Power and Sample Size
+
Principles of Simulation

Benjamin Neale
March 8th, 2012
International Twin Workshop, Boulder, CO

- What is power?
- What affects power?
- How do we calculate power?
- What is simulation?
- Why do we simulate?
- How do we simulate?

Slide of Questions

What is power?

- 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

- Null Hypothesis
- Alternative Hypothesis
- Distribution of test statistics

Key Concepts and Terms

Key Concepts and Terms

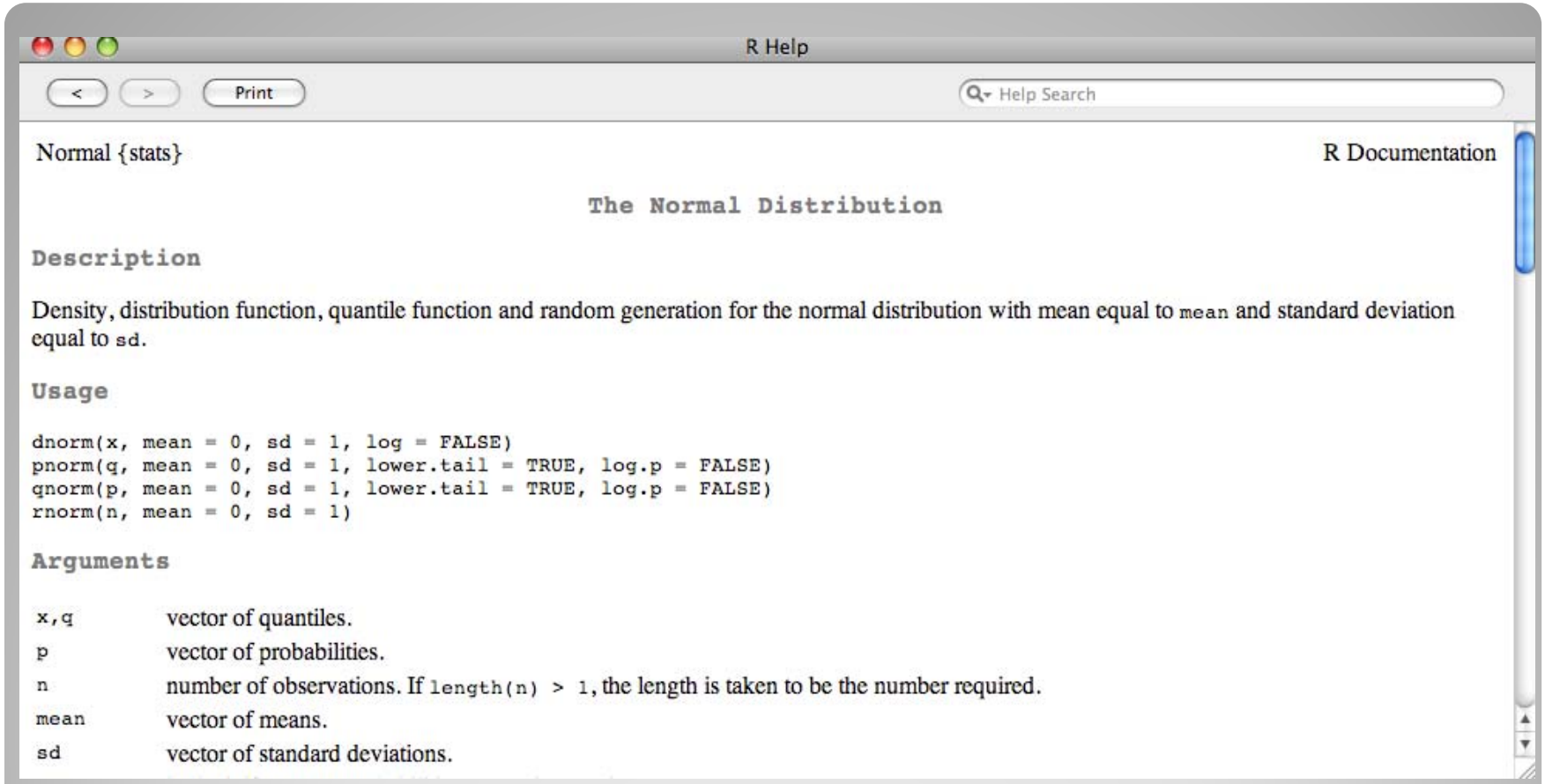
- Null Hypothesis
  - The baseline hypothesis, generally assumed to be the absence of the tested effect
- Alternative Hypothesis
  - The hypothesis for the presence of an effect
- Distribution of test statistics
  - The frequencies of the values of the tests statistics under the null and the alternative

- 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

Simulation functions

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

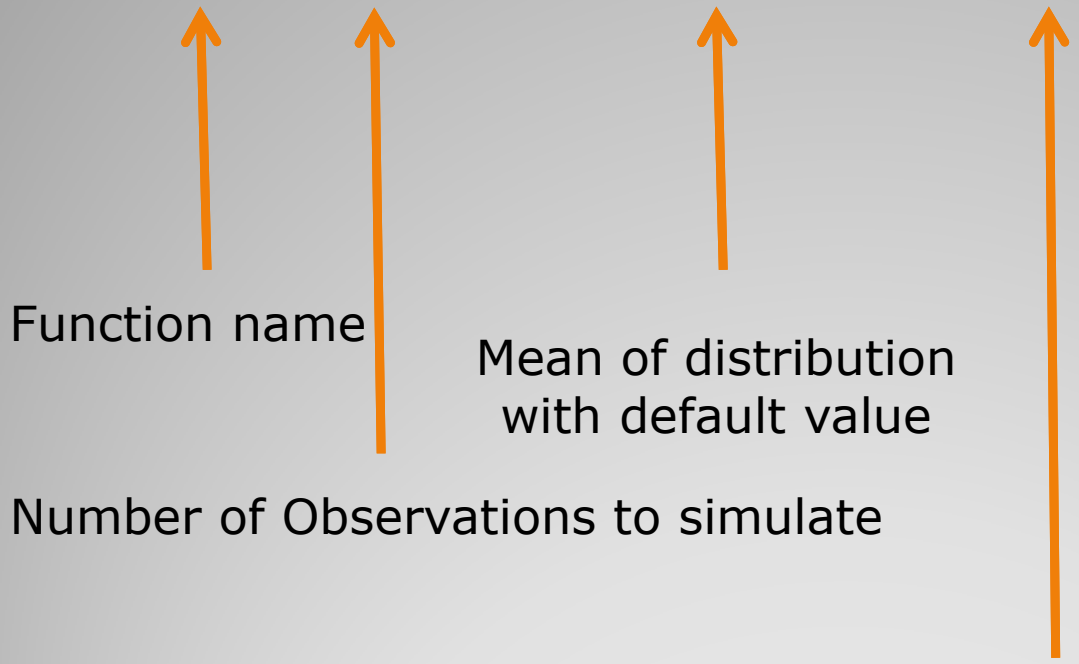


In R: ?rnorm

rnorm Documentation

rnorm syntax

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



Function name

Number of Observations to simulate

Mean of distribution with default value

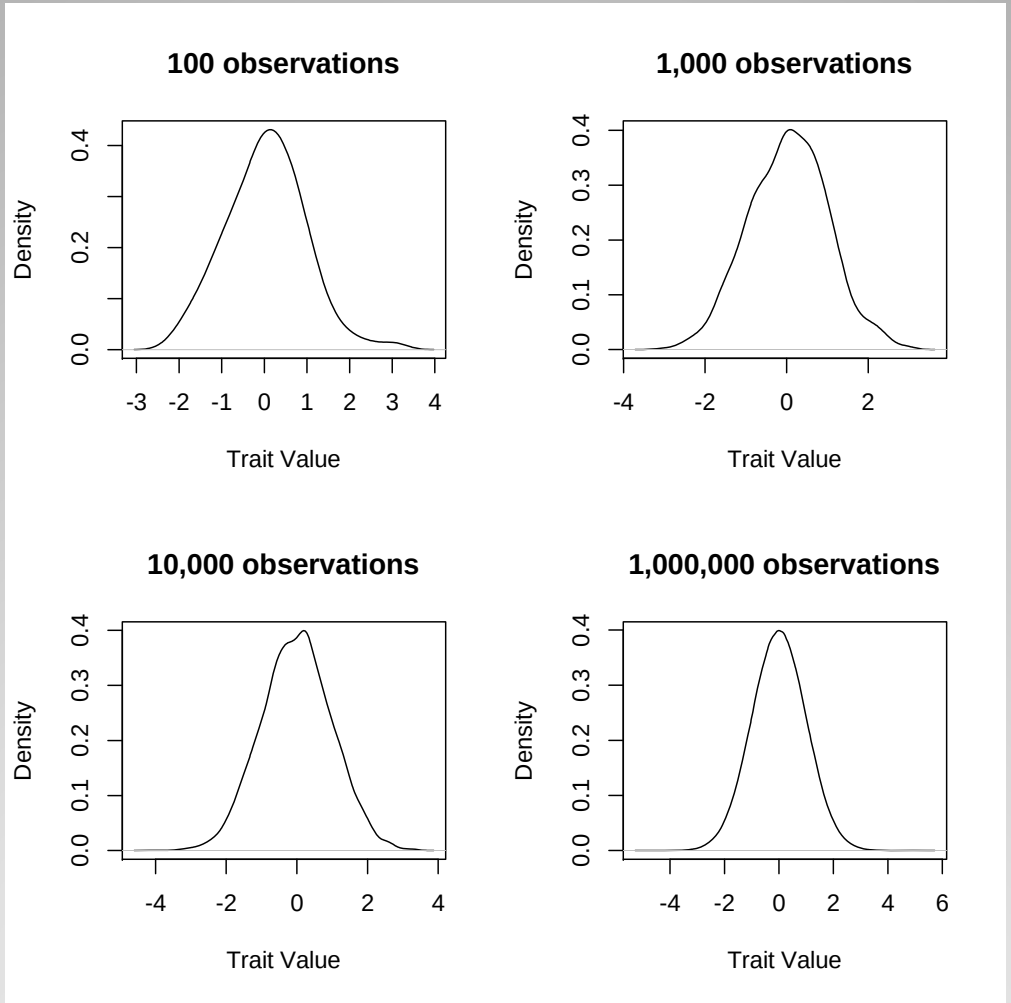
Standard deviation of distribution with default value

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

R script: Norm_dist_sim.R

- You have to comment
- The presentation will not continue without audience participation
  - No this isn’t a game of chicken

What did we learn?



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

Mean estimation

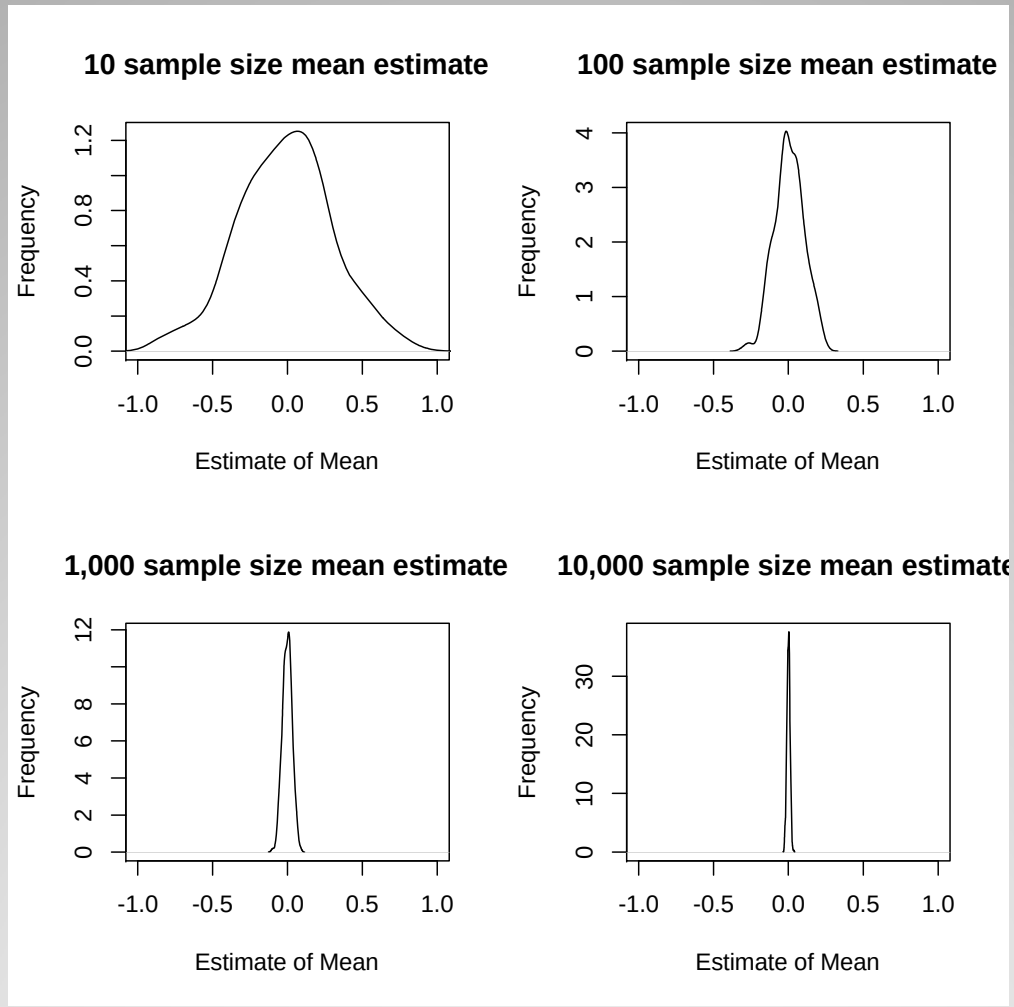
- Rather than just simulating the normal distribution, let's simulate what our estimate of a mean looks like as a function of sample size
- We will run the R script mean_estimate_sim.R

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

R script: mean_estimate_sim.R

- You have to comment
- The presentation will not continue without audience participation
  - No this isn’t a game of chicken

What did we learn?



One I made earlier

Standard Error

- 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

Existential Crisis!

What does this variability mean?

Again—this is where you comment

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

Key Concept 1

Power definition again

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

Hypothesis Testing

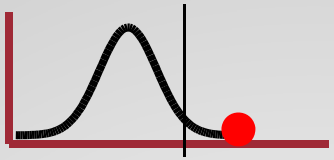
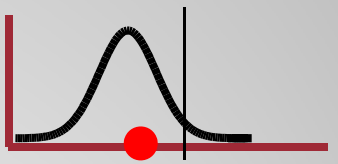
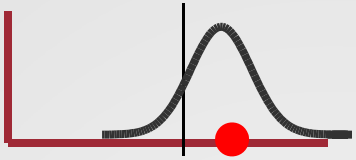
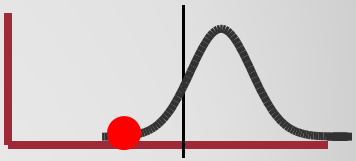
- Mean different from 0 hypotheses:
  - h_o (null hypothesis) is $\mu=0$
  - h_a (alternative hypothesis) is $\mu \neq 0$
    - Two-sided test, where $\mu > 0$ or $\mu < 0$ are one-sided
- Null hypothesis usually assumes no effect
- Alternative hypothesis is the idea being tested

Possible scenarios

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

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

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 

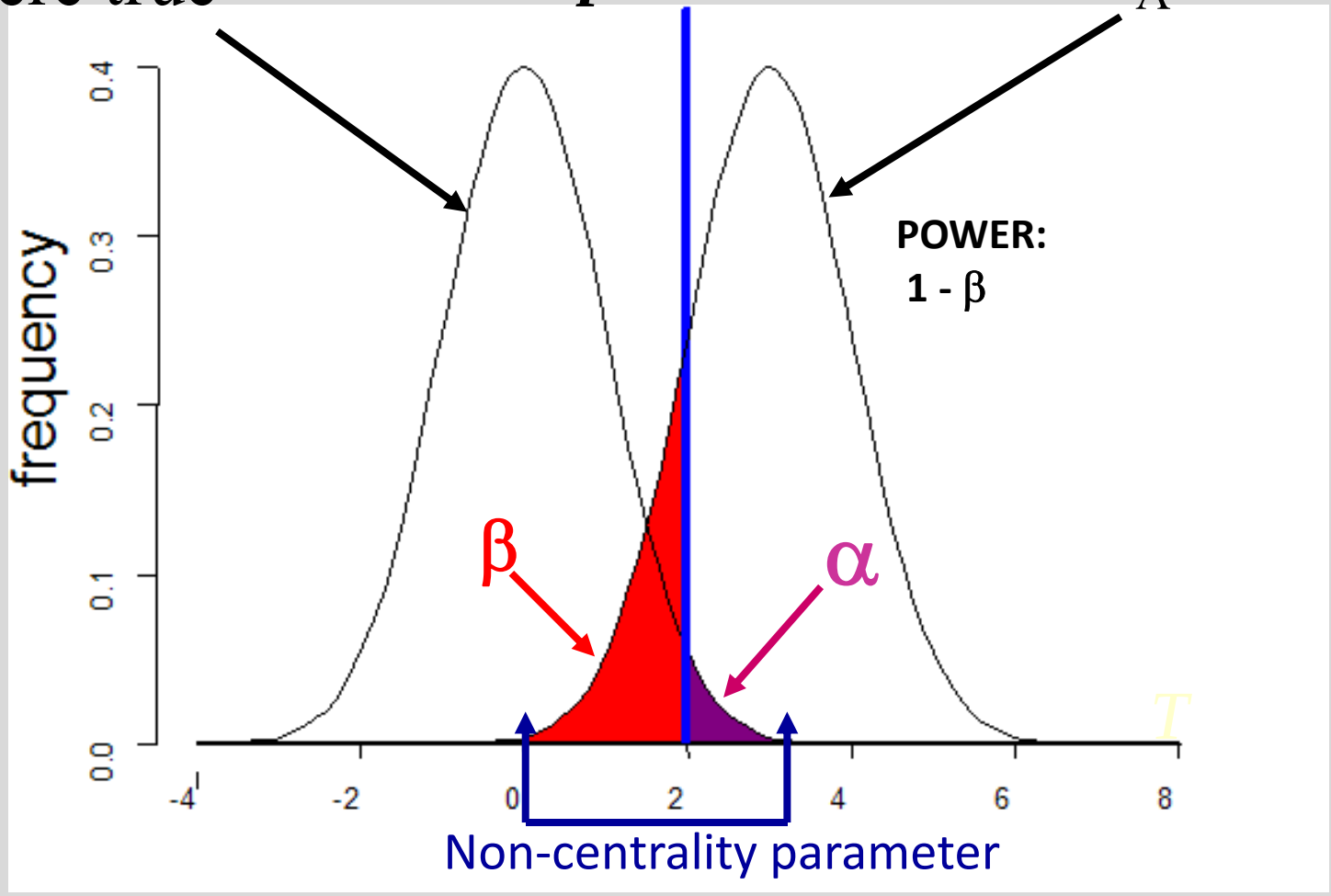
Expanded Power Definition

- The probability of rejection of the null hypothesis depends on:
  - The significance criterion (α)
  - The sample size (N)
  - The effect size (Δ)
- The probability of detecting a given effect size in a population from a sample size, N, using a significance criterion, α.

Standard Case

Sampling distribution if H_0 were true

Sampling distribution if H_A were true

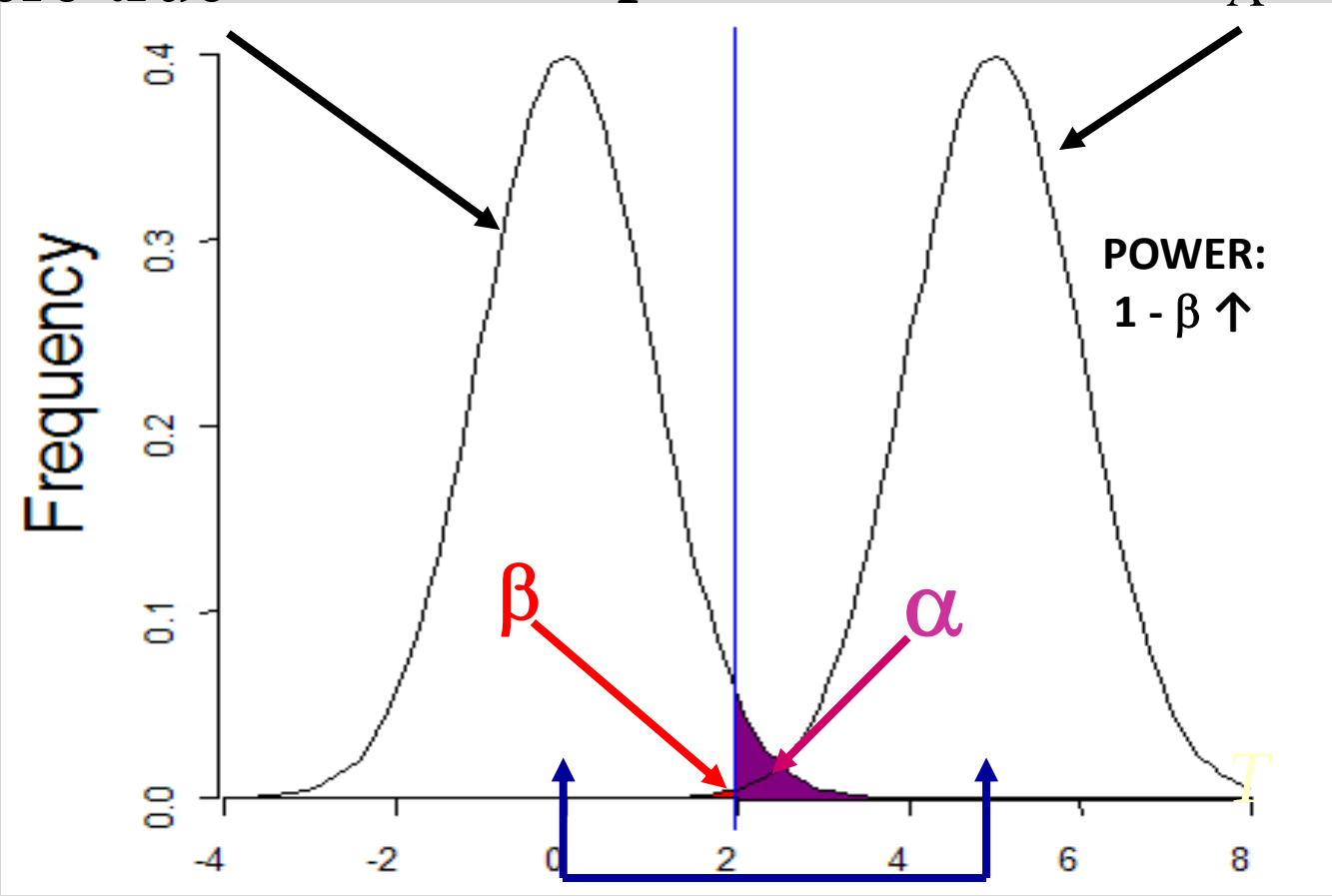


Increased effect size

Sampling distribution if H_0 were true

alpha 0.05

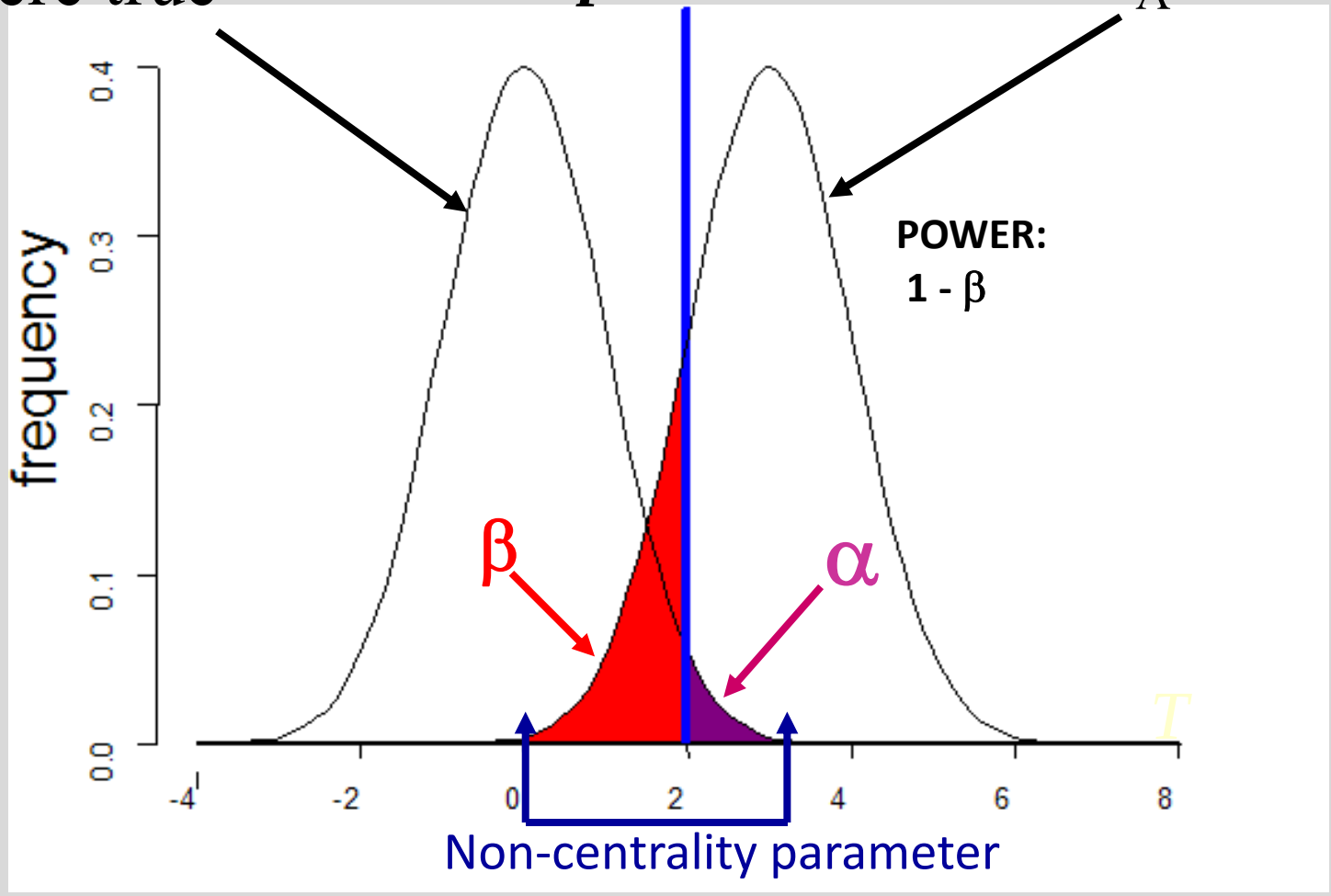
Sampling distribution if H_A were true



Standard Case

Sampling distribution if H_0 were true

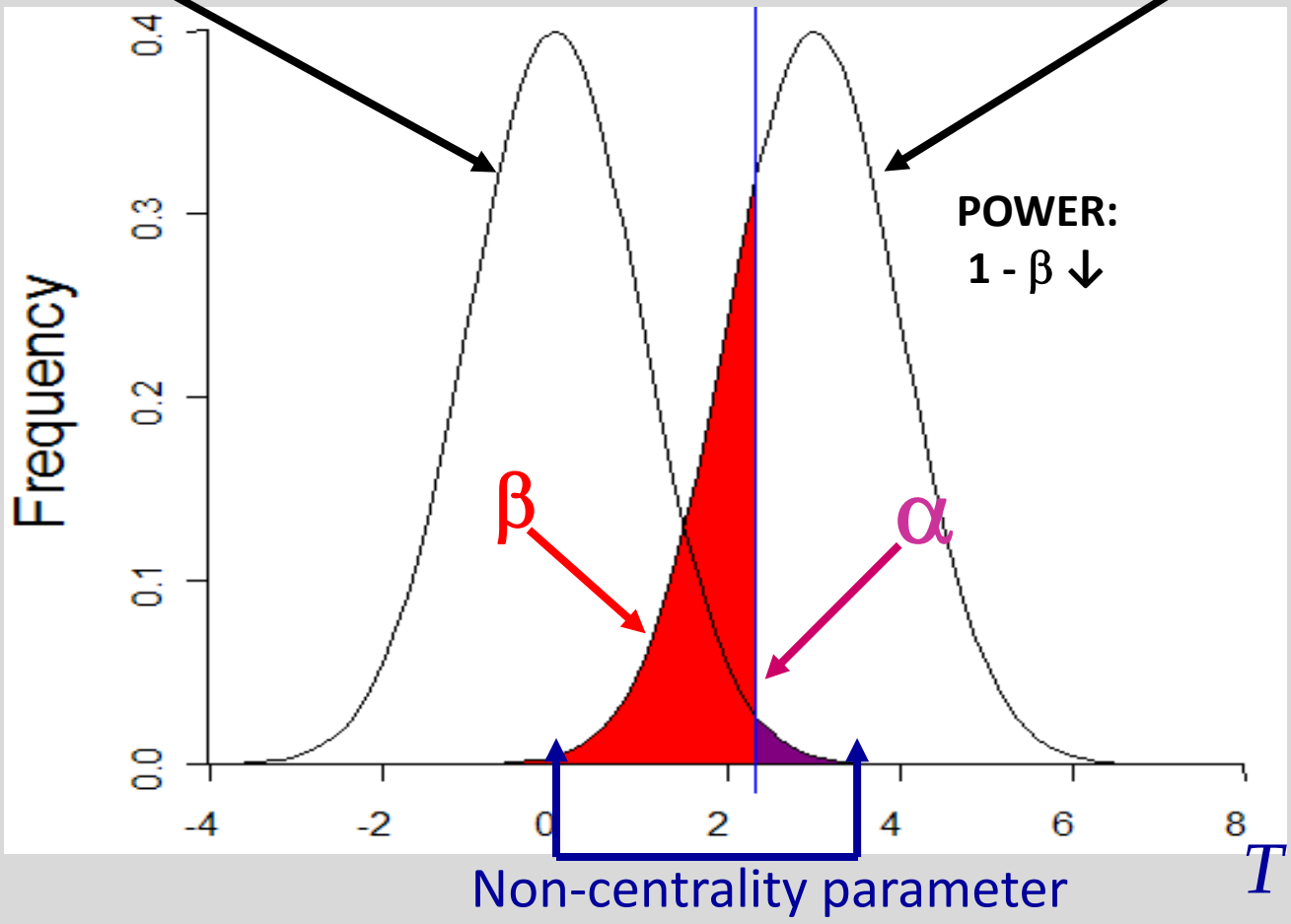
Sampling distribution if H_A were true



More conservative α

Sampling distribution if H_0 were true

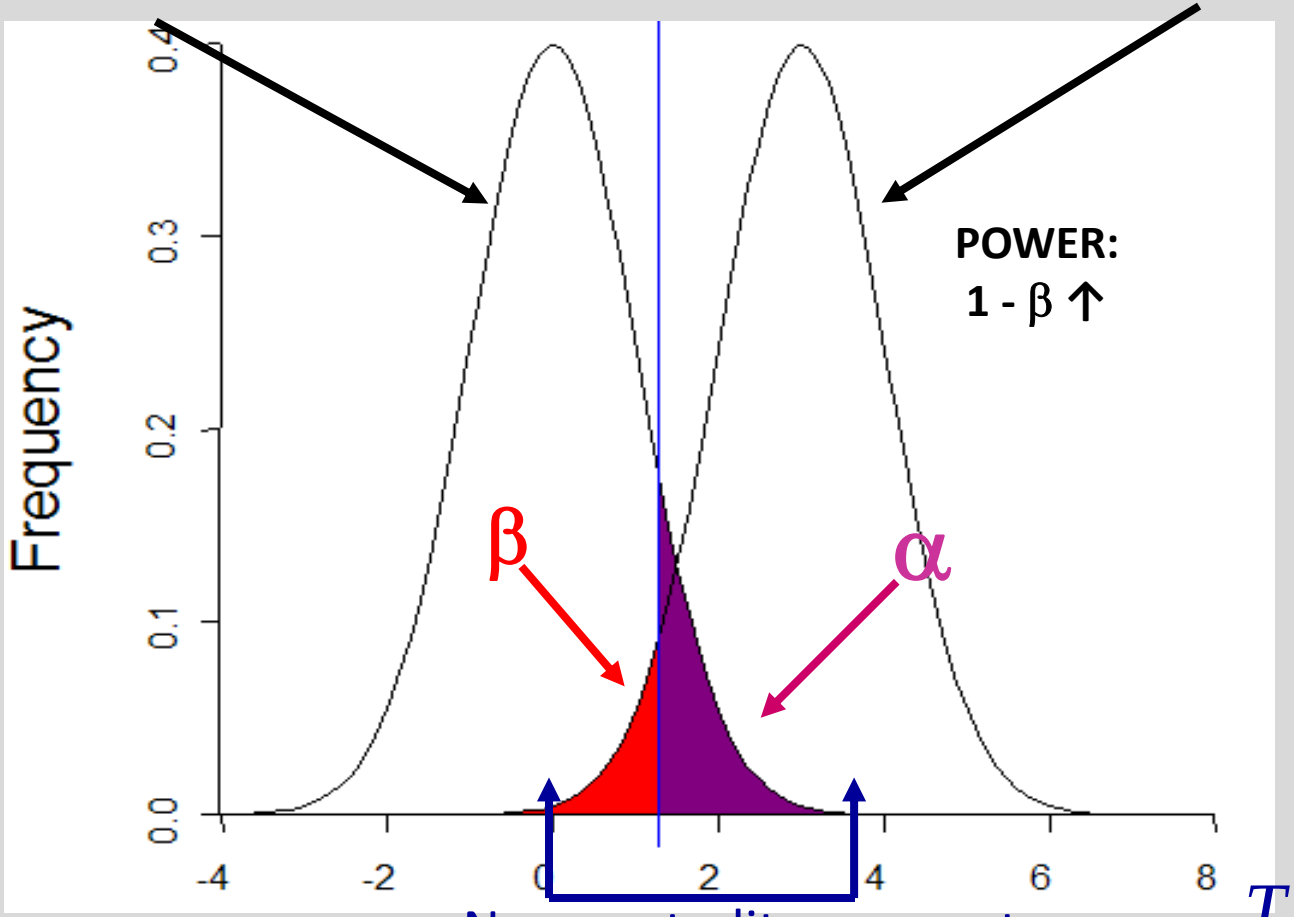
Sampling distribution if H_A were true



Less conservative α

Sampling distribution if H_0 were true

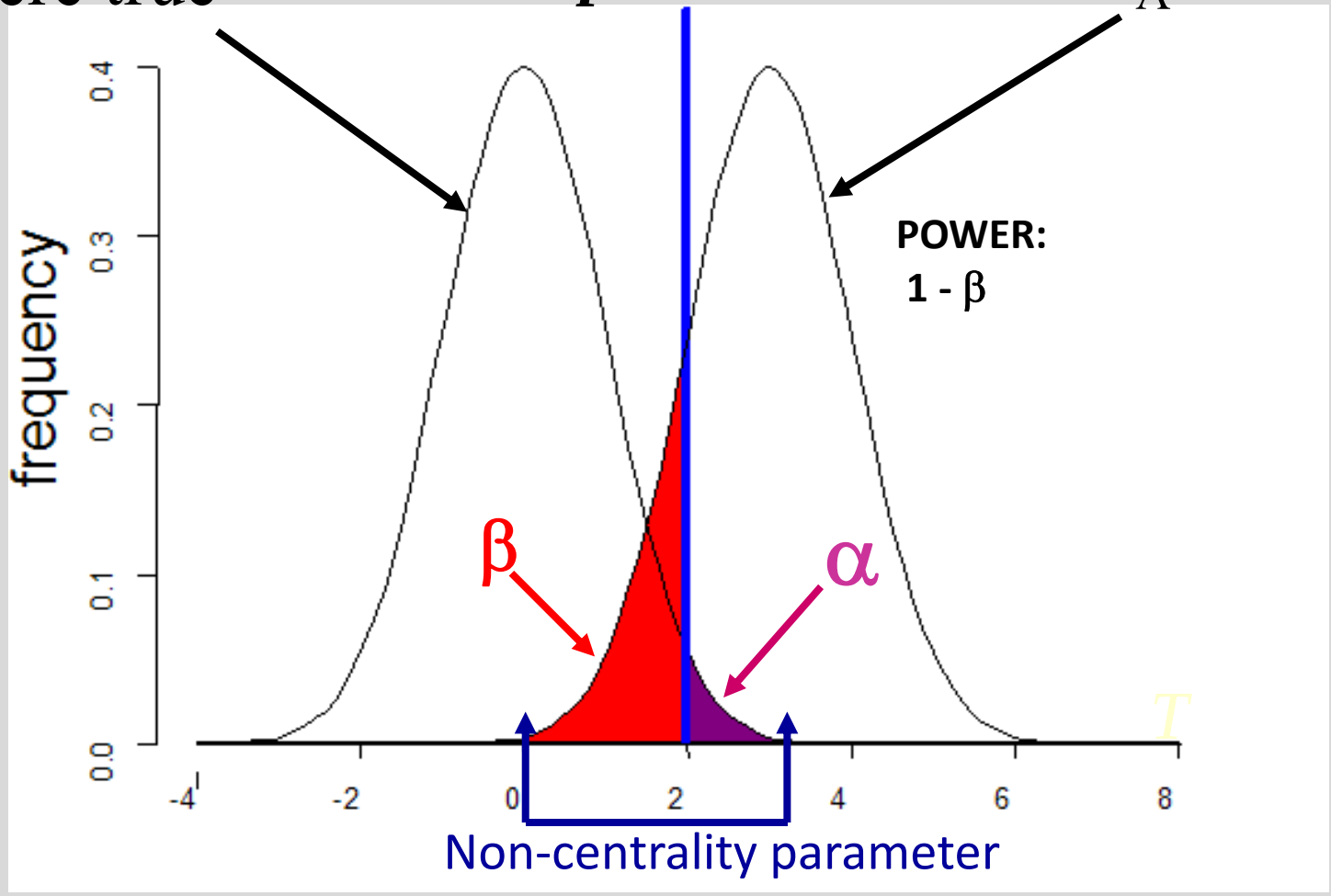
Sampling distribution if H_A were true



Standard Case

Sampling distribution if H_0 were true

Sampling distribution if H_A were true

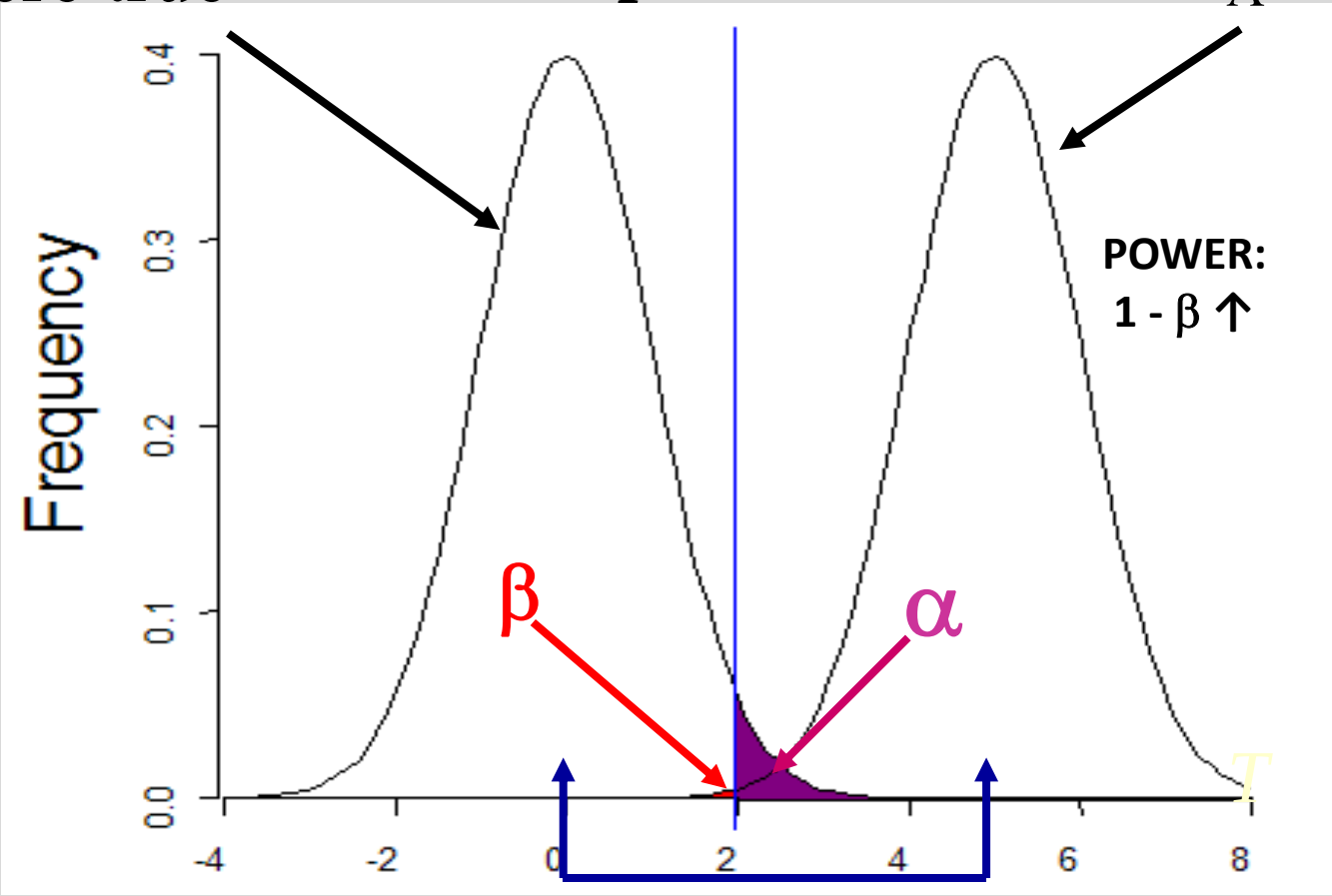


Increased sample size

Sampling distribution if H_0 were true

alpha 0.05

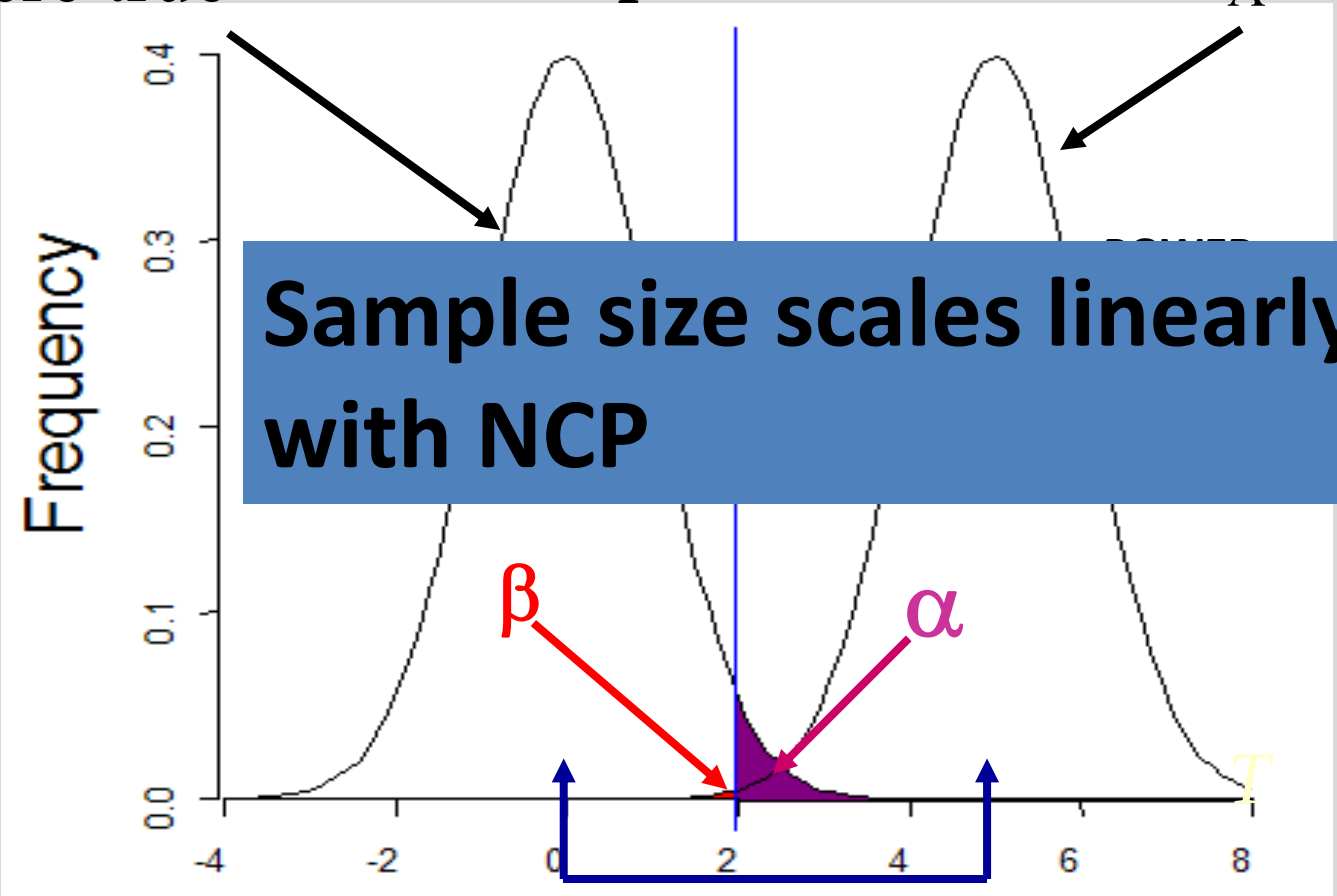
Sampling distribution if H_A were true



Increased sample size

Sampling distribution if H_0 were true

Sampling distribution if H_A were true



Sample size scales linearly with NCP

Additional Factors

- Type of Data
  - Continuous > Ordinal > Binary
  - Do not turn “true” binary variables into continuous
- Multivariate analysis
- Remove confounders and biases
- MZ:DZ ratio

Effects Recap

- Larger effect sizes
  - Reduce heterogeneity
- Larger sample sizes
- Change significance threshold
  - False positives may become problematic

2. The power of discriminating between simple models

(i) *Theory*

If we have a set of observed mean squares 0_i, their expected values calculated on the basis of the " true " model of variation F_i and the expected values calculated on the basis of a " false " model, E_i, then we wish, for each i to test the null hypothesis

$$H_0: \mathscr{E}(0_i) = E_i$$

If H_0 is true and the degrees of freedom ν_i are large then $0_i \sim N(E_i, 2E_i^2/\nu_i)$ approximately. We assume H_0 is to be tested by the statistic $\frac{\nu_i(0_i-E_i)^2}{2E_i^2}$ which is approximately chi-square with one degree of freedom. The power function of this test is not known for all alternative hypotheses. Consequently we follow Mitra (1958) and consider the limiting power function of the test for large sample sizes and alternative models not too far from the hypothesised model. To express this idea suppose

$$H_1: \mathscr{E}(0_i) = F_i = E_i + \mu_i/\sqrt{\nu_i}$$

where $\mu_i = \sqrt{\nu_i}(F_i - E_i)$ is the deviation between the two models. Given that ν_i is large, then if H_0 is true

$$0_i \sim N(E_i, 2E_i^2/\nu_i)$$

while if H_1 is true then

$$0_i \sim N(F_i, 2F_i^2/\nu_i) \sim N(F_i, (2E_i^2/\nu_i)(1+o(\nu_i^{-\frac{1}{2}}))).$$

where $o(\nu_i^{-\frac{1}{2}})$ denotes a term of the order of $\nu_i^{-\frac{1}{2}}$.
Thus

$$\frac{\sqrt{\nu_i}(0_i-E_i)}{\sqrt{2E_i^2}} \sim N\left(\frac{\sqrt{\nu_i}(F_i-E_i)}{\sqrt{2E_i^2}},\ 1+o(\nu_i^{-\frac{1}{2}})\right)$$

and the asymptotic power function of

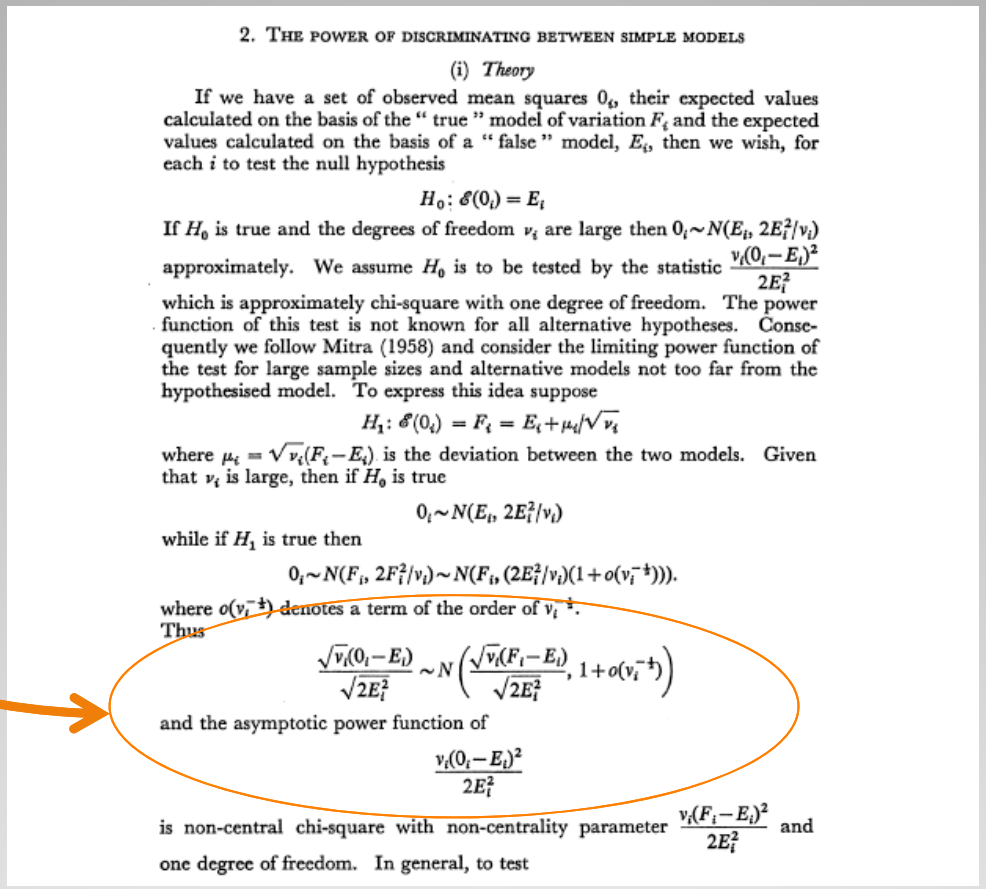
$$\frac{\nu_i(0_i-E_i)^2}{2E_i^2}$$

is non-central chi-square with non-centrality parameter $\frac{\nu_i(F_i-E_i)^2}{2E_i^2}$ and one degree of freedom. In general, to test

Power of Classical Twin Model

You can do this by hand

But machines can be much more fun



2. THE POWER OF DISCRIMINATING BETWEEN SIMPLE MODELS

(i) *Theory*

If we have a set of observed mean squares 0_i, their expected values calculated on the basis of the " true " model of variation F_i and the expected values calculated on the basis of a " false " model, E_i, then we wish, for each i to test the null hypothesis

$$H_0: \mathscr{E}(0_i) = E_i$$

If H_0 is true and the degrees of freedom ν_i are large then $0_i \sim N(E_i, 2E_i^2/\nu_i)$ approximately. We assume H_0 is to be tested by the statistic $\frac{\nu_i(0_i-E_i)^2}{2E_i^2}$ which is approximately chi-square with one degree of freedom. The power function of this test is not known for all alternative hypotheses. Consequently we follow Mitra (1958) and consider the limiting power function of the test for large sample sizes and alternative models not too far from the hypothesised model. To express this idea suppose

$$H_1: \mathscr{E}(0_i) = F_i = E_i + \mu_i/\sqrt{\nu_i}$$

where $\mu_i = \sqrt{\nu_i}(F_i - E_i)$ is the deviation between the two models. Given that ν_i is large, then if H_0 is true

$$0_i \sim N(E_i, 2E_i^2/\nu_i)$$

while if H_1 is true then

$$0_i \sim N(F_i, 2F_i^2/\nu_i) \sim N(F_i, (2E_i^2/\nu_i)(1+o(\nu_i^{-\frac{1}{2}}))).$$

where $o(\nu_i^{-\frac{1}{2}})$ denotes a term of the order of $\nu_i^{-\frac{1}{2}}$. Thus

$$\frac{\sqrt{\nu_i}(0_i-E_i)}{\sqrt{2E_i^2}} \sim N\left(\frac{\sqrt{\nu_i}(F_i-E_i)}{\sqrt{2E_i^2}}, 1+o(\nu_i^{-\frac{1}{2}})\right)$$

and the asymptotic power function of

$$\frac{\nu_i(0_i-E_i)^2}{2E_i^2}$$

is non-central chi-square with non-centrality parameter $\frac{\nu_i(F_i-E_i)^2}{2E_i^2}$ and one degree of freedom. In general, to test

Power of Classical Twin Model

Martin, Eaves, Kearsey, and Davies Power of the Twin Study, Heredity, 1978

How can we determine our power to detect variance components?

A trio of scripts

- We'll run through three different scripts
- power_approximation.R
  - This lays out a theoretical consideration for correlations
- test_C_sim.R
  - Creates a function for running a simulated ACE model to drop C
- run_C_sim.R
  - This will run the function from test_C_sim.R

- We'll walk through the script to explore the core statistical concepts we discussed earlier
- Most of the rest of this session will be done in R ☺.

power_approximation.R

Simulation is king

- Just like we simulated estimates of means we can simulate chi squares from dropping C
- We get to play God [well more than usual]
  - We fix the means and variances as parameters to simulate
  - We fit the model ACE model
  - We drop C
  - We generate our alternative sampling distribution of statistics

- This script does not produce anything, but rather creates a function
- I will walk through the script explaining the content
- We will make this function and then run the function once to see what happens
- Then we can generate a few more simulations (though I'm not super keen on the computation power of the machines)

Script: test_C_sim.R

- Now we understand what the function is doing, we can run it many times
- We are going to use sapply again
- This time we will sapply the new function we made
- In addition to generating our chis, we create an object of the chis to assess our power and see what our results look like

Script: run_C_sim.R

Principles of Simulation

"*All models are wrong but some are useful*"
--George Box

- Simulation is useful for refining intuition
- Helpful for grant writing
- Calibrates need for sample size
- Many factors affect power
  - Ascertainment
  - Measurement error
  - Many others

Why R is great

- Simulation is super easy in R
- ClassicMx did not have such routines
- We can evaluate our power easily using R as well
- We can generate pictures of our power easily

Additional Online Resources

- Genetic Power Calculator
  - Good for genetic studies with genotype data
  - Also includes a probability function calculator
  - http://pngu.mgh.harvard.edu/~purcell/gpc/
- Wiki has pretty good info on statistical power
  - http://en.wikipedia.org/wiki/Statistical_power

Citations for power in twin modeling

Visscher PM, Gordon S, Neale MC., Power of the classical twin design revisited: II detection of common environmental variance. Twin Res Hum Genet. 2008 Feb;11(1):48-54.

Neale BM, Rijsdijk FV, Further considerations for power in sibling interactions models. Behav Genet. 2005 Sep 35(5):671-4

Visscher PM., Power of the classical twin design revisited., Twin Res. 2004 Oct;7(5):505-12.

Rietveld MJ, Posthuma D, Dolan CV, Boomsma DI, ADHD: sibling interaction or dominance: an evaluation of statistical power. Behav Genet. 2003 May 33(3): 247-55

Posthuma D, Boomsma DI., A note on the statistical power in extended twin designs. Behav Genet. 2000 Mar;30(2):147-58.

Neale MC, Eaves LJ, Kendler KS. The power of the classical twin study to resolve variation in threshold traits. Behav Genet. 1994 May;24(3):239-58.

Nance WE, Neale MC., Partitioned twin analysis: a power study. Behav Genet. 1989 Jan;19(1):143-50.

Martin NG, Eaves LJ, Kearsey MJ, Davies P., The power of the classical twin study. Heredity. 1978 Feb;40(1):97-116.