Power Calculations

Joe

galvanize

Introduction

galvanize

Morning Objectives

1. Define Power and relate it to type 2 error

2. Compute Power for a dataset and a given problem

3. Demonstrate the relationship between sample size, effect size, and significance on power

galvanize

Review of Key Concepts

galvanize

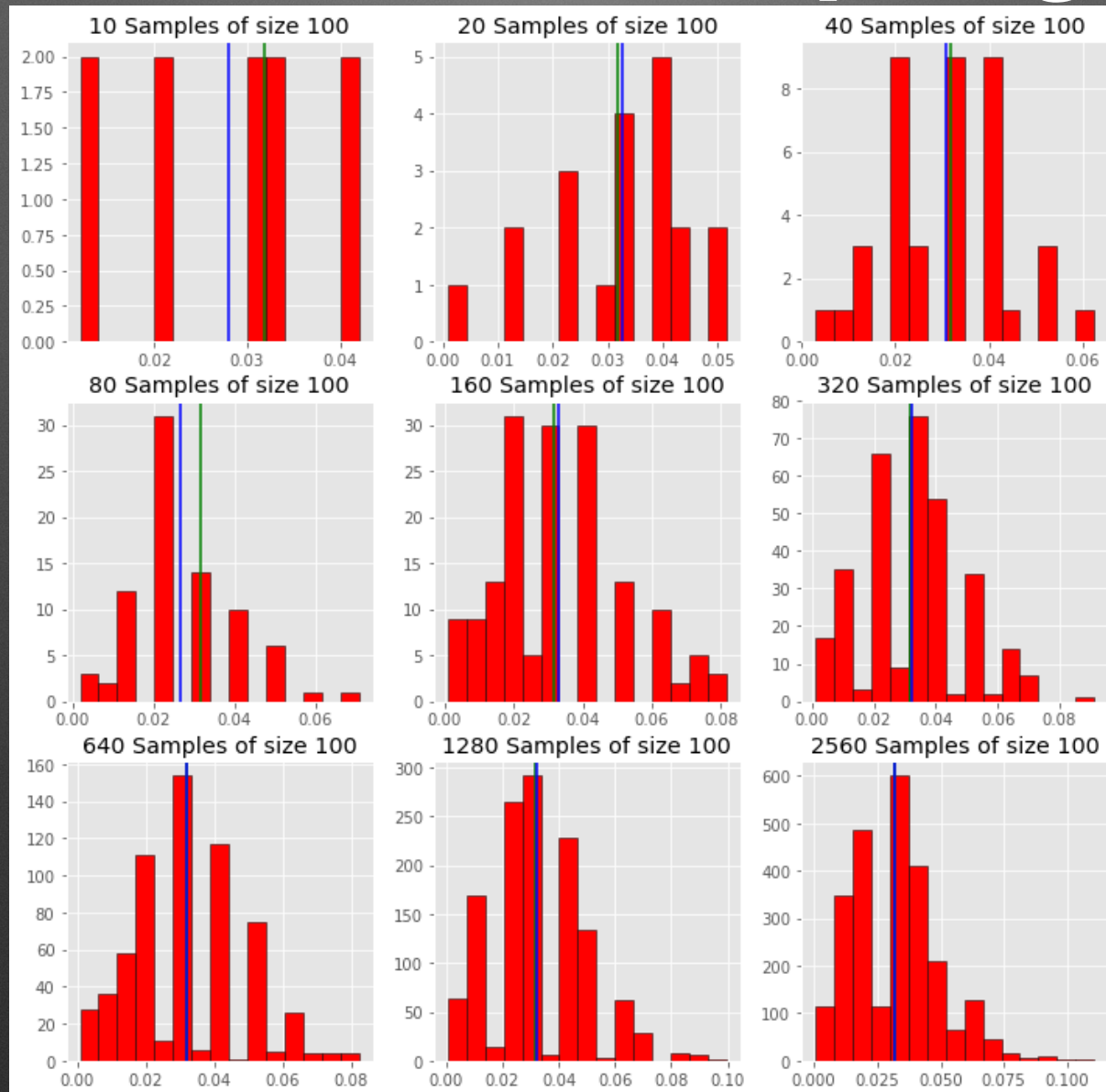# Recall: Online Spending from CLT Lecture

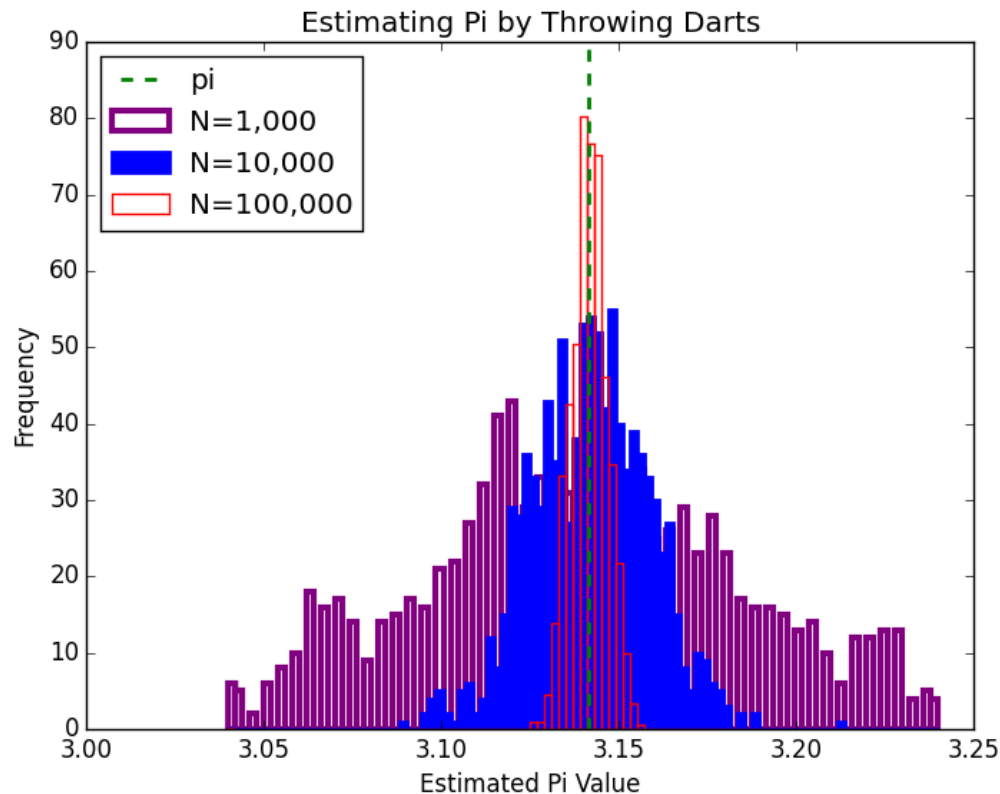| Revenue Per Visitor	P(X)
X = $0.00	90%
X = $0.10 (ad-click)	7%
X = $0.99 (app purchase)	3%
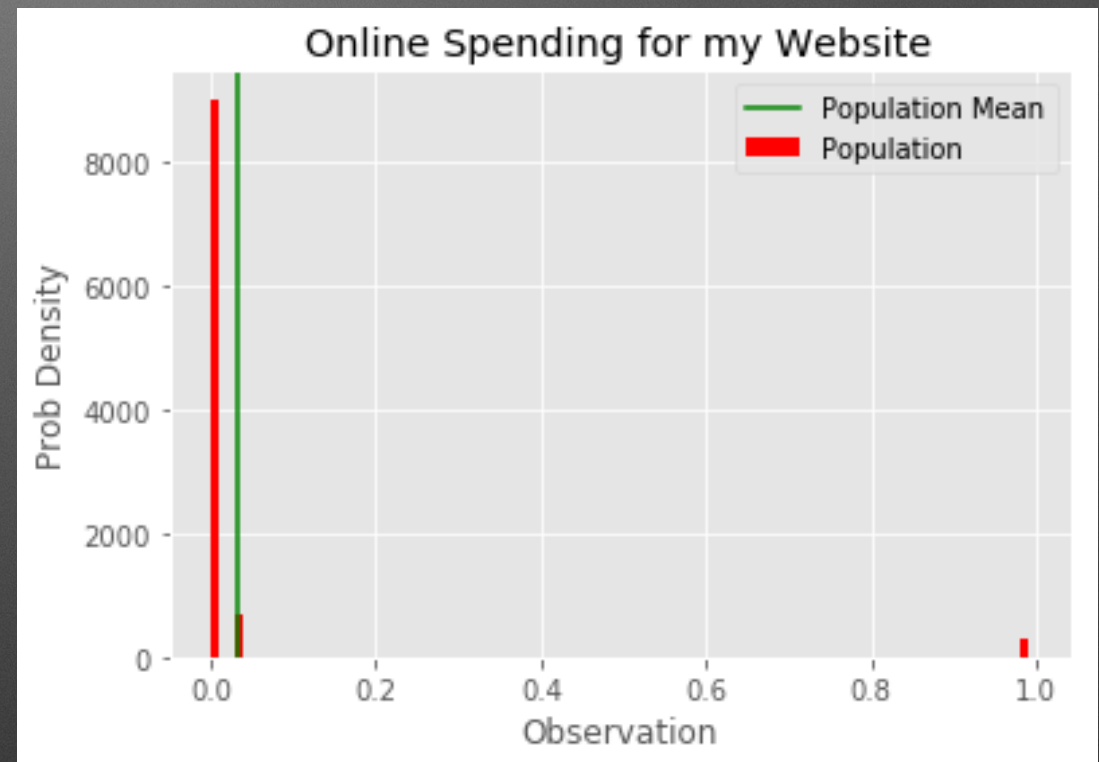
What does the CLT say about the mean of a sample?
What does the CLT say about the variance of the sample mean?

galvanize

Random Sampling

CLT in Action

Estimating Pi by Throwing Darts

Legend:
- pi
- N=1,000
- N=10,000
- N=100,000

CLT States that the mean of a distribution is normally distributed, with:

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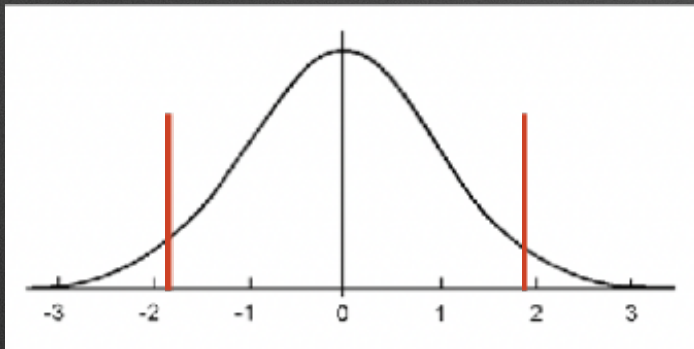$$\sigma' = \sigma/\sqrt{n}$$

where n is the sample size

galvanize

Recall: Hypothesis Testing One Tail vs Two Tail Tests
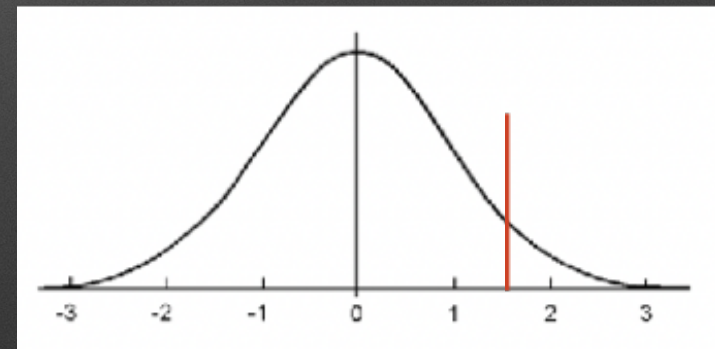
Two Sided Test:
$H_0 : \mu = 0$
$H_A : \mu \neq 0$

Two Sided Test:
$H_0 : \mu = 0$
$H_A : \mu \geq 0$

$z_{\alpha/2} = [-1.96, 1.96]$
For $\alpha = 0.05$

$z_{\alpha/2} = 1.65$
For $\alpha = 0.05$

galvanize

Recall: Error Types

Type 1:
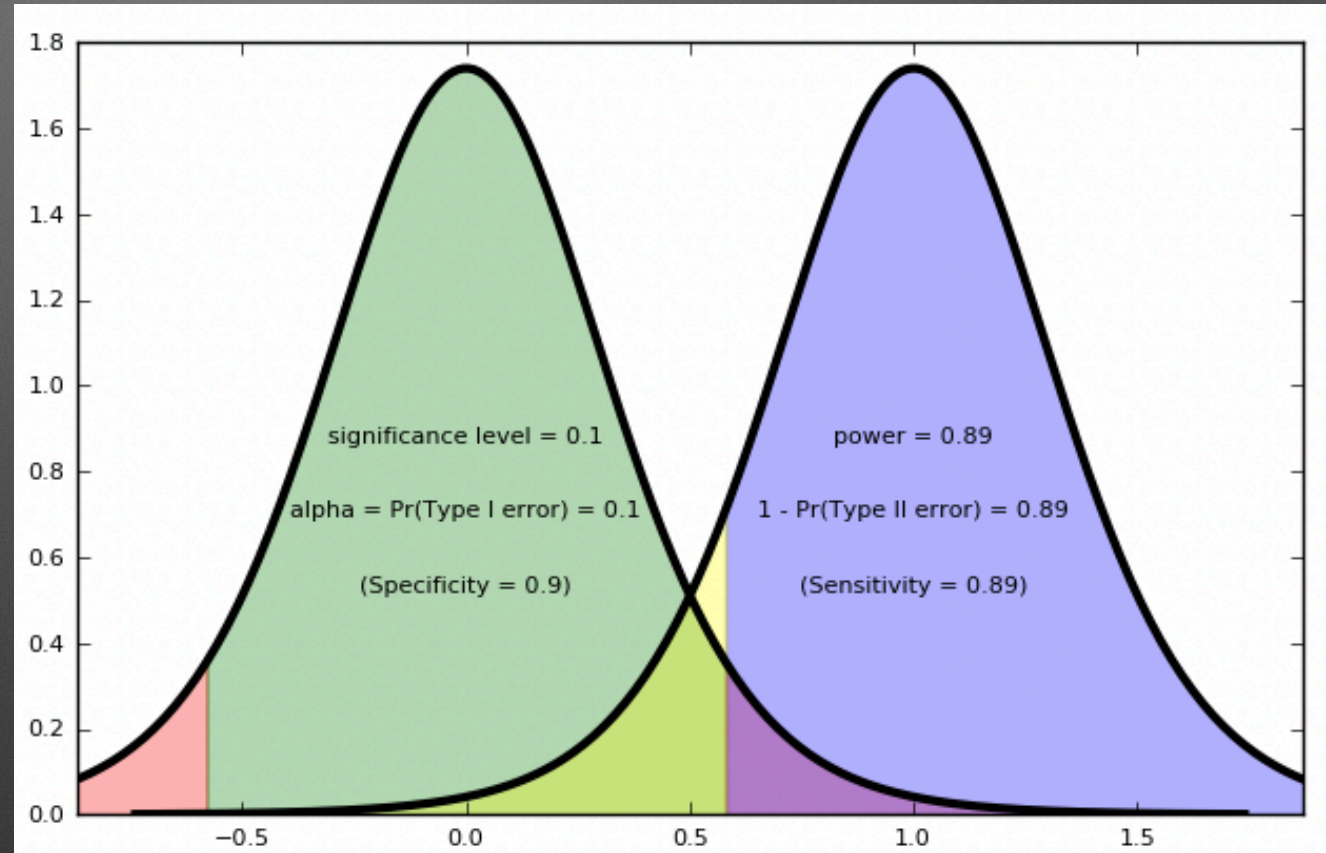
α =

Type 2:

β =

galvanize
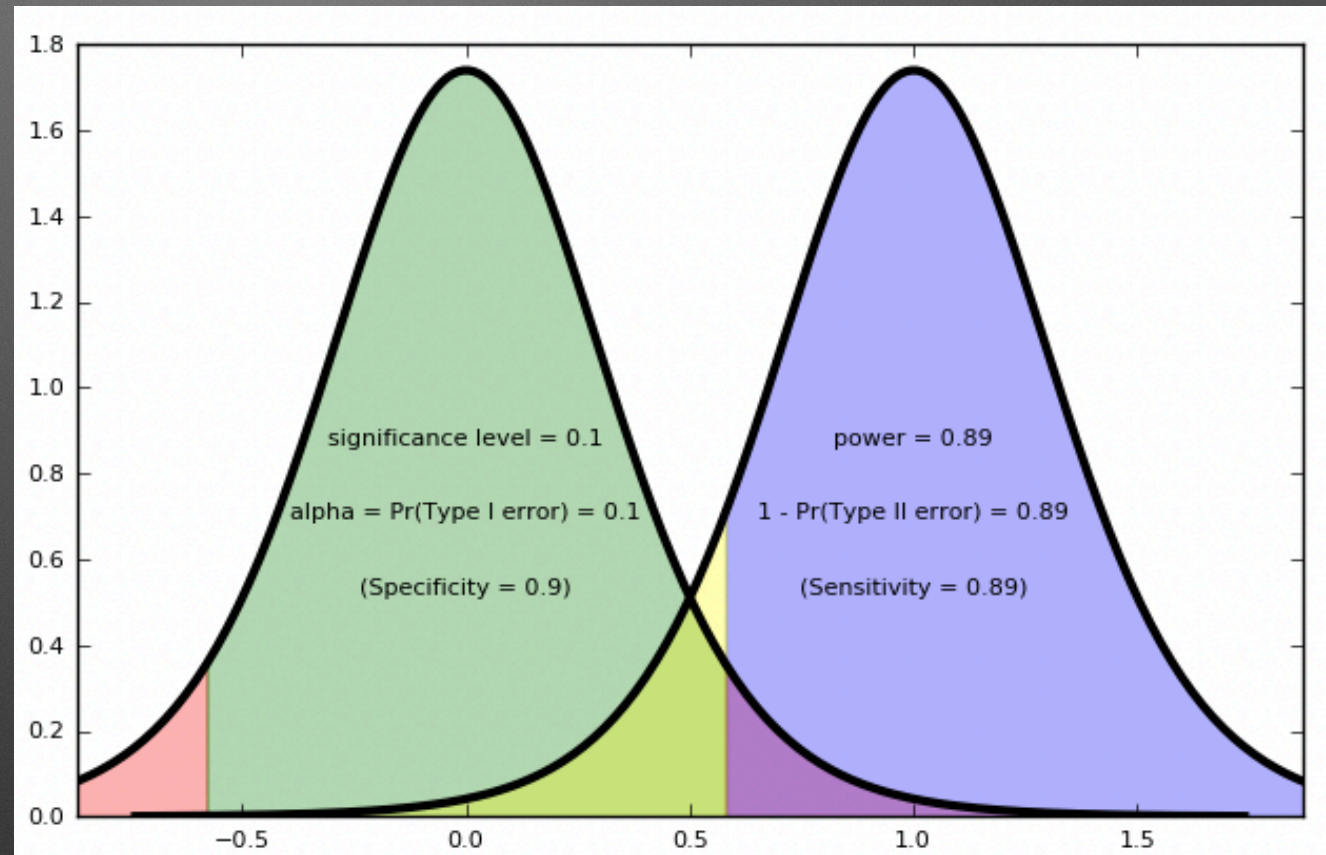
Recall: Error Types

Type 1:
$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$

Type 2:
$\beta = P(\text{Fail to reject } H0 \mid H_A \text{ is true})$

significance level = 0.1

alpha = Pr(Type I error) = 0.1

(Specificity = 0.9)

power = 0.89

1 - Pr(Type II error) = 0.89

(Sensitivity = 0.89)

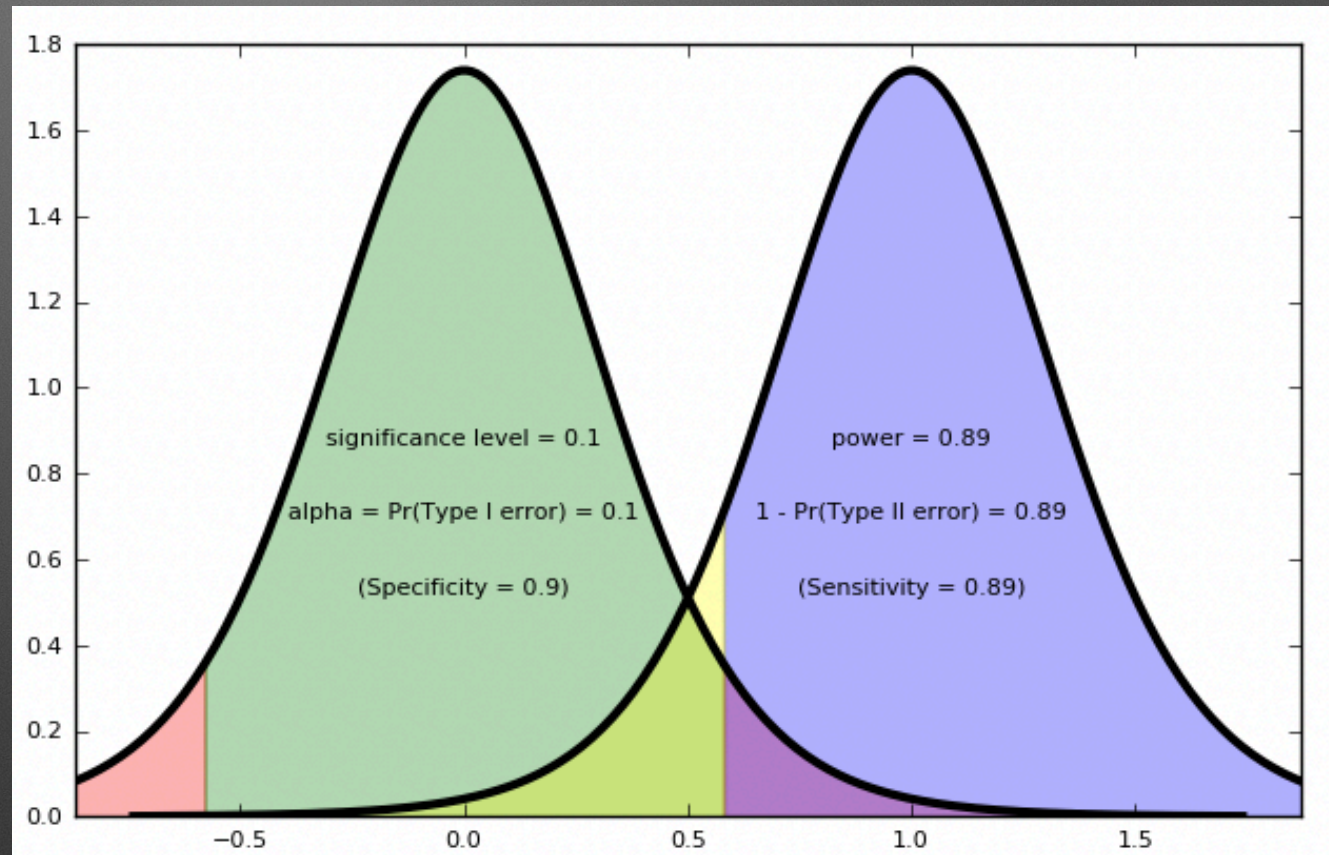galvanize

Recall: Error Types

Type 1:
$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$
We call α our significance level

Type 2:
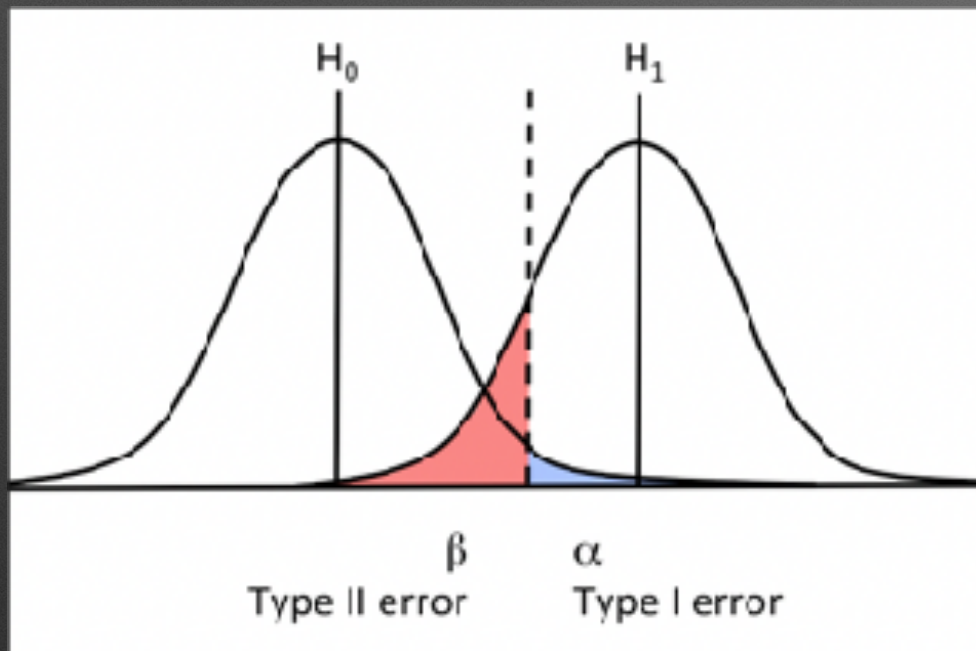$\beta = P(\text{Fail to reject } H0 \mid H_A \text{ is true})$
We call $(1-\beta)$ our test power

galvanize

Power Region

galvanize

Type 2 Error Region

	H_0 Is True	H_a Is True
Fail To Reject H_0	Correct Decision (1 - α)	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision (1 - β)

Test power is always defined w.r.t. a specific alternate hypothesis.
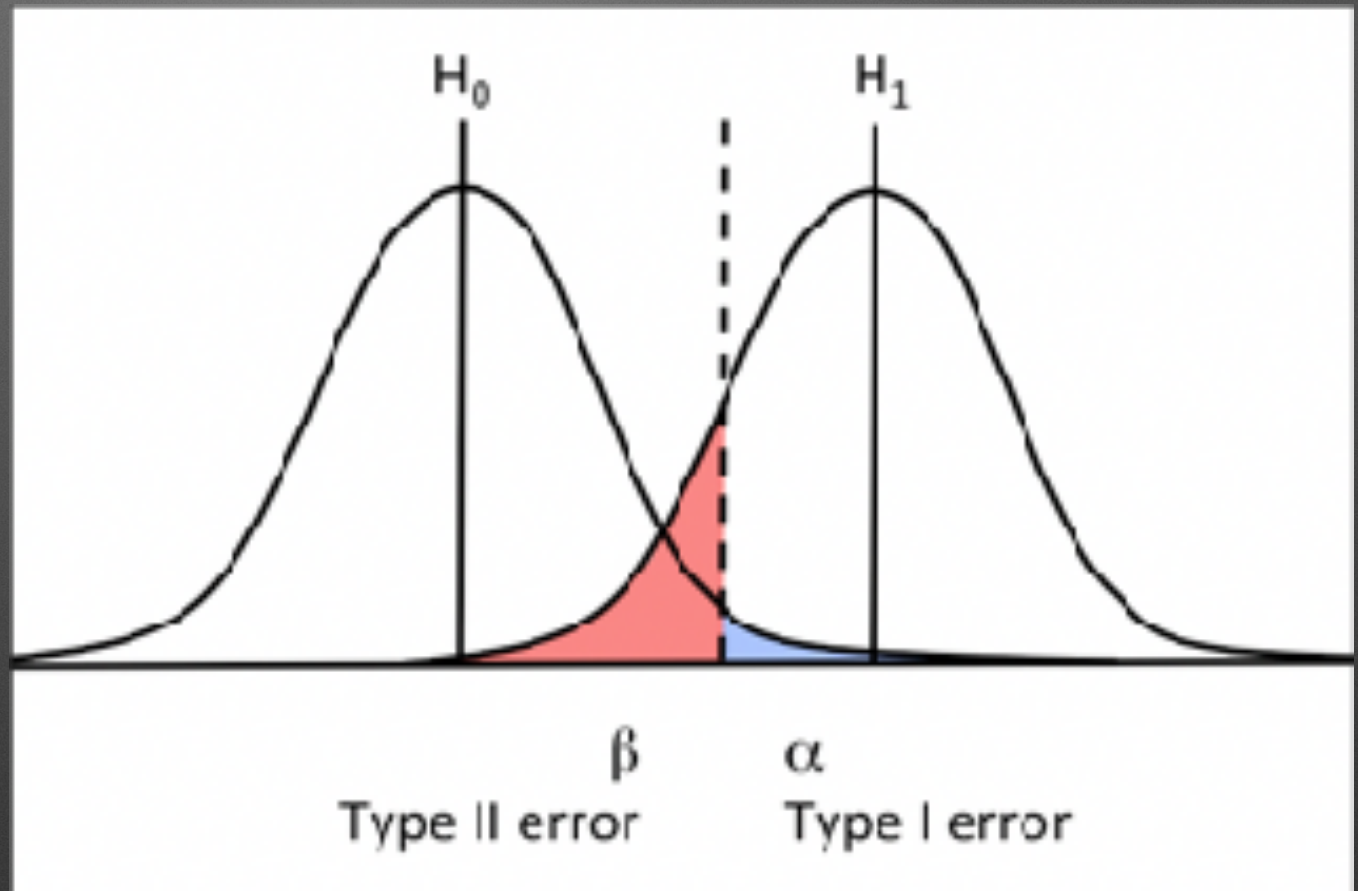
In pink, we have what is called the error region.

Let's hop into notebook 0 and do a quick calculation.

13

galvanize

Z Test Error Region

Majority of hypothesis testing occurs via CLT (i.e. normal approximation of the population mean).

What region should we hope to minimize/eliminate?

What 'controls' do we have to do such a thing?

galvanize

Z Test Error Region

Majority of hypothesis testing occurs via CLT (i.e. normal approximation of the population mean).

What region should we hope to minimize/eliminate?

What 'controls' do we have to do such a thing?

sample size, effect size, and significance - back to the notebook!

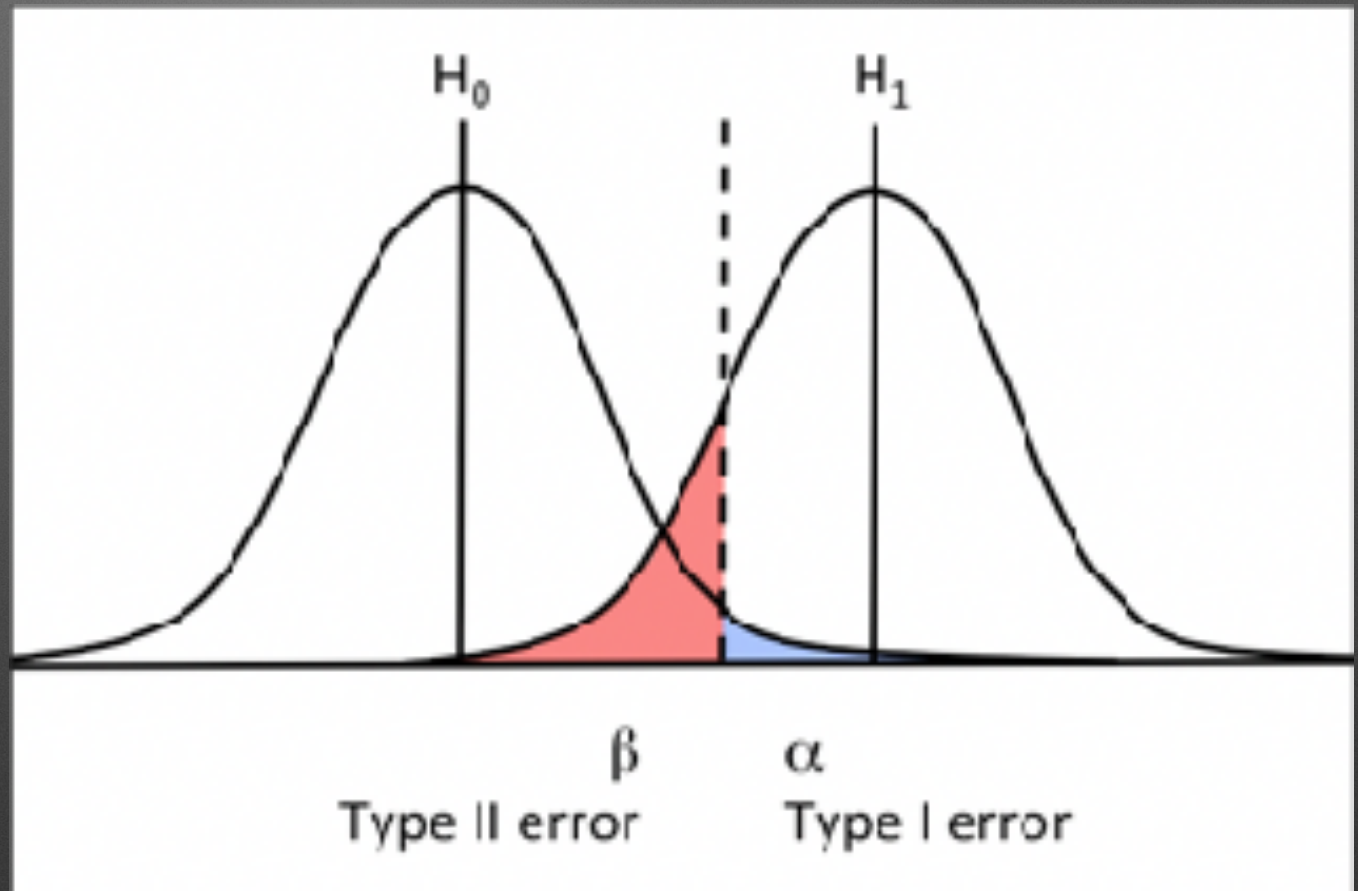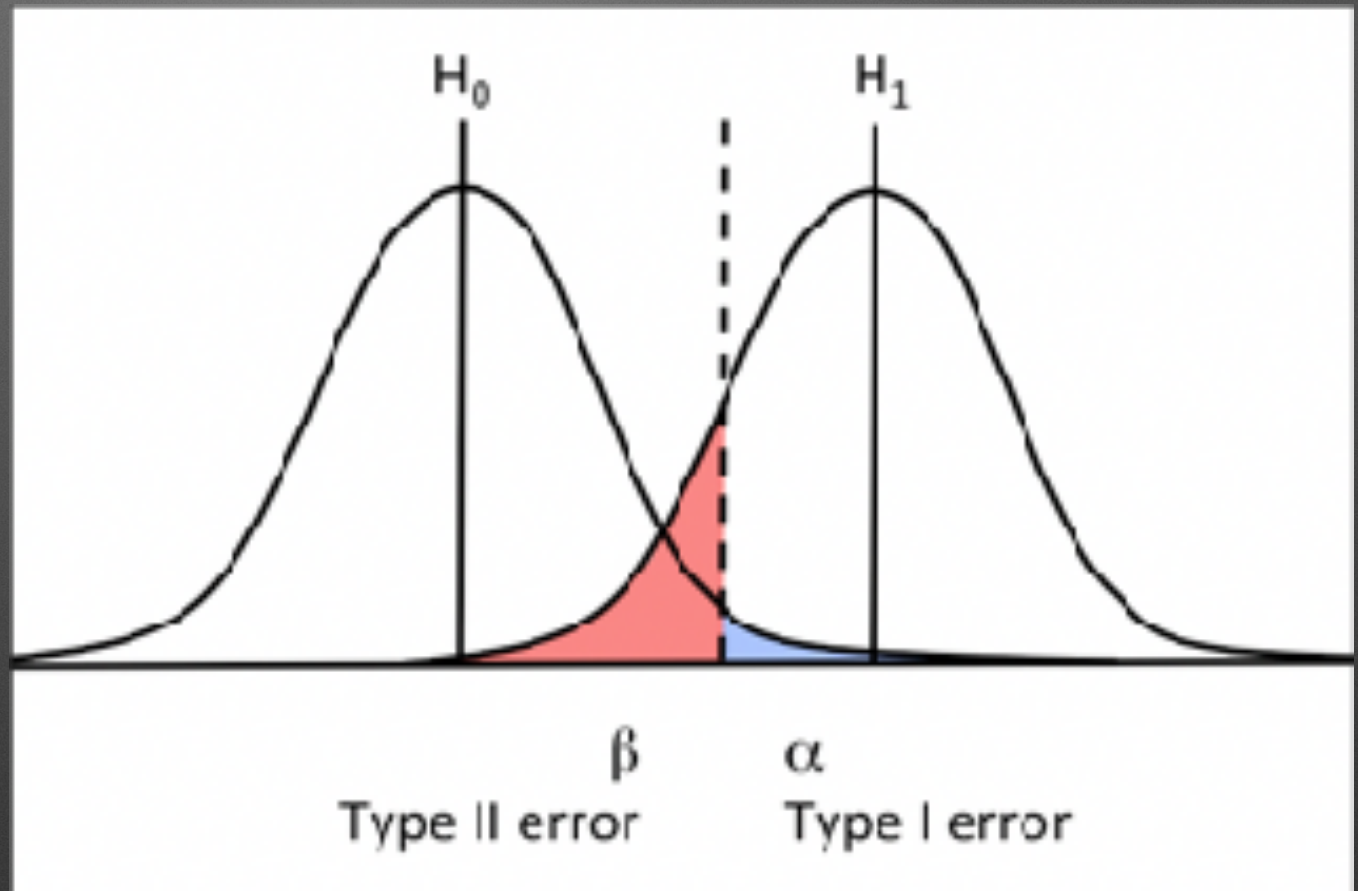15

galvanize

A/B Testing

galvanize

Z Test Error Region

Majority of hypothesis testing occurs via CLT (i.e. normal approximation of the population mean).

What region should we hope to minimize/ eliminate?

What 'controls' do we have to do such a thing?

sample size, effect size, and significance - which of the 3 is a reasonable control?

galvanize

Experimental Setup Revisited

Elements of a statistical test:
1. H_0
2. H_a
3. A test statistic
4. A significance level, α

Steps to run a test
1. Determine H_0
2. Choose α
3. Take sample, μ and σ
4. Accept or reject H_0

galvanize

Experimental Setup Revisited

Elements of a statistical test:
1. H_0
2. H_a
3. A test statistic
4. A significance level, α

Steps to run a test
1. Determine H_0, H_a
2. Choose α, (1-β)
3. Calculate required sample size
4. Take sample, μ and σ
5. Accept or reject H_0

galvanize

A/B Testing

Our current homepage has a signup conversion rate of 6% (what is the s.d.?)

We want to test if a new layout results in an increase of 1%.

How much data do we need to collect?

galvanize

Morning Objectives

1. Define Power and relate it to type 2 error

2. Compute Power for a dataset and a given problem

3. Demonstrate the relationship between sample size, effect size, and significance on power

galvanize