Monte Carlo Simulations for Inference and Power analyses II

October 14th 2014

Readings.

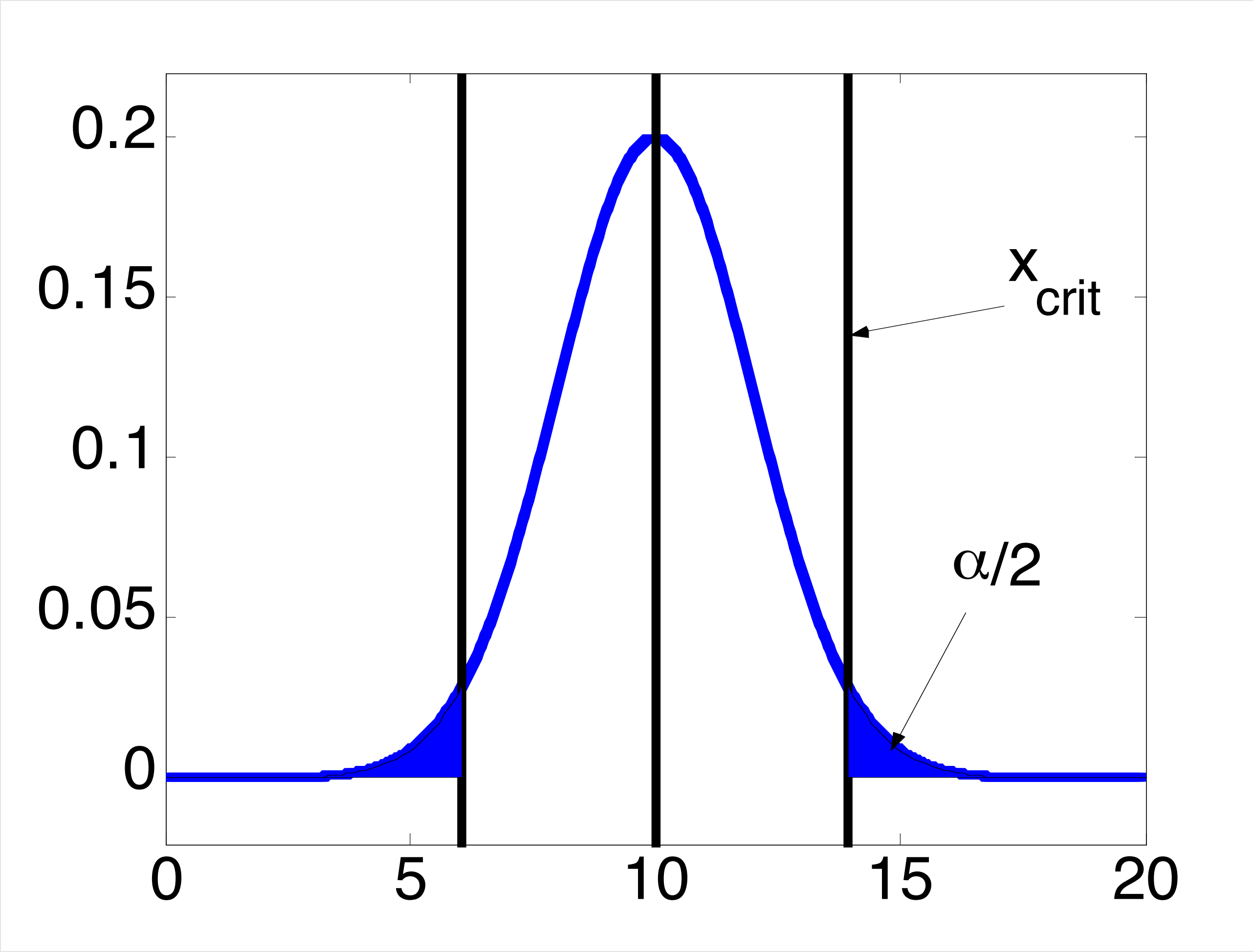
- There are three PDF’s on the ANGEL site in the simulation/monte carlo folder.
- These provide some very deep insight into thinking about power.

Readings for Thursday Non-parametric resampling (bootstrap and permutations)

- Moore (this is a very introductory and gentle introduction if needed).
- Crawley (R Book): 284, 287, 418-421
- Fox_appendix, Crowley

Statistical power

Critical values



t-distribution

4 Possible Outcomes

	Reject null	Accept null
Null correct	Type I error (α)	Correct ($1-\alpha$)
Null wrong	Correct ($1-\beta$)	Type II error (β)

- ($1 - \beta$) is power - probability of detecting a true difference
- ($1 - \alpha$) is confidence - probability of correctly accepting the null

Purpose of Power

- *A priori*:
  - How big does my experiment need to be?
  - What can I test with this experiment?

- *Post hoc*: if we fail to reject the null hypothesis then want to differentiate between:
    - Our sample size was too small/variance is very large making false rejection of large effect
    - The effect is nonexistent or real but trivially small
    - Confidence interval's give you a lot (maybe all) of this information.

Components of Power

- Five parameters
  - Effect size (d or Δ_x)
  - Sample size (n)
  - Variance (s^2)
  - α (type I error) - can be either 1- or 2-tailed
  - β (type II error) - always evaluated as 1-tailed
- Power increases with increased:
  - Effect size
  - Sample size
  - α

  But decreased:
  - Variance

- Power (1-β, ranges from 0 to 1, close to 1 is good)

  > 0.8 is generally considered to be high

A Priori Power Analysis

- Given β, variance and α, what n is needed?
  - If I want to have a certain level of power how big of an experiment do I need to plan?
- Given d, α and n, what is β?
  - I can afford this experiment, what's my prob. of rejecting the null when it is in fact false?

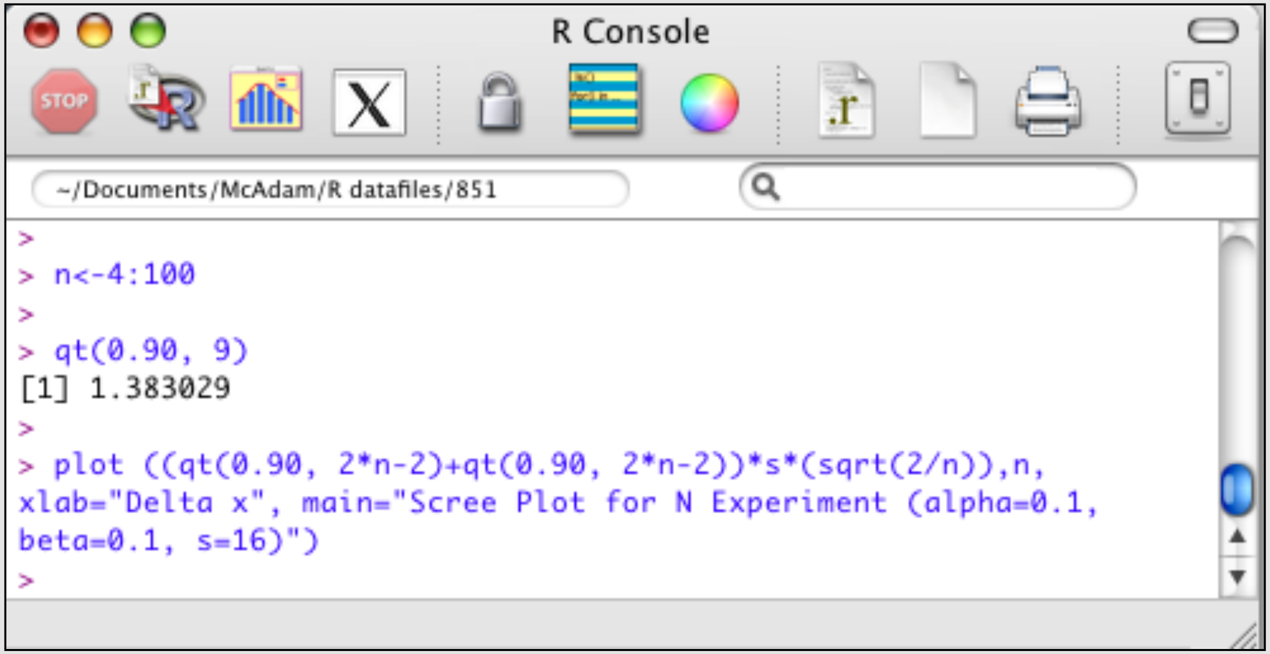
Scree Plots

- Useful to plot relationship between n and detectable effect size d
- Non-linear
- Area of large returns
- Area of diminishing return

Using R to Generate Scree Plots.

$$t_\alpha = \frac{\Delta_x}{s\sqrt{2/n}} \qquad \Delta_x = (t_\alpha + t_{1-\beta}) * s\sqrt{2/n}$$

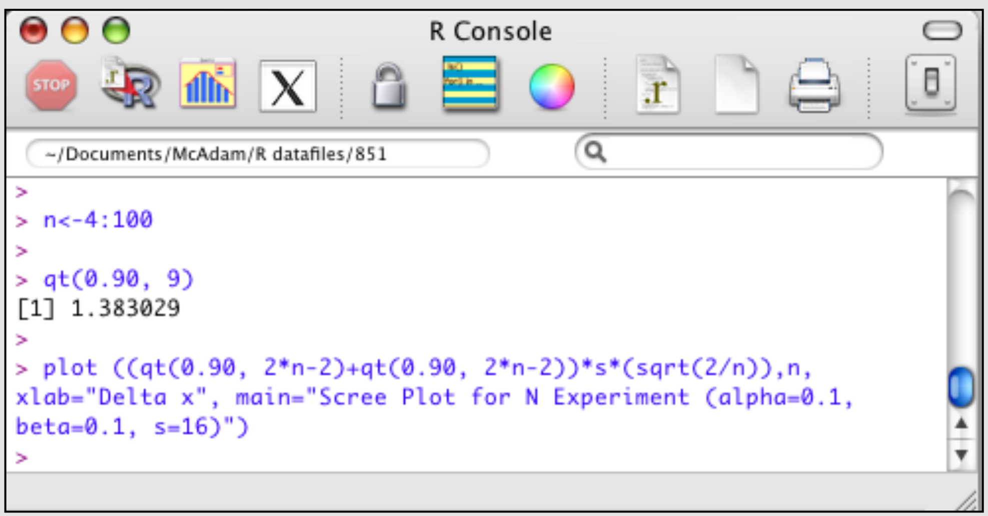
Recall from N example that s = 16, α = 0.10 (one-tailed), 1-β = 0.90

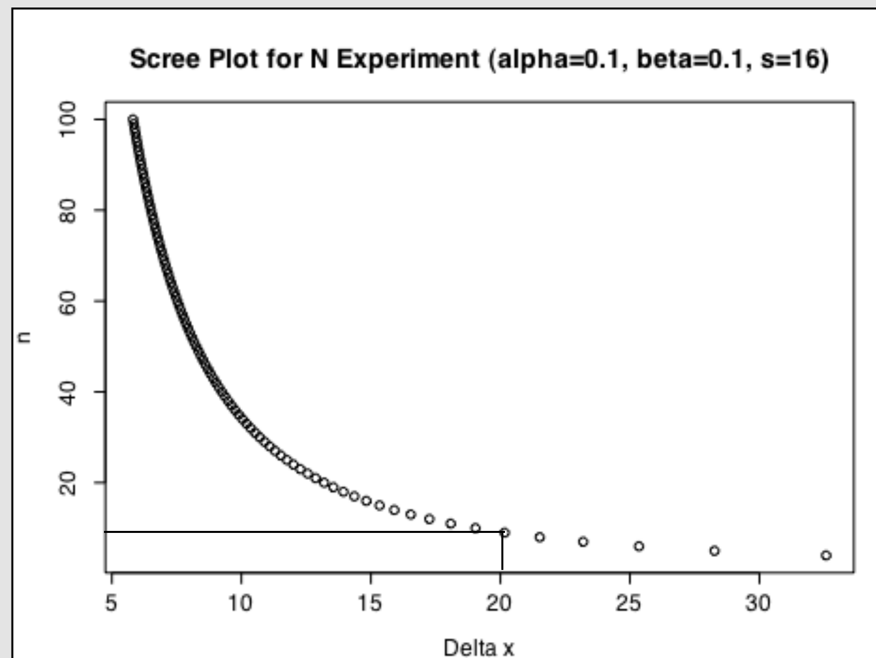


Using R to Generate Scree Plots.

$$t_\alpha = \frac{\Delta_x}{s\sqrt{2/n}} \qquad \Delta_x = (t_\alpha + t_{1-\beta}) * s\sqrt{2/n}$$

Recall from N example that s = 16, α = 0.10 (one-tailed), 1-β = 0.90





A Posteriori (Retrospective) Power Analysis

- Reviewers often mistakenly ask for and authors often mistakenly provide an estimate of power for a non-significant analysis already performed "Observed Power".
  - This is not useful!
  - This approach will always yield low power (< 0.5)
  - This is just a fancy way of stating the P value!

A Posteriori Power Analysis

$$t_\alpha = \frac{\Delta_x}{s/\sqrt{n}}$$

$$t_\alpha + t_{1-\beta} = \frac{\Delta_x}{s/\sqrt{n}}$$

When effect size, sd and n are fixed then t_α and hence P are related negatively to power

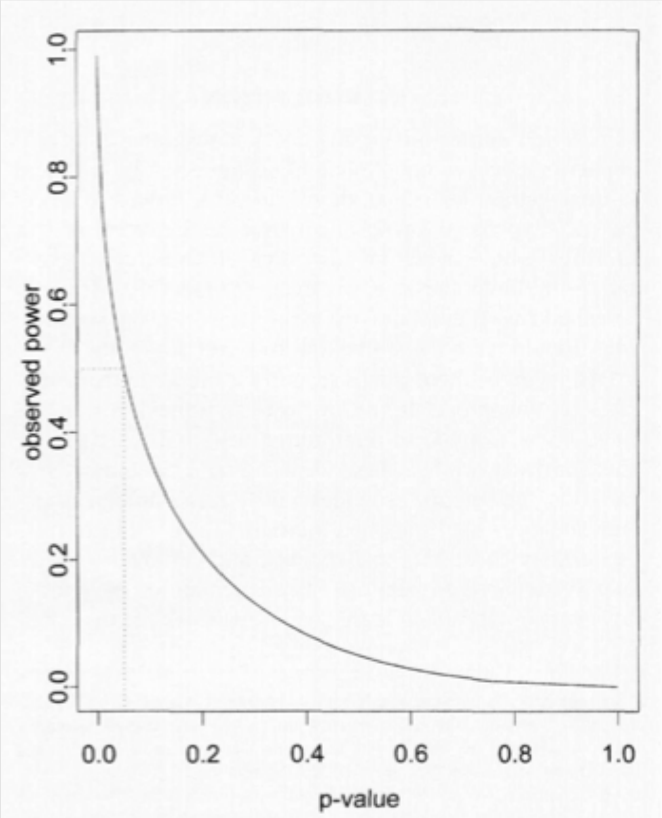


Figure 1. "Observed" Power as a Function of the p Value for a One-Tailed Z Test in Which α is Set to .05. When a test is marginally significant (P = .05) the estimated power is 50%.

- from Hoenig & Heisey, 2001

A Posteriori Power Analysis

- When are *a posteriori* power analyses useful?
  - Given β, n, what effect size (d) can be detected?
  - Could my experiment reject the null hypothesis in interesting cases?
- Report n, effect sizes, variances, β or α that were used.
- Most interesting questions can be answered with effect sizes and consideration of biological significance
- Best to always think of power analyses as prospective
  - Retrospective analyses are just prospective for future studies.

if you're interested in figuring out whether the variation is so high that you need larger samples to truly capture the effect sizes;
or
if you're studying hyena clans with a fixed clan size, can extrapolate and determine how large the clan would need to be before you can detect effects

Confidence Intervals

- In general an idea that captures power is the confidence interval
  - High power = small confidence interval
- Gives idea of range of possible effect sizes
  - May or may not include null hypothesis and biologically important values
- 95% CI will capture the true parameter 95% of the time
    - Subtle difference from 95% chance that the true value lies within CI

- Calculating Confidence Intervals
  - Based on two-tailed alpha, and beta = 0.5
  - Mean +/- $t_{0.025,\ df}$ * se
  - Presented as (lower 95% CL, upper 95% CL)
    e.g. (-1.005, 0.034)

Power

- Power is the probability of rejecting the null when it is false
    - Increased by sample size, effect size, α
    - Decreased by variance
- *A Priori* power analyses are essential
- *A Posteriori* power analyses are often asked for, but are of limited use
- Confidence intervals incorporate much of what we want to achieve with analysis post-mortem and biological significance

Given 4 components, can calculate the 5th

Example: Test of whether a standard is exceeded

$$t_\alpha = \frac{\Delta_x}{s/\sqrt{n}} \qquad t_\alpha + t_{1-\beta} = \frac{\Delta_x}{s/\sqrt{n}} \qquad \hat{n} = (t_\alpha + t_{1-\beta})^2 \frac{s^2}{\Delta_x{}^2}$$

Defaults to $\beta = 0.5$

Note that $t_{0.5} = 0$

Example: planning an experiment of N on plant biomass. How many samples are needed?

- Pilot study or literature indicates that average biomass is ~ 103 kg/ha and sd = 16 kg/ha
- Biological expertise and costs of type I and II error used to determine:
  - α = 0.10 (one-tailed)
  - β = 0.10 (power = 0.90)
  - Biological effect = 20%
- How many samples needed to achieve 90% power with an effect size of a 20.6 kg/ha <u>increase</u> (20% of 103 kg/ha)?

Example: planning an experiment of N on plant biomass. How many samples are needed?

$$t_\alpha = \frac{\Delta_x}{s\sqrt{2/n}}$$

$$\hat{n} = 2(t_\alpha + t_{1-\beta})^2 \frac{s^2}{\Delta_x^{\ 2}}$$

- $s^2 = 256$
- $\Delta_x^2 = 424.36$
- n is size of each group
- In this case $t_{0.10,\ df} = t_{0.90,\ df}$, but a function of n
- df = 2n-2
- Start by assuming t for a large n ($t_{0.1,\ df=large} = 1.28$)

$n^* = 2*(1.28 + 1.28)^2 * 0.60$
$= 7.86$
$t_{0.10,\ 14} = 1.35$

$n^{**} = 2*(1.35 + 1.35)^2 * 0.60$
$= 8.75$
$t_{0.10,\ 16} = 1.34$

$n^{***} = 2*(1.34 + 1.34)^2 * 0.60$
$= 8.62$

Converges on 8.7

Power analysis

- For all but the simplest models there is no closed form for power analysis.
- For most complex models, we need to rely on…..

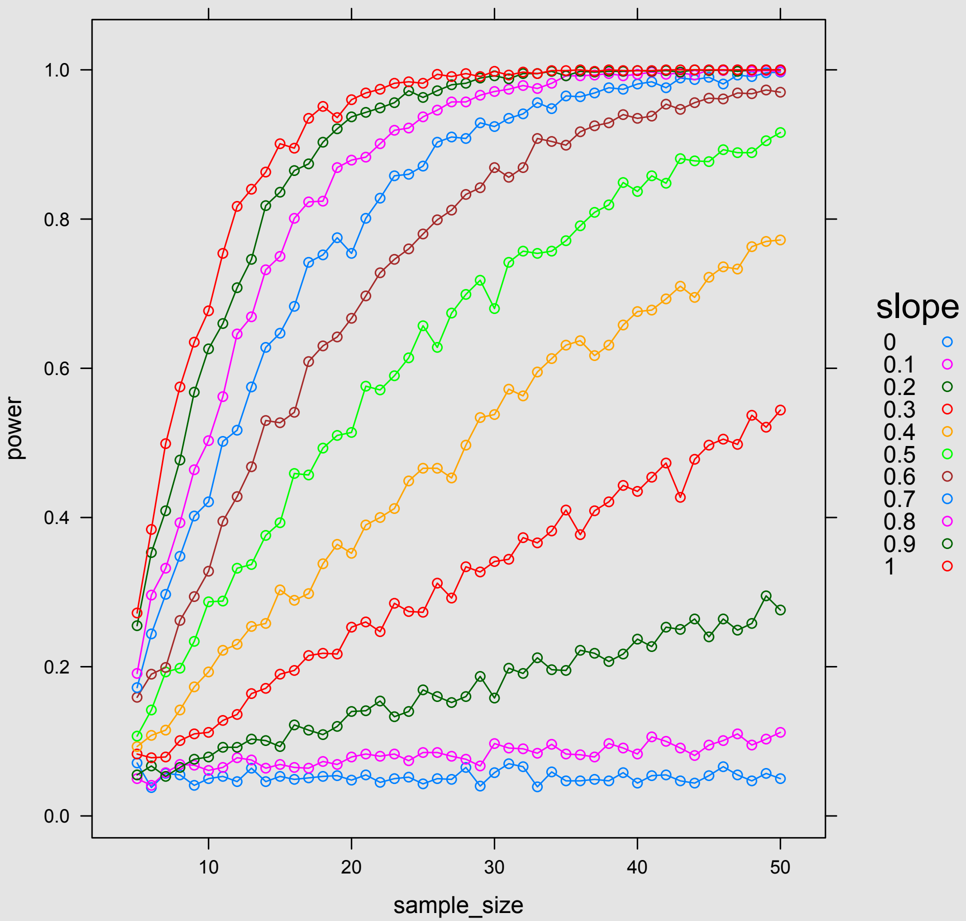
Monte Carlo simulation..

- How would you use monte carlo simulations to perform a power analysis?

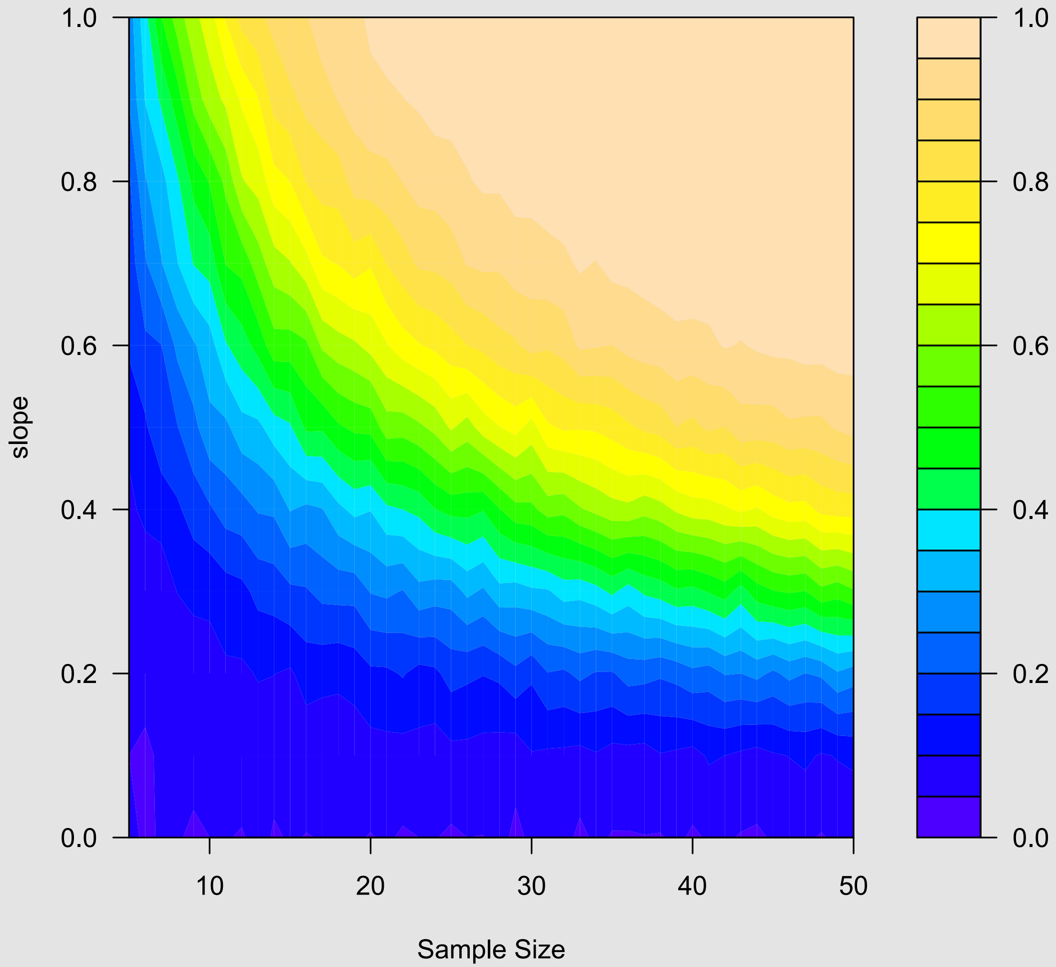
set of models for several parameters (ß1, ß2, etc), have variance, CI, sample size. aka, what are the biologically plausible parameters.

Steps for Monte Carlo simulations for Power

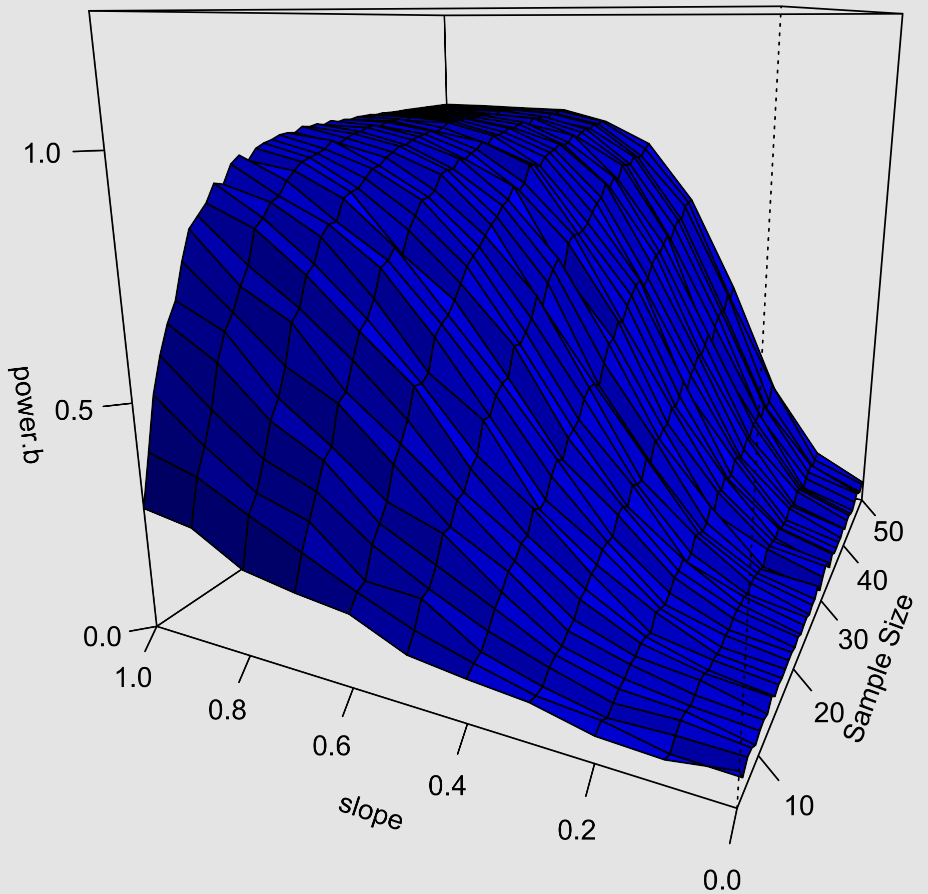
Simulations for power analyses



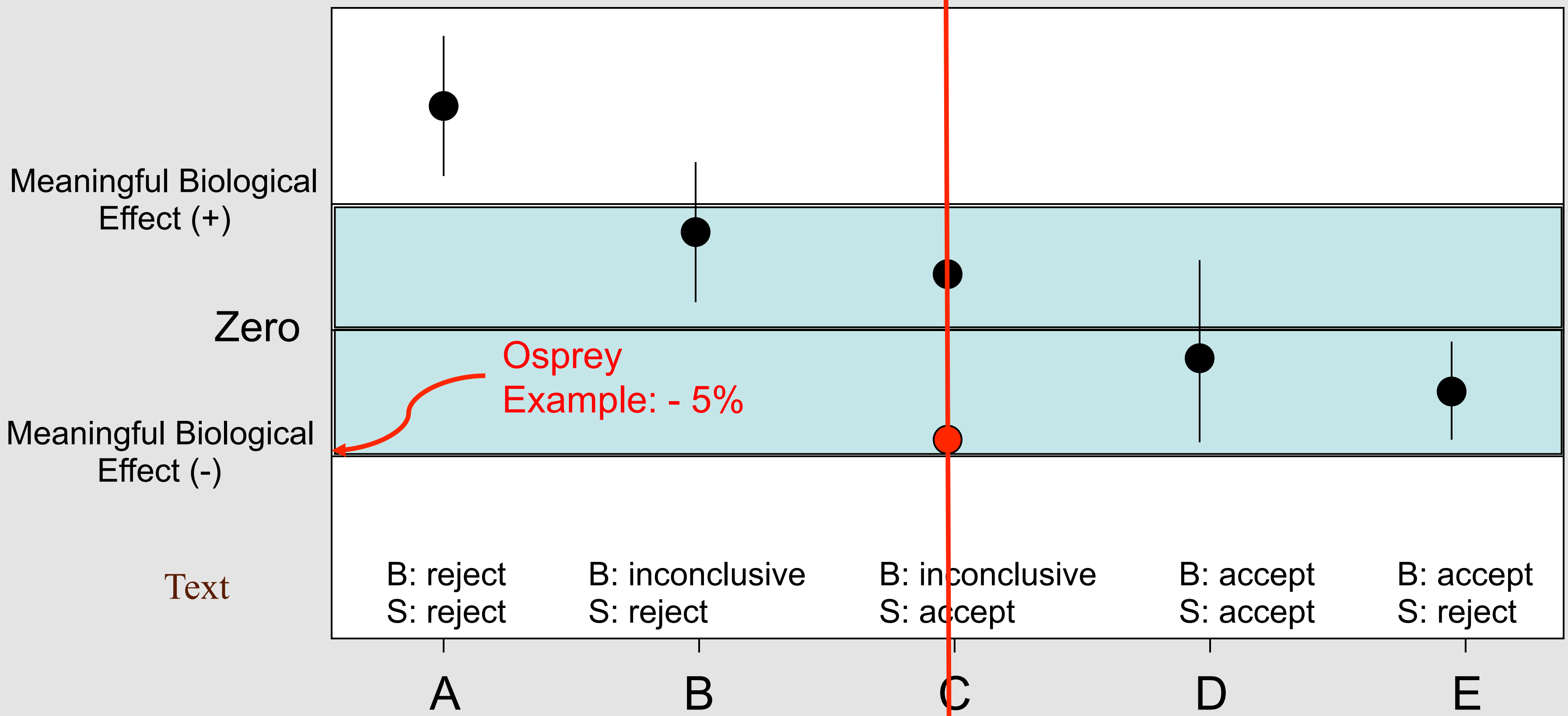
Simulations for power analyses



Simulations for power analyses



Statistical and Biological Decisions



- after Steidl and Thomas, 2001