



# PSYCH 201B

## *Statistical Intuitions for Social Scientists*

## **Statistical Power & Group Discussion**

You can download these slides:  
course website > Week 8 > Overview

# Faculty life....current mood....

NEWS > EDUCATION

SUBSCRIBER ONLY

## UC San Diego, bracing for more than \$200M in funding cuts, freezes faculty hiring

'This is the biggest financial challenge in UCSD's history,' said Terry Gaasterland, a longtime faculty member.



# Announcements

- Today's plan
  - Mini-lecture: Statistical Power
  - Break & demos to play with
  - Mini-Group and Class Discussion
- We're giving you a **mini HW** due **Monday**
  - No coding! Shouldn't take more than 1hr!
  - Just reading an interactive article and **testing yourself** on questions
  - Writing up a brief response using google doc template

# **What is statistical power?**

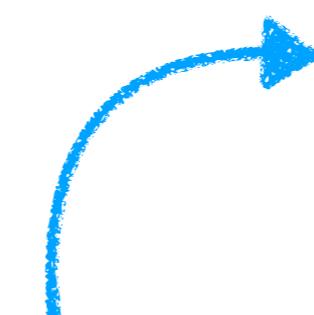
**...some background first**

# P-values are **evidence** for hypotheses

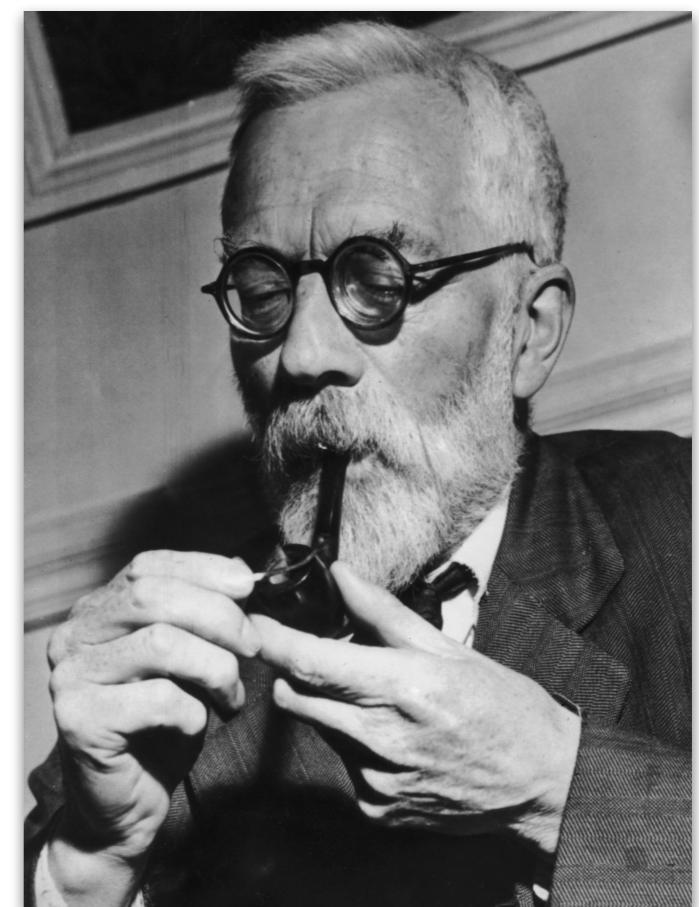
"If  $P$  is between .1 and .9 there is certainly no reason to suspect the hypothesis tested. If it is below .02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. **We shall not often be astray if we draw a conventional line at .05** ... it is convenient to draw the line at about the level at which we can say: Either there is something in the treatment, or a coincidence has occurred such as does not occur more than once in twenty trials..."

BUT

**"no scientific worker has a fixed level of significance** at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas"



**"invented"  $p < .05$**



Sir Ronald **Fisher**

# No! P-values are **not** evidence for hypotheses

*"no test based upon a theory of probability can by itself provide any valuable evidence of the truth or falsehood of a hypothesis. But we may look at the purpose of tests from another viewpoint. **Without hoping to know** whether each separate **hypothesis is true or false**, we may search for **rules** to govern our behavior with regard to them, in following which we insure that, **in the long run of experience, we shall not often be wrong**"*

Dad (Karl Pearson) invented... like everything  
r correlation coefficient, Chi-square, PCA ...



**invented**  
**confidence intervals**

Jerzy Neyman    Karl Pearson

We can't know which specific decisions are right or wrong...

We can **control how often** our decisions will be **wrong in the long run**

"no test based upon a theory of probability can by itself provide any valuable evidence of the truth or falsehood of a hypothesis. But we may look at the purpose of tests from another viewpoint. **Without hoping to know** whether each separate **hypothesis is true or false**, we may search for **rules** to govern our behavior with regard to them, in following which we insure that, **in the long run of experience, we shall not often be wrong**"

Dad (Karl Pearson) invented... like everything  
r correlation coefficient, Chi-square, PCA ...



**invented**  
**confidence intervals**

Jerzy Neyman   Karl Pearson

# Neyman-Pearson Approach

- Two possible states of the world
  - $H_0$  (null) is **true**
  - $H_0$  (null) is **false**
- Two decisions we can make
  - **reject**  $H_0$
  - **retain**  $H_0$
- Two ways we can be wrong
  - **Reject** when  $H_0$  is **true**
  - **Retain** when  $H_0$  is **false**



False Positive **Type I** error  $\alpha$   
False Negative **Type II** error  $\beta$

Imagine you're a (medical) doctor...

$H_0$ : Not pregnant

$H_1$ : Pregnant

Type I Error



Type II Error



**Type I Error:** Falsely rejecting the null hypothesis (even though it is true)

**Type II Error:** Failing to reject the null hypothesis (even though it is false)

Imagine you're a (medical) doctor...

$H_0$ : Not pregnant     $H_1$ : Pregnant

Ways to  
be wrong



<b>false positive</b>	Type I error	$\alpha$
-----------------------	--------------	----------

$p(\text{reject } H_0 | H_0 \text{ is true})$

<b>false negative</b>	Type II error	$\beta$
-----------------------	---------------	---------

$p(\text{not reject } H_0 | H_1 \text{ is true})$

Ways to  
be right



<b>true positive</b>	Sensitivity	$1 - \beta$
----------------------	-------------	-------------

$p(\text{reject } H_0 | H_1 \text{ is true})$

<b>true negative</b>	Specificity	$1 - \alpha$
----------------------	-------------	--------------

$p(\text{not reject } H_0 | H_0 \text{ is true})$

The **power** of a binary hypothesis test is the probability that the test rejects the null hypothesis ( $H_0$ ) when a **specific** alternative hypothesis ( $H_1$ ) is true.

---

$H_0$ : Students and non-students have the same balance.

**Model C**

$$Y_i = \beta_0 + \epsilon_i$$

$$\beta_1 = 0$$

$H_1$ : Students and non-students have different balances.

**Model A**

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

$$\beta_1 \neq 0$$

We cannot calculate power in this case.

We need a **specific** alternative hypothesis!

The **power** of a binary hypothesis test is the probability that the test rejects the null hypothesis ( $H_0$ ) when a **specific** alternative hypothesis ( $H_1$ ) is true.

---

$H_0$ : Students and non-students have the same balance.

**Model C**

$$Y_i = \beta_0 + \epsilon_i$$

$$\beta_1 = 0$$

$H_1$ : Students and non-students have different balances.

**Model A**

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

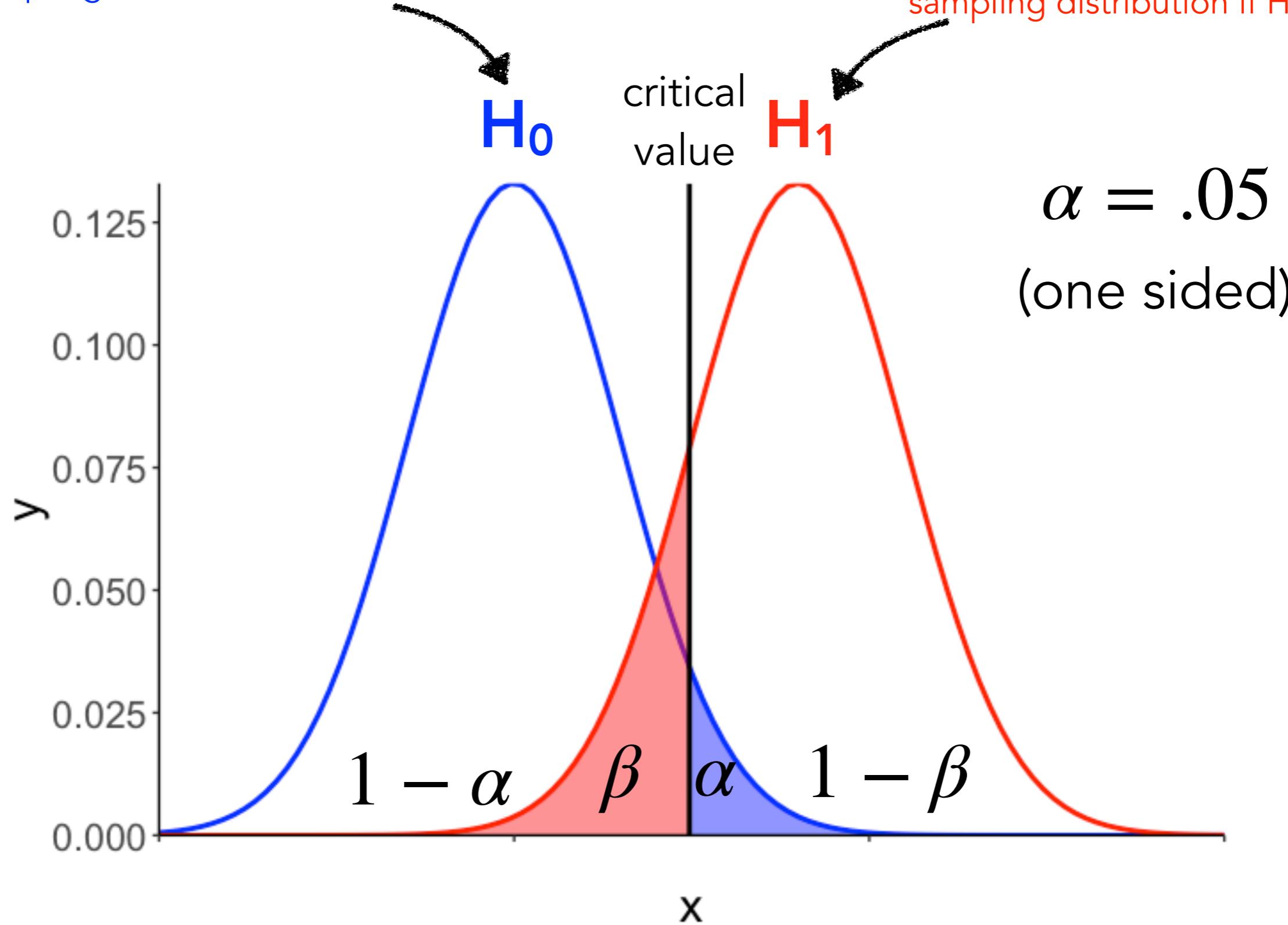
$$\beta_1 = 300$$

We can calculate power in this case since we have a **specific** alternative hypothesis!

# What affects power?

false positive	Type I error $p(\text{reject } H_0   H_0 \text{ is true})$	$\alpha$	Power $1 - \beta$
false negative	Type II error $p(\text{not reject } H_0   H_1 \text{ is true})$	$\beta$	Sensitivity $p(\text{reject } H_0   H_1 \text{ is true})$ Specificity $1 - \alpha$

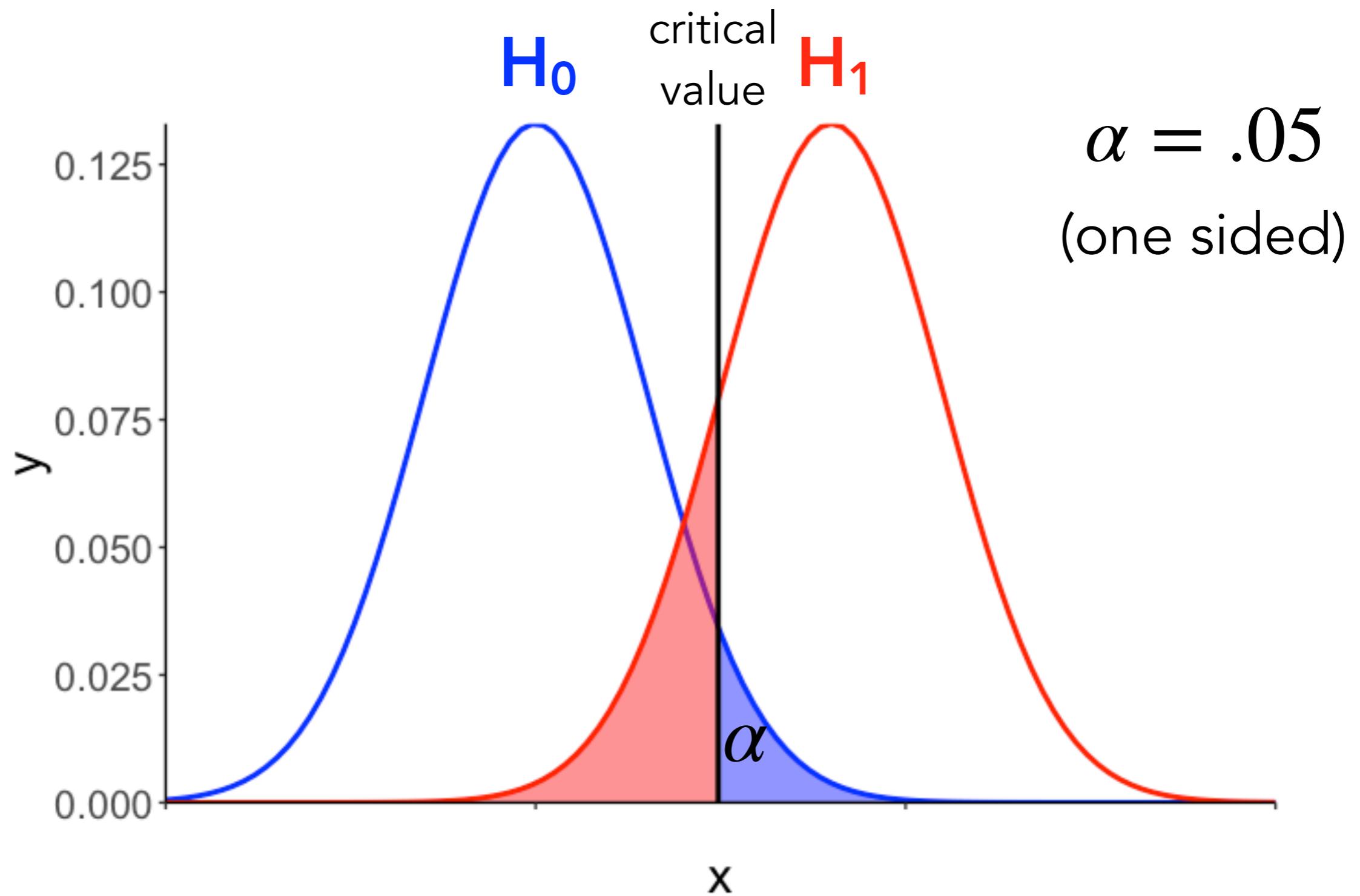
sampling distribution if  $H_0$  is true



# a affects power

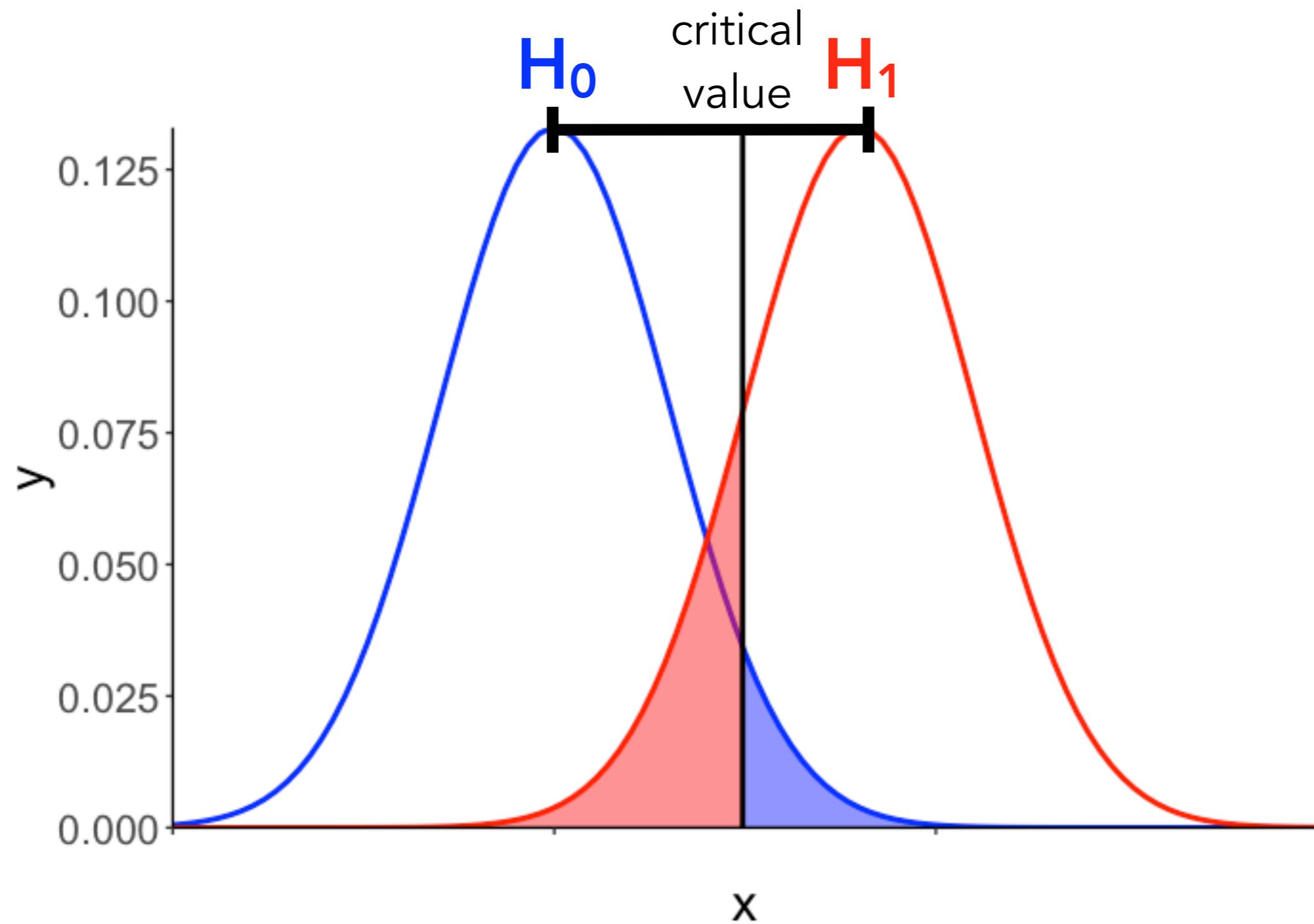
(e.g. p-value threshold)

false positive	Type I error $p(\text{reject } H_0   H_0 \text{ is true})$	$\alpha$	true positive	Sensitivity $p(\text{reject } H_0   H_1 \text{ is true})$	Power $1 - \beta$
false negative	Type II error $p(\text{not reject } H_0   H_1 \text{ is true})$	$\beta$	true negative	Specificity $p(\text{not reject } H_0   H_0 \text{ is true})$	$1 - \alpha$



distance between  
means affects power

<b>false positive</b>	<b>Type I error</b>	$\alpha$	<b>Power</b>
	$p(\text{reject } H_0   H_0 \text{ is true})$		$1 - \beta$
<b>false negative</b>	<b>Type II error</b>	$\beta$	
	$p(\text{not reject } H_0   H_1 \text{ is true})$		$1 - \alpha$



Power

 $1 - \beta$ 

Sensitivity

 $p(\text{reject } H_0 | H_1 \text{ is true})$ 

true positive

Specificity

 $p(\text{not reject } H_0 | H_0 \text{ is true})$ 

true negative

 $\alpha$ 

Type I error

 $p(\text{reject } H_0 | H_0 \text{ is true})$ 

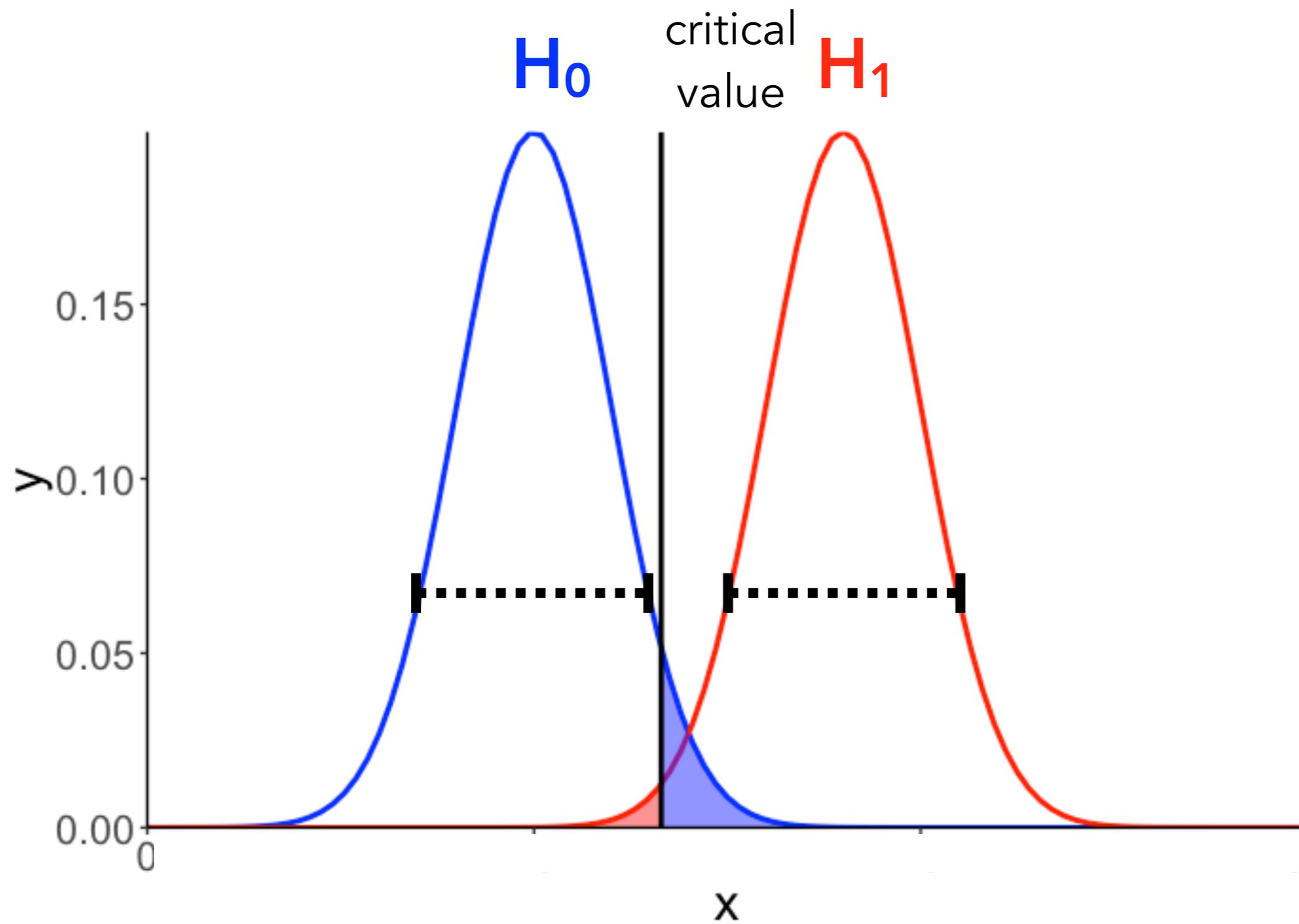
false positive

Type II error

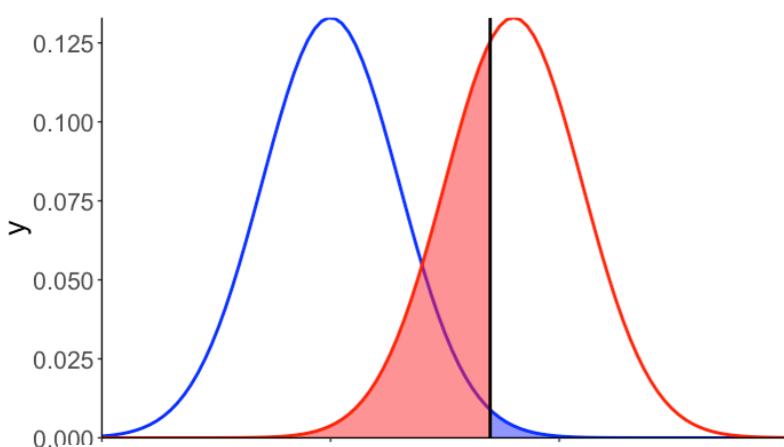
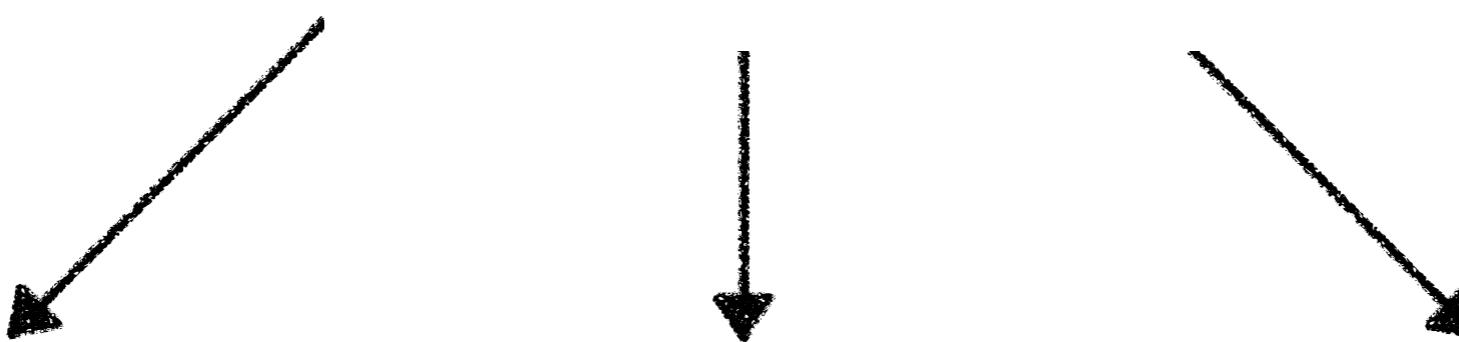
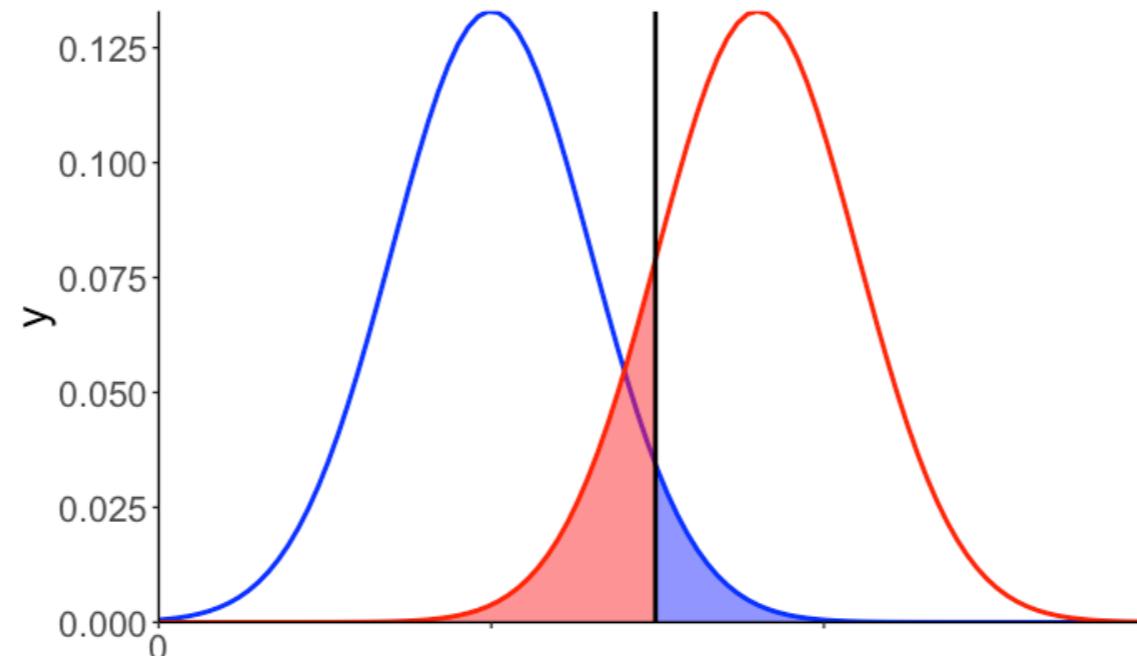
 $p(\text{not reject } H_0 | H_1 \text{ is true})$ 

false negative

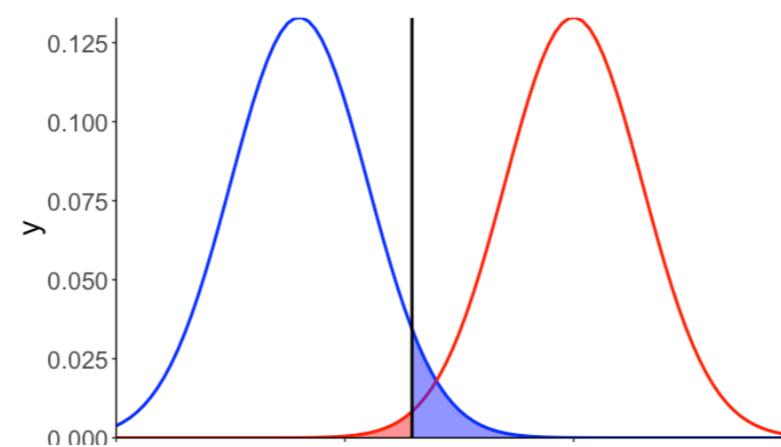
# variance affects power



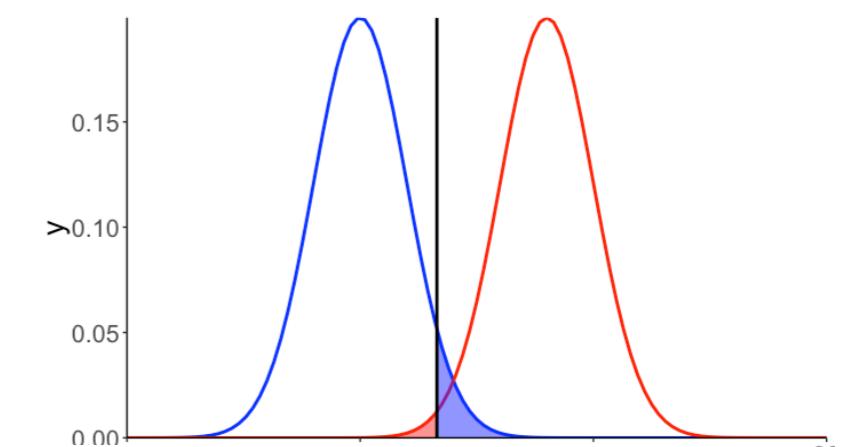
# The knobs we can turn to affect power



a threshold



effect size



sample size

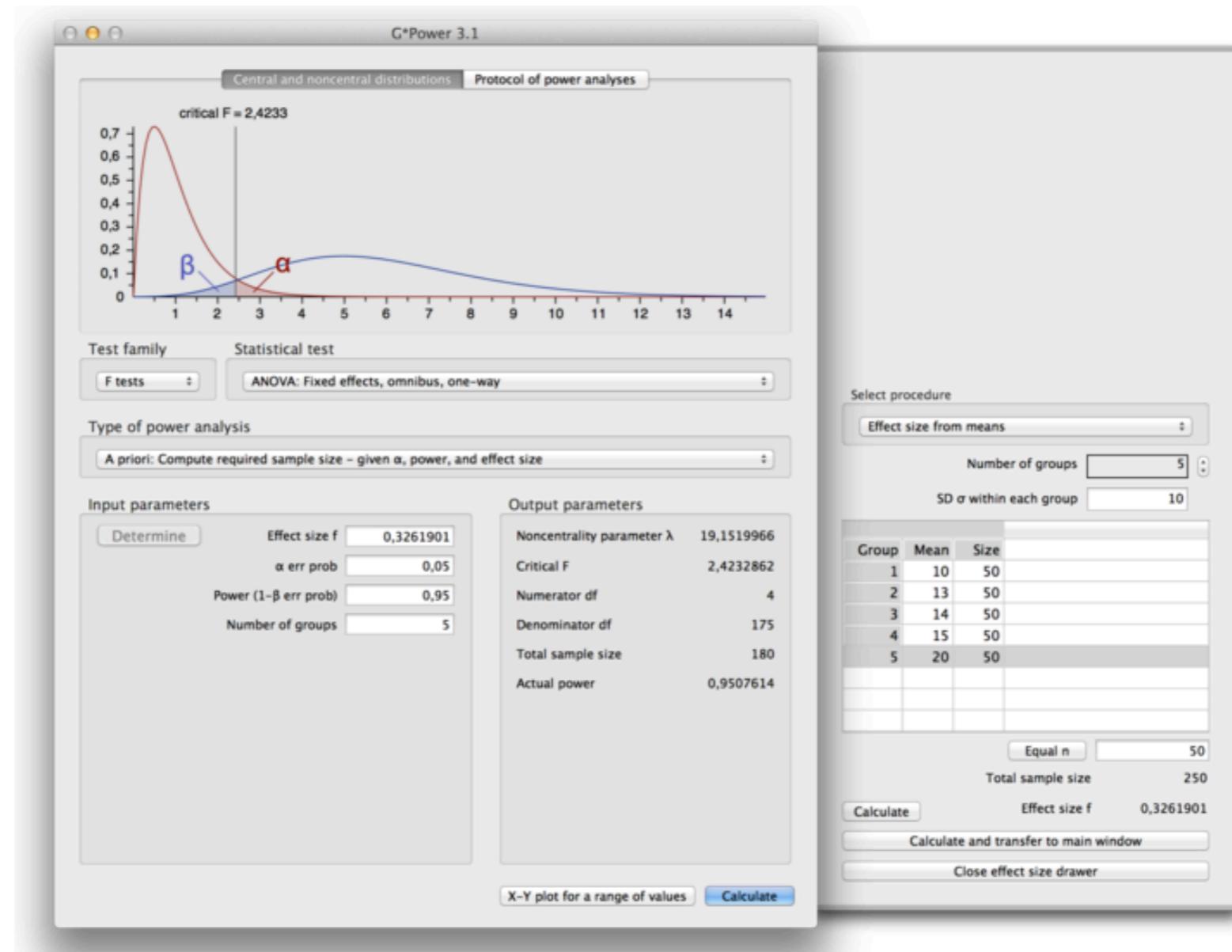
# How do we “turn the knobs” (determine power)?

*...how many participants do I need to run to have a good chance of detecting a true effect?*

# How do we “turn the knobs” (determine power)?

## Option 1

**G\*Power 3.1:** Graphical program for power calculations



<http://www.gpower.hhu.de/>

# How do we “turn the knobs” (determine power)?

## Option 2

**statsmodels**: The same library we've been using for **ols** has functions to calculate power

The screenshot shows the statsmodels 0.14.4 documentation page. The left sidebar contains a navigation menu with sections like "Installing statsmodels", "Getting started", "User Guide" (expanded), "Background", "Regression and Linear Models", "Time Series Analysis", "Other Models", "Statistics and Tools" (expanded), "Statistics stats" (selected), "Residual Diagnostics and Specification Tests", "Sandwich Robust Covariances", "Goodness of Fit Tests and Measures", "Non-Parametric Tests", "Descriptive Statistics", "Interrater Reliability and Agreement", "Multiple Tests and Multiple Comparison Procedures", "Basic Statistics and t-Tests with frequency weights", "Power and Sample Size Calculations" (selected), and "Proportion". The main content area is titled "Power and Sample Size Calculations" and describes the "power" module's implementation. It lists several functions with their descriptions:

Function	Description
<code>TTestIndPower (**kwds)</code>	Statistical Power calculations for t-test for two independent sample
<code>TTestPower (**kwds)</code>	Statistical Power calculations for one sample or paired sample t-test
<code>GofChisquarePower (**kwds)</code>	Statistical Power calculations for one sample chisquare test
<code>NormalIndPower ([ddof])</code>	Statistical Power calculations for z-test for two independent samples.
<code>FTestAnovaPower (**kwds)</code>	Statistical Power calculations F-test for one factor balanced ANOVA
<code>FTestPower (**kwds)</code>	Statistical Power calculations for generic F-test of a constraint
<code>normal_power_het (diff, nobs, alpha[, ...])</code>	Calculate power of a normal distributed test statistic
<code>normal_sample_size_one_tail (diff, power, alpha)</code>	explicit sample size computation if only one tail is relevant
<code>tt_solve_power ([effect_size, nobs, alpha, ...])</code>	solve for any one parameter of the power of a one sample

The top right corner shows the statsmodels logo, version 0.14.4, 10.5k stars, and 3.2k forks. A search bar is also present at the top.

# How do we “turn the knobs” (determine power)?

## Option 3

Power analysis via **simulation**

- “4 horsemen” of computational statistics (wk2 01-15)
  - Monte-carlo simulation
  - Bootstrapping
  - Permuting
  - Cross-validation

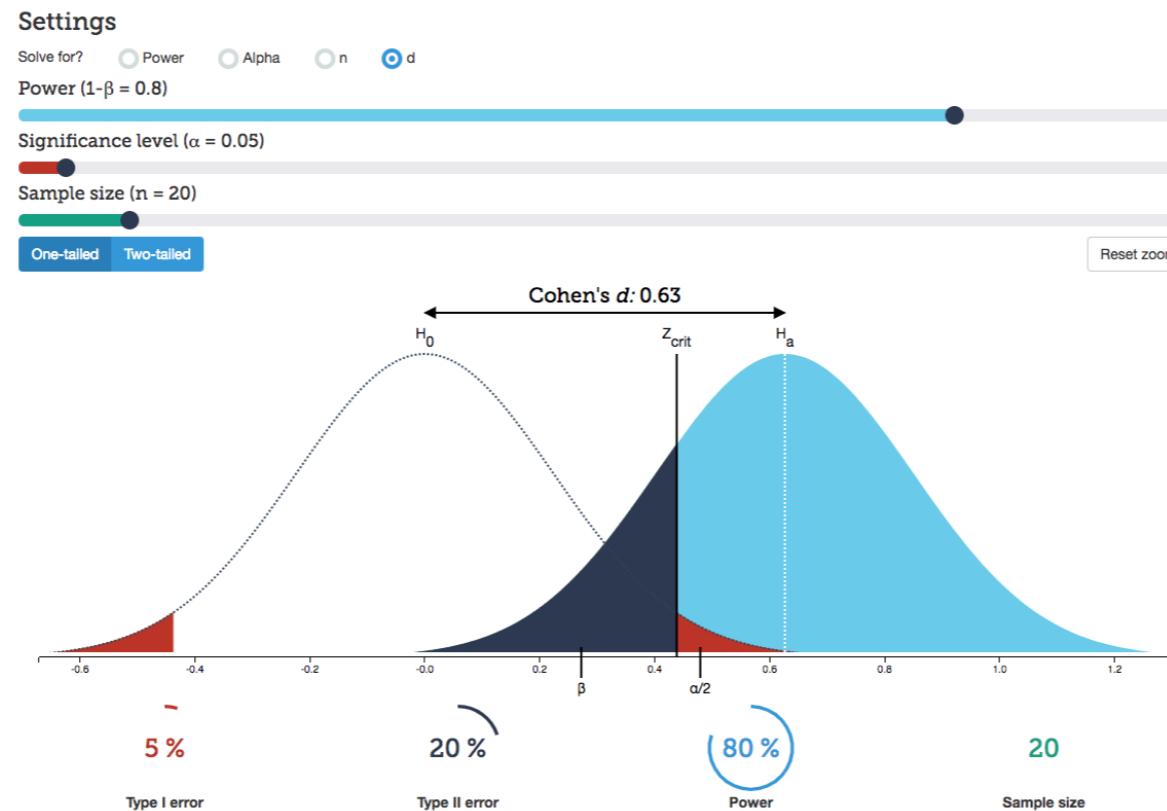
Next week's  
notebook

### Recipe

- pick:  $\alpha$ ,  $n$ , **effect-size**
- simulate a large number of data sets of size  $n$  with the specified **effect-size**
- for each data set, run a statistical test to calculate the *p-value*
- calculate **sensitivity/power**: proportion of times we *correctly rejected  $H_0$*  (given that  $H_1$  is true)

# Now: Break and interactive demos on your own (15min)

## After-break: Mini-group discussions



<https://tinyurl.com/power-interactive>

<https://tinyurl.com/power-exploreable>

# Mini-group Discussion (20 - 30 min)

- Please get in groups of 3-4
- Take turns discussing
  1. Most **interesting** thing you've learned
  2. Most **applicable** thing you've learned
  3. What **statistics** concept(s) are confusing?
  4. What **Python** concept(s) are confusing?
- **Individually:** fill out survey with your responses  
<https://tinyurl.com/struggle-survey>
- Nominate group member to share answers with the class

# Class Discussion

1. Most **interesting** thing you've learned
2. Most **applicable** think you've learned
3. What **statistics** concept(s) are confusing?
4. What **Python** concept(s) are confusing?

# Wrap-up

- **Mini-HW** (~1hr no coding!) <https://tinyurl.com/mini-hw>
  - Read article and answer Qs in **section 1.8.2 only**
  - Submit PDF to #mini-hw-submission channel on Slack
  - Deadline: **Monday**
- Upcoming
  - **Monday**: TA “struggles session” (Eshin away)
  - **March 12th**: Proposal Approval Deadline
    - You **must** meet with us **before** this to discuss proposal