

STATISTICAL INFERENCE

Olga Vitek

College of Science
College of Computer and Information Science



Northeastern University

SCHEDULE

	March 10	March 11
9:00-9:30	Section1 : Lecture : motivating example, experimental design	Section6 : Differential Analysis of DDA Data with Skyline
9:30-10:00		Skyjam
10:00-10:30	break	break
10:30-11:00	Section2 : Introduction to Skyline	Section7: Lecture - Msstats (45 mins)
11:00-11:30	Section3 : Processing DDA Data with Skyline	Section8 : Hands-on : different abundance analysis
11:30-12:00		
12:00-12:30	Lunch	Lunch
12:30-13:00		
13:00-13:30	Section4 : Lecture : statistical inference, multiple testing	Section9 : Manual Inspection of Differential Results with Skyline
13:30-14:00		
14:00-14:30	break	break
14:30-15:00		
15:00-15:30	Section5 : Intro R, data exploration, basic statistics	Section10 : From DDA Quantification to SRM, PRM and DIA
15:30-16:00		

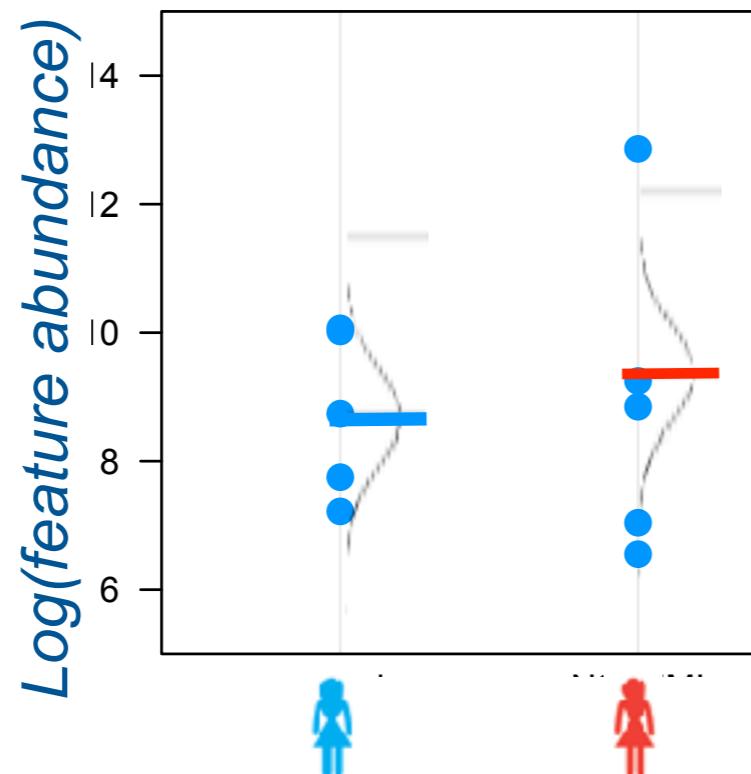
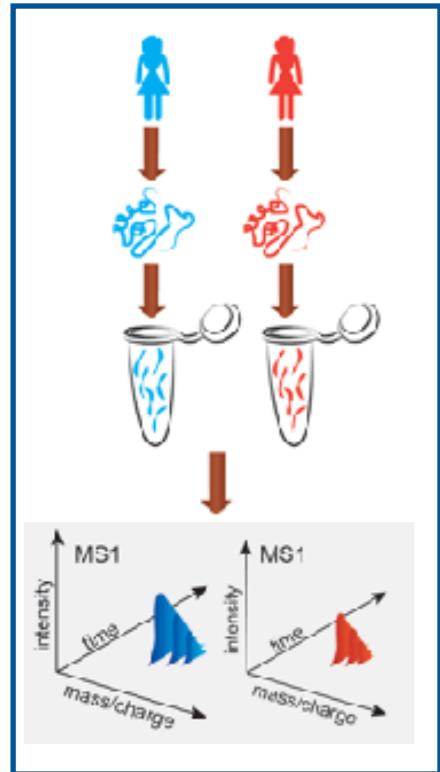
<https://goo.gl/EyPprx>

OUTLINE

- Basic statistical inference
 - T-test and p-values
- P-values: a word of caution
 - Instability, multiplicity, alternative approaches
- So how many replicates do I need?
 - Design of complex experiments

TWO-SAMPLE T-TEST

Simple example: label-free experiment, one feature/protein



$FoldChange = \frac{'typical' value in group 1}{'typical' value in group 2}$

$\log_2(FoldChange) = \log_2('typical' value in group 1) - \log_2('typical' value in group 2)$

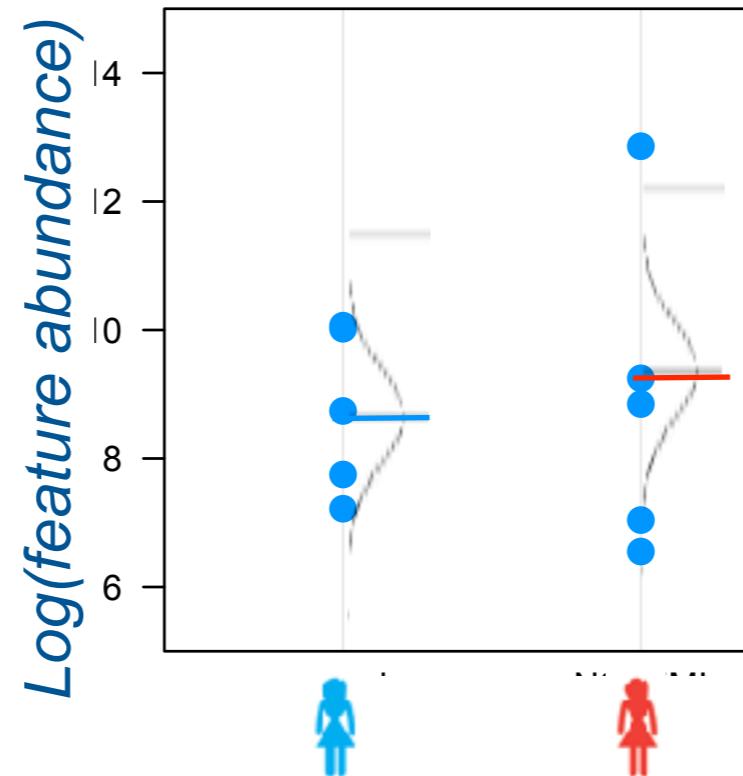
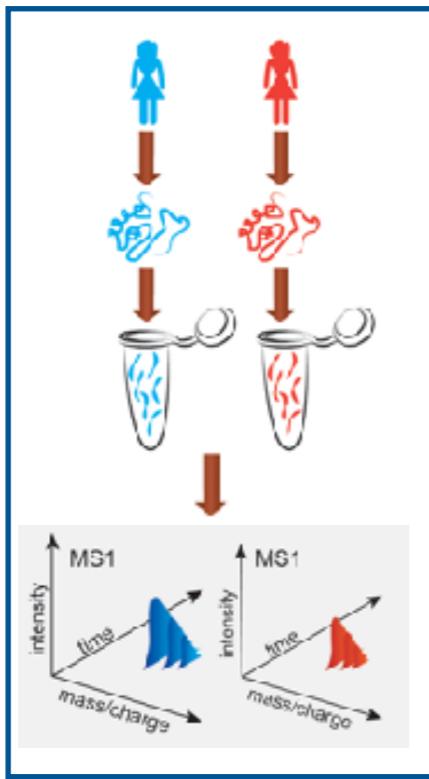
- $\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}$ = estimates log-fold change

$$\frac{1}{n_1} \sum_j Y_{1j} - \frac{1}{n_2} \sum_j Y_{2j} = \frac{1}{n_1} \sum_j \log_2 X_{1j} - \frac{1}{n_2} \sum_j \log_2 X_{2j} =$$
$$\log_2 \left(\prod_j X_{1j} \right)^{\frac{1}{n_1}} - \log_2 \left(\prod_j X_{2j} \right)^{\frac{1}{n_2}} = \log_2 \frac{\left(\prod_j X_{1j} \right)^{\frac{1}{n_1}}}{\left(\prod_j X_{2j} \right)^{\frac{1}{n_2}}}$$

Conclusion:
On log scale, estimates of FC are ratios of geometric means

TWO-SAMPLE T-TEST

Simple example: label-free experiment, one feature/protein



Sample means
in each group

H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$

H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

observed $t = \frac{\text{difference of group means}}{\text{estimate of variation}}$

$$\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}$$

$$\sqrt{s_1^2/n_1 + s_2^2/n_2}$$

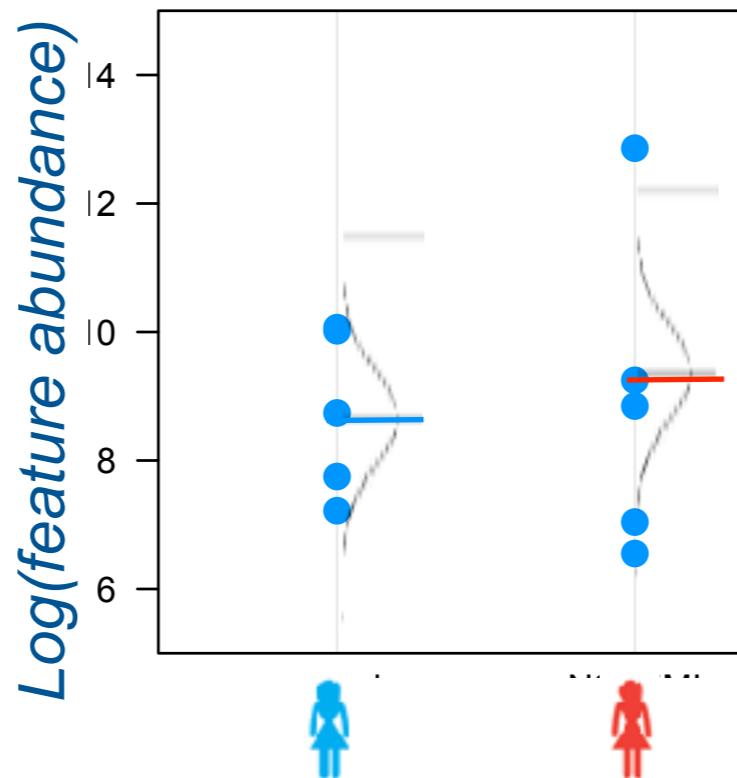
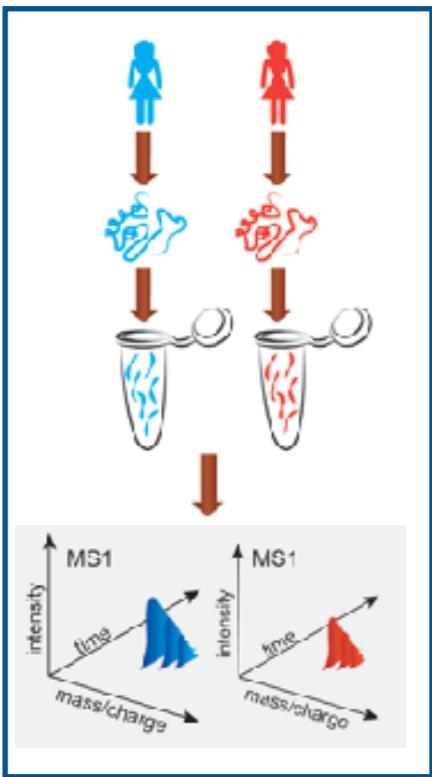
Number of
replicates

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_{1\cdot})^2$$

Sample variance

TWO-SAMPLE T-TEST

Simple example: label-free experiment, one feature/protein



Properties of the means

$$\frac{s_1^2}{n_1}$$

Variance of the sampling distribution of first mean

$$\sqrt{\frac{s_1^2}{n_1}}$$

Standard error of the first mean

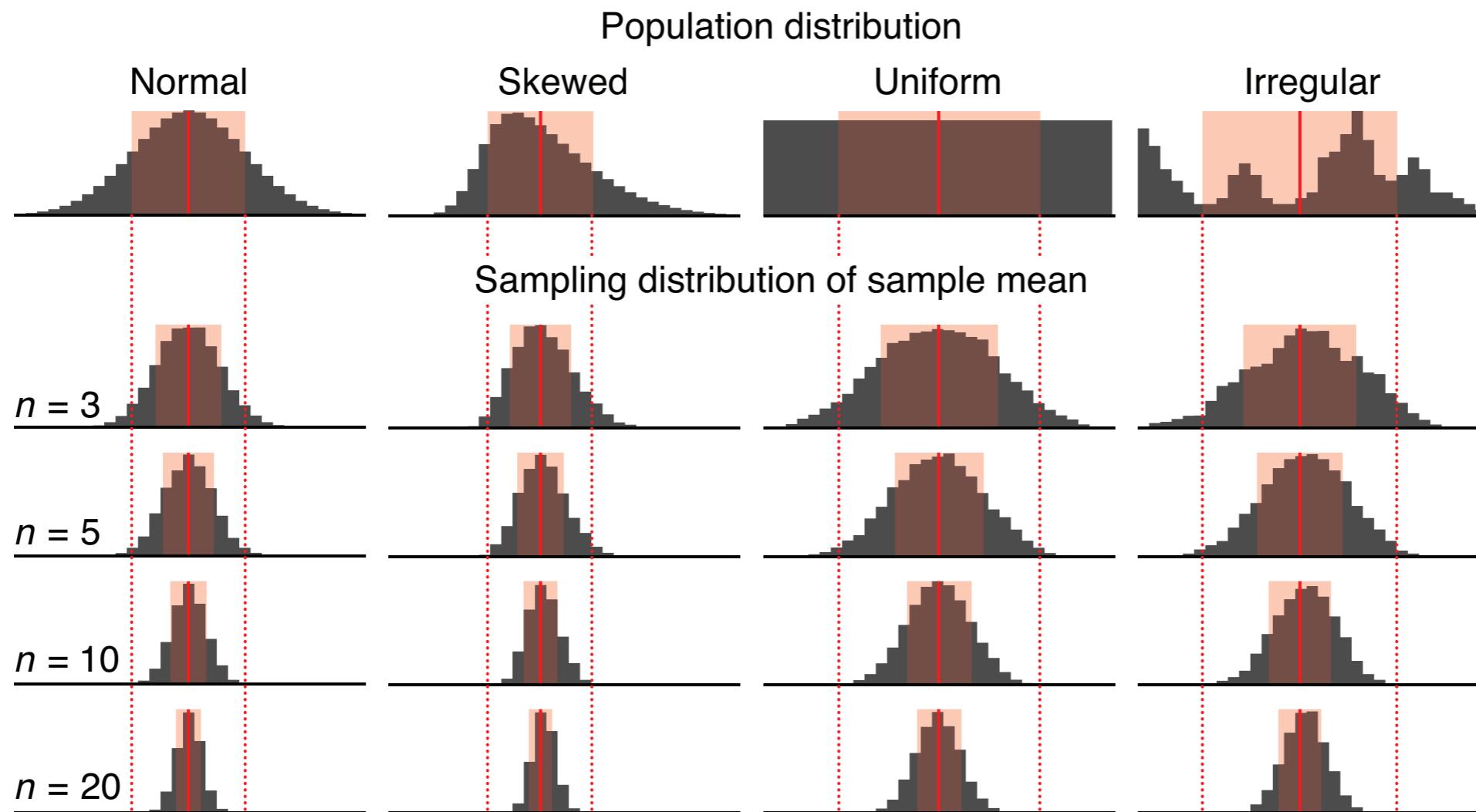
H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$
 H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

$$\text{observed } t = \frac{\text{difference of group means}}{\text{estimate of variation}} = \frac{\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_{1\cdot})^2$$

ASSUMPTION: NORMAL DISTRIBUTION

The Central Limit Theorem



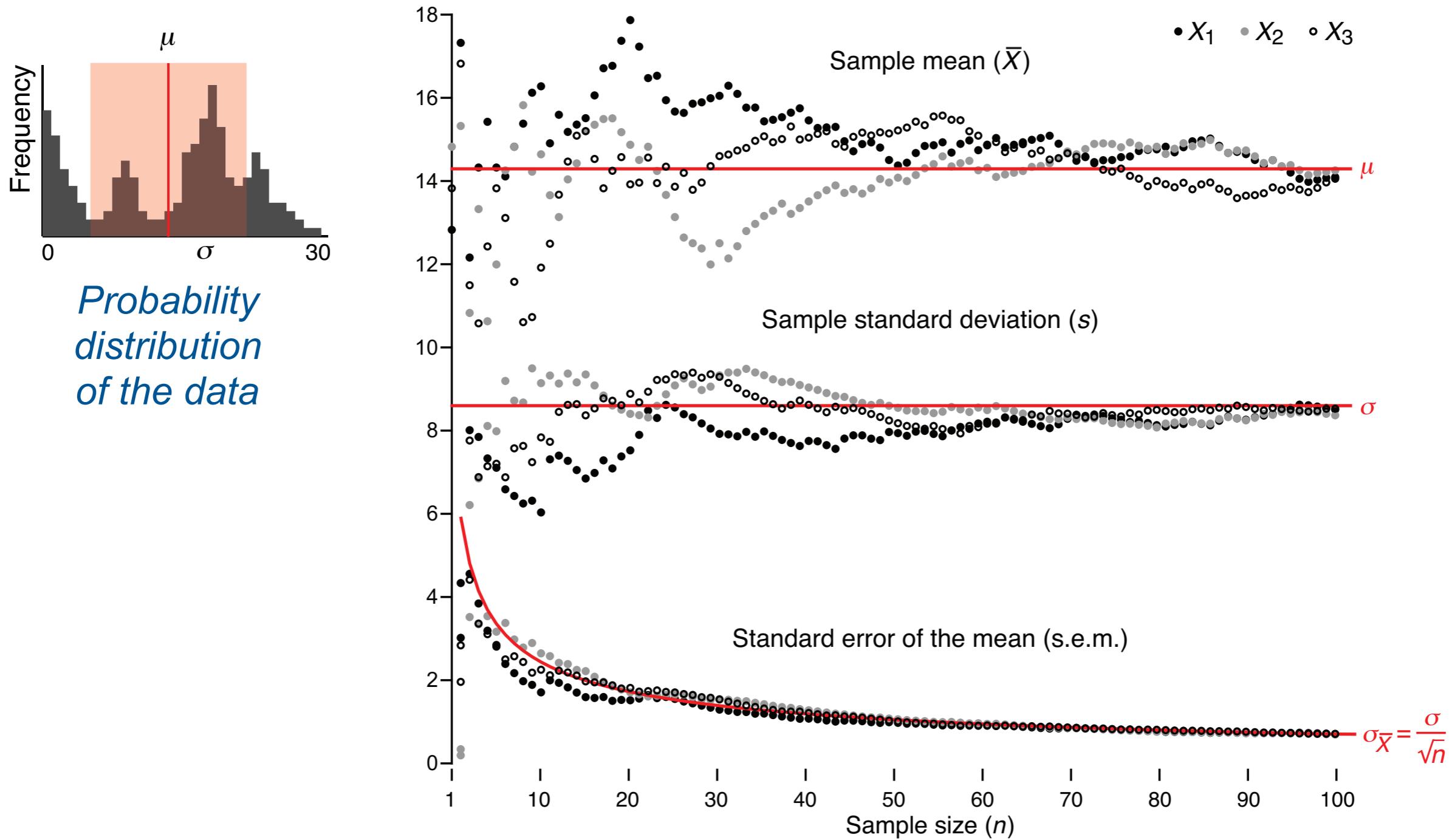
Probability distribution of the data

Repeatedly selecting n data points and calculating means

Conclusion:
As n increases, the mean is less variable and more Normal

EFFECT OF SAMPLE SIZE

As n increases, the estimates stabilize

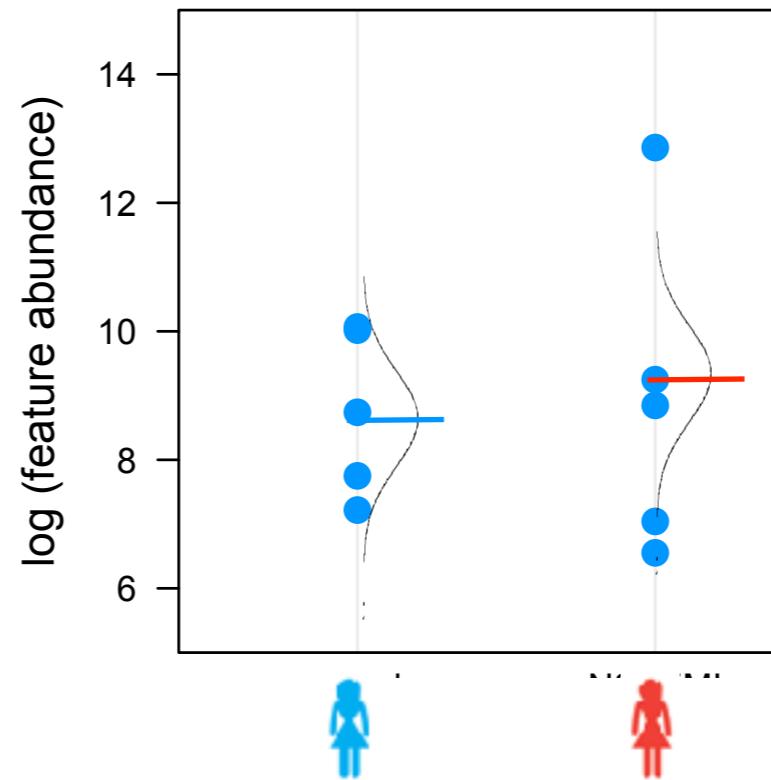
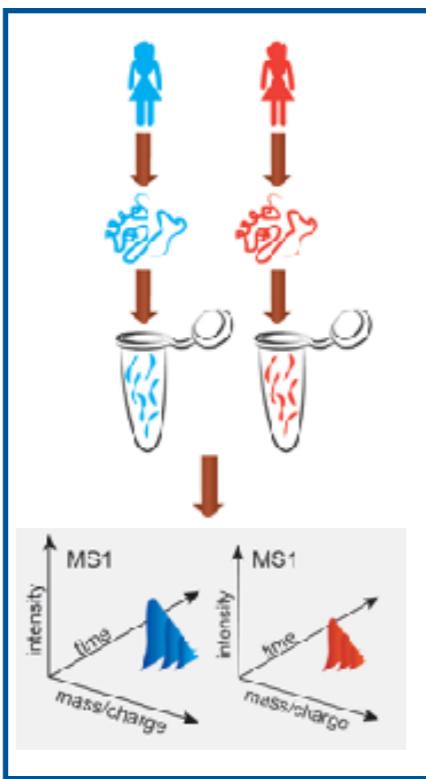


Simulated example

Krzywinski and Altman, Points of Significance Collection, *Nature Methods*

FINDING DIFFERENTIALLY ABUNDANT PROTEINS

False positive rate

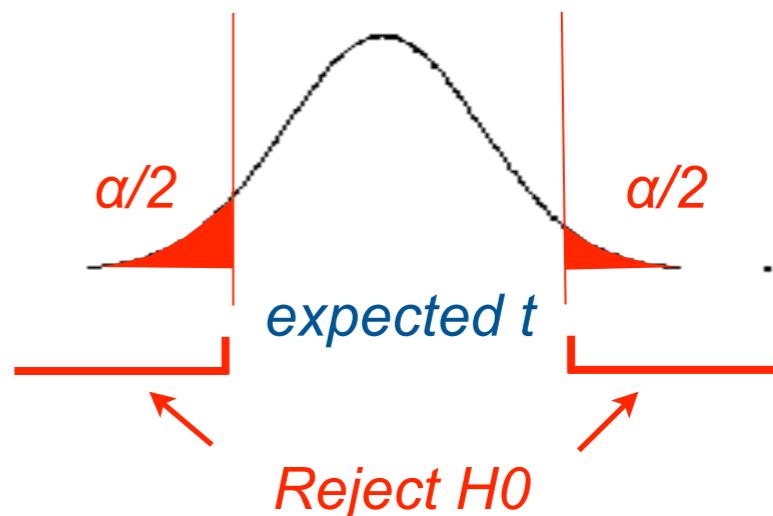


H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$
 H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

Distribution of the score if H_0 is true

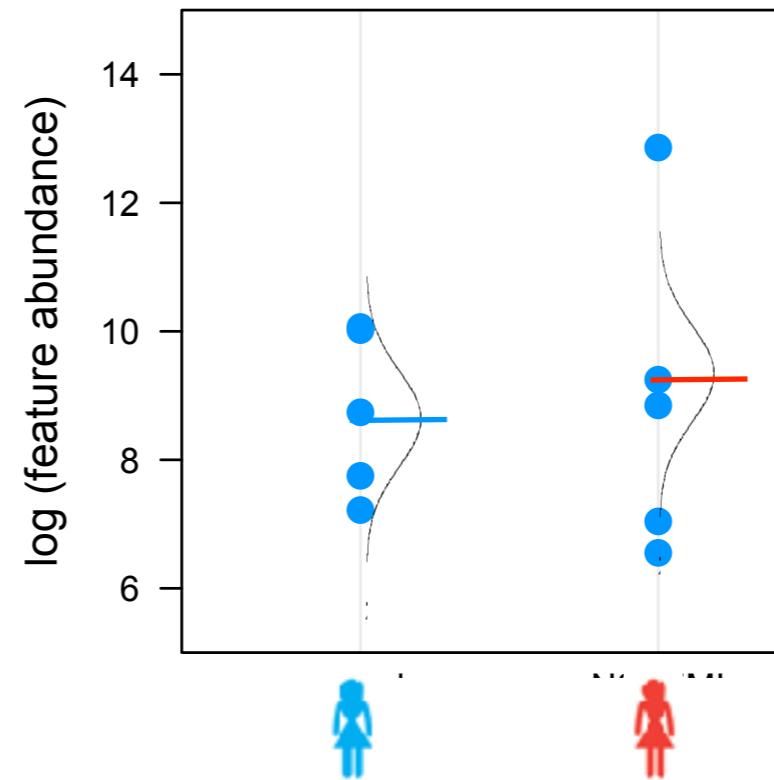
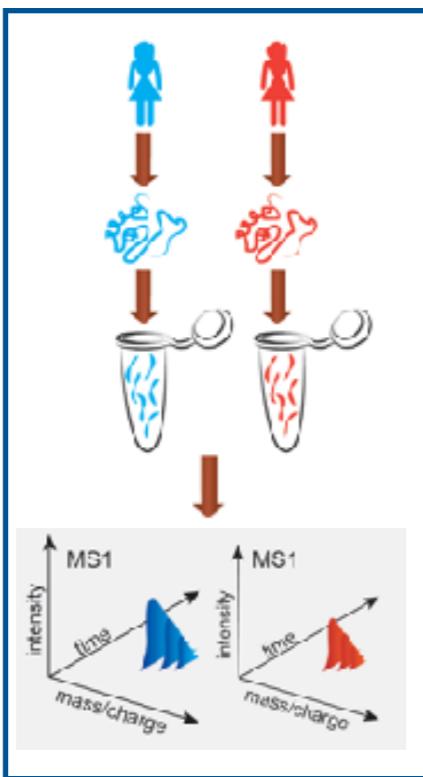
α = False Positive Rate

observed $t = \frac{\text{difference of group means}}{\text{estimate of variation}}$
no difference \sim Student distribution



FINDING DIFFERENTIALLY ABUNDANT PROTEINS

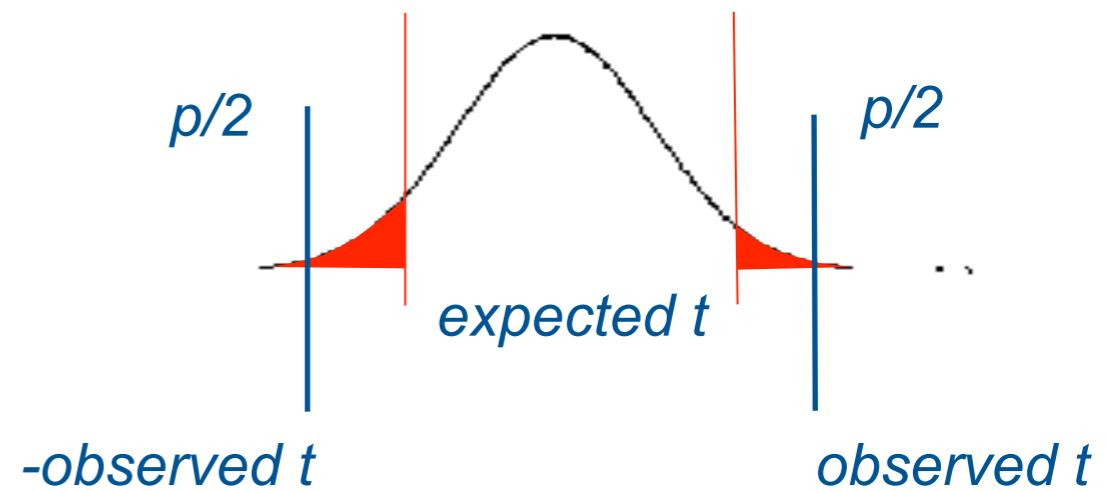
P-value



H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$
 H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

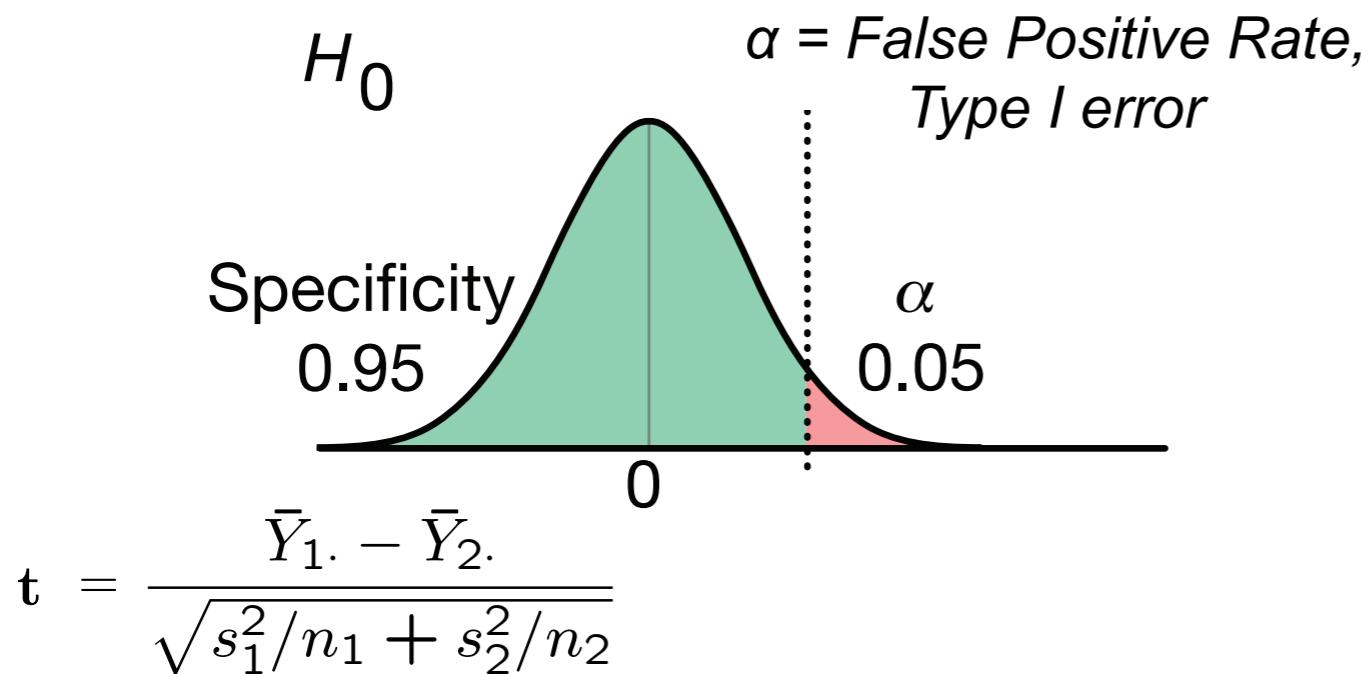
observed $t = \frac{\text{difference of group means}}{\text{estimate of variation}}$
no difference \sim Student distribution

Distribution of the score if H_0 is true



STATISTICAL POWER

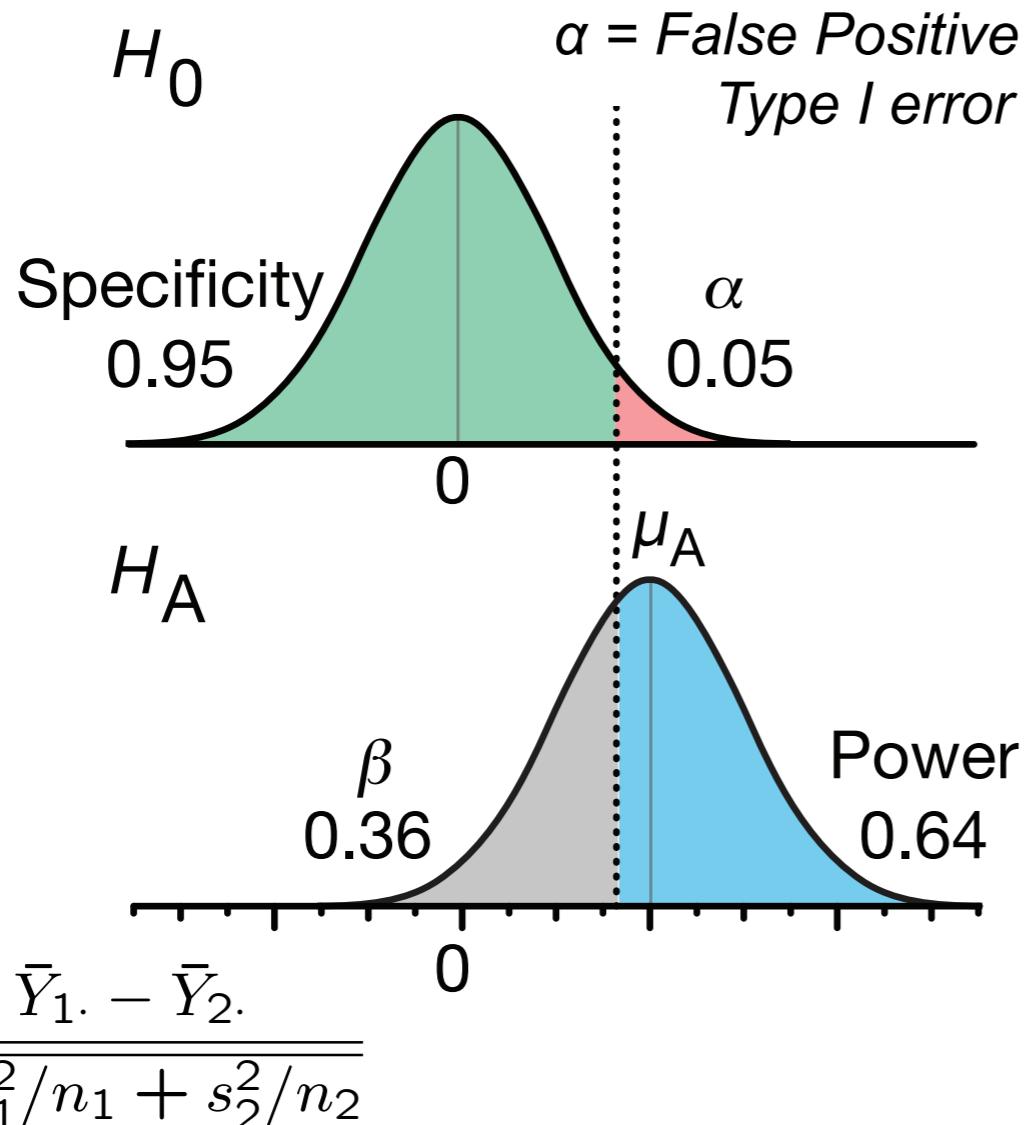
Probability to detect a difference when it exists



- Correct inference
 - Specificity, $1 - \alpha$
 - Power, sensitivity, $1 - \beta$
- Incorrect inference
 - Type I error, α
 - Type II error, β

STATISTICAL POWER

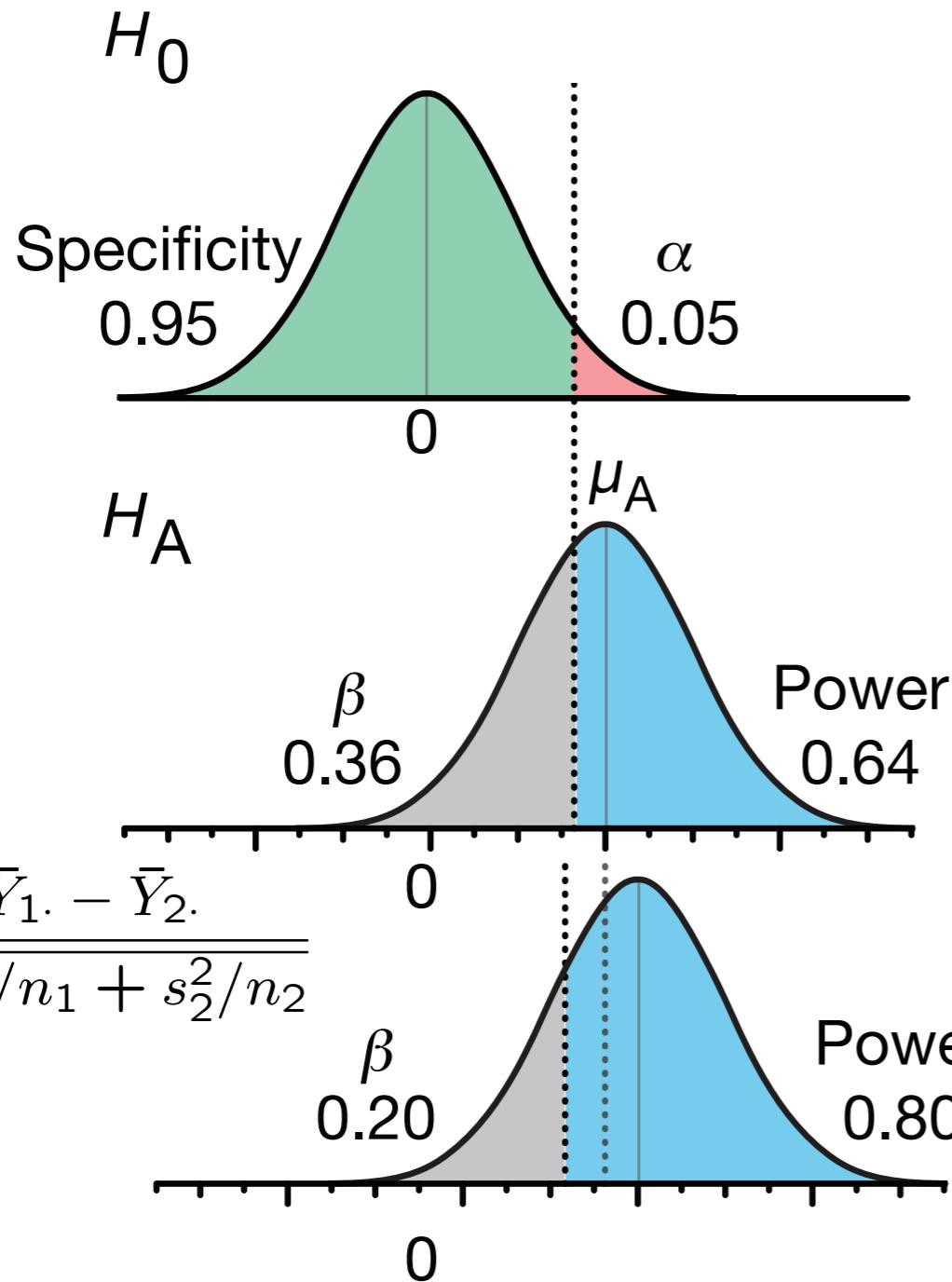
Probability to detect a difference when it exists



- Correct inference
 - Specificity, $1 - \alpha$
 - Power, sensitivity, $1 - \beta$
- Incorrect inference
 - Type I error, α
 - Type II error, β

STATISTICAL POWER

Probability to detect a difference when it exists

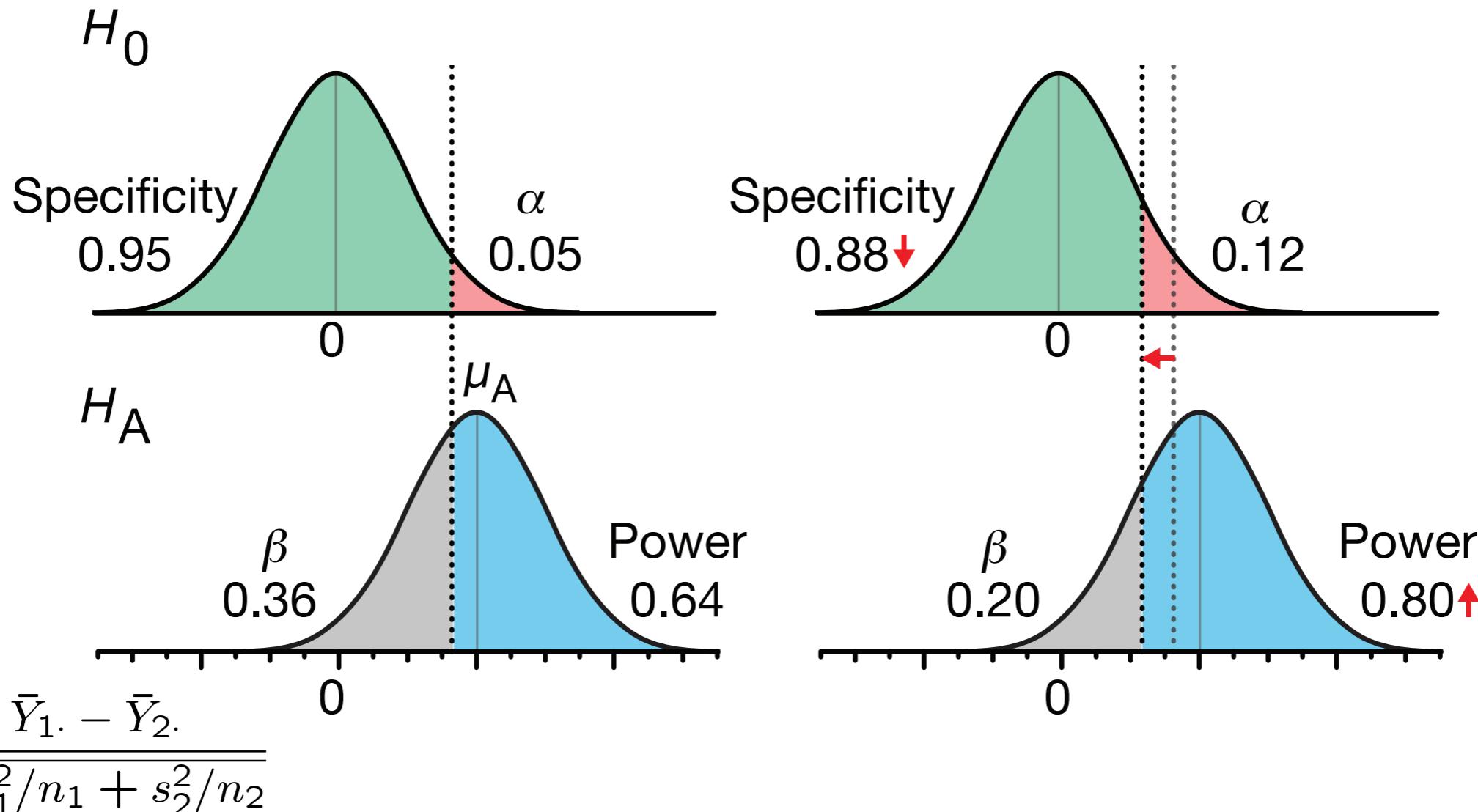


- Correct inference
 - Specificity, $1 - \alpha$
 - Power, sensitivity, $1 - \beta$
- Incorrect inference
 - Type I error, α
 - Type II error, β

sample size increases
statistical power

STATISTICAL POWER

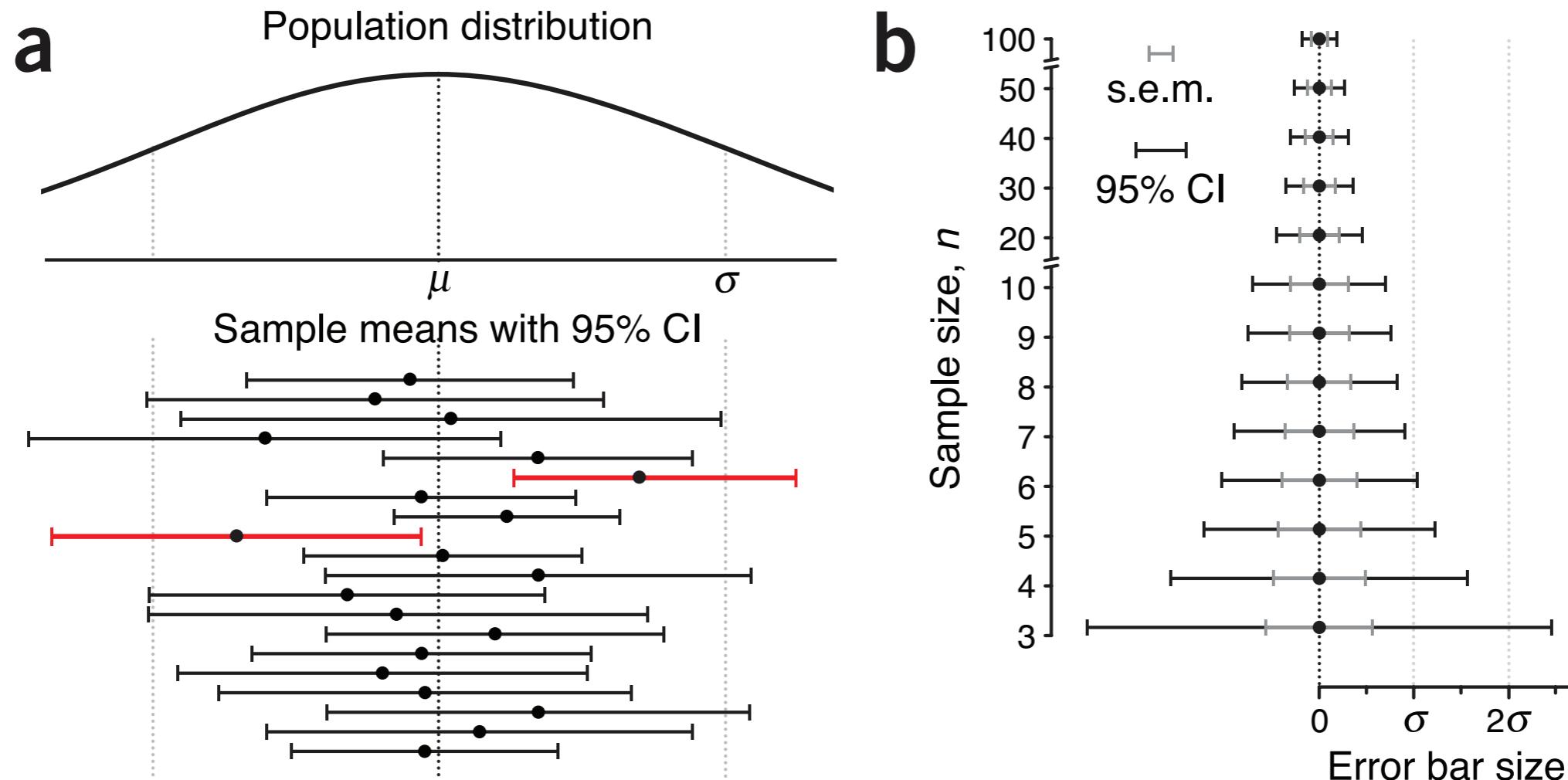
Probability to detect a difference when it exists



α trades off the sensitivity and the specificity of the test

ALTERNATIVE TO TESTING: CONFIDENCE INTERVALS

Not all error bars are made equal



A 95% CI: if we repeatedly collect data and draw confidence intervals, then 95% of them will contain the true mean

$$\left[(\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}) - t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}) + t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

CI are wider than bars indicating standard error of the mean!

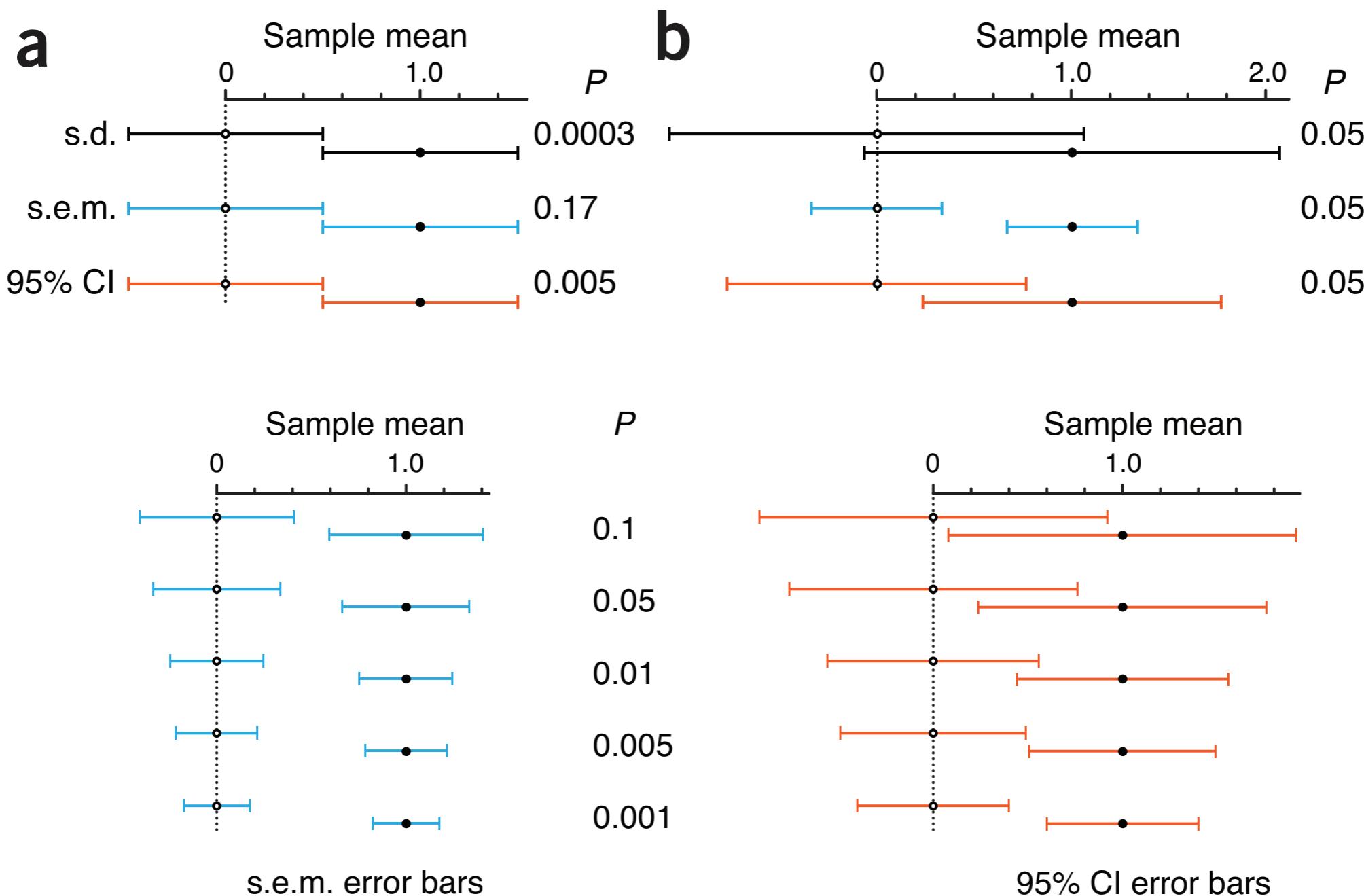
Width of the intervals depends on the sample size

Simulated example

Krzywinski and Altman, Points of Significance Collection, *Nature Methods*

ERROR BARS PROVIDE DIFFERENT INSIGHT

Absence of overlap does not always mean stat. significance



Simulated example

Krzywinski and Altman, Points of Significance Collection, *Nature Methods*

OUTLINE

- Basic statistical inference
 - T-test and p-values
- P-values: a word of caution
 - Instability, multiplicity, alternative approaches
- So how many replicates do I need?
 - Design of complex experiments

AMERICAN STATISTICAL ASSOCIATION (ASA) STATEMENT ON STATISTICAL SIGNIFICANCE AND P-VALUES

The American Statistician, February 2016

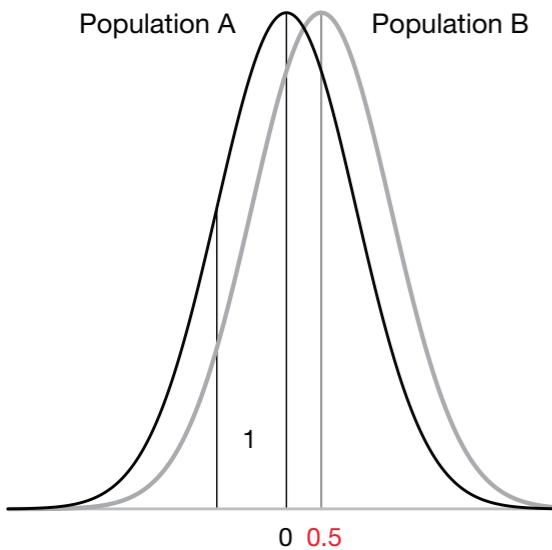
- P-values can indicate how incompatible the data are with a specified statistical model
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance
- Scientific conclusions and business policy decisions should not be based only on whether a p-value passes a specific threshold

AMERICAN STATISTICAL ASSOCIATION (ASA) STATEMENT ON STATISTICAL SIGNIFICANCE AND P-VALUES

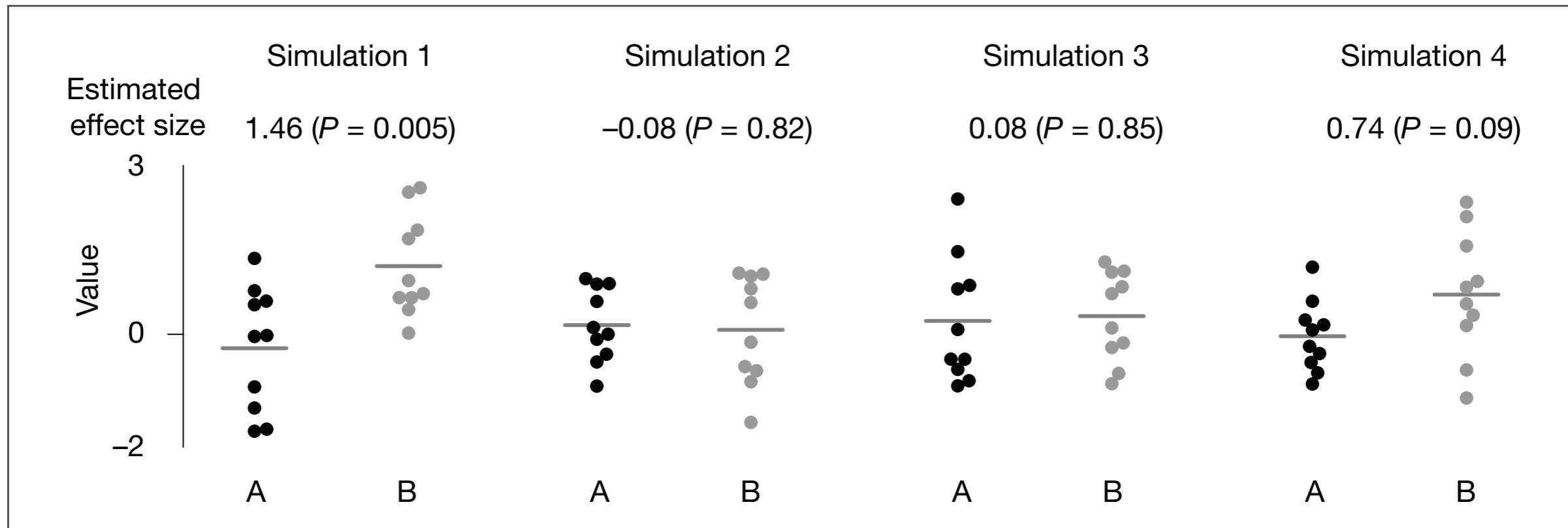
The American Statistician, February 2016

- Proper inference requires full reporting and transparency
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result
- By itself, a p-value does not provide a good measure of evidence regarding a model or a hypothesis

WITH SMALL SAMPLE SIZE, P-VALUES ARE UNSTABLE



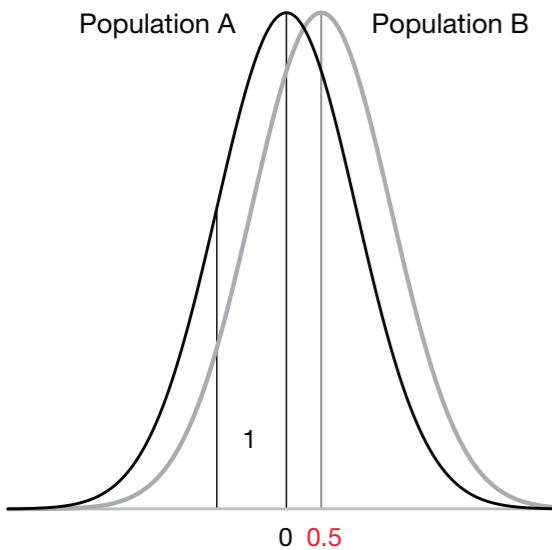
- Repeatedly sampling data leads to different results
- The problem worsens when testing many proteins
- Partial solutions:
 - Larger sample size
 - Adjustment for multiple testing



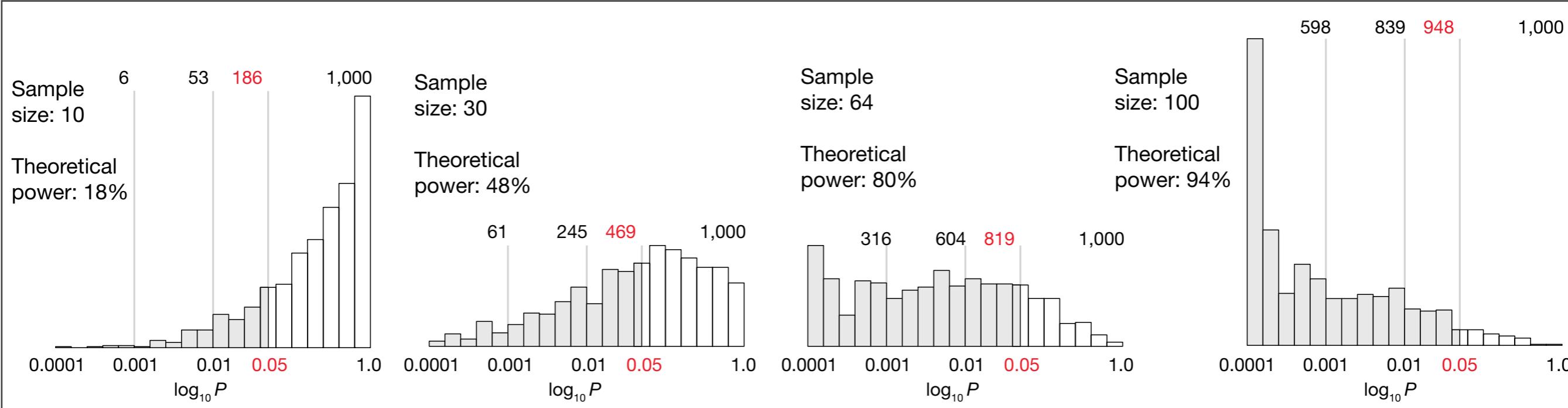
Simulated example

Halsey, Curran-Everett, Volwer and Drummond, *Nature Methods*, 2015

WITH SMALL SAMPLE SIZE, P-VALUES ARE UNSTABLE



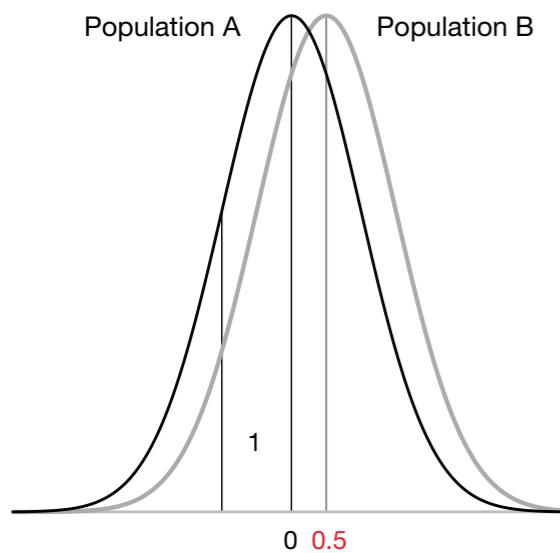
- Repeatedly sampling data leads to different results
- The problem worsens when testing many proteins
- Partial solutions:
 - Larger sample size
 - Adjustment for multiple testing



Simulated example

Halsey, Curran-Everett, Volwer and Drummond, *Nature Methods*, 2015

WITH SMALL SAMPLE SIZE, CONCLUSIONS ARE BIASED



Simulated example

Halsey, Curran-
Everett, Volwer
and Drummond,
Nature Methods,
2015

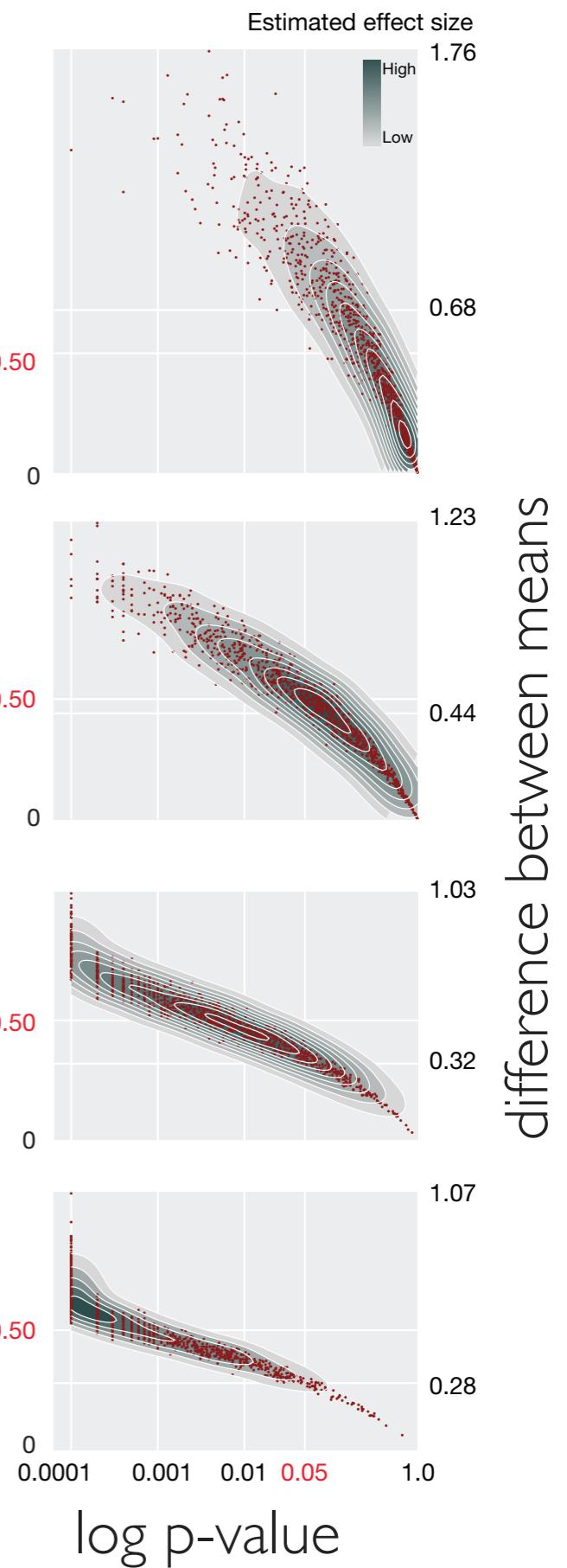
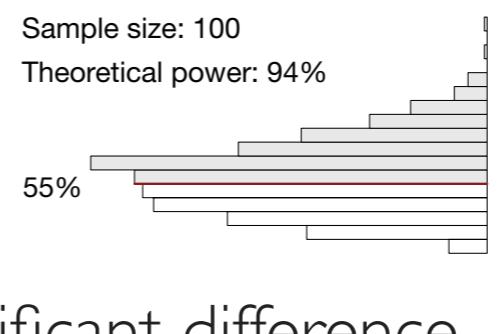
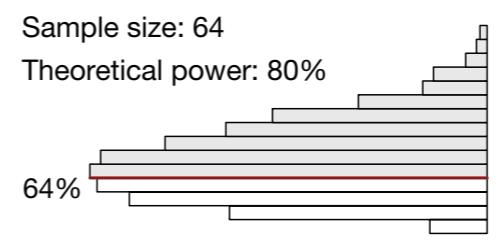
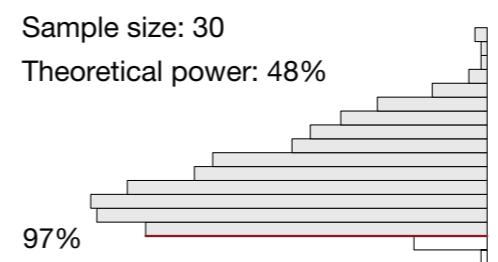
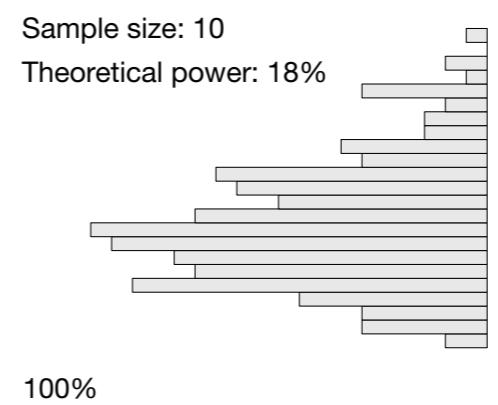
10 replicates

30 replicates

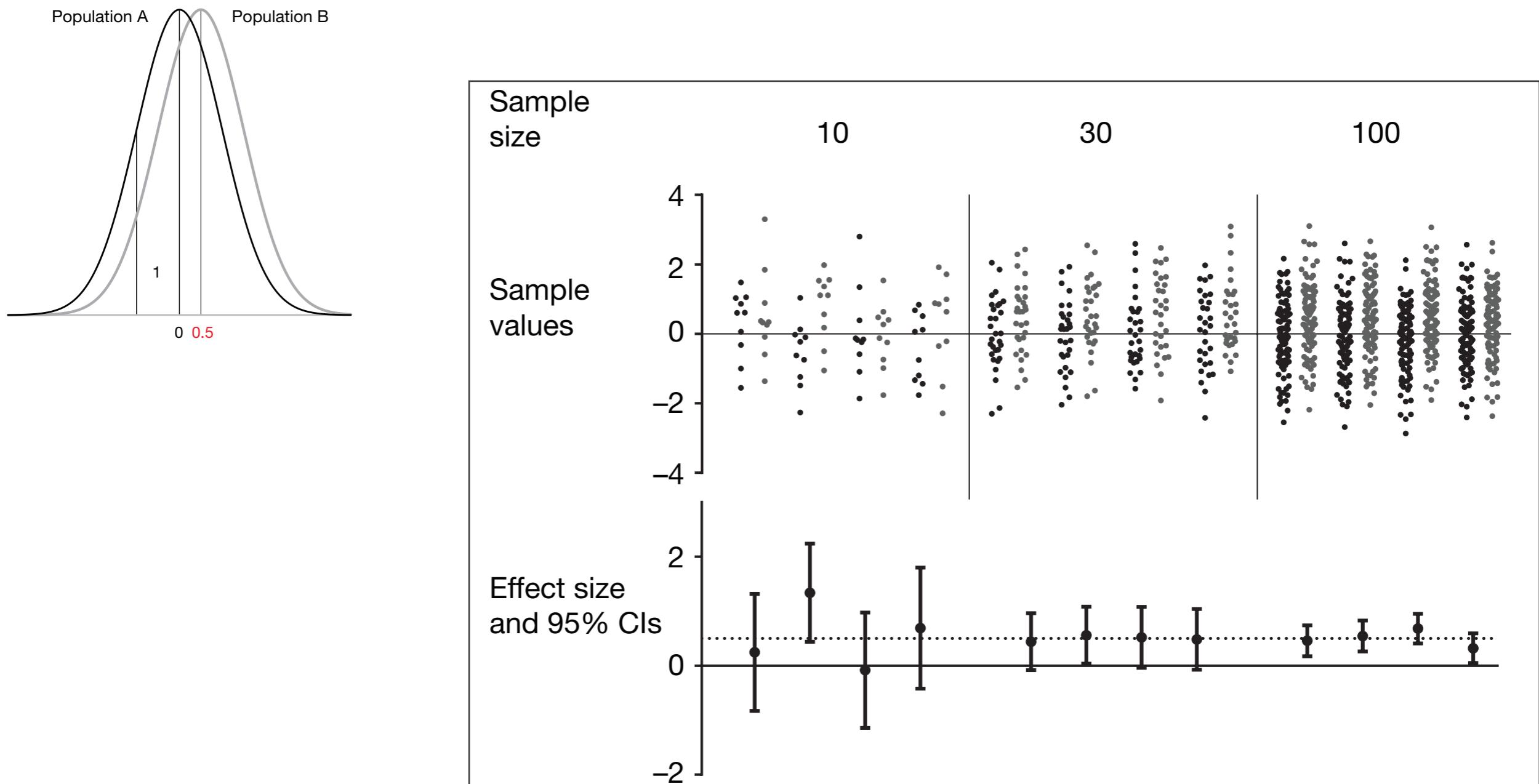
64 replicates

100 replicates

significant difference
between means



CONFIDENCE INTERVALS PROVIDE COMPLEMENTARY INSIGHT



Simulated example

Halsey, Curran-Everett, Volwer and Drummond, *Nature Methods*, 2015

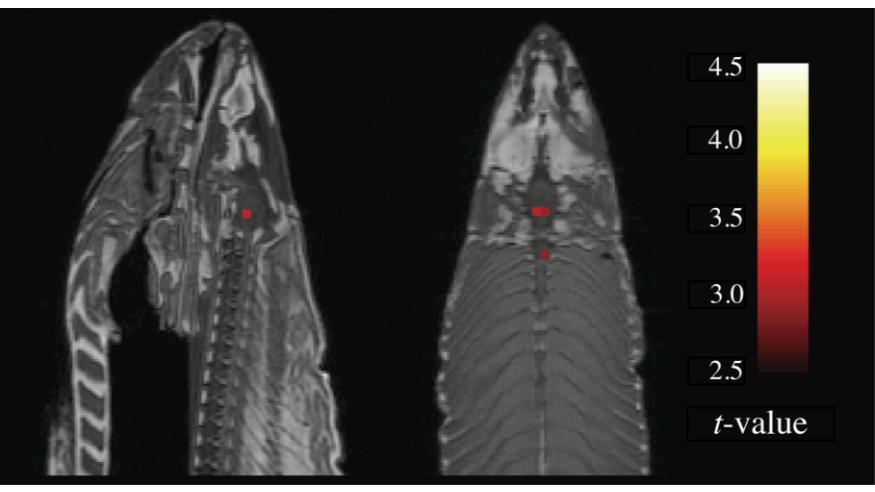
PITFALL: MULTIPLE TESTING

- An fMRI on dead fish
- Found many active brain regions
 - All background noise and random variation

 **Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction**
Craig M. Bennett¹, Abigail A. Baird², Michael B. Miller¹, and George L. Wolford³
¹ Psychology Department, University of California Santa Barbara, Santa Barbara, CA; ² Department of Psychology, Vassar College, Poughkeepsie, NY;
³ Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION
With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we carried out a real experiment that demonstrates the danger of not correcting for chance properly.

METHODS
Subject. One mature Atlantic Salmon (*Salmo salar*) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs. and was not alive at

GLM RESULTS


Source: a blog by Jeff Leek, Biostatistics, John Hopkins University

<http://simplystatistics.org/2016/02/01/a-menagerie-of-messed-up-data-analyses-and-how-to-avoid-them/>

MULTIPLE TESTING

Control False Positive Rate for two proteins

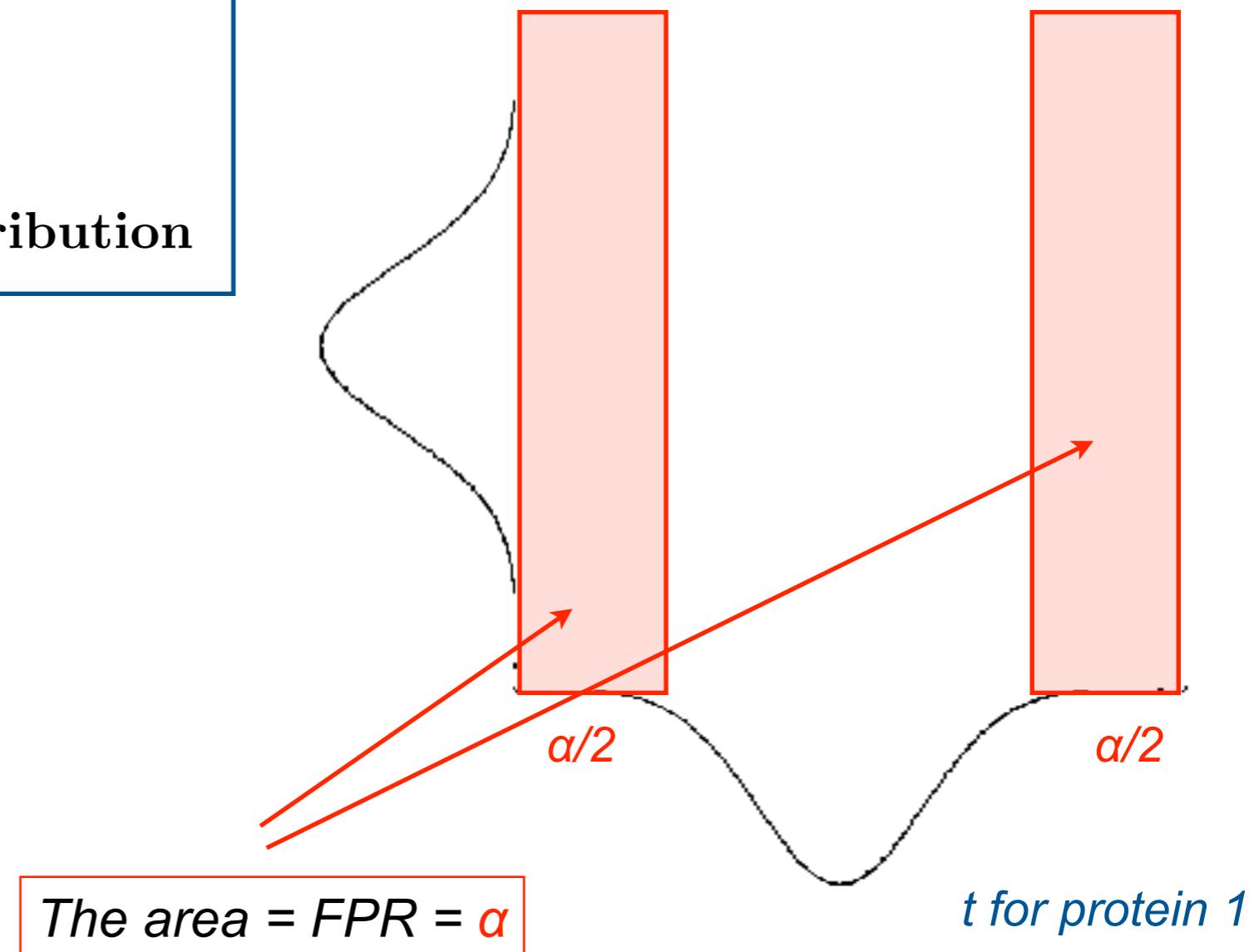
For each protein:

H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$

H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

$$\text{observed } t = \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

no difference \sim Student distribution



MULTIPLE TESTING

Control False Positive Rate for two proteins

For each protein:

H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$

H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

$$\text{observed } t = \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

no difference \sim Student distribution

t for protein 2

$\alpha/2$

$\alpha/2$

The area = FPR = α

MULTIPLE TESTING

Control False Positive Rate for two proteins

For each protein:

H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$

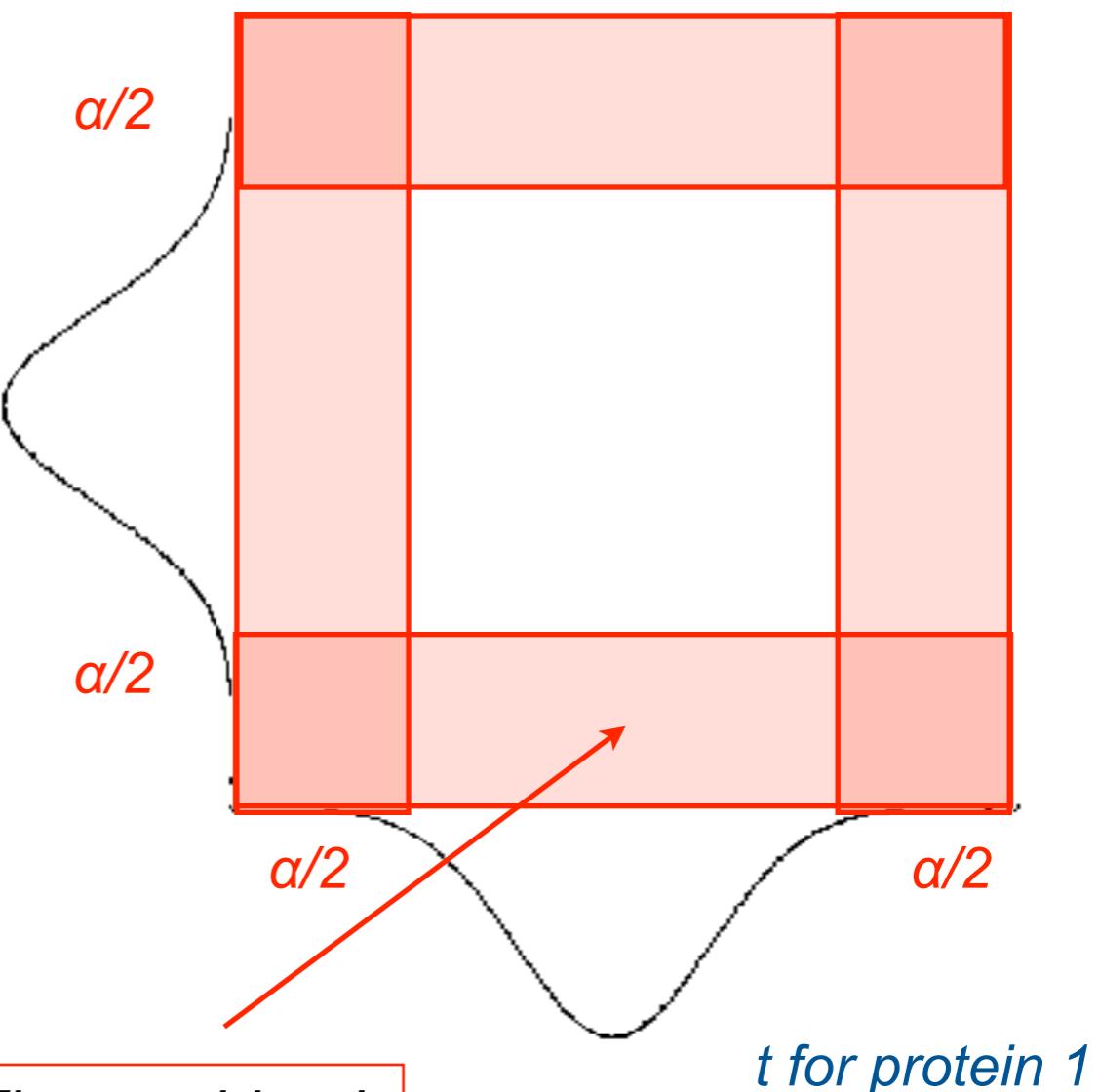
H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

$$\text{observed } t = \frac{\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

no difference \sim Student distribution

- $P(\text{at least one incorrect decision}) > \alpha$
- The univariate FPR does not hold for the list
- Need to define a *multivariate* error rate

t for protein 2



The combined area is > alpha !

MULTIPLE TESTING

Control False Positive Rate for two proteins

For each protein:

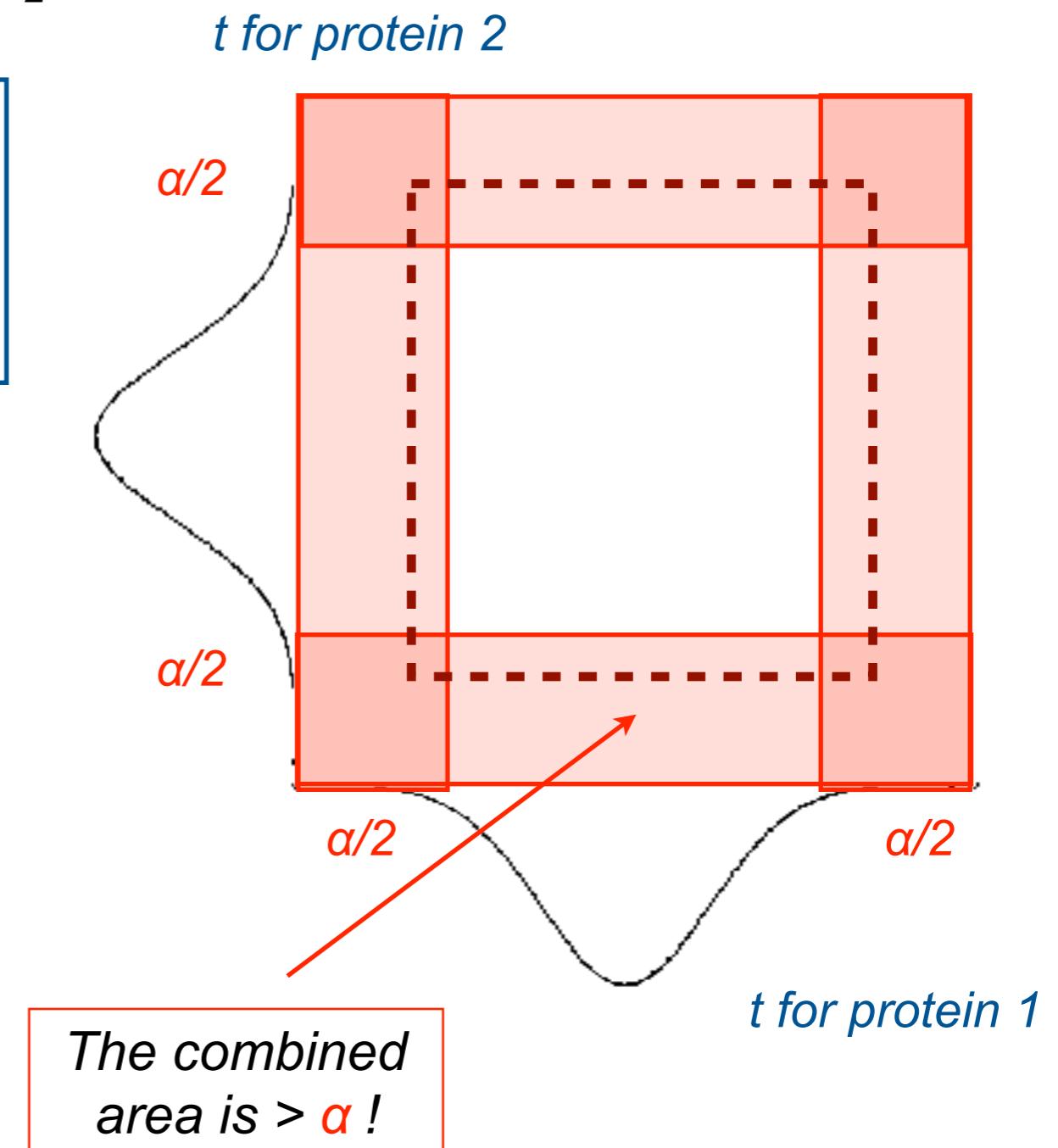
H_0 : 'status quo', no change in abundance, $\mu_1 - \mu_2 = 0$

H_a : change in abundance, $\mu_1 - \mu_2 \neq 0$

$$\text{observed } t = \frac{\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

no difference \sim Student distribution

- $P(\text{at least one incorrect decision}) > \alpha$
- The univariate FPR does not hold for the list
- Need to define a *multivariate* error rate



TESTING M PROTEINS

Change criteria from False Positive Rate to False Discovery Rate

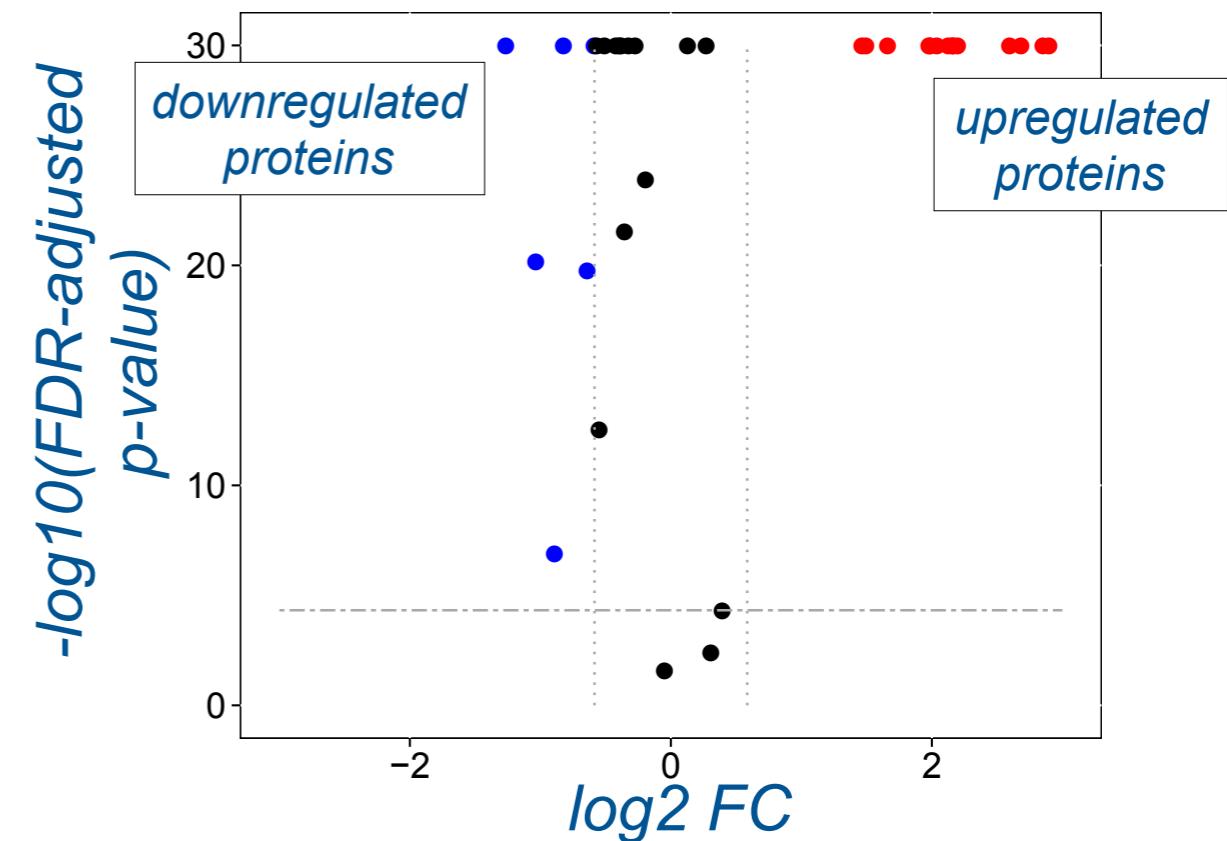
	# of proteins with no detected difference	# of proteins with detected difference	Total
# true non-diff. proteins	U	V	m_0
# true diff. proteins	T	S	$m_1 = m - m_0$
Total	$m - R$	R	m

- False discovery rate (FDR)

- ◆ An infinite number of measurements on same proteins
- ◆ FDR: the *average* proportion of false discoveries

$$FDR = E \left[\frac{V}{\max(R, 1)} \right]$$

Bonferroni approach
controls family-wise error
rate = $P(V > 0)$



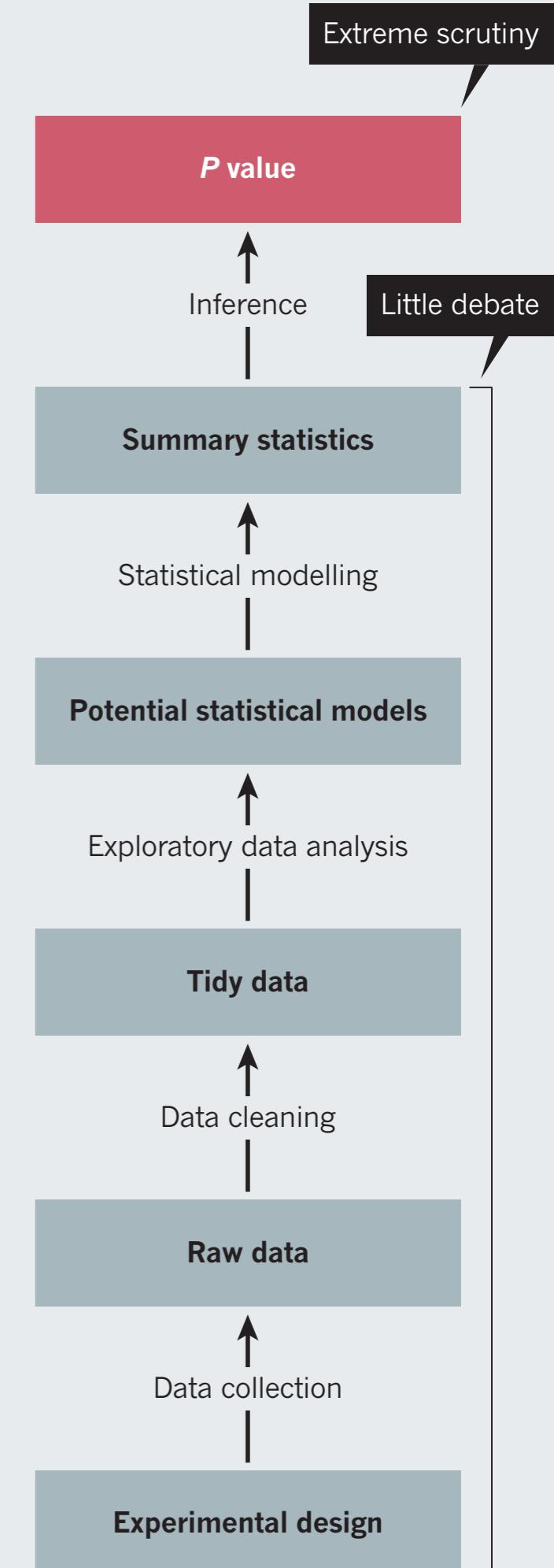
NATURE | COMMENT

Statistics: *P* values are just the tip of the iceberg

Jeffrey T. Leek & Roger D. Peng

28 April 2015

Statistical
considerations are key
at every step



PITFALL: OUTCOME SWITCHING

- Anti-depressant Paxil was studied for several main outcomes
 - None showed an effect
 - Some secondary outcomes did
- Switched the outcome of the trial and used to market the drug

Vox SCIENCE & HEALTH

How researchers dupe the public with a sneaky practice called "outcome switching"

Updated by Julia Belluz on December 29, 2015, 8:10 a.m. ET
✉ julia.belluz@voxmedia.com



Source: a blog by Jeff Leek, Biostatistics, John Hopkins University

<http://simplystatistics.org/2016/02/01/a-menagerie-of-messed-up-data-analyses-and-how-to-avoid-them/>

PITFALL: NOT PRE-SPECIFIED DATA SELECTION AND ANALYSIS

- Compare 2 groups: women at peak and off peak fertility cycle
 - A series of choices of which women to include in which comparison group
 - Conclude that at peak fertility women are more likely to wear red or pink shirts

The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time*

Andrew Gelman[†] and Eric Loken[‡]

14 Nov 2013

Researcher degrees of freedom can lead to a multiple comparisons problem, even in settings where researchers perform only a single analysis on their data. The problem is there can be a large number of *potential* comparisons when the details of data analysis are highly contingent on data, without the researcher having to perform any conscious procedure of fishing or examining multiple p-values. We discuss in the context of several examples of published papers where data-analysis decisions were theoretically-motivated based on previous literature, but where the details of data selection and analysis were not pre-specified and, as a result, were contingent on data.

Source: a blog by Jeff Leek, Biostatistics, John Hopkins University

<http://simplystatistics.org/2016/02/01/a-menagerie-of-messed-up-data-analyses-and-how-to-avoid-them/>

RESEARCHER DEGREE OF FREEDOM



The image shows the cover of the journal "Psychological Science". The title "Psychological SCIENCE" is prominently displayed in large white letters on a dark red background. Below the title, it says "A Journal of the Association for Psychological Science". A navigation bar at the bottom includes links for "Home", "OnlineFirst", "All Issues", "Subscribe", "RSS", and "Email Alerts". A search bar with the placeholder "Search this journal" is also present.

False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons¹,

Leif D. Nelson² and

Uri Simonsohn¹

Author Affiliations

Joseph P. Simmons, The Wharton School, University of Pennsylvania, 551 Jon M. Huntsman Hall, 3730 Walnut St., Philadelphia, PA 19104 E-mail:
jsimmo@wharton.upenn.edu

Leif D. Nelson, Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900 E-mail: leif_nelson@haas.berkeley.edu

Uri Simonsohn, The Wharton School, University of Pennsylvania, 548 Jon M. Huntsman Hall, 3730 Walnut St., Philadelphia, PA 19104 E-mail: uws@wharton.upenn.edu

[« Previous](#) | [Next Article »](#)
[Table of Contents](#)

This Article

Published online before print October 17, 2011, doi:

[10.1177/0956797611417632](https://doi.org/10.1177/0956797611417632)

Psychological Science November 2011
vol. 22 no. 11 1359-1366

» Abstract **Free**

Full Text **Free**

Full Text (PDF)  **Free**

All Versions of this Article:

» Version of Record - Nov 7, 2011
[0956797611417632v1](https://doi.org/10.1177/0956797611417632v1) - Oct 17, 2011

What's this?

Services

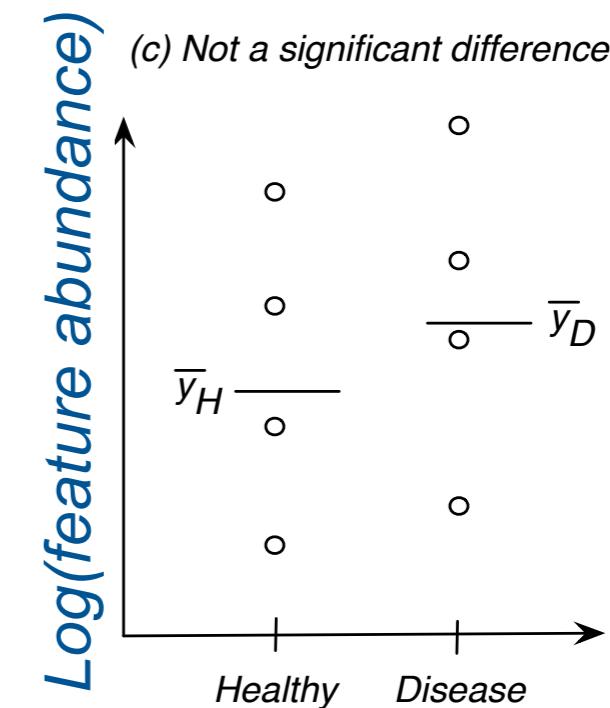
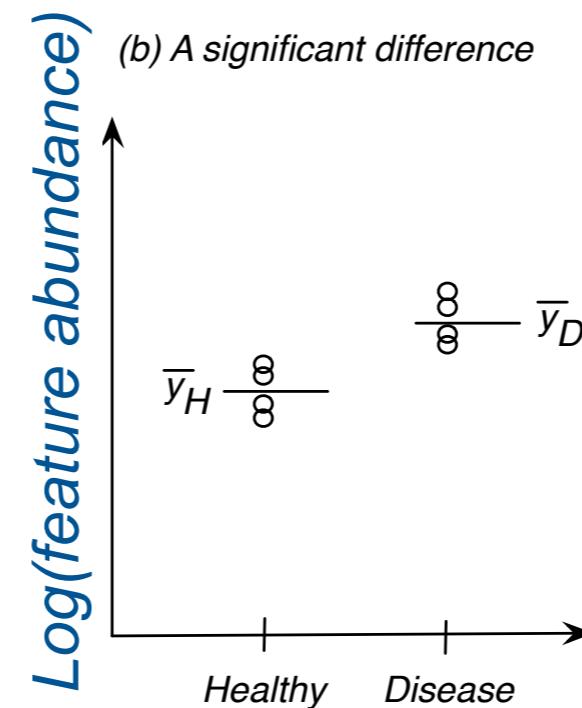
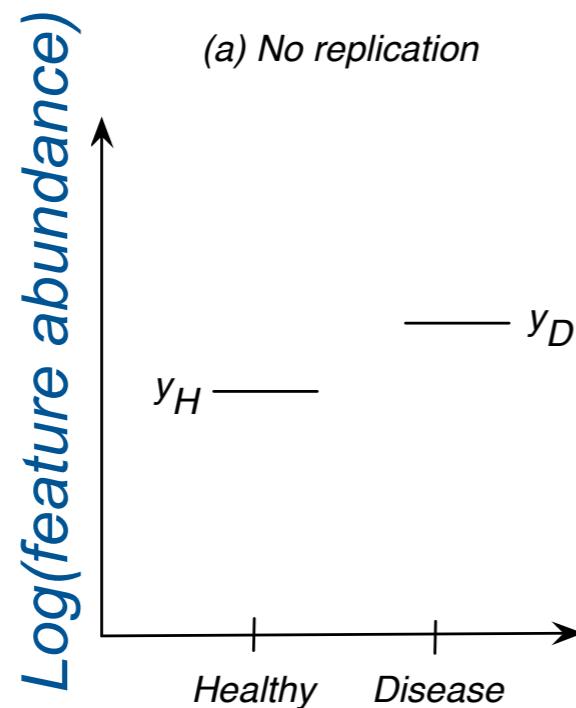
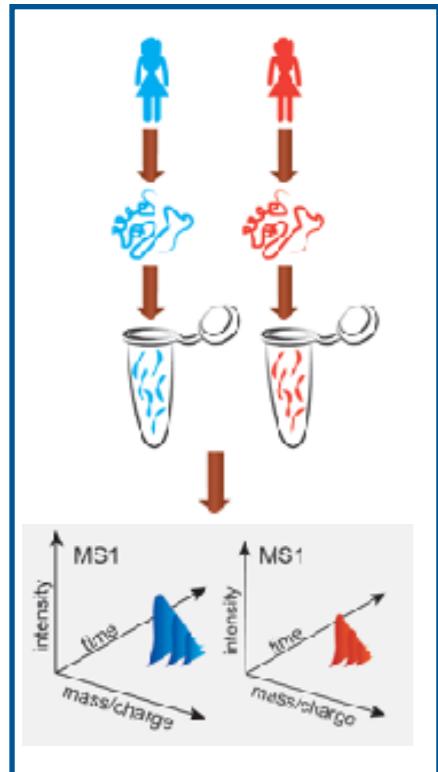
- » Email this article to a colleague
- » Alert me when this article is cited

OUTLINE

- Basic statistical inference
 - T-test and p-values
- P-values: a word of caution
 - Instability, multiplicity, alternative approaches
- So how many replicates do I need?
 - Design of complex experiments

PRINCIPLE I: REPLICATION

(1) carries out the inference and (2) minimizes inefficiencies



Two levels of randomness imply two types of replication:

- ◆ *Biological replicates*: selecting multiple subjects from the population
- ◆ *Technical replicates*: multiple runs per subject

BIOLOGICAL REPLICATION IS MOST IMPORTANT

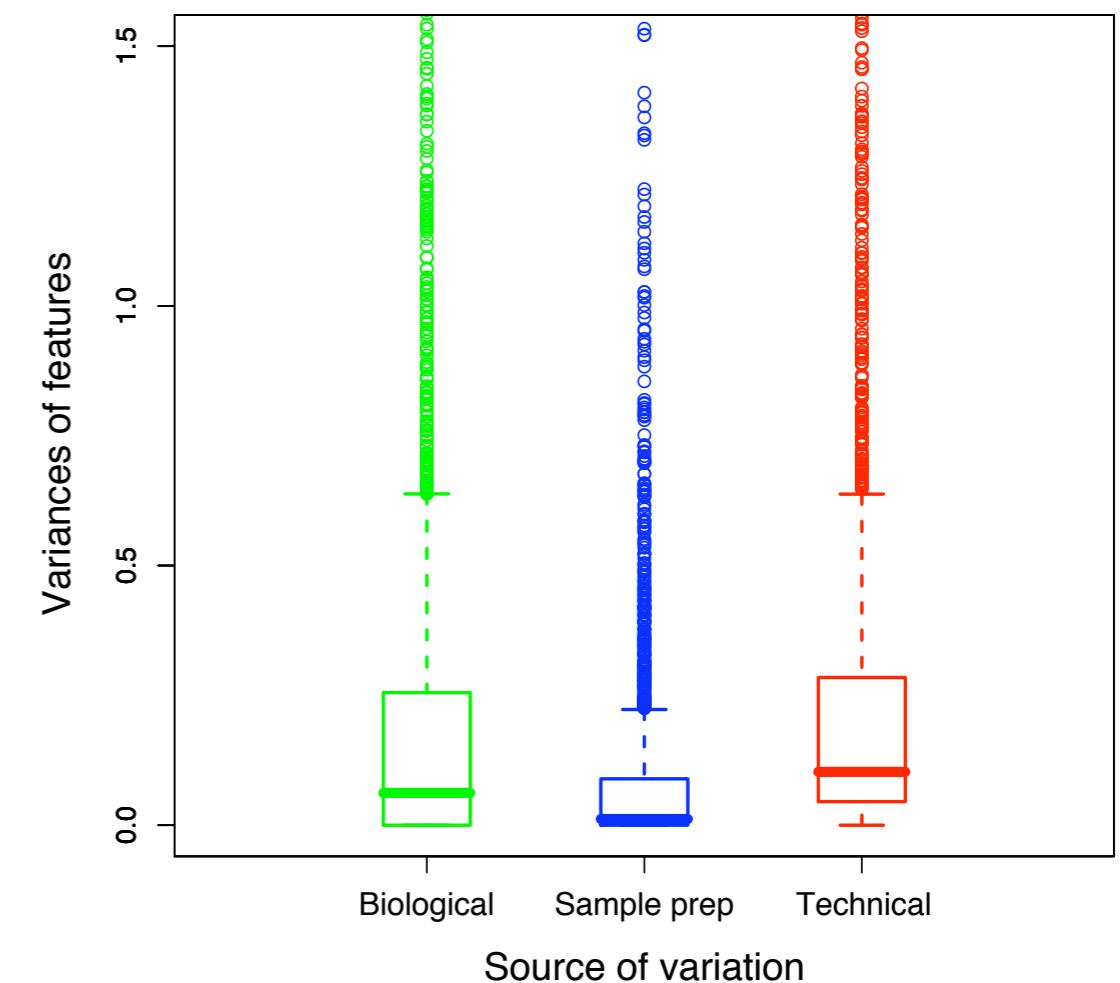
Observed feature intensity	=	Systematic mean signal of disease group	+	Random deviation due to individual	+	Random deviation due to sample preparation	+	Random deviation due to measurement error
y_{ijkl}	=	Group mean _i	+	Indiv(Group) _{j(i)} $\sim N(0, \sigma_{\text{Indiv}}^2)$	+	Prep(Indiv) _{k(ij)} $\sim N(0, \sigma_{\text{Prep}}^2)$	+	Error _{l(ijk)} $\sim N(0, \sigma_{\text{Error}}^2)$

$$\text{Var}(\bar{y}_H - \bar{y}_D) = 2 \left(\frac{\sigma_{\text{Indiv}}^2}{I} + \frac{\sigma_{\text{Prep}}^2}{IJ} + \frac{\sigma_{\text{Error}}^2}{IJK} \right)$$

I: # individuals per disease group
J: # sample preps
K: # replicate runs

A pilot experiment

- 2 healthy individuals, 2 with diabetes
- multiple sample preparations
- multiple LC-MS replicates



BIOLOGICAL REPLICATION IS MOST IMPORTANT

Observed feature intensity	=	Systematic mean signal of disease group	+	Random deviation due to individual	+	Random deviation due to sample preparation	+	Random deviation due to measurement error
y_{ijkl}	=	Group mean _i	+	Indiv(Group) _{j(i)} $\sim N(0, \sigma_{\text{Indiv}}^2)$	+	Prep(Indiv) _{k(ij)} $\sim N(0, \sigma_{\text{Prep}}^2)$	+	Error _{l(ijk)} $\sim N(0, \sigma_{\text{Error}}^2)$

$$\text{Var}(\bar{y}_H - \bar{y}_D) = 2 \left(\frac{\sigma_{\text{Indiv}}^2}{I} + \frac{\sigma_{\text{Prep}}^2}{IJ} + \frac{\sigma_{\text{Error}}^2}{IJK} \right)$$

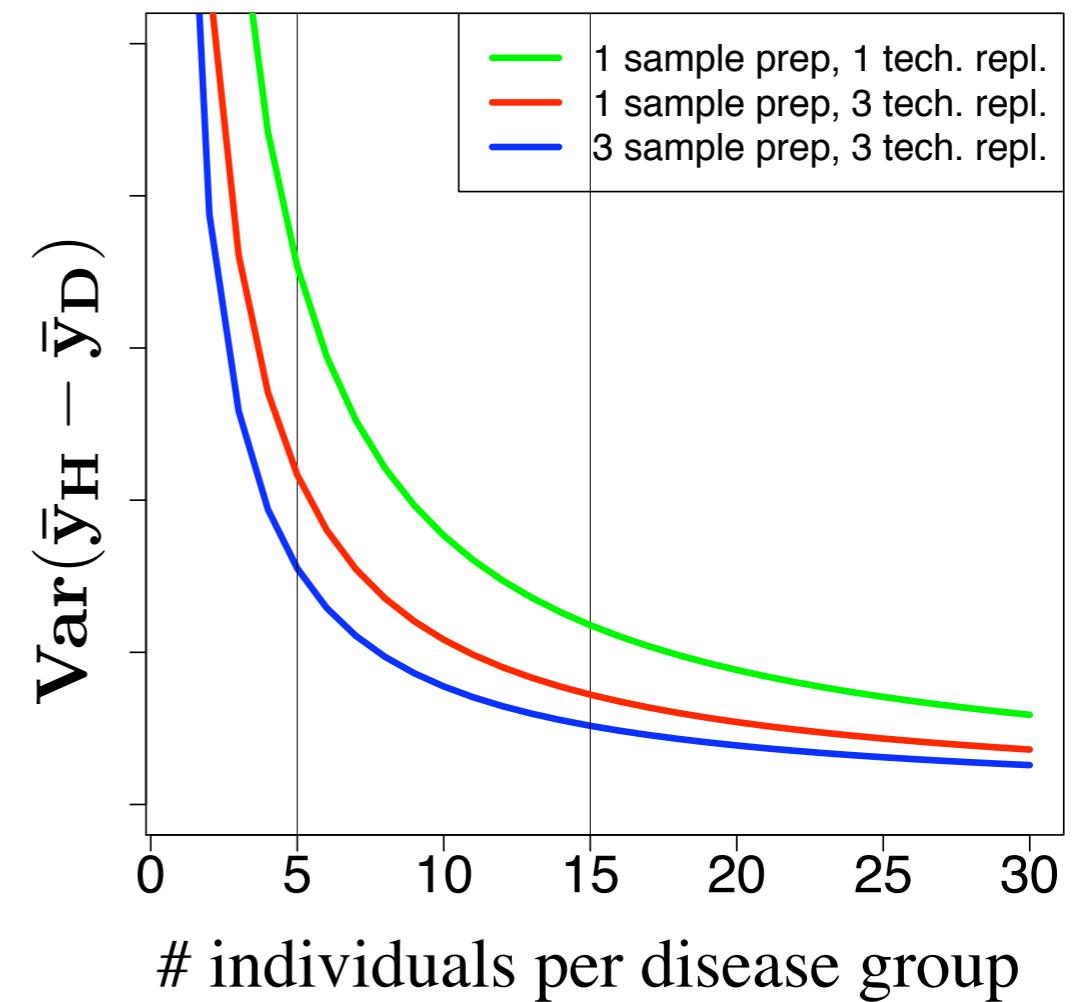
I: # individuals per disease group

J: # sample preps

K: # replicate runs

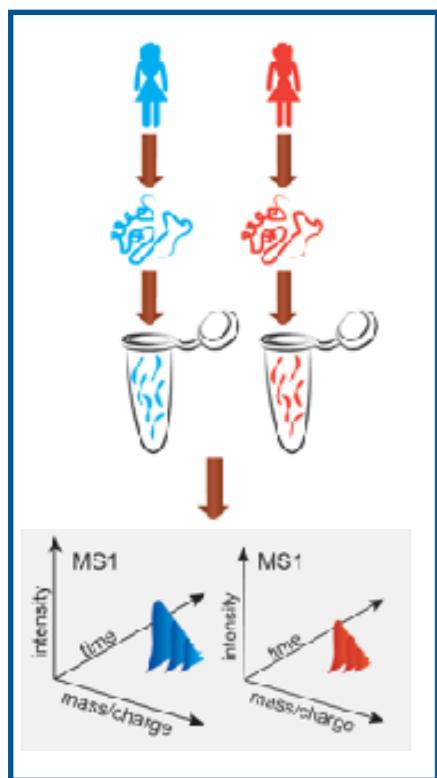
Conclusion 1:

Maximize the number
of biological replicates

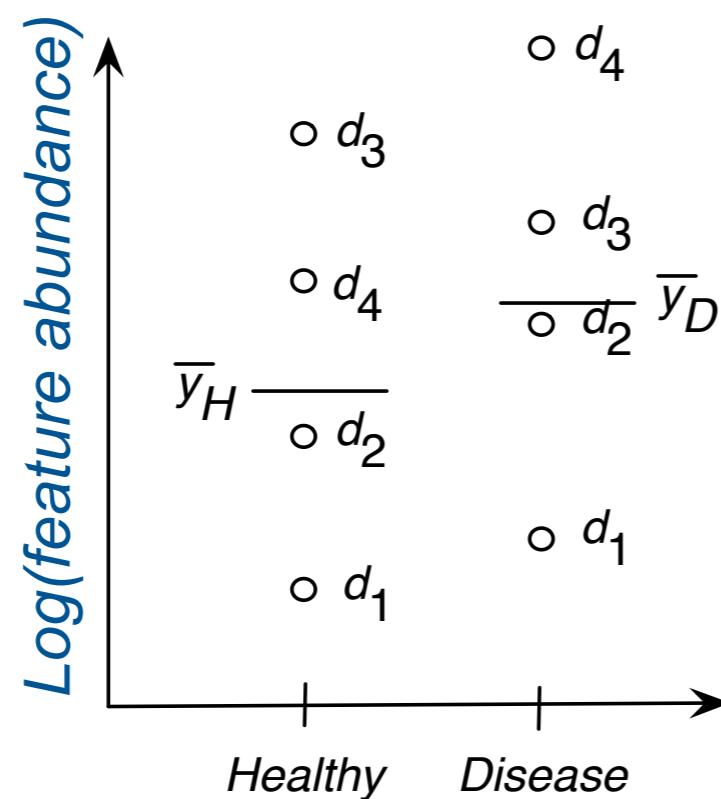


PRINCIPLE 3: BLOCKING

Helps reduce both bias and inefficiency

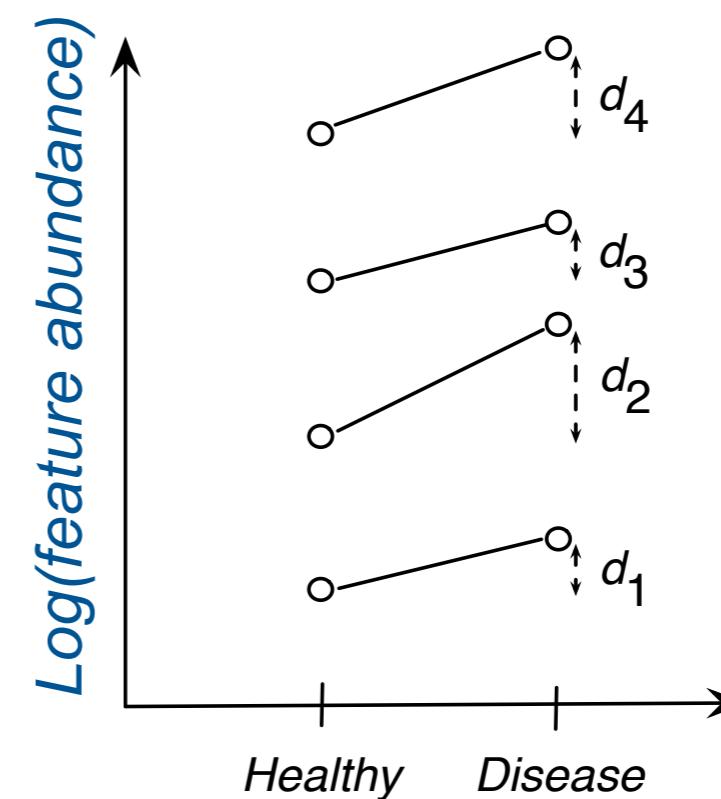


(b) Complete randomization



Complete randomization
= inflated variance

(c) Day = block



Block-randomization
= restriction on randomization
= systematic allocation

Two levels of randomness imply two types of blocks:

- ◆ *Biological replicates*: subjects having similar characteristics (e.g. age)
- ◆ *Technical replicates*: samples processed together (e.g. in a same day)

BLOCKING IS HELPFUL

Especially when between-block variance is large

Observed feature intensity	Systematic mean signal of disease group	Random deviation due to block (e.g. plate or day)	Random deviation due to individual	Random deviation due to measurement error
y_{ijkl}	Group mean _i	$\sim N(0, \sigma_{\text{Block}}^2)$	$\sim N(0, \sigma_{\text{Indiv}}^2)$	$\sim N(0, \sigma_{\text{Error}}^2)$

A completely randomized design

I: # individuals per disease group

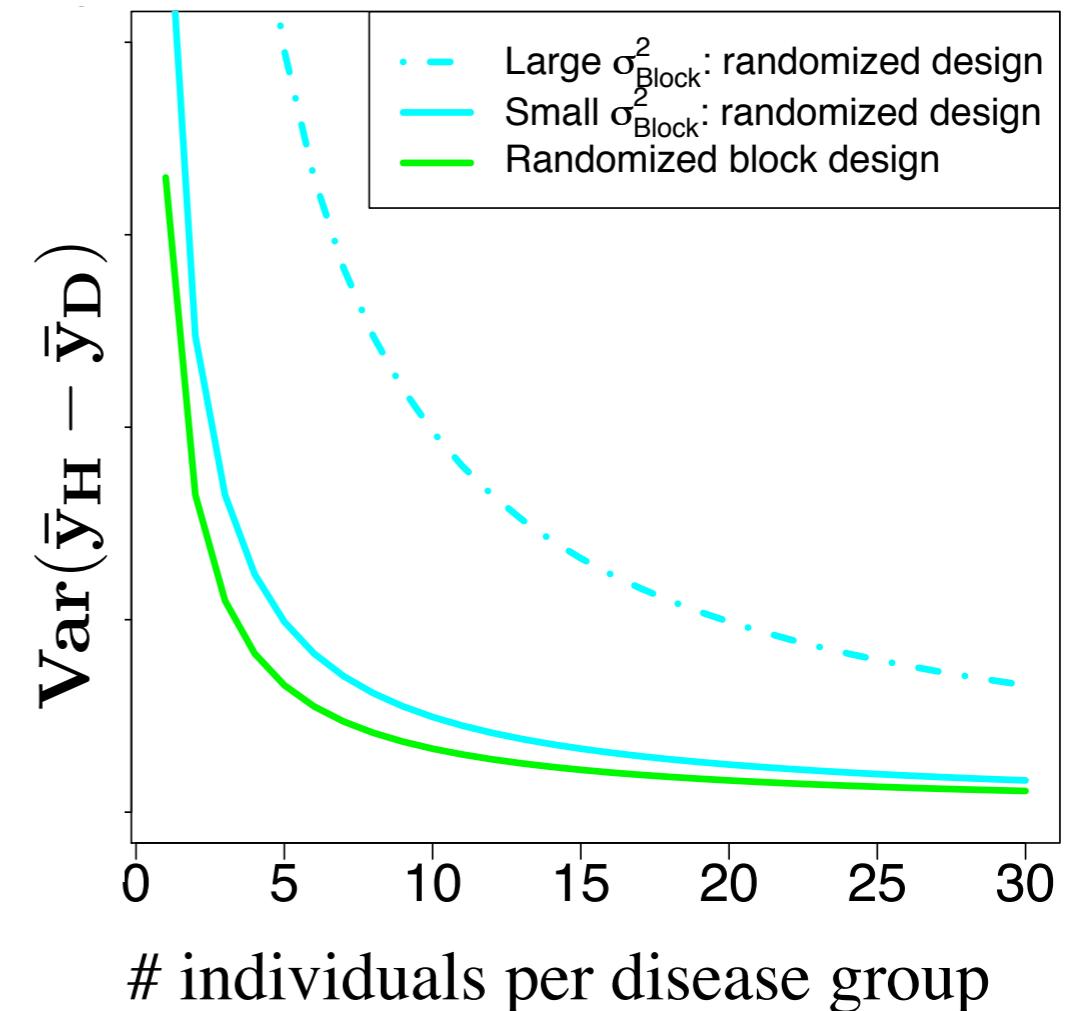
$$\text{Var}(\bar{y}_H - \bar{y}_D) = 2 \left(\frac{\sigma_{\text{Block}}^2 + \sigma_{\text{Indiv}}^2 + \sigma_{\text{Error}}^2}{I} \right)$$

A block-randomized design

$$\text{Var}(\bar{y}_H - \bar{y}_D) = 2 \left(\frac{\sigma_{\text{Indiv}}^2 + \sigma_{\text{Error}}^2}{I} \right)$$

Conclusion 2: Block-randomize

- if can not control a large source of variation
- if moderate sample size



SO HOW MANY REPLICATES DO I NEED?

If we only have one feature:

Fix: α - probability of a false positive discovery

β - probability of a true positive discovery

Δ - anticipated fold change

σ_{Indiv}^2 and σ_{Error}^2 - anticipated variability

Write:

$$\text{Var}(\bar{y}_H - \bar{y}_D) \leq \left(\frac{\Delta}{z_{1-\beta} + z_{1-\alpha/2}} \right)^2$$

where $z_{1-\beta}$ and $z_{1-\alpha/2}$ are quantiles of the Normal distribution

A completely randomized design

I: # individuals per disease group

$$\text{Var}(\bar{y}_H - \bar{y}_D) = 2 \left(\frac{\sigma_{\text{Block}}^2 + \sigma_{\text{Indiv}}^2 + \sigma_{\text{Error}}^2}{I} \right)$$

A block-randomized design

$$\text{Var}(\bar{y}_H - \bar{y}_D) = 2 \left(\frac{\sigma_{\text{Indiv}}^2 + \sigma_{\text{Error}}^2}{I} \right)$$



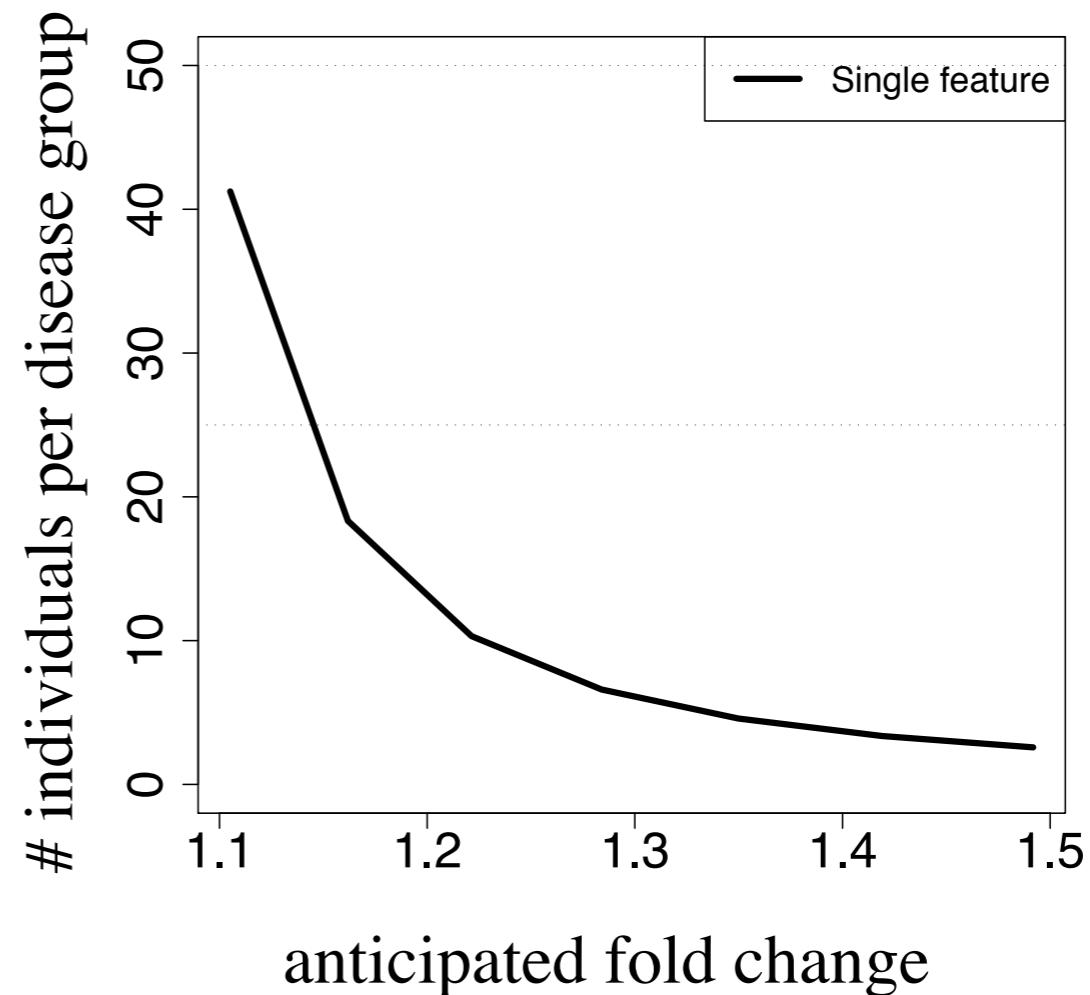
solve for the number of individuals I

SO HOW MANY REPLICATES DO I NEED?

Example: pilot study with diabetes patients.

A block-randomized design

If we only had one feature:



Conclusion 3:

The smaller the anticipated difference, the larger the sample size

DIFFICULTY: MANY FEATURES ARE OF INTEREST

Would like to control the False Discovery Rate:

	# of features with no detected difference	# of features with detected difference	Total
# true non-diff. features	U	V	m_0
# true diff. features	T	S	$m_1 = m - m_0$
Total	$m - R$	R	m

$q = E \left[\frac{V}{\max(R, 1)} \right]$ = the “average” proportion of false positives

This changes the sample size calculation:

Fix: q - the False Discovery Rate

m_0/m_1 - anticipated ratio of unchanging features

This defines

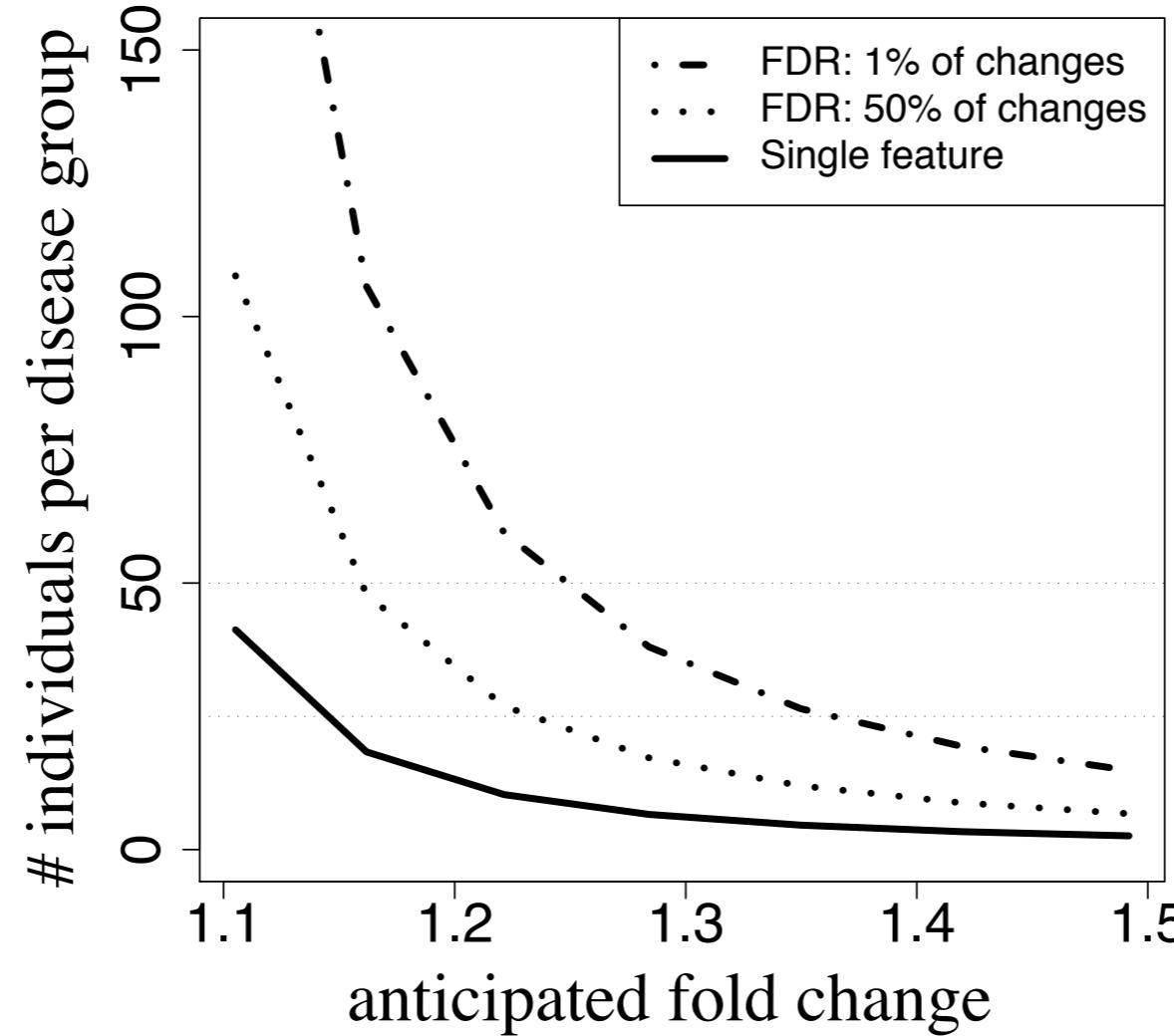
$$\alpha_{ave} \leq (1 - \beta)_{ave} \cdot q \frac{1}{1 + (1 - q) \cdot m_0/m_1},$$

- average probability of a false positive discovery

SO HOW MANY REPLICATES DO I NEED?

Example: pilot study with diabetes patients.

A block-randomized design



Conclusion 4:

The fewer changes we expect, the larger the sample size

Oberg and
Vitek, *JPR*,
2009

WHY R?

- *Researcher:*
 - Lightweight: minimal software/hardware requirements
 - Interactive: easy data exploration on a laptop
 - Useful results with minimal coding and fast
- *Developer:*
 - Large community: leverage state-of-the-art
 - Easy to customize/extend: e.g. include in existing pipelines
- *Science:*
 - Full transparency: open algorithms and code
 - Infrastructure for fully documentable workflows