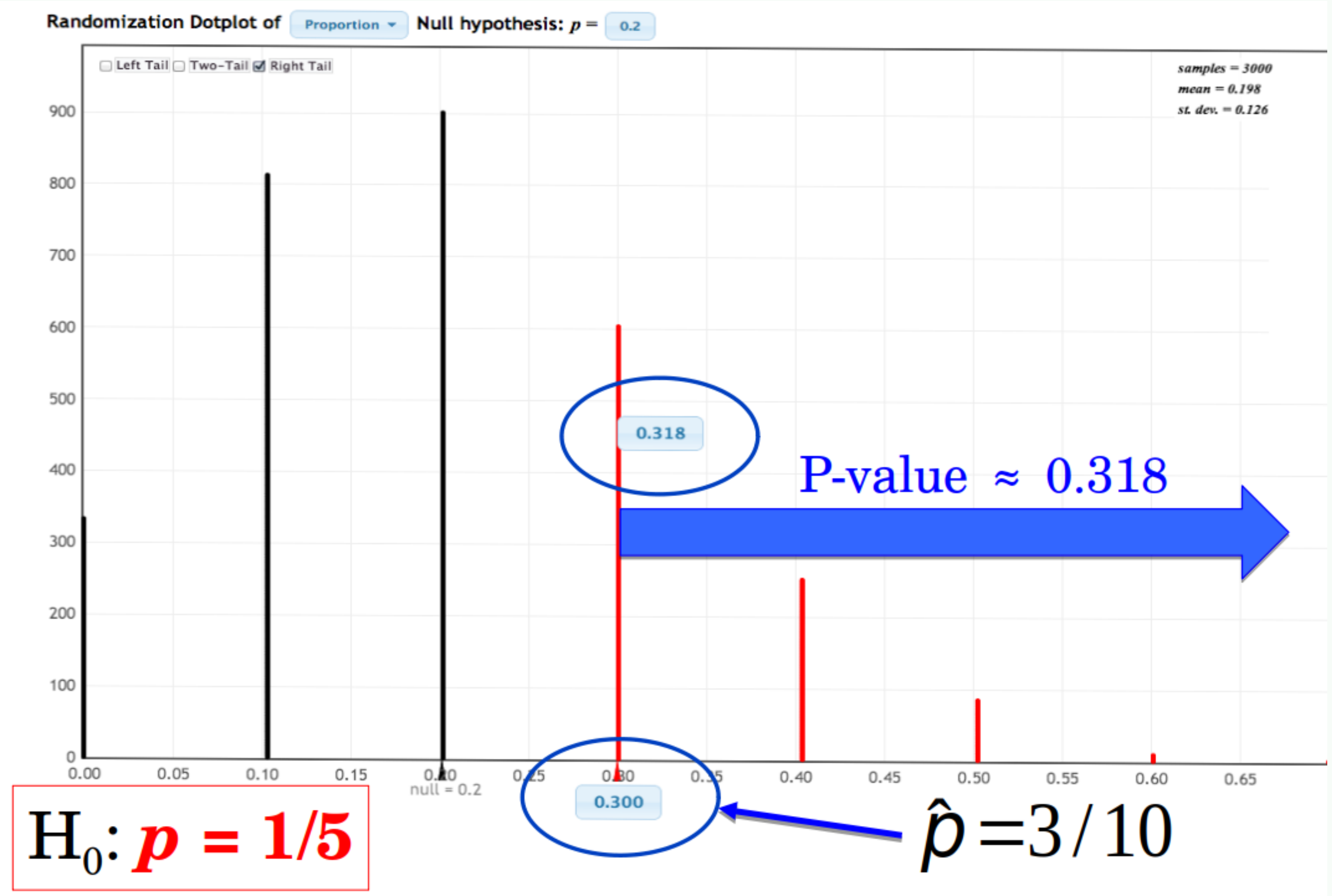
*The **p-value** is the chance of obtaining a sample statistic as extreme (or more extreme) than the observed sample statistic, if the null hypothesis is true*

P-value

The p-value is the chance of obtaining a sample statistic as extreme (or more extreme) than the observed sample statistic, if the null hypothesis is true

- The p-value can be calculated as the proportion of statistics in a randomization distribution that are as extreme (or more extreme) than the observed sample statistic
- "extreme" is determined by the alternative hypothesis

StatKey ESP



Randomization Dotplot of Proportion
Null hypothesis: p = 0.2
Left Tail
Two-Tail
Right Tail
samples = 3000
mean = 0.198
st. dev. = 0.126
900
800
700
600
500
400
300
200
100
0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
null = 0.2
0.318
0.300
P-value ≈ 0.318
H0: p = 1/5
p̂ = 3/10

p-value for ESP

The p-value is the chance of getting at least 3 out of 10 guesses correct, if p = 0.2.

- *P-value is about 0.318.*
- *About 31% of the time we would get at least 3 out 10 guesses correct just by chance (no ESP). (interpretation)*

p-value for ESP

The p-value is the chance of getting at least 3 out of 10 guesses correct, if p = 0.2.

- ***P-value is about 0.318.***
- ***About 31% of the time we would get at least 3 out 10 guesses correct just by chance (no ESP). (interpretation)***

Which conclusion does this p-value support?

A. Inconclusive, little evidence that supports $\mathbf{ESP(H_a)}$

B. Borderline, weak evidence for $\mathbf{ESP(H_a)}$

C. Strong statistically significant evidence for $\mathbf{ESP(H_a)}$

▶ Click for answer

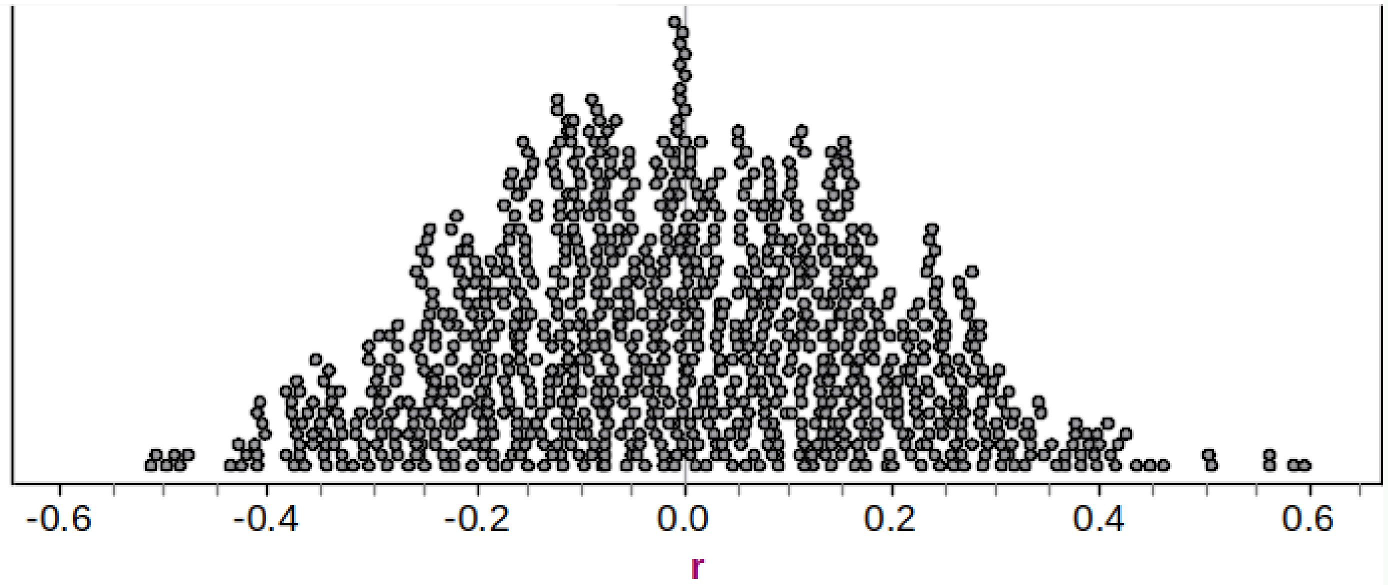
Randomization Distribution: Correlation

Using the randomization distribution below to test

$$\mathrm{H_0} : \rho = 0 \quad \text{vs} \quad \mathrm{H_a} : \rho > 0$$

Match the sample statistics: $r = 0.1$, $r = 0.3$, and $r = 0.5$

With the p-values: 0.005, 0.15, and 0.35, which sample statistic goes with which p-value?

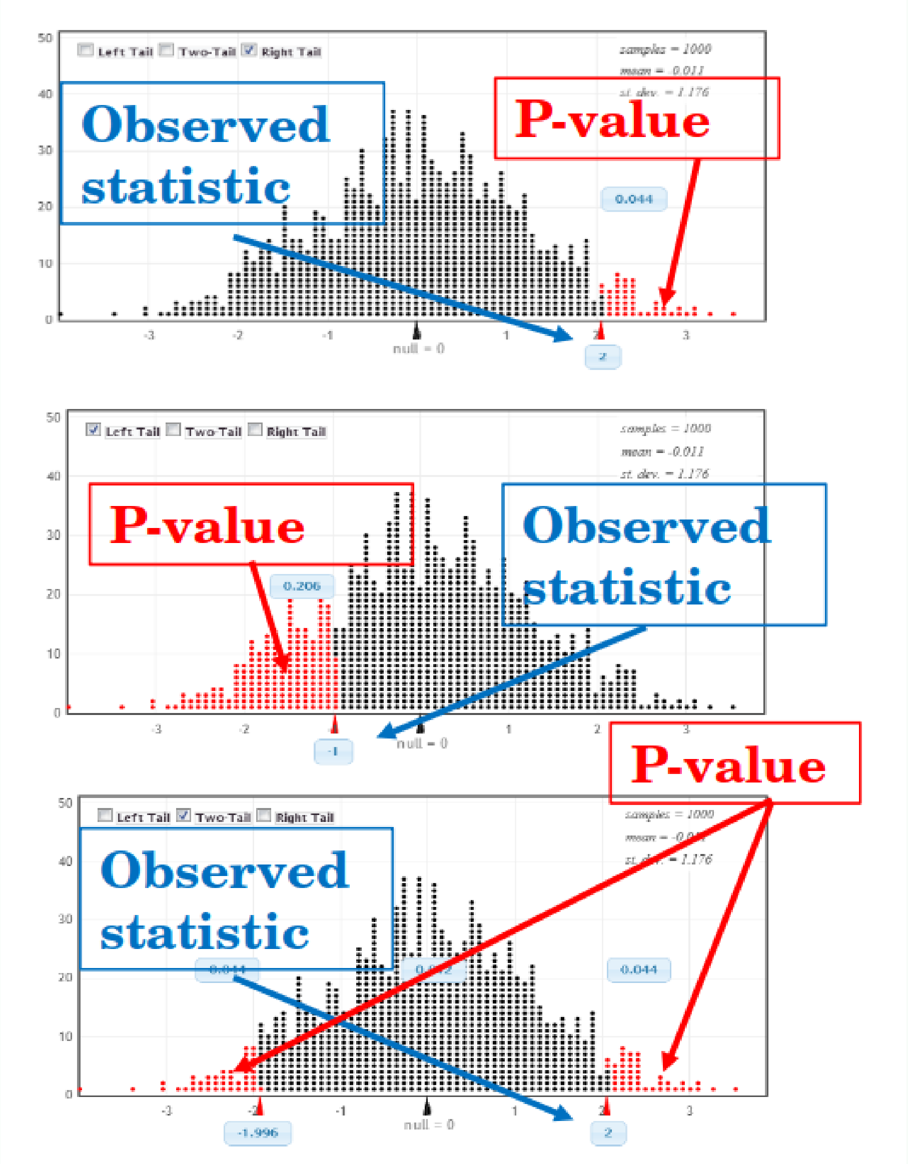


Alternative Hypothesis

- *A **one-sided** alternative contains either > or <*
- *A **two-sided** alternative contains $\neq$*
- *The p-value is the proportion in the tail in the direction specified by H_a*
- *For a two-sided alternative, the p-value is twice the proportion in the smallest tail*

p-value and H_a

Upper-tail (Right Tail)	$H_0 : \mu = 0$ $H_a : \mu > 0$ $\bar{x} = 2$
Lower-tail (Left Tail)	$H_0 : \mu = 0$ $H_a : \mu < 0$ $\bar{x} = -1$
Two-tailed	$H_0 : \mu = 0$ $H_a : \mu \neq 0$ $\bar{x} = 2$



Coccaine Addiction

In a randomized experiment, 48 cocaine addicts attempting to quit were randomly assigned to take either desipramine (a new drug) or lithium (an existing drug) to test if desipramine is more effective than lithium at treating cocaine addiction, with relapse as the response variable.

	Relapse	No Relapse	total
Desipramine	10	14	24
Lithium	18	6	24

Is desipramine more effective than lithium at treating **cocaine addiction**?

StatKey Coccaine Addiction

	Relapse	No Relapse	total
Desipramine	10	14	24
Lithium	18	6	24

- $\hat{p}_D$: proportion relapsed in Desipramine
- $\hat{p}_L$: proportion relapsed in Lithium

$$\hat{p}_D = \frac{10}{24} = 0.42 \quad \hat{p}_L = \frac{18}{24} = 0.75$$

$$\mathrm{H}_0 : p_D = p_L$$
$$\mathrm{H}_a : p_D < p_L$$

So the sample statistic is:

$$\hat{p}_D - \hat{p}_L = 0.42 - 0.75 = -0.33$$

StatKey Coccaine Addiction

	Relapse	No Relapse	total
Desipramine	10	14	24
Lithium	18	6	24

- $\hat{p}_D$: proportion relapsed in Desipramine
- $\hat{p}_L$: proportion relapsed in Lithium

$$\hat{p}_D = \frac{10}{24} = 0.42 \quad \hat{p}_L = \frac{18}{24} = 0.75$$

$$\mathrm{H}_0 : p_D = p_L$$
$$\mathrm{H}_a : p_D < p_L$$

So the sample statistic is:

$$\hat{p}_D - \hat{p}_L = 0.42 - 0.75 = -0.33$$

How extreme is -0.33, if $p_D = p_L$?

Randomization steps?

a. Randomly assign treatments (desipramine/lithium) to 48 participants.

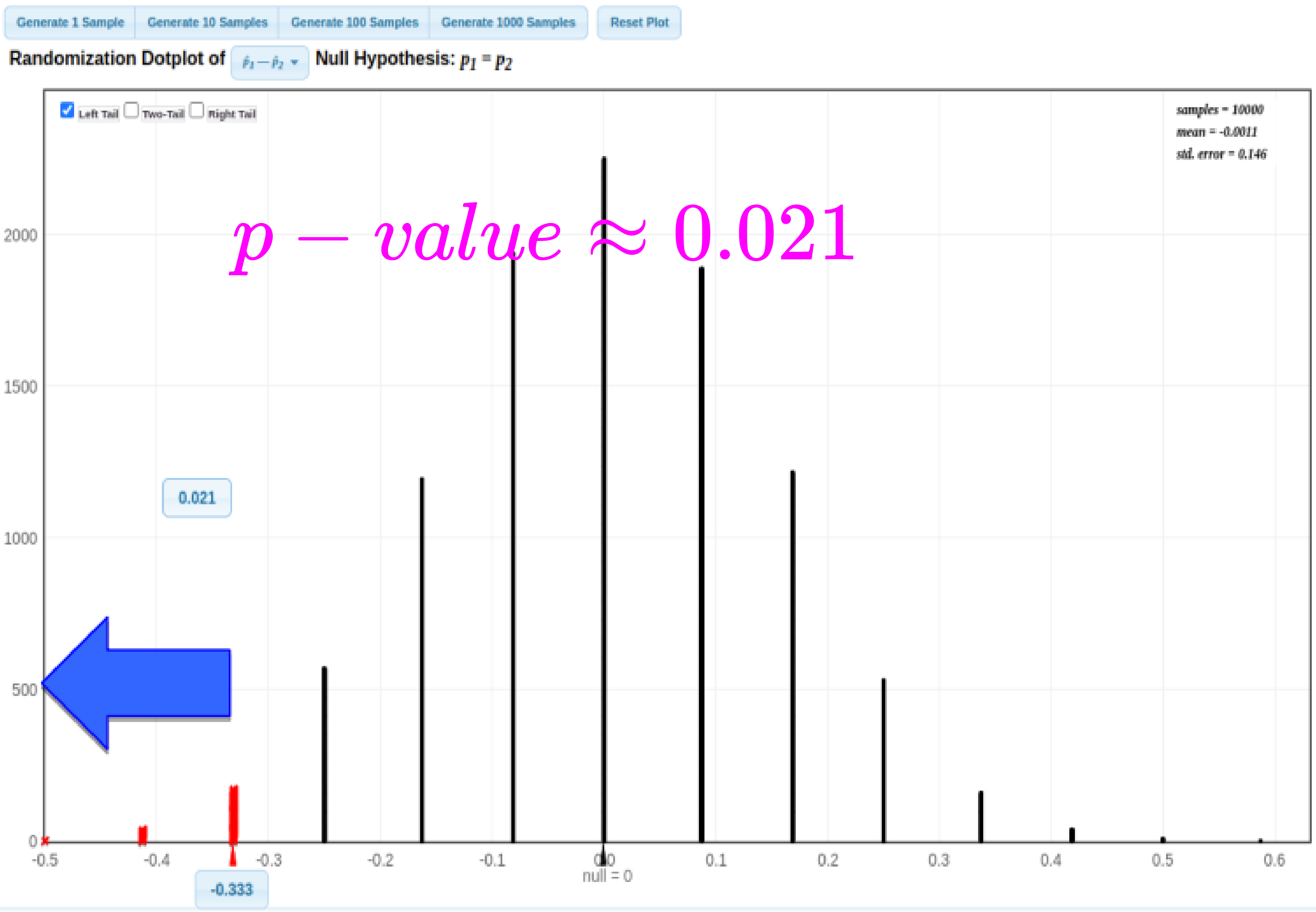
b. Create many simulated samples, assuming null hypothesis is true.

c. Shuffle outcomes (relapse/no relapse) without regard to treatment.

d. Calculate the difference in proportion of relapses for each sample.

e. Repeat steps c and d multiple times (1000+).

StatKey Coccaine Addiction



$$p - value \approx 0.021$$

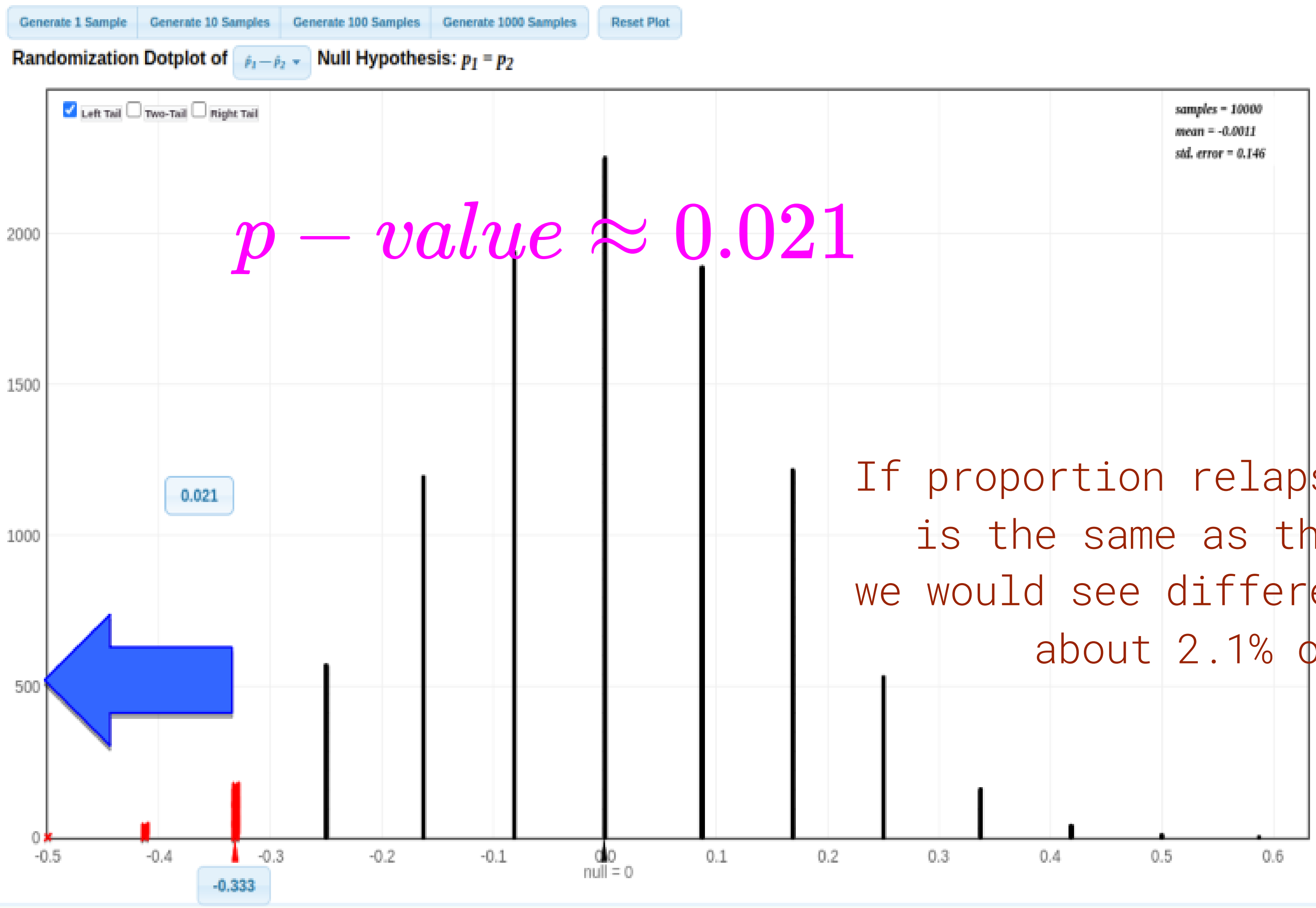
Original Sample

Group	Count	Sample Size	Proportion
Group 1	10	24	0.417
Group 2	18	24	0.750
Group 1-Group 2	-8	n/a	-0.333

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	11	24	0.458
Group 2	17	24	0.708
Group 1-Group 2	-6	n/a	-0.250

StatKey Coccaine Addiction



$$p - value \approx 0.021$$

Original Sample

Group	Count	Sample Size	Proportion
Group 1	10	24	0.417
Group 2	18	24	0.750
Group 1-Group 2	-8	n/a	-0.333

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	11	24	0.458
Group 2	17	24	0.708
Group 1-Group 2	-6	n/a	-0.250

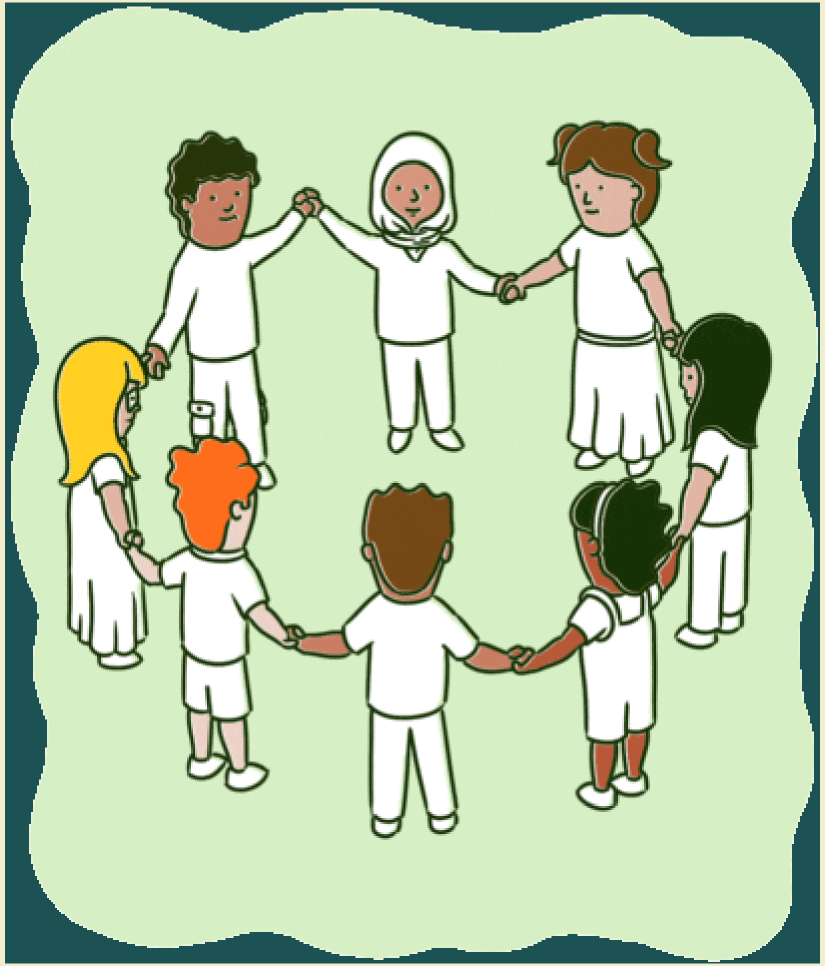
If proportion relapse for desipramine
is the same as that for lithium,
we would see differences this extreme
about 2.1% of the time

p-value and H_0

- *If the p-value is small, then a statistic as extreme as that observed would be unlikely if the null hypothesis were true, providing significant evidence against* H_0
- *The smaller the p-value, the stronger the evidence against the null hypothesis and in favor of the alternative*

YOUR TURN 1

10:00



- ***Please work on the in-class activity and we will discuss this together!***
- ***Feel free to talk to your neighbor***