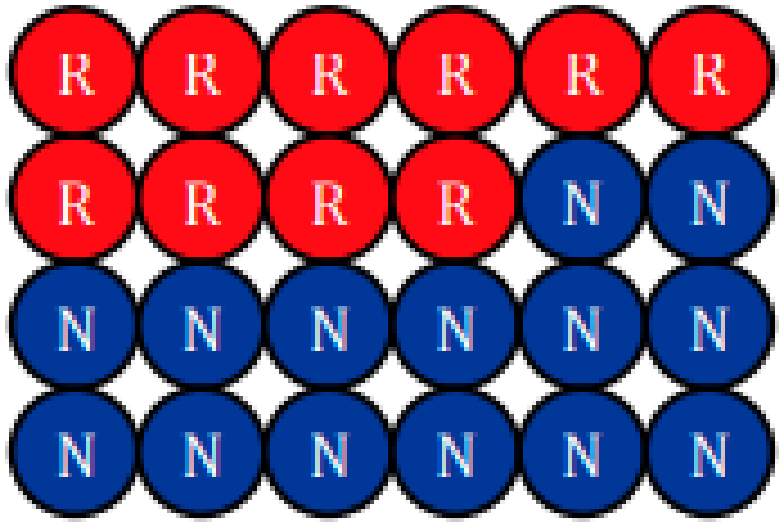
Hypothesis Testing: Significance

Stat 120

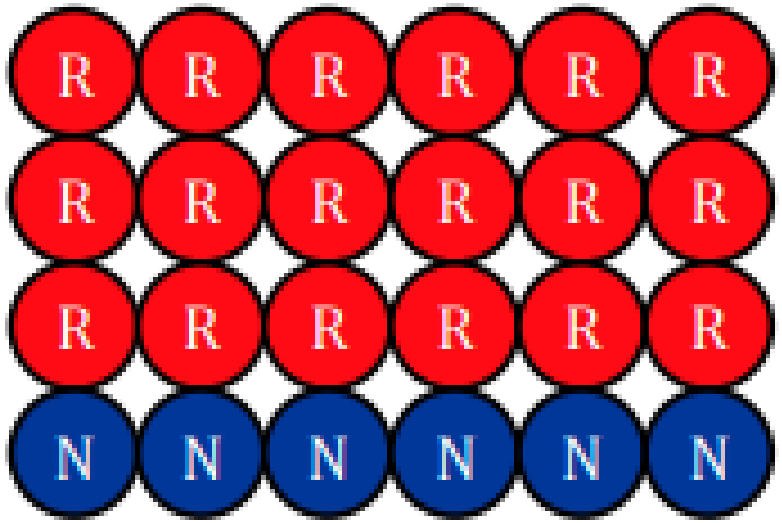
April 24 2023

Recap: Coccaine Addiction

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



10 relapse, 14 no relapse



18 relapse, 6 no relapse

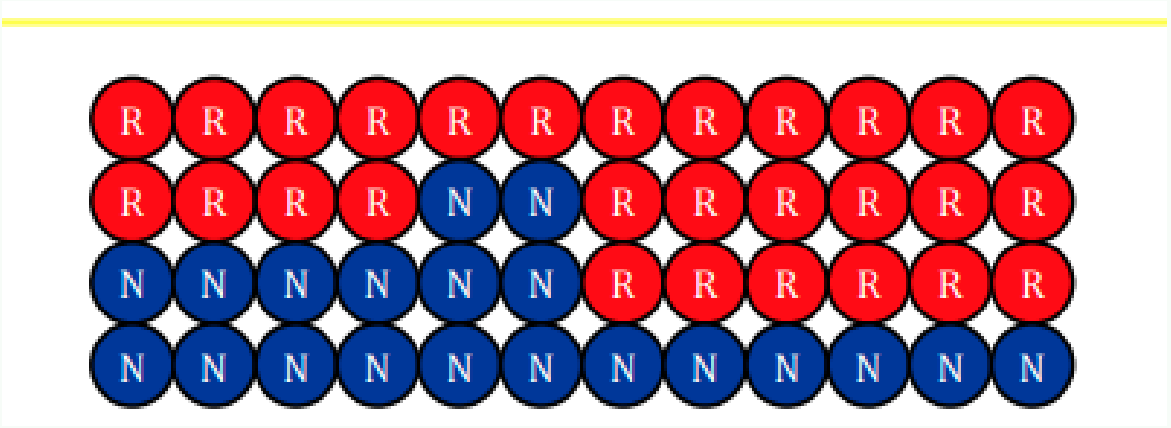
Desipramine

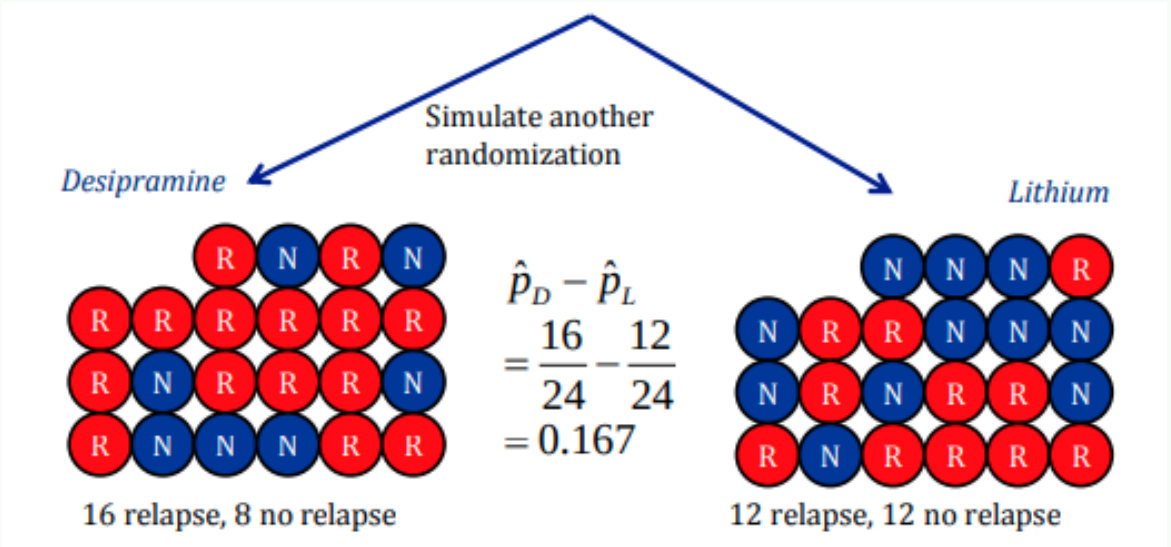
Lithium

Recap: Randomization Distribution

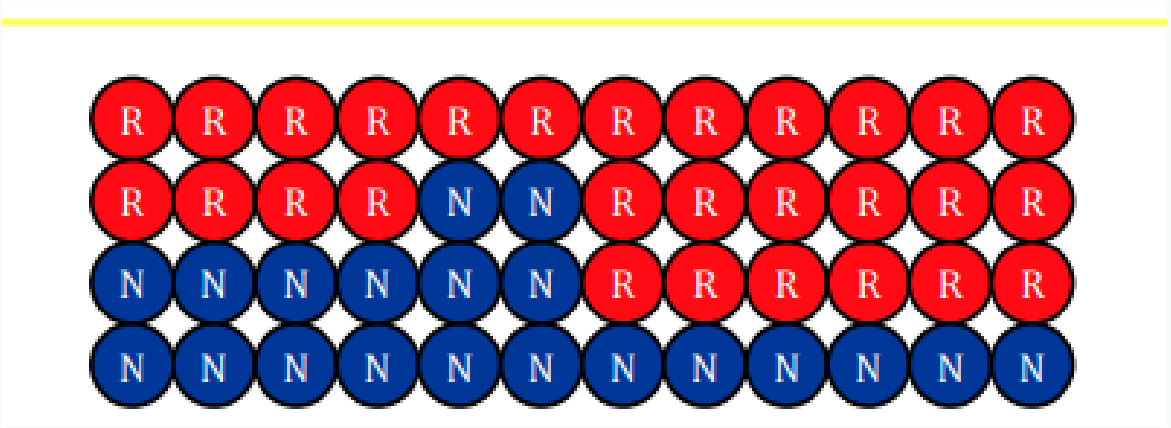
- *In the experiment, 28 people relapsed and 20 people did not relapse. Create cards or slips of paper with 28 "R" values and 20 "N" values.*
- *Pool these response values together, and randomly divide them into two groups (representing Desipramine and Lithium)*
- *Calculate your difference in proportions*
- *Plot your statistic on a dotplot like how Statkey does*
- *To create an entire randomization distribution, we simulate this process many more times with technology:*

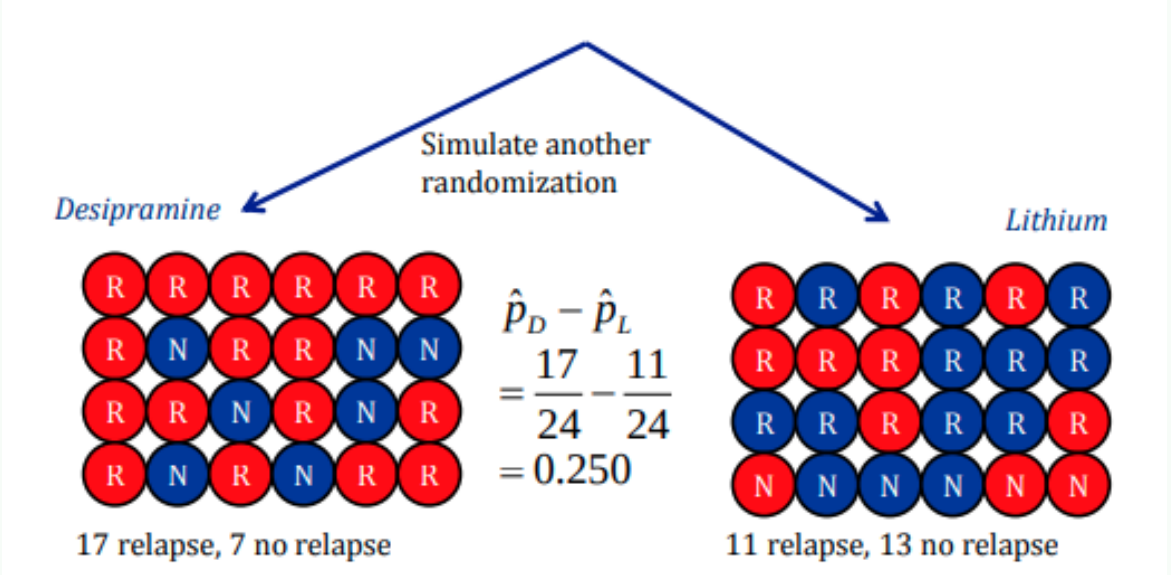
Randomization





Randomization

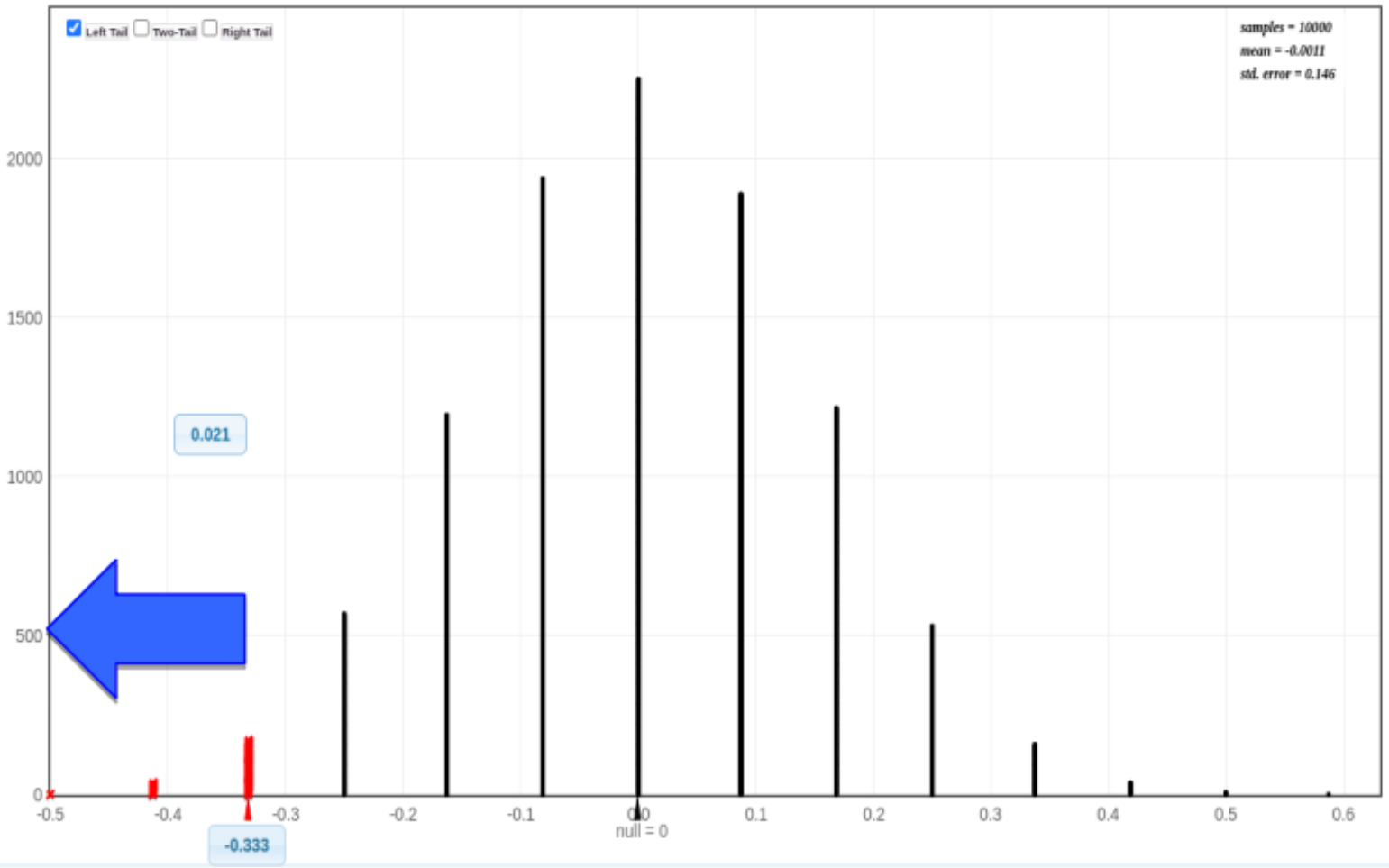




Randomization: Statkey

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

Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$



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

Formal Decisions

If the p-value is small:

- REJECT H_0
- the sample would be extreme if H_0 were true
- the results are statistically significant
- we have evidence for H_a

If the p-value is not small:

- DO NOT REJECT H_0
- the sample would not be too extreme if H_0 were true
- the results are not statistically significant
- the test is inconclusive; either H_0 or H_a may be true

Significance Level & Formal Decisions

The significance level, α is the threshold below which the p-value is deemed small enough to reject the null hypothesis (evidence is statistically significant).

$$\text{p-value} < \alpha \implies \text{Reject } H_0$$

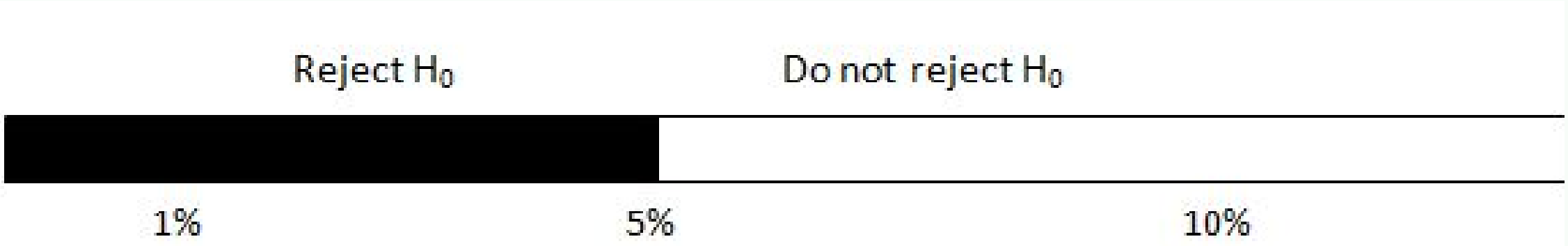
$$\text{p-value} \geq \alpha \implies \text{Do not Reject } H_0$$

Common levels:

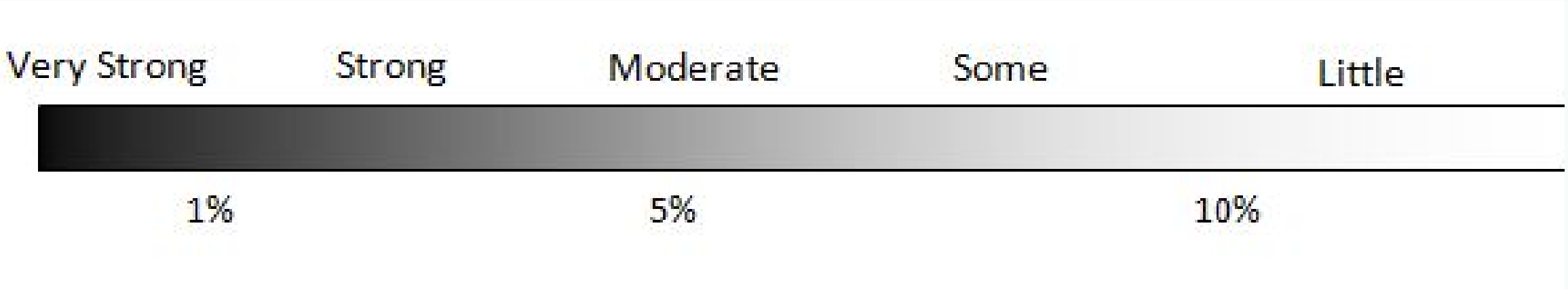
- 10% : **need some evidence to reject the null**
- 5% : **need moderate evidence to reject the null**
- 1% : **need strong evidence to reject the null**

Statistical Conclusions

Formal decision of hypothesis test, based on $\alpha = 0.05$:



Informal strength of evidence against H0:

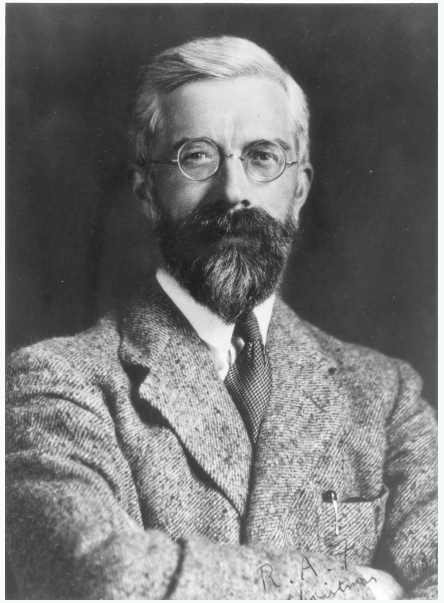


Never Accept H_0

- *"Do not reject H_0 " is not the same as "accept H_0 "!*
- *Lack of evidence against H_0 is NOT the same as evidence for H_0 !*

For the logical fallacy of believing that a hypothesis has been proved to be true, merely because it is not contradicted by the available facts, has no more right to insinuate itself in statistical than in other kinds of scientific reasoning …"

Sir R. A. Fisher



Errors in Hypothesis Testing

	Reject H_0	Do not reject H_0
H_0 true	TYPE I ERROR	😄
H_0 false	😄	TYPE II ERROR

- A Type I Error is rejecting a true null (false positive)
- A Type II Error is not rejecting a false null (false negative)

Analogy to law

A person is innocent until proven guilty.

Evidence must be beyond the shadow of a doubt

Types of mistakes in a verdict?

$$\text{Convict an innocent} \Rightarrow \text{Type I error}$$
$$\text{Release a guilty} \Rightarrow \text{Type II error}$$

Analogy to law

A person is innocent until proven guilty.

H_0

Evidence must be beyond the shadow of a doubt

Types of mistakes in a verdict?

$$\text{Convict an innocent} \Rightarrow \text{Type I error}$$
$$\text{Release a guilty} \Rightarrow \text{Type II error}$$

Analogy to law

A person is innocent until proven guilty.

H_0 H_a

Evidence must be beyond the shadow of a doubt

Types of mistakes in a verdict?

$$\text{Convict an innocent} \Rightarrow \text{Type I error}$$
$$\text{Release a guilty} \Rightarrow \text{Type II error}$$

Analogy to law

A person is innocent until proven guilty.

H_0 H_a

Evidence must be beyond the shadow of a doubt

p-value

Types of mistakes in a verdict?

$$\text{Convict an innocent} \Rightarrow \text{Type I error}$$
$$\text{Release a guilty} \Rightarrow \text{Type II error}$$

Analogy to law

A person is innocent until proven guilty.

H_0 H_a

Evidence must be beyond the shadow of a doubt

p-value α

Types of mistakes in a verdict?

$$\text{Convict an innocent} \Rightarrow \text{Type I error}$$
$$\text{Release a guilty} \Rightarrow \text{Type II error}$$

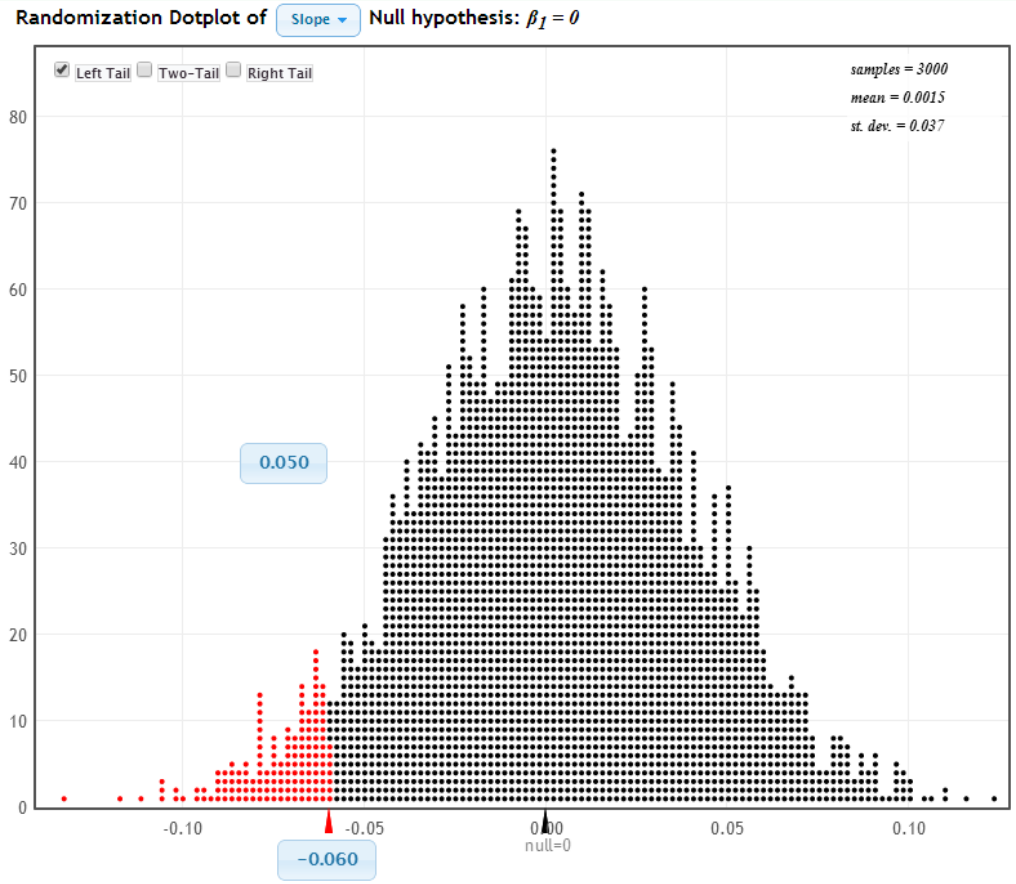
$\alpha =$ Probability of Type I Error

The significance level α controls the type I error rate.

- **Recall the Florida Lakes slope test:**

$$\mathrm{H}_0 : \beta = 0 \quad \mathrm{H}_\mathrm{a} : \beta < 0$$

- **If H_0 is true and $\alpha = 0.05$, then 5% of sample slopes will be lower red tail $(b \leq 0.06)$.**
- **5% of the sample slopes will give p-values less than 0.05, so 5% of statistics will lead to rejecting H_0 if it is true (Type I error)!!!**



Null distribution

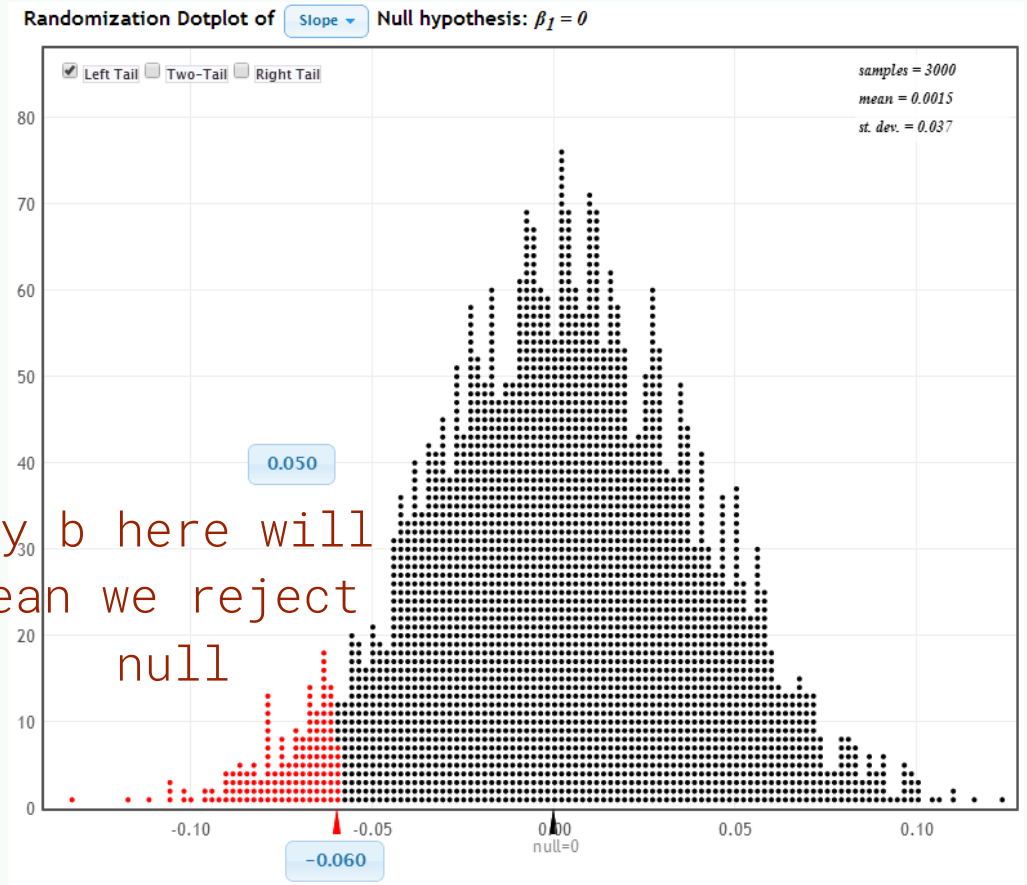
$\alpha =$ Probability of Type I Error

The significance level α controls the type I error rate.

- **Recall the Florida Lakes slope test:**

$$\mathrm{H}_0 : \beta = 0 \quad \mathrm{H}_\mathrm{a} : \beta < 0$$

- **If H_0 is true and $\alpha = 0.05$, then 5% of sample slopes will be lower red tail $(b \leq 0.06)$.**
- **5% of the sample slopes will give p-values less than 0.05, so 5% of statistics will lead to rejecting H_0 if it is true (Type I error)!!!**



Null distribution

Selecting a significance level

Decreasing α will lower your Type I error rate (makes it harder to reject the null)

- *but it will also increase your type II error rate (makes it harder to accept a true alternative)*

Selecting a significance level

If a Type I error (rejecting a true null) is much worse than a Type II error, we may choose a smaller α, like $\alpha = 0.01$ (need lots of evidence to reject null).

- E.g. sending an innocent person to jail

Selecting a significance level

If a Type II error (not rejecting a false null) is much worse than a Type I error, we may choose a larger α, like $\alpha = 0.10$

- E.g. a false negative test for a serious disease

Probability of Type II Error

Not as simple to compute since the alternative is assumed to be true

- **E.g. which value in $\mathrm{H_a} : \beta < 0$ do we select to create an "alternative" randomization distribution?**

The probability of making a Type II Error (not rejecting a false null) depends on

- **Effect size (how far the truth is from the null)**
- **Sample size (bigger n means less uncertainty)**
- **Variability of measurements**
- **Significance level (bigger α means more false positives but fewer false negatives)**

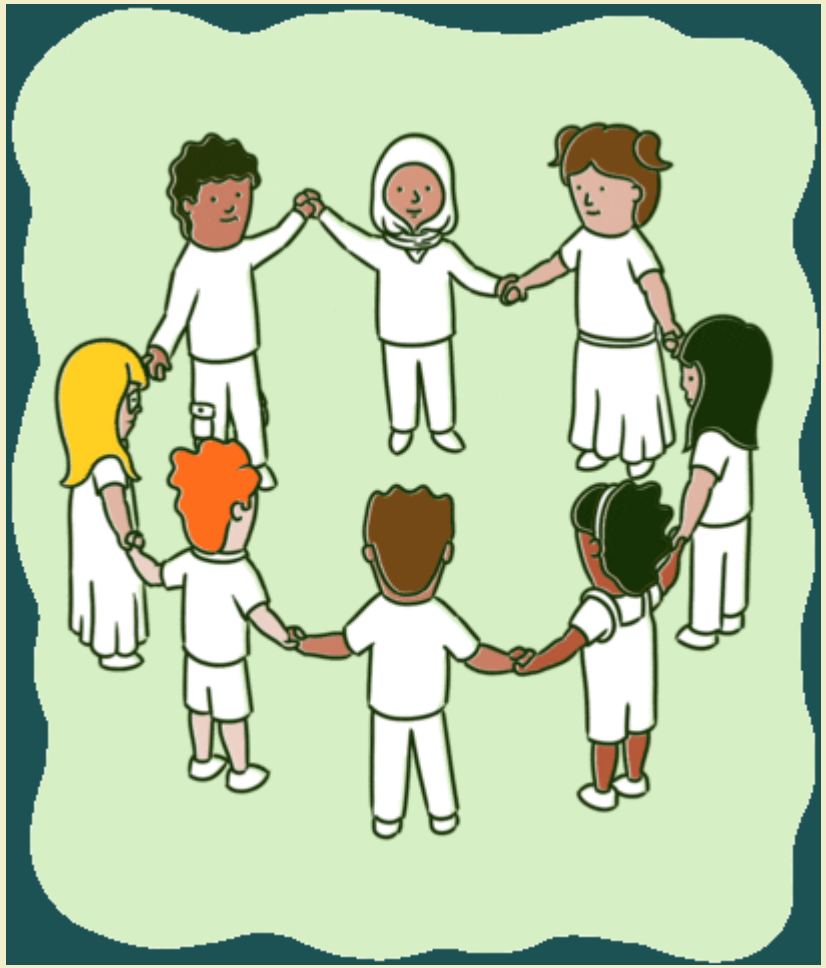
Power of a test

The power of a test is the chance that it will correctly reject the null, or

$$1 - \text{Prob}(\text{Type II error})$$

YOUR TURN 1

10:00



Please go over the class activity for today and attempt the problems.