# Power Calculation

Natalie Hunt

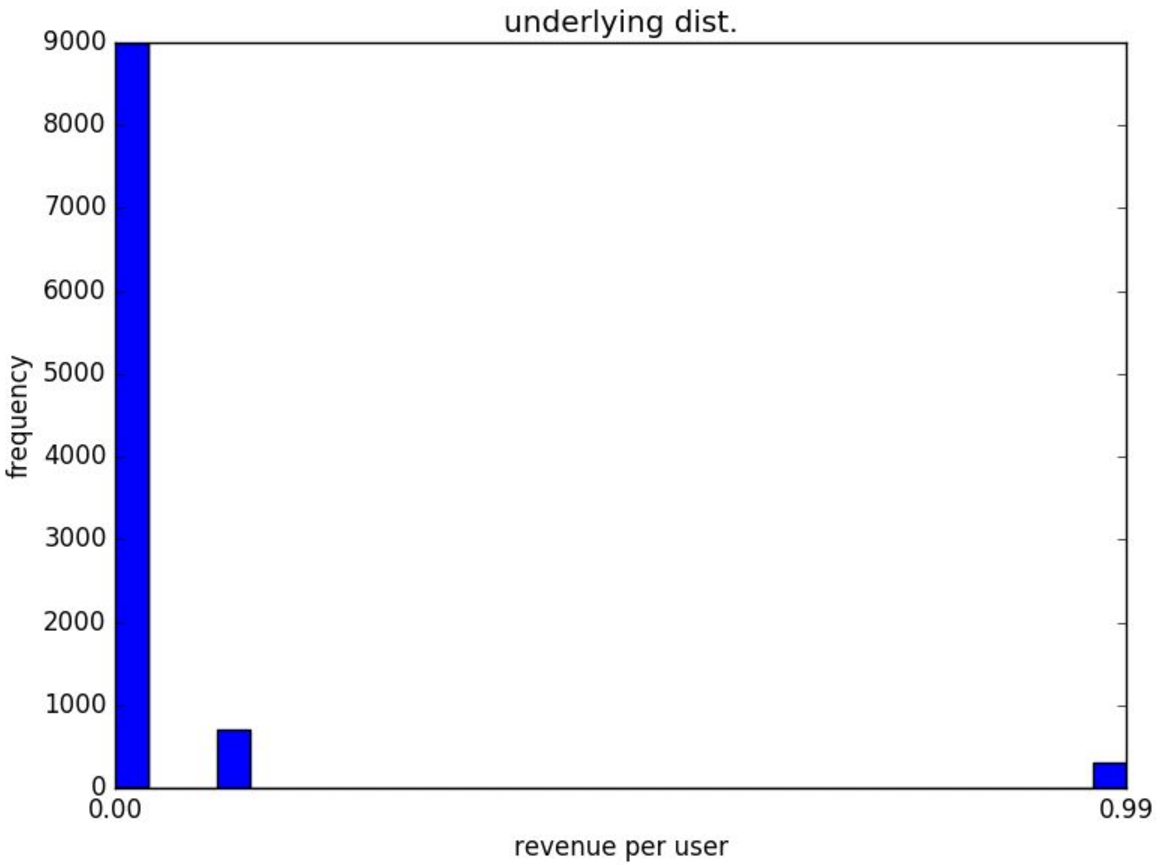
galvanize

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

# Distribution of website revenue per visitor

## Underlying Distribution:

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



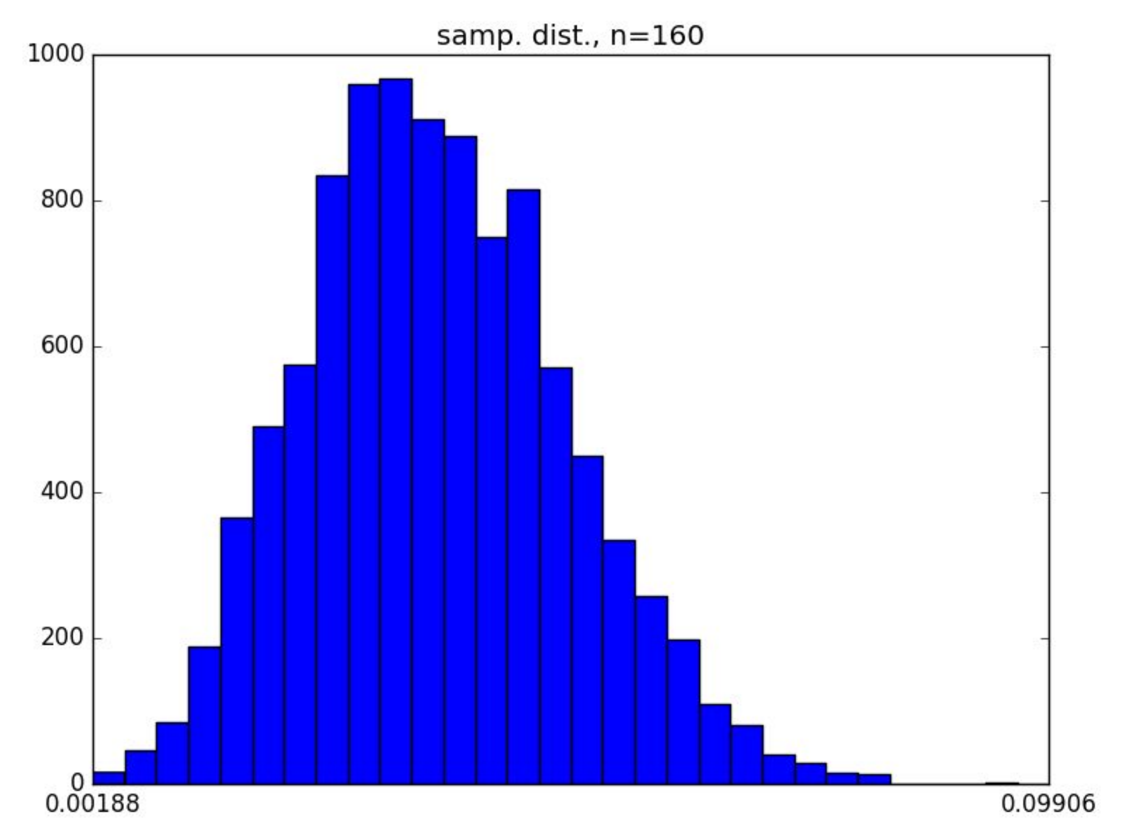
# Distribution of sample means

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

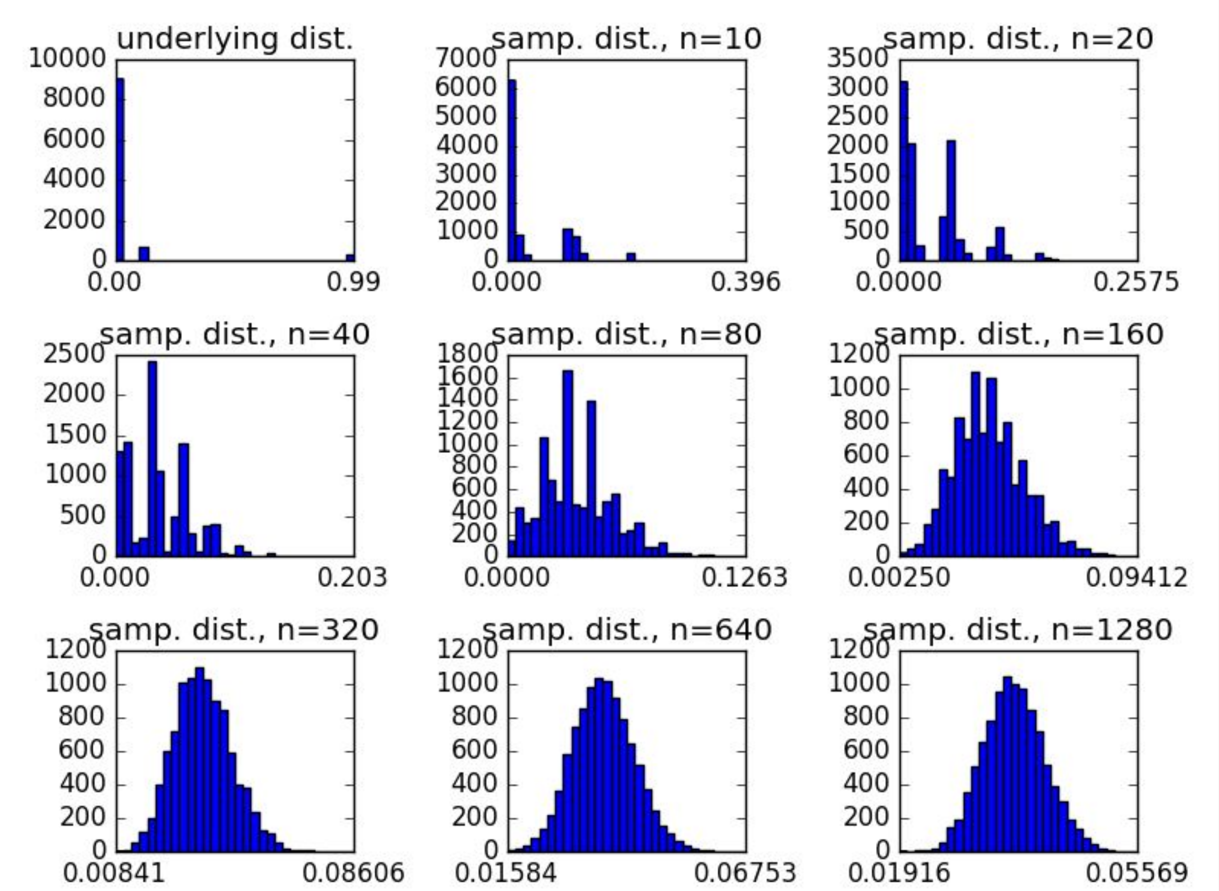
Repeat 10,000 times, we get:

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

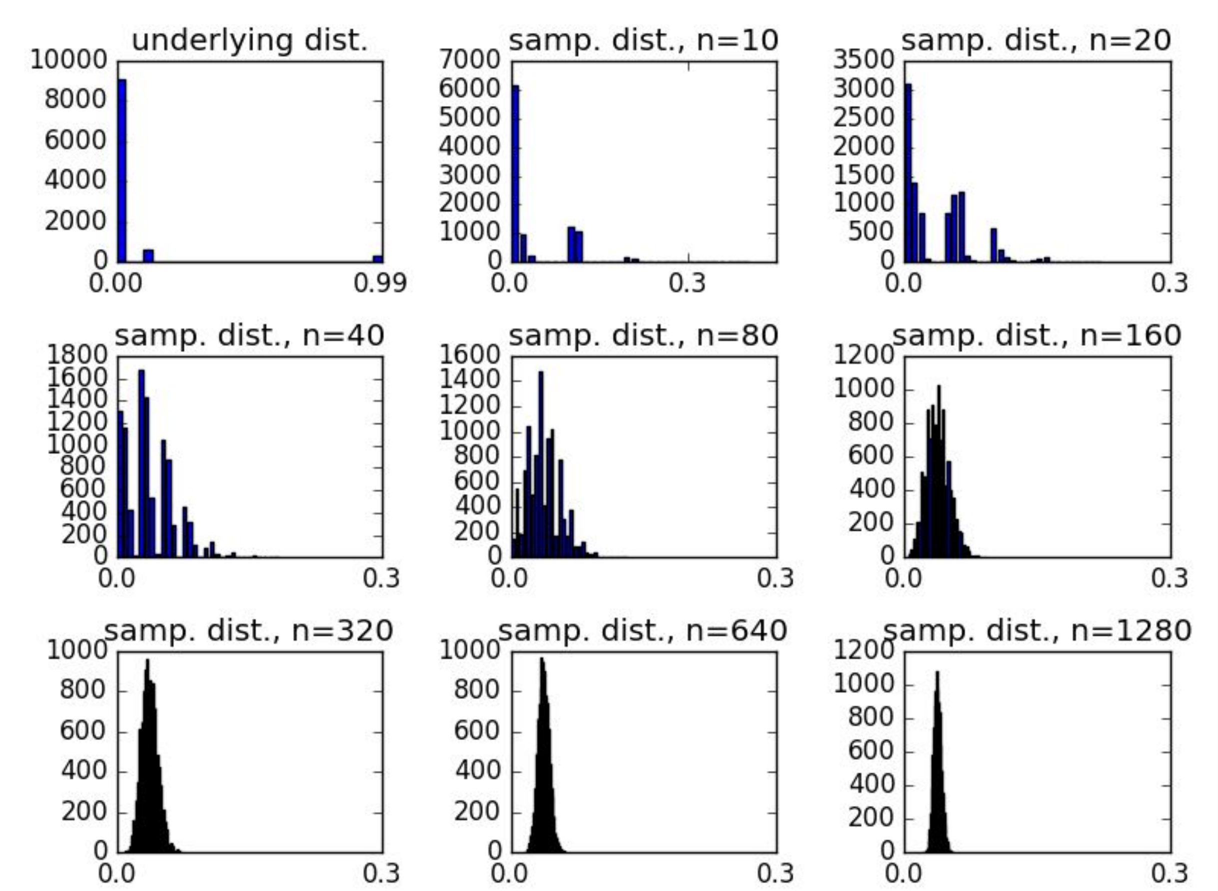
Plot all 10,000 sample means.



# Central Limit Theorem



# Central Limit Theorem: What happens when the sample size increases?



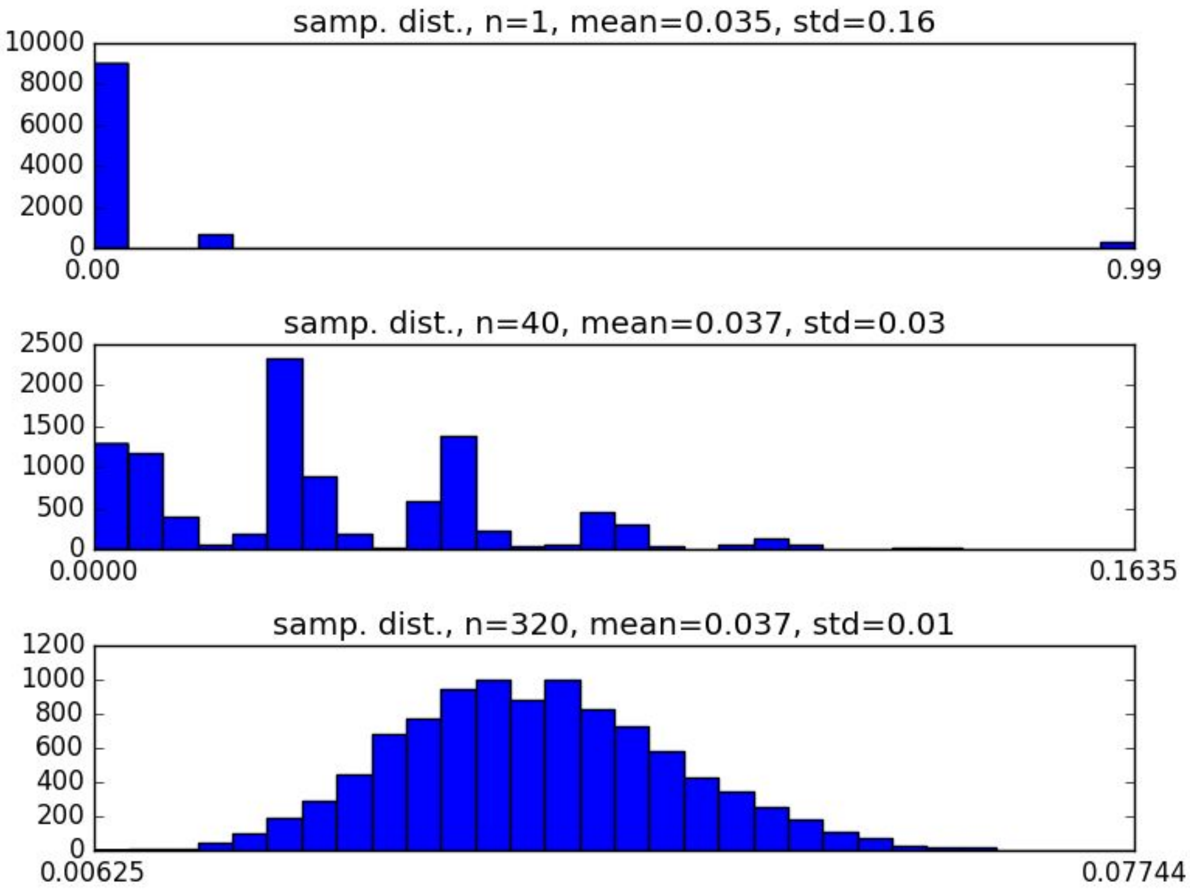
# Central Limit Theorem: Std. Dev precise relationship to sample size

Let the underlying distribution have mean and std. dev.

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

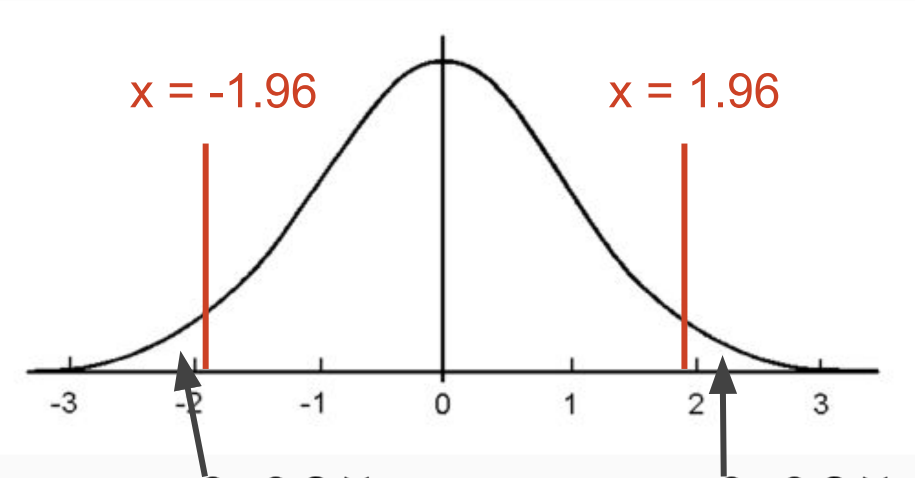
The sampling distribution's mean and std. dev. will equal:

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



## Two-sided test:

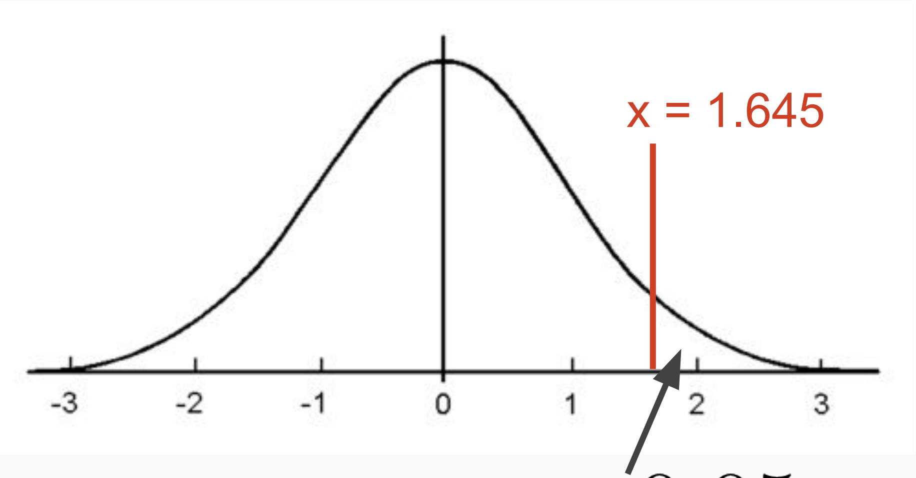
$$H_0 : \mu = 0$$
$$H_A : \mu \neq 0$$




$$\alpha = 0.025 \qquad \alpha = 0.025$$

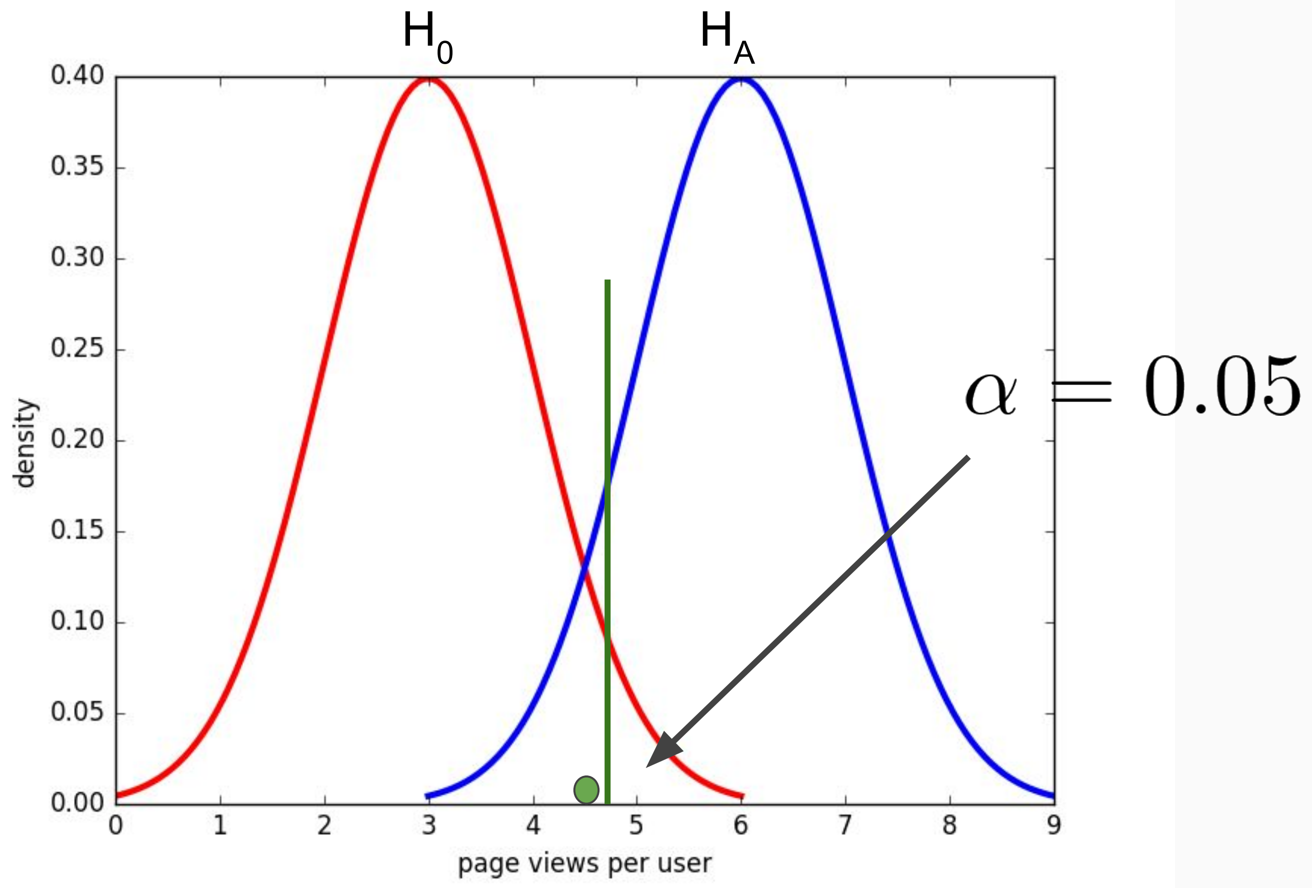
## One-sided test:

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




$$\alpha = 0.05$$

# Guessing the unknown



# Hypothesis Testing: Possible Outcomes

galvanize

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-$\alpha$) | Type II Error ($\beta$) |
| Reject $H_0$ | Type I Error ($\alpha$) | Correction 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.

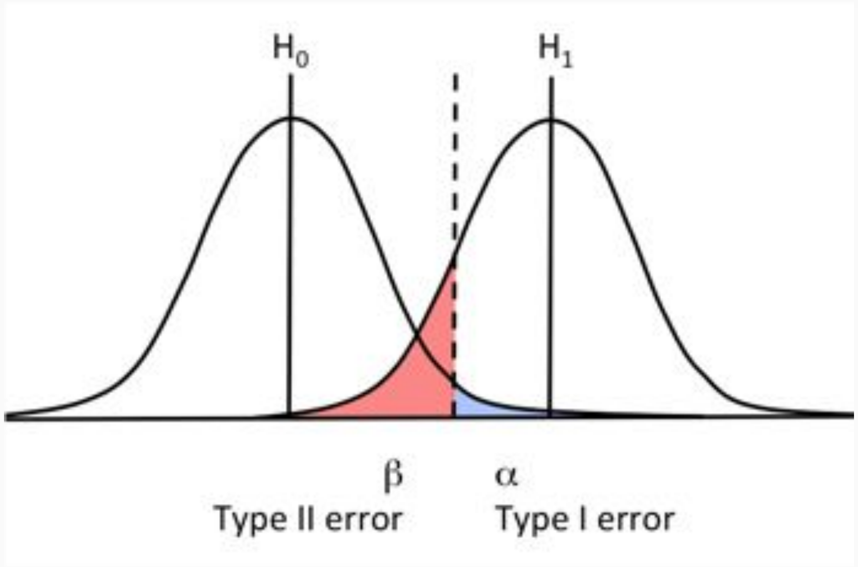
# Hypothesis Testing: Possible Outcomes

| | $H_0$ is true<br>true - | $H_0$ is false<br>true + |
|---|---|---|
| Accept $H_0$<br>predict - | Correct Decision<br>$(1-\alpha)$ | Type II Error<br>$(\beta)$ |
| Reject $H_0$<br>predict + | Type I Error<br>$(\alpha)$ | Correction Decision<br>$(1-\beta)$ |

false positive rate
(aka, 1 - specificity)

true positive rate
(aka, sensitivity)

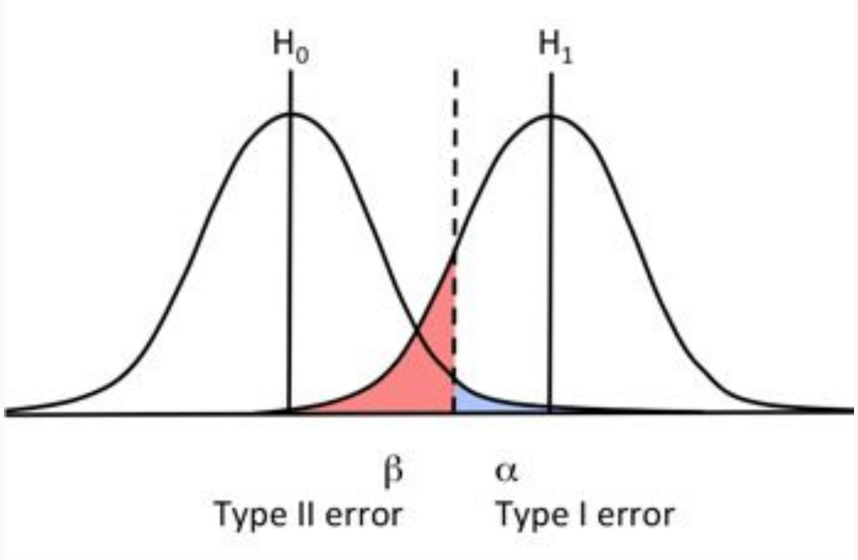
# Hypothesis testing: the *power* region




| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction 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".

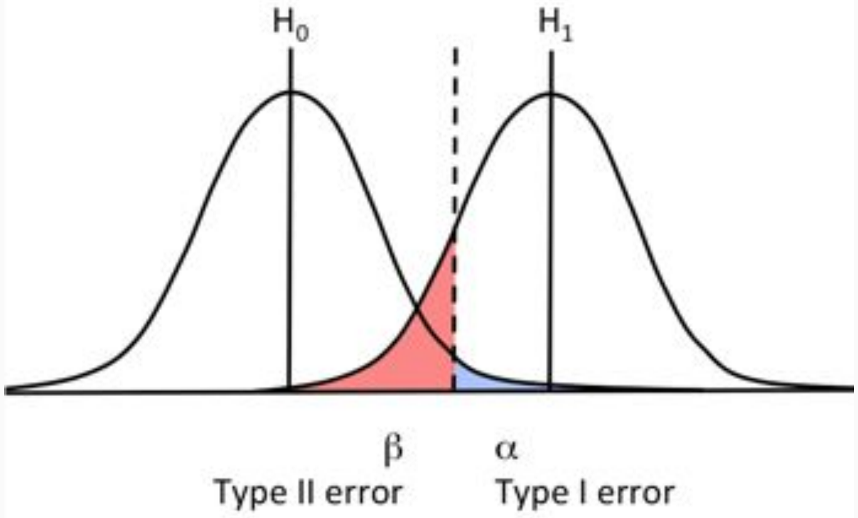
# Hypothesis testing: the *power* region




| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction Decision (1-β) |

**What happens to *power* when we increase alpha?**

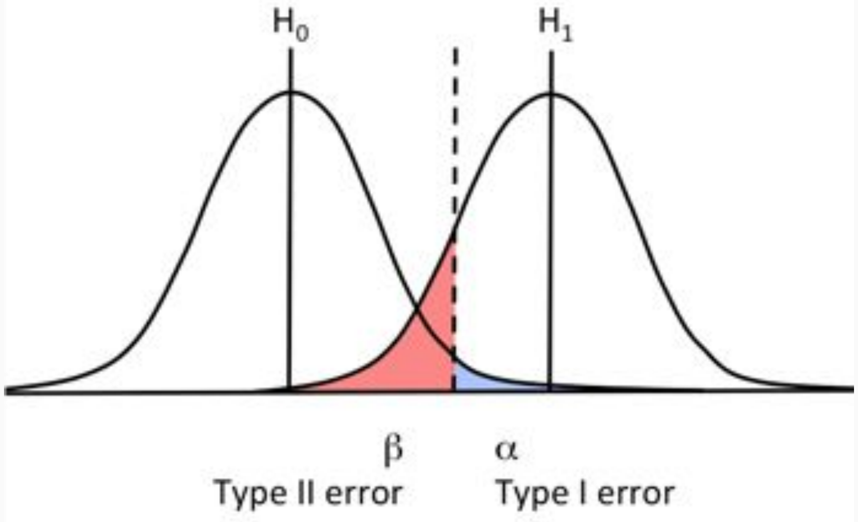
# Hypothesis testing: the *power* region




| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction Decision (1-β) |

**What happens to *power* when we increase the effect size?**

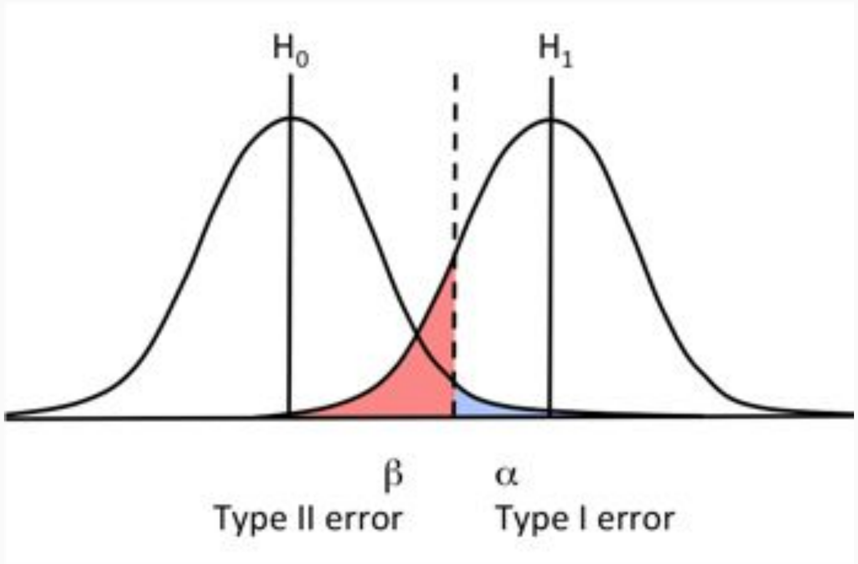
# Hypothesis testing: the *power* region




| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction Decision (1-β) |

**What happens to *power* when we increase the sample std. deviation?**

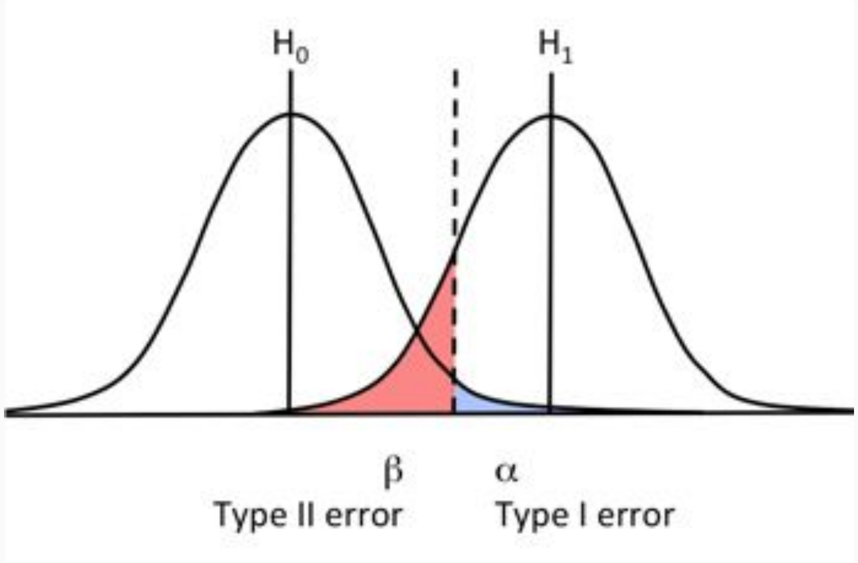
# Hypothesis testing: the *power* region




| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction Decision (1-β) |

**What happens to *power* when we increase the sample size?**

# Hypothesis testing: the *power* region



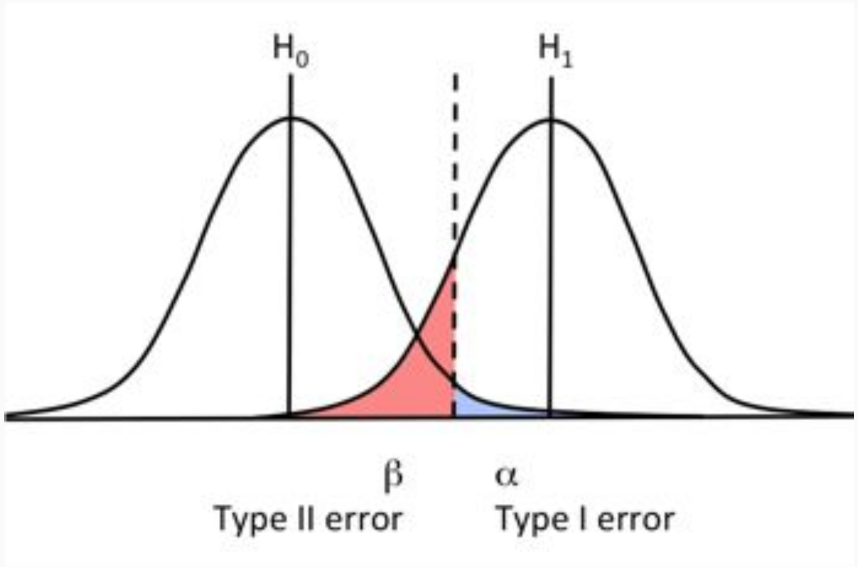

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction Decision (1-β) |

**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 |
|---|---|---|
| Accept $H_0$ | Correct Decision (1-α) | Type II Error (β) |
| Reject $H_0$ | Type I Error (α) | Correction Decision (1-β) |

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

(new steps)

# Calculating the required sample size

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

```python
import scipy.stats as st
```

```python
st.norm.ppf(alpha)
```

```python
st.norm.ppf(1 - beta)
```

# A/B Testing

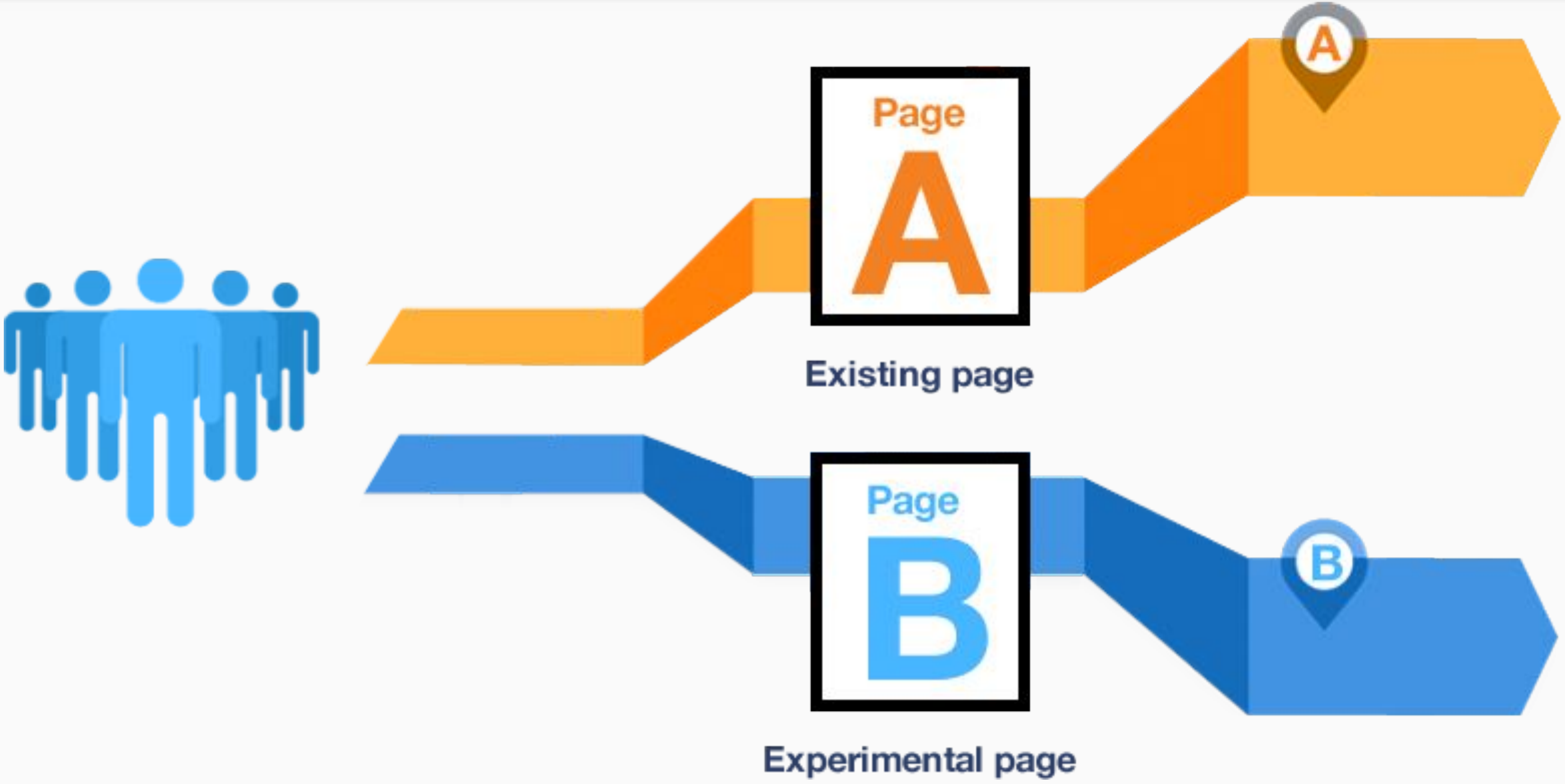



Image from: http://techcrunch.com/2014/06/29/ethics-in-a-data-driven-world/

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

Our current homepage has a signup conversion rate of 6%. (The standard deviation 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?

$$n \geq 9,084$$

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

Our current homepage has a signup conversion rate of 1%. (The standard deviation would be 0.099.)

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

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

$$n \geq 39,427$$

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

Our current homepage has a signup conversion rate of 20%. (The standard deviation would be 0.4.)

We want to test a new homepage design to see if we can get a <u>30% signup rate</u>. We'll want an experiment where <u>alpha is 1%</u> and <u>power is 95%</u>.

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

$$n \geq 253$$