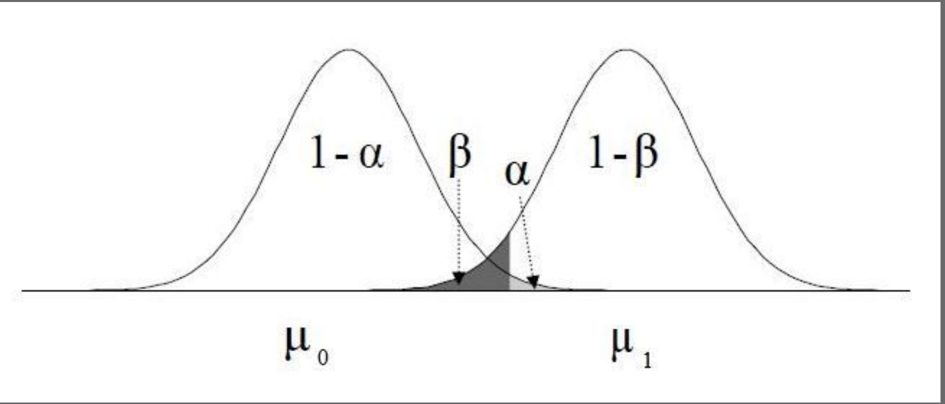
Power Calculation



Taryn Heilman
Ryan Henning
Frank Burkholder

galvanize

Learning Objectives

- **Review** relevant concepts
- **Define** statistical power
- Learn how to **calculate** power and how to integrate into hypothesis
- **Understand** different factors that affect power

Review

- What does the Central Limit Theorem state?
- What are the implications of this for hypothesis testing?
- Describe the steps necessary to compute 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

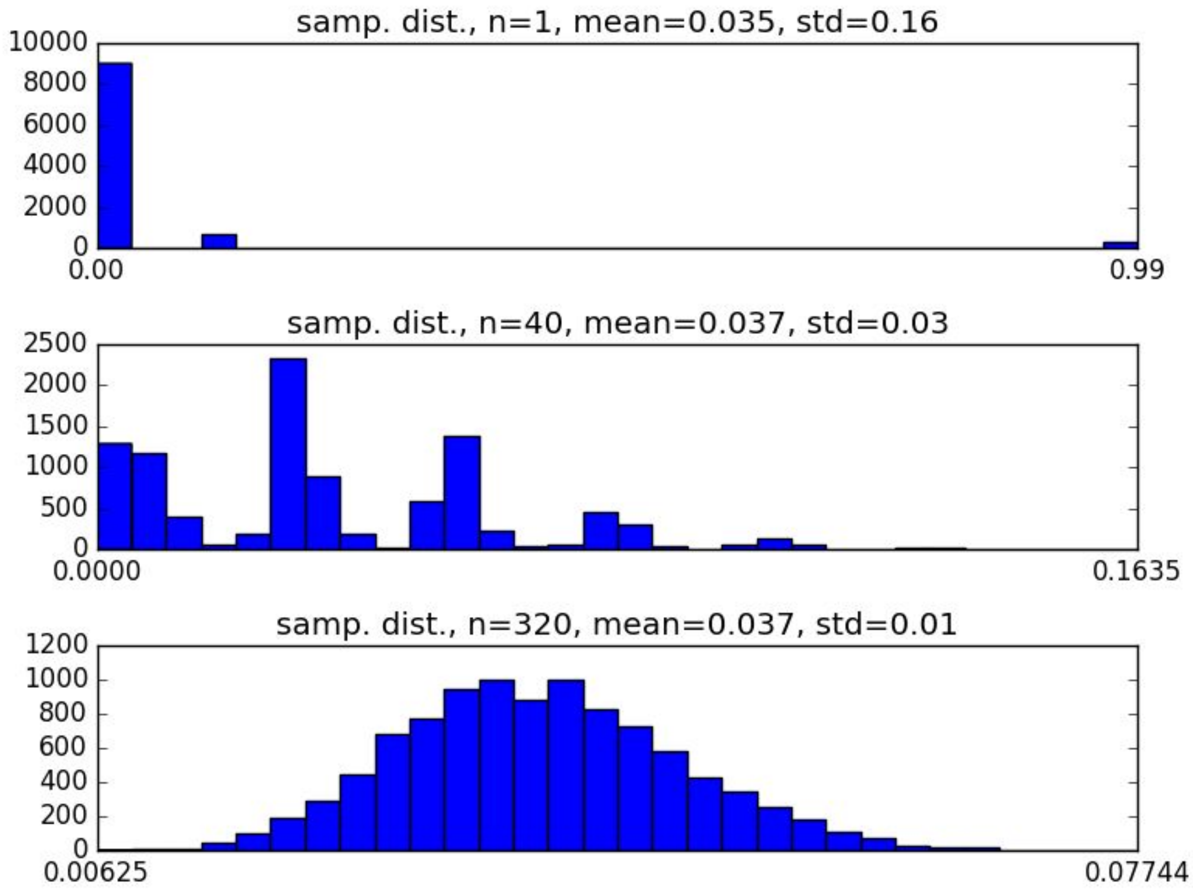
One more time: Central Limit Theorem

Let the underlying distribution have mean and std. dev.

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

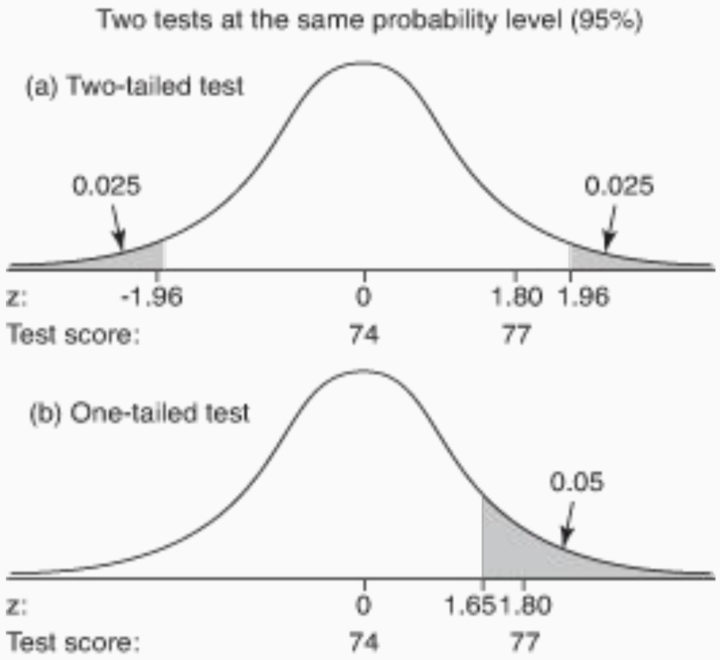
The sampling distribution's mean and std. dev. will equal:

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



Review: One-tailed vs. Two-tailed tests

Direction	H_0	H_A	P-value
2-sided Test	$=$	$\neq$	One half of P-value in each tail
Left-Tail	$\geq$	$<$	All of P-value in left tail
Right-Tail	$\leq$	$>$	All of P-value in right tail



Statistical Power

galvanize

Hypothesis Testing: Possible Outcomes

galvanize

Reality

What we choose to do

	Null (N) H_0 is true	Alternative (P) H_0 is false
Accept H_0	Correct Decision (1-α)	False negative Type II Error (β)
Reject H_0	False positive Type I Error (α)	Correction Decision (1-β)

Think of H_a as positive.

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

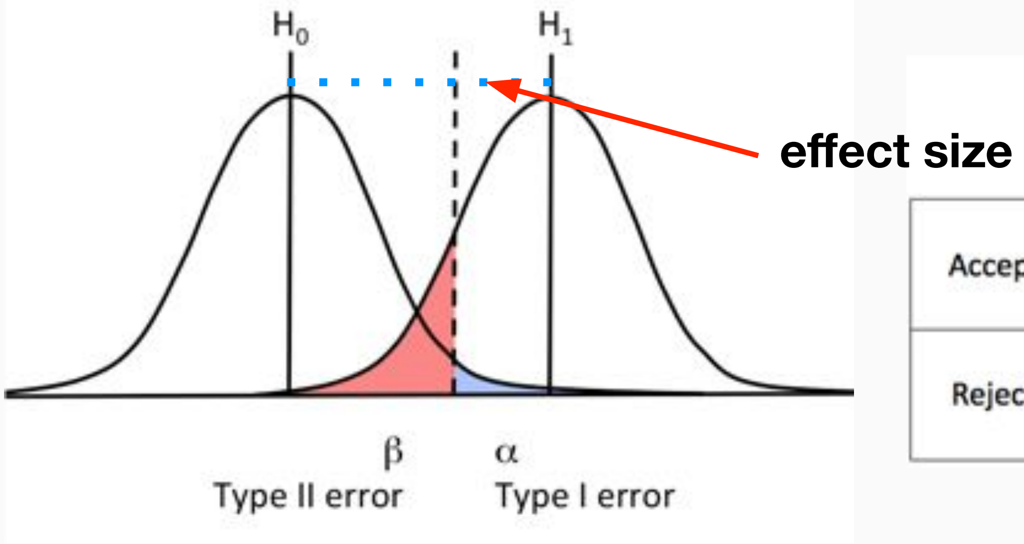
Hypothesis Testing: Possible Outcomes

	H_0 is true 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)

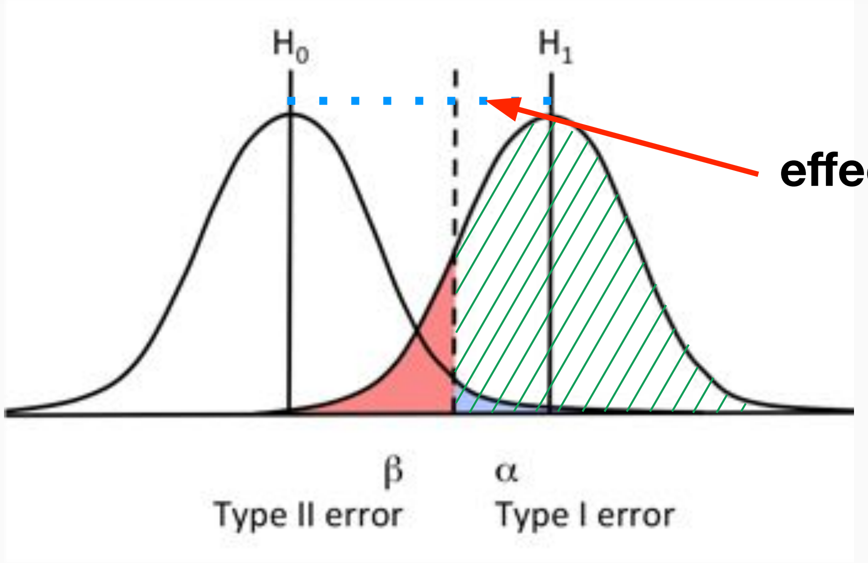
Hypothesis testing: the *power* region



	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 specific 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)	Power Correction Decision (1-β)

Power:

a) the probability that your test will reject a false null hypothesis
b) the ability of your test to detect an effect, if it actually exists

typically set to 80%

Explaining Power Interactively

Interactive exploration

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

Another resource

http://powerandsamplesize.com/Calculators/Other/1-Sample-Normal

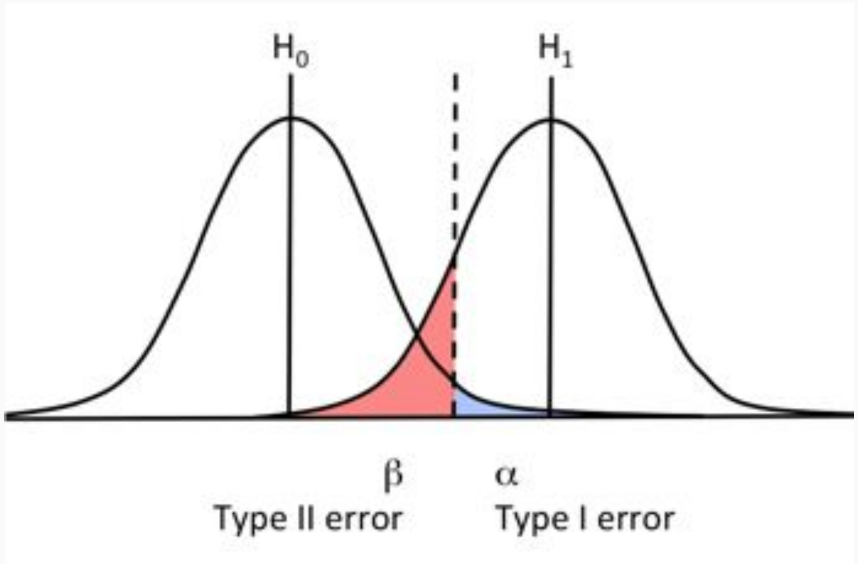
Factors that affect Power

- Significance level (alpha)
- Effect size
- Sample size (n)

Can do this calculation to solve for any of the 4 factors (alpha, effect size, n, power)

TYPICALLY you set an alpha, desired effect size, and statistical power, and calculate the required sample size based on these

Check Understanding: factors that affect power



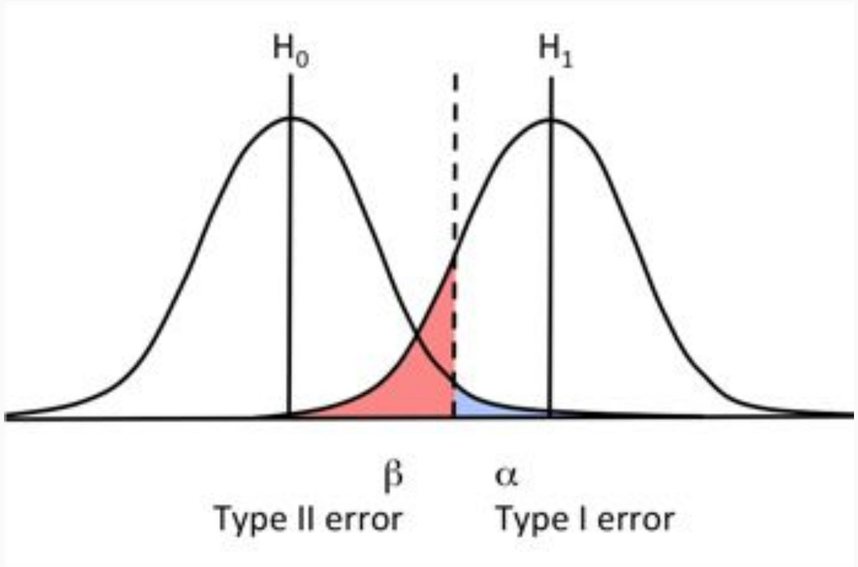
	H_0 is true	H_0 is false
Accept H_0	Correct Decision (1-α)	Type II Error (β)
Reject H_0	Type I Error (α)	Correction Decision (1-β)

What happens to *power* when we:

- *increase alpha?*
- *Increase effect size?*
- *increase the sample size?*

(Hint, what happens to standard deviation as we increase sample size?)

Hypothesis testing: in practice



	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 Steps : revised

1. State the null (H0) and the alternative (H1) hypotheses
2. Choose a level of significance (alpha) *and power (1 - beta)*
   i. *Compute number of samples required for desired effect size*
   ii. Collect data
3. Compute the test statistic
4. Calculate p-value
5. Draw conclusions
   - Reject H0 in favor of H1
   - Fail to reject H0

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 1: calculate sample size

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?

Breakout 1: calculate sample size

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?

$$n \geq 9,084$$

Breakout 2: calculate sample size

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 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 sample size

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 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?

$$n \geq 38{,}642$$

Learning Objectives

- **Review** relevant concepts
- **Define** statistical power
- Learn how to **calculate** power and how to integrate into hypothesis
- **Understand** different factors that affect power

Recap: Learning Objectives

galvanize

- ✓ **Review** relevant concepts
- ✓ **Overview** of other test statistics:
  - ✓ Chi-squared tests
  - ✓ Paired t-test
  - ✓ KS test
- ✓ **Discuss** Experimental design and Confounding Variables
- ✓ **Define** statistical power
- ✓ Learn how to **calculate** power and how to integrate into hypothesis
- ✓ **Understand** different factors that affect power