Power Calculation

Miles Erickson

(with Ryan Henning and Hutch Brock)

Power Calculation

Miles Erickson

(with Ryan Henning and Hutch Brock)

1. Review:
    a. Central Limit Theorem
    b. Hypothesis Testing
2. Type I vs Type II errors
3. What is "Power"?
4. Calculating Power / Sample Size
5. A/B Testing w/ Power

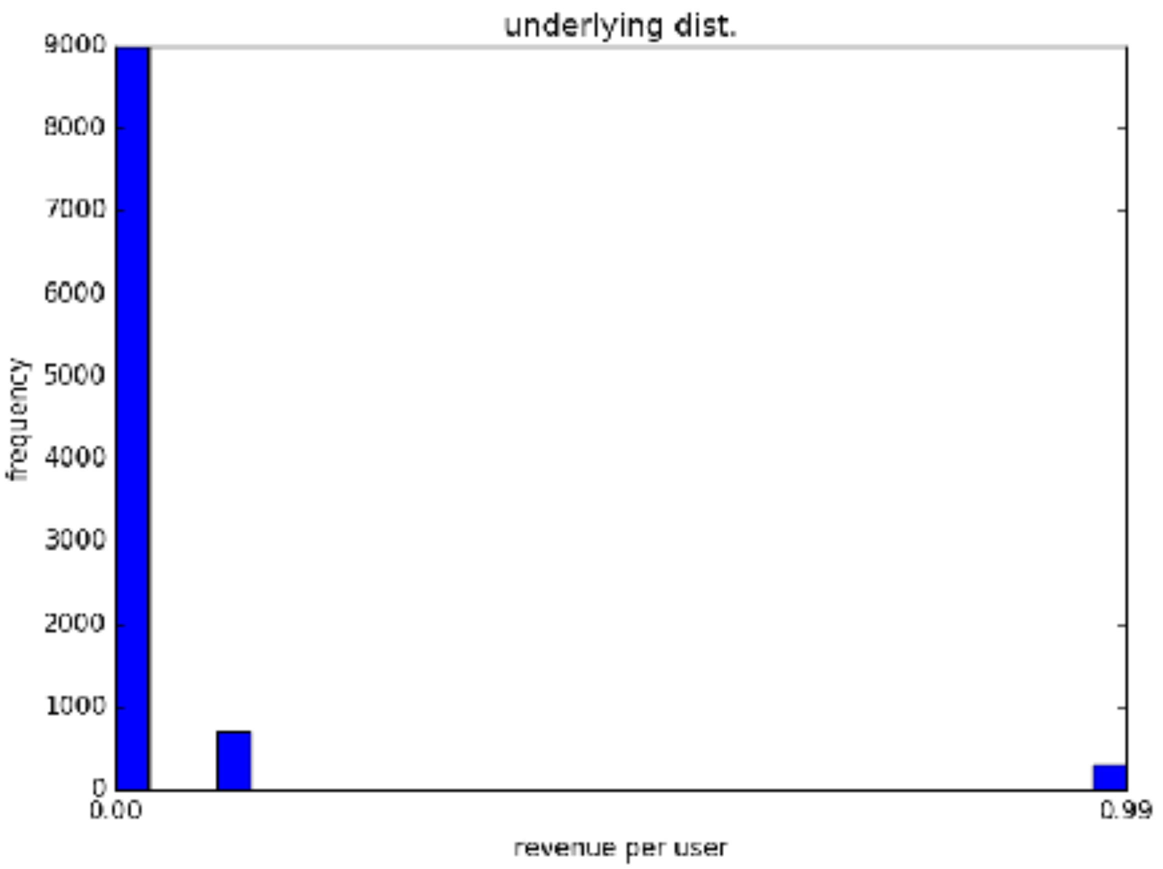
Standards

- Define Power and relate it to the Type II error
- Compute power given a dataset and a problem
- Explain how sample size, effect size, and significance contribute to power
- Identify what can be done to increase power
- Estimate sample size required of a test
- Define power - Be able to draw the picture with two normal curves with different means and highlight the section that represents Power
- Explain trade off between significance and power

Distribution of website revenue per visitor

Underlying Distribution:

Random variable: ***X* = revenue per visitor**	**P(*X*):**
X = $0.00 (no revenue)	90%
X = $0.10 (ad-click)	7%
X = $0.99 (app purchase)	3%



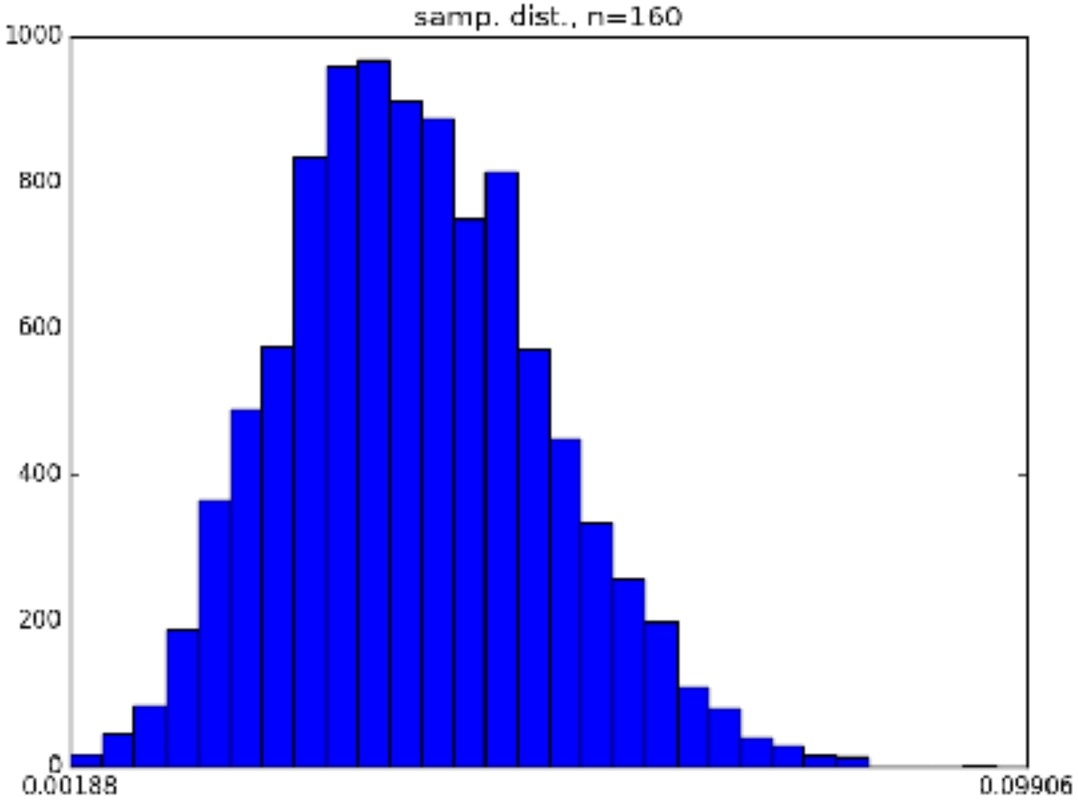
Distribution of sample means

Collect *n* samples from the website revenue distribution, calculate the sample mean $\bar{x}$

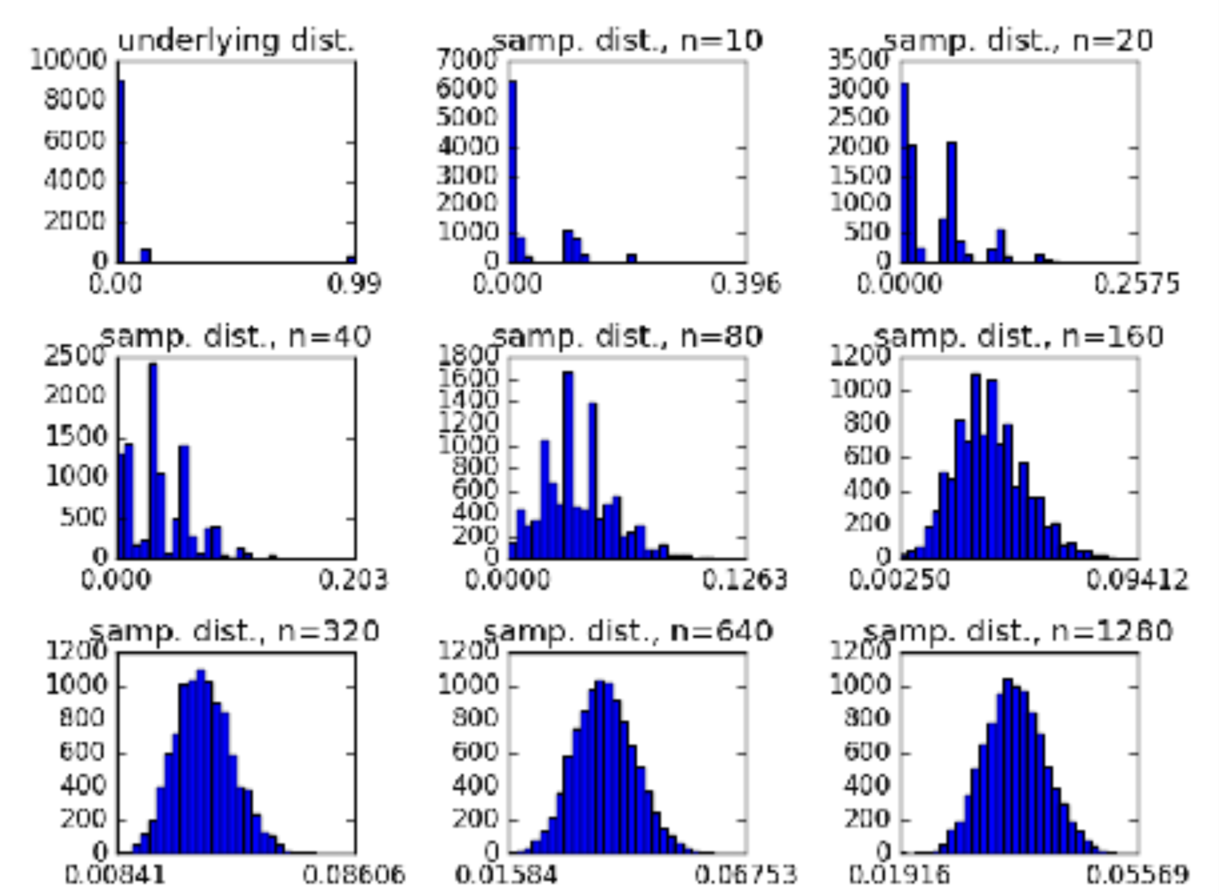
Repeat 10,000 times, we get:

$$\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{9999}$$

Plot all 10,000 sample means.



Central Limit Theorem



Central Limit Theorem: Std. Dev precise relationship to sample mean

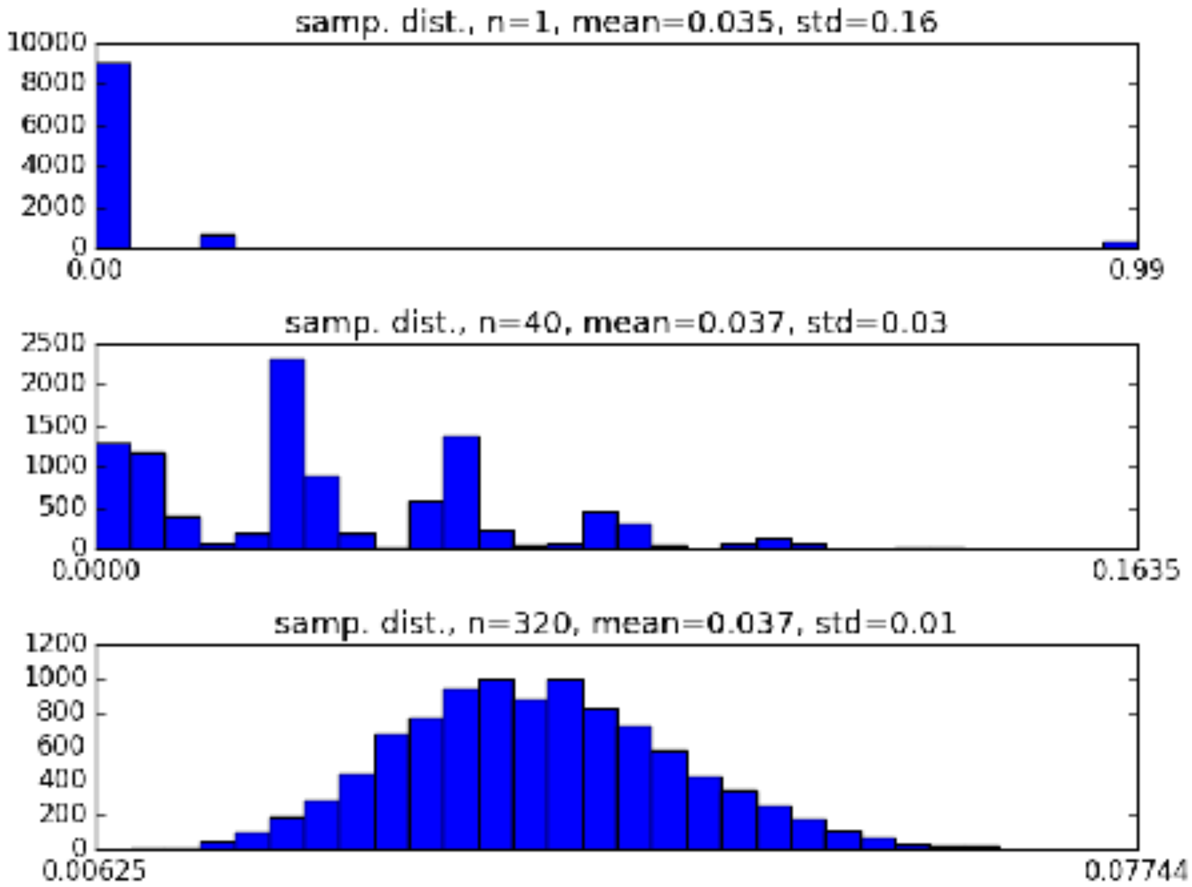
Let the underlying distribution have mean and std. dev.

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

The sampling distribution's mean and std. dev. will equal:

$$\mu' = \mu$$

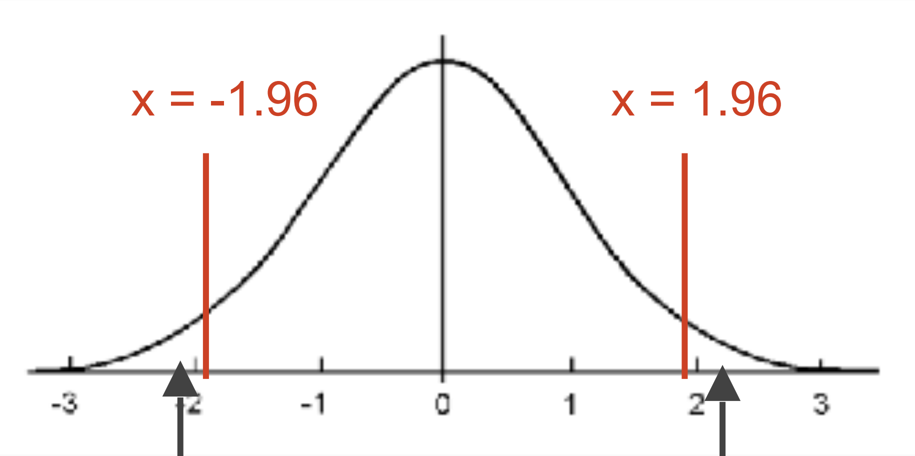
$$\sigma' = \sigma / \sqrt{n}$$



Hypothesis Testing: Review

Two-sided test:

$$H_0 : \mu = 0$$
$$H_A : \mu \neq 0$$

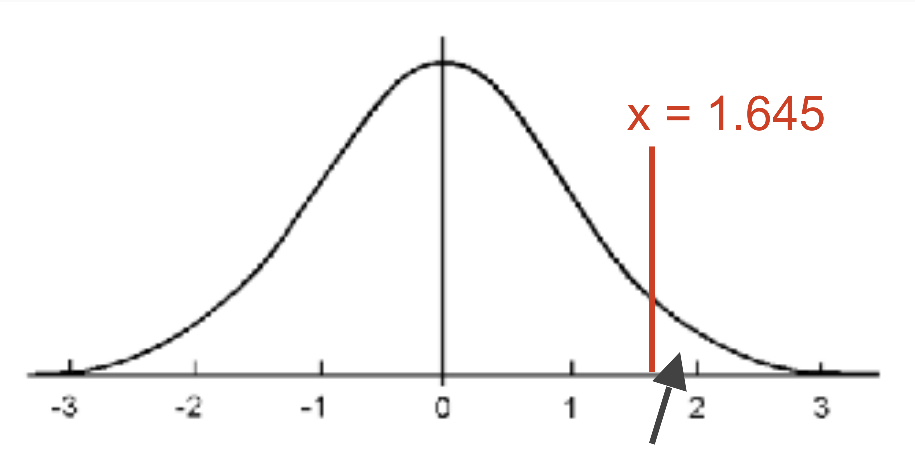


α/2 = 0.025 α/2 = 0.025

One-sided test:

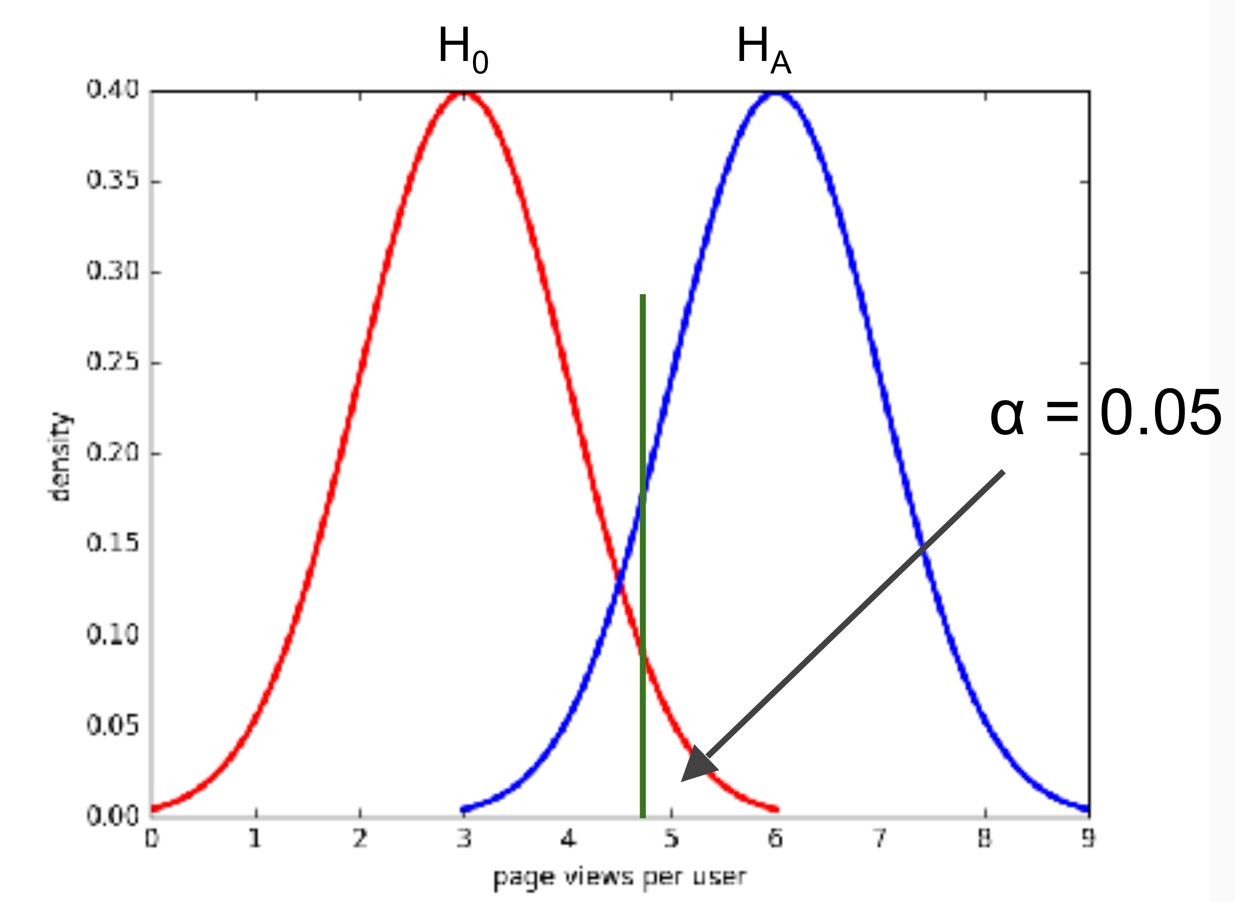
$$H_0 : \mu = 0$$
$$H_A : \mu > 0$$

α = 0.05



α = 0.05

Guessing the unknown



Hypothesis Testing: Possible Outcomes

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

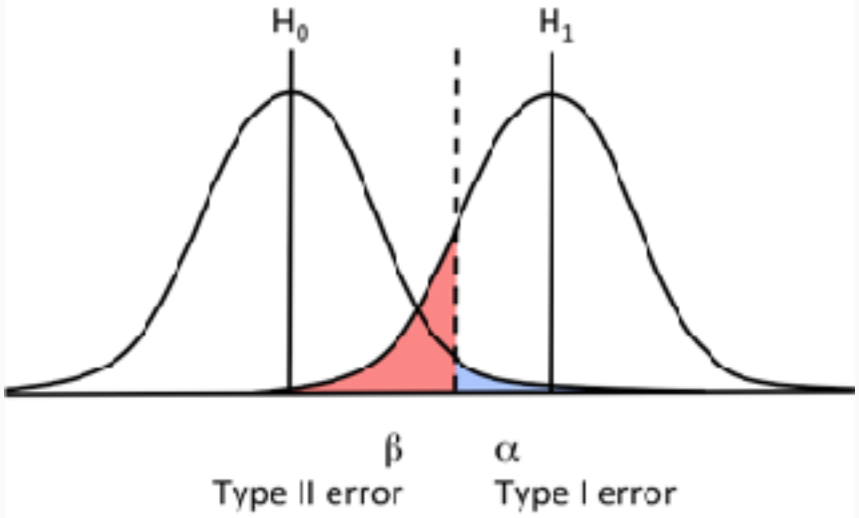
We call this the experiment's "Power". It is the probability that we **correctly reject H_0** when the null hypothesis is false.

Hypothesis Testing: Possible Outcomes

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

$$\text{Power} = P(\text{Reject } H_0 \mid H_a \text{ Is True})$$

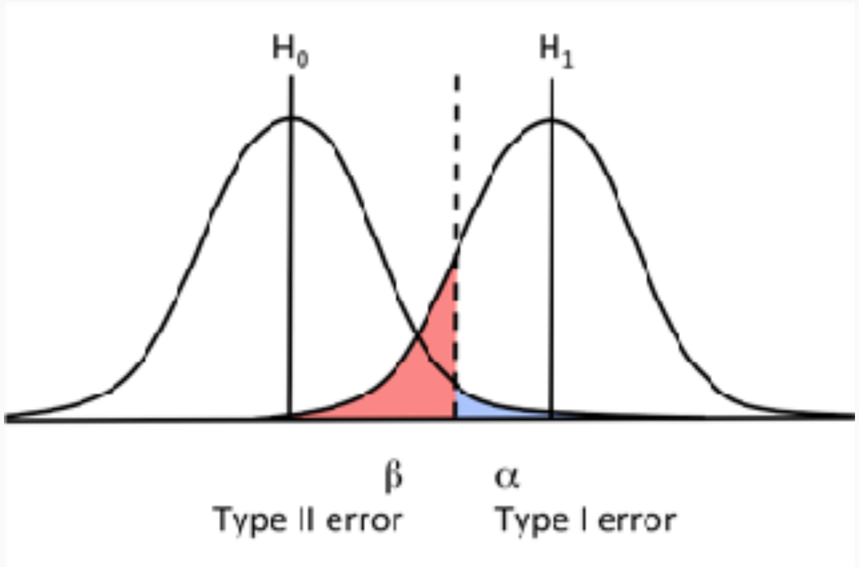
Hypothesis testing: the *power* region



	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)**

The *power* measurement is in relationship to a <u>specific</u> alternative hypothesis. Think of it as the *power* to detect a particular "effect size".

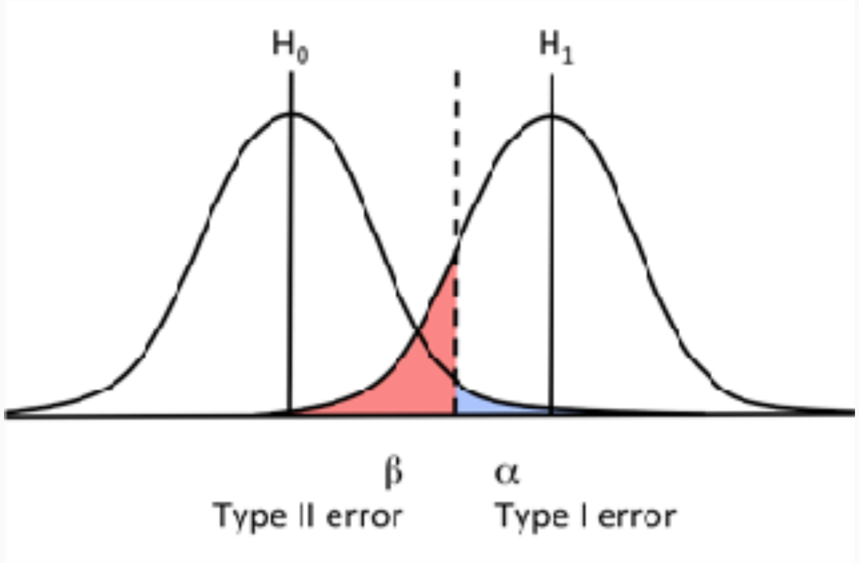
Hypothesis testing: the *power* region



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

What is power? How is it related to sample size, variance, effect size, and significance level?

Hypothesis testing: the *power* region



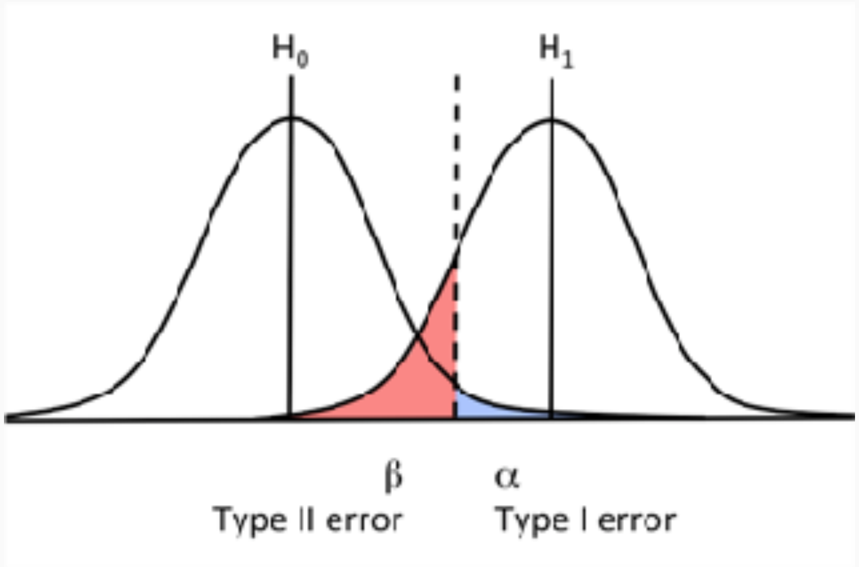
	H_0 Is True	H_0 Is False
Fail To Reject H_0	Correct Decision (1 - α)	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision (1 - β)

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 (revised with power calculation)



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

1. Decide to run an experiment, choose α $\boxed{\text{and } (1-\beta)}$
2. $\boxed{\text{Calculate required sample size } n}$
3. Take sample, obtain $\bar{x}$ and s
4. Accept or reject H_0

(new steps)

Calculating the required sample size

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

```
import scipy.stats as st
st.norm.ppf(alpha)
st.norm.ppf(1 - beta)
```

Hands-on Exploration

(power-exploration.ipynb)

A/B Testing

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

Our current homepage has a signup conversion rate of 6%. (What is the standard deviation?)

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 >= ?

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

Our current homepage has a signup conversion rate of 1%. (What is the standard deviation?)

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 5% and power is 80%.

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

n >= ?

A/B Testing

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

Our current homepage has a signup conversion rate of 20%. (What is the standard deviation?)

We want to test a new homepage design to see if we can get a 30% signup rate. We'll want an experiment where alpha is 10% and power is 99%.

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

$$n >= ?$$

Factors Influencing Power

1. Effect Size
2. Standard Deviation
3. Sample Size
4. Significance Level (α)

Review Questions

- What is the relationship between power and Type II error?
- What is the tradeoff between significance and power?
- What can be done to increase power?