Denotes
He-man

Denotes
practical

Power and p-values

Benjamin Neale
March 5th, 2020
International Twin Workshop, Boulder, CO

FiveThirtyEight

Politics | Sports | **Science & Health** | Economics | Culture

SCIENTIFIC METHOD | 10:23 AM | MAR 7, 2016

Statisticians Found One Thing They Can Agree On: It's Time To Stop Misusing P-Values

By CHRISTIE ASCHWANDEN

p-values are in the press

The New York Times

SCIENCE TIMES AT 40

Essay: The Experiments Are Fascinating. But Nobody Can Repeat Them.

Science is mired in a “replication” crisis. Fixing it will not be easy.

https://www.nytimes.com/2018/11/19/science/science-research-fraud-reproducibility.html

In February, 2014, George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

> Q: Why do so many colleges and grad schools teach $p = .05$?
>
> A: Because that's still what the scientific community and journal editors use.
>
> Q: Why do so many people still use $p = 0.05$?
>
> A: Because that's what they were taught in college or grad school.

What we’ve been teaching

What exactly is a p-value?

- A way to ask if the data are consistent with a null model

- The baseline model for comparison, usually no effect [e.g. no heritability]

What's a null model?

- Distrust of his aunt’s claims of being able to discriminate between milk in first or tea in first

Whose fault is it anyway?

Alternative hypothesis	Null hypothesis
Some effect	No effect

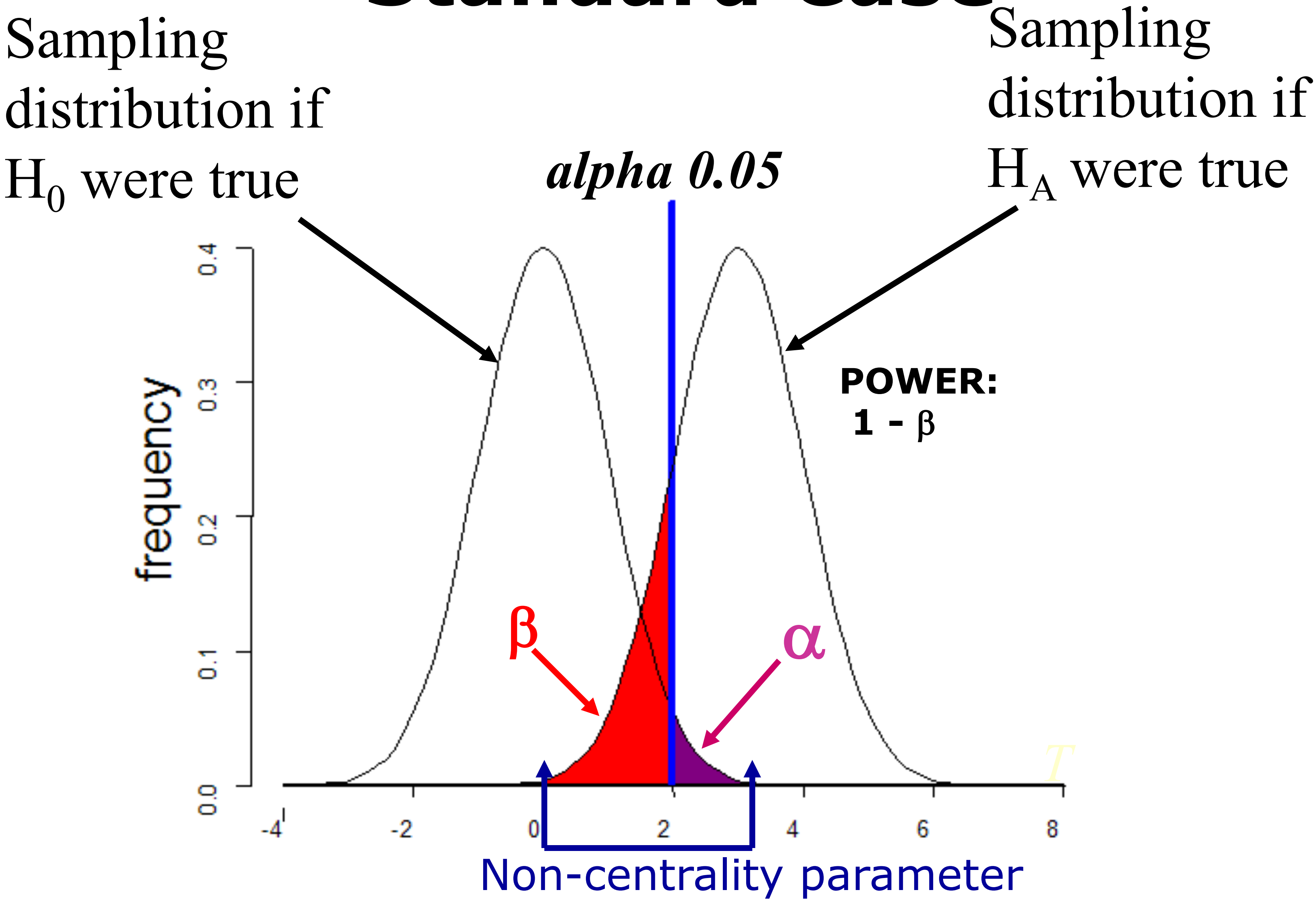
Hypothesis testing

Possible scenarios

	Statistics: Reject H_0	Statistics: Fail to reject H_0
Truth: H_0 is true	α	$1-\alpha$
Truth: H_a is true	$1-\beta$	β

α=type 1 error rate
β=type 2 error rate
$1-\beta$=statistical power

Standard Case

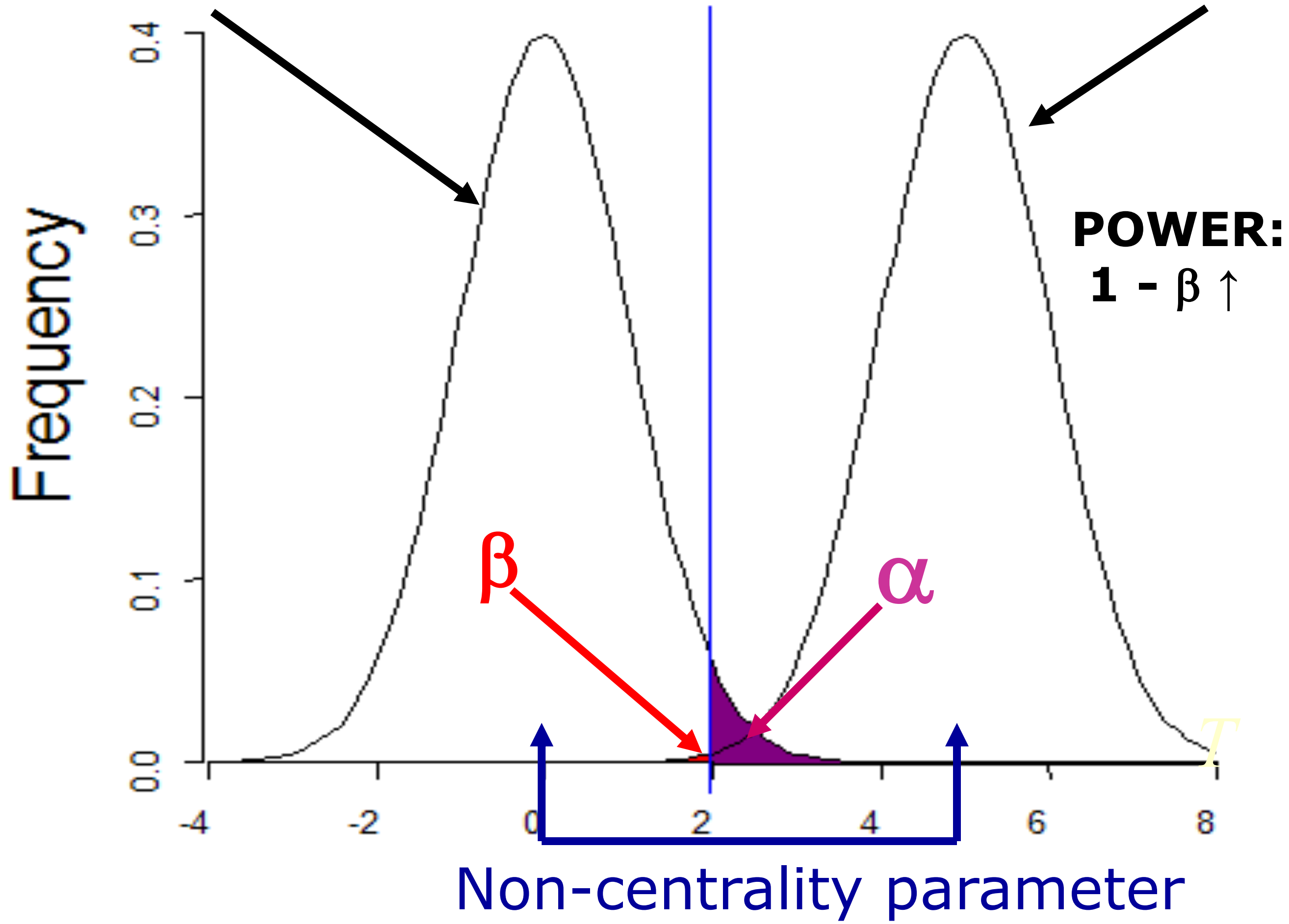


Increased effect size

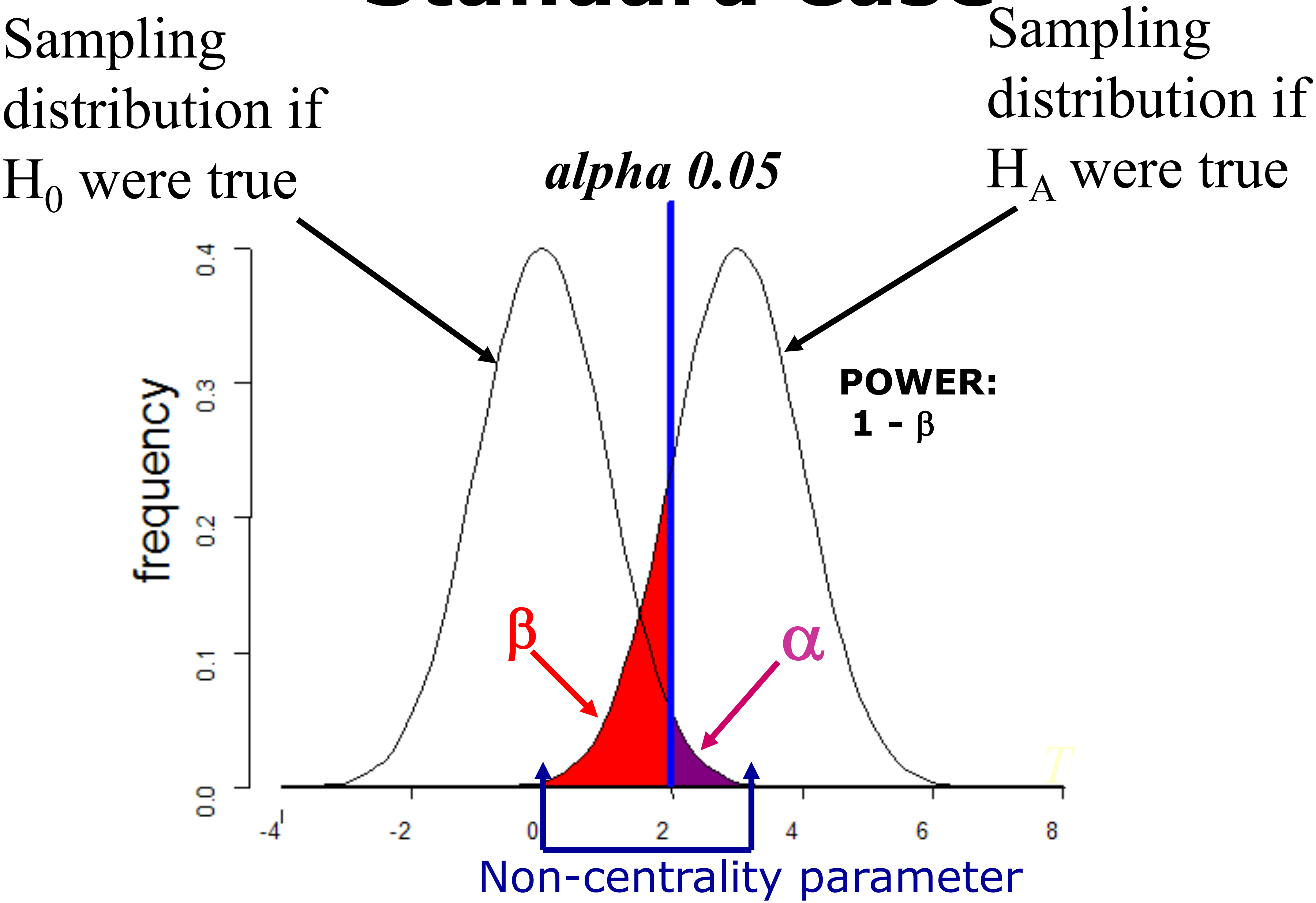
Sampling distribution if H_0 were true

alpha 0.05

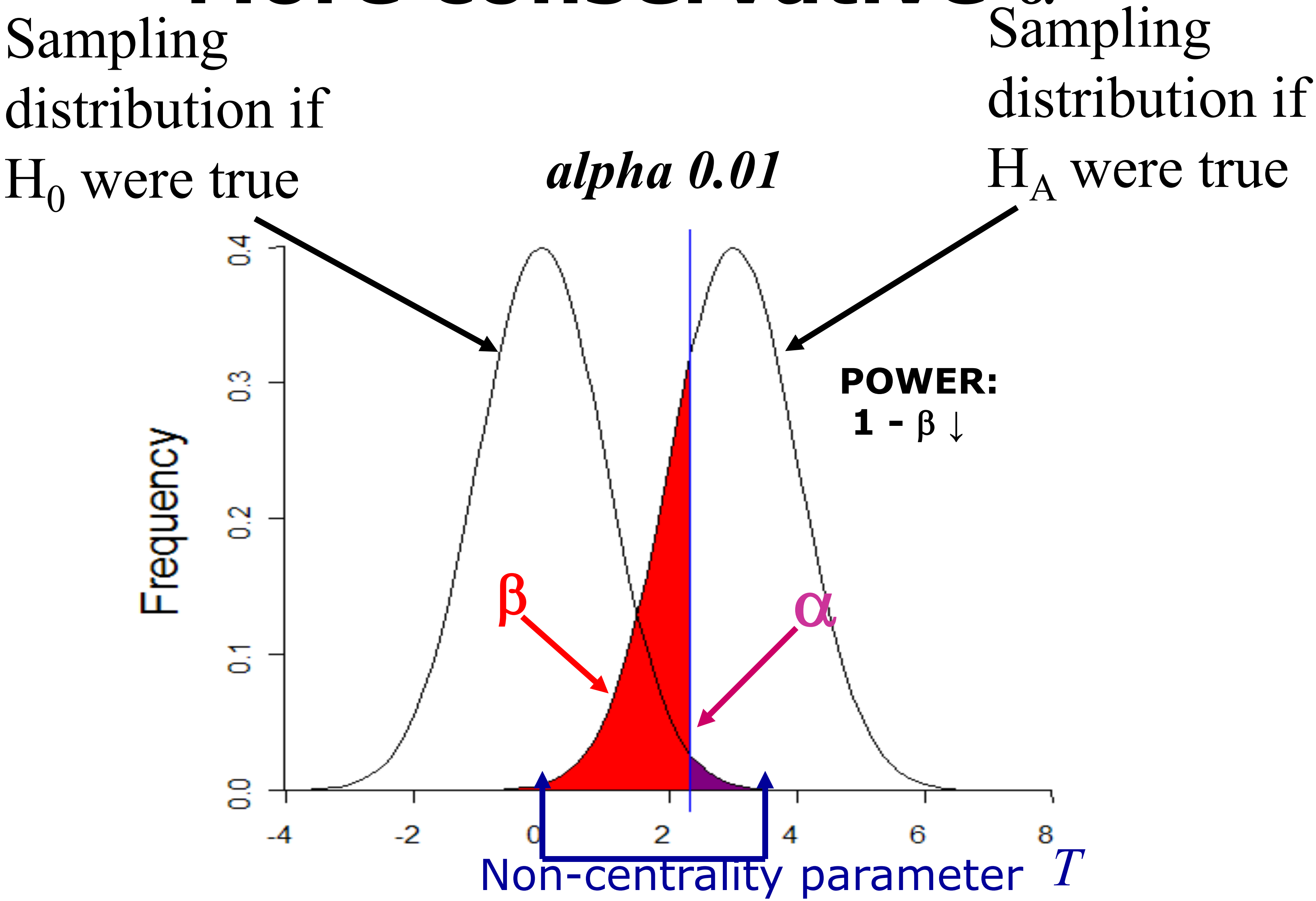
Sampling distribution if H_A were true



Standard Case



More conservative α

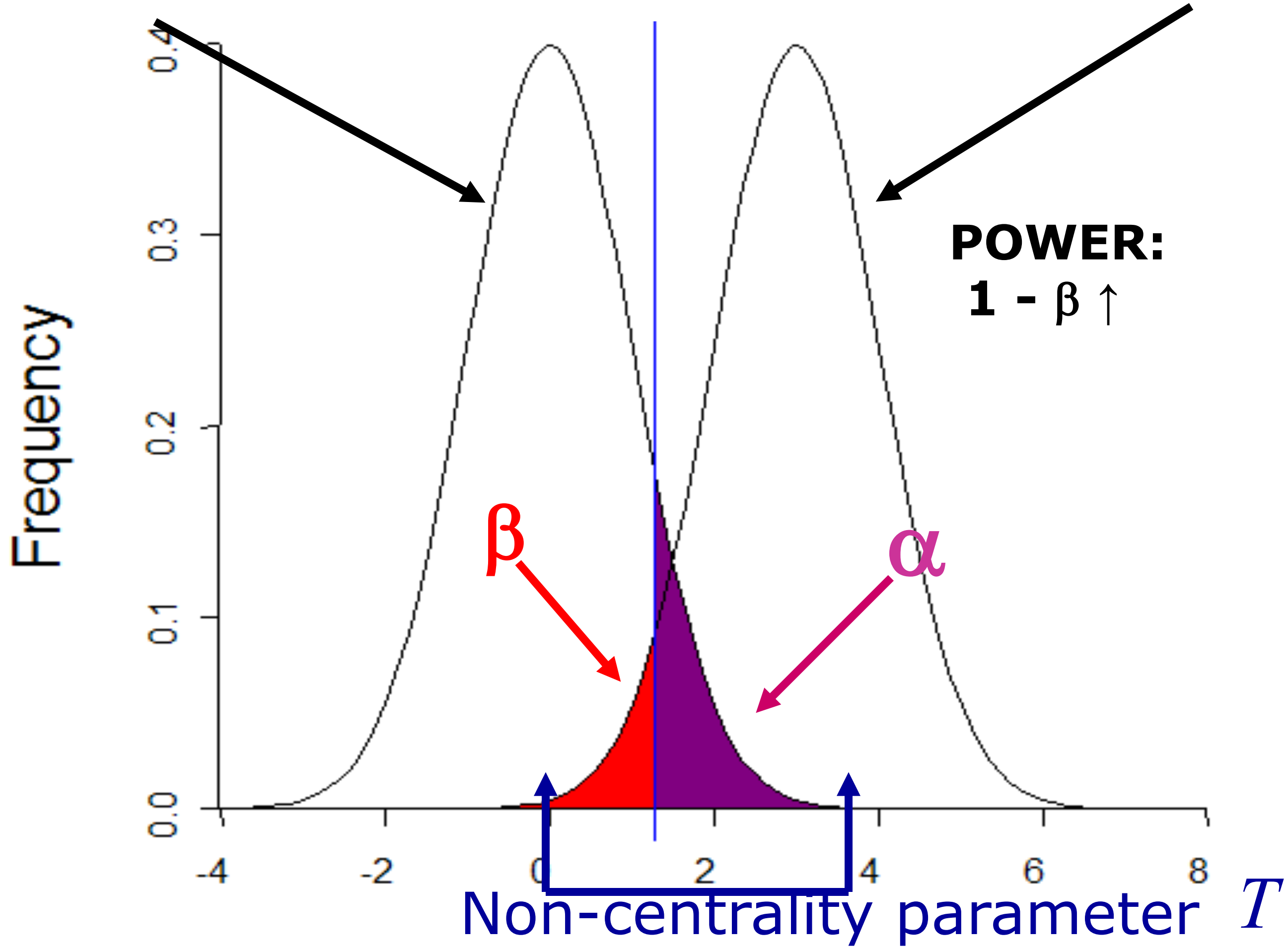


Less conservative α

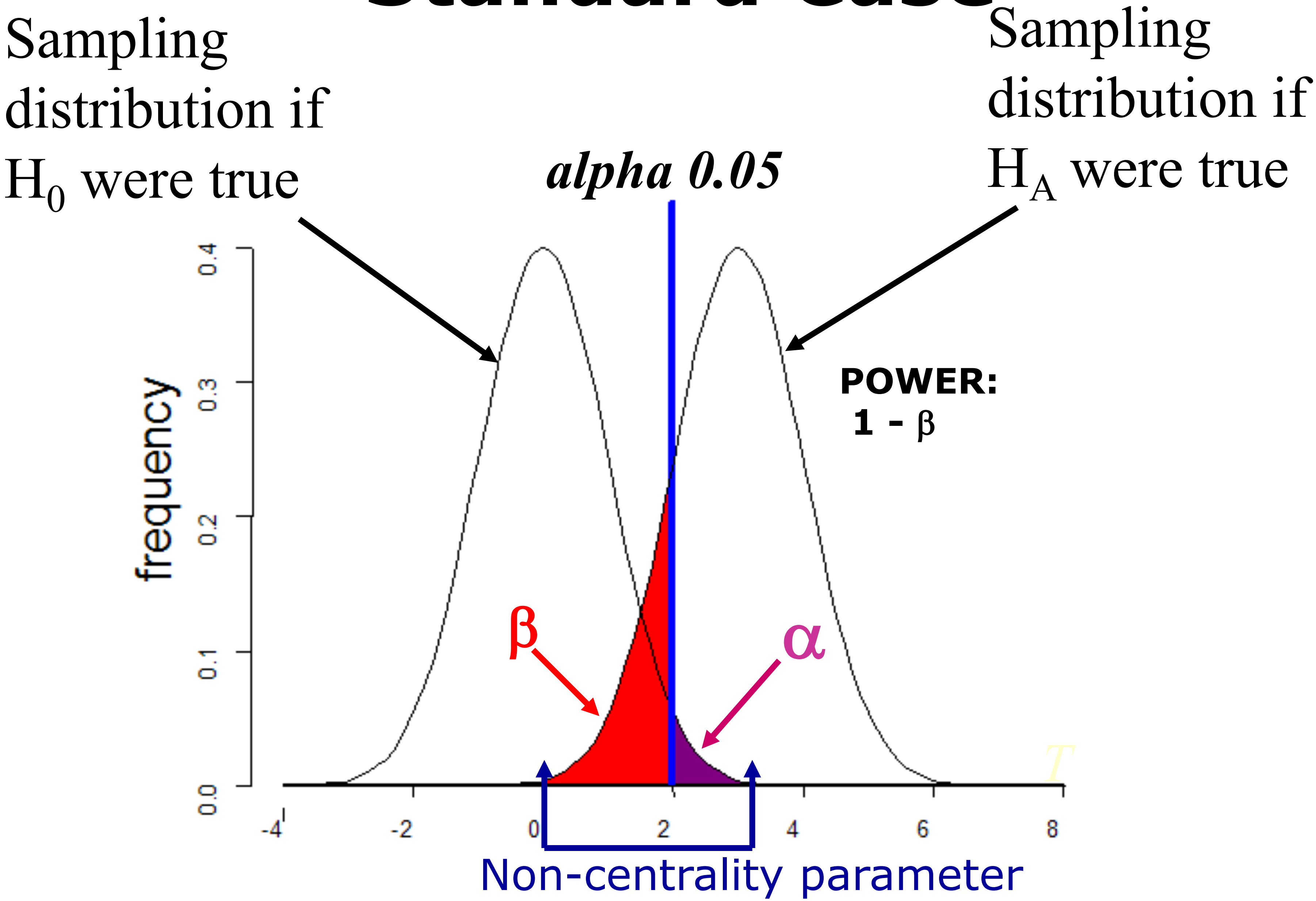
Sampling distribution if H_0 were true

alpha 0.10

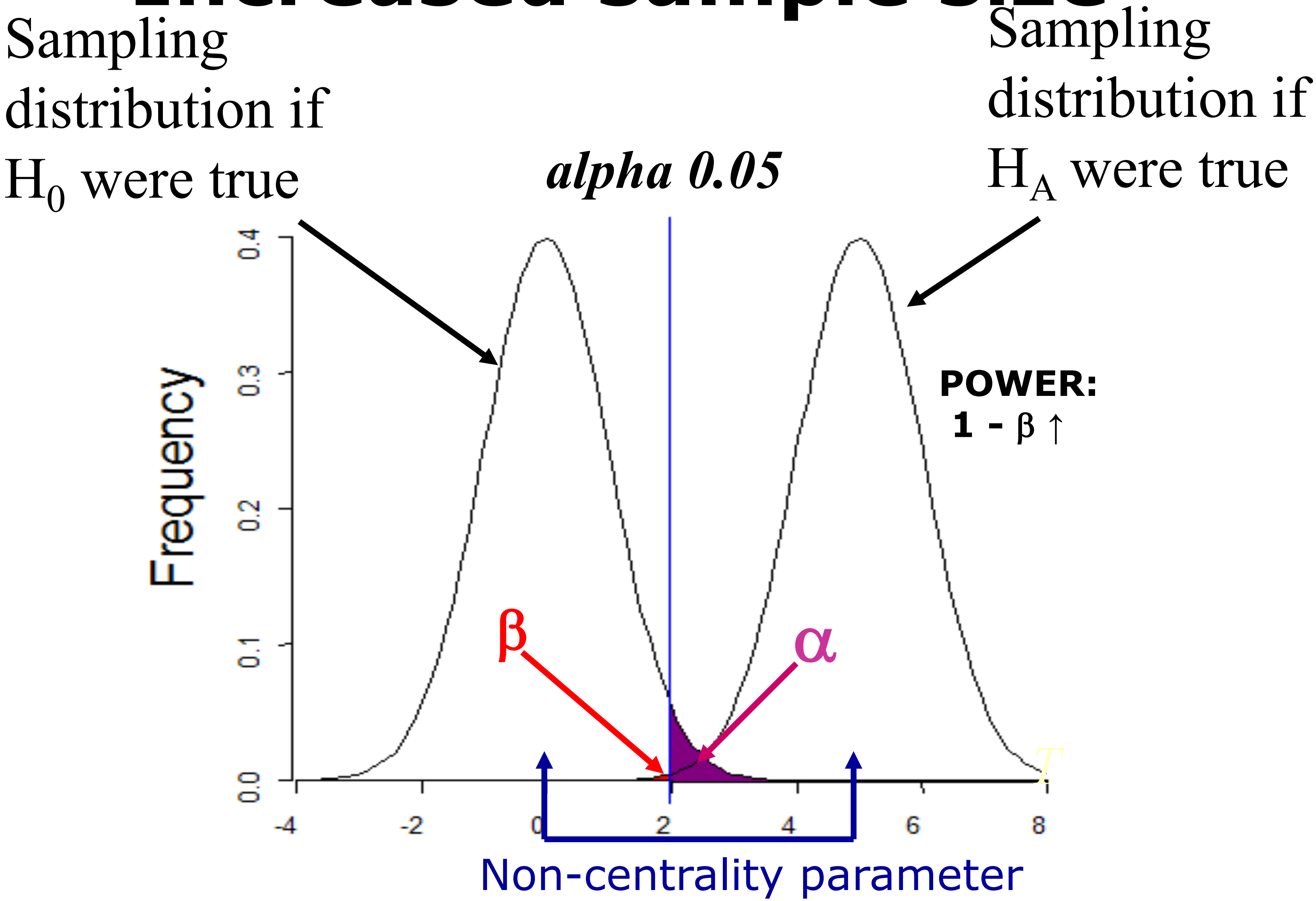
Sampling distribution if H_A were true



Standard Case



Increased sample size

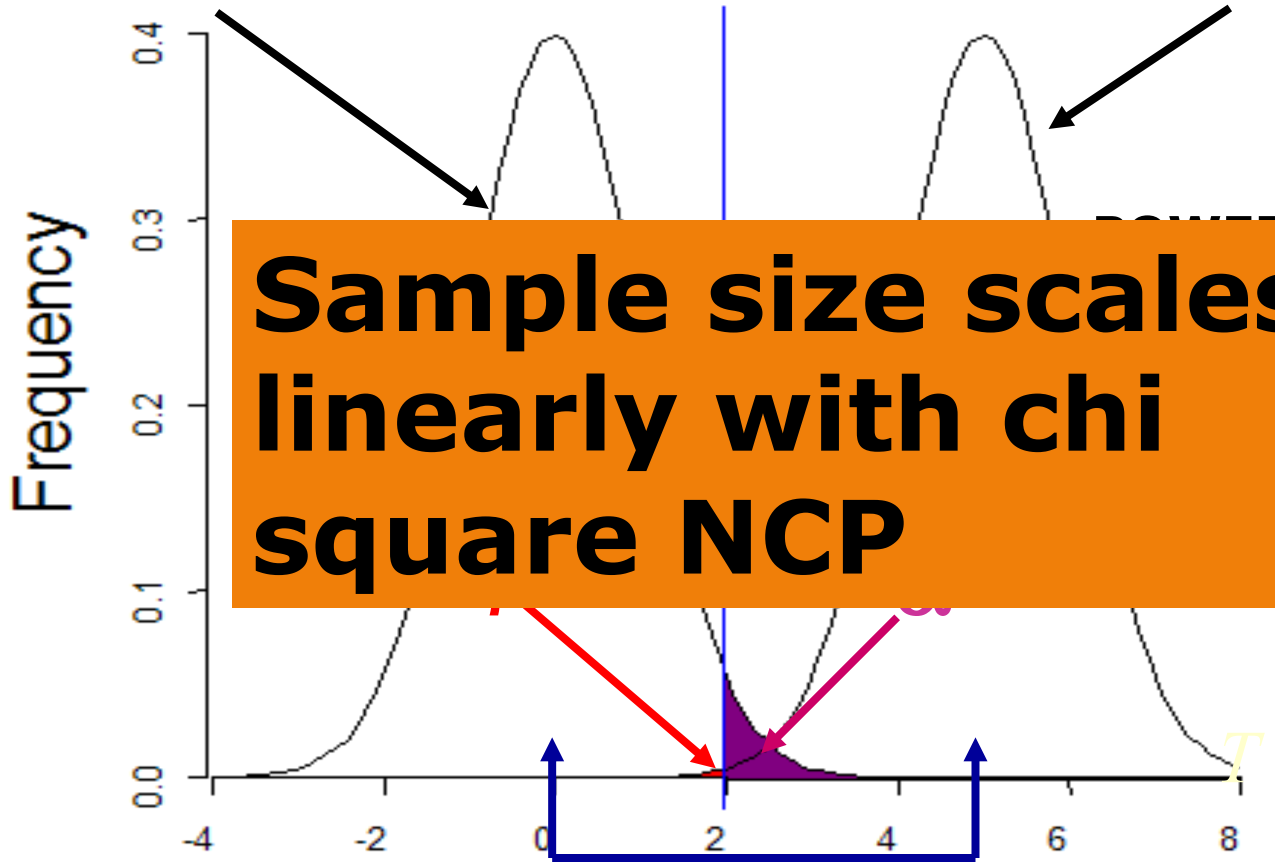


Increased sample size

Sampling distribution if H_0 were true

alpha 0.05

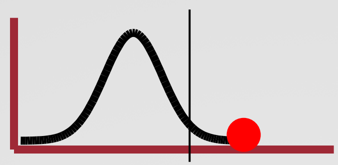
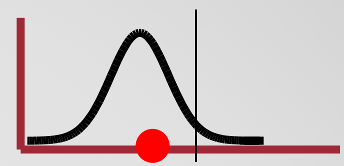
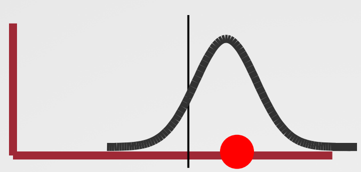
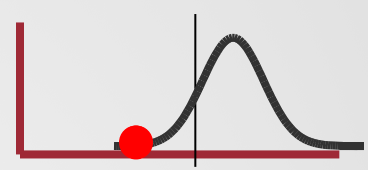
Sampling distribution if H_A were true



Sample size scales linearly with chi square NCP

Non-centrality parameter

Statistical Analysis

Truth	Rejection of H_0	Non-rejection of H_0
H_0 true	Type I error at rate α 	Nonsignificant result $(1-\alpha)$ 
H_A true	Significant result $(1-\beta)$ 	Type II error at rate β 

- Definitions of power
- The probability that the test will reject the null hypothesis if the alternative hypothesis is true

- The chance the your statistical test will yield a significant result when the effect you are testing exists

What is power?

- We are going to simulate a normal distribution using R

- We can do this with a single line of code, but let’s break it up

Practical 1

- R has functions for many distributions
- Normal, χ^2, gamma, beta (others)
- Let's start by looking at the random normal function: rnorm()

Simulation functions

rnorm Documentation

R Help

Normal {stats} R Documentation

The Normal Distribution

Description

Density, distribution function, quantile function and random generation for the normal distribution with mean equal to `mean` and standard deviation equal to `sd`.

Usage

```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

Arguments

`x,q`	vector of quantiles.
`p`	vector of probabilities.
`n`	number of observations. If `length(n) > 1`, the length is taken to be the number required.
`mean`	vector of means.
`sd`	vector of standard deviations.

In R: ?rnorm

rnorm syntax

- rnorm(n, mean = 0, sd = 1)

Function name

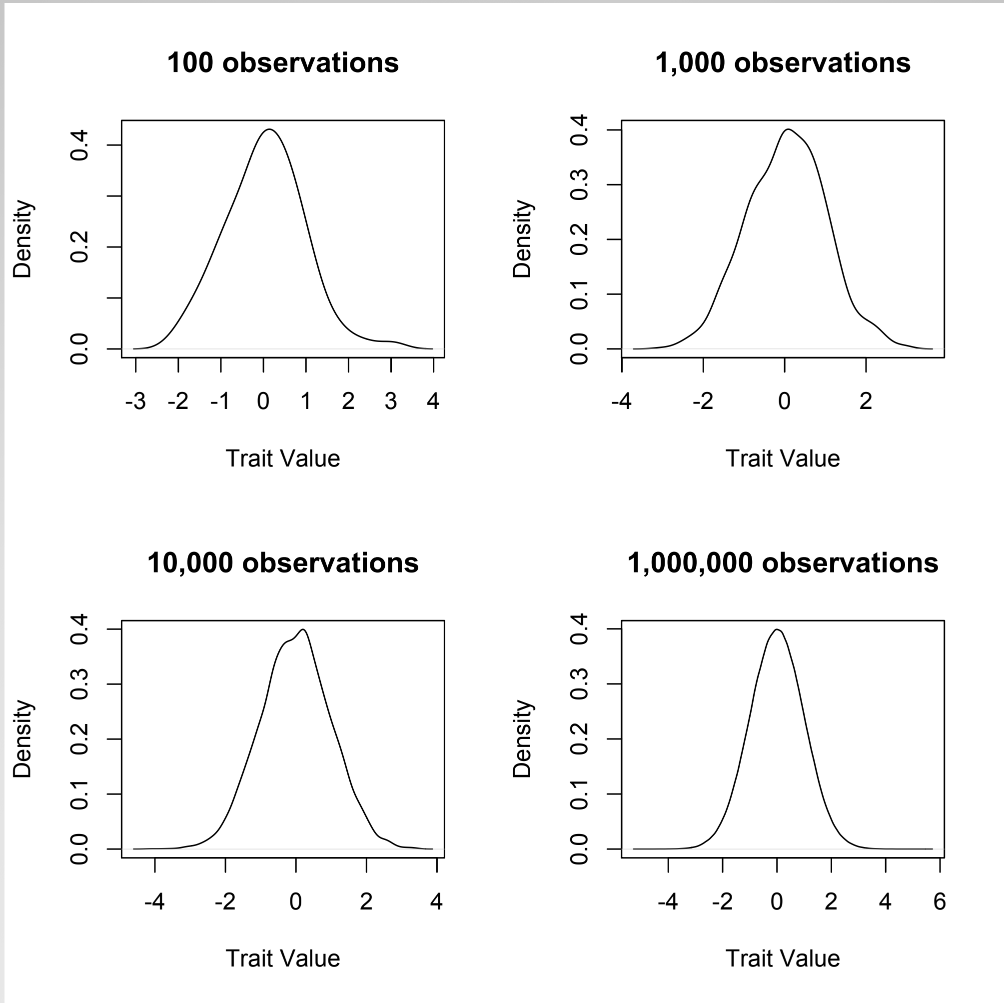
Number of observations to simulate

Mean of distribution with default value

Standard deviation of distribution with default value

- This script will plot 4 samples from the normal distribution
- Look for changes in shape
- Thoughts?

R script: Norm_dist_sim.R



One I made earlier

- Sampling variance
  - We saw that the ‘normal’ distribution from 100 observations looks stranger than for 1,000,000 observations
- Where else may this sampling variance happen?
- How certain are we that we have created a good distribution?

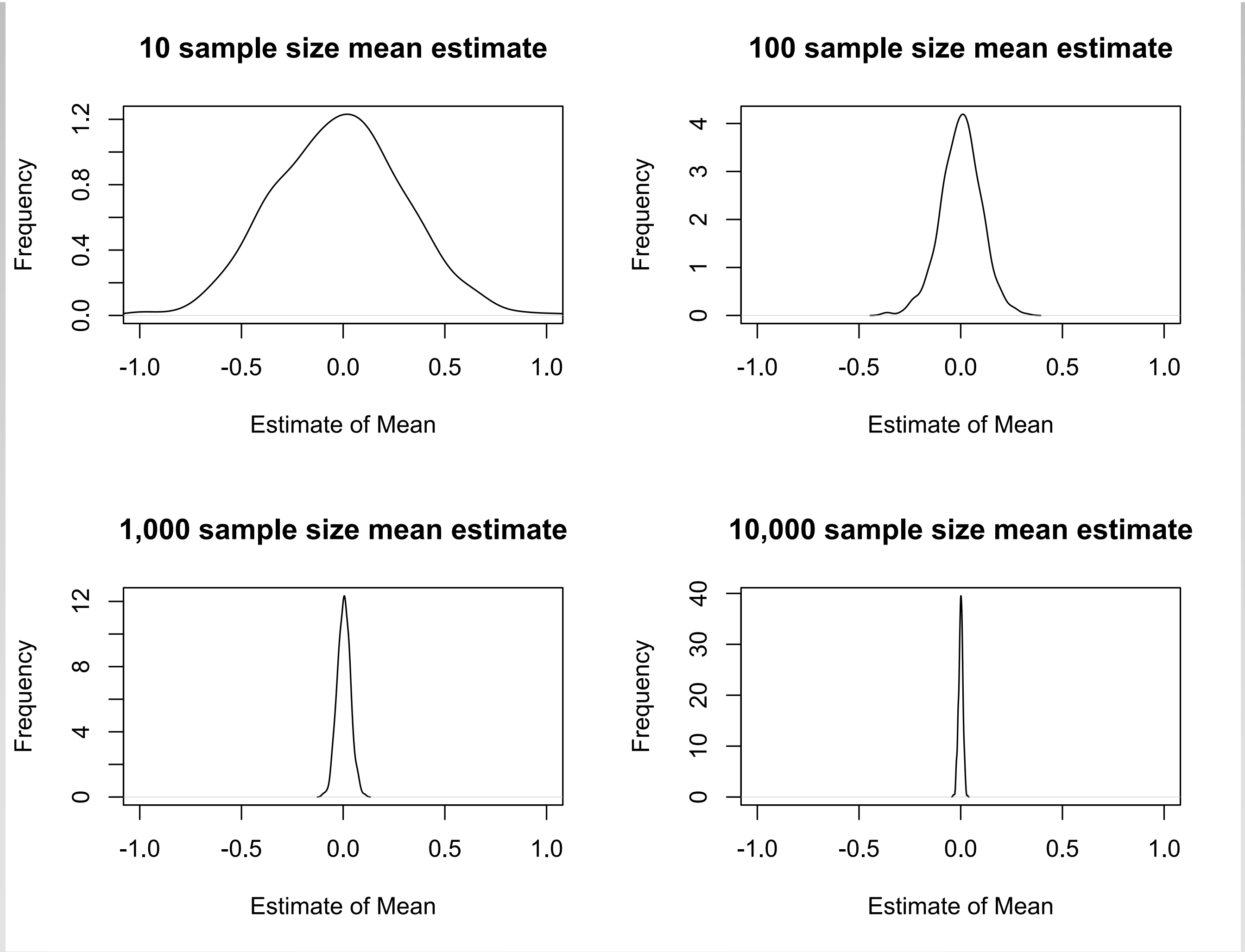
Concepts

- Rather than just simulating the normal distribution, let's simulate what our estimate of a mean looks like as a function of sample size
- We will run the R script mean_estimate_sim.R

Mean estimation

- This script will plot 4 samples from the normal distribution
- Look for changes in shape
- Thoughts?

R script: mean_estimate_sim.R



One I made earlier

- We see an inverse relationship between sample size and the variance of the estimate
- This variability in the estimate can be calculated from theory
- $SE_x = s/\sqrt{n}$
- SE_x is the standard error, s is the sample standard deviation, and n is the sample size

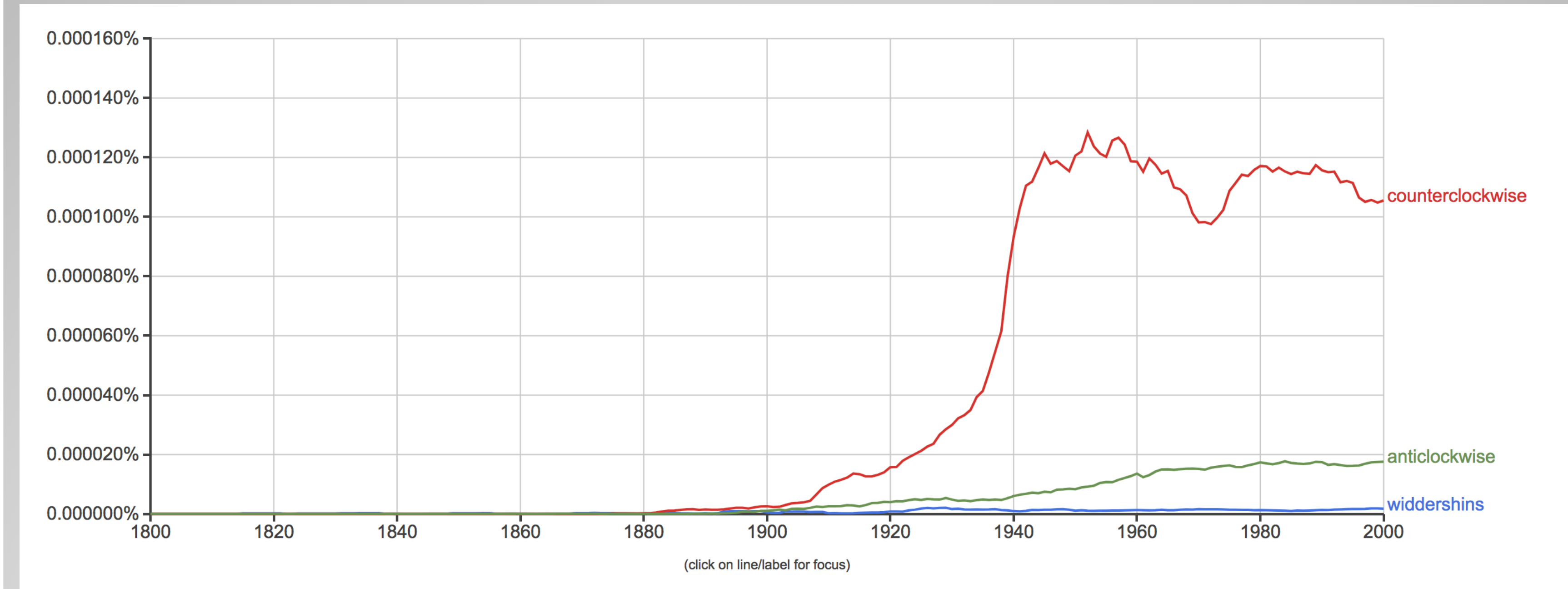
Standard Error

The sampling variability in my estimate affects my ability to declare a parameter as significant (or significantly different)

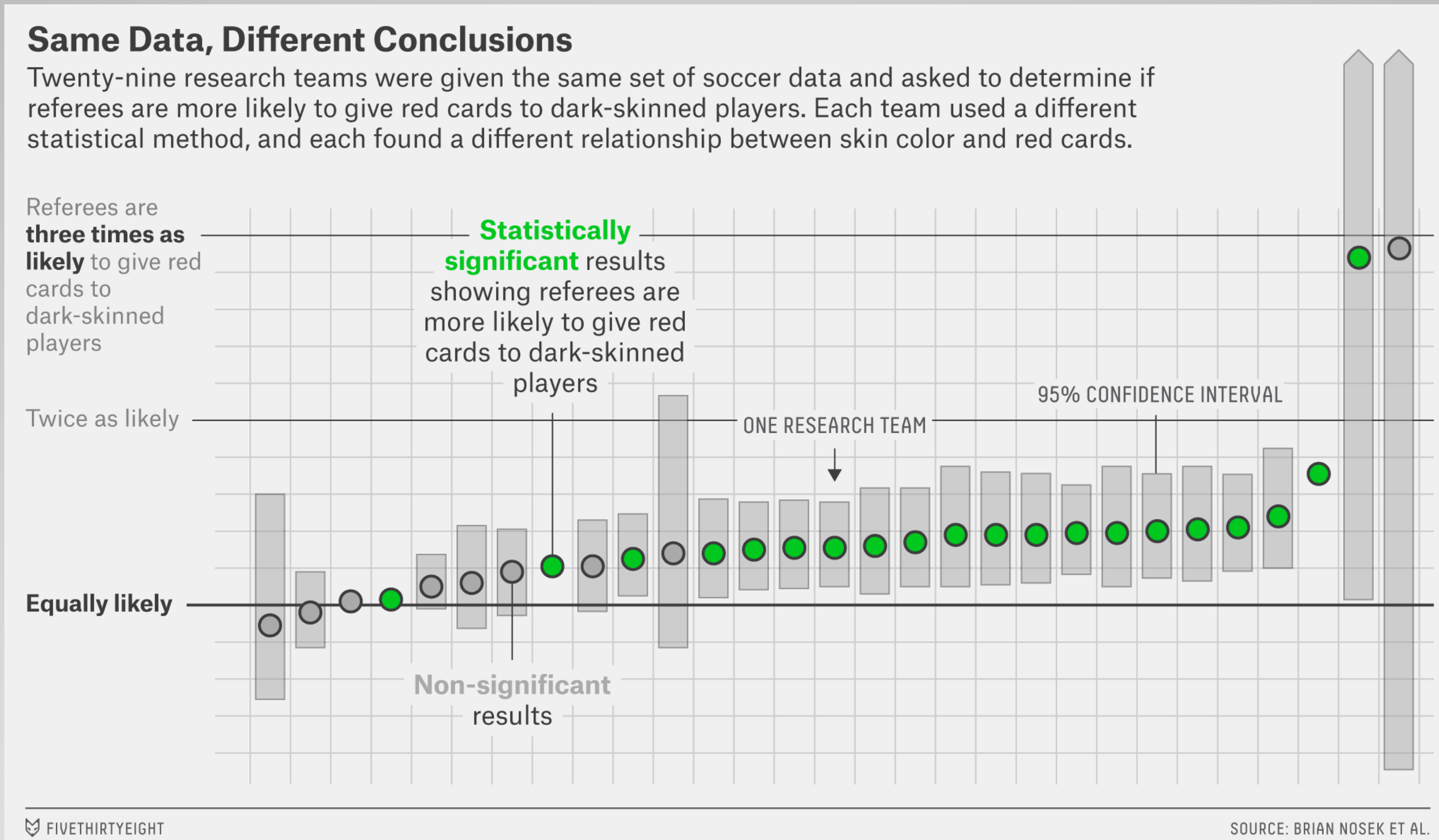
Key Concept 1

The probability that the test will reject the null hypothesis if the alternative hypothesis is true

Power definition again



What about when you have the total population?



Analytic choices matter a lot

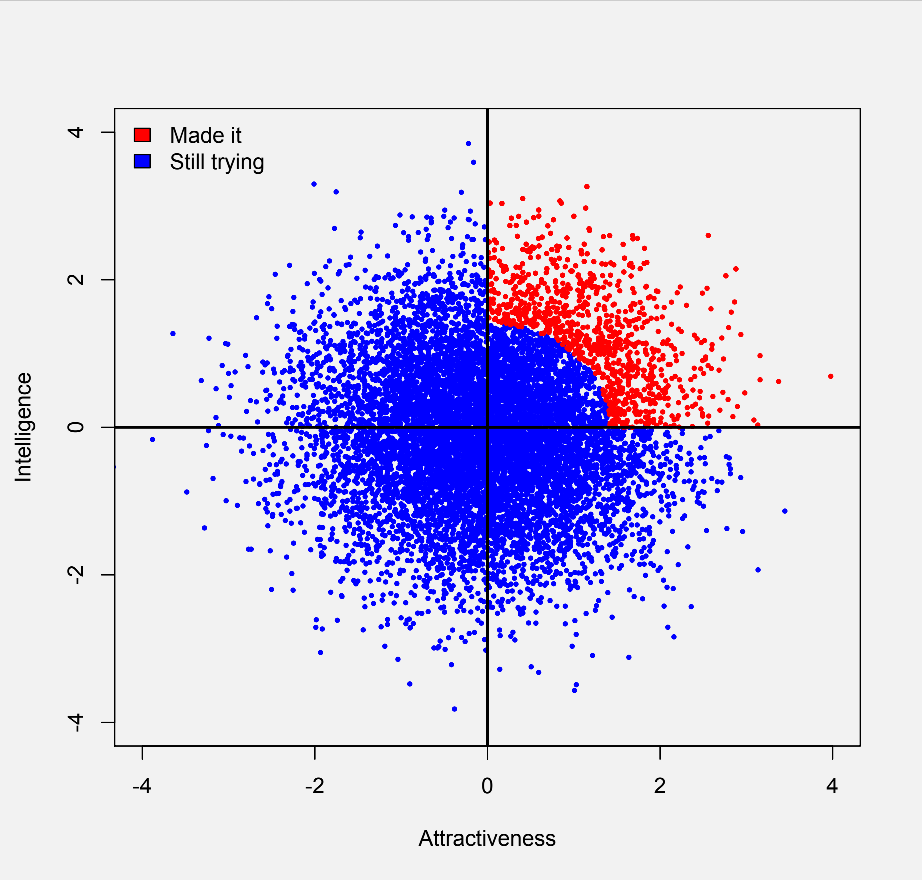
https://fivethirtyeight.com/features/science-isnt-broken/#part1

- Rather than disprove the null – estimate posterior probability
- Attempt to condition on the range of possibilities

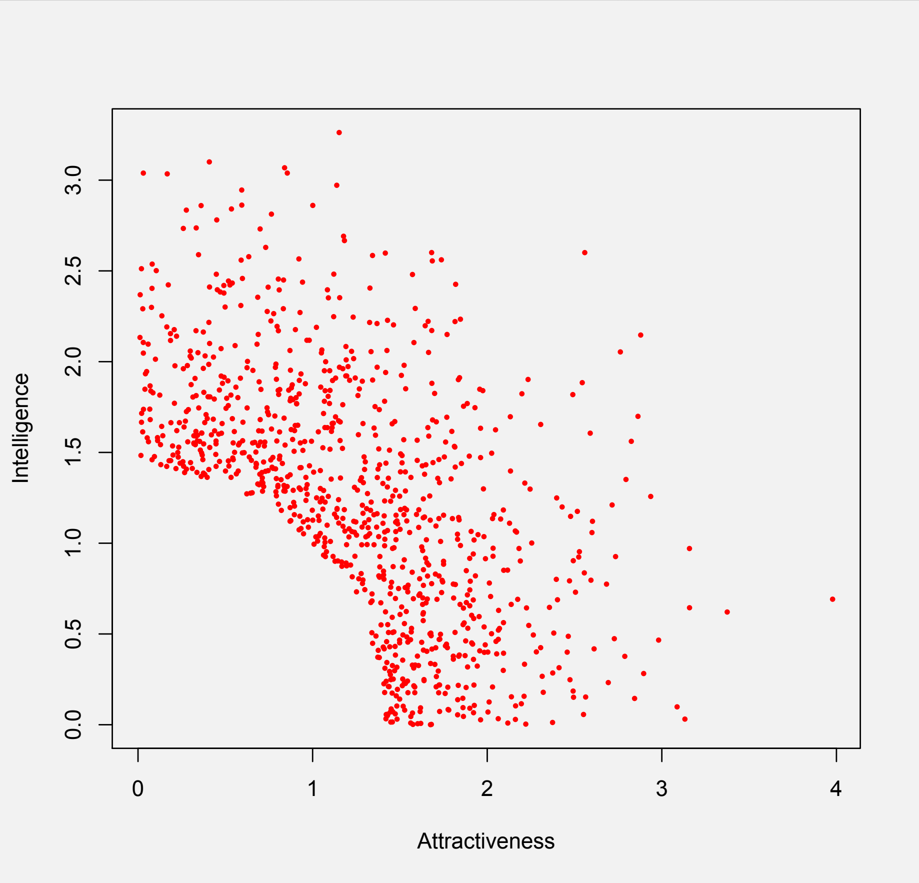
Bayesian philosophy

Ascertainment

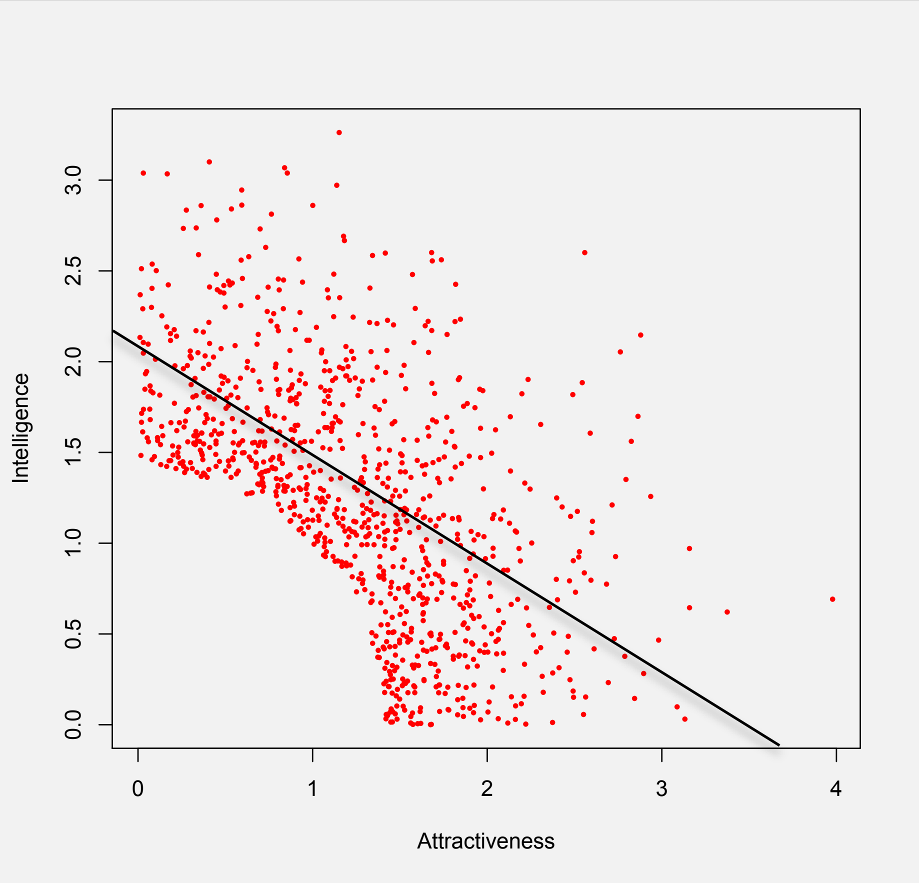
Why being picky can be good and bad



Bivariate plot for actors in Hollywood



Bivariate plot for actors who "made it" in Hollywood



P<2e-16

Bivariate plot for actors who "made it" in Hollywood

- Bias in your parameter estimates
  - Bias is a difference between the “true value” and the estimated value
- Can apply across a range of scenarios
  - Bias estimates of means, variances, covariances, betas etc.

Ascertainment

When might we want to ascertain?

- For testing means, ascertainment increases power
- For characterizing variance:covariance structure, ascertainment can lead to bias

- For testing means, ascertainment increases power
- For characterizing variance:covariance structure, ascertainment can lead to bias

Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps

Eric S. Lander,[*,†,‡] and David Botstein[‡,§]

[*]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, [†]Harvard University, Cambridge, Massachusetts 02138, [‡]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and [§]Genentech, South San Francisco, California 94080

Manuscript received August 2, 1988
Accepted for publication October 6, 1988

When might we want to ascertain?

- Power calculations using NCP
- We create the model
  - specifying our effect sizes
- We then simulate data
  - empirical = T means that the simulated data matches the specifications [within some error]
- The chi square can then be used to generate power

Practical!