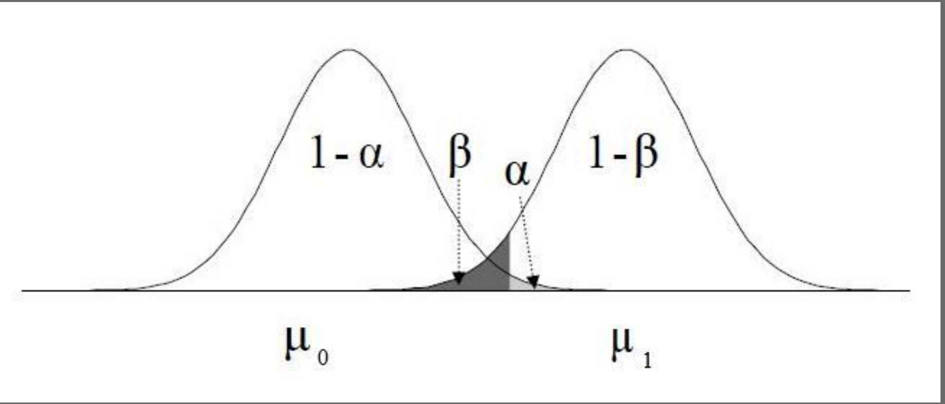
Power Calculation

Taryn Heilman
Ryan Henning
Frank Burkholder
Alex Rose

galvanize



Learning Objectives

- **Review** relevant concepts
- **Define** statistical power
- **Understand** different factors that affect power
- **Learn** how to apply power to hypothesis testing

Review

- What is the Central Limit Theorem?
- What are the implications of this for hypothesis testing?
- Describe the steps necessary to construct a hypothesis test.
- What is a p-value?
- Describe Type I and Type II Errors and name a situation where each would be the worst error to make.
- Describe the Bonferroni correction and why we would use it.

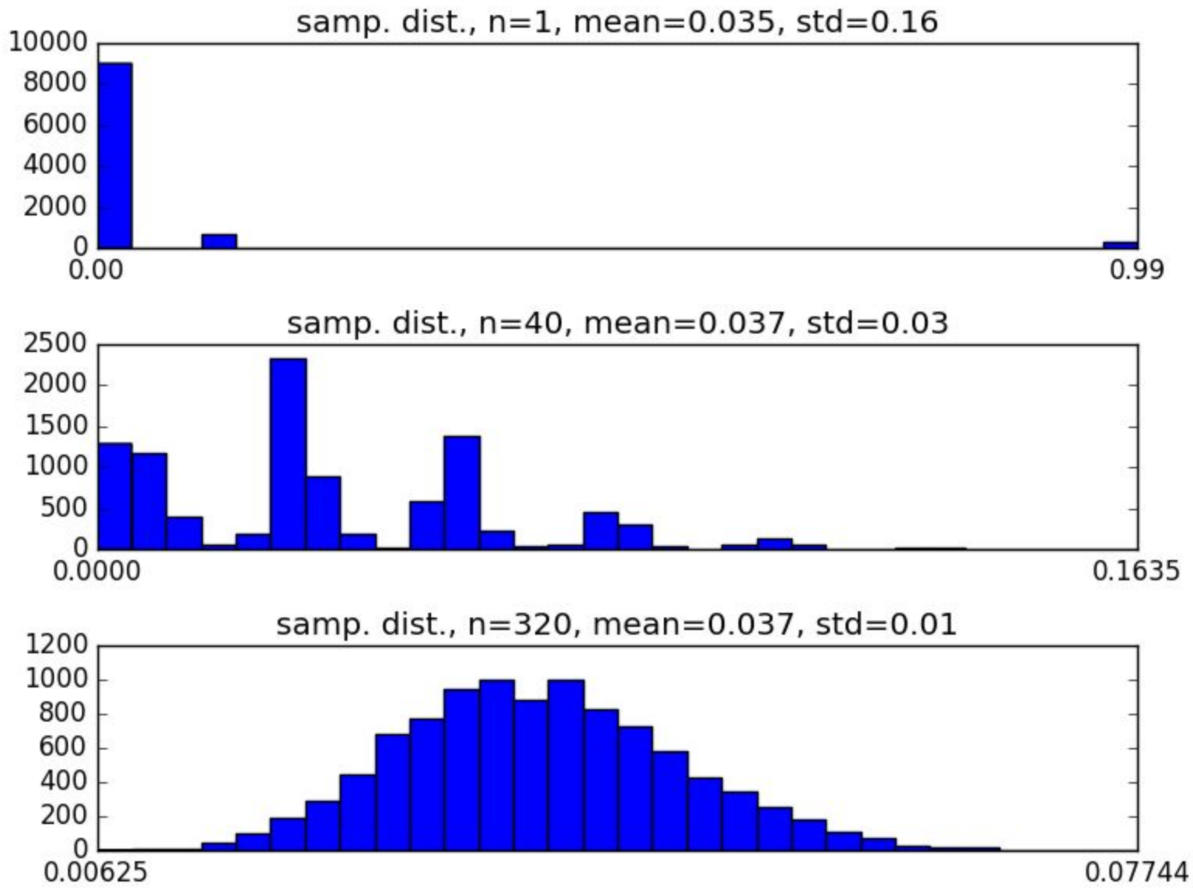
Review: Central Limit Theorem

Let the underlying distribution have mean and std. dev.

$$\mu \text{ and } \sigma$$

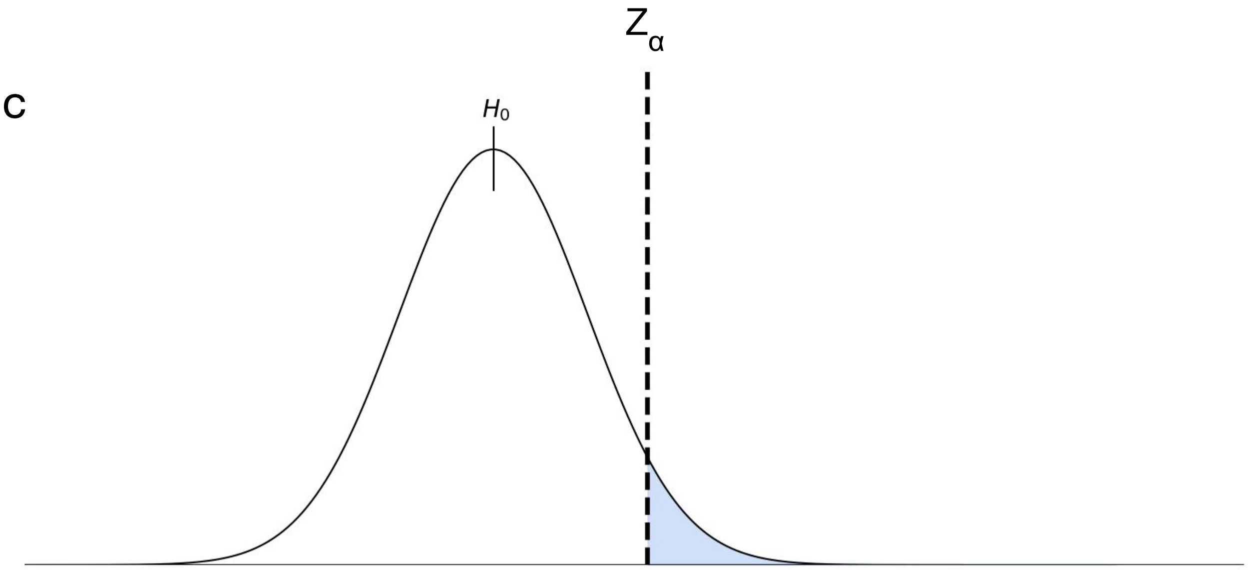
The sample mean will be normally distributed, with mean and std. dev.

$$\mu' = \mu$$
$$\sigma' = \sigma / \sqrt{n}$$



Review: Hypothesis Testing Steps

1. State the null (H_0) and the alternative (H_1) hypotheses
2. Choose a level of significance (α)
3. Collect data
4. Compute the test statistic
5. Calculate the p-value
6. Draw conclusions
   - Reject H_0
   - Fail to reject H_0



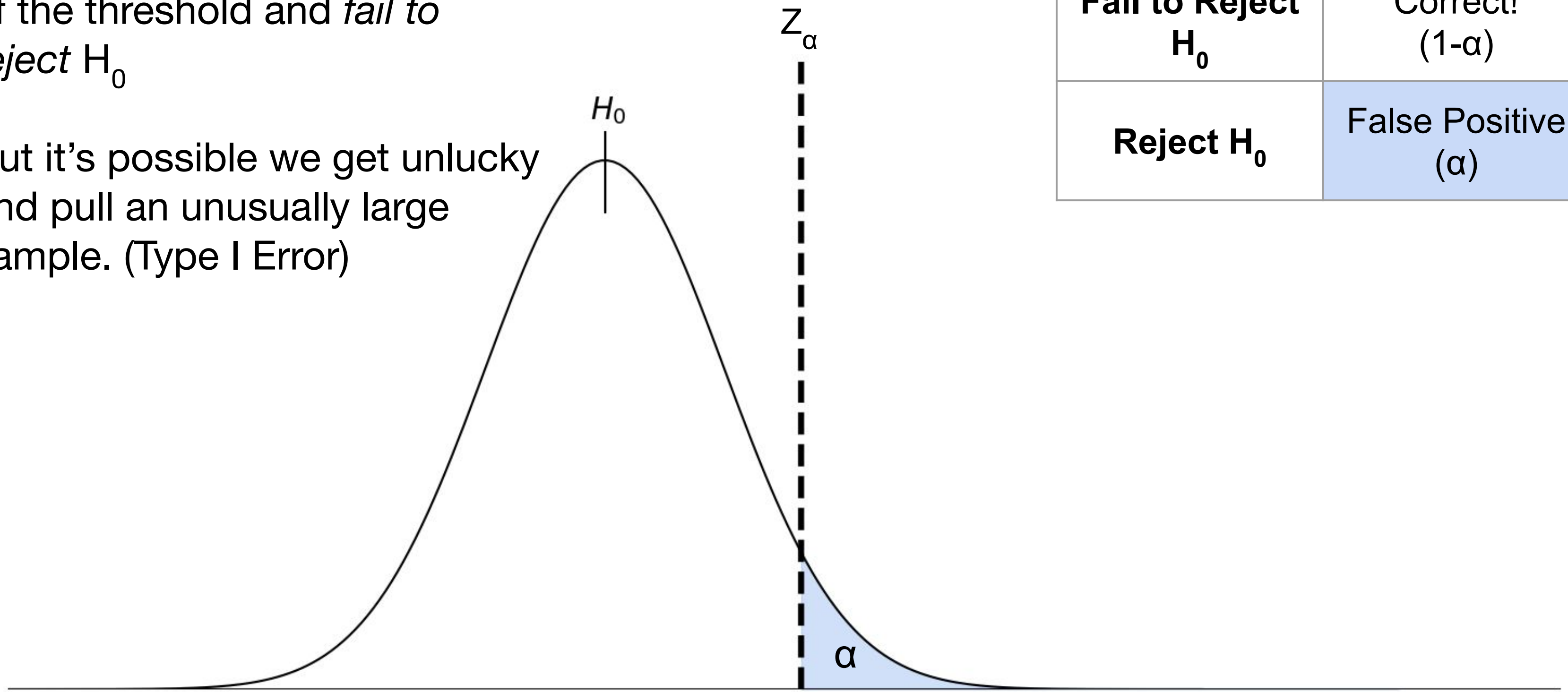
Statistical Power

galvanize

Suppose H_0 is the True Distribution

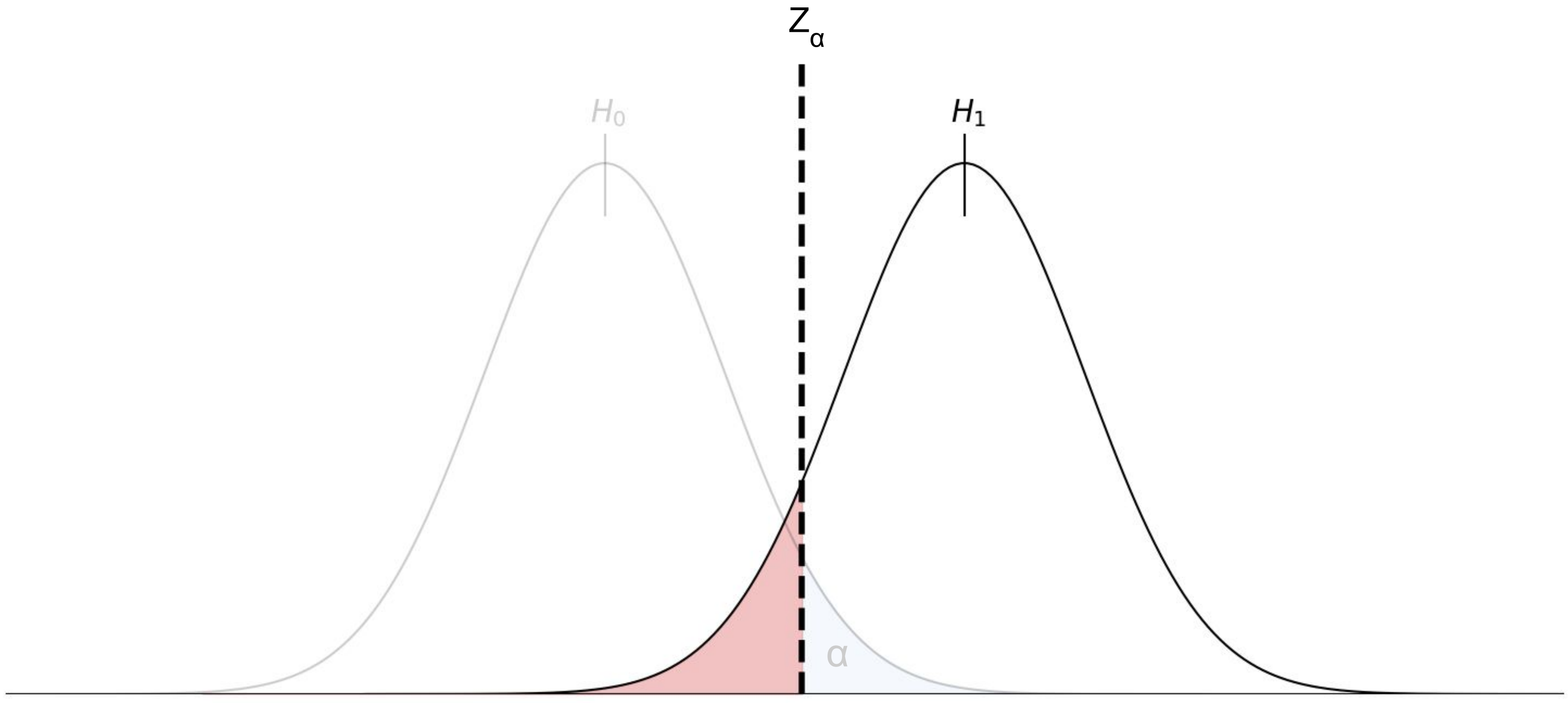
- We want to be on the left side of the threshold and *fail to reject* H_0

- But it's possible we get unlucky and pull an unusually large sample. (Type I Error)



Fail to Reject H_0	Correct! (1-α)
Reject H_0	False Positive (α)

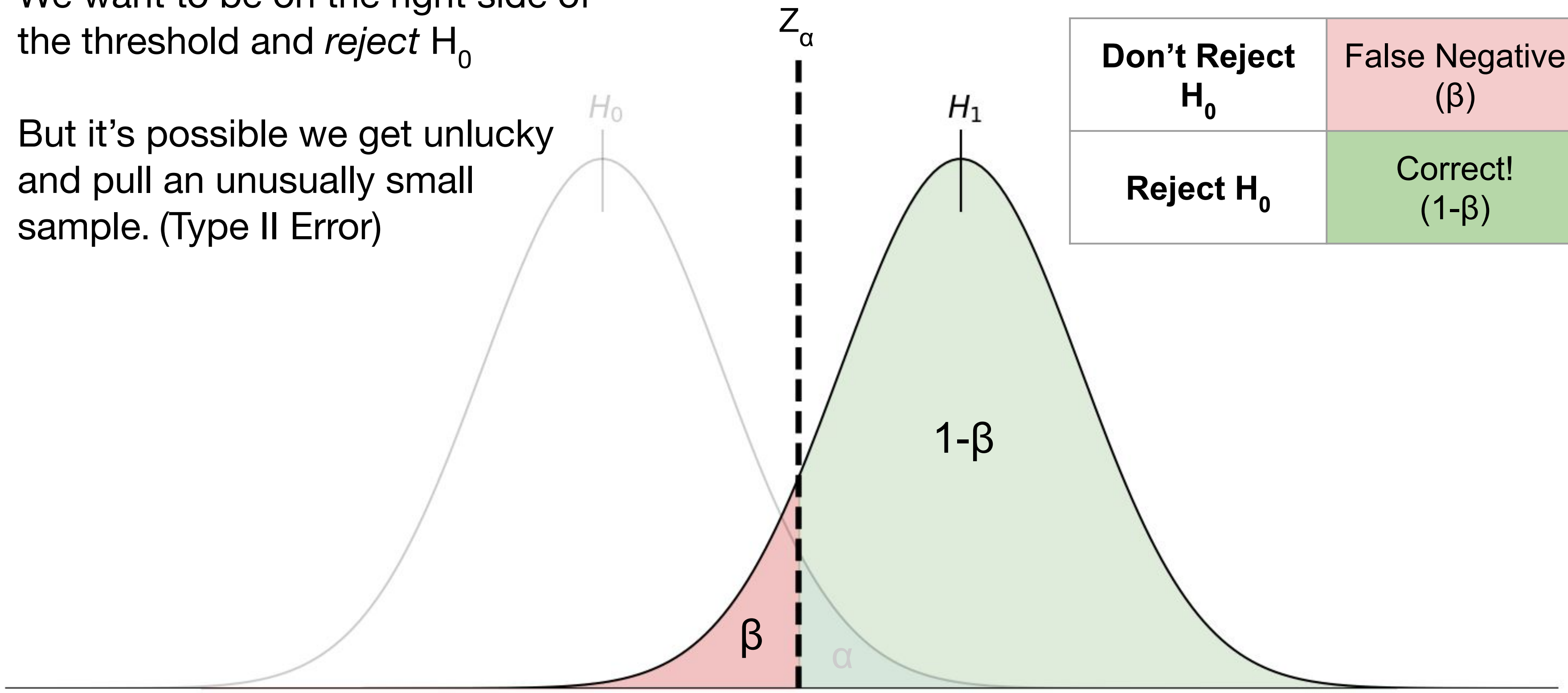
What if H_1 was the True Distribution



What if H_1 was the True Distribution

- We want to be on the right side of the threshold and *reject* H_0
- But it's possible we get unlucky and pull an unusually small sample. (Type II Error)

Don't Reject H_0	False Negative (β)
Reject H_0	Correct! ($1-\beta$)



galvanize

Power

- The “power” of our hypothesis test to detect an effect if there actually is one.
- Given that the true distribution is H_1, the probability our test rejects the (false) null hypothesis.

Power

- The “power” of our hypothesis test to detect an effect if there actually is one.
- Given that the true distribution is H_1, the probability our test rejects the (false) null hypothesis.

Power Formula for One-Tailed z-test:

$$Power = \Phi\left(\frac{\mu_1 - \mu_0}{s/\sqrt{n}} - Z_{1-\alpha}\right)$$

Hypothesis Test Outcomes

galvanize

Reality

		H_0 **is True** (Null)	H_0 **is False** (Alternate)
What we choose to do	**Don't Reject H_0**	Correct! $(1-\alpha)$	False Negative Type II Error (β)
	Reject H_0	False Positive Type I Error (α)	Correct! $(1-\beta)$

This is the experiment's **Power**: the probability that we correctly reject H_0 when the null hypothesis is false.

Aside: Other Terminology

galvanize

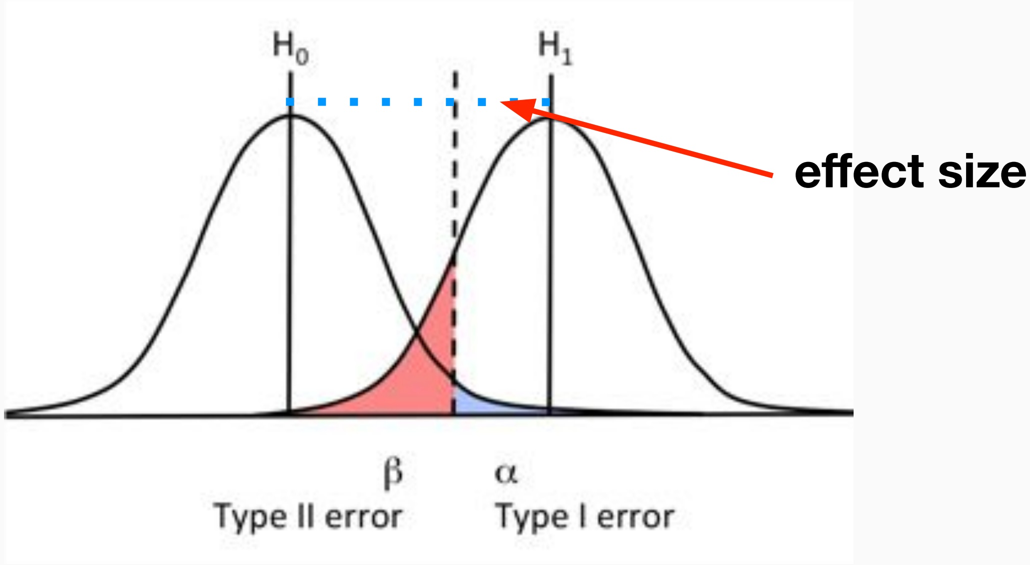
	H_0 is true true N	H_0 is false true P
Accept H_0 predict N	Correct Decision ($1-\alpha$)	Type II Error (β)
Reject H_0 predict P	FP Type I Error (α)	TP Correction Decision ($1-\beta$)

False Positive Rate
= FP / N
(aka, 1 - specificity)

True Positive Rate
= TP / P
(aka, sensitivity)

Effect Size

- Power is in relation to a *specific* alternate hypothesis.
- The **effect size** is how far the mean of the alternate hypothesis is away from the mean of the null.



Breakout: Factors that Affect Power

galvanize

- Effect Size
- Significance Level (α)
- Sample Size (n)
- Beta

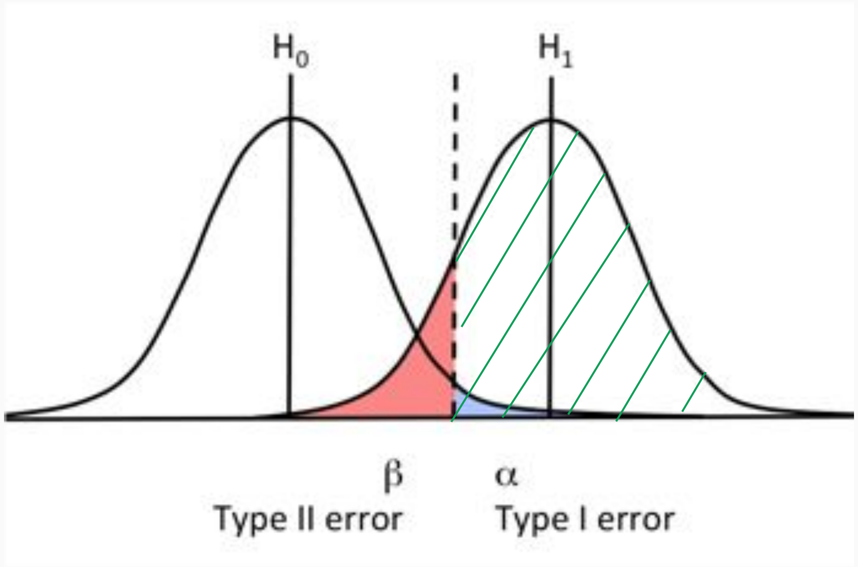
How does varying each of these affect Power?

You may use this interactive tool to help answer!

Interactive exploration

http://rpsychologist.com/d3/NHST/

Power In Practice: Required Sample Size



	H_0 is True	H_0 is False
Don't Reject H_0	Correct! (1-α)	False Negative (β)
Reject H_0	False Positive (α)	Correct! (1-β)

Often, we know:

1. **The "effect size" that we want to detect, and**
2. **The *power* that we want to achieve.**

We then calculate the *sample size* needed to get what we want!

Hypothesis Testing Steps: Revised

1. State the null (H_0) and the alternative (H_1) hypotheses
2. Choose a level of significance (α) and power (1 - β) (typically 80%)
    i. Compute the number of samples required for your desired α, power and effect size.
3. Collect data
4. Compute the test statistic
5. Calculate p-value
6. Draw conclusions

- Reject H_0
- Fail to reject H_0

Calculating the Required Sample Size (2 means)

$$n > \left((Z_{(1-\beta)} - Z_{\alpha}) \frac{s}{\mu_b - \mu_a} \right)^2$$

FYI: [derivation]

```
from scipy import stats

alpha = 0.05 # allowable Type I error rate (incorrectly rejecting H0)
beta = 0.2   # allowable Type II error rate (failing to reject H0 when we should)
power = 1 - beta

mu_a = val_a  # the mean value of a
mu_b = val_b  # the mean value of b
s = val_s     # effective standard deviation of the difference between a & b distributions

n = ((stats.norm.ppf(1-beta) - stats.norm.ppf(alpha)) * s / (mu_b - mu_a))**2
```

Breakout 2: Calculate Required Sample Size

Setup: A/B Test our website's homepage.

Our current homepage has a signup conversion rate of 6%. (The standard deviation s would be 0.24)

We want to test a new homepage design to see if we can get a 7% signup rate. We'll want an experiment where alpha is 1% and power is 95%.

How many visitors must visit the new homepage in order to fulfill the requirements of this experiment?

Breakout 2: Calculate Required Sample Size

galvanize

Setup: A/B Test our website's homepage.

Our current homepage has a signup conversion rate of 6%. (The standard deviation s would be 0.24)

We want to test a new homepage design to see if we can get a 7% signup rate. We'll want an experiment where alpha is 1% and power is 95%.

How many visitors must visit the new homepage in order to fulfill the requirements of this experiment?

$$n \geq 9,084$$

Breakout 2: Calculate Required Sample Size

Setup: A/B Test our website's homepage.

Our current homepage has a signup conversion rate of 1%. (The standard deviation s would be 0.10)

We want to test a new homepage design to see if we can get a 1.2% signup rate. We'll want an experiment where alpha is 1% and power is 95%.

How many visitors must visit the new homepage in order to fulfill the requirements of this experiment?

Breakout 2: Calculate Required Sample Size

Setup: A/B Test our website's homepage.

Our current homepage has a signup conversion rate of 1%. (The standard deviation s would be 0.10)

We want to test a new homepage design to see if we can get a 1.2% signup rate. We'll want an experiment where alpha is 1% and power is 95%.

How many visitors must visit the new homepage in order to fulfill the requirements of this experiment?

$$n \geq 39,427$$

Breakout 3: Calculating Power

Setup: Testing a New Treatment

A hospital has a 70% success rate when treating a rare disease (The standard deviation s would be .46)

They want to test a new treatment plan and see if they can obtain a 75% success rate. They're willing to tolerate an alpha of 10%, but they currently only have 100 patients available for this extended medical trial.

What would be the power of their trial? Do you think they should run it?

Breakout 3: Calculating Power

Setup: Testing a New Treatment

A hospital has a 70% success rate when treating a rare disease (The standard deviation s would be .46)

They want to test a new treatment plan and see if they can obtain a 75% success rate. They're willing to tolerate an alpha of 10%, but they currently only have 100 patients available for this extended medical trial.

What would be the power of their trial? Do you think they should run it?

Power = 42%

Learning Objectives

- **Review** relevant concepts
- **Define** statistical power
- **Understand** different factors that affect power
- **Learn** how to calculate power and how to apply it to hypothesis testing