

Supplementary Material for README: Mistakes of Significance

Nathan (Nat) Goodman

June 12, 2019

Introduction

This document contains supplementary material for the [README document](#) associated with [this repo](#). It's not meant to be read as a standalone document. It's more of a "usage test" demonstrating that all major functions work in the limited setting of this document. It proved it's worth in this regard, driving out several bugs and leading to software improvements.

The document exercises all capabilities of the program: simulation and other data generation, plotting of figures, and saving of summary results. It has figures showing each plot function applied to each kind of data that can be sensibly displayed using that function. Some figures are included for completeness even though they may not be terribly informative. It also illustrates custom plots from the blog posts.

The figures as displayed in the document are tiny - drawn at 25% scale - to reduce screen real estate. The underlying image data is full scale, however, and readers can expand the figures if desired.

The simulations driving the figures too small to yield stable results - 10^3 iterations vs. 10^4 or 10^5 in the blog posts. This explains why some of the figures are so jumpy.

Overview

Nomenclature

in code	in text	meaning
<code>n</code>	n	sample size
<code>d.pop</code>	d_{pop}	population effect size
<code>d.sdz</code>	d_{sdz}	standardized observed effect size, aka Cohen's d
<code>d.het</code>	d_{het}	mean of heterogeneous population distribution
<code>d0</code>	d_0	"center" of non-central t-distribution, more precisely, the non-centrality parameter (ncp) expressed in d units; equivalently, the presumed true effect size for which this is the sampling distribution
<code>d</code>	d	variously means either d_{pop} , d_{sdz} , or d_{het} ; hopefully clear from context
<code>sd.het</code>	sd_{het}	standard deviation of heterogeneous population distribution
<code>meand</code>	$meand$	mean significant observed effect size
<code>pval</code>	$pval$	p-value
<code>power</code>	$power$	power
<code>ci</code>	ci	confidence interval
<code>simu</code>	$simu$	statistic computed from simulated data
<code>theo</code>	$theo$	statistic computed analytically

The program implements two data generation models.

1. *fixed effect (fix)* which imagines we're repeating a study an infinite number of times, drawing a different sample each time from a population with a fixed true effect size
2. *heterogeneous effect (het)* which assumes that each time we do the study, we're sampling from a different population with a different true effect size

The program explores the models with three types of simulations.

1. *fixd*. Fixes d_{pop} to a few values of interest, sets n to a range of values, and simulates many studies for each d_{pop} and n .
2. *rand*. Randomly selects a large number of d_{pop} s, sets n to a range of values, and simulates one study for each d_{pop} and n .
3. *hetd*. Sets d_{het} and sd_{het} to a few values of interest, and selects a large number of d_{pop} s from normal distributions with mean d_{het} and standard deviation sd_{het} . Then sets n to a range of values, and simulates one study for each d_{pop} and n .

The program calculates four statistics across the data generation models and simulation types.

1. *meand* (mean significant effect size)
2. *p-values*
3. *power*
4. *confidence intervals*

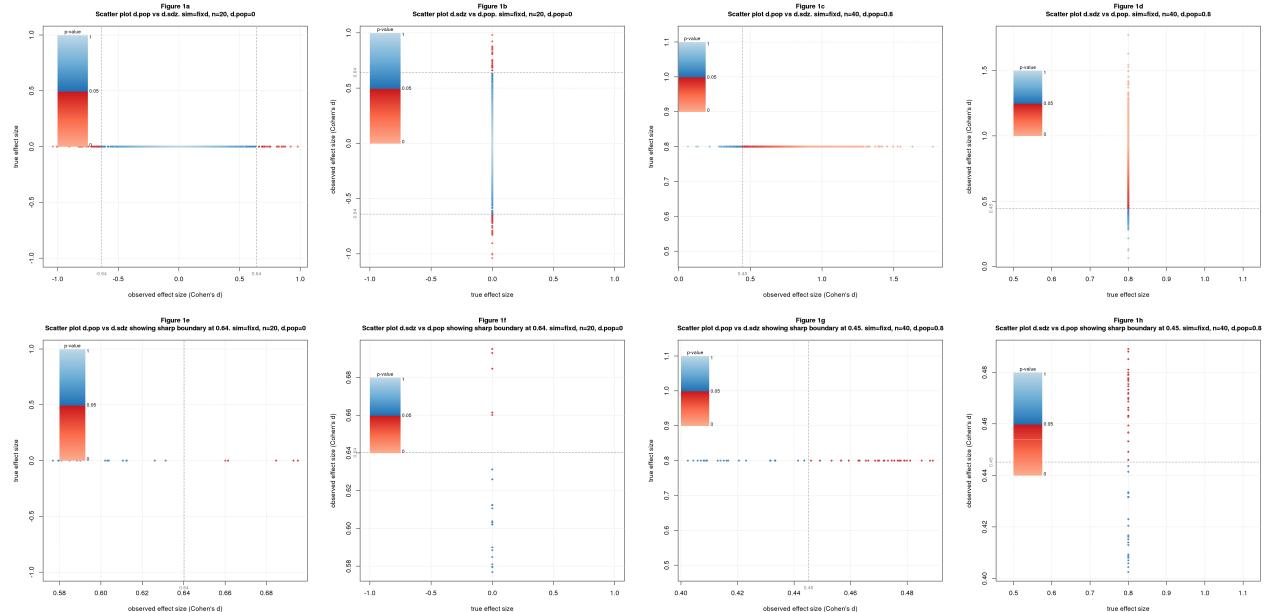
plotdvsd - Plot one effect size vs another

The first group of figures are scatter plots colored by p-value of d_{pop} vs d_{sdz} and vice versa for all three simulation types. The program plots results for two parameter sets and has full scale and zoomed-in figures. The zoomed-in figures illustrate the sharp boundary between significant and non-significant results.

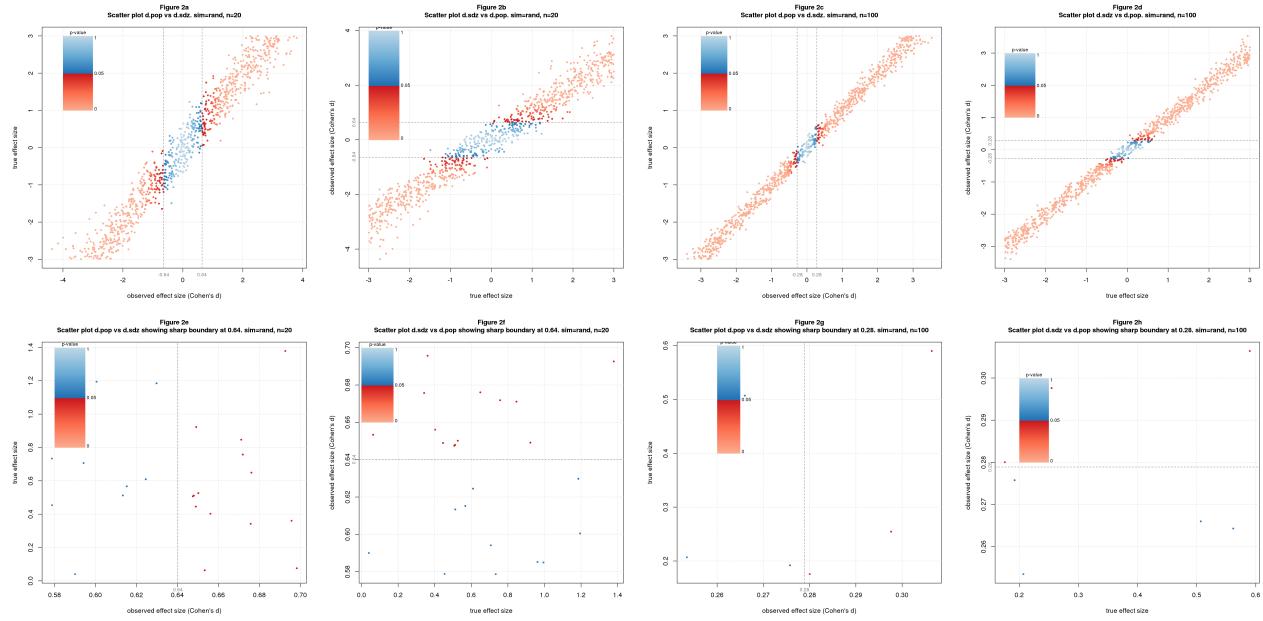
Technical note: The sharpness of the boundary is due to the use of Cohen's d in conjunction with the t-test. This pairing is mathematically natural because both are *standardized*, meaning both are relative to the sample standard deviation. In fact, Cohen's d and the t-statistic are essentially the same statistic, related by the identities $d = t\sqrt{2/n}$ and $t = d\sqrt{n/2}$ (for my simulation scenario).

There are blocks of eight figures for each type of simulation. The *fixd* figures aren't very informative since there's no d_{pop} spread; I include them for completeness only. You'll notice the often-poor placement of the p-value legend; this is a limitation of the program.

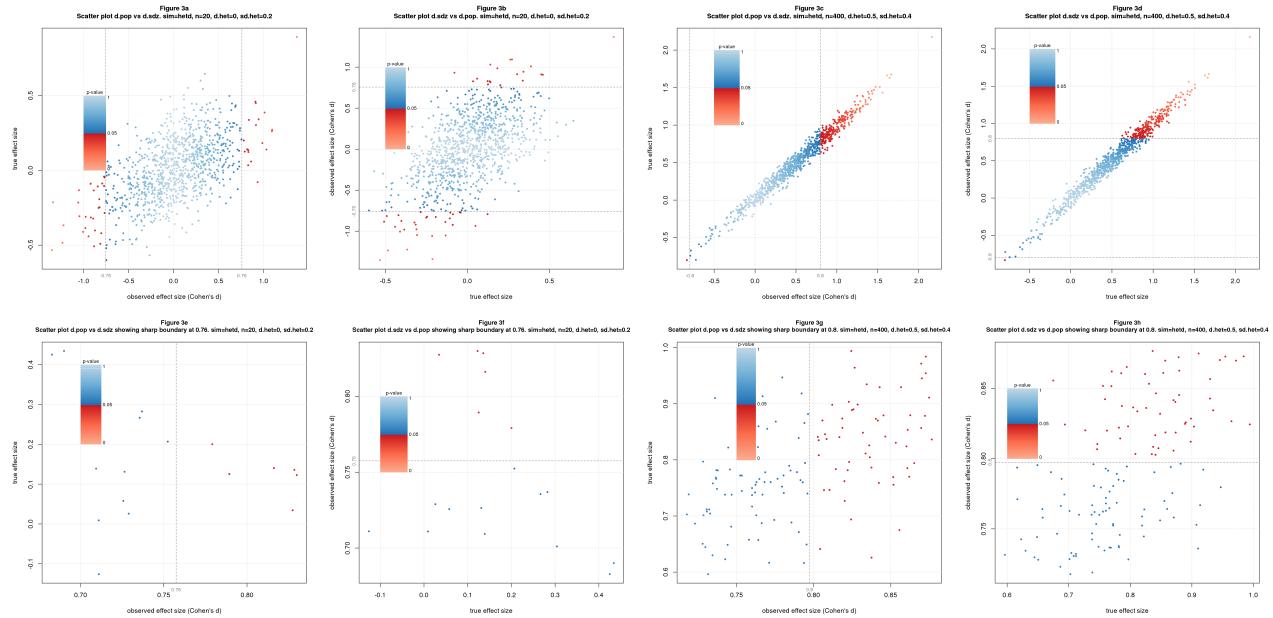
fixd simulation



rand simulation



hetd simulation

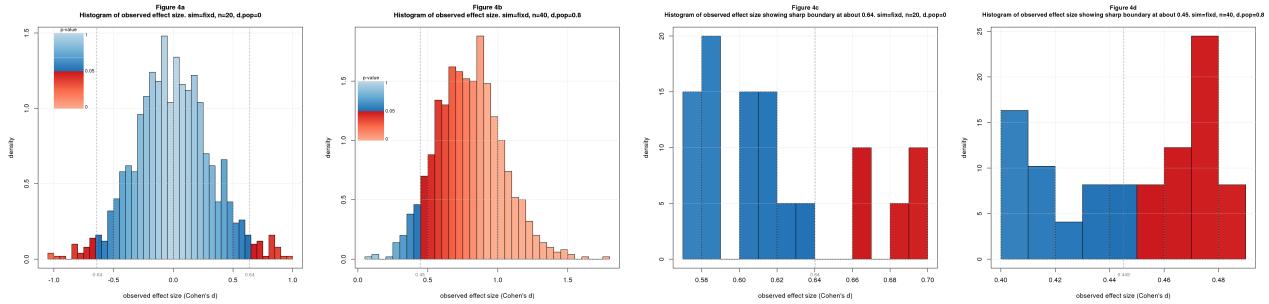


plothist - Plot histogram of simulation results

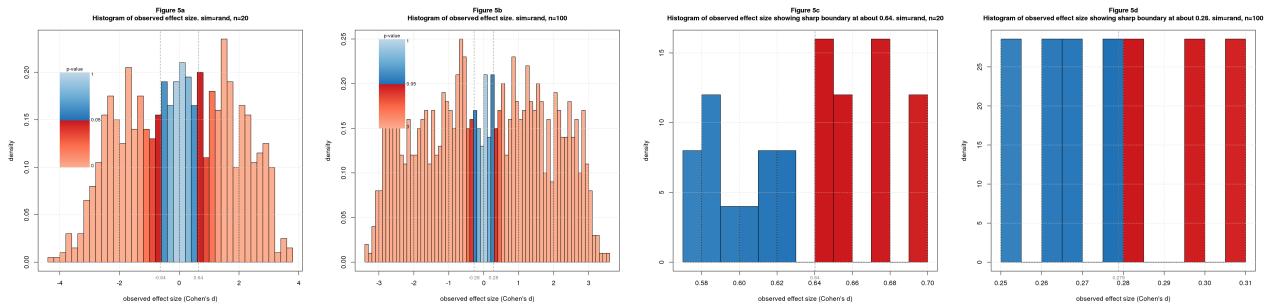
The next group are histograms colored by p-value of d_{sdz} for all three simulation types. The program plots results for two parameter sets and has full scale and zoomed-in figures. The zoomed-in figures again illustrate the sharp boundary between significant and non-significant results.

There are blocks of four figures for each type of simulation. As noted in the preceding section, the p-value legend is poorly placed in some figures, reflecting a limitation of the program.

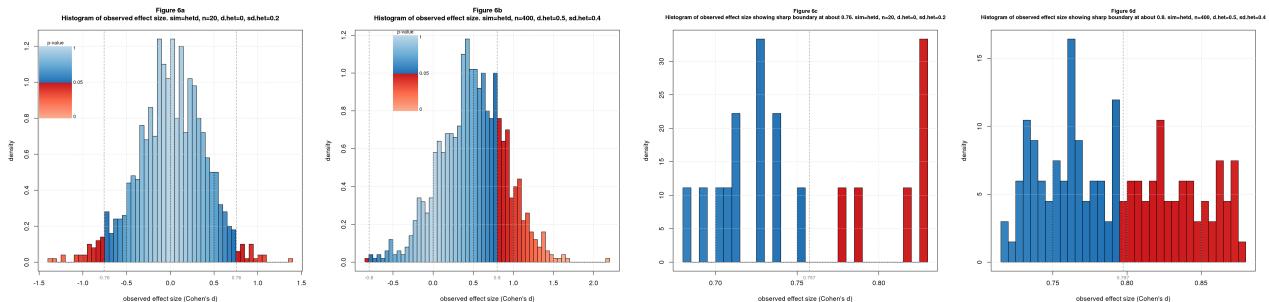
fix simulation



rand simulation



hetd simulation

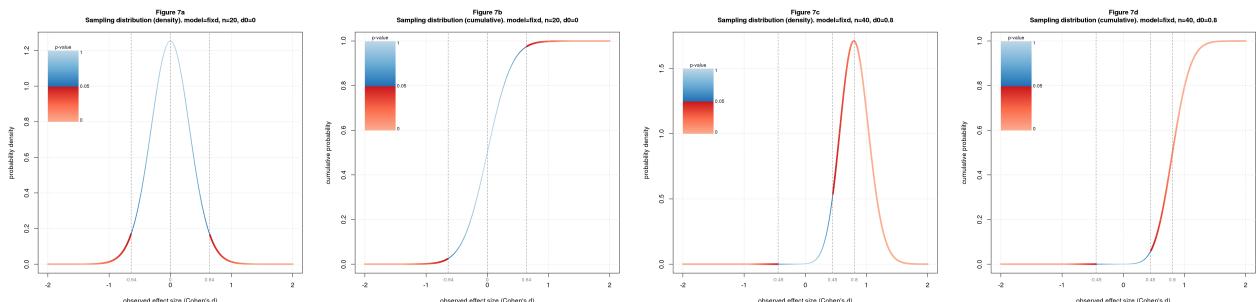


plotpvsd - Plot sampling distribution vs. observed effect size

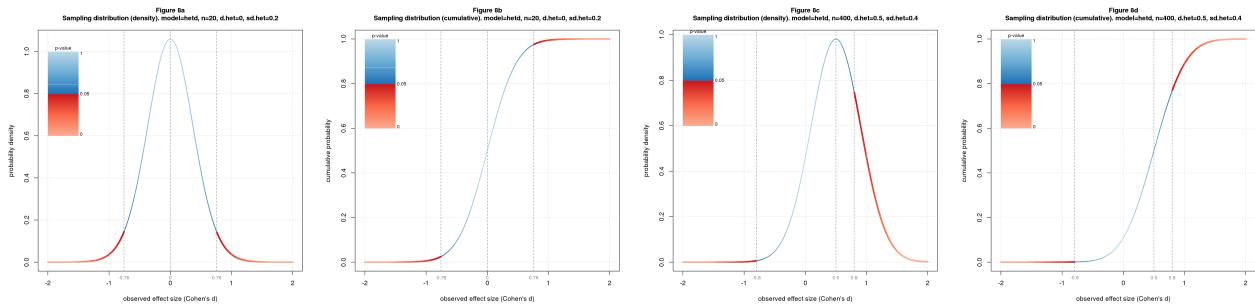
Next are sampling distributions colored by p-value for the two data generation models. The program plots results for two parameter sets and has figures showing probability density and cumulative probability.

There are four figures for each type of simulation. The p-value legend is reasonably placed in these figure, but the sizes don't quite match; this is again a limitation of the program.

fixed effect model



heterogeneous effect model

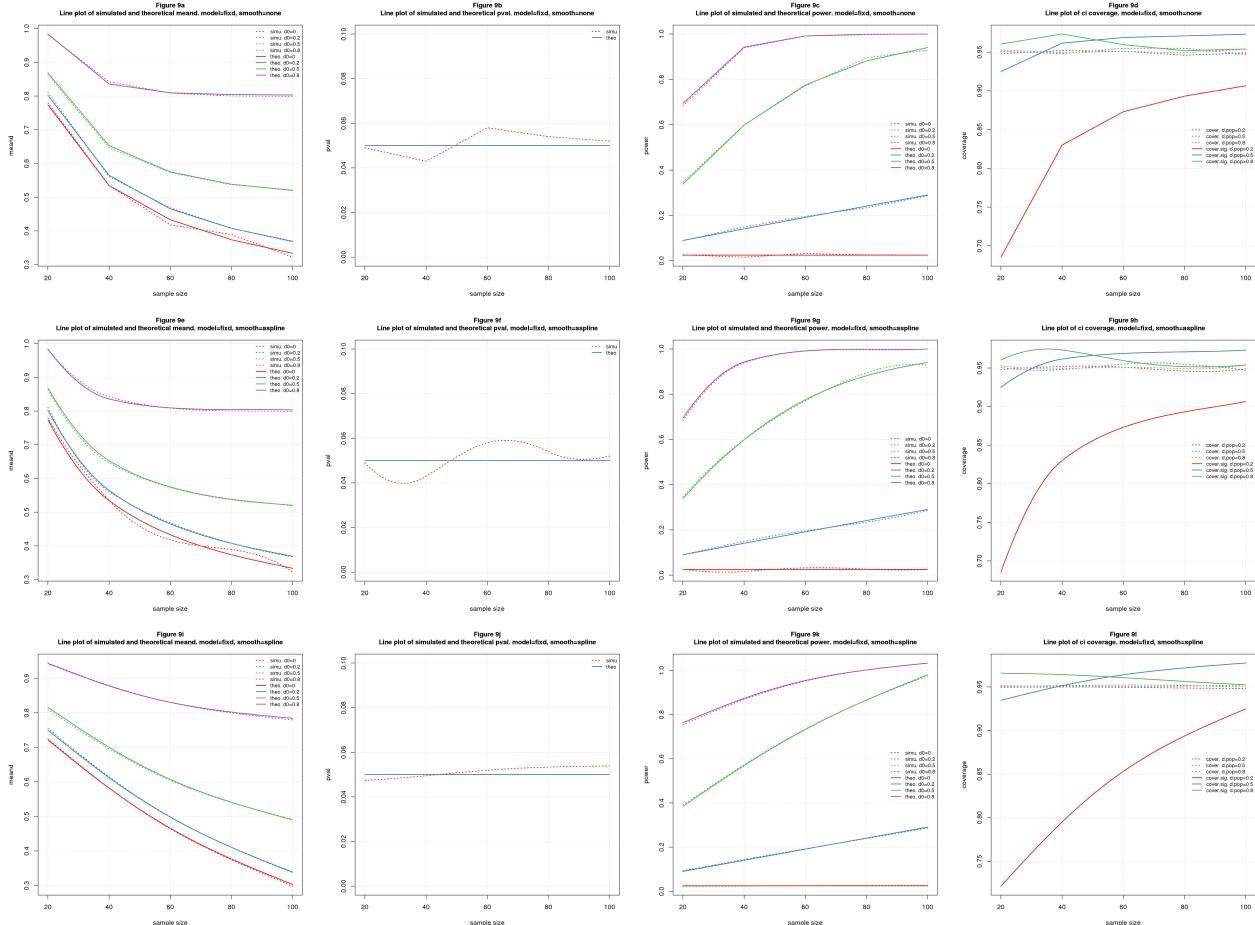


plotm - Plot multiple lines - wrapper for R's matplot

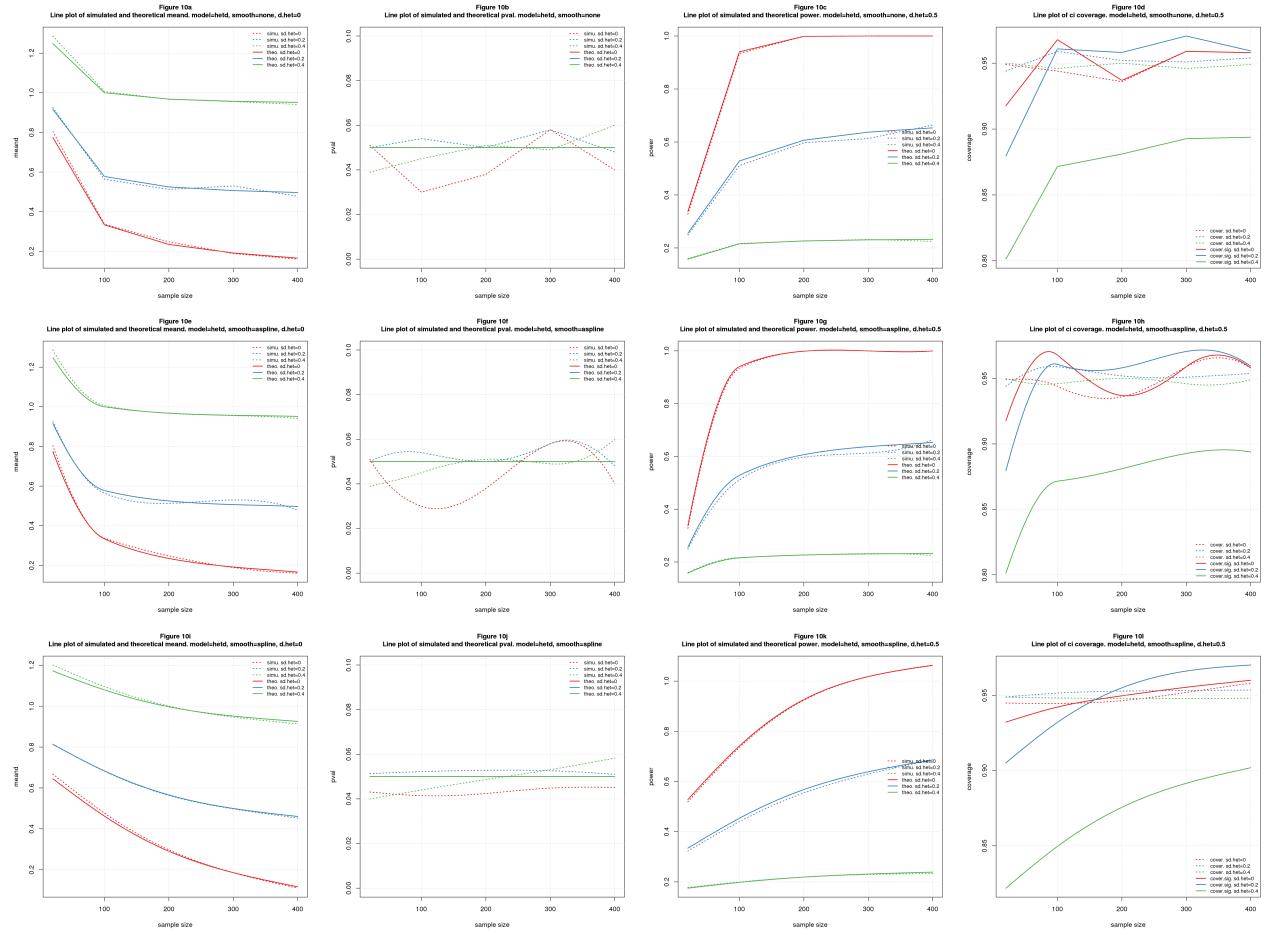
The next figures are line plots for each of the four statistics across the two data generation models. For *meand*, *p-values*, and *power*, the figures show concordance of simulated and theoretical results. For *confidence intervals*, the figures show coverage (the fraction of intervals that contain the true mean, d_{pop}) for all results and significant results. For each statistic, there are figures using three different smoothing methods:

1. none 2 *aspline* is from the *akima* package
2. *spline* is base R's *smooth.spline*

fixed effect model



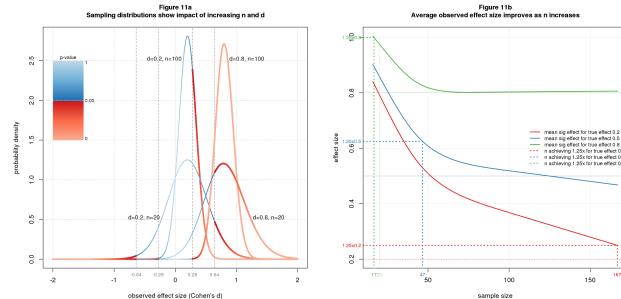
heterogeneous effect model



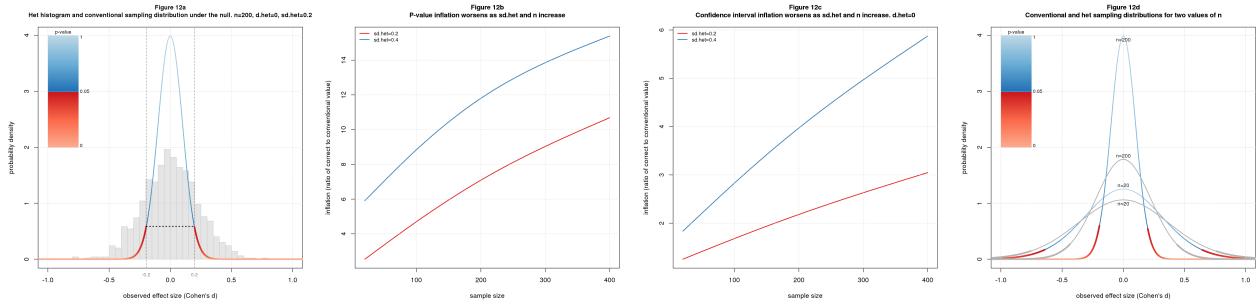
Custom plots from blog posts

The final group are figures adapted from the blog posts associated with this repo.

[from ovrfx blog post - When You Select Significant Findings, You're Selecting Inflated Estimates](#)



[from ovrht blog post - Your P-Values are Too Small! And So Are Your Confidence Intervals!](#)



Tables

The program creates files of supporting data in two formats: flat table and R list. The flat table is below. The document omits the list version as there's no easy way to display in Rmarkdown.

what	d.crit	md20_0.2	md20_0.5	md20_0.8	ov20_0.2	ov20_0.5	ov20_0.8	nov_0.2	nov_0.5	nov_0.8
meand.fixd	0.64	0.80	0.85	0.98	4.00	1.71	1.22	161.58	45.85	16.00
meand.d2t	0.64	0.79	0.86	0.98	3.97	1.72	1.22	166.70	46.64	16.00

Comments Please!

Please post comments on [Twitter](#) or [Facebook](#), or contact me by email natg@shore.net.

Please report bugs, other software problems, and feature requests using the [GitHub Issue Tracker](#). I will be notified, and you'll be apprised of progress. As already noted, the software is still rough and software documentation nonexistent.

Copyright & License

Copyright (c) 2019 Nathan Goodman

The software is **open source and free**, released under the [MIT License](#). The documentation is **open access**, released under the [Creative Commons Attribution 4.0 International License](#).