PSYCH 201B

Statistical Intuitions for Social Scientists

Statistical Power & Group Discussion

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

02/26/2025

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

Final Project Proposal

- https://tinyurl.com/201b-proposal
- **March 12th:** proposal approval deadline
    - you must have **met with us and received approval** by this date
- **March 20th:** final project deadline; you will submit
    - Methods & Results section write-up **PDF**
    - Data Analysis **Notebook(s)**

Announcements

- Final Project Proposal
- HW 3 **Due tomorrow by 4pm**
- **No new HW** this week
- Instead, please check-out this week's readings

Readings (Monday & Tuesday's materials)

- Data Analysis: A Model Comparison Approach
  - Chapter 8: One-Way ANOVA: Models with a Single Categorical Predictor
  - Chapter 9: Factorial ANOVA: Models with Multiple Categorical Predictors and Product Terms
  - Chapter 10: ANCOVA: Models with Continuous and Categorical Predictors
- Regression and Other Stories
  - Chapter 10: Linear Regression with Multiple Predictors

Announcements

- Final Project Proposal
- HW 3 **Due tomorrow by 4pm**
- **No new HW** this week
- Instead, please check-out this week's readings
- After the BREAK we have *solution* notebooks from last week and a new notebook to complete
  - `06_models_solutions` (check your work today)
  - `07_models_solutions` (check your work today)
  - `08_models` (work through today)

What is statistical power?

…some background first

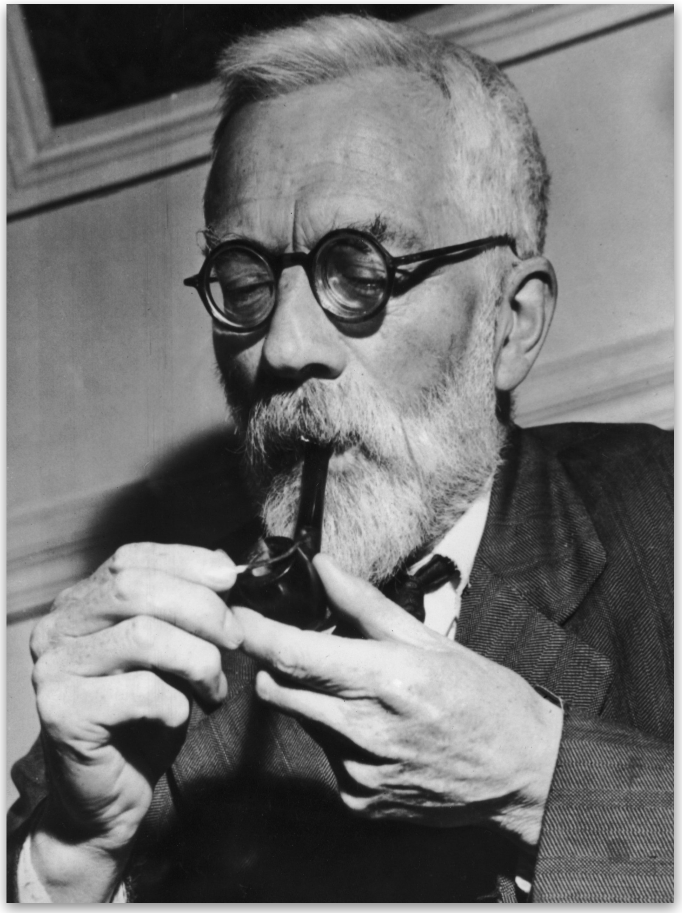
P-values are **evidence** for hypotheses

"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 **F**isher

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

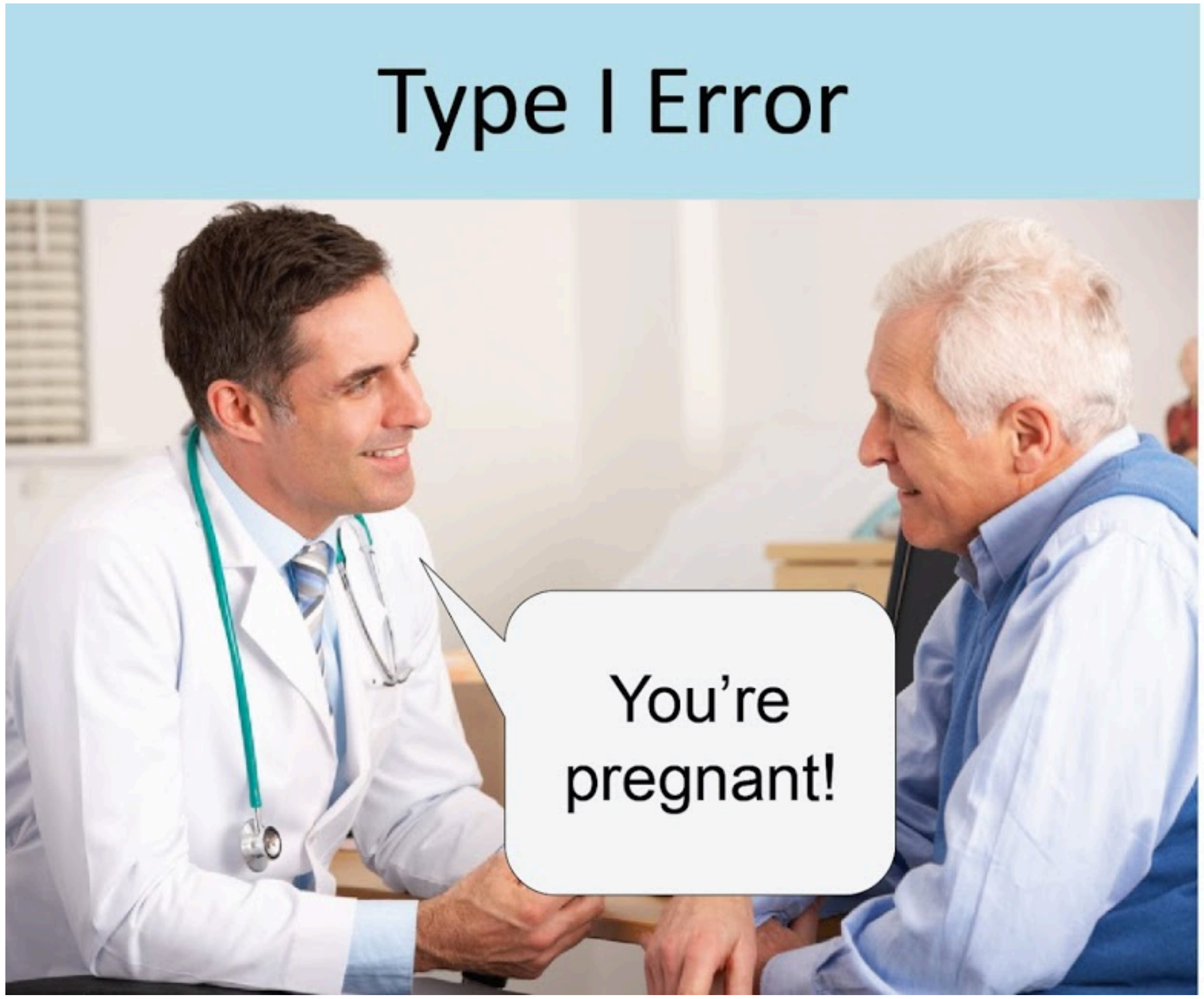
Jerzy Neyman   Karl Pearson

Neyman-Pearson Approach

- Two possible states of the world
  - H0 (null) is **true**
  - H0 (null) is **false**
- Two decisions we can make
  - **reject** H0
  - **retain** H0
- Two ways we can be wrong
  - **Reject** when H0 is **true** — False Positive **Type I** error α
  - **Retain** when H0 is **false** — False Negative **Type II** error β

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

H_0: Not pregnant **H_1**: Pregnant



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	**false positive**	Type I error $p(\text{reject } H_0 \mid H_0 \text{ is true})$	α
	false negative	Type II error $p(\text{not reject } H_0 \mid H_1 \text{ is true})$	β
Ways to be right	true positive	Sensitivity $p(\text{reject } H_0 \mid H_1 \text{ is true})$	Power $\boxed{1-\beta}$
	true negative	$p(\text{not reject } H_0 \mid H_0 \text{ is true})$ Specificity	$1-\alpha$

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

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

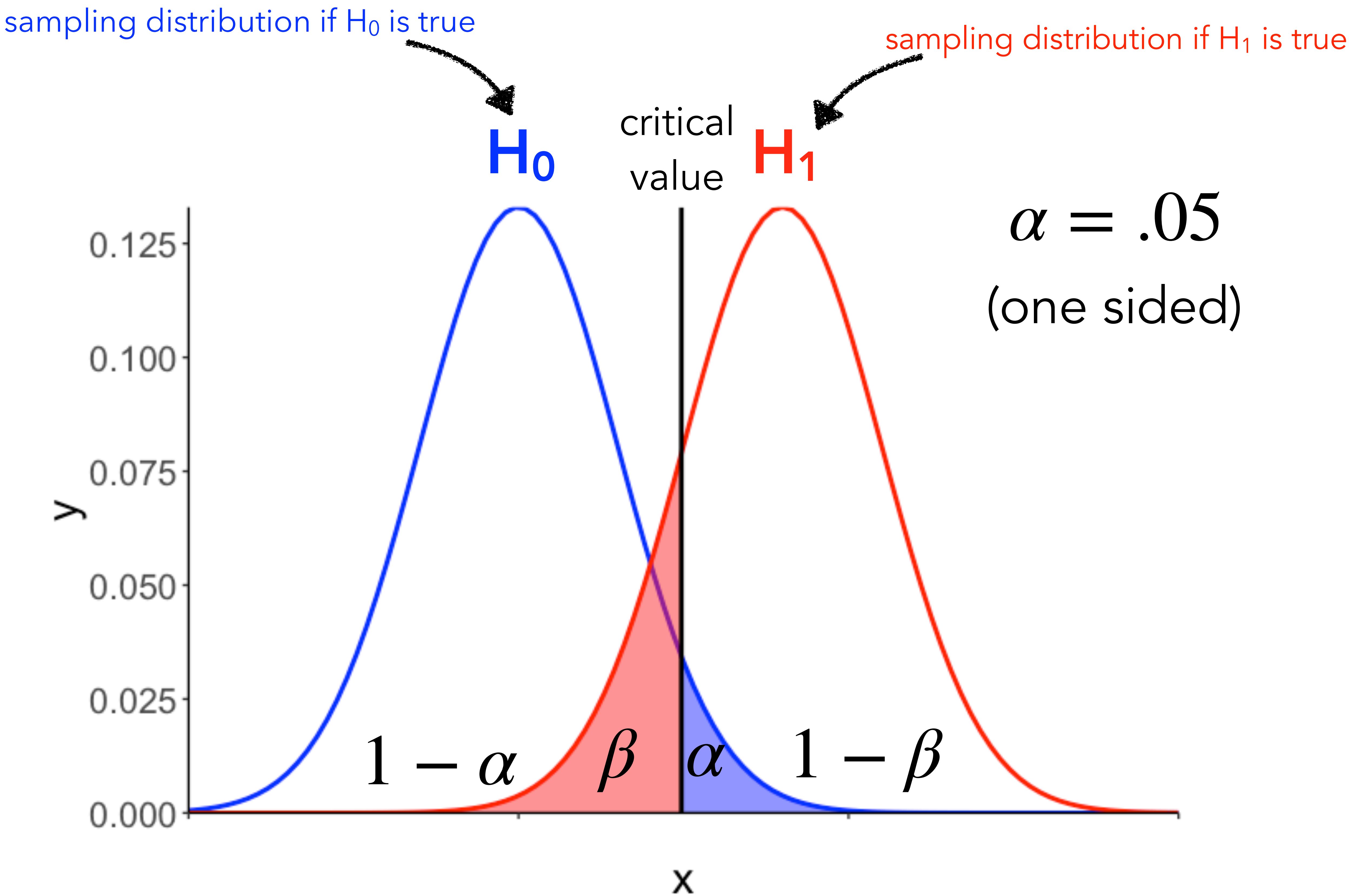
$$\beta_1 = 300$$

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

What affects power?

false positive	Type I error $p(\text{reject } H_0 \mid H_0 \text{ is true})$	α
false negative	Type II error $p(\text{not reject } H_0 \mid H_1 \text{ is true})$	β

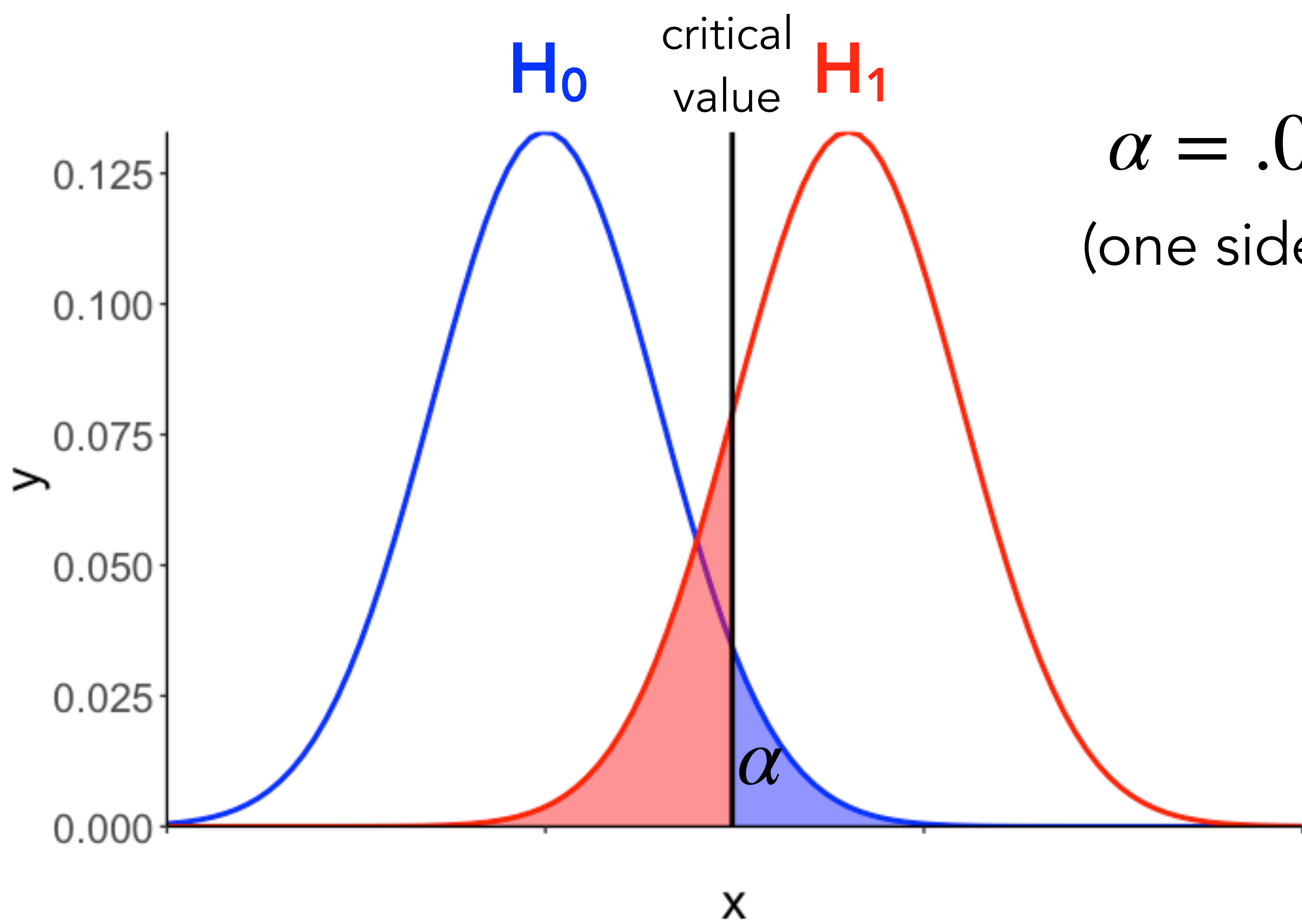
		Power
true positive	Sensitivity $p(\text{reject } H_0 \mid H_1 \text{ is true})$	$\boxed{1-\beta}$
true negative	$p(\text{not reject } H_0 \mid H_0 \text{ is true})$ Specificity	$1-\alpha$



α affects power

(e.g. p-value threshold)

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

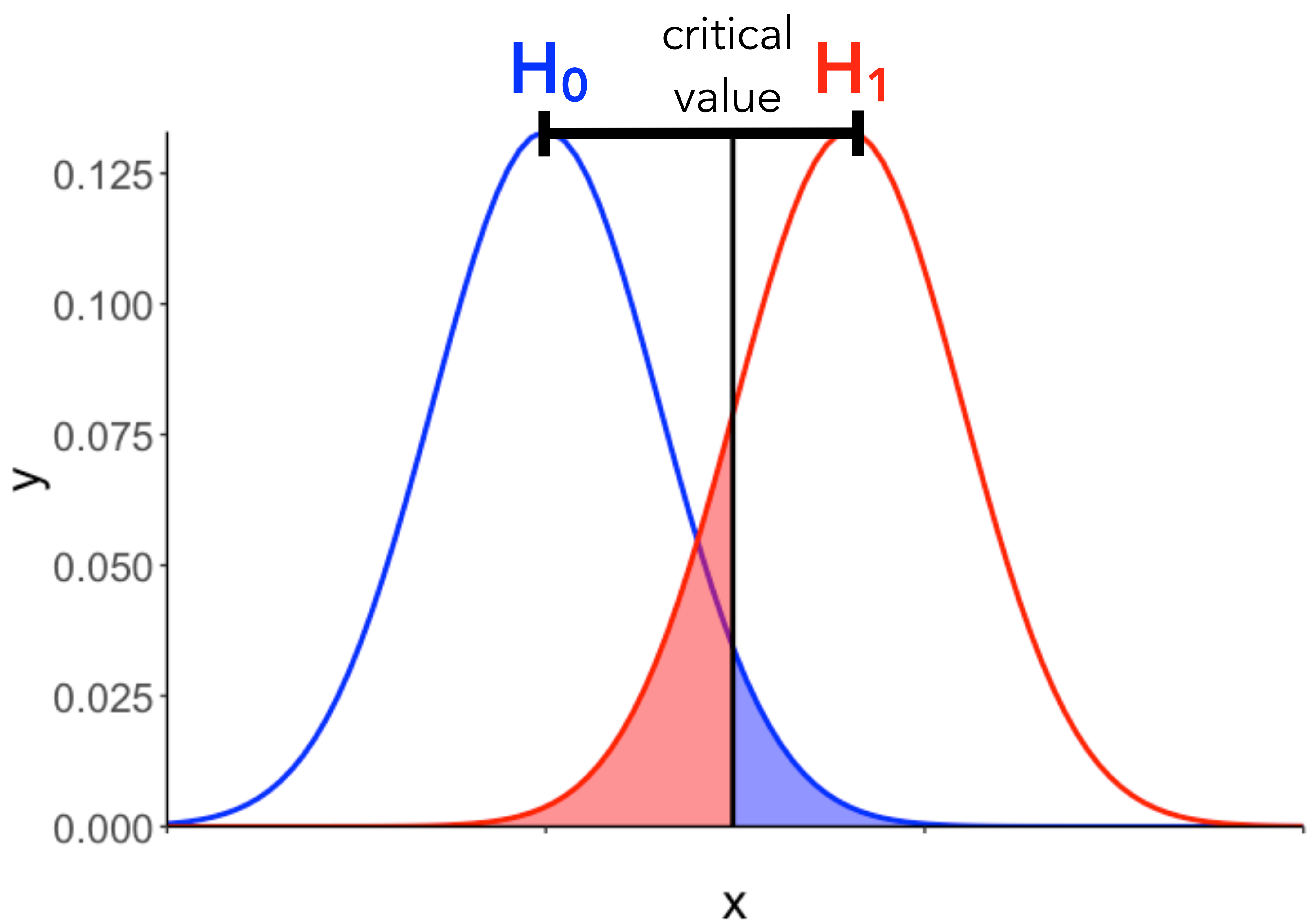


$\alpha = .05$

(one sided)

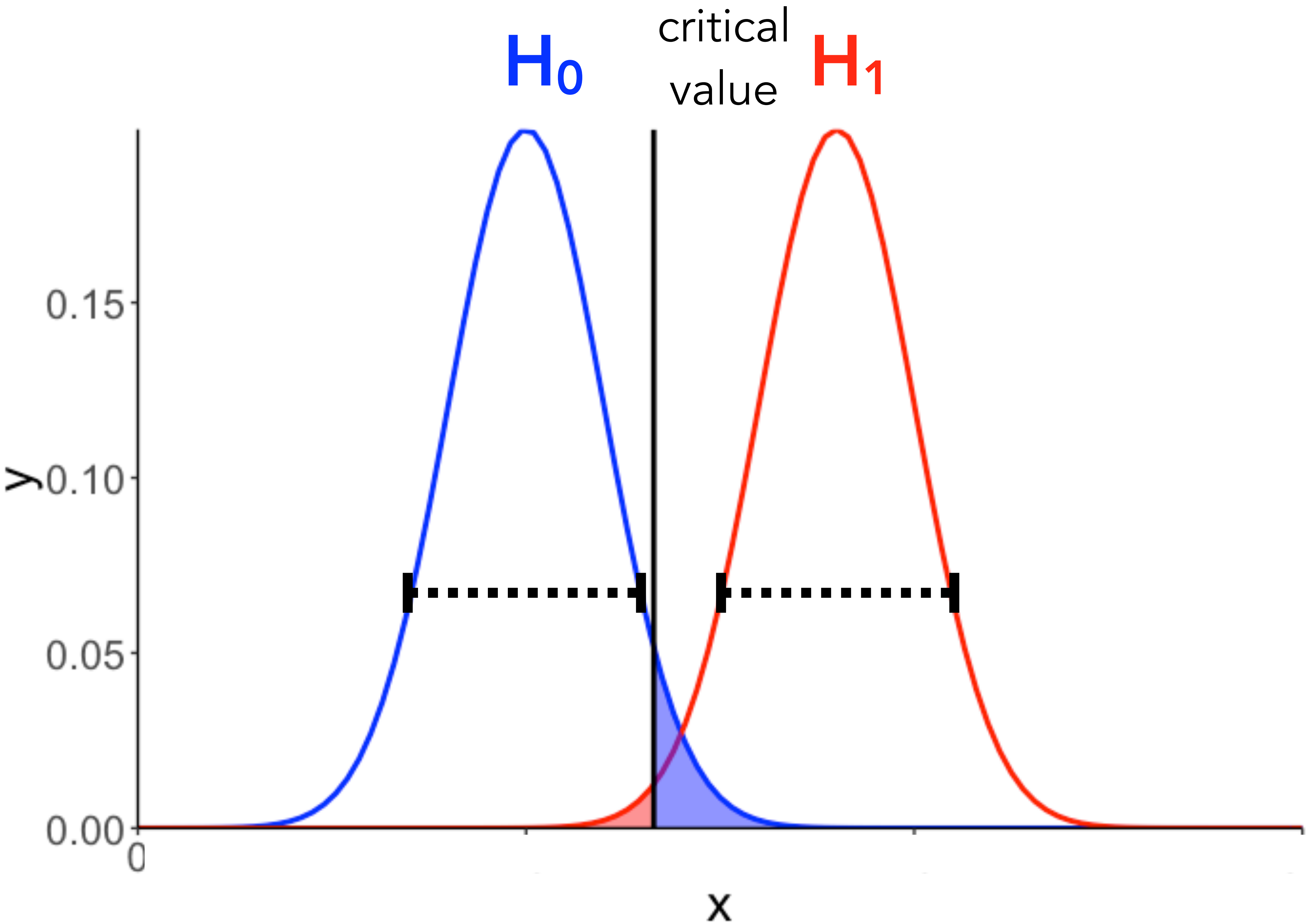
distance between means affects power

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

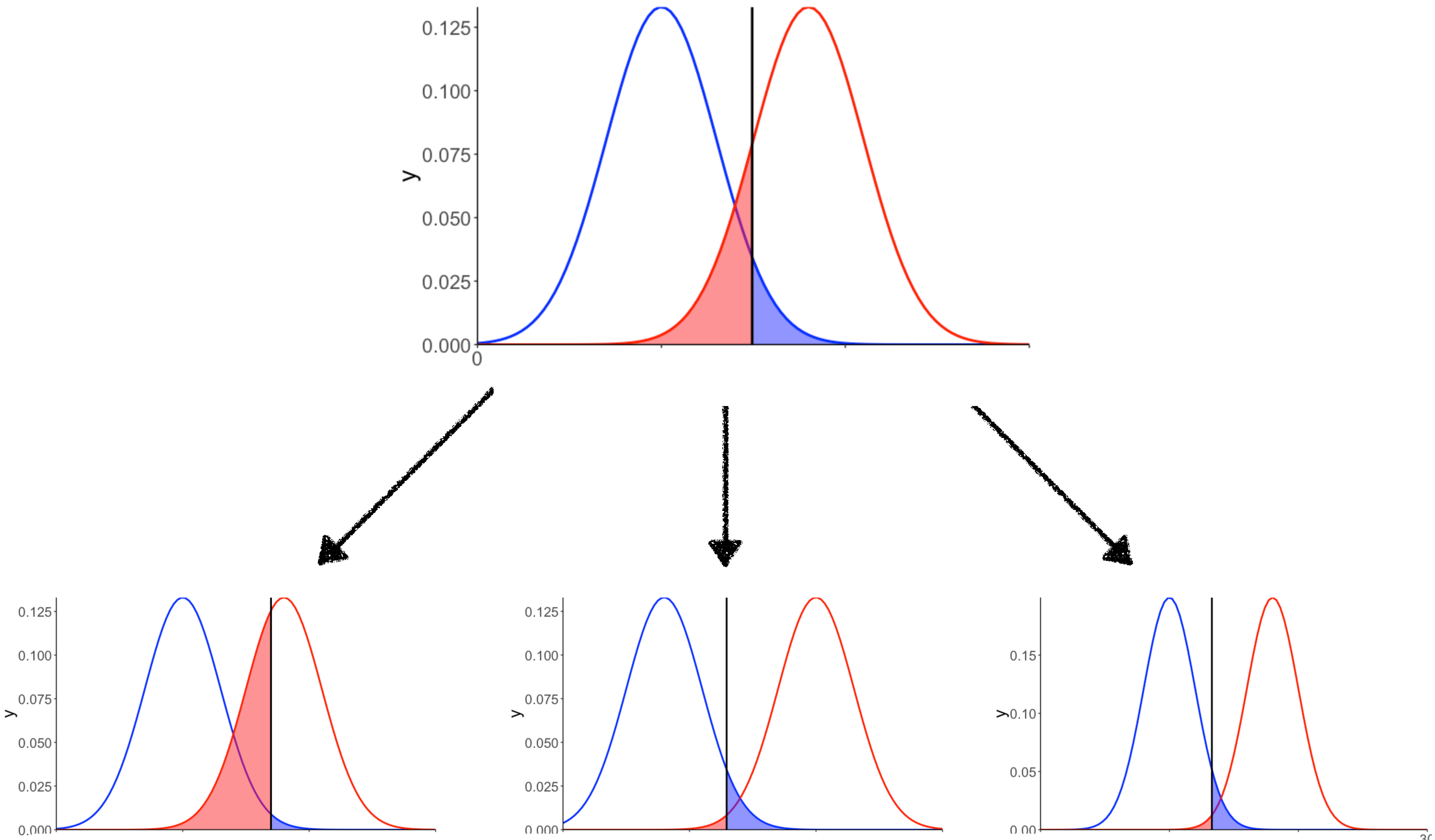


variance affects power

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



The knobs we can turn to affect power



α threshold effect size sample size

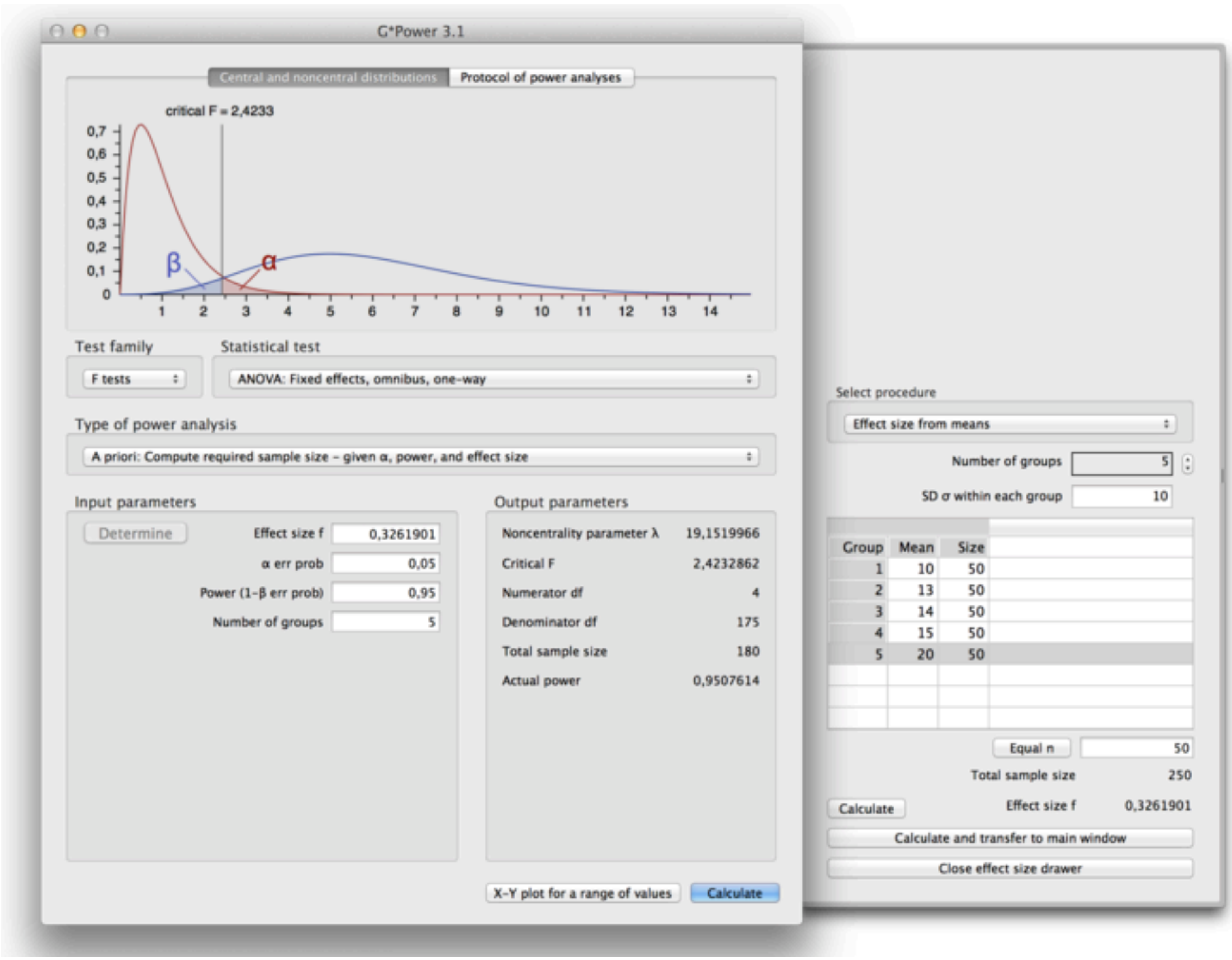
How do we “turn the knobs” (determine power)?

…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

Statistics stats   Search   statsmodels v0.14.4 10.5k 3.2k

statsmodels 0.14.4

Installing statsmodels
Getting started
User Guide
Background
Regression and Linear Models
Time Series Analysis
Other Models
Statistics and Tools
Statistics stats
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
Proportion

Power and Sample Size Calculations

The `power` module currently implements power and sample size calculations for the t-tests, normal based test, F-tests and Chisquare goodness of fit test. The implementation is class based, but the module also provides three shortcut functions, `tt_solve_power`, `tt_ind_solve_power` and `zt_ind_solve_power` to solve for any one of the parameters of the power equations.

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

Contents

Residual Diagnostics and Specification Tests
Outliers and influence measures
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
Proportion
Rates
Multivariate
Oneway Anova
Robust, Trimmed Statistics
Moment Helpers
Mediation Analysis
Oaxaca-Blinder Decomposition
Distance Dependence Measures
Meta-Analysis

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

Option 3

Power analysis via simulation

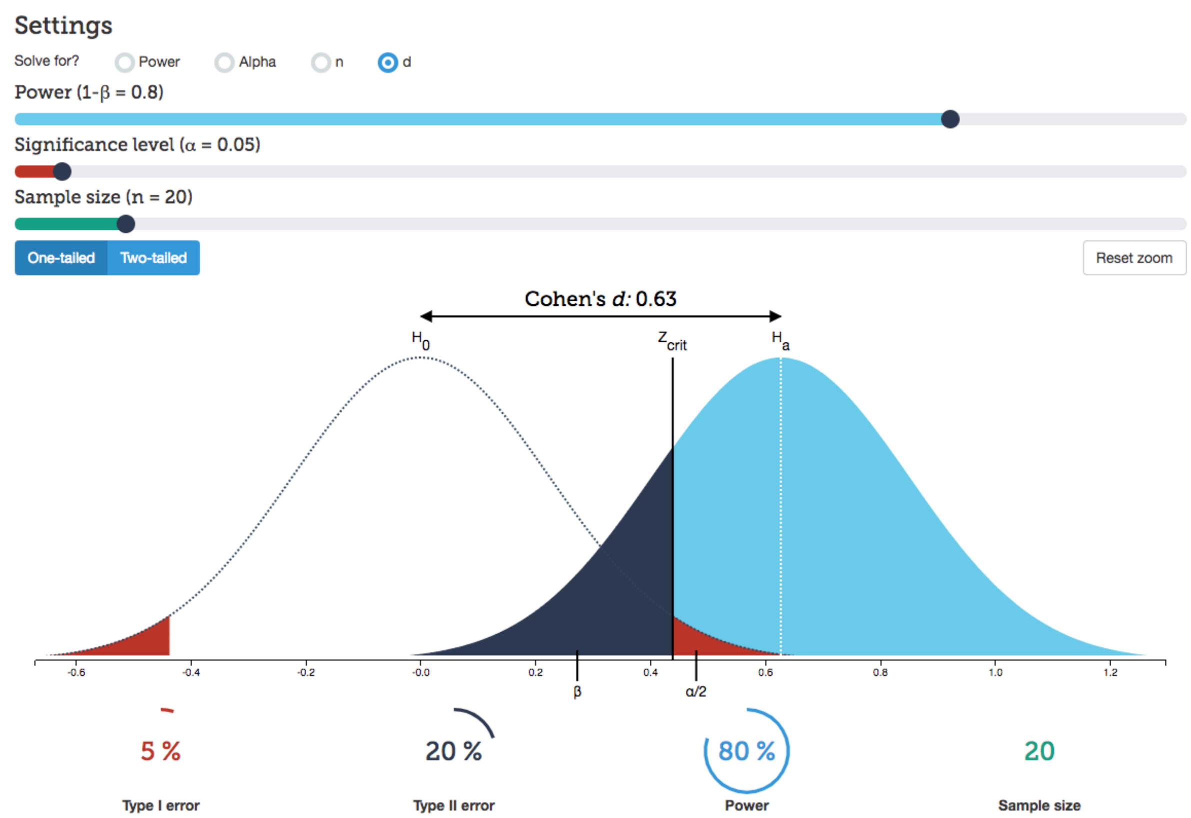
- “4 horsemen” of computational statistics (wk2 01-15)
  - **Monte-carlo simulation**
  - Bootstrapping
  - Permuting
  - Cross-validation

Next week’s notebook

Recipe

- pick: α, 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** think 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 multiple choice Qs at the end
  - 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 to get final approval